Ninth Valencia International Meeting on Bayesian Statistics

2010 ISBA World Meeting

Co-sponsored by the *Universitat de València* (Spain), and by the *International Society for Bayesian Analysis* (ISBA), the *9th Valencia International Meeting on Bayesian Statistics* and the *2010 World Meeting of the International Society for Bayesian Analysis* are being jointly held in Benidorm (Alicante, Spain) from Thursday June 3rd to Tuesday June 8th, both inclusive. This is preceded by a half-day set of postgraduate tutorials on the morning of Thursday June 3rd.

The scientific programme includes 24 lectures invited by the Valencia 9 Scientific Committee, whose discussion will be initiated by an invited discussant, 40 plenary talks selected through a blinded review process by the ISBA 2010 Program Committee, and about 350 contributed papers to be presented in poster form in 5 evening plenary sessions. This booklet contains the complete joint programme, the abstracts of all contributions (in alphabetical order by first named author), the list of participants, and an abbreviated version of the Bayesian singalong book.

Social Programme

The opening ceremony will take place in the Auditorium of Gran Hotel Bali, Benidorm, at 09h15 of Friday June 4th, 2010. The ISBA General Body Meeting will be on afternoon of Sunday June 6th. All members are invited to attend for discussion of ISBA business including planning and decisions for future World Meetings. The conference banquet will be held on the evening of Tuesday June 8th. This will be followed by the traditional Valencia meeting cabaret. The seaside location of the venue provides an ideal leisure setting for accompanying persons during the conference working hours.

Proceedings

Bayesian Statistics 9, the Valencia 9 proceedings, will be published by Oxford University Press. This will include the invited papers and their discussions. By June 28th 2010, the invited discussants and all delegates wishing to participate in the discussion should send their written discussions to the author(s) of the invited papers, and to the local organiser. Contributions should not exceed *six* typeset pages (including figures) for invited discussions, and *three* pages for contributed discussions. By August 23d, 2010, just after the summer vacation, authors of invited papers should have sent the local organizer a final manuscript, together with a rejoinder to the discussion of their paper. The final version should not exceed *twenty four* typeset pages (including figures and references), and the rejoinder should not exceed *six*. The LaTeX source of any accepted contribution to the discussion, separate **eps** postscript files of all figures, and the **pdf** files of the output (to be used for control), should be sent by e-mail to **valenciameeting@uv.es**. All contributions are expected to use the purpose-built LaTeX Valencia macros, which may be downloaded from the conference website **www.uv.es/valenciameeting**. Typeset proofs will be sent to authors as they become available. We expect the proceedings to be finalized and sent to the printers by November 2010.

Bayesian Analysis (**ba.stat.cmu.edu**) will publish selected contributed papers presented at the World Meeting (as talks or posters). Submissions are subject to the standard refereeing process of Bayesian Analysis. Please make sure to indicate in your cover letter that the paper was presented at ISBA 2010. All such papers received by August 16 will be eligible for the Lindley prize (**www.bayesian.org/awards/LindleyPrize.html**).

Acknowlegements

We are grateful to the Universitat de València for its continuous support, and to the US agencies National Science Foundation (NSF), National Institutes of Health (NIH) and Office of Naval Research Global (ONRG), for their partial funding.

Welcome to Spain!

José M. Bernardo

Thursday, June 3rd (Morning)

Postgraduate Tutorial Seminars

Chair: **Schmidt, Alexandra** (Univ. Federal do Rio de Janeiro, Brazil)

09h00–10h20: ISBA 2010 Tutorial 1.

- **Petrone, Sonia*** (Università Bocconi, Italy)
  Introduction to Bayesian inference

10h25-11h45: ISBA 2010 Tutorial 2.

- **Mena, Ramsés*** (Universidad Nacional Autonóma de México, Mexico)
  Some topics on Bayesian nonparametrics and mixture models

11h50-13h10: ISBA 2010 Tutorial 3.

- **Inoue, Lurdes*** (University of Washington, USA)
  A quick tour to the principles and approaches of decision theory

13h10–17h00: *Break*

Thursday, June 3rd (Afternoon)

ISBA 2010 Invited Programme

17h00-18h40: ISBA 2010 Session 1.
Stochastic models and stochastic Monte Carlo

Chair: **Lee, Herbie** (University of California, Santa Cruz, USA)

- **Miguez, Joaquin*** (Universidad Carlos III, Spain)
  Crisan, Dan (Imperial College London, UK)
  Djurić, Petar (Stony Brook University, USA)
  Solving a class of global optimization problems by way of Bayesian estimation methods

- **Gonçalves, Flávio*** (University of Warwick, UK)
  Roberts, Gareth (University of Warwick, UK)
  Exact simulation and Bayesian Inference for jump-diffusion processes

- **Beskos, Alexandros** (University College London, UK)
  Mattingly, Jonathan (Duke University, USA)
  Pillai, Natesh (University of Warwick, UK)
  Roberts, Gareth* (University of Warwick, UK)
  Sanz, Jesús (Universidad de Valladolid, Spain)
  Stuart, Andrew (University of Warwick, UK)
  MCMC in high dimensions: A new perspective

18h40–19h10: *Coffee Break*

Thursday, June 3rd (Evening)

19h10-20h50: ISBA 2010 Session 2.
Biostatistics

Chair: **Lurdes, Inoue** (University of Washington, USA)

- **Savelieva, Marina*** (Novartis, Switzeland)
  Demin, Ivan (Novartis, Switzeland)
  Neuenschwander, Beat (Novartis, Switzeland)
  Bayesian approach to population PKPD modelling: advantages and drawbacks

- **Krnjajic, Milovan*** (National University of Ireland, Galway, Ireland)
  Kottas, Athanasios (University of California, Santa Cruz, USA)
  Bayesian flexible models for censored data

- **Draper, David*** (University of California, Santa Cruz, USA)
  A Bayesian decision-theoretic alternative to standard multiple-comparisons adjustments in clinical trials

- **Geneletti, Sara*** (London School of Economics, UK)
  Best, Nicky (Imperial College London, UK)
  Toledano, Mireille (Imperial College London, UK)
  Elliott, Paul (Imperial College London, UK)
  Richardson, Sylvia (Imperial College London, UK)
  Uncovering selection bias in case control studies using Bayesian poststratification

21h00–22h00: *Dinner*

22h00-01h00: *Poster Session 1.*

Mainly posters whose first author family name is in [A-C]. See later.

Friday, June 4th, Morning

09h15-09h40: **Opening ceremony**

Valencia 9 Invited Programme

09h45-11h15: Valencia 9 Session 1.
Uncertainty quantification and optimization

Chair: **French, Simon** (University of Manchester, UK)

- **Goldstein, Michael*** (Durham University, UK)
  External Bayesian analysis for computer simulators

  Discussant: **Paulo, Rui** (Universidade Técnica de Lisboa, Portugal)

- **Gramacy, Robert B.** (University of Cambridge, UK)
  Lee, Herbie* (University of California, Santa Cruz, USA)
  Optimization under unknown constraints

  Discussant: **Holmes, Christopher** (University of Oxford, UK)

11h15–11h45: *Coffee Break*

11h45-13h15: Valencia 9 Session 2.
Applications

Chair: **Migon, Helio** (Universidade Federal do Rio de Janeiro, Brazil)

- **Tebaldi, Claudia***
  (Climate Central, USA and University of British Columbia, Canada)
  Smith, Richard L. (University of North Carolina, Chapel Hill, USA)
  Sansó, Bruno (University of California, Santa Cruz, USA)
  Characterizing uncertainty of future climate change projections using hierarchical Bayesian models

  Discussant: **Ferreira, Marco** (University of Missouri, USA)

- **Madigan, David*** (Columbia University, USA)
  Bayesian methods in pharmacovigilance

  Discussant: **DuMouchel, William**. (Phase Forward Inc., USA)

13h15–17h00: *Break*

Friday, June 4th (Afternoon)

ISBA 2010 Invited Programme

17h00-18h40: ISBA 2010 Session 3.
Bioinformatics

Chair: **Christen, Andrés** (CIMAT, Mexico)

- **Rancoita, Paola M.V.*** (IDSIA, Switzerland)
  Hutter, Marcus (Australia National University, Australia)
  Bertoni, Francesco (IOSI, Switzerland)
  Kwee, Ivo (IOSI, Switzerland)
  Bayesian integrated genomics

- **Jensen, Shane*** (The Wharton School, USA)
  Braunstein, Alexander (The Wharton School, USA)
  McAuliffe, Jon (The Wharton School, USA)
  Bayesian modeling of the evolutionary escape response of HIV

- **Yau, Christopher*** (University of Oxford, UK)
  Holmes, Christopher (University of Oxford, UK)
  Decision theoretic Bayesian nonparametric inference for the molecular characterisation and stratification of colorectal cancer using genome-wide microarrays

- **Lenarcic, Alan*** (Harvard University, USA)
  Airoldi., Edo (Harvard University, USA)
  A Bayesian application of lasso in covariance network selection

18h40–19h10: *Coffee Break*

Friday, June 4th (Evening)

19h10-20h50: ISBA 2010 Session 4.
Decision theory and prior specification

Chair: **Clyde, Merlise** (Duke University, USA)

- **Mohamed, Shakir*** (University of Cambridge, UK)
  Heller, Katherine (University of Cambridge, UK)
  Ghahramani, Zoubin (University of Cambridge, UK)
  Sparse exponential family latent variable models

- **Hans, Chris*** (The Ohio State University, USA)
  Penalized regression via orthant normal priors

- **Hahn, Richard*** (Duke University, USA)
  Carvalho, Carlos (Univ. Chicago and Univ. Texas at Austin, USA)
  Mukherjee, Sayan (Duke University, USA)
  Sparse partial factor regression

- **Lacoste-Julien, Simon*** (University of Cambridge, UK)
  Ghahramani, Zoubin (University of Cambridge, UK)
  Approximate inference for the loss-calibrated Bayesian

21h00–22h00: *Dinner*

22h00-01h00: *Poster Session 2.*

Mainly posters whose first author family name is in [D-J]. See later.

Saturday, June 5th

Valencia 9 Invited Programme

09h00-11h15: Valencia 9 Session 3.
Computation

Chair: **Robert , Chris** (Université Paris- Dauphine, France)

- **McCallum, Andrew*** (University of Massachusetts, USA)
  Probabilistic programming with imperatively-defined factor graphs

  Discussant: **Ghahramani, Zoubin** (University of Cambridge, UK)

- **Huber, Mark*** (Claremont McKenna College, USA)
  Schott, Sarah (Duke University, USA)
  Using TPA for Bayesian inference

  Discussant: **Roberts, Gareth** (University of Warwick, UK)

- **Lopes, Hedibert*** (University of Chicago, USA)
  Carvalho, Carlos (Univ. Chicago and Univ. Texas at Austin, USA)
  Johannes, Michael (Columbia University, USA)
  Polson, Nicholas (University of Chicago, USA)
  Particle learning for sequential Bayesian computation

  Discussant: **Pitt, Michael** (University of Warwick, UK)

11h15–11h45: *Coffee Break*

11h45-13h15: Valencia 9 Session 4.
Sparsity modelling

Chair: **Armero, Carmen** (Universitat de València, Spain)

- **Polson, Nicholas*** (University of Chicago, USA)
  Scott, James (University of Texas at Austin, USA)
  Shrink globally, act locally: Sparse Bayesian regularization and prediction

  Discussant: **Clarke, Bertrand** (University of Miami, USA)

- **Richardson, Sylvia** (Imperial College London, UK)
  Bottolo, Leonardo* (Imperial College London, UK)
  Rosenthal, Jeffrey (University of Toronto, Canada)
  Bayesian models for sparse regression analysis of high dimensional data

  Discussant: **Mallick, Bani** (Texas A&M University, USA)

13h15–17h00: *Break*

Saturday, June 5th (Afternoon)

ISBA 2010 Invited Programme

17h00-18h40: ISBA 2010 Session 5.
Regression Models

Chair: **Gamerman, Dani** (Universidade Federal do Rio de Janeiro, Brazil)

- **Ghosh, Sujit*** (North Carolina State University, USA)
  Wang, Jiangdian (North Carolina State University, USA)
  Bayesian shape restricted regression with multivariate Bernstein polynomials

- **Cabras, Stefano** (Università di Cagliari, Italy)
  Castellanos, María E. (Universidad Rey Juan Carlos, Spain)
  Quirós, Alicia* (Universidad Rey Juan Carlos, Spain)
  Assessing the fit of regression models for multiple imputation

- **Bhattacharya, Anirban*** (Duke University, USA)
  Dunson, David (Duke University, USA)
  Sparse Bayesian infinite factor models

- **Astle, William*** (Imperial College London, UK)
  de Iorio, María (Imperial College London, UK)
  Ebbels, Timothy (Imperial College London, UK)
  Richardson, Sylvia (Imperial College London, UK)
  A Bayesian model of NMR spectra for the deconvolution and quantification of metabolites in complex biological mixtures

18h40–19h10: *Coffee Break*

Saturday, June 5th (Evening)

19h10-20h50: ISBA 2010 Session 6.
Statistical Computing

Chair: **Ferreira, Marco** (University of Missouri, USA)

- **Holmes, Christopher*** (University of Oxford, UK)
  Doucet, Arnaud (University of British Columbia, Canada)
  Lee, Anthony (University of Oxford, UK)
  Giles, Michael (University of Oxford, UK)
  Yau, Christopher* (University of Oxford, UK)
  Bayesian computation on graphics cards

- **Haslett, John** (Trinity College Dublin, Ireland)
  Parnell, Andrew* (University College Dublin, Ireland)
  Fast joint posterior modelling through marginal posterior mixtures

- **Wauthier, Fabian*** (University of California, Berkeley, USA)
  Jordan, Michael (University of California, Berkeley, USA)
  Sparse process classification via the Gaussian copula

- **Atchadé, Yves** (University of Michigan, USA)
  Wang, Jing* (University of Michigan, USA)
  Approximate MCMC simulation from doubly-intractable distributions

21h00–22h00: *Dinner*

22h00-01h00: *Poster Session 3.*

Mainly posters whose first author family name is in [K-Na]. See later.

Sunday, June 6th, Morning

Valencia 9 Invited Programme

09h00-11h15: Valencia 9 Session 5.
Nonparametrics

Chair: **Wago, Hagime** (Kyoto Sangyo University, Japan)

- **Bhattacharya, Abhishek** (Duke University, USA)
  Dunson, David* (Duke University, USA)
  Nonparametric Bayes classification and testing on manifolds

  Discussant: **Griffin, James** (University of Kent, UK)

- **Ickstadt, Katja*; Bornkamp, Björn**
  (Technische Universität Dortmund, Germany)
  Grzegorczyk, Marco; Wieczorek, Jakob
  (Technische Universität Dortmund, Germany)
  Sheriff, Rahuman; Grecco, Hernán; Zamir, Eli
  (Max-Planck-Institut für Molekulare Physiologie, Dortmund, Germany)
  Nonparametric Bayesian networks

  Discussant: **Jordan, Michael** (University of California, Berkeley, USA)

- **Loredo, Tom*** (Cornell University, USA)
  Rotating stars and revolving planets: Bayesian exploration of the pulsating sky

  Discussant: **Müller, Peter** (MD Anderson Cancer Center, Texas, USA)

11h15–11h45: *Coffee Break*

11h45-13h15: Valencia 9 Session 6.
Time series and spatial modelling

Chair: **Peña, Daniel** (Universidad Carlos III, Spain)

- **Carvalho, Carlos*** (Univ. Chicago and Univ. Texas at Austin, USA)
  Lopes, Hedibert (University of Chicago, USA)
  Aguilar, Omar (Financial Engines, USA)
  Dynamic stock selection strategies: A structured factor model framework

  Discussant: **Mendoza, Manuel** (ITAM, Mexico)

- **Schmidt, Alexandra*** (Universidade Federal do Rio de Janeiro, Brazil)
  Rodríguez, Marco (Université du Québec à Trois-Rivières, Canada)
  Modelling multivariate counts varying continuously in space

  Discussant: **Boys, Richard** (University of Newcastle, UK)

13h15–17h00: *Break*

Sunday, June 6th (Afternoon and Evening)

ISBA 2010 Invited Programme

17h00-18h40: ISBA 2010 Session 7.
Spatial and Temporal Modeling

Chair: **Pruenster, Igor** (Università di Torino, Italy)

- **Pati, Debdeep*** (Duke University, USA)
  Reich, Brian (North Carolina State University, USA)
  Dunson, David (Duke University, USA)
  Bayesian geostatistical modeling with informative sampling locations

- **Paciorek, Chris*** (University of California, Berkeley, USA)
  Flexible spatial latent variable modeling for combining information sources while accounting for systematic errors in proxies

- **Almeida, Carlos*** (Technische Universität München, Germany)
  Czado, Claudia (Technische Universität München, Germany)
  Bayesian inference for time-varying pair-copula constructions

- **Manolopoulou, Ioanna*** (Duke University, USA)
  Kepler, Thomas B. (Duke University, USA)
  West, Mike (Duke University, USA)
  Dynamic spatial modelling in inhomogeneous force fields

18h40–19h10: *Coffee Break*

19h10–20h50: *ISBA General Body Meeting*

21h00–22h00: *Dinner*

22h00-01h00: *Poster Session 4.*

Mainly posters whose first author family name is in [Ne-So]. See later.

Monday, June 7th, Morning

Valencia 9 Invited Programme

09h00-11h15: Valencia 9 Session 7.
Modelling and causality

Chair: **Makov, Udi** (University of Haifa, Israel)

- **Chopin, Nicolas*** (ENSAE, France)
  Jacob, P. (ENSAE, France)
  Free energy sequential Monte Carlo, application to mixture modelling

  Discussant: **Green, Peter** (University of Bristol, UK)

- **Wexler, Ydo** (Microsoft Research, USA)
  Meek, Christopher* (Microsoft Research, USA)
  Improved approximate sum-product inference using multiplicative error bounds

  Discussant: **Mira, Antonietta** (Università dell'Insubria, Italy and Università della Svizzera italiana, Switzerland)

- **Richardson, Thomas*** (University of Washington, USA)
  Evans, Robin (University of Washington, USA)
  Robins, James (Harvard School of Public Health, USA)
  Transparent parametrizations of models for potential outcomes

  Discussant: **Fienberg, Stephen** (Carnegie-Mellon University, USA)

11h15–11h45: *Coffee Break*

11h45-13h15: Valencia 9 Session 8.
Modelling for applications in molecular biology

Chair: **Moreno, Elías** (Universidad de Granada, Spain)

- **Stingo, Francesco** (Rice University, USA)
  Vannucci, Marina* (Rice University, USA)
  Bayesian models for variable selection that incorporate biological information

  Discussant: **Berzuini, Carlo** (University of Cambridge, UK)

- **Wilkinson, Darren*** (University of Newcastle, UK)
  Parameter inference for stochastic kinetic models of bacterial gene regulation: a Bayesian approach to systems biology

  Discussant: **Kou, Samuel** (Harvard University, USA)

13h15–17h00: *Break*

Monday, June 7th (Afternoon)

ISBA 2010 Invited Programme

17h00-18h40: ISBA 2010 Session 8.
Nonparametrics

Chair: **Petrone, Sonia** (Università Bocconi, Italy)

- **Teh, Yee-Whye*** (University College London, UK)
  Wood, Frank (Columbia University, USA)
  Gasthaus, Jan (University College London, UK)
  Archambeau, Cédric (Xerox Research Centre Europe, France)
  James, Lancelot (University of Science and Technology, Hong Kong)
  Hierarchical Bayesian nonparametric models for language and text

- **Guha, Subharup*** (University of Missouri, USA)
  Posterior simulation in countable mixture models for large datasets

- **Kolossiatis, Michalis*** (Cyprus University of Technology, Cyprus)
  Griffin, Jim (University of Kent, UK)
  Steel, Mark (University of Warwick, UK)
  Modelling via normalisation for Bayesian nonparametric inference

- **Huszar, Ferenc*** (University of Cambridge, UK)
  Bayesian kernel machines: the third way of going nonparametric

18h40–19h10: *Coffee Break*

Monday, June 7th (Evening)

19h10-20h50: ISBA 2010 Session 9.

Chair: **Rousseau, Judith** (Université Paris Dauphine, France)

- **Airoldi, Edoardo*** (Harvard University, USA)
  Representation and Bayesian analysis of integer-valued networks
- **Mogapi, Ometse** (University of Botswana, Botswana)
  Snijders, Tom (Oxford University, UK)
  Nicholls, Geoff* (Oxford University, UK)
  Silverman, Bernard (Oxford University, UK)
  Johnson, David (Oxford University, UK)
  Karn, Nicholas (University of Southampton, UK)
  Bayesian inference for a partial order from random linear extensions
- **Henao, Ricardo*** (Danmarks Tekniske Universitet, Denmark)
  Winther, Ole (Københavns Universitet, Denmark)
  Learning structure in directed acyclic graphs with latent variables
- **McCormick, Tyler*** (Columbia University, USA)
  Zheng, Tian (Columbia University, USA)
  Latent structure models for social networks using aggregated relational data

21h00–22h00: *Dinner*

22h00-01h00: *Poster Session 5.*

Mainly posters whose first author family name is in [Sp-Z]. See later.

Tuesday, June 8th, Morning

Valencia 9 Invited Programme

09h00-11h15: Valencia 9 Session 9.
Model choice and ranking problems

Chair: **Sun, Dongchu** (University of Missouri, USA)

- **Consonni, Guido*** (Università di Pavia, Italy)
  La Rocca, Luca (Università di Modena e Reggio Emilia, Italy)
  On moment priors for Bayesian model choice with applications to directed acyclic graphs

  Discussant: **Smith, James Q.** (University of Warwick, UK)

- **Louis, Tom*** (Johns Hopkins University, USA)
  Li, Qing (NIH-National Human Genome Research Institute, USA)
  Carvalho, Benilton (University of Cambridge, UK)
  Fallin, Danielle (Johns Hopkins University, USA)
  Irizarry, Rafael (Johns Hopkins University, USA)
  Ruczinski, Ingo (Johns Hopkins University, USA)
  Association tests that accommodate genotyping uncertainty

  Discussant: **Dukic, Vanja** (University of Chicago, USA)

- **Frühwirth-Schnatter, Sylvia*** (J. Kepler Universität Linz, Austria)
  Bayesian variable selection for random intercept modeling of Gaussian and non-Gaussian data

  Discussant: **Brown, Philip** (University of Kent, UK)

11h15–11h45: *Coffee Break*

11h45-13h15: Valencia 9 Session 10.
Foundations

Chair: **Ruggeri, Fabrizio** (CNR IMATI, Italy)

- **Meng, Xiao-Li*** (Harvard University, USA)
  What's the H in H-likelihood: A Holy Grail or an Achilles' Heel?

  Discussant: **George, Edward** (University of Pennsylvania, USA)

- **Bernardo, José M.*** (Universitat de València, Spain)
  Integrated objective Bayesian estimation and hypothesis testing

  Discussant: **Pericchi, Luis**
  (Universidad de Puerto Rico, Rio Piedras, Puerto Rico)

13h15–17h00: *Break*

Tuesday, June 8th (Afternoon, Evening and Night)

ISBA 2010 Invited Programme

17h00-18h40: ISBA 2010 Session 10.
L.J. Savage award finalists

Chair: **Vanucci, Marina** (Rice, University, USA)

- **Fox, Emily*** (Duke University, USA)
  Bayesian nonparametric time series models for complex dynamic phenomena
- **Taddy, Matt*** (Chicago Booth USA)
  Dynamic point process modeling with a DDP
- **Adams, Ryan*** (University of Toronto, Canada)
  Generative modeling of probability densities with Gaussian processes
- **Scott, James*** , (University of Texas at Austin, USA)
  Robust Bayesian shrinkage in sparse signal-extraction problems

Conclusion

22h00–23h30: *Gala Dinner*

Keynote Speaker:
Dawid, A. Philip (University of Cambridge, UK)

23h30–Sunrise: *The Valencia 9 Cabaret and Dance Party*

Master of Ceremonies:
Reese, Shane (Brigham Young University, USA)
Musical Director:
Carlin, Brad (University of Minnesota, USA)

Contributed Papers

Poster Session 1

Date and time: **Thursday June 3rd, 22h00**
Coordinator: **Schmidt, Alexandra** (UF do Rio de Janeiro, Brazil)

- **Abellán, Juan-José** (CSISP, Valencia, Spain)
  Coscollá, Mireia (UPF-CSIC Barcelona, Spain)
  González-Candelas, Fernando (CIBERESP, Spain)
  Bayesian analysis of the Phylogeography of bacteria L. Pneumophila in the Valencian region of Spain.

- **Adhikari, Kaustubh** (Harvard University, USA)
  Lange, Christoph (Harvard University, USA)
  Is it rare or common? A coalescent tree-based Bayesian approach to identify the genetic types of variants underlying complex diseases.

- **Akerkar, Rupali** (Norges Teknisk-Natur. Univ, Norway)
  Approximate Bayesian Inference for non-homogeneous Poisson processes with application to survival analysis

- **Aktekin, Tevfik** (The George Washington University, USA)
  Soyer, Refik (Georgia Southwestern State University, USA)
  Xu, Feng (Georgia Southwestern State University, USA)
  Bayesian state space modeling of mortgage default risk

- **Alkema, Leontine** (National University of Singapore, Singapore)
  Raftery, Adrian (University of Washington, USA)
  Gerland, Patrick (United Nations Population Division, USA)
  Clark, Samuel (University of Washington, USA)
  Pelletier, François (United Nations Population Division, USA)
  Buettner, Thomas (United Nations Population Division, USA)
  Probabilistic projections of the total fertility rate using a Bayesian hierarchical model

- **Alston, Clair** (Queensland University of Technology, Australia)
  Kerrie Mengersen (Queensland University of Technology, Australia)
  Bayesian mixture models for monitoring reef health

- **Altuzarra, Alfredo** (Universidad de Zaragoza, Spain)
  Gargallo, Pilar (Universidad de Zaragoza, Spain)
  Bayesian Ahp-priorities in a global context

- **Amorós, Rubén** (Universitat de València, Spain)
  Conesa, David (Universitat de València, Spain)
  López-Quílez, Antonio (Universitat de València, Spain)
  Martínez-Beneito, Miguel A. (CSISP, Generalitat Valenciana, Spain)
  Poisson model with a hidden Markov structure for the detection of influenza outbreaks

- **Anacleto-Junior, Osvaldo** (Open University, UK)
  Queen, Catriona (Open University, UK)
  Bayesian forecasting models for traffic management systems

- **Ancelet, Sophie** (Met@risk unit, France)
  Abellan, Juan José (CIBERESP, Spain)
  Bayesian shared spatial-component models to strengthen inference in disease mapping from multiple disease-related surveillance sources

- **Andrianakis, I.** (National Oceanography Centre, UK)
  Challenor, P.G. (National Oceanography Centre, UK)
  The effect of the nugget on Gaussian process emulators of computer models

- **Antunes, Marilia** (University of Lisbon, Portugal)
  Sousa, Lisete (University of Lisbon, Portugal)
  EMB: a combination of EM-algorithm and Bayesian classification model for gene identification

- **Arbel, Julyan** (CREST, Université Paris Dauphine, France)
  Bayesian adaptive estimation using a sieve prior

- **Arellano-Valle, Reinaldo B.** (Pontificia Universidad Catolica, Chile)
  Combining Bayesian Test Procedures

- **Argiento Raffaele** (CNR-IMATI, Italy)
  Alessandra Guglielmi (Politecnico di Milano, Italy)
  Bayesian density estimation by mixture models using epsilon-approximations of homogeneus normalized random measures.

- **Aslett, Louis JM** (Trinity College Dublin, Ireland)
  Wilson, Simon P (Trinity College Dublin, Ireland)
  Modelling and inference for networks with repairable redundant subsystems

- **Asseburg, Christian** (Itä-Suomen Yliopisto, Finland)
  Ramsberg, Joakim (I3 Innovus, Sweden)
  Peura, Piia (Itä-Suomen Yliopisto, Finland)
  Henriksson, Martin (Astra Zeneca AB, Sweden)
  Purmonen, Timo (Itä-Suomen Yliopisto, Finland)
  Turunen, Juha (Itä-Suomen Yliopisto, Finland)
  Martikainen, Janne (Itä-Suomen Yliopisto, Finland)
  A Bayesian assessment of decision uncertainty in the choice between triptans for the treatment of migraine in Finland

- **Ausín, Concepción** (Universidad Carlos III de Madrid, Spain)
  Gómez-Villegas, Miguel A. (Univ. Complutense de Madrid, Spain)
  González-Pèrez, Beatriz (Universidad Complutense de Madrid, Spain)
  Rodríguez-Bernal, Mayte (Universidad Complutense de Madrid, Spain)
  Salazar, Isabel (Universidad Complutense de Madrid, Spain)
  Sanz, Luis (Universidad Complutense de Madrid, Spain)
  Bayesian analysis of multiple hypothesis testing with applications to microarray experiments

- **Azadi, Nammam A.** (Kurdistan University of Medical Sciences, Iran)
  Ridall, Gareth (Lancaster University, UK)
  Fearnhead, Paul (Lancaster University, UK)
  Adaptive sequential Monte Carlo for optimal Bayesian experimental design

- **Azevedo Caio** (Universidade Estadual de Campinas, Brazil)
  Bolfarine, Heleno (University of São Paulo, Brazil)
  Andrade, Dalton F. (Federal University of Santa Catarina, Brazil)
  Bayesian inference through a hierarchical framework for an IRT model with centred skew distribution

- **Bacallado, Sergio A.** (Stanford University, USA)
  Pande, Vijay S. (Stanford University, USA)
  Bayesian analysis of reversible, higher-order Markov chains

- **Banks, David** (Duke University, USA)
  Datta, Gauri S. (University of Georgia, USA)
  Heaton, Matt (National Institute of Statistical Science, USA)
  Karr, Alan (University of South Carolina, USA)
  Lynch, James (Clemson University, USA)
  Vera, Francisco (Clemson University, USA)
  Zou, Frank (National Institute of Statistical Science, USA)
  Hierarchical Bayesian modeling in syndromic surveillance

- **Barber, Xavier** (CIO-Miguel Hernandez University, Spain)
  Morales, Javier (University of Valencia, Spain)
  Mayoral, Asuncion (CIO-Miguel Hernandez University, Spain)
  López-Quilez, Antonio (University of Valencia, Spain)
  Barber, Antoni (IDENTIA Institute, Spain)
  A bioclimatic classification method using a conditional corregionalized linear model

- **Barney, Bradley J.** (Texas A M University, USA)
  Johnson, Valen E. (M. D. Anderson Cancer Center, USA)
  Cleeland, Charles S. (M. D. Anderson Cancer Center, USA)
  Post-processing of factor loadings in Bayesian factor analysis

- **Barrantes Ramiro** (University of Vermont, USA)
  Jeffrey Bond (University of Vermont, USA)
  Using shifts in amino acid frequency or substitution rate to identify latent structural characters improves our understanding of the structure, function and evolution of base-excision repair enzymes

- **Beamonte, Asunción** (Universidad de Zaragoza, Spain)
  Alonso, M. Pilar (Universidad de Lleida, España)
  Chakraborty, Avishek (Duke University, USA)
  Gargallo, Pilar (Universidad de Zaragoza, España)
  Gelfand, Alan (Duke University, USA)
  Salvador, Manuel (Universidad de Zaragoza, Spain)
  Delineation of local labor markets. An application to the Spanish region of Aragón

- **Bermúdez, José D.** (Universidad de Valencia, Spain)
  Corberán-Vallet, Ana (Universidad de Valencia, Spain)
  Segura, José V. (Universidad Miguel Hernández de Elche, Spain)
  Vercher, Enriqueta (Universidad de Valencia, Spain)
  Bayesian analysis of a generalized exponential smoothing model

- **Berrett, Candace** (The Ohio State University, USA)
  Calder, Catherine (The Ohio State University, USA)
  Bayesian probit regression models for spatially-dependent categorical data

- **Berrocal, Veronica** (SAMSI, USA)
  Gelfand, Alan (Duke University, USA)
  Holland, David (Environment Protection Agency, USA)
  Fusing space-time data under measurement error for computer model output

- **Berry, Robert** (Sandia National Laboratories, USA)
  Najm, Habib (Sandia National Laboratories, USA)
  Debusschere, Bert (MIT Boston, USA)
  Adalsteinsson, Helgi (MIT Boston, USA)
  Marzouk, Youssef (MIT Boston, USA)
  Data-free inference of the joint distribution of uncertain model parameters

- **Bhattacharya, Arnab** (Trinity College Dublin, Ireland)
  Wilson, Simon (Trinity College Dublin, Ireland)
  Sequential integrated nested Laplace approximation

- **Bickis, Mikelis** (University of Saskatchewan, Canada)
  Empirical calibration of p-values

- **Bliznyu, Nikolay** (Texas A&M University, USA)
  Paciorek, Christopher (University of California at Berkeley, USA)
  Coull, Brent (Harvard School of Public Health)
  Nonlinear latent process models for addressing temporal change of support in spatio-temporal studies of environmental exposures

- **Bolstad, William** (University of Waikato, New Zealand)
  Efficient MCMC using Metropolis-Hastings

- **Bornkamp, Björn** (Technische Univ. Dortmund University, Germany)
  Katja Ickstadt (Technische Univ. Dortmund University, Germany)
  Fitting B-splines to quantiles using a Brier entropy penalty

- **Bortz, David** (University of Colorado, USA)
  A sequential Bayesian experimental design/model selection methodology for validating models of disease pathogenesis

- **Botella, Paloma** (Universidad CEU-Cardenal Herrera, Spain)
  Martínez-Beneito, Miguel-Ángel (CSISP, Spain)
  A kernel-based spatio-temporal surveillance system for monitoring influenza incidence

- **Bouriga, Mathilde** (Université Paris Dauphine, France)
  Féron, Olivier (EDF R&D Clamart, France)
  Marin, Jean-Michel (Université Montpellier, France)
  Robert, Christian (Université Paris Dauphine, France)
  Bayesian estimation of a covariance matrix: Application for asset and liabiliy management

- **Brynjarsdóttir, Jenný** (The Ohio State University, USA)
  Berliner, Mark (The Ohio State University, USA)
  Bayesian hierarchical modeling for paleoclimate reconstruction from geothermal data

- **Byrne, Simon** (University of Cambridge, UK)
  Dawid, A. Philip (University of Cambridge, UK)
  The structural Markov property

- **Caglar, Toros** (The George Washington University, USA)
  Soyer, Refik (The George Washington University, USA)
  Bayesian analysis of Markov modulated discrete time queues

- **Calder, Catherine A.** (The Ohio State University, USA)
  Darnieder, William F. (The Ohio State University, USA)
  Bayesian inference for incomplete marked spatial point patterns

- **Cano, Javier** (Universidad Rey Juan Carlos, Spain)
  Moguerza, Javier M. (Universidad Rey Juan Carlos, Spain)
  Ríos Insua, David (Real Academia de las Ciencias, Spain)
  Bayesian analysis of semimarkovian models with applications to hardware reliability

- **Cano, Juan A.** (Universidad de Murcia, Spain)
  Salmerón, Diego (Consejería de Sanidad de Murcia, Spain)
  Integral priors for Bayesian model selection in practice

- **Carta, Alessandro** (University of Warwick, UK)
  Steel, Mark (University of Warwick, UK)
  Modelling multi-output stochastic frontiers using copulas

- **Cassese, Alberto** (Università di Firenze, Italy)
  Blangiardo, Marta (Imperial College, UK)
  Richardson, Sylvia (Imperial College, UK)
  Synthesizing lists of differentially expressed features in related experiments

- **Chan, Jennifer** (The University of Sydney, Australia)
  Wan, Wai (The University of Sydney, Australia)
  Bayesian approach to analysing longitudinal bivariate binary data with informative dropout

- **Chang, Rui** (University of California, San Diego, USA)
  Bayesian network parameter learning with informative Dirichlet prior

- **Chen, Carla** (Queensland University of Technology, Australia)
  Keith, Jonathan (Queensland University of Technology, Australia)
  Mengersen, Kerrie (Queensland University of Technology, Australia)
  From phenotype to genotype: Reconciling approaches through Bayesian model averaging

- **Chen, Qian** (University of Sydney, Australia)
  Gerlach, Richard (University of Sydney, Australia)
  Lu, Zudi (University of Adelaide, Australia)
  Bayesian value-at-risk and expected shortfall forecasting via the asymmetric Laplace distribution

- **Chen, Zhen** (National Institutes of Health, USA)
  Tadesse, Mahlet (Georgetown University, USA)
  Wang, Kai (University of Pennsylvania, USA)
  Li, Mingyao (University of Pennsylvania, USA)
  An integrated Bayesian hidden Markov model approach to CNVs incorporating pedigree information

- **Chikobvu, Delson** (University of Free State, South Africa)
  van der Merwe, Abrie (University of Free State, South Africa)
  De Beer, Jesse (University of Free State, South Africa)
  A Bayesian approach for comparing the relative returns of securities using the reciprocal of the coefficient of variation

- **Choy, Boris** (The University of Sydney, Australia)
  A modified multivariate Student-t distribution: An application to stochastic volatility models

- **Christen, J. Andrés** (CIMAT, Guanajuato, Mexico)
  Blaauw, Maarten (University Belfast, UK)
  Sample design in the Bayesian analysis of radiocarbon dates in paleoecology

- **Christensen, Anette** (Aarhus University Hospital, Denmark)
  Rasmussen, Lars (Aarhus University Hospital, Denmark)
  Overvad, Kim (Aarhus University Hospital, Denmark)
  Lundbye-Christensen, Søren (Aarhus University Hospital, Denmark)
  Dethlefsen, Claus (Aarhus University Hospital, Denmark)
  A new method in detecting dynamic seasonal variation using state space models

- **Commeau, Natalie** (AFSSA, France)
  Cornu, Marie (Agroparistech, INRA MIA, France)
  Parent, Eric (Agroparistech, INRA MIA, France)
  The need for informative priors in hierarchical models for microbiological control in food industry

- **Corander, Jukka** (Helsingin Yliopisto, Finland)
  Cui, Yaqiong (Helsingin Yliopisto, Finland)
  Koski, Timo (Royal Institute of Technology, Sweden)
  Sirén, Jukka (Helsingin Yliopisto, Finland)
  Have I seen you before? Principles of Bayesian predictive classification revisited

- **Cornebise, Julien** (University of British Columbia, Canada)
  Ratmann, Oliver (Duke University, USA)
  Auxiliary sequential Monte Carlo samplers with applications to approximate Bayesian computation

- **Corradi, Fabio** (Università di Firenze, Italy)
  Ricciardi, Federico (Università di Firenze, Italy)
  Informativeness and experimental design issues in definite identification systems based on genetic markers

- **Cosma, Ioana A.** (University of Cambridge, UK)
  Gramacy, Robert B. (University of Cambridge, UK)
  Polson, Nicholas G. (University of Chicago, USA)
  Particle learning for regularization

- **Costain, Deborah** (Lancaster University, UK)
  Bayesian partitioning for mapping residual disease risk: an 'individually-matched' case-control design

- **Cui, Tiangang** (The University of Auckland, New Zealand)
  Fox, Colin (University of Otago, New Zealand)
  O'Sullivan, Mike (The University of Auckland, New Zealand)
  An adaptive delayed acceptance MCMC algorithm for massive inverse problem

- **Cumming, Jonathan** (Durham University, UK)
  Goldstein, Michael (Durham University, UK)
  Dimension reduction for Bayesian analysis of high-dimensional computer models using principal variables

- **Czado, C.** (Technische Universität München, Garching, Germany)
  Bayesian inference and model selection for models involving pair copula constructions

Poster Session 2

Date and time: **Friday June 4th, 22h00**
Coordinator: **Petrone, Sonia** (Università Bocconi, Italy)

- **Arnold, Richard** (Victoria Unversity of Wellington, New Zealand)
  Townend, John (Victoria Unversity of Wellington, New Zealand)
  Bias in the estimation of tectonic stress from earthquake focal mechanisms

- **Dahl, David** (Texas A&M University, USA)
  Day, Ryan (University of the Pacific, USA)
  Tsai, Jerry (University of the Pacific, USA)
  Distance-based probability distributions on set partitions for Bayesian nonparametric models

- **Dahyot, Rozenn** (Trinity College Dublin, Ireland)
  Averaged likelihoods

- **de la Horra, Julián** (Universidad Autónoma de Madrid, Spain)
  Rodríguez-Bernal, M. Teresa (Univ. Complutense de Madrid, Spain)
  Comparing and assessing discrepancy measures for Bayesian model selection

- **Delatola, Eleni-Ioanna** (University of Kent, UK)
  Griffin, Jim (University of Kent, UK)
  Bayesian semiparametric modelling of volatility

- **Dimitrakopoulou, Vasiliki** (University of Kent, UK)
  Brown, Philip (University of Kent, UK)
  Griffin, James (University of Kent, UK)
  Bayesian variable selection in cluster analysis.

- **Domijan, Katarina** (National Univ. of Ireland Maynooth, Ireland)
  Wilson, Simon (Trinity College Dublin, Ireland)
  Variable selection for classification of high dimensional data

- **Donald, Margaret** Queensland Univ. of Technology, Australia)
  Alston, Clair (Queensland University of Technology, Australia)
  Young, Rick (Tamworth Agricultural Institute, Australia)
  Mengersen, Kerrie (Queensland Univ. of Technology, Australia)
  Bayesian analysis of a complex agricultural field trial

- **Donnet, Sophie** (Université Paris Dauphine University, France)
  Foulley, Jean-Louis (INRA-Jouy, France)
  Samson, Adeline (Université Paris Descartes, France)
  Bayesian analysis of growth curves using mixed models defined by stochastic differential equations

- **Drovandi, Christopher** (Queensland Univ. of Technology, Australia)
  Pettitt, Anthony (Queensland Univ. of Technology, Australia)
  Faddy, Malcolm (Queensland Univ. of Technology, Australia)
  Approximate Bayesian computation using indirect inference

- **Dukic, Vanja** (University of Chicago, USA)
  Bayesian model for assessing smoking exposure during pregnancy

- **DuMouchel, William** (Phase Forward Lincoln Safety Group, USA)
  Multivariate Bayesian logistic regression for analysis of clinical trial safety issues

- **Elfadaly, Fadlalla** (The Open University, UK)
  Garthwaite, Paul (The Open University, UK)
  On eliciting priors for correlated coefficients in generalized linear models

- **Engelhardt, Barbara** (University of Chicago, USA)
  Stephens, Matthew (University of Chicago, USA)
  Sparsity-inducing priors: A practical comparison across different types of data

- **Everitt, Richard** (University of Bristol, UK)
  Didelot, Xavier (University of Warwick, UK)
  Johansen, Adam M. (University of Warwick, UK)
  Lawson, Dan (University of Bristol, UK)
  Likelihood-free estimation of model evidence

- **Fall, Mame Diarra** (Université Paris-Sud 11, France)
  Barat, Eric (Commissariat à l'Energie Atomique, LIST, France)
  Comtat, Claude (Commissariat à l'Energie Atomique, SHFJ, France)
  Dautremer, Thomas (Commissariat à l'Energie Atomique, LIST, France))
  Montagu, Thierry (Commissariat à l'Energie Atomique, LIST, France)
  Djafari, Ali (CNRS/SUPELEC/UPS, France)
  Trebossen, Régine (Commissariat à l'Energie Atomique, SHFJ, France)
  Bayesian nonparametric modeling in space-time emission tomography

- **Farah, Marian** (University of California, Santa Cruz, USA)
  Kottas, Athanasios (University of California, Santa Cruz, USA)
  Morris, Robin D. (Res. Inst. Advanced Computer Science, USA)
  Bayesian inference for global sensitivity analysis and calibration of radiative transfer models

- **Faria, Álvaro** (The Open University, UK)
  Dynamic Bayesian smooth transition auto-regression models

- **Ferreira, Marco** (University of Missouri, USA)
  Holan, Scott (University of Missouri, USA)
  Bertolde, Adelmo (Universidade Federal do Espírito Santos, Brazil)
  Dynamic multiscale spatio-temporal models for areal data

- **Figueira, Manuela** (Instituto Politécnico da Guarda, Portugal)
  Rosado, Fernando (Universidade de Lisboa, Portugal)
  Bayes in the court

- **Filippone, Maurizio** (University of Glasgow, UK)
  Calderhead, Ben (University of Glasgow, UK)
  Girolami, Mark (University of Glasgow, UK)
  Mohamed, Linah (Heriot-Watt University, UK)
  Christie, Mike (Heriot-Watt University, UK)
  Inference for Gaussian process emulation of oil reservoir simulation codes

- **Ford, Ashley** (Warwick University, UK)
  Roberts, Gareth (Warwick University, UK)
  Indian buffet epidemics: A Bayesian approach to modelling heterogeneity

- **Ford, Eric** (University of Florida, USA)
  Ghosh, Malay (University of Florida, USA)
  Guo, Pengcheng (University of Florida, USA)
  Moorhead, Althea (University of Florida, USA)
  Nelson, Benjamen (University of Florida, USA)
  Payne, Matthew (University of Florida, USA)
  Veras, Dimitri (University of Florida, USA)
  Wang, Ji (University of Florida, USA)
  Bayesian methods for analyzing extrasolar planet observations

- **Forte, Anabel** (Universitat de València, Spain)
  Bayarri, Susie (Universitat de València, Spain)
  Berger, James (Duke University, USA)
  García-Donato, Gonzalo (Universidad de Castilla-La Mancha)
  Invariant priors for variable selection

- **Fortunato, Lèa** (Imperial College London, UK)
  Best, Nicky (Imperial College London, UK)
  Richardson, Sylvia (Imperial College London, UK)
  Area-specific decision rules for classifying unusual spatial risks of disease

- **Fouskakis, Dimitris** (National Technical University of Athens, Greece)
  Ntzoufras, Ioannis (Athens Univ. of Economics and Business, Greece)
  Draper, David (University of California at Santa Cruz, USA)
  Incorporating Cost in Bayesian Variable Selection, with application to cost-effective measurement of quality of health care

- **French, Simon** (Manchester Business School, UK)
  Combining expert judgement

- **Fried, Roland** (Technische Universität Dortmund, Germany)
  Agueusop, Inoncent (Technische Universität Dortmund, Germany)
  Bornkamp, Björn (Technische Universität Dortmund, Germany)
  Fokianos, Konstantinos (University of Cyprus, Cyprus)
  Ickstadt, Katja (Technische Universität Dortmund, Germany)
  Detection of outliers and interventions in INGARCH time series

- **Fronczyk, Kassandra** (University of California, Santa Cruz, USA)
  Kottas, Athanasios (University of California, Santa Cruz, USA)
  A Bayesian nonparametric modeling framework for developmental toxicity studies

- **Fu, Audrey Q** (University of Cambridge, UK)
  Krejci, Alena (University of Cambridge, UK)
  Housden, Ben (University of Cambridge, UK)
  Fischer, Bettina (University of Cambridge, UK)
  Bernard, Fred (University of Cambridge, UK)
  Bray, Sarah (University of Cambridge, UK)
  Russell, Steve (University of Cambridge, UK)
  Tavaré, Simon (University of Cambridge, UK)
  Clustering and characterizing temporal profiles of gene expression

- **Gargallo, Pilar** (Universidad de Zaragoza, Spain)
  Altuzarra, Alfredo (Universidad de Zaragoza, Spain)
  Moreno, José María (Universidad de Zaragoza, Spain)
  Salvador, Manuel (Universidad de Zaragoza, Spain)
  The influence of the criteria in the Bayesian AHP

- **Garthwaite, Paul** (Open University, UK)
  Fan, Yanan (University of New South Wales, Australia)
  Sisson, Scott (University of New South Wales, Australia)
  Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro process

- **Germain, Sarah** (Newcastle University, UK)
  Farrow, Malcolm (Newcastle University, UK)
  Boys, Richard (Newcastle University, UK)
  Building genuine beliefs into a prior distribution for the covariance matrix

- **Ghosh, Joyee** (Univ. of North Carolina at Chapel Hill, USA)
  Herring, Amy (Univ. of North Carolina at Chapel Hill, USA)
  Siega-Riz, Anna (Univ. of North Carolina at Chapel Hill, USA)
  Bayesian variable selection for latent class models

- **Ghosh, Sujit** (North Carolina State Univerisity, USA)
  White, Gentry (The University of Queensland, USA)
  Hristopulos, Dionissios (Technical University of Crete, Greece)
  Bayesian inference for spartan spatial random fields

- **Gillespie, Colin** (Newcastle University, UK)
  Golightly, Andrew (Newcastle University, UK)
  Bayesian inference for generalized stochastic population growth models with application to aphids

- **Girolami, Mark** (University of Glasgow, UK)
  Calderhead, Ben (University of Glasgow, UK)
  Roberts, Gareth (University of Warwick, UK)
  Differential geometric structure in MCMC

- **Girón, F. Javier** (Universidad de Málaga, Spain)
  Moreno, Elías (Universidad de Granada, Spain)
  Martínez, M. Lina (Universidad de Málaga, Spain)
  Approximations to Bayes factors for intrinsic priors: Generalizing the BIC

- **Golightly, Andrew** (Newcastle University, UK)
  Gillespie, Colin (Newcastle University, UK)
  Bayesian inference for hybrid discrete-continuous stochastic kinetic models

- **Gómez-Villegas, Miguel** (Universidad Compluense de Madrid, Spain)
  González Pérez, Beatriz (Universidad Compluense de Madrid, Spain)
  Approaching p-values by means of least favorable answers in the multivariate two-sided testing problem

- **Gordillo, Marisol** (Universidad del Valle, Colombia)
  Olaya, Javier (Universidad del Valle, Colombia)
  Delvasto, Silvio (Universidad del Valle, Colombia)
  Mejía de Gutiérrez, Ruby (Universidad del Valle, Colombia)
  Analysis of a prior distribution of resistance compression of a alkali-activated concrete

- **Gormley, Claire** (University College Dublin, Ireland)
  Murphy, Brendan (University College Dublin, Ireland)
  A mixture of experts latent position cluster model

- **Goyal, Philip** (Perimeter Institute, Canada)
  Knuth, Kevin (SUNY at Albany, USA)
  Skilling, John (Maximum Entropy Data Consultants, Ireland)
  Foundations of probability and quantum theory

- **Grapsa, Erofili** (University of Southampton, UK)
  Forster, Jonathan (University of Southampton, UK)
  Bayesian inference for categorical survey data.

- **Griffiths, William** (The University of Melbourne, Australia)
  Zhang, Xiaohui (Monash University, Australia)
  Zhao, Xueyan (Monash University, Australia)
  A stochastic frontier model for discrete ordinal outcomes: A health production function

- **Gsteiger, Sandro** (Novartis Pharma AG, Switzerland)
  Neuenschwander, Beat (Novartis Pharma AG, Switzerland)
  Mercier, François (Novartis Pharma AG, Switzerland)
  Schmidli, Heinz (Novartis Pharma AG, Switzerland)
  Using historical placebo information for the design and analysis of clinical trials in multiple sclerosis.

- **Guglielmi, Alessandra** (Politecnico di Milano, Italy)
  Ieva, Francesca (Politecnico di Milano, Italy)
  Paganoni, Anna Maria (Politecnico di Milano, Italy)
  Ruggeri, Fabrizio (CNR-IMATI, Italy)
  A random-effects model for survival analysis after acute myocardial infarction

- **Han, Chao** (Virginia Tech, USA)
  Leman, Scotland (Virginia Tech, USA)
  House, Leanna (Virginia Tech, USA)
  Bayesian interactive clustering for high dimensional ata

- **Haran, Murali** (Pennsylvania State University, USA)
  Bhat, K. Sham (Pennsylvania State University, USA)
  Tonkonojenkov, Roman (Pennsylvania State University, USA)
  Keller, Klaus (Pennsylvania State University, USA)
  Inferring likelihoods and climate system characteristics from climate models and multiple tracers

- **Harris, Keith** (University of Glasgow, UK)
  Girolami, Mark (University of Glasgow, UK)
  Inferring meta-covariates in classification via Gibbs sampling

- **Heaton, Matthew J.** (Duke University, USA)
  Gelfand, Alan (Duke University, USA)
  Spatial regression using kernel averaged predictors

- **Heaton, Timothy** (University of Sheffield, UK)
  Blackwell, Paul (University of Sheffield, UK)
  Buck, Caitlin (University of Sheffield, UK)
  Bayesian non-parametric regression with observations at imprecise times: radiocarbon calibration

- **Henderson, Daniel** (Newcastle University, UK)
  Boys, Richard (Newcastle University, UK)
  Wilkinson, Darren (Newcastle University, UK)
  Bayesian calibration of a stochastic computer model of mitochondrial DNA population dynamics using multiple data sources

- **Heron, Eleisa** (Trinity College Dublin, Ireland)
  Anney, Richard (Trinity College Dublin, Ireland)
  Gallagher, Louise (Trinity College Dublin, Ireland)
  Gill, Michael (Trinity College Dublin, Ireland)
  Bayes factors for the transmission disequilibrium test (TDT) for trio genome-wide association data

- **Hill, Carter** (Louisiana State University, USA)
  Jeliazkov, Ivan (University of California-Irvine, USA)
  Weathers, Danny (Louisiana State University, USA)
  A Bayesian model of choice with heterogeneity

- **Hofmann, Mathias** (Technische Universität Muenchen, Germany)
  Joint Bayesian inference for D-vine based copula models with GARCH margins

- **Holand, Anna Marie** (Norges Teknisk-Natur. Univ, Norway)
  Steinsland, Ingelin (Norges Teknisk-Natur. Univ, Norway)
  Jensen, Henrik (Norges Teknisk-Natur. Univ, Norway)
  Bayesian animal model using integrated nested Laplace approximations: a wild House Sparrow population case study

- **Holloman, Christopher** (The Ohio State University, USA)
  Nittrouer, Susan (The Ohio State University, USA)
  Chapman, Chris (The Ohio State University, USA)
  Welling, Brad (The Ohio State University, USA)
  A Bayesian latent growth model of language acquisition in deaf children

- **Holmes, Chris** (University of Oxford, UK)
  Yau, Chris (University of Oxford, UK)
  Variable selection for Bayesian mixture models

- **Houlding, Brett** (Trinity College Dublin, Ireland)
  Coolen, Frank (Durham University, UK)
  Nonparametric predictive utility inference

- **House, Leanna** (Virginia Tech, USA)
  Reinforcing reification with application to a rainfall-runoff computer model

- **Huan, Xun** (Massachusetts Institute of Technology, USA)
  Marzouk, Youssef (Massachusetts Institute of Technology, USA)
  Simulation-based optimal experimental design with polynomial surrogates on sparse grids

- **Hurtado, Sandra** (University of Connecticut, USA)
  Dey, Dipak (University of Connecticut, USA)
  Semiparametric Bayesian modeling of spatio-survival data under cure fraction

- **Hutter, Marcus** (Australian National University, Australia)
  Predictive hypothesis identification

- **Jarocinski, Marek** (European Central Bank, Germany)
  Marcet, Albert (London School of Economics and CEPR, UK)
  Autoregressions and priors about initial growth rates

- **Jo, Seongil** (Seoul National University, Korea)
  Lim, Yaeji (Seoul National University, Korea)
  Lee, Jaeyong (Seoul National University, Korea)
  Oh, Hee-Seok (Seoul National University, Korea)
  Kang, Hyun-Suk (Korea Meteorological Administration, Korea)
  Bayesian regression models for seasonal forecast of precipitation over Korea

- **Johnson, Sandra** (Queensland University of Technology, Australia)
  Mengersen, Kerrie (Queensland University of Technology, Australia)
  Object oriented Bayesian networks designing for simplicity and integration

- **Johnson, Valen** (M.D. Anderson Cancer Center, USA)
  On the consistency of Bayesian model selection procedures

- **Jones, Emma** (University of Sheffield, UK)
  Buck, Caitlin (University of Sheffield, UK)
  Litton, Clifford. (University of Nottingham, UK)
  Tyers, Cathy (University of Sheffield, UK)
  Bayliss, Alex (English Heritage, UK)
  Using a Bayesian hierarchical model for tree-ring dating

- **Joshi, Chaitanya** (Trinity College Dublin, Ireland)
  Wilson, Simon (Trinity College Dublin, Ireland)
  O'Brien, Eugene (University College Dublin, Ireland)
  Taheri, Abdolrahim (University College Dublin, Ireland)
  Stochastic differential equation models for dynamic force

- **Juárez, Miguel** (University of Warwick, UK)
  Morrissey, Edward (University of Warwick, UK)
  Burroughs, Nigel (University of Warwick, UK)
  Inferring the topology of a non-linear sparse gene regulatory network using fully Bayesian spline regression

Poster Session 3

Date and time: **Saturday June 5th, 22h00**
Coordinator: **Christen, Andrés** (CIMAT, México)

- **Bornn, Luke** (University of British Columbia, Canada)
  Caron, François (INRIA Bordeaux Sud-Ouest, France)
  The product graphical model

- **Gómez-Rubio, Virgilio** (Universidad de Castilla-La Mancha, Spain)
  Fast Bayesian detection of disease clusters using generalized linear mixed models

- **Haas, Armin** (Potsdam-Institut für Klimafolgenforschung, Germany)
  Fucik, Markus (Universität Potsdam , Germany)
  Ötsch, Rainald (Universität Potsdam , Germany)
  Bayesian risk management

- **Kazianka, Hannes** (Technische Universität Wien, Austria)
  Pilz, Juergen (Alpen Adria Universität Klagenfurt, Austria)
  Objective Bayesian analysis of spatially correlated data including measurement error

- **Keller, Merlin** (EDF R&D, Chatou, France)
  Parent, Eric (AgroParisTech, Paris, France)
  Bousquet, Nicolas (EDF R&D, Chatou, France)
  Pasanisi, Alberto (EDF R&D, Chatou, France)
  Bayesian and frequentist parametric predictions of a tail probability in an industrial reliability context

- **Klauenberg, Katy** (PTB-Berlin, Germany)
  Ebert, Bernd (PTB-Berlin, Germany)
  Voigt, Jan(PTB-Berlin, Germany)
  Elster, Clemens (PTB-Berlin, Germany)
  Measurement uncertainty of ELISA concentration estimates in biochemistry

- **Kohonen, Jukka** (Helsingin Yliopisto, Finland)
  Representation of interactions in protein annotation

- **Koizumi, Daiki** (Waseda University, Japan)
  Suko, Tota (Waseda University, Japan)
  Matsushima, Toshiyasu (Waseda University, Japan)
  On the Bayesian forecasting algorithm under the non-stationary binomial distribution with the hyper parameter estimation

- **Kottas, Athanasios** (University of California, Santa Cruz, USA)
  Fellingham, Gilbert W. (Brigham Young University, USA)
  Bayesian semiparametric modeling with mixtures of symmetric distributions

- **Kruijer, Willem** (Université Paris Dauphine and CREST, France)
  Rousseau, Judith (Université Paris Dauphine and CREST, France)
  On Bayesian estimation of the long-memory parameter in the FEXP-model for Gaussian time series

- **Laine, Marko** (Ilmatieteen Laitos, Finland)
  Tamminen, Johanna (Ilmatieteen Laitos, Finland)
  Adaptive reversible jump MCMC and applications to model selection and averaging in geophysical re-mote sensing

- **Lam, Connie P. Y.** (The University of Sydney, Australia)
  Chan, Jennifer S. K. (The University of Sydney, Australia)
  Yu, Philip L. H. (The University of Hong Kong, HK)
  Choy, Boris S. T. (The University of Sydney, Australia)
  Bayesian conditional autoregressive geometric process model for range data

- **Lambert, Philippe** (Université de Liège, Belgium)
  Estimation of conditional quantiles from bivariate histogram data

- **Latuszynski, Krzysztof** (University of Warwick, UK)
  Papaspiliopoulos, Omiros (Universitat Pompeu Fabra, Spain)
  Pillai, Natesh (University of Warwick, UK)
  Roberts, Gareth (University of Warwick, UK)
  Stability of the Gibbs sampler for hierarchical generalized linear model

- **Launay, Tristan** (Université de Nantes, France)
  Philippe, Anne (Université de Nantes, France)
  Lamarche, Sophie (EDF R&D, France)
  A non linear regression model applied to electricity load forecasting

- **Lee, Dominic** (University of Canterbury, New Zealand)
  Darlow, Brian (Christchurch School of Medicine, New Zealand)
  Zahari, Marina (University of Canterbury, New Zealand)
  Bayesian nonparametric modelling of risk factors for retinopathy of prematurity

- **Lee, J. Jack** (University of Texas, USA)
  Yin, Guosheng Nan (University of Hong Kong, Hong Kong)
  Chen, Nan (University of Hong Kong, Hong Kong)
  Randomized phase II trial design with Bayesian adaptive randomization and predictive probability

- **Leininger, Thomas J.** (Brigham Young University, USA)
  Reese, Shane (Brigham Young University, USA)
  Burraston, Bert (Brigham Young University, USA)
  Rutkowski, Michael (Brigham Young University, USA)
  Criminal career trajectory latent class estimation using a Bayesian hierarchical dynamic linear model

- **Leman, Scotland** (Virginia Tech, USA)
  House, Leanna (Virginia Tech, USA)
  North, Chris (Virginia Tech, USA)
  Bayesian visual analytics (BaVA): A formal visual updating procedure

- **Leon-Novelo, Luis** (University of Florida, USA)
  Casella, George (University of Florida, USA)
  Model selection in probit models with non-intrinsic Bayes factors

- **Li, Fan** (Duke University, USA)
  Schwartz, Scott (Duke University, USA)
  Mealli, Fabrizia (Università di Firenze, Italy)
  A Bayesian semiparametric approach to causal inference with intermediate variables

- **Li, Feng** (Stockholms Universitet, Sweden)
  Villani, Mattias (Sveriges Riksbank, Sweden)
  Kohn, Robert (University of New South Wales, Australia)
  Flexible modeling of conditional distribu tions using smooth mixtures of asymmetric student t densities

- **Li, Guangquan** (Imperial College London, UK)
  Best, Nicky (Imperial College London, UK)
  Richardson, Sylvia (Imperial College London, UK)
  Hansell, Anna (Imperial College London, UK)
  A Bayesian mixture model for detecting unusual time trends in small area estimates: application to COPD mortality in England

- **Li, Yingbo** (Duke University, USA)
  Clyde, Merlise A. (Duke University, USA)
  Bayesian hierarchical shrinkage prior for sparse signals

- **Liang, Li-Jung** (University of California at Los Angeles, USA)
  Weiss, Robert (University of California at Los Angeles, USA)
  Suchard, Marc (University of California at Los Angeles, USA)
  Stitching time to analyze nine: Hierarchical modeling of longitudinally collected HIV-1 phylogenies

- **Lin, Lin** (Duke University, USA)
  Chan, Cliburn (Duke University, USA)
  West, Mike (Duke University, USA)
  Structured prior modelling for Bayesian analysis of complex mixtures in combinatorial marker-based flow cytometry studies

- **Liverani, Silvia** (University of Bristol, UK)
  Smith, Jim Q. (University of Warwick, UK)
  Bayesian partition model selection on high dimensional dependent time series

- **Loizides, Charalambos** (University of Oxford, UK)
  Yau, Christopher (University of Oxford, UK)
  Holmes, Chris (University of Oxford, UK)
  Bayesian inference in a genome-wide association kin-cohort studies

- **Lopes, Danilo** (Duke University, USA)
  Application of Gaussian predictive process to the evaluations of a volcanic flow computer model

- **López-Quílez, Antonio** (Universitat de València, Spain)
  Botella, Paloma (Universidad CEU-Cardenal Herrera, Spain)
  Martínez-Beneito, Miguel-Angel (CSISP, Spain)
  Moving average smoothing for spatio-temporal disease mapping

- **Loschi, Rosangela** (Universidade Federal de Minas Gerais, Brazil)
  Monteiro, João (University of Minnesota, USA)
  Assunção, Renato (Universidade Federal de Minas Gerais, Brazil)
  Product partition models with correlated parameters

- **Love, Tanzy** (University of Rochester, USA)
  Carriquiry, Alicia (Iowa State University, USA)
  Clustering Posterior Densities - Application in Microarrays

- **Low-Choy, Samantha** (CRC National Plant Biosecurity, Australia)
  Mengersen, Kerrie (Queensland University of Technology, Australia)
  Murray, Justine (CSIRO Entomology, Australia)
  Rousseau, Judith (CEREMADE, Université Dauphine, France)
  How should we combine expert opinions: on elicitation, encoding, priors or posteriors?

- **Lum, Kristian** (Duke University, USA)
  Gelfand, Alan (Duke University, USA)
  Bayesian spatial quantile regression: An application to conditional birthweight distributions

- **Lutz, Sharon** (Universität Bonn, Germany)
  Weiss, Scott (Harvard University, USA)
  Lange, Christoph (Harvard University, USA)
  Using expression data in genetic association studies: An integrated Bayesian approach to determine genetic pathways

- **Lykou, Anastasia** (Lancaster University, UK)
  Ntzoufras, Ioannis (Athens Univ. of Economics and Business, Greece)
  On Bayesian variable selection using lasso and related methods

- **Lytton, Diane** (Private Practice, USA)
  Bayesian analyses in USA sexually violent predator civil commitments

- **Madi, Mohamed** (UAE University, UAE)
  Raqab, Mohamed (University of Jordan, Jordan)
  Inference for the generalized Rayleigh distribution based on progressively censored data

- **Main, Paloma** (Universidad Complutense de Madrid, Spain)
  Gómez-Villegas, Miguel (Universidad Complutense de Madrid, Spain)
  Navarro, Hilario (UNED, Spain)
  Susi, Rosario (Universidad Complutense de Madrid, Spain)
  Sensitivity to prior hyperparameters in Gaussian Bayesian networks

- **Maiti, Dipayan** (Virginia Tech, USA)
  Leman, Scotland (Virginia Tech, USA)
  Multiset model selection

- **Malsiner-Walli, Gertraud** (Johannes Kepler Univ. Linz, Austria)
  Wagner, Helga (Johannes Kepler Univ. Linz, Austria)
  Bayesian variable selection in normal regression models

- **Marín, J. Miguel** (Universidad Carlos III de Madrid, Spain)
  Rodríguez-Bernal, M. Teresa (Univ. Complutense de Madrid, Spain)
  Bayesian clustering based multiple hypotheses testing

- **Martínez-Beneito, Miguel A.** (CSISP Valencia, Spain)
  García-Donato, Gonzalo (Universidad de Castilla-La Mancha, Spain)
  Variable selection with Gibbs samplers and Zellner-Siow priors

- **Martínez-Beneito, Miguel A.** (CSISP Valencia, Spain)
  Sun, Dongchu (University of Missouri, USA)
  A computationally convenient framework for multivariate disease mapping

- **Martino, Luca** (Universidad Carlos III, Spain)
  Míguez, Joaquín (Universidad Carlos III, Spain)
  Two adaptive rejection sampling methods for sampling a posteriori probability distributions

- **Martins, Camila** (Universidade de São Paulo, Brazil)
  Tomazella, Vera L. (Universidade Federal de São Carlos, Brazil)
  Polpo, Adriano (Universidade Federal de São Carlos, Brazil)
  Bayesian reference analysis for the Weibull model in the application competing risk

- **Marttinen, Pekka** (Aalto-Universitetets Teckniska Högskola, Finland)
  Corander, Jukka (Helsingin Yliopisto, Finland)
  Bayesian multi-locus model for gene-gene interactions

- **Marzouk, Youssef** (Massachusetts Institute of Technology, USA)
  Paul Boggs (Sandia National Laboratories, USA)
  Nonparametric Bayesian density estimation using polynomial chaos expansions

- **Mason, Alexina** (Imperial College London, UK)
  Best, Nicky (Imperial College London, UK)
  Richardson Sylvia (Imperial College London, UK)
  Plewis, Ian (University of Manchester, UK)
  Strategy for modelling non-random missing data mechanisms in longitudinal studies: Application to in-come data from the millennium cohort study

- **Mayoral, Asunción** (Universidad Miguel Hernández de Elche, Spain)
  Quesada, M. (Universidad Miguel Hernández de Elche, Spain)
  Morales, J. (Universidad Miguel Hernández de Elche, Spain)
  Barber, Xavier (Universidad Miguel Hernández de Elche, Spain)
  OBANSoft, a Bayesian software for objective Bayesian analysis: Its first days.

- **Mayrink, Vinicius** (Duke University, USA)
  Lucas, Joseph (Duke University, USA)
  A Bayesian factor model for gene expression detection on oligonucleotide microarrays.

- **McCandless, Lawrence** (Simon Fraser University, Canada)
  Sequential versus full-Bayesian regression adjustment for the propensity score.

- **Mattei, Alessandra** (Università degli Studi di Firenze, Italy)
  Mealli, Fabrizia (Università degli Studi di Firenze, Italy)
  Bayesian inference in augmented designs for assessing principal strata causal effects

- **Mendoza, Manuel** (ITAM, Mexico)
  Gutiérrez-Peña, Eduardo (IIMAS-UNAM, Mexico)
  Some thoughts on the Bayesian robustness of location-scale models

- **Merl, Daniel** (Lawrence Livermore National Laboratory, USA)
  Sequential clustering and anomaly detection for streaming data

- **Meyer, Renate** (University of Auckland, New Zealand)
  Cai, Bo (University of South Carolina, USA)
  Adaptive Metropolis-Hastings-within-Gibbs algorithms using copulae

- **Migon, Helio** (Universidade Federal do Rio de Janeiro, Brazil)
  Salazar, Esther (Universidade Federal do Rio de Janeiro, Brazil)
  Alves, Larissa (Universidade Federal do Rio de Janeiro, Brazil)
  A Bayesian spatial model for panel time series data

- **Miller, Kurt** (University of California at Berkeley, USA)
  Jordan, Michael (University of California at Berkeley, USA)
  Griffiths, Thomas (University of California at Berkeley, USA)
  Non-exchangeable Bayesian nonparametric latent feature models

- **Min, Xiaoyi** (University of Missouri, USA)
  Sun, Dongchu (University of Missouri, USA)
  Bayesian model selection for 2-way ANOVA models

- **Mira, Antonietta** (Università dell'Insubria, Italy and Università della Svizzera italiana, Switzerland)
  Solgi, Reza* (Swiss Finance Institute at the Università della Svizzera Italiana, Switzerland)
  Zero variance simulation: From Monte Carlo to Markov chain Monte Carlo simulation

- **Miyawaki, Koji** (National Institute for Environmental Studies, Japan)
  Omori, Yasuhiro (University of Tokyo, Japan)
  Hibiki, Akira (National Institute for Environmental Studies, Japan)
  A Bayesian estimation of the residential gas demand on the nonconvex budget set

- **Mohammadzadeh, Mohsen** (Tarbiat Modares University, Iran)
  Omid Karimi (Tarbiat Modares University, Iran)
  Bayesian spatial prediction for closed skew Gaussian random fields

- **Montes Diez, Raquel** (Universidad Rey Juan Carlos, Spain)
  Quirós, Alicia (Universidad Rey Juan Carlos, Spain)
  Modelling the hemodynamic response in fMRI using Gaussian processes

- **Morris, Robin D.** (USRA-RIACS, USA)
  Farah, Marian (University of California, Santa Cruz, USA)
  Kottas, Athanasios (University of California, Santa Cruz, USA)
  Predicting failures due to spaceborne radiation: A comparison of parametric and semiparametric Bayesian isotonic regression

- **Moura, Fernando** (Universidade Federal do Rio de Janeiro, Brazil)
  Silva, Valmária (Universidade Federal do Rio de Janeiro, Brazil)
  Small area estimation using skew-normal models

- **Mukherjee, Chiranjit** (Duke University, USA)
  West, Mike (Duke University, USA)
  Novel Bayesian models and inference for high-resolution lattice data

- **Mukhopadhyay, Nitai** (Virginia Commonwealth University, USA)
  Dass, Sarat (Michigan State University, USA)
  Bayesian model selection on inverse correlation matrices with application to sparse networks

- **Mukhopadhyay, Sabyasachi** (Indian Statistical Institute, India)
  Roy, Sisir (Indian Statistical Institute, India)
  Bhattacharya, Sourabh (Indian Statistical Institute, India)
  Fast and efficient Bayesian semi-parametric curve-fitting and clustering in massive data

- **Muñoz, Facundo** (Universitat de València, Spain)
  Amorós, Rubén (Universitat de València, Spain)
  López-Quílez, Antonio (Universitat de València, Spain)
  Martínez-Beneito, Miguel A. (CSISP Valencia, Spain)
  Fast methods of spatio-temporal disease mapping with INLA

- **Murdoch, Duncan** (University of Western Ontario, Canada)
  MCMC in R

- **Murphy, Brendan** (University College Dublin, Ireland)
  Salter-Townshend, Michael (University College Dublin, Ireland)
  Variational Bayes for the latent position cluster model in social network analysis.

- **Murray, Iain** (University of Edinburgh, UK)
  Adams, Ryan (University of Toronto, Canada)
  Slice sampling with latent Gaussian models

- **Nakajima, Jouchi** (Duke University, USA)
  Omori, Yasuhiro (University of Tokyo, Japan)
  Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution

- **Nakatsuma, Teruo** (Keio University, Japan)
  A Markov chain Monte Carlo implementation of the Bayesian method of moments for linear regression models

- **Nascimento, Fernando** (Universidade Federal do Rio de Janeiro, Brazil
  Gamerman, Dani (Universidade Federal do Rio de Janeiro, Brazil)
  Lopes, Hedibert (University of Chicago, USA)
  A semiparametric Bayesian approach to extreme value estimation

- **Natário, Isabel** (Universidade Nova de Lisboa, Portugal)
  Carvalho, Lucília (Universidade de Lisboa, Portugal)
  Bayesian estimation of ILI incidence using non-representative data

- **Reiter, Jerome** (University of Southampton, UK)
  Mitra, Robin (University of Southampton, UK)
  Estimating risks of identification disclosure in partially synthetic data

Poster Session 4

Date and time: **Sunday June 6th, 22h00**
Coordinator: **Lee, Herbie** (University of California, Santa Cruz, USA)

- **Myung, Jay** (Ohio State University, USA)
  Cavagnaro, Daniel (Ohio State University, USA)
  Pitt, Mark (Ohio State University, USA)
  Bayesian adaptive optimal design for discriminating models of cognition

- **Neuenschwander, Beat** (Novartis Pharma, Switzerland)
  Capkun-Niggli, Gorana (Novartis Pharma, Switzerland)
  Branson, Michael (Novartis Pharma, Switzerland)
  Spiegelhalter, David (MRC, Biostatistics Unit, UK)
  The meta-analytic-predictive approach to deriving priors from historical data in clinical trials

- **Khatab, Khaled** (Aachen University, Germany)
  Ngianga-Bakwin, Kandala (University of Warwick, UK)
  Malnutrition among children under the age of five in the Democratic Republic of Congo: Do geography matters?

- **Nicholson, George** (University of Oxford, UK)
  Holmes, Chris (University of Oxford, UK)
  Robust Bayesian models for ANOVA

- **Nicolette, Raquel** (Universidade de Aveiro, Portugal)
  Pereira, Isabel (Universidade de Aveiro, Portugal)
  Scotto, Manuel (Universidade de Aveiro, Portugal)
  Bayesian inference in integer-valued self-exciting threshold autoregressive models SETINAR(2;p)

- **Nipoti, Bernardo** (Università di Pavia, Italy)
  A construction of vectors of dependent Dirichlet processes

- **Nixon, Richard** (Novartis AG, Switzerland)
  Using short-term evidence to predict six-month outcomes in clinical trials of signs and symptoms in rheumatoid arthritis

- **Ottavi, Alessandro** (Norges Teknisk-Natur. Univ, Norway)
  Bayesian spatial smoothing over complex domains

- **Overstall, Antony** (University of Southampton, UK)
  Woods, Dave (University of Southampton, UK)
  Bayesian emulators for multivariate computer models with categorical inputs

- **Oxlade, Rachel** (University of Durham, UK)
  Bayesian emulators as a tool for comparing ecosystem models

- **Page, Garritt** (Duke University, USA)
  Dunson, David (Duke University, USA)
  Bayesian local contamination models for multivariate outliers

- **Palacios, Ana** (Universidad Carlos III de Madrid, Spain)
  Marín, Juan (Universidad Carlos III de Madrid, Spain)
  Wiper, Michael (Universidad Carlos III de Madrid, Spain)
  Bayesian estimation of bacterial growth curves using neural networks and hierarchical Gompertz models

- **Pamminger, Christoph** (Johannes Kepler Univ. Linz, Austria)
  Frühwirth-Schnatter, Sylvia (Johannes Kepler Univ. Linz, Austria)
  Winter-Ebmer, Rudolf (Johannes Kepler Univ. Linz, Austria)
  Weber, Andrea (RWI Essen, Germany)
  Model-based clustering of categorical time series using finite mixtures of Markov chain models with logit extension

- **Panagiotelis, Anastasios** (Technische Universität München, Germany)
  Czado, Claudia (Technische Universität München, Germany)
  Bayesian estimation of pair copula constructions with discrete margins

- **Papathomas, Michail** (Imperial College London, UK)
  Molitor, John (Imperial College London, UK)
  Richardson, Sylvia (Imperial College London, UK)
  Riboli, Elio (Imperial College London, UK)
  Vineis, Paolo (Imperial College London, UK)
  Bayesian profile regression with an application to the study of lung cancer in a large cohort study

- **Parry, Matthew** (University of Cambridge, UK)
  Dawid, A. Philip (University of Cambridge, UK)
  Lauritzen, Steffen (University of Oxford, UK)
  Local proper scoring rules for discrete outcome spaces

- **Peiris, Shelton** (The University of Sydney, Australia)
  Chan, Jennifer (The University of Sydney, Australia)
  Feng, Derek (The University of Sydney, Australia)
  Bayesian stochastic volatility model for leverage effect

- **Pereira, Isabel** (Universidade de Aveiro, Portugal)
  Silva, M. Eduarda (Universidade do Porto, Portugal)
  Detection of additive outliers in integer-valued autoregressive models

- **Pérez, María-Eglée** (Universidad de Puerto Rico, Puerto Rico)
  Pericchi, Luis-Raúl (Universidad de Puerto Rico, Puerto Rico)
  Limiting the shrinkage for the exceptional: A solution to the "Clemente Problem"

- **Pirinen, Matti** (University of Oxford, UK)
  Hellenthal, Garrett (University of Oxford, UK)
  Spencer, Chris (University of Oxford, UK)
  Donnelly, Peter (University of Oxford, UK)
  Meta-analysis in genome-wide association studies

- **Poirier, Dale J.** (University of California, Irvine, USA)
  Bayesian analysis of partial observability in bivariate probit models

- **Poleto, Frederico** (Universidade de São Paulo, Brazil)
  Paulino, Carlos (IST-Universidade Técnica de Lisboa, Portugal)
  Molenberghs, Geert (Katholieke Universiteit Leuven, Belgium)
  Singer, Julio (Universidade de São Paulo, Brazil)
  Inferential implications of over-parameterization: a case study in incomplete categorical data

- **Poppe, Stephan** (Max-Planck Gesellschaft, Germany)
  Bourguignon, Pierre-Yves (Max-Planck Gesellschaft, Germany)
  A continuum of inductive methods for discrete universes of uncertain constitution and cardinality

- **Posekany, Alexandra** (Universität für Bodenkultur Wien, Austria)
  Felsenstein, K. (Technische Universität Wien, Austria)
  Sykacek, P. (Technische Universität Wien, Austria)
  Assessing robustness issues in microarray data analysis

- **Prangle, Dennis** (Lancaster University, UK)
  Fearnhead, Paul (Lancaster University, UK)
  Summary statistics for approximate Bayesian computation

- **Puig, Xavier** (Universitat Politècnica de Catalunya, Spain)
  Josep Ginebra (Universitat Politècnica de Catalunya, Spain)
  Multinomial cluster analysis of electoral data

- **Queen, Catriona** (The Open University, UK)
  Albers, Casper (Rijksuniversiteit Groningen, The Netherlands)
  Forecasting traffic flow series: A comparison of dynamic Bayesian approaches.

- **Rahaman Khan, Hasinur** (University of Warwick, UK)
  Shaw, J. Ewart H. (University of Warwick, UK)
  Bayesian and non-Bayesian variable selection for parametric AFT models in high dimensions

- **Ramos, Sandra** (Politécnico do Porto, Portugal)
  Turkman, Antónia Amaral (Universidade de Lisboa, Portugal)
  Antunes, Marília (Universidade de Lisboa, Portugal)
  A nonparametric Bayesian approach for supervised classification based on pairs of covariates

- **Rantalainen, Mattias** (University of Oxford, UK)
  Nicholson, George (University of Oxford, UK)
  Holmes, Chris (University of Oxford, UK)
  Bayesian approaches to the analysis of spectral metabolic profiling data

- **Ratmann, Oliver** (Duke University, USA)
  Peters, Gareth (University of New South Wales, Australia)
  Richardson, Sylvia (Imperial College London, UK)
  Guiding Bayesian model choice within ABC

- **Ray, Jaideep** (Sandia National Laboratories, USA)
  McKenna, Sean (Sandia National Laboratories, USA)
  van BloemenWaander, Bart (Sandia National Laboratories, USA)
  Marzouk, Youssef (MIT, Boston, USA)
  Multiscale Bayesian reconstruction of binary permeability fields from sparse static and dynamic data

- **Renfro, Lindsay** (Baylor University, USA)
  Carlin, Bradley (University of Minnesota, USA)
  Bayesian adjusted R^2 for the meta-analytic evaluation of tme-to-event surrogate endpoints in clinical trials

- **Riebler, Andrea** (Universität Zürich, Switzerland)
  Held, Leonhard (Universität Zürich, Switzerland)
  Rue, Håvard (Norges Teknisk-Naturvitenskapelige Univ, Norway)
  Schrödle, Birgit (Universität Zürich, Switzerland)
  Correlated GMRF priors and INLA

- **Rios, Jesus** (Manchester Business School, UK)
  French, Simon (Manchester Business School, UK)
  Modelling the sustainability of nuclear power

- **Robert, Christian** (Université Paris-Dauphine, France)
  Marin, Jean-Michel (Université de Montpellier, France)
  On resolving the Savage-Dickey paradox

- **Rodrigues, Josemar** (Universidade Federal de São Carlos, Brazil)
  Castro, Mário de (Universidade de São Paulo, Brazil)
  Cancho, Vicente (Universidade de São Paulo, Brazil)
  Balakrishnan, N. (McMaster University, Ontario, Canada)
  A Bayesian destructive weighted Poisson cure rate model and an application to cutaneous melanoma data

- **Rodriguez, Abel** (University of California, Santa Cruz, USA)
  Dynamic blockmodels for social networks through Bayesian nonparametric mixtures

- **Román, Lizbeth** (Universitat de València, Spain)
  Reference priors for prediction

- **Roodaki, Alireza** (SUPELEC, France)
  Bect, Julien (SUPELEC, France)
  Fleury, Gilles (SUPELEC, France)
  Comparison of Fully Bayesian and Empirical Bayes approaches for joint Bayesian model selection and estimation of sinusoids via reversible jump MCMC

- **Rossell, David** (Instituto de Investigación Biomédica, Barcelona, Spain)
  Johnson, Valen (M.D. Anderson Cancer Center, USA)
  Tellesca, Donatello (University of California at Los Angeles, USA)
  Non-local priors for variable selection

- **Rousseau, Judith** (Université Paris Dauphine, France)
  Mengersen, Kerrie (Queensland University of Technology, Australia)
  Asymptotic behaviour of the posterior distribution in mixture models with too many components

- **Rueda, Raúl** (IIMAS-UNAM, Mexico)
  Gutiérrez-Peña, Eduardo (IIMAS-UNAM, Mexico)
  Reference prior for the shape parameter of the exponential power distribution

- **Rufo, María-Jesús** (Universidad de Extremadura, Spain)
  Martín, Jacinto (Universidad de Extremadura, Spain)
  Pérez, Carlos (Universidad de Extremadura, Spain)
  Kullback-Leibler divergence to pool prior opinions

- **Sabanés Bové, Daniel** (Universität Zürich, Switzerland)
  Held, Leonhard (Universität Zürich, Switzerland)
  Hyper-g priors for generalized linear models

- **Safta, Cosmin** (Sandia National Laboratories, USA)
  Ray, Jaideep (Sandia National Laboratories, USA)
  Cheng, Karen (Applied Research Associates, Inc., USA)
  Crary, David (Applied Research Associates, Inc., USA)
  Bayesian classification of partially observed outbreaks using time-series data

- **Salimans, Tim** (Erasmus Universiteit Rotterdam, The Netherlands)
  Paap, Richard (Erasmus Universiteit Rotterdam, The Netherlands)
  Variable selection and functional form uncertainty in cross-country growth regressions

- **Salvador, Manuel** (Universidad de Zaragoza, Spain)
  Gargallo, Pilar (Universidad de Zaragoza, Spain)
  Gallizo, José-Luis (Universitat de Lleida, Spain)
  Saladrigues, Ramó (Universitat de Lleida, Spain)
  The persistence of abnormal returns at industry and firm levels: A Bayesian approach.

- **Sargsyan, Khachik** (Sandia National Laboratories, USA)
  Safta, Cosmin (Sandia National Laboratories, USA)
  Debusschere, Bert (Sandia National Laboratories, USA)
  Najm, Habib (Sandia National Laboratories, USA)
  Bayesian methods for discontinuity detection in climate model predictions

- **Sauleau, Erik** (Université de Strasbourg, France)
  Mameli, Valentina (Università degli studi di Cagliari, Italy)
  Musio, Monica (Università degli studi di Cagliari, Italy)
  Using integrated nested Laplace approximation for modeling spatial healthcare utilization

- **Saville, Benjamin** (Vanderbilt University School of Medicine, USA)
  Herring, Amy (University of North Carolina at Chapel Hill, USA)
  Kaufman, Jay (McGill University, USA)
  Testing variance components in multilevel linear models using approximate Bayes factors

- **Schäfer, Christian** (ENSAE, CREST, France)
  Cross-entropy and sequential Monte Carlo methods for large scale Bayesian model choice problems

- **Schmidt, Kristian** (Iowa State University, USA)
  Carriquiry, Alicia (Iowa State University, USA)
  A multivariate hierarchical Poisson approach to construct candidate lists of sites for possible improvements

- **Schmitz, Susanne** (Trinity College Dublin, Ireland)
  Adams, Roisin (National Centre for Pharmacoeconomics, Ireland)
  Walsh, Cathal (Trinity College Dublin, Ireland)
  Combining evidence for cost effectiveness analysis: An application in rheumatoid arthritis

- **Schoergendorfer, Angela** (University of Kentucky, USA)
  Branscum, Adam (University of Kentucky, USA)
  Hanson, Timothy (University of Minnesota, USA)
  A Bayesian nonparametric test for logistic distribution and odds ratio estimation without dichotomizing

- **Schott, Sarah** (Duke University, USA)
  Huber, Mark (Claremont McKenna College, USA)
  Speeding up the product estimator using random temperatures

- **Schrödle, Birgit** (Universität Zürich, Switzerland)
  Held, Leonhard (Universität Zürich, Switzerland)
  Riebler, Andrea (Universität Zürich, Switzerland)
  Bayesian inference for spatio-temporal models using INLA and MCMC: A case study on animal disease counts from Switzerland

- **Seheult, Allan** (Durham University, UK)
  Assessing model discrepancy for simulators of complex systems

- **Sepúlveda, Nuno** (Instituto Gulbenkian de Ciência, Portugal)
  Sousa, Vitor (M.D. Anderson Cancer Center, USA)
  Guindani, Michele (University of New Mexico, USA)
  Mueller, Peter (M. D. Anderson Cancer Center, USA)
  Paulino, Carlos Daniel (Universidade Técnica de Lisboa, Portugal)
  Carneiro, Jorge (Instituto Gulbenkian de Ciência, Portugal)
  Biodiversity estimation: unraveling the T-cell receptor repertoire in the body's cellular ecosystem

- **Sermaidis, Giorgos** (Lancaster University, UK)
  Exact inference for discretely observed diffusions

- **Shemyakin, Arkady** (University of St. Thomas, Minnesota, USA)
  Reference priors, information inequalities, and Hellinger distance

- **Sherlock, Chris** (Lancaster University, UK)
  Optimal scaling of the random walk Metropolis: General criteria for the 0.234 *acceptance rule*

- **Shi, Minghui** (Duke University, USA)
  Dunson, David (Duke University, USA)
  Regularized decompositions for sparse MAP estimation in factor analysis

- **Silva, Giovani** (Universidade Técnica de Lisboa, Portugal)
  Dias, Maria-Inés (Universidade de Evora, Portugal)
  Modelling the proportion of burned area in Portuguese forest fires

- **Silva-Fortes, Carina** (Universidade de Lisboa, Portugal)
  Amaral-Turkman, M. Antónia (Universidade de Lisboa, Portugal)
  Sousa, Lisete (Universidade de Lisboa, Portugal)
  Comparison between Bayesian and classical methodologies for discovering transcription-factor targets from RIP-chip experiments

- **Silva, Ricardo** (University College London, UK)
  Gramacy, Robert B. (University of Cambridge, UK)
  Gaussian process structural equation models with latent variables

- **Silva, Nélia** (Universidade de Aveiro, Portugal)
  Pereira, Isabel (Universidade de Aveiro, Portugal)
  Forecasting in INBL(1,0,1,1) Model

- **Simpson, Shawn** (Columbia University, USA)
  Madigan, David (Columbia University, USA)
  A Bayesian self-controlled method for drug safety surveillance in large-scale longitudinal data

- **Sirén, Jukka** (Helsingin Yliopisto, Finland)
  Marttinen, Pekka (Aalto-Universitetets Teckniska Högskola, Finland)
  Corander, Jukka (Helsingin Yliopisto, Finland)
  Reconstructing population histories from single-nucleotide polymorphism data

- **Sisson, Scott** (University of New South Wales, Australia)
  Fan, Yanan (University of New South Wales, Australia)
  Peters, Gareth (University of New South Wales, Australia)
  Some comments on "likelihood-free" samplers and models

- **Skilling, John** (Maximum Entropy Data Consultants, Ireland)
  Foundations of computational inference

- **Smith, Murray** (Nat. Inst. Water Atmos. Research, New Zealand)
  Baird, Suze (Nat. Inst. Water Atmos. Research, New Zealand)
  Bayesian prediction of non-fish bycatch in commercial fisheries using hierarchical models

- **Soares, Paulo** (Universidade Técnica de Lisboa, Portugal)
  Achcar, Jorge (Universidade Estadual de Maringá, Brazil)
  Paulino, Carlos (Universidade Técnica de Lisboa, Portugal)
  How much can change at a change point? A note on modelling change points in non-homogeneous Poisson processes

- **Solonen, Antti** (Lappeenrannan Teknillinen Yliopisto, Finland)
  Haario, Heikk (Lappeenrannan Teknillinen Yliopisto, Finland)
  Laine, Marko (Ilmatieteen Laitos, Finland)
  Simulation-based optimal design using a response variance criterion

- **Wiper, Michael P.** (Universidad Carlos III de Madrid, Spain)
  Justel, Ana (Universidad Autónoma de Madrid, Spain)
  Bayesian nonparametric modeling of the INAR(1) process

Poster Session 5

Date and time: **Monday June 7th, 22h00**
Coordinator: **Müller, Peter** (M. D. Anderson Cancer Center, USA)

- **Spiller, Elaine** (Marquette University, USA)
  Bayesian data assimilation for dynamical systems

- **Spitzner, Dan** (University of Virginia, USA)
  Neutral-data comparisons for Bayesian testing

- **Sprenger, Jan** (Universiteit van Tilburg, The Netherlands)
  Evidence and stopping rules in sequential trials

- **Stathopoulos, Vassilios** (University of Glasgow, UK)
  Girolami, Mark (University of Glasgow, UK)
  Efficient Bayesian inference for mixture models

- **Strimenopoulou, Foteini** (University of Kent, UK)
  Brown, Philip (University of Kent, UK)
  Robust Bayesian functional mixed effects modeling

- **Sutton, Charles** (University of Edinburgh, UK)
  Jordan, Michael (University of California, Berkeley, USA)
  Inference in queueing networks with missing data

- **Sweeney, James** (Trinity College Dublin, Ireland)
  Approximate joint statistical inference for large spatial datasets

- **Tabet, Aline** (University of British Columbia, Canada)
  Doucet, Arnaud (University of British Columbia, Canada)
  Gustafson, Paul (University of British Columbia, Canada)
  Bayesian estimation of a time-varying correlation matrix using the Wishart process

- **Tamminen, Johanna** (Ilmatieteen Laitos, Finland)
  Laine, Marko (Ilmatieteen Laitos, Finland)
  Haario, Heikki (Lappeenrannan Teknillinen Yliopisto, Finland)
  Solonen, Antti (Lappeenrannan Teknillinen Yliopisto, Finland)
  Oja, Erkki (Aalto Yliopiston Teknillinen Korkeakoulu, Finland)
  Järvinen, Heikki (Ilmatieteen Laitos, Finland)
  Novel advanced mathematical and statistical methods for understanding climate (NOVAC)

- **Tang, Jing** (VTT Technical Research Center, Finland)
  Kohonen, Jukka (Helsingin Yliopisto, Finland)
  Corander, Jukka (Helsingin Yliopisto, Finland)
  Oresic, Matej (VTT Technical Research Center, Finland)
  A Bayesian mixture model for partitioning zero-inflated continuous feature vectors

- **Tardella, Luca** (Università di Roma La Sapienza, Italy)
  Serena, Arima (Università di Roma La Sapienza, Italy)
  Pecora,Valentina (Policlinico A. Gemelli, Italy)
  A Bayesian hierarchical model for identifying epitopes in peptide microarray data

- **Taylor, Benjamin** (Lancaster University, UK)
  Fearnhead, Paul (Lancaster University, UK)
  An adaptive sequential Monte Carlo method

- **Teng, Gloria** (University of Canterbury, New Zealand)
  Harlow, Jennifer (University of Canterbury, New Zealand)
  Lee, Dominic (University of Canterbury, New Zealand)
  Sainudiin, Raazesh (University of Canterbury, New Zealand)
  Statistical regular sub-pavings for multivariate density estimation

- **ter Horst, Enrique** (Euromed Management, France)
  Rodriguez, Abel (University of California, Santa Cruz, USA)
  Gzyl, Henryk (IESA, Venezuela)
  Molina, German (Tudor Hedge Fund, UK)
  Stochastic volatility models including open, close, high and low prices

- **Thomas, Andrew** (Carnegie Mellon University, USA)
  Blitzstein, Joseph (Harvard University, USA)
  Marginally specified hierarchical models for relational data

- **Tokdar, Surya** (Duke University, USA)
  A Goodness of fit test of the normal distribution against a Dirichlet process mixture alternative

- **Tomazella, Vera L.** (Universidade Federal de São Carlos, Brazil)
  Bernardo, José M. (Universitat de València, Spain)
  Reference analysis and precise hypothesis testing of the Hardy-Weinberg equilibrium

- **Torrado, Nuria** (Universidad Carlos III de Madrid, Spain)
  Wiper, Michael (Universidad Carlos III de Madrid, Spain)
  Lillo, Rosa (Universidad Carlos III de Madrid, Spain)
  Software failure prediction using metrics information via Gaussian process models.

- **Torres-Avilés, Francisco** (Universidad de Santiago de Chile, Chile)
  González, Nelly (Universidad de Santiago de Chile, Chile)
  Carrasco, Elena (Universidad de Santiago de Chile, Chile)
  A temporal Bayesian model with discrete response: An application in epidemiology

- **Tuyl, Frank** (University of Newcastle, Australia)
  Credible intervals that should impress frequentists

- **Valenzuela, Carmen** (Centro Ing. Genética y Biotecnología, Cuba)
  García, Elizeth (Centro Ing. Genética y Biotecnología, Cuba)
  Tuero, Angela (Centro Ing. Genética y Biotecnología, Cuba)
  López, Pedro (Centro Ing. Genética y Biotecnología, Cuba)
  Baladron, Idania (Centro Ing. Genética y Biotecnología, Cuba)
  González, Lidia (Centro Ing. Genética y Biotecnología, Cuba)
  García, Idrian (Center for Genetic Eng. and Biotechnology, Cuba)
  Acevedo, Boris (Centro Ing. Genética y Biotecnología, Cuba)
  Perea, Silvio (Centro Ing. Genética y Biotecnología, Cuba)
  Bayesian adaptive designs of clinical trials. Example of CIGB-300 in Phase I studies.

- **van der Linde, Angelika** (Universität Bremen, Germany)
  Functional data analysis with latent variables

- **van Dyk David** (University of California, Irvine, USA)
  Park, Taeyoung (University of Pittsburgh, USA)
  Metropolis Hastings with partially collapsed Gibbs samplers with application in high-energy astrophysics

- **Vatsa, Richa** (Trinity College Dublin, Ireland)
  Wilson, Simon (Trinity College Dublin, Ireland)
  The variational Bayes method for inverse regression problems with application to the palaeoclimate reconstruction

- **Velasco, Ciro** (Virginia Tech, USA)
  Leman, Scotland (Virginia Tech, USA)
  Bayesian Spatial Model Selection

- **Ventura, Laura** (Università degli Studi di Padova, Italy)
  Racugno, Walter (Università degli Studi di Cagliari, Italy)
  Recent advances on Bayesian inference for $P(X > Y)$

- **Venturini, Sergio** (Università Bocconi, Italy)
  Consonni, Guido (Università di Pavia, Italy)
  Moreno, Elías (Universidad de Granada, Spain)
  Testing Hardy-Weinberg equilibrium: An objective Bayesian analysis

- **Vernon, Ian** (Durham University, UK)
  Goldstein, Michael (Durham University, UK)
  Bower, Richard (Durham University, UK)
  Visualising the input space of a galaxy formation simulation

- **Vicente, Leonel** (Instituto Politécnico de Lisboa, Portugal)
  Turkman, Kamil Feridun (Universidade de Lisboa, Portugal)
  Hierarchical Bayesian models for manpower planning

- **Vidal, Ignacio** (Universidad de Talca, Chile)
  Bolfarini, Heleno (Universidade de São Paulo, Brasil)
  Posterior modal estimation in elliptical measurement error models

- **Vlasakakis, Georgios** (MRC Human Nutrition Research, UK)
  Bluck, Leslie (MRC Human Nutrition Research, UK)
  Jackson, Sarah (MRC Human Nutrition Research, UK)
  Mander, Adrian (MRC Biostatistics Unit, UK)
  A Bayesian hierarchical ^{13}Carbon-Octanoate breath test to assess gastric emptying

- **Vrousai, Dina** (Trinity College Dublin, Ireland)
  Haslett, John (Trinity College Dublin, Ireland)
  Sampling from the posterior- MCMC. Importance resampling or numerical integration?

- **Vukcevic, Damjan** (University of Oxford, UK)
  Holmes, Chris (University of Oxford, UK)
  Donnelly, Peter (University of Oxford, UK)
  Do rare genetic variants lead to greater disease risk? Inferring the relationship between allele frequency and relative risk using replicated disease loci

- **Vuppala, Radhakrishna** (University of California, Santa Cruz, USA)
  Rodriguez, Abel (University of California, Santa Cruz, USA)
  Nested Dirichlet process with shared mixture components

- **Wade, Sara** (Università Bocconi, Italy)
  Mongelluzzo, Silvia (Università Bocconi, Italy)
  Petrone, Sonia (Università Bocconi, Italy)
  A new generalization of the Dirichlet process with application to mixture models

- **Wang, Joanna** (The University of Sydney, Australia)
  Choy, Boris (The University of Sydney, Australia)
  Chan, Jennifer (The University of Sydney, Australia)
  Robust Bayesian analysis of stochastic volatility model with generalised t-distribution via scale mixtures

- **Wang, Shouqiang** (Duke University, USA)
  Banks, David (Duke University, USA)
  Network routing in counterterrorism: An adversarial risk analysis framework

- **Wang, Xiaojing** (Duke University, USA)
  Berger, James (Duke University, USA)
  Burdick, Donald (MetaMetrics, Inc., USA)
  Bayesian dynamic Item response model

- **Wetzels, Ruud** (Univ. van Amsterdam, The Netherlands)
  Grasman, Raoul (Univ. van Amsterdam, The Netherlands)
  Wagenmakers, Eric-Jan (Univ. van Amsterdam, The Netherlands)
  An encompassing prior generalization of the Savage-Dickey density ratio

- **White, Nicole** (Queensland University of Technology, Australia)
  Johnson, Helen (Queensland University of Technology, Australia)
  Silburn, Peter (St. Andrew's War Memorial Hospital, Australia)
  Mengersen, Kerrie (Queensland University of Technology, Australia)
  Bayesian hidden Markov models for the analysis of extracellular recordings

- **Williamson, Daniel** (Durham University, UK)
  Goldstein, Michael (Durham University, UK)
  Decision support for policy makers using the Bayesian analysis of computer models

- **Williamson, Sinead** (University of Cambridge, UK)
  Orbanz, Peter (University of Cambridge, UK)
  Ghahramani, Zoubin (University of Cambridge, UK)
  Dependent beta processes

- **Wilson, Kevin** (Newcastle University, UK)
  Farrow, Malcolm (Newcastle University, UK)
  Bayes linear kinematics in the design of experiments

- **Wilson, Melanie** (Duke University, USA)
  Iversen, Edwin (Duke University, USA)
  Clyde, Merlise (Duke University, USA)
  Model prior choice and multiplicity correction in Bayesian model and variable selection

- **Wilson, Simon** (Trinity College Dublin, Ireland)
  Houlding, Brett (Trinity College Dublin, Ireland)
  Costello, Mark (University of Auckland, New Zealand)
  Bayesian estimation of the number of unknown species: Incorporating a model for the discovery process

- **Wilson, Simon P.** (Trinity College Dublin, Ireland)
  Kuruoğlu, Ercan (CNR Pisa, Italy)
  Yoon, Jiwon (Trinity College Dublin, Ireland)
  Quirós Carretero, Alicia (Universidad Rey Juan Carlos, Spain)
  Source separation for multi-spectral image data with Gaussian mixture priors, with application to the cosmic microwave background

- **Womack, Andrew** (Washington University in St. Louis, USA)
  The posterior predictive information criterion

- **Wraith, Darren** (INRIA Grenoble Rhone-Alpes, France)
  Forbes, Florence (INRIA Grenoble Rhone-Alpes, France)
  Doyle, Senan (INRIA Grenoble Rhone-Alpes, France)
  A Bayesian weighted approach for clustering of spatial and/or temporal data

- **Xu, Chang** (University of Missouri, USA)
  Sun, Dongchu (University of Missouri, USA)
  Bayesian methods of estimating population size under type I censoring

- **Yang, Hongxia** (Duke University, USA)
  Dunson, David (Duke University, USA)
  O'Brien, Sean (Duke University, USA)
  Nonparametric Bayes stochastically ordered latent class models

- **Yen, Ester** (Institute of Information Management, Taiwan)
  Chung, Chun-Fu (Institute of Information Management, Taiwan)
  Novel algorithms for Bayesian hierarchical clustering

- **Yıldırım, Sinan** (University of Cambridge, UK)
  Singh, Sumeetpal (University of Cambridge, UK)
  Doucet, Arnaud (University of British Columbia, Canada)
  Offline and online expectation maximization via forward smoothing in changepoint systems

- **Yoon, Ji-Won** (Trinity College Dublin, Ireland)
  Wilson, Simon (Trinity College Dublin, Ireland)
  Efficient Gibbs sampler approaches for reconstructing multidimensional NMR spectra

- **Yoshimura, Kenichi** (Kyoto University Hospital, Japan)
  A Bayesian analysis for biomarker-adaptive threshold design

- **Zwiernik, Piotr** (Warwick University, UK)
  Asymptotic model selection and identifiability of directed tree models with hidden variables

Abstracts

(by alphabetic order of first-named author)

Bayesian Analysis of the Phylogeography of Bacteria L. Pneumophila in the Valencia Region of Spain

Juan J. Abellán
CSISP Valencia, Spain
abellan_jua@gva.es

Mireia Coscollá
UPF-CSIC, Barcelona, Spain
mireia.coscolla@upf.edu

Fernando González
CIBERESP Spain
gonzalef@uv.es

Phylogeography is defined as "the field of study concerned with the principles and processes governing the geographical distributions of lineages, especially those within and among closely related species" (Avise JC (2000) Phylogeography: the History and Formation of Species. Harvard University Press, London). In the last years several papers have been published dealing with the phylogeography of rapidly evolving viruses with high public health impact such as HIV, dengue, rabies, human papilomavirus, etc. This reflects the need to comprehend better the spatial, temporal and evolutionary processes underlying disease dynamics and (epidemics) propagation, in order to improve strategies for control and prevention. Bayesian models are used in several of those analyses due to their flexibility to integrate the different processes involved in virus dissemination.

The phylogeography of bacteria, as opposed to that of viruses, is rather novel however. Here we present a Bayesian analysis of the phylogeography of circa 300 environmental and clinical samples of bacteria *Legionella Pneumophila*, which causes pneumonia, collected between 1998 and 2006 at several points of the Valencian Region in Spain. We use data on genetic sequences from six genes, as well as the location and date of collection.

We show the usefulness of this Bayesian approach to address this type of analysis. In particular, we illustrate how our Bayesian hierarchical model is able to integrate the spatial, temporal and evolutionary information to make inference about the propagation dynamics of bacteria *Legionella Pneumophila*.

Generative Modeling of Probability Densities with Gaussian Processes

Ryan P. Adams
University of Toronto, Canada
rpa@cs.toronto.edu

The Gaussian process is an appealing prior for Bayesian modeling of functions. Unfortunately, applying the GP to probability density functions has been difficult due to the requirement of integrating over an infinite-dimensional random function. This difficulty is similar to that encountered when attempting to compute the partition function of undirected graphical models. I will discuss a new GP-based prior for probability density functions that helps overcome these problems, by enabling simulation of exact and exchangeable data from the model. With this capability, it is possible to take advantage of recent developments in Markov chain Monte Carlo to perform inference without intermediate approximation of intractable integrals. I will also discuss how this model can be adapted to the inhomogeneous Poisson process and the machine learning task of semi-supervised learning.

Is it Rare or Common? A Coalescent Tree-based Bayesian Approach to Identify the Genetic Types of Variants Underlying Complex Diseases

Kaustubh Adhikari and Christoph Lange
Harvard University, USA
kadhikar@hsph.harvard.edu clange@hsph.harvard.edu

An important problem in modern genetic studies is whether a given disease is caused by a few underlying common variants or by several rare variants in a gene. This paper presents a unique method of identifying the type of variants in a gene segment (marked for detailed analysis through candidate-gene, functional pathway or GWAS study) responsible for a disease through SNP genotyping in that segment.

Based on case-control data of SNP genotypes in a gene segment, the method obtains joint posterior distribution of the number of common and rare variants present in the segment (phenocopy is allowed). We do a Bayesian modeling using coalescent genealogical trees to model the unknown ancestral history of both the cases and the controls jointly. A particular type of Ancestral Recombination Graphs ñ– minARG –ñ is used for the modeling. Such trees have been used frequently in the past to identify SNPs associated with causal variants, but the complexity grows out of hand with thousands of genome-wide SNPs. This method, applied to a candidate-gene segment, utilizes the ability of ARGs to model the ancestral structural history (with mutation and recombination) by employing the information present in the SNP markers.

Representation and Bayesian Analysis of Integer-Valued Networks

Edoardo M. Airoldi
Harvard University, USA
airoldi@fas.harvard.edu

Mobile carriers such as AT&T and T-Mobile handle several hundred millions calls each week. An important class of questions for these carriers has to do with quantifying the effects of structural aspects of the call graph and its dynamics on outcomes of interest, such as adoption of mobile plan features, or response rates to targeted advertising. These analyses necessitate methods to represent large call graphs, sample targeted sub-graphs with desired properties, and identify and quantify structural aspects of the call graph. In this talk we outline a family of statistical models for integer-valued call graphs. These models provide the simplest step-up in complexity from the random graph model of Erdos, Renyi and Gilbert (1959) by introducing a weak form of dependence among the probability of sampling edges (i.e., exchangeability) that is due to latent node- specific bit strings. Full model specification is obtained by imposing dependence among the bits at each node. Dependence among the bits can be induced by generating sets of dependent probabilities from a family of distributions on the unit hypercube, and then generating the bits independently given these dependent probabilities. These exchangeable graph models help focus the analysis, whether empirical or theoretical, on the interplay between connectivity of a graph and node- specific sources of variability. In theory, they provide a new tool to explore variability of graph connectivity. In practice, they lead to an information-based approach to goodness-of-fit and model comparison for models of call graphs, and to assess complexity of observed call graphs.

Approximate Bayesian Inference for Non-homogeneous Poisson Processes with Application to Survival Analysis

Rupali Akerkar
Norges Teknisk-Naturvitenskapelige Universitet, Norway
akerkar@math.ntnu.no

A common class of models in survival analysis are based on non-homogeneous Poisson processes. One example is when multiple events are observed on each individual like in various medical studies, for example multiple attacks of cardiac arrest. We decompose the hazard-rate in the Poisson process into a product of the base-line hazard, which we model using a piecewise constant model, and a term which include the effect of various covariates, and frailty to take care of heterogeneity among the intensity processes in different individuals. Using this formulation, we demonstrate that we can rewrite the model into a latent Gaussian model which allow us to do the Bayesian inference using integrated nested Laplace approximations (Rue, Martino and Chopin, 2009). The big benefit is computational speed, as most models do not require more than a few seconds to run, but also accuracy in the results as the errors in the approximation are relative

and not additive as with Monte Carlo based inference. We illustrate our approach using the well known data set on rat tumours of Gail, Santner and Brown (1980) where we also show comparison using approximate inference and MCMC, which support our claims. This is joint work with Sara Martino and Håvard Rue.

Bayesian State Space Modeling of Mortgage Default Risk

Tevfik Aktekin
The George Washington University, USA
taktekin@gwu.edu

Refik Soyer and Feng Xu
Georgia Southwestern State University, USA
soyer@gwu.edu engxu@gwmail.gwu.edu

In this paper, we consider discrete time Bayesian state space models with Poisson measurements to model the mortgage default risk at the aggregate level. Managing risk at the aggregate level is crucial for banks and financial institutions as required by the Basel II framework which encourages banks to identify and manage present and future risks. At the aggregate level, in modeling the mortgage default risk issues such as whether the default rate exhibits static or dynamic behavior and the effect of macroeconomic variables on default risk are of concern to both practitioners and researchers. In addressing these issues, we propose Bayesian state space models whose rates are evolving over time and as a function of covariates. In doing so we present updating and estimation of system parameters using Markov chain Monte Carlo methods. In assessing the dynamic nature of the mortgage default rate, we compare the forecasting performance of the proposed models with a Bayesian Poisson regression model used as a benchmark. In order to show the implementation of our models we use actual U.S. residential mortgage default data and discuss implications of the proposed models.

Probabilistic Projections of the Total Fertility Rate using a Bayesian Hierarchical Model

Leontine Alkema
National University of Singapore, Singapore
alkema@nus.edu.sg

Adrian E. Raftery
University of Washington, USA
raftery@u.washington.edu

Patrick Gerland
UN Population Div., USA
gerland@un.org

Samuel J. Clark
University of Washington, USA
work@samclark.net

Thomas Buettner
UN Population Div., USA
buettner@un.org

We describe a Bayesian projection model to produce country-specific projections of the total fertility rate (TFR) for all countries. It models the evolution of TFR

in three phases: pre-transition high fertility, the fertility transition, and post-transition low fertility. The model for the fertility decline builds on the United Nations Population Division's current deterministic projection methodology, which assumes that fertility will eventually fall below replacement level. It models the decline in TFR as the sum of two logistic functions that depend on the current TFR level, and a random term. A Bayesian hierarchical model is used to project future TFR based on both the country's TFR history and the pattern of all countries. It is estimated from United Nations estimates of past TFR in all countries using a Markov chain Monte Carlo algorithm. The post-transition low fertility phase is modeled using an autoregressive model, in which long-term TFR projections converge toward and oscillate around replacement level. The method is evaluated using out-of-sample projections for the period since 1980 and the period since 1995, and is found to be reasonably well calibrated.

Bayesian Inference for Time-Varying Pair-Copula Constructions

Carlos Almeida and Claudia Czado
Technische Universität München, Germany
calmeida@ma.tum.de cczado@ma.tum.de

This work proposes a flexible Bayesian modelling and framework for posterior inference of a time varying dependence in multidimensional time series. The time dependence model extends the Stochastic Copula Auto Regressive model (SCAR) of Hafner and Manner (2008) from the bivariate to the multivariate case. The extension is operated by pair-copula construction (PCC) in line with Ass et al. (2009). The time-varying parameters are treated as latent variables.

In a Bayesian framework, the parametric space becomes huge because it involves also the latent variables. The properties of the PCC and an adequate prior specification permit to build a Bayesian sampler for the posterior distribution. By using the results of Liu and Sabbati (2000), the sampler for the latent variables is improved by updating overlapped blocks. The proposed sampler is shown to be operational and applicable to real data sets.

Bayesian Mixture Models for Monitoring Reef Health

Clair Alston and Kerrie Mengersen
Queensland University of Technology, Australia
c.alston@qut.edu.au k.mengersen@qut.edu.au

In this paper we look at statistical modelling techniques that are useful in summarising and monitoring the levels of chlorophyll in a reef environment. We focus on the use of a Bayesian mixture model to aid in determining appropriate clustering. These models provide estimates of chlorophyll mean and proportions of levels of interest which serve as a suitable summarisation of these large data sets. The resulting parameter estimates from these models can then be monitored for changes over time using standard control charting techniques.

The usual Gibbs sampling framework of the Bayesian mixture model is extended to account for binned data. This model involves the addition of a latent variable which represents simulated values from the believed true distribution at each iteration of the algorithm. The technique results in better model fit and recognition of the more subtle aspects of the density of the data.

Control charts are chosen as a management tool because they are useful for tracking environmental parameters over time. In this paper, they are used to monitor trends in reef water quality, for example, are chlorophyll levels stable, increasing or decreasing over time. They are also useful in detecting out of control events, such as a surge in chlorophyll levels on a particular day, which can then be investigated by managers to determine if there is a problem needing to be addressed.

Control charts are also useful in helping to set compliance levels. As the initial phase of building control charts involves examining the system to determine if it is currently stable it allows realistic limits on "acceptable" levels to be set. The process of building a control chart also involves determining the current level of natural variability, which again helps determine acceptable compliance levels.

Bayesian AHP-Priorities in a Global Context

Alfredo Altuzarra
Universidad de Zaragoza, Spain
altuzarr@unizar.es

Pilar Gargallo
Universidad de Zaragoza, Spain
pigarga@unizar.es

José María Moreno
Universidad de Zaragoza, Spain
moreno@unizar.es

Manuel Salvador
Universidad de Zaragoza, Spain
salvador@unizar.es

This work extends the Bayesian prioritization procedure suggested for use in a local context (one criterion) by Altuzarra et al. (2007) to a global context (a hierarchy). The procedure involves the systematic treatment of the uncertainty inherent in the judgment elicitation process associated with the pairwise comparison matrices used in the Analytic Hierarchy Process (AHP). In addition, we propose different measures for evaluating the influence of a specific set of criteria in the final priorities of the global Bayesian prioritization procedure. The methodology is illustrated by means of a numerical example.

Poisson Model with a Hidden Markov Structure for the Detection of Influenza Outbreaks

Rubén Amorós
Universitat de València, Spain
ruasal@alumni.uv.es

David Conesa
Universitat de València, Spain
David.V.Conesa@uv.es

Antonio López-Quílez
Universitat de València, Spain
Antonio.lopez@uv.es

Miguel-Angel Martínez-Beneito
Centro Superior de Investigación en Salud Pública, Spain
martinez_mig@gva.es

Considerable effort has been devoted to the development of statistical algorithms for the automated monitoring of influenza surveillance data. This work introduces a Bayesian hierarchical model with a hidden Markov structure that allows to early detect the onset of an influenza epidemic. In particular, the process of the observed cases is modeled as a Poisson distribution being its intensity parameter a function of the incidence rate. The key point is to consider this incidence rate as a truncated normal distribution in which both parameters are modeled differently depending if the system is in epidemic or non epidemic phase. Transition between both phases is modelled as a Markovian process. Bayesian inference is carried out to provide the probability of being in an epidemic state at any given moment. Methodology is evaluated in terms of sensitivity, specificity and timeliness, on influenza illness data obtained from different epidemiologic Sentinel Networks. Results indicate that our method outperforms some of the most popular ones.

Bayesian Forecasting Models for Traffic Management Systems

Osvaldo Anacleto-Junior and Catriona Queen
Open University, UK
O.Anacleto-Junior@open.ac.uk C.Queen@open.ac.uk

Many roads have real-time traffic flow data available which can be used as part of a traffic management system. In a traffic management system, traffic flows are monitored over time with the aim of reducing congestion by taking actions, such as imposing variable speed limits or diverting traffic onto alternative routes, when problems arise. Reliable short-term forecasting models of traffic flows are crucial for monitoring traffic flows and, as such, are crucial to the ultimate success of any traffic management system.

The model used here for forecasting traffic flows uses a directed acyclic graph (DAG) in which the nodes represent the time series of traffic flows at the various data collection sites, and the links between nodes represent the conditional independence and causal structure between flows at different sites. The DAG breaks

the multivariate model into simpler univariate components, each one being a dynamic linear model. This makes the model computationally simple, no matter how complex the traffic network is, and allows the forecasting model to work in real-time, as required by any traffic management system.

This poster will report current research in the development of this class of model with particular reference to a busy motorway junction in the UK.

Bayesian Shared Spatial-Component Models to Strengthen Inference in Disease Mapping from Multiple Disease-Related Surveillance Sources

Sophie Ancelet
French National Institute for Agricultural Research, France
sophie.enjalric@agroparistech.fr

Juan Jose Abellan
CIBERESP-CSISP, Spain
juan.j.abellan@uv.es

Victor Del Rio Vilas
Environ. and Rural Affairs, UK
vdelriovilas@yahoo.co.uk

Colin Birch
Veterinary Labs. Agency, UK
c.birch@vla.defra.gsi.gov.uk

Sylvia Richardson
Imperial College, UK
sylvia.richardson@imperial.ac.uk

Bayesian hierarchical models have recently flourished in geographical epidemiology. Methodological developments have mainly focused on describing the spatial and spatio-temporal variations of disease risks using a single data source. However, it is not rare to have more than one surveillance source (e.g., incidence/mortality register and health surveys) informing about the spatial distribution of a same disease risk. The contrasting processes generating different data sources may complicate and discourage their combination into a coherent global statistical model to exploit all the information that is available on a single disease. In this context, we present two novel extensions of the class of so-called Bayesian shared component models to strengthen inference in the analysis of the geographical variation of disease risk, when several sources of information are available for (i) one disease (ii) two diseases potentially sharing common disease determinants. Our joint models split the risk of disease not explained by individual exposure factors into one (or two) shared spatial components and specific unstructured components that control the possible differential between the multiple sources for each disease. We applied our models to jointly analyse the spatial variation of risk of two forms of scrapie, an animal disease, from multiple surveillance data collected in Wales (UK). Our joint models allowed to clearly identify clusters of either type of scrapie. This case study illustrates, especially in the presence of sparseness, the benefits of the joint analyses compared to separate analyses: reduced uncertainty and better goodness-of-fit, predictive abilities and robustness to prior choice.

The Effect of the Nugget on Gaussian Process Emulators of Computer Models

I.Andrianakis and P.G.Challenor
National Oceanography Centre, UK
I.Andrianakis@soton.ac.uk P.Challenor@noc.soton.ac.uk

Building a Gaussian Process based computer model emulator requires the estimation of a number of parameters, such as regression coefficients, a scaling factor, and the parameters of the correlation function. A nugget term can be included when the model outputs are stochastic, or when some of the inputs are considered inactive, and therefore ignored. However, a small, fixed nugget term can also be added to the correlation function, even when none of the above conditions hold, simply because it improves the numerical stability of the computations. The effect of the nugget in this case is considered in the present work. Specifically, we demonstrate that the inclusion of a nugget term introduces a second mode in the likelihood of the correlation function parameters (correlation lengths), which is typically found at large correlation length values. Using such large values, results in an emulator that does not interpolate the training points, but rather approximates them, treating local variation as noise. In this work we present a correlation length prior that avoids the approximating mode, and favours the mode that interpolates the training data. The prior can be useful for devising posterior sampling schemes, or when searching for the location of the interpolating mode. Theoretically, using a nugget implies that the training data are in any case approximated, since the nugget term models the existence of noise in the observations. Practically, we require that this level of noise is very small, so that the approximation is satisfactory. We conclude this work by proposing methods for evaluating the quality of this approximation, and strategies for remedying the situation if necessary.

EMB: A Combination of EM-Algorithm and Bayesian Classification Model for Gene Identification

Marilia Antunes and Lisete Sousa
Universidade de Lisboa and CEAUL, Portugal
marilia.antunes@fc.ul.pt lmsousa@fc.ul.pt

In experiments involving DNA microarrays, the objective is usually to identify genes with differential expression levels. A current approach consists of assuming that the data, usually intensity ratios or log-intensity ratios, come from a mixture of populations. In most of the cases, genes that are differentially expressed are not known a priori and hence the problem can be seen as an unsupervised classification task. In such scenario, EM-algorithm has proven to be a very useful tool (Dean and Raftery, 2005). The main advantage of the method is that it converges easily and no prior information is needed. The disadvantage is that the method does not produce a classification rule.

An alternative consists in considering a Bayesian hierarchical model for the data, still assuming data coming from a mixture model, and to classify the genes in

the group for which the predictive conditional probability is the highest (Antunes and Sousa, 2008). Moreover, this method produces a simple classification rule. As a disadvantage, it requires pre-classification of the genes.

EMB is an approach that combines the two methods: EM-algorithm is first used to classify the genes and this information is used as input for the Bayesian classifier. The method gathers the advantages of both methods, none of the disadvantages, and has proved to produce better results than each of the methods used alone.

Bayesian Adaptive Estimation Using a Sieve Prior

Julyan Arbel
Université Paris Dauphine, France
julyan.arbel@ensae.fr

We study the Bayes estimation of an infinite dimensional parameter from a Sobolev smoothness class. We propose a family of sieve priors whose resulting Bayes estimators are *adaptive optimal*, both in posterior concentration rate and in risk convergence for the inherent distance of the model, denoted d_n (e.g. the Hellinger distance). This result is applied to several examples, including white noise model, regression function and density model.

Sieve priors were introduced by Zhao (*Ann Statist*, 2000), and recently used by Rivoirard and Rousseau (2009). A sieve prior on the parameter θ of a model $(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, P_\theta^{(n)} : \theta \in \Theta)$ has the additive form $\Pi = \sum_k \lambda_k \Pi_k$. The λ_k's form the prior weights on the cutoff parameter k, and the Π_k's are prior distributions of θ conditional to k, that is on the first k coordinates of θ.

The main theorem we state is in the form of the results of Ghosal and van Der Vaart (*Ann Statist*, 2007), where we make assumptions explicit in the case of the sieve prior. They are written in terms of distance d_n and of $P_\theta^{(n)}$ (in terms of Kullback-Leibler divergence). In addition, we show that in the white noise model, global and local risks of the same estimation procedure differ of a power n term. In Cai, Low and Zhao (*Bernoulli*, 2007), those rates are shown to differ of at least a $\log(n)$ term.

Combining Bayesian Test Procedures

Reinaldo B. Arellano-Valle
Pontificia Universidad Catolica de Chile, Chile
reivalle@mat.puc.cl

Cristiano C. Santos and Rosangela H. Loschi
Universidade Federal de Minas Gerais, Brazil
criscarvalhosan@yahoo.com loschi@est.ufmg.br

Jeffreys and Pereira-Stern (or Full Bayesian Significance Test) procedures for testing provide measures of evidence in favour the null hypothesis. They are both

Bayesian tests, say, they minimizes the posterior expectated loss function. Jeffreys and Pereira-Stern measures of evidence are both usefull posterior summaries and, in general, they lead to the same decision. Different decisions are expected whenever the null hypothesis is precise and improper priors or conjugate priors with variance going to infinity are elicited to describe the prior uncertainty about the parameter. Under such priors, Jeffreys test can lead to the Jeffreys-Lindley paradox which is overcome if the FBST is assumed. We introduce two measures of evidence for the null hypothesis H_0 which consist of pooling the Pereira-Stern one and $P(H_0|\mathbf{x})$. To aggregate these two measures of evidence, we consider the linear and the logarithmic operators. Such operators have been widely used in Group Decision Theory in order to obtain consensus probability measures. Thus, these proposed procedures are intermediate measures of evidence for H_0. We verify the existence of a loss function which renders decision theoretic aspects to the test procedure built assuming the linear operator. That is, we prove that it is also a Bayesian test. We also prove that this procedure does not lead to the Jeffreys-Lindley paradox. Although we could not prove that the procedure constructed taking into consideration the logarithmic operator is a Bayesian test, we verify that it is a generalization of Jeffreys test. All four procedures are applied to test precise hypothesis in some probability families, including the skew-normal one. In such a family, inference for the shape parameter considering the usual maximum likelihood approach has some problems. Particularly, the information matrix for the skew-normal distribution is singular whenever the skewness parameter is zero, preventing the use of likelihood-based methods for testing normality. Thus, as a by- product of the proposed methodologies, we obtain normality tests under the standard skew-normal distribution.

Bayesian Density Estimation by Mixture Models Using Epsilon-Approximations of Homogeneus Normalized Random Measures

Raffaele Argiento
CNR-IMATI, Italy
raffaele@mi.imati.cnr.it

Alessandra Guglielmi
Politecnico di Milano, Italy
alessandra.guglielmi@polimi.it

In this paper a class of discrete nonparametric priors is studied. Following a constructive approach, we show that this class approximates homogeneous normalized random measures. A normalized random measure can be viewed as a discrete probability measure: the weights are obtained by normalization of the countable jumps of a Poisson process, while the support consists of a countable number of random points. In this case, the posterior inference is complicated by the presence of infinite unknown parameters. Such problem can be avoided by considering only jumps larger than a threshold ϵ, which turns out to control the approximation to the infinite-parameters prior. In particular, we will consider a random ϵ, letting the data to choose the degree of approximation. We illustrate our results through a density estimation problem, when the prior is a mixture of

parametric densities with an ϵ-approximation of a normalized generalized gamma (ϵ-NGG) process as mixing measure. The ϵ-NGG process is identified by a parametric distribution and a pair of positive parameters ($sigma$,$kappa$). In order to robustify the inference, a bivariate prior distribution for the parameters (σ,κ) is assumed. Using a Gibbs sampler algorithm, a Markov chain on the parameters space is built. We do not analytically integrate out the mixing component, but impute the ϵ-NGG and update it as a component of the Gibbs sampler. In fact, conditionally to ϵ, only a finite number of jumps has to be considered, so that the support of the mixing measure becomes finite. We pursue a full nonparametric Bayesian inference, obtaining posterior estimates of linear and non linear functionals of the population distribution.

Bias in the Estimation of Tectonic stress from Earthquake Focal Mechanisms

RICHARD ARNOLD and JOHN TOWNEND
Victoria Unversity of Wellington, New Zealand
richard.arnold@vuw.ac.nz john.townend@vuw.ac.nz

The stress in the earth's crust is a three-dimensional symmetric tensor, the estimation of which leads to some interesting statistical and geomechanical problems. Most methods of estimation rely on sets of earthquakes, each characterised by a pair of orthonormal vectors: the normal to the fault plane $\widehat{\mathbf{n}}$ and the slip vector in that plane $\widehat{\mathbf{u}}$.

Standard estimation methods for the orientations of the principal stresses (the eigenvectors of the stress tensor) adjacent to a major fault (such as the San Andreas fault in California) usually yield an estimate for the axis of maximum compressive stress that lies at 45° to such a fault.

Bayesian estimation (e.g. Arnold and Townend 2007) also prefers this 45° direction if a uniform prior for the orientation of earthquake fault normals $\widehat{\mathbf{n}}$ is assumed.

All methods may therefore lead to significantly biased results in the vicinities of major faults, with important ramifications for the physics of earthquake faulting, if the distribution of fault normals is not uniform. Such non-uniformity can be physically motivated by considering the effects of friction within fault populations.

In this paper, we investigate the extent and significance of this estimation bias, and methods by which it can be ameliorated, both in the analogous Lighthouse toy problem (e.g. Gull 1988, Sivia 1996) and in the full tectonic stress estimation problem.

Modelling and Inference for Networks with Repairable Redundant Subsystems

Louis J.M. Aslett and Simon P. Wilson
Trinity College Dublin, Ireland
louis@maths.tcd.ie swilson@tcd.ie

We consider the problem of modelling the reliability of a network of subsystems where each subsystem has redundancy and is repairable. The motivation for this work is large-scale telecommunications networks.

The time to failure of the subsystem hardware is modeled by an appropriate Markov process and is hence a phase-type distribution. The network structure defines a failure rule in terms of the states of the subsystems, allowing computation by Monte Carlo simulation of the time to failure distribution for the network. When data on the reliability of the subsystems are available, this can be incorporated via modifications to an existing Bayesian inference approach to update the prediction of network reliability.

A Bayesian Assessment of Decision Uncertainty in the Choice between Triptans for the Treatment of Migraine in Finland

Christian Asseburg
University of Eastern Finland
christian.asseburg@uef.fi

Joakim Ramsberg
I3 Innovus, Sweden

Piia Peura
University of Eastern Finland

Martin Henriksson
Astra Zeneca AB, Sweden

Timo Purmonen
University of Eastern Finland

Juha Turunen
University of Eastern Finland

Janne Martikainen
University of Eastern Finland

Migraine is a potentially debilitating neurological disorder that manifests usually in recurring severe headaches, nausea and altered bodily perceptions. Triptans (serotonin 5-HT1B/1D receptor agonists) were introduced in the 1990s and are used to manage these symptoms in patients who do not achieve sufficient control with NSAIDs or other analgesics. Several triptans are currently marketed for migraine in Finland, but their health-economic impact has not been assessed. Such a study is even more warranted after the expiry of the sumatriptan patent in 2008.

Methods. Using systematic review and Bayesian meta-analysis, we estimated the effectiveness of triptans with regard to three clinical endpoints (2-hour pain-free, recurrence, any adverse event, assumed uncorrelated) from published clinical trial reports, thereby updating an earlier meta-analysis (Ferrari *et al*. Cephalalgia 2002). We modelled the health-economic consequences of almotriptan (12.5mg), eletriptan (40mg), rizatriptan (10mg), sumatriptan (50 and 100mg)

and zolmitriptan (2.5 and 5mg) from a Finnish societal perspective. Costs included drug acquisition and productivity losses, utility was evaluated using QALY weights (QWB-SA, Sieber *et al.* Headache 2000). The model time horizon was one acute migraine attack (24 hours). In sensitivity analyses different meta-analytic approaches are compared in terms of decision uncertainty (value of information).

Results. Thirty-eight trials were included in the evidence base, of which twenty-four compared an active agent against placebo alone and fourteen included more than one active arm. The highest probability of 2-hour pain-free was estimated for rizatriptan 10mg (0.391, 95% credibility interval: 0.334 to 0.449), and the highest probability of recurrence was also estimated for rizatriptan 10mg (0.391, 95% CrI 0.330 to 0.453). Sumatriptan 100mg was estimated as the cheapest treatment option, with total costs of EUR 20.47 per day (95% CrI EUR 15.60 to 26.82). At an additional cost of EUR 4.65 per day (95% CrI EUR 3.27 to 5.85), rizatriptan 10mg provided a QALY gain of 0.00016 (95% CrI 0.00005 to 0.00028), resulting in an ICER of approximately EUR 28.400 per QALY gained. The other treatments were dominated or extended-dominated. The probability of achieving sustained pain free and no adverse events (SNAE, defined here as the product of the probabilities of 2-hour pain free, no recurrence and no adverse event) was estimated to be highest with eletriptan 40mg (0.177, 95% CrI 0.143 to 0.221), and the number of migraine attacks needed to treat (NNT) with eletriptan 40mg instead of sumatriptan 100mg to achieve an additional SNAE is estimated at 23 (95% CrI 12 to 155).

Conclusions. At a willingness-to-pay per QALY of EUR 28.400 or higher, rizatriptan 10mg appears to provide the best cost-utility, whereas sumatriptan 100mg appears to be the most cost-effective triptan at lower willingness-to-pay per QALY. The ranking of treatments by effectiveness differs if SNAE is used instead of QALY, which suggests a mismatch between available QALY weights and clinical consensus on the gold standardard of treatment success.

A Bayesian Model of NMR Spectra for the Deconvolution and Quantification of Metabolites in Complex Biological Mixtures.

William Astle Maria de Iorio
Imperial College London, UK
wja@ic.ac.uk m.deiorio@ic.ac.uk
Timothy Ebbels & Sylvia Richardson
Imperial College London, UK
t.ebbels@ic.ac.uk s.richardson@ic.ac.uk

Nuclear Magnetic Resonance (NMR) spectroscopy is a technique for obtaining structural and quantitative information about molecules from the resonant radio frequencies of their constituent nuclei in a magnetic field. The chemical environment of a magnetic nucleus perturbs the local magnetic field and this leads to characteristic shifts in and splitting of the absorption frequencies of the nucleus. The aggregated shift and splitting signals from nuclei in multiple chemical groups

can be used for molecular identification while the strength of the resonance signals can be used for molecular quantification. We present a novel model for proton NMR spectra, which allows us to identify and quantify the metabolic compounds present in complex mixtures such as biofluids.

Resonance signals combine additively in NMR spectra so our modelling is based on a Bayesian linear regression. We assume the spectral data are generated with Gaussian error from $\sum_j w_j S_j(\delta) + \eta(\delta)$, sampled at regular intervals of the chemical shift δ (a parameter proportional to the frequency of exposure radiation). Here, S_j is a template function determined by the chemistry of the j^{th} metabolite, through the nuclear shifts and multiplet splittings. We aim to deconvolve a spectrum into components corresponding to individual metabolites, which will allow us to make inference about the concentration parameters $\{w_j\}$. We take advantage of substantial prior information about the shape of these templates for selected metabolites. NMR theory implies that each S_j is a mixture of Lorentzian curves, the position, shape and relative heights of which are determined by the number of hydrogen nuclei in the compound and their shift and splitting parameters. The shift and splitting parameters for a particular biofluid can be determined empirically and for many metabolites these data are available from online databases.

To complement our model for the spectral signal generated by compounds with well characterised NMR signatures we model the residual signal, mostly due to unidentified compounds (i.e. those for which no template S_j is available), by a flexible non-parametric component η. We aim to use the parameters w_j and the component η, representing unidentified metabolites, for supervised and unsupervised classification of individuals by their biofluid metabolite profile.

Posterior inference is performed by MCMC methods. We will demonstrate the performance of our model using simulated and real data.

Approximate MCMC Simulation from Doubly-intractable Distributions

Yves F. Atchadé & Jing Wang
University of Michigan, USA
yvesa@umich.edu jingjw@umich.edu

When the likelihood has an intractable normalizing constant that is a function of the unknown parameters, the posterior distribution is doubly intractable. Doubly-intractable distributions arise in many statistical problems, such as Markov random fields, image analysis, protein design, and social network modeling. Existing Markov chain Monte Carlo (MCMC) algorithms for sampling from such distributions are restricted by either strict model assumptions or long computing time. We propose an MCMC algorithm that improves both aspects.

Our sampler is built on the exchange algorithm [Murray *et al.* (2006), UAI] and the auxiliary variable method [Moller *et al.* (2006), Biometrika, 93]. Unlike these two methods, our algorithm does not require exact sampling, which is computationally expensive and impossible for many useful models. We use approximate sampling instead and replace the estimator of the ratio of the normalizing constants with a dynamic version of the path sampling identity [Gelman and Meng (1998), Statistical Science, 13].

We examine the algorithm in a simulation study with various models and a real large-scale friendship network of 1461 students. Results show that our method outperforms existing ones in computing time and retains good mixing properties. We also explore theoretical aspects of the sampler and prove its convergence to the target distribution as well as a law of large numbers. Given its simplicity to implement, the ability to deal with large-scale data, and the flexibility to be applied to most models, our new algorithm will facilitate the modeling with doubly-intractable distributions and reduce computational load.

Bayesian Analysis of Multiple Hypothesis Testing with Applications to Microarray Experiments

Concepción Ausín
Universidad Carlos III de Madrid, Spain
causin@est-econ.uc3m.es

Miguel A. Gómez-Villegas Beatriz González-Pérez
Universidad Complutense de Madrid, Spain
ma.gv@mat.ucm.es beatrizg@mat.ucm.es

Mayte Rodríguez Isabel Salazar and Luis Sanz
Univ. Complutense de Madrid, Spain
{mayter, isalazar, luis_sanz}@mat.ucm.es

Recently, the field of multiple hypothesis testing has experienced a great expansion, basically because of the new methods developed in the field of genomics. These new methods allow scientists to simultaneously process thousands of hypothesis tests. The frequentist approach to this problem is made by using different testing error measures that allow to control the Type I error rate at a certain desired level. Alternatively, in this paper, a Bayesian hierarchical model based on mixture distributions and an empirical Bayes approach are proposed in order to produce a list of rejected hypotheses that will be declared significant and interesting for a more detailed posterior analysis. In particular, we develop a straightforward implementation of a Gibbs sampling scheme where all the conditional posterior distributions are explicit. The results are compared with the frequentist False Discovery Rate (FDR) methodology. Simulation examples show that our model improves the FDR procedure in the sense that it diminishes the percentage of false negatives keeping an acceptable percentage of false positives.

Adaptive Sequential Monte Carlo for Optimal Bayesian Experimental design

Nammam A. Azadi
Kurdistan University of Medical Sciences, Iran
n.azadi@lancaster.ac.uk

Gareth Ridall and Paul Fearnhead
Lancaster University, UK
g.ridall@lancaster.ac.uk p.fearnhead@lancaster.ac.uk

Threshold tracking is an electro-physiological technique formulated by Bostok *et al. Muscle Nerve*, (1998, vol 21(2)) to study the properties of human nerves. Geleijnse, *University Medical Center Rotterdam*, PhD thesis, (2008) is currently using the same technique for studying the excitability properties of single axons where the response is binary. Currently the stimuli to estimate these properties are collected according to certain rules. We develop a sequential Monte Carlo method for the selection of single stimuli assuming a squared loss function on the percentile or parameter of interest. The particles are allowed to move stochastically using the scheme of Lui and West. We compare the efficiency of our technique to standard practice, Polya urns and several other semi-Bayesian methods.

Bayesian Inference through Hierarchical Framework for an IRT Model with Centre Skew Distribution

Caio L. N. Azevedo
Universidade Estadual deCampinas, Brazil
cnaber@ime.unicamp.br

Heleno Bolfarine
Universidade de São Paulo, Brazil
hbolfar@ime.usp.br

Dalton F. Andrade
Universidade Federal de Santa Catarina, Brazil

Item Response Models (IRM) comprises useful models for psychometrical measurements. One of the most usual assumptions is to consider that the latent traits as random variables and to assign a standard normal distributions to them. In the last years, some alternatives to the normal distribution have been proposed. Among those proposals, one that deserves attention is the skew normal distribution. The main goal of this work is to consider a Henze's stochastic representation to the skew normal IRT model proposed by Azevedo *et al.* (2010). This is done in order to facilitate the inference process. A MCMC algorithm, more specifically a Metropolis-Hastings within Gibbs Sampling algorithm is proposed for parameter estimation. We conduct a simulation study which shows that the estimation algorithm recoveries properly.

Bayesian Analysis of Reversible, Higher-Order Markov Chains

Sergio A. Bacallado and Vijay S. Pande
Stanford University, USA
sergiob@stanford.edu pande@stanford.edu

Reversible Markov chains are central to a number of fields. They underlie problems in applied probability like card-shuffling and queueing networks, and pervade computational statistics through the many variants of Markov chain Monte Carlo; in physics, they are natural stochastic models for time-reversible dynamics. However, the notion of reversibility in stochastic processes with memory is not as widely discussed, and statistical problems like determining the order of a reversible process remain a challenge. We define a conjugate prior for higher-order, reversible Markov chains from a generalized Polya-urn, which extends a prior for reversible Markov chains by Diaconis and Rolles ("Bayesian analysis for reversible Markov chains". *Ann. Stat.*, 34, p. 1270, 2006). The prior facilitates Bayes factor tests for the order of a reversible process. An extension to variable-order, reversible Markov chains enriches our model space, providing a parsimonious alternative to full high-order specifications. The prior is also endowed with a natural sampling scheme.

Hierarchical Bayesian Modeling in Syndromic Surveillance

David Banks
Duke University, USA
banks@stat.duke.edu

Gauri S. Datta
University of Georgia, USA
gaurisdatta@gmail.com

Matt Heaton
Duke University, USA
matt@stat.duke.edu

Alan Karr
NISS, USA
karr@niss.org

James Lynch
University of South Carolina, USA
lynch@stat.sc.edu

Francisco Vera
Clemson University, USA
veraf@clemson.edu

Frank Zou
NISS, USA
frankzou@niss.org

Syndromic surveillance is an emerging research area. Quick detection of emerging geographical clusters of disease is important to provide swift intervention to prevent a pandemic. To overcome the lag inherent in traditional public health notification structures, the Centers for Disease Control plans to implement a monitoring system to which all hospitals in USA report, in real or near-real time, admissions associated with various symptoms.

Syndromic surveillance has grown opportunistically, relying on cusum charts, regression, exponentially weighted moving averages and standard time series models. A combination of time series and control charts has been used for early detection of anthrax outbreaks by tracking over-the-counter medication sales. A statistically advanced procedure, the scan statistic, has been used to detect hotspots in which disease reports are significantly higher than expected. And spatio-temporal scan statistic are used to detect disease clusters in time and space. All of these methods use frequentist statistics, and suffer from issues of multiple testing, poor interpretability, high false alarm rates versus low power, and inadequate use of covariate information. A hierarchical Bayesian approach that is presented in this poster that assesses the posterior probability of an that there has been a disease outbreak in a particular location at a particular time. Our method successfully resolves all of the difficulties faced by the frequentist solutions. We use a Poisson distribution to model the disease counts and a log-linear spatio-temporal hierarchical model to assess the posterior probability of an epidemic at each reporting hospital.

A Bioclimatic Classification Method Using a Conditional Corregionalized Linear Model

Xavier Barber
Universidad Miguel Hernandez, Elche, Spain
xbarber@umh.es

Javier Morales Asunción Mayoral
Universidad Miguel Hernandez, Elche, Spain

Antonio Lopez-Quilez
Universitat de València, Spain

Antoni Barber
DENTIA Institute, Spain

Bioclimatology is essential for the comprehension of the close relationship between climate and vegetation, and therefore, configuration of the landscape. The main aim of this work is to establish a Bioclimatic classification of the Island of Cyprus, based on the current Worldwide Bioclimatic Classification System (Rivas-Martinez, 2004). The Bayesian spatial modeling will serve to improve identification of bioclimates and bioclimatics belts through predictions on the bioclimatics indexes continentality (CI) and ombrothermicity (OI).

We propose a conditional corregionalized Bayesian linear model that makes use of the reparametization of the variance. We stablish new default priors for the variances on this multivariate bioclimatic model.

We obtain inferences on the posterior models parameters and predict the OI and CI indexes over the whole island. The Bayesian framework provides interesting summaries as probabilities for any location to be classified within any Bioclimatic-subtype: the Spatial Distribution Function(Barber, 2009).

This models allows us to easily incorporate the relationship between OI and CI indexes, so improving results from the univariate modeling.

Post-Processing of Factor Loadings in Bayesian Factor Analysis

Bradley J. Barney
Texas A&M University, USA
bjbstat@stat.tamu.edu

Valen E. Johnson and Charles S. Cleeland
M. D. Anderson Cancer Center, USA
vejohnson@mdanderson.org ccleeland@mdanderson.org

It is widely recognized that without constraints or prior assumptions, factor loadings (and factor scores) are not identifiable in a factor analysis. It is common to assume a priori that certain elements of the factor loading matrix are zero while some elements are assumed to be positive (e.g., Lopes and West 2004, Fahrmeir and Raach 2007). Lopes and West (2004) found that inference on the number of common factors can be affected by such a prior assumption (pp. 45–46). Without substantive prior knowledge regarding the manifest variables that load highly on each factor and the relative importance of each factor, the exact constraints used might be unduly influential. We propose a slight modification of commonly used priors to eliminate such prior constraints on the factor loadings. We advocate post-processing of factor loadings and factor scores to identify useful summaries. In particular, an MCMC algorithm can be used to simulate draws from the posterior distribution, and then the draws of factor scores and factor loadings can be rotated to achieve a particular structure for the loadings matrix, such as lower triangular. Because any permutation of the ordering for the manifest variables has an associated factor loading matrix that is lower triangular (Lopes and West 2004, p. 45), this technique serves to mitigate the effect of which elements are constrained to be zero/positive. Many permutations can be used, in each case identifying a lower triangular structure for the associated factor loading matrix draws, to determine which of various arbitrary constraints needed to obtain identifiability produce more readily summarized results. We illustrate this method using simulated and real data. This research was supported by NIH grant P01 CA124787, Cleeland (PI).

Using Shifts in Amino Acid Frequency or Substitution Rate to Identify Latent Structural Characters Improves our Understanding of the Structure, Function and Evolution of Base-Excision Repair Enzymes

Ramiro Barrantes and Jeffrey Bond
University of Vermont, USA
ramiro.barrantes@uvm.edu jeffrey.bond@uvm.edu

We describe a method for identification of the gain and loss of structural features in a protein phylogeny. We then apply this method to obtain a more comprehensive understanding of the evolution of protein sequence, structure, and function of a family of DNA glycosylases. Protein structure evolution includes transitions

between character states that are latent in sequence data, for example, the gain or loss of a salt bridge. Our first goal is to annotate the phylogeny of the Fpg/Nei family of base excision repair enzymes with latent structural characters (LSC) that varied during evolution of the family. First, we identified instances in which amino acid frequencies or overall substitution rates change during evolution using a bayesian extension of methods developed by Xun Gu. Second, we found sets of amino acids near each other in the structure exhibiting correlations of such changes. Third, we used these sets of amino acids to identify LSC in clades within the Fpg/Nei phylogeny. We developed a method to identify LSC in any family of proteins. Our methods identified the Fpg substrate specificity loop as an LSC in one clade, quantified conservation in each clade, and predicted that the role of this region in specificity varies substantially throughout the family. Since LSC may evolve more slowly than amino acid identities, they may be useful as phylogenetic characters. Phylogenetic inference based on LSC resolved a key edge in the Fpg/Nei phylogeny, providing convincing evidence of independent losses of Zn fingers.

Delineation of Local Labor Markets. An Application to the Spanish Region of Aragón

Asunción Beamonte
Universidad de Zaragoza, Spain
beamonte@unizar.es

M. Pilar Alonso
Universitat de Lleida, Spain
p.alonso@geosoc.udl.cat

Avishek Chakraborty
Duke University, USA
ac103@stat.duke.edu

Pilar Gargallo
Universidad de Zaragoza, Spain
pigarga@unizar.es

Alan E. Gelfand
Duke University, USA
alan@stat.duke.edu

Manuel Salvador
Universidad de Zaragoza, Spain
salvador@unizar.es

Populations are increasingly mobile. In this regard, it is important to illuminate the role of commuting in explaining employment patterns for a region. In particular, it is essential to understand the new territorial configurations that commuting creates, configurations which we refer to as local labor markets (LLMs). The delineation of these areas has usually been attacked by using algorithmic methods. Our contribution to this problem is to design and fit a suitable hierarchical spatial interaction model.

Our starting point is a minimum market configuration (MMC), obtained through a constrained agglomerative clustering procedure. Given this MMC, we model the joint probabilities of the workers living in one unit, working in another (possibly the same) unit. We apply an initial filter to constrain maximum commuting time. Then, in addition to commuting time, we introduce spatial random effects with regard to residence unit, employment unit along with spatial interaction effects for each pair. We also introduce commuter level covariate information including sex, age, and level of education. Finally, we introduce further spatial random effects to enable interactions between these covariates and residence unit as well as employment unit. We discretize the covariate information in order to

present the entire dataset in the form of a multiway contingency table. Therefore, the joint probabilities are modeled through normalizing conditionally independent Poisson regressions with loglinear models. From these joint probablilities we can obtain any conditional and marginal probabilities of interest, including customary transition probabilities.

Finally, we propose a criterion for merging the MMC configuration into the creation of local labor markets. Then, based upon the model fitting, using suitable posterior probabilities, we develop the final LLM map. We illustrate the methodology with an analysis of the Spanish region of Aragón.

Bayesian Analysis of a Generalized Exponential Smoothing Model

José D. Bermúdez
Universitat de València, Spain
bermudez@uv.es

Ana Corberán-Vallet
Universitat de València, Spain
ana.corberan@uv.es

José V. Segura
Univ. Miguel Hernández, Spain
jvsh@umh.es

Enriqueta Vercher
Universitat de València, Spain
enriqueta.vercher@uv.es

We introduce an extension of exponential smoothing to deal with covariates and double seasonality which could easily be adapted to more than two seasonal cycles. Assuming additive effects and a stochastic component given by independent, homoscedastic, normal errors, the generalized exponential smoothing model can be expressed as an equivalent linear dynamic model with a very peculiar structure of the covariance matrix. This covariance matrix is a function of only the unknown smoothing parameters, while the mean vector only depends on the unknown initial conditions. That characteristic of the model simplifies its statistical analysis. Following the Bayesian paradigm, we obtain the joint posterior distribution of all the unknowns. Only the marginal posterior of the smoothing parameters is analytically intractable and has to be approached using simulation techniques. The conditional distribution of initial conditions giving the smoothing parameters is well known and can be integrated out exactly in order to compute the predictive distribution. Finally, we propose to integrate out the smoothing parameters using Monte-Carlo techniques, obtaining an estimate of the predictive distribution as well as their main characteristics: point forecasts and prediction intervals. We empirically evaluate our forecasting procedure using a real data set.

Integrated Objective Bayesian Estimation and Hypothesis Testing

José M. Bernardo
Universitat de València, Spain
jose.m.bernardo@uv.es

The complete final product of Bayesian inference is the posterior distribution of the quantity of interest. Important inference summaries include point estimation, region estimation, and precise hypotheses testing. Those summaries may appropriately be described as the solution to specific decision problems which depend on the loss function chosen. The use of a continuous loss function leads to an integrated set of solutions where the same prior distribution may be used throughout. Objective Bayesian methods are those which use a prior distribution which only depends on the assumed model and the quantity of interest. As a consequence, objective Bayesian methods produce results which only depend on the assumed model and the data obtained. The combined used of the intrinsic discrepancy, an invariant information-based loss function, and appropriately defined reference priors, provides an integrated objective Bayesian solution to both estimation and hypothesis testing problems. The ideas are illustrated with a large collection of non-trivial examples.

Bayesian Probit Regression Models for Spatially-Dependent Categorical Data

Candace Berrett and Catherine A. Calder
The Ohio State University, USA
cberrett@stat.osu.edu calder@stat.osu.edu

In the Bayesian setting, data augmentation/latent variable strategies are often used to facilitate model specification and fitting. One well-known application of this technique is the introduction of latent response variables in the probit generalized linear model for categorical outcomes. This specification of the probit regression model can be used to analyze spatially-referenced categorical response variables (for example, land-cover/land-use observations where regions of the Earth are classified into one of several land-cover categories such as forest, agriculture, urban, or water); however, the assumption of residual independence required by this model will not be appropriate. In this poster, we show how the data augmentation approach for the Bayesian probit regression model can be readily extended to incorporate standard covariance models for spatially-dependent continuous data. Analogous to the standard (independent) version of the augmented probit regression model, the spatial variance parameter is not identifiable. This identifiability problem can be overcome either by fixing the value of the variance parameter to a constant, as is typically done, or by specifying a proper prior on the spatial variance parameter and reporting inferences on the properly normalized regression coefficients. Within the general class of data augmentation MCMC algorithms, these two approaches can be viewed as conditional and marginal data augmentation strategies, respectively, where the non-identifiable variance serves

as a working parameter. We compare various versions of the conditional and marginal data augmentation MCMC algorithms for the Bayesian spatial probit regression model in terms of MCMC convergence and mixing using both a simulation study and an analysis of satellite-derived land-cover data.

Fusing Space-time Data under Measurement Error for Computer Model Output

Veronica J. Berrocal
SAMSI, USA
vjb2@stat.duke.edu

Alan E. Gelfand
Duke University, USA
alan@stat.duke.edu

David M. Holland
Environ. Protec. Agency, USA
Holland.David@epa.gov

Computer models are widely used in environmental sciences to predict spatio-temporal processes. Given a set of inputs and equations that describe the evolution in time of the underlying physical processes, these sophisticated computer models produce outputs at very high spatial and temporal resolutions, usually given in terms of averages over a large number of grid cells. Validation and calibration of computer model outputs require comparison with observations collected at point level, rather than areal level.

In this paper, we present an approach to combine computer model output and observational data that addresses the difference in spatial resolution between the two sources of data while accounting for uncertainty in the computer model output itself. More specifically, we define a weighted linear regression model where observations at each site are regressed on the entire computer model output itself. The latter is weighted via a spatio-temporal weighting process that assigns to each grid cell a weight that varies in time and that depends on the observation site, thus allowing a local calibration of the computer model output. As an illustration, we have applied our modeling approach to daily ozone concentration for the Eastern United States for the months of June, July and August 2001.

Data-Free Inference of the Joint Distribution of Uncertain Model Parameters

Robert D. Berry
Sandia National Laboratories, USA
rdberry@sandia.gov

Habib N. Najm Bert J. Debusschere
Sandia National Laboratories, USA
hnnajm@sandia.gov bjdebus@sandia.gov

Helgi Adalsteinsson
Sandia National Laboratories, USA
hadalst@sandia.gov

Youssef M. Marzouk
MIT Boston, USA
ymarz@mit.edu

It is known that, in general, the correlation structure in the joint distribution of model parameters is critical to the uncertainty analysis of that model. Very often, however, studies in the literature only report nominal values for parameters inferred from data, along with confidence intervals for these parameters, but no details on the correlation or full joint distribution of these parameters. When neither posterior nor data are available, but only summary statistics such as nominal values and confidence intervals, a joint PDF must be chosen. Given the summary statistics it may not be reasonable nor necessary to assume the parameters are independent random variables.

We demonstrate, using a Bayesian inference procedure, how to construct a posterior density for the parameters exhibiting self consistent correlations, in the absence of data, given (1) the fit-model, (2) nominal parameter values, (3) bounds on the parameters, and (4) a postulated statistical model, around the fit-model, for the missing data. Our approach ensures external Bayesian updating while marginalizing over possible data realizations.

We then address the matching of given parameter bounds through the choice of hyperparameters, which are introduced in postulating the statistical model, but are not given nominal values. We discuss some possible approaches, including (1) inferring them in a separate Bayesian inference loop and (2) optimization.

We also perform an empirical evaluation of the algorithm showing the posterior obtained with this data free inference compares well with the true posterior obtained from inference against the full data set.

MCMC in High Dimensions: A New Perspective

Alexandros Beskos
University College London, UK
alex@stats.ucl.ac.uk

Jonathan C. Mattingly
Duke University, USA
jonm@math.duke.edu

Natesh S. Pillai
University of Warwick
n.pillai@warwick.ac.uk

Gareth O. Robert
University of Warwick
g.o.roberts@warwick.ac.uk

Jesús Sanz
Universidad deValladolid, Spain
sanzsern@mac.uva.es

Andrew M. Stuart
University of Warwick, UK
a.m.stuart@warwick.ac.uk

In this talk we will discuss the efficiency of certain MCMC algorithms in high dimensions. In the first part of the talk we discuss the diffusion limits for Random Walk Metropolis Algorithm, and the in the second part we discuss the optimal tuning for the Hybrid Monte Carlo algorithm.

A pioneering paper in this context is Roberts, Gelman and Gilks (1997) in which diffusion limits are derived for the random walk Metropolis algorithm. For the random walk Metropolis algorithm, they proved that the proposal variance should be scaled as $O(N^{-1})$ in order to get a $O(1)$ acceptance probability, as N-the dimension of the state space, goes to infinity. Furthermore, for this choice of the proposal variance, maximizing the speed measure of the corresponding diffusion (related to the autocorrelation function of the Markov chain) leads to a universal acceptance probability of 0.234, leading to a clear criteria for the practitioners to optimize the performance of the RWM algorithm, for instance, by tuning the proposal to have an acceptance probability of 0.234. However, to date such results have only been proved for target measures with a product structure, severely limiting their applicability to real applications. In the first part of the talk, we will discuss recent results on diffusion limits for a wide class of *correlated* high dimensional measures which occur naturally in real applications, in the context of statistical inference for data coming from physical processes such as the heat flow, the Navier-Stokes equations or the Wave equation, modeled using partial differential equations.

HMC uses Hamilton's ordinary differential equations for energy conservation, and exploits the information on the derivative of the target measure to deliver guided, *global* moves, of higher acceptance probability. Hamilton's differential equations cannot be solved explicitly in general and are usually discretized using a symplectic scheme such as the leapfrog (Verlet) method. Although quite extensively used by an entire gamut of researchers, as noted by Qin and Liu (2000), to date no rigorous results are available regarding the performance of HMC in high dimensions. We present a theory for choosing the step size as a function of dimension for implementing the HMC algorithm. In particular we show that the step size for the leapfrog scheme should be scaled as $O(N^{-\frac{1}{4}})$ so as to obtain a $O(1)$ acceptance probability. Moreover for this choice of the step size, tuning the proposal to minimize the corresponding computational cost leads to an optimal acceptance probability of 0.651 confirming the predictions made in Roberts, Gelman and Gilks (1997), based on empirical calculations.

Nonparametric Bayes Classification and Testing on Manifolds

Abhishek Bhattacharya and David Dunson
Duke University, USA
dunson@stat.duke.edu

We develop general Bayes methods for density estimation, classification and testing on known manifolds. For example, the manifold may correspond to the surface of a hypersphere or a planar shape space. We propose a general kernel mixture model for the joint density of the response and predictors, with the kernel expressed in product form and dependence induced through the unknown mixing measure. We provide simple sufficient conditions on the prior that lead to L_1 and Kullback-Leibler (KL) support on the space of continuous joint densities for the predictors and response. Focusing on a Dirichlet process prior for the mixing measure, these conditions hold using von Mises-Fisher kernels when the manifold is the unit hypersphere and complex Watson kernels for planar shape spaces. Bayesian methods are developed for efficient posterior computation using an exact block Gibbs sampler. We also develop nonparametric methods for testing for differences between groups in variables having an unknown density on a manifold, with efficient computational methods proposed for Bayes factor calculation. The methods are evaluated using simulation examples and applied to spherical data and shape applications.

Sparse Bayesian Infinite Factor Models

Anirban Bhattacharya
Duke University, USA
ab179@stat.duke.edu
David B. Dunson
Duke University, USA
dunson@stat.duke.edu

We propose a sparse shrinkage prior on factor loadings for modeling large covariance matrices. The proposed multiplicative gamma process shrinkage (MGPS) prior allows introduction of infinitely many factors, with sparsity manifest through an increasing degree of shrinkage with the column index. The shrinkage prior enables efficient posterior computation through block updating of the loadings. Theoretical results are provided on the support of the prior and truncation approximation bounds. The challenging issue of inference on the number of factors is addressed using the shrinkage approach. To make the approach robust with respect to the initial number of factors, an adaptive Gibbs sampler is proposed to automatically select the truncation level. A fast algorithm is proposed for calculating an approximate maximum a posteriori estimate of the covariance matrix. Latent factor regression methods are developed for prediction and variable selection in applications with high-dimensional correlated predictors. Operating characteristics are assessed through simulation studies and the approach is applied

to predict survival after chemotherapy from gene expression data. An extension of the linear factor model based on a flexible non-linear embedding is proposed to achieve higher degree of dimension reduction. The proposed approach places a nonparametric prior on the distribution of the latent factors, while allowing nonlinearity in the unknown function relating the latent factors to the response variables using a Gaussian process prior. Theoretical properties are explored and an approach is developed for factor selection in the non-linear setting. The method is illustrated through simulation examples and an application in compressive sensing.

Sequential Integrated Nested Laplace Approximation

Arnab Bhattacharya and Simon Wilson
Trinity College Dublin, Ireland
bhattaca@tcd.ie simon.wilson@tcd.ie

This work addresses the problem of sequential inference of time series in real time, which will be improved further to deal with spatio-temporal models. The idea is to develop a fast functional approximation scheme so as to perform real-time data analysis of unknown quantities, given observations, which are dependent on some underlying latent variable.

The problem is defined as follows: the observed variables $\mathbf{Y}_t, t \in \mathbb{N}$, $\mathbf{Y}_t \in \mathcal{Y}$ are assumed to be conditionally independent given the latent process $\mathbf{X}_t$ (assumed to be a GMRF) and the unknown hyperparameters Θ, can have any distribution. The primary aim is to estimate the posterior distribution $\mathbb{P}(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}, \theta)$ and also the filtering density $\mathbb{P}(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta)$. The computation of these two terms necessarily requires the estimation of the posterior density of Θ. We are interested in providing sequential solution for both $\mathbb{P}(\theta|\mathbf{y}_{1:t})$ and $(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta)$. The new method is motivated by a recently published technique known as Integrated Nested Laplace Transformation (INLA) developed by by Rue *et al.* 2009.

The procedure has already been implemented on Linear Gaussian state-space models with unknown state of the system and covariance parameters and has proved to be very accurate and fast. We consider implementing it in the generalized case where there is nonlinearity and non-Gaussianity.

Empirical Calibration of p-Values

Mikelis G. Bickis
University of Saskatchewan, Canada
bickis@snoopy.usask.ca

The p -value is a strange animal. While widely used in applied statistics, it has no Bayesian interpretation, nor does it have any relevance to Neyman-Pearson hypothesis testing. It is commonly described as quanitifying the amount of evidence against the null hypothesis, although this interpretation is disputed by Bereger and Sellke (1987). However, one may attempt to calibrate a p -value by transforming it to a posterior probability (Sellke et al., 2001). In the case of multiple p -values, Efron (2007) proposed the local false discovery rate as the (estimated)

posterior probability of the null hypothesis given the p -value. A similar idea was presented by Bickis (2004).

In the case of multiple p -values from related hypotheses, inferences about the proportion of true null hypotheses can be made from the empirical distribution of p -values. Under certain regularity conditions, the posterior probability of the null hypotheses can be calculated as the ratio of slopes of the actual distribution of p -values. Using this property empirically requires a smooth estimate of this distribution.

The p -values can be modelled as arising from normally-distributed test statistics in which the location parameter itself has an underlying distribution, consisting of an atom at zero mixed with a distribution of alternatives. The prior of this distribution of alternatives is modelled as a Dirichlet process. The posterior mean of the distribution of alternatives is then used to calibrate the p -values as posterior probabilities.

Nonlinear Latent Process Models for Addressing Temporal Change of Support in Spatio-Temporal Studies of Environmental Exposures

Nikolay Bliznyuk
Texas A&M University, USA
nbliznyuk@stat.tamu.edu

Christopher Paciorek
Univ. California at Berkeley, USA
paciorek@stat.berkeley.edu

Brent Coull
Harvard Sch. Public Health, USA
bcoull@hsph.harvard.edu

Spatio-temporal prediction of levels of an environmental exposure is an important problem in environmental epidemiology. When multiple sources of exposure information are available, a joint model that pools information across sources maximizes data coverage over both space and time, thereby reducing the prediction error. We consider a Bayesian hierarchical framework where a joint model consists of a set of submodels, one for each data source, and a model for the latent process that serves to relate the submodels to one another. However, if a submodel depends on the latent process nonlinearly, inference using standard MCMC techniques can be computationally prohibitive. To make such problems tractable, we 'linearize' the nonlinear components with respect to the latent process and induce sparsity in the covariance matrix of the latent process using compactly supported covariance functions. We propose an efficient MCMC scheme that takes advantage of these approximations. We then apply our methods to motivating data on the spatio-temporal distribution of mobile source particles in the greater Boston area. We use our model to address a temporal change of support problem whereby interest focuses on pooling daily and weekly black carbon readings in order to maximize the spatial coverage of the study region.

Efficient MCMC Using Metropolis-Hastings

William M. Bolstad
University of Waikato, New Zealand
bolstad@waikato.ac.nz

Markov chain Monte Carlo (MCMC) is a method for drawing a sample from the Bayesian posterior. A Markov chain is set up that has the posterior as its long run distribution, and after a sufficiently long burn-in period a draw from the chain is a draw from the Bayesian posterior. However subsequent draws from the chain are serially dependent. Thinning the sample by only including every k^{th} draw is used so the thinned sample will be approximately a random sample from the posterior. We use the thinned sample for inference since the usual formulae for standard errors apply. We need to determine how long the burn-in period should be, and how much thinning will we use. Efficient MCMC sampling requires that we find a chain with good mixing properties. It should be able to move from any state to any other state very quickly so that the effect of the starting position dies out quickly. It will not require long burn-in nor very much thinning to get an approximate random sample for inference.

There is great flexibility in MCMC sampling. We can set up the Metropolis-Hastings algorithm using a random walk candidate density, Metropolis-Hastings using an independent candidate density, or blockwise Metropolis-Hastings algorithm. (The Gibbs sampling algorithm is a special case of blockwise Metropolis-Hastings.) After we decide on the type of candidate density, we still have flexibility in choosing a particular one.

We will show that the Metropolis-Hastings algorithm with an independent matched curvature Student's t candidate density based on the matching the curvature at the node of the posterior will have excellent mixing properties. We will show using a Monte Carlo simulation that this method works very well both in the highly correlated parameters case, and in the case where the posterior distribution is highly skewed with a heavy tail. We will use mean absolute step size to show that only a small burn-in and small amount of thinning will be required using this candidate density

Fitting B-Splines to Quantiles Using a Brier Entropy Penalty

Björn Bornkamp and Katja Ickstadt
Technische Universität Dortmund, Germany
{bornkamp, ickstadt}@statistik.tu-dortmund.de

In this work we present a novel method for transferring information about quantiles of an uncertain bounded quantity into a fully specified probability distribution. A typical application is expert elicitation, for example in the context of health economics.

This is a challenging problem, as the stated quantiles are usually sparse and imprecise. We will approach both issues by penalization, such that the final distribution will contain the stated information, but subject to that, distributes its probability mass as uniformly as possible. As a functional basis for the modeled

distribution we use linear combinations of B-splines. B-splines are highly flexible, have excellent shape-preserving properties, and are also well suited for evaluation, differentiation, and integration. (Bornkamp and Ickstadt, *Amer. Statist,* 2009).

The Product Graphical Model

Luke Bornn
Univ. British Columbia, Canada
l.bornn@stat.ubc.ca

François Caron
INRIA Bordeaux SO, France
Francois.Caron@inria.fr

In this paper, the authors propose a class of prior distributions on decomposable graphs, allowing for improved modeling flexibility. While existing methods solely penalize the number of edges, the proposed work allows practitioners to control clustering, level of separation, and other features of the graph. Emphasis is placed on a particular prior distribution which derives its motivation from the class of product partition models; the properties of this prior relative to existing priors is examined through theory and simulation. The authors then explore the use of graphical models in the field of agriculture, showing how the proposed prior distribution alleviates the inflexibility of previous approaches in properly modeling the interactions between the yield of different crop varieties.

A Sequential Bayesian Experimental Design/Model Selection Methodology for Validating Models of Disease Pathogenesis

D. M. Bortz
University of Colorado, USA
dmbortz@colorado.edu

In our previous work we have proposed a model selection methodology for selecting among candidate compartmental (differential equation) models which balances goodness-of-fit with a measure of statistical complexity (Bortz and Nelson, Bull. of Math. Bio., 2006). In this work we propose a Sequential Bayesian Algorithm, iterating between computing model rankings via selection criteria and designing experiments to maximize model distinguishability. We will present results illustrating selection between competing models for HIV pathogenesis and Staphylococcus epiermidis bacteremia. We also show how to identify an optimal sampling strategy for incorporation of salient immune system features into the model. A theoretical investigation of the properties of this algorithm will also be presented.

A Kernel-Based Spatio-Temporal Surveillance System for Monitoring Influenza Incidence

Paloma Botella-Rocamora
Universidad CEU-Cardenal Herrera, Spain
pbotella@uch.ceu.es

Miguel-Angel Martínez-Beneito
Centro Superior de Investigación en Salud Pública, Valencia, Spain
martinez_mig@gva.es

The threat of pandemics has made influenza surveillance systems a priority in epidemiology services around the world. The emergence of A-H1N1 influenza has required accurate spatio-temporal surveillance systems in order to undertake specific actions only when and where they are necessary. In that sense, the main goal of this work is to describe a novel Bayesian model for monitoring the geographical distribution of the incidence of influenza, based on information from sentinel surveillance networks.

A Bayesian Poisson mixed linear model is proposed in order to describe the observed cases of influenza for every sentinel and week of surveillance. This model includes a spatio-temporal random effect that shares information in space by means of a kernel convolution process and in time by means of a first order autoregressive process. The extrapolation of this term to sites where information on incidence is not available will allow us to visualize the geographical distribution of the disease for every week of study. The following work also shows the performance of this model in the Comunitat Valenciana's Sentinel Network (one of the 17 autonomous regions of Spain) during the 2009-2010 winter season as a real case study of this methodology.

Bayesian Estimation of a Covariance Matrix: Application for Asset and Liabiliy Management

Mathilde Bouriga
Université Paris Dauphine, France
mathilde.bouriga@hotmail.fr

Olivier Féron
EDF R&D Clamart, France
feron@lss.supelec.fr

Jean-Michel Marin
Université Montpellier, France

Christian P.Robert
Université Paris Dauphine, France

Many authors have considered the problem of estimating a covariance matrix in small samples. In this framework the sample covariance matrix is not robust, the solution is to impose some ad hoc structure on the covariance matrix to force it to be well-conditioned. This method is known as shrinkage. Here we approach the problem with an hierarchical bayesian perspective : we propose hierarchical priors for the covariance matrix to shrink toward diagonality. This approach draws on the works of M.J. Daniels and R.E. Kass (Nonconjugate bayesian estimation of covariance matrices and its use in hierarchical models, 1999) but we have searched to avoid any influence of the hyperparameters on the inference. We use an Inverse

Wishart prior and we place flat priors on the logarithm of its hyperparameters. The problem of estimating is under the Stein's loss function and a Markov Chain Monte Carlo sampling scheme is used to implement posterior inference in the proposed model. A simulation study allows us to assess the performance of the estimator in terms of small-sample risk. Our bayesian estimator is then applied to a real longitudinal example from portfolio selection, in which the dimension of the covariance matrix is large relative to the sample size.

Bayesian Hierarchical Modeling for Paleoclimate Reconstruction from Geothermal Data

Jenný Brynjarsdóttir and Mark Berliner
The Ohio State University, USA
jennyb@stat.osu.edu mb@stat.osu.edu

Information about past climate has an important role in learning about climatic change. Paleoclimate reconstructions have been made from many different types of proxies, including ice-cores, pollen and tree-rings. We have developed a Bayesian hierarchical modeling approach to temperature reconstruction from borehole temperature-depth profiles. This approach relies on the physical model of heat conduction in solids and uses the solution to the heat equation with a step function as a boundary condition. We have applied our method to data from eight borehole temperatures from the Colorado Plateau and produced temperature histories with uncertainty estimates, going back 400 years. We have shown that combining these boreholes in one hierarchical Bayesian model as opposed to fitting eight separate models, reduces the posterior uncertainty for each reconstruction.

The Structural Markov Property

Simon Byrne and A. Philip Dawid
University of Cambridge, UK
S.Byrne@statslab.cam.ac.uk A.P.Dawid@statslab.cam.ac.uk

We consider the problem of model determination for undirected decomposable graphical models. Also referred to as structural learning, this involves inferring the structure of the underlying graph from the observed data. This problem has been the focus of much recent work in both Bayesian and frequentist statistics, particularly in the context of covariance matrix estimation and contingency table analysis.

A Bayesian approach to this problem requires the specification of a probability distribution over a set of graphs. We introduce a "structural Markov property" for such distributions: the structure of induced subgraphs should be independent conditional on the existence of a complete separating subgraph.

The form of the structural Markov property is analogous to the Markov property of the sampling distribution, and the hyper Markov properties of the parameter distributions. By exploiting these properties we can obtain an efficient MCMC methods for determining the posterior based on local computations.

Furthermore, we propose an exponential family distributions on the set of decomposable that satisfy the structural Markov property, and show that it forms the conjugate prior for sampling from a family of compatible Markov distributions.

Assessing the Fit of Regression Models for Multiple Imputation

Stefano Cabras
University of Cagliari, Italy
s.cabras@unica.it

María E. Castellanos and Alicia Quirós
Universidad Rey Juan Carlos, Spain
maria.castellanos@urjc.es alicia.quiros@urjc.es

Multiple imputation techniques, first introduced by Rubin (*Biometrika*, 1978), have become popular in the last decades, and nowadays there are a variety of multiple imputation models and software available (for instance, the MICE package in R). These techniques consist in filling missing data values of a variable with replicates, which usually come from the conditional predictive model for that variable given all the others in the data set. In this paper we concentrate on sequential regression multiple imputation (SRMI).

Despite the popularity of multiple imputation methods, assessing their goodness-of-fit (GOF) is not a common practice and only a few papers address this problem. Gelman *et al.* (*Biometrics*, 2005) and Abayomi *et al.* (*JRSSC*, 2008) proposed Bayesian posterior predictive checks for completed datasets, i.e. observed data and imputed unobserved data. He *et al.* (2007) suggested to use the posterior predictive p-value (ppp) of Gelman *et al.* (*Statistica Sinica*, 1996) to assess the GOF of multiple imputation models (MIM).

In this work, we focus on the latter approach, based on p-values, pointing out that it cannot be interpreted under the usual uniform distribution in (0,1) (Bayarri and Berger, *JASA*, 2004) and showing that ppp is conservative for GOF of SRMI, when the GOF is based on suitable discrepancy measures. In order to overcome these drawbacks, in this paper we propose to assess the GOF of SRMI with the calibrated posterior predictive p-value (cppp) of Hjort *et al. JASA*, 2006) This approach has proven to be powerful for the GOF of SRMI with discrepancy measures based on regression residuals. This evidence comes from simulations and theoretical results as we are able to find the theoretical null distribution of some of the discrepancy measures used. The behavior of cppp under several types of departures from the null model has been obtained through a simulation study for several proportions of missing data showing a better performance when compared with the ppp. For illustration purposes we further present an application to a data set with missing values included in MICE package.

Bayesian Analysis of Markov Modulated Discrete Time Queues

Toros Caglar and Refik Soyer
The George Washington University, USA
toros@gwu.edu soyer@gwu.edu

Markov modulated queues have previously been addressed in areas needing to account for varying arrival and service rates dependent on a randomly changing environment. Discrete time queues and their Markov modulated extensions have also been studied previously, with an emphasis on the computer and communication system applications. In this paper, we will be focusing on the Bayesian analysis of the Markov modulated single and multiple server discrete time queues. Specifically, we are interested in the Geo/Geo/1 and Geo/Geo/s which are the discrete counterparts of the M/M/1 and M/M/s queues, respectively. Initially, we will allow the service and arrival processes to be modulated by independent environments. We will also look at the case where the service and arrival processes modulate based on a common environment. Situations where the environmental process is observable and unobservable will be dealt with separately. In order to model the aforementioned systems, we will be considering two experimental designs. The first design will utilize n_a inter-arrival and n_s service times and result in two independent geometric processes describing the arrival and the service processes. The second design will view both the arrival and service processes as independent Bernoulli processes, resembling instantaneous replacement reliability models with periodic inspection. Lastly, beyond providing Bayesian inferential methods for the unknown parameters, predictive distributions of certain performance measures will be obtained.

Bayesian Inference for Incomplete Marked Spatial Point Patterns

Catherine A. Calder and William F. Darnieder
The Ohio State University, USA
calder@stat.osu.edu darnieder@stat.osu.edu

We consider the problem of inferring the intensity function of a marked inhomogeneous spatial Poisson process that is observed incompletely as a result of a mark-dependent censoring mechanism. We formulate a general latent variable strategy for model specification and develop an efficient reversible jump MCMC framework for model fitting. Through a series of simulation studies, we explore how the form and extent of the censoring affects our ability to make inference on the underlying Poisson intensity function. In addition, we provide two illustrative examples of our methodology. First, we consider a forestry example in which the location and size (diameter at breast height) of all trees in a forest are known. We compare estimates of the underlying spatial intensity function based on the complete data to that obtained when we impose a size-dependent censoring mechanism. We also analyze incomplete human activity pattern data collected as part of the Los Angeles Family and Neighborhood Survey (L.A.FANS). In large-scale

sample surveys such as L.A. FANS, complete records of individuals' activities cannot be collected due to the burdensome nature of data collection, processing, and storage. Instead, survey participants are only asked to provide location and time-use information for particular types of regular activities (e.g., working and shopping). Using our methodology, we infer L.A. FANS participants' activity intensities across space (Los Angeles county) based on their observed activity locations, as well as from information about the number of unknown activities provided by the amount of time not accounted for by their observed activities.

Bayesian Analysis of Semimarkovian Models with Applications to Hardware Reliability

Javier Cano
Universidad Rey Juan Carlos, Spain
javier.cano@urjc.es

Javier M. Moguerza
Univ. Rey Juan Carlos, Spain
javier.moguerza@urjc.es

David Ríos Insua
Royal Academy of Sciences, Spain
david.rios@urjc.es

We provide a unified framework for Bayesian analysis (inference, prediction and decision making) with semimarkovian processes, providing new perspectives on the area. We consider both short-term and long-term forecasting. We then consider reliability, availability and maintenance applications, in which semimarkovian processes provide further flexibility with still a reasonably feasible computational treatment. We use flowgraph models to deal with generalized phase-type absorbing times. We illustrate our study with a HW reliability problem, analyzing the performance of our university Enterprise Resource Planner (ERP) system, making inference for the system parameters, performing RAM forecasting, and discussing some maintenance problem.

Integral Priors for Bayesian Model Selection in Practice

Juan A. Cano
Universidad de Murcia, Spain
jacano@um.es

Diego Salmerón
Consej. de Salud de Murcia, Spain
diego.salmeron@carm.es

In Bayesian model selection when the prior information on the parameters of the models is vague default priors should be used. Unfortunately, these priors are usually improper yielding indeterminate Bayes factors that preclude the comparison of the models. Among the many attempts for solving this difficulty intrinsic priors were introduced in Berger and Pericchi (1996) instead of the original default priors; however, there are situations where the class of intrinsic priors is too large and a problem of robustness arises. To overcome this problem integral priors that are solutions of a system of integral equations which is derived to calibrate the initial default priors were proposed as prior distributions for Bayesian model selection in Cano *et al.* (2008), where it was stated that under some general assumptions

integral priors are unique and some illustrative examples were provided. Here, we present some examples to illustrate how this new methodology works. In some cases integral priors are explicitly found and they can be used to directly obtain Bayes factors, nevertheless in the cases where we are not able to explicitly find the integral priors they have associated Markov chains that can be used to obtain approximated Bayes factors.

Modelling Multi-Output Stochastic Frontiers Using Copulas

Alessandro Carta and Mark Steel
University of Warwick, UK
A.Carta@warwick.ac.uk M.F.Steel@stats.warwick.ac.uk

A major challenge in the study of production and efficiency analysis through stochastic frontiers is the treatment of economic units producing multiple outputs. In this work we introduce an econometric model to deal with this through a system of stochastic frontier equations. The main feature of the stochastic frontier model is the presence of an individual random effect which characterises the distance of an economic unit from the hypothetical frontier. This so-called inefficiency term is defined on the positive orthant of the real line. The sum of the inefficiency and the usual symmetric measurement error is referred to as "composed error" and clearly has a skewed distribution. The main interest in these models is often in the inefficiency terms, as they are directly related to firm-specific efficiencies. Our approach makes use of a multivariate regression model, but we model the two components of the composed error separately through different multivariate distributions. The main innovation is in the choice of the multivariate distribution for the inefficiencies; this distribution has to be such that each marginal is modelling the inefficiency of the corresponding output with respect to the frontier. So we need multivariate distributions that allow us to control the margins defined on the positive real line while also being able to model different degrees of dependence. A flexible tool to deal with such complex distribution is represented by the copula function, which allows us to choose the marginal distributions independently from the dependence structure which is represented by the copula. An earlier approach in the literature directly models the composed error as a skewed multivariate distribution. The main advantage of our modelling through copulas is the possibility to make separate inference on the specific efficiency for each output and study their dependence properties. In a Bayesian inference setting with flexible generalised gamma marginals for the inefficiencies, we compare different copulas through posterior odds, using prior matching for the copula parameters. We prove posterior propriety under a convenient and commonly used improper prior structure. The model is applied to a dataset of Dutch dairy farms and estimated using MCMC. The flexibility of our model is not only useful for the case of the stochastic frontier model, but immediately extends to other cases where the usual normal assumption for random effects is not appropriate.

Dynamic Stock Selection Strategies: A Structured Factor Model Framework

Carlos M. Carvalho
University of Chicago and University of Texas at Austin, USA
carlos.carvalho@mccombs.utexas.edu

Hedibert F. Lopes
The University of Chicago, USA
hlopes@chicagobooth.edu

Omar Aguilar
Financial Engines, USA
o_aguilar@ymail.com

We propose a novel framework for estimating the time-varying covariation among stocks. Our work is inspired by asset pricing theory and associated developments in Financial Index Models. We work with a family of highly structured dynamic factor models that seek the extraction of the latent struc- ture responsible for the cross-sectional covariation in a large set of financial securities. Our models incorporate stock specific information in the estima- tion of commonalities and deliver economically interpretable factors that are used both, as a vehicle to estimate large time-varying covariance matrix, and as a potential tool for stock selection in portfolio allocation problems. In an empirically oriented, high-dimensional case study, we showcase the use of our methodology and highlight the flexibility and power of the dynamic factor model framework in financial econometrics.

Synthesizing Lists of Differentially Expressed Features in Related Experiments

Alberto Cassese
Università di Firenze, Italy
cassese@ds.unifi.it

Marta Blangiardo and Sylvia Richardson
Imperial College, UK
{m.blangiardo, sylvia.richardson}@imperial.ac.uk

In microarray experiments, a common problem is the comparison of similar experiments involving different tissue/treatment/species. The aim is to find a restricted set of interesting features perturbed in all experiments to be further investigated and validated by direct experimentation. Ideally such problem should involve the joint re-analysis of the experiments, but it is not always feasible and it is computationally demanding. We propose a Bayesian model to compare two or more experiments with the aim of finding a common list of features (e.g. genes). For each experiment we rank the features using a measure of differential expression. A multinomial distribution is specified on the number of features perturbed in any combination of the experiments under study (e.g. all experiments, only one experiment, etc.), with a weakly informative Dirichlet prior. The quantity of interest

is the ratio between the probability of being perturbed in all experiments and the probability of being perturbed by chance under the hypothesis that the experiments are independent. Two decision rules have been investigated to choose a cut-off on the ranked lists: the first one selects the maximal list of common features (strongest deviation from the hypothesis of independence); the second finds the cut-off where the probability of being in the common list doubles the chance. Extensive simulations proved that this second rule leads to a smaller number of false positives and false negatives (Blangiardo and Richardson, *Genome Biology*, 2007). This method has been illustrated on several publicly available datasets and an R package is available on the R website (www.r-project.org).

Bayesian Approach to Analysing Longitudinal Bivariate Binary Data with Informative Dropout

Jennifer S.K. Chan and Wai Y. Wan
The University of Sydney, Australia
jennifer.chan@sydney.edu.au waiyin.wan@sydney.edu.au

To study the effect of methadone treatment in reducing multiple drug uses while controlling for their joint dependency and non-random dropout, we propose a bivariate binary model with a separate informative dropout (ID) model. In the model, the logit of the probabilities of each type of drug-use and dropout indicator as well as the log of the odds ratio of both drug-uses are linear in some covariates and outcomes. The model allows the evaluation of the joint probabilities of bivariate outcomes. To account for the heterogeneity of drug use across patients, the model is further extended to incorporate mixture and random effects. Parameter estimation is conducted using a Bayesian approach and is demonstrated using a methadone treatment data. A simulation experiment is conducted to evaluate the effect of including an ID modeling to parameters in the outcome models.

Bayesian Network Parameter Learning with Informative Dirichlet Prior

Rui Chang
University of California, USA
chang.rui@hotmail.com

Parameter learning task under Bayesian framework requires prior distribution in the parameter space. The use of prior distribution becomes critical and indispensable to achieve good estimation results especially when the data amount is scarce relative to the size of the network. Dirichlet prior with pre-defined hyperparameters is often used to provide initial estimate of the Bayesian network parameters. Maximization-a-posterior(MAP) produces a local maximum based on this initial guess. However, one essential problem in Dirichlet prior is that it is often impossible for a domain expert to accurately specify the Dirichlet hyperparameters. Instead, these hyperparameters are usually set arbitrarily which may

result in biased estimation results. The distortion becomes more significant when data sample is scarce.

On the other hand, there is usually rich amount of qualitative prior knowledge available in a domain. This knowledge contains validated information describing certain functional relationships between the variables in a Bayesian network, such as the presence/absence of a variable causes its child variable more/less likely. This type of prior knowledge has not been well studied and used in Bayesian network learning due to its lacking of quantitative information. However, this type of prior domain knowledge possesses valuable information which can be very useful in determining the hyperparameters of the Dirichlet prior.

In this paper, we propose a novel framework for learning parameters in a Bayesian network by integrating generic qualitative domain knowledge with the data. Based on simple sampling scheme, we reconstruct the hyperparameters of Dirichlet prior distribution from the qualitative constraints and these hyperparameters are combined with data statistics (based on MAP score) to form a Bayesian network model space. Then, we design two model selection schemes to estimate the parameters. We test our algorithms (QMAP) in estimating the parameters of a real world computer vision network and a real biological network. We showed that the informatively set Dirichlet hyperparameters based on the qualitative domain knowledge can significantly improve Bayesian network learning and generalization accuracy comparing to MAP and ML.

From Phenotype to Genotype: Reconciling Approaches Through Bayesian Model Averaging

Carla Chen Jonathan Keith and Kerrie Mengersen
Queensland University of Technology, Australia
k.mengersen@qut.edu.au

Genetic research of diseases with complex etiology poses extra uncertainty due to lack of clear biomakers for phenotype ascertainment. Although the phenotype for these diseases are often identified on the basis of clinically defined criteria, this may not be suitable for understanding the genetic composition of the diseases. Different statistical approaches have been proposed for phenotype definition. However, the results of our previous studies have shown that difference in the phenotype estimated by different approach have substantial impact on the subsequent linkage analysis.

Instead of obtaining the results based upon a single parsimonious model, we explored ways of combining these models, using Bayesian model averaging as the foundation. We also explored the use of this same foundation for combining different goodness of fit criteria, adopted as preference weights for the models.

Our approach is illustrated using a real problem of better understanding migraine. The models considered are latent class analysis and grade of membership. We also evaluate the approach using simulated data. The model evaluation criteria include BIC, DIC, Bayes factors and posterior predictive probabilities.

Our approach shows promising ability in consolidation of the cores of phenotype clusters and reflecting the uncertainty due to model uncertainty by increasing the fuzziness at the boundaries of the clusters. Thus, in the subsequent genetic

linkage analysis, loci which are strongly differentiated at the cluster cores may have stronger LOD scores under the combined model than under an individual model. The formal combination of different model evaluation criteria also provides a compromise if there is *a priori* uncertainty about the criterion to be chosen.

Bayesian Value-at-Risk and Expected Shortfall Forecasting via the Asymmetric Laplace Distribution

Qian Chen
University of Sydney, Australia
qian.chen@sydney.edu.au

Richard Gerlach
University of Sydney, Australia
richard.gerlach@sydney.edu.au

Zudi Lu
University of Adelaide, Australia
zudi.lu@adelaide.edu.au

A parametric approach to estimating and forecasting Value-at-Risk (VaR) and Expected Shortfall (ES) for a heteroskedastic financial return series is proposed. A GJR-GARCH models the volatility process, capturing the leverage effect. To take account of potential skewness and heavy tails, the model assumes an asymmetric Laplace as the conditional distribution of the financial return series. Furthermore, dynamics in higher moments are captured by allowing the shape parameter in this distribution to be time-varying. Estimation is via an adaptive Markov Chain Monte Carlo (MCMC) sampling scheme, employing the Metropolis-Hastings (MH) algorithm with a mixture of Gaussian proposal distributions. A simulation study highlights accurate estimation and improved inference of parameters compared to a single Gaussian proposal MH method. We illustrate the model by applying it to forecast return series from four international stock market indices, as well as two exchange rates, and generating one step-ahead forecasts of VaR and ES. We apply standard and non-standard tests to these forecasts, as well as to those from some competing methods, and find that the proposed model performs favourably compared to many popular competitors: in particular it is the only conservative model of risk over the period studied, which includes the financial crisis.

An integrated Bayesian Hidden Markov Model Approach to CNVs Incorporating Pedigree Information

Zhen Chen
National Institutes of Health, USA
chenzhe@mail.nih.gov

Mahlet Tadesse
Georgetown University, USA
mgt26@georgetown.edu

Kai Wang
University of Pennsylvania, USA
kai@mail.med.upenn.edu

Mingyao Li
University of Pennsylvania, USA
mingyao@mail.med.upenn.edu

CNVs refer to gains and losses of genomic elements compared to a reference genome assembly. Studies have demonstrated the heritability of CNVs but few

incorporated family structures. We develop an integrated Bayesian approach that incorporates family relationships when inferring CNVs. We assume copy number sequence along the chromosome follows a hidden Markov model (HMM) with transition probabilities dependent on genetic distances between adjacent SNPs. We also allow for de novo events in offspring's CNVs and use another HMM to account for the dependence with neighboring SNPs in the same de novo CNV region. The two HMMs are linked together using Mendellian rules, hence permitting the incorporation of family relationship when inferring about the hidden CNV states. Our approach yields posterior distributions of CNV configurations, thus providing an uncertainty measure for the inferred CNVs. We evaluate the performance of the method by applying it to simulated datasets and the CEU trio data from HapMap.

A Bayesian Approach for Comparing the Relative Returns of Securities using the Reciprocal of the Coefficient of Variation

Delson Chikobvu
University of Free State, South Africa
chikobvu@ufs.ac.za

Abrie J. van der Merwe and Jesse A. De Beer
University of Free State, South Africa

Bayesian statistics is employed to simulate and estimate the reciprocal of the coefficient of variation and used compare the returns of three securities. Bayesian inference has a number of advantages. A full Bayesian analysis provides a natural way of taking into account all sources of uncertainty in the estimation of the parameters. Uncertainty about the true value of the reciprocal of the coefficient of variation is incorporated into the analysis through the choice of a vague prior distribution. The Bayesian simulation procedure is employs the posterior distribution in doing the simulations. The procedure can be useful in solving the portfolio selection problem. Results show that the Bayesian simulation approach is just as good if not better than the standard classical statistical approach in assessing the performance of an investment. The added advantage of the Bayesian approach is that, from the posterior distribution of the reciprocal of the coefficient of variation, we are in a position to obtain quantiles, credible regions and perform other inferential tasks.

Free Energy Sequential Monte Carlo. Application to Mixture Modelling

Nicolas Chopin and Pierre Jacob
CREST (ENSAE), France
nicolas.chopin@ensae.fr pierre.jacob@ensae.fr

We introduce a new class of Sequential Monte Carlo (SMC) methods, which we call free energy SMC. This class is inspired by free energy methods, which originate from Physics, and where one samples from a biased distribution such that a

given function $\xi(\theta)$ of the state θ is forced to be uniformly distributed over a given interval. From an initial sequence of distributions (π_t) of interest, and a particular choice of $\xi(\theta)$, a free energy SMC sampler computes sequentially a sequence of biased distributions $(\tilde{\pi}_t)$ with the following properties: (a) the marginal distribution of $\xi(\theta)$ with respect to $\tilde{\pi}_t$ is approximatively uniform over a specified interval, and (b) $\tilde{\pi}_t$ and π_t have the same conditional distribution with respect to ξ . We apply our methodology to mixture posterior distributions, which are highly multimodal. In the mixture context, forcing certain hyper-parameters to higher values greatly faciliates mode swapping, and makes it possible to recover a symetric output. We illustrate our approach with univariate and bivariate Gaussian mixtures and two real-world datasets.

A Modified Multivariate Student-t Distribution: An Application to Stochastic Volatility Models

S.T. Boris Choy
The University of Sydney, Australia
boris.choy@sydney.edu.au

Stochastic volatility (SV) models with leverage have been widely used to model time varying volatility and the correlation between the return and volatility in financial time series. In a univariate SV setting, the errors of the return and log-volatility are assumed to follow a robustifying bivariate Student-t distribution. From distribution theory, the marginal univariate t-distributions must have the same degrees of freedom. This restriction may not be realistic in real data analysis because the return and log-volatility may have very different degrees of freedom for their marginal t-distributions. This paper proposed a modified Student-t distribution which allows the marginal t-distributions to have different degrees of freedom. In the setup of a univariate SV model with leverage and modified t-distribution, the return and log-volatility can have marginal t distributions with different degrees of freedom and at the same time are allowed to be correlated. In an empirical study of different stock market log-return data in this paper, we show that the modified t-distribution can accurately estimate the degrees of freedom for the return and log-volatility and hence provides a much better model fit than the bivariate t-distribution. The construction of the modified t-distribution is based on the scale mixture of normal (SMN) form of the t-distribution and the scale mixing variable, the by-product of the SMN form, can be used as a global diagnostics for outliers in return and volatility. We demonstrate that the model can be easily implemented using WinBUGS package. Keywords: Scale mixture of normal; scale mixing variable; Markov chain Monte Carlo; Gibbs sampler; Outlier diagnostics; WinBUGS

Sample Design in the Bayesian Analysis of Radiocarbon Dates in Paleoecology

J. Andrés Christen
CIMAT, Guanajuato, Mexico
jac@cimat.mx

Maarten Blaauw
Queen's University Belfast, UK
maarten.blaauw@qub.ac.uk

Radiocarbon dating is routinely used in Palaeoecology to build chronologies of lake and peat cores, aiming at inferring a model that would relate the core depth with its age. Recently, we have developed a new approach for chronology building that has received enthusiastic attention by palaeoecologists. Our methodology is based on controlling core accumulation rates using a gamma autoregressive semiparametric model with an arbitrary number of subdivisions along the core. Using prior knowledge about accumulation rates is crucial, and informative priors are routinely used. Since many cores are currently analyzed, using different data sets and prior distributions, a robust (addaptive) MCMC is very useful. We use the t-walk (Christen and Fox, 2010), which has proved to be a very robust MCMC sampling algorithm, that works acceptably well in many situations. However, radiocarbon dates have a cost of c. 400 euro each and it is therefore important to carefully select which and how many radiocarbon samples to date. Recently a sequential sampling procedure has been proposed (Christen and Sansó, 2010) in the context of Gaussian Processes and computer models, that is computationally simple and provides sound designs. Using the above mentioned gamma autoregressive processes, we estimate the (posterior) covariance structure at any core depth using the MCMC sample. This covariance structure is then used to calculate a simple score at every depth; by maximizing such score we obtain the next depth to be radiocarbon dated. Either using the true date or a predicted one, the covariance structure is updated to continue with the sequential design. This represents a feasible alternative to the full solution of maximizing expected utility over all possible designs. We explain the autoregressive semiparametric gamma model and the adaptive MCMC used, present a sequential design example from a Dutch peat core and discuss further generalizations into radiocarbon dating and other design problems.

A New Method in Detecting Dynamic Seasonal Variation Using State Space Models

Anette L. Christensen
Aarhus University Hospital, Denmark
anluc@rn.dk

Lars H. Rasmussen, Kim Overvad,
Aarhus University Hospital, Denmark
lhr@rn.dk oko@rn.dk

Søren Lundbye-Christensen and Claus Dethlefsen
Aarhus University Hospital, Denmark
solc@rn.dk cld@rn.dk

Introduction: Several studies have reported that the frequencies of incident venous thromboembolism (VTE) exhibit seasonal variation within the calendar year (Boulay *et al.* 2001). These reports are based on assuming the seasonal variation being described as a single cycled sinusoidal curve and static, i.e. the seasonal variation is identical from year to year. Studies investigating changes over time in seasonality of coronary diseases have been published (Crawford *et al.* 2003; Seretakis *et al.* 1997). These studies rely on fitting a static model to data within each year of the study period and using a linear regression model to test for trend. As proposed by Lundbye-Christensen *et al.* 2009, dynamic seasonal variation can be modelled by a state space model (Harrison & Stevens 1976). We formulated a state space model describing the potentially dynamic seasonal variation exhibited by VTE in the Danish population. To assess whether the seasonal variation is dynamic or static we used a class I multi process state space model (West & Harrison 1997). Method and Materials: We conducted a cohort study consisting of the Danish population from 1977 to 2009. Daily cases of VTE were identified through the Danish National Patient Registry, which holds information of all hospitalisations in Denmark. During follow-up more than 93,000 cases developed VTE. We formulated a Poisson state space model for VTE described by a seasonality component, a secular trend and a weekday effect. The seasonality was modelled by a four cycled sinusoidal curve, the secular trend as a restricted cubic spline and the weekday effect as a sum-to-zero effect with a period of seven days. Data was analysed by fitting a class I multi process state space model, which discriminate between state space models, one with all parameters being static and a series of models with dynamic components, with equal prior probability. From extended Kalman smoothing we obtained estimates of the posterior probabilities for each model, and the model with highest final estimated posterior probability was chosen. The analyses were performed using R (R Development Core Team 2008).

The Need for Informative Priors in Hierarchical Models for Microbiological Control in Food Industry

Natalie Commeau
Agroparistech/INRA MIA, France
natalie.commeau@agroparistech.fr

Marie Cornu
AFSSA, France
m.simon-cornu@afssa.fr

Eric Parent
Agroparistech/INRA MIA, France
eric.parent@agroparistech.fr

Data are poorly informative in food industry: few records can be made because of a limited budget, it is difficult to overcome confidentiality constraint, and when available, most of them are imprecise due to methods used for microbiological analyses. For the pathogen *L. monocytogenes*, two types of data can be obtained: either binary (detection of presence) or counts (enumeration of colony forming units). Food is produced by batches and random effect have to be modeled to take into account between and within batch variability. Additional hidden layers to the hierarchical model can describe other levels of similarity among units, e.g. units produced from the same animal Among various types of possible models, we focus in this paper on the following one: $[\mu] \sim \mathcal{N}(m, s^2)$; $[\tau] \sim \mathcal{G}(a, b)$; $[z|\mu, \tau] \sim \mathcal{N}(\mu, \tau)$; $[x|z] \sim \mathcal{B}er(1 - e^{-10^z * S})$, and $[y|z] \sim \mathcal{P}ois(10^z * S * d)$, where x stands for the detection result, y for the enumeration one and z is the mean concentration in 1 g. In the Bayesian setting, it is not an easy task to assess the mean and the precision of such random effect.

In a first application (diced bacon), we rely on a physically based simulator of the process and very imprecise past data to obtain priors for μ and τ but we enlarge by the rule of thumb their variances. In a second application (cold smoked salmon), nothing precise could be obtained from the expert for τ and we had to set a vague prior ($a = b = 0.1$), which prevented us from comparing competing models through Bayes factors.

In his recent book *Uncertain Judgements*, O'Hagan (2006) deals with prior elicitation for mean and variance but questions remain for elicitation of the between batch variance: we discuss this issue and its consequences on both food industry applications.

On Moment Priors for Bayesian Model Choice with Applications to Directed Acyclic Graphs

Guido Consonni
Università di Pavia, Italy
guido.consonni@unipv.it

Luca La Rocca
Università di Modena e Reggio Emilia, Italy
luca.larocca@unimore.it

We propose a new method for ob jective Bayesian comparison of nested models, based on the notion of fractional Bayes factor applied to a particular class of

non-local priors named moment priors. These have been recently introduced to obtain a better separation between models and accelerate the learning behaviour of the Bayes factor, when the smaller model holds, relative to currently used local priors. Although the argument showing the superior performance of non-local priors is asymptotic, the improvement they produce is already apparent for small to moderate samples sizes, which makes them a useful and practical tool. As a by-product, it turns out that routinely used ob jective methods, such as ordinary fractional Bayes factors, are alarmingly slow in learning the true model, when this happens to be the smaller one. On the downside, when the larger model holds, non-local priors exhibit a weaker discriminatory power for sampling distributions close to the smaller model. This drawback, however, becomes rapidly negligible as the sample size grows, because the learning rate of the Bayes factor under the larger model is ex- ponentially fast, whether one uses local or non-local priors. We apply our methodology to directed acyclic graph models having a Gaussian distribu- tion. Because of the recursive nature of the joint density and the assumption of global parameter independence embodied in our prior, all calculations can be done at the level of individual vertices; additionally, only vertices admit- ting a distinct parent structure under the two graphs need be compared. We provide illustrations for a simple three-variable case, as well as for a more elaborate seven-variable situation. Although we concentrate on pairwise com- parisons of nested models, our methodology can be adapted to a full model search using the notion of model encompassing.

Have I Seen You Before? Principles of Bayesian Predictive Classification Revisited

Jukka Corander
Helsingin Yliopisto, Finland
jukka.corander@helsinki.fi

Yaqiong Cui
Helsingin Yliopisto, Finland
yaqiong.cui@helsinki.fi

Timo Koski
Största Tek. Univ., Sweden
tjtkoski@kth.se

Jukka Sirén
Helsingin Yliopisto, Finland
jukka.siren@helsinki.fi

A general Bayesian classification framework is introduced for data from multiple finite alphabets using predictive representations based on random urn models and generalized exchangeability. We develop a novel principle of supervised and semi-supervised probabilistic classification, which has attractive theoretical properties and is illustrated to achieve better correct classification rates in numerical examples. Optimal simultaneous and marginal supervised predictive classifiers are shown to become equivalent classification rules under generalized exchangeability when the amount of training data increases. It is also explicitly shown that the predictive framework can coherently handle missing observations without added computational complexity.

Auxiliary Sequential Monte Carlo Samplers with Applications to Approximate Bayesian Computation

Julien Cornebise
Univ.of British Columbia, Canada
cornebis@cs.ubc.ca

Oliver Ratmann
Duke University, USA
oliver.ratmann@duke.edu

Approximate Bayesian Computation (ABC) is at the center of much recent interest – see e.g. [Beaumont, *et al.* (2009)] – and extends Bayesian analysis to the most intricate models where the likelihood of the observations is either intractable or extremely expensive to compute. Successful applications range from population genetics to financial modeling. Most recent developments rely on Sequential Monte Carlo (SMC) Samplers, as introduced in [Del Moral, *et al.* (2006)], who extend classical SMC methodology to arbitrary sequences of distributions – not requiring any underlying Markovian recurrence relations such as occur in state-space models. We develop an auxiliary variable framework for SMC samplers similar to the one [Peters, *et al.* (2008)] established for classical SMC filters.

We first bring new theoretical interpretation of the cutting edge ABC algorithms, unifying their many variants in a common framework relying on newly developed Auxiliary SMC Samplers similar to those brought by [Cornebise, *et al.* (2008)] to classical SMC. We cast SMC-Partial Rejection Control of [Pitt *et al.* (1999)] and ABC-specific selection schemes based on the error threshold [Ratmann, *et al.* (2009)] as different approximations strategies of auxiliary weights. We then build upon this new understanding some new methodological extensions, using extensions of the risk-theoretic criteria to provide a unified framework for optimal choice of the key algorithmic parameters: the forward and backward kernels, the error thresholds, and those newly introduced adjustment multiplier weights.

Informativeness and Experimental Design Issues in Definite Identification Systems Based on Genetic Markers

Fabio Corradi and Federico Ricciardi
Università di Firenze, Italy
corradi@ds.unifi.it ricciardi@ds.unifi.it

Traditionally indirect identification systems consider some specified family's members looking for the identification of an individual as a certain member of the family making use of the genetic markers conventionally employed in the country and the laboratory at hand.

From a pre-experimental/design-of-experiment point of view, we propose two alternatives to determine if a certain system satisfies the requirements of all parties involved and we compare different experiments varying according to the number and the position of the genetic loci to be used and the individuals in the familial pedigree providing their DNA evidence, taking also into account the genetic populations involved in the analysis.

In the Bayesian approach, the aim of an identification system is to update the probability of the hypotheses describing different familial pedigrees from their

subjectively determined prior to their posterior. Since such probability update depends on the LR, we focus on the distribution of the LR induced by the identification systems taken into account and by their sample spaces.

In this perspective, a choice among the systems can be made by defining a decision rule which consider those LR's probability regions able to reach high hypotheses posterior probabilities. Furthermore also the costs related to each identification system will be taken into account.

In addition we consider some computationally efficient measures of information ordering among the considered identification systems. Finally a case-study will be extensively detailed, to show a possible application of the proposed framework.

Particle Learning for Regularization

Ioana A. Cosma
University of Cambridge, UK
ioana@statslab.cam.ac.uk

Robert B. Gramacy
University of Cambridge, UK
bobby@statslab.cam.ac.uk

Nicholas G. Polson
University of Chicago, USA
Nicholas.Polson@chicagobooth.edu

In this paper we develop a simulation-based approach to online inference for regression models under regularization. We begin by illustrating how standard approaches to batch regularized regression may be adapted for sequential inference in n, the number of responses. Unfortunately, this inferential scheme degenerates for high dimensional regression problems, i.e., $n \ll p$, where p is the number of predictors. However, we argue that, by turning the problem on its head and proceeding sequentially in p, the $p \gg n$ inference is again tractable. We employ particle learning methods, which are sequential Monte Carlo methods for state and parameter estimation that exploit the availability of recursively defined sufficient statistics for the parameters and/or states of the model, and allow for better estimation by inverting the order of the resampling and propagate steps.

Essentially, we develop a novel methodology for the estimation of a sparse set of regression coefficients by recasting the problem as sequential inference for a hidden state from a state space model whose dimension depends upon the length of the parameter vector. In doing so, we are positioned to apply fast, parallelizeable, particle learning methods to uncover the sparse regression vector as the mode of a suitably defined pseudo-posterior distribution. Moreover, our framework can also view the choice of the optimal amount of regularization as the mode of the observed data marginal likelihood. We show that our sequential methods naturally extend to learning this parameter under a variety of regularization penalties. Incorporating missing data is also easily handled in our state space representation. We provide an application to high dimensional portfolio mimicking – tracking a passive index – using high frequency intra-day data on the individual components of the S&P500 stock index.

Bayesian Partitioning for Mapping Residual Disease Risk: An 'Individually-Matched' Case-Control Design

Deborah Costain
Lancaster University, UK
d.costain@lancaster.ac.uk

Methods for exploring geographical heterogeneity in health outcomes continue to motivate much research. Sub-dynamics in atypical regions can be used filter resources, generate hypotheses and motivate research. In general, variation due to 'known' risk factors is not of intrinsic interest and the ability to accommodate 'confounders' is thus paramount. In addition, models which are flexible and make few assumptions about the functional form are sought after. Assuming individual-level geo-referenced health data confounding can be handled at the analysis stage or alternatively, at the design stage by using 'confounder-matched' case and controls. A consequence of this design-based approach is that the number of parameters increases with the sample size and analysis typically proceeds based on the matched, conditional likelihood (Breslow and Day, 1980). This work presents a Bayesian partitioning approach to handling matched case-control data. The model assumes that the region of interest can be decomposed into disjoint sub-regions in which the residual disease risk is constant. Random Voronoi tessellations are used for the decomposition and a Poisson model re-formulation, which is proportional to the matched conditional likelihood, is used. Posterior sampling is performed using Reversible jump MCMC (Green, 1995); Bayesian least-squares iterates (Gamerman, 1996) and re-parameterisation is performed to efficiently propose candidates. Key features of the approach include relaxing the customary assumptions of a distance defined covariance structure, and moreover, the ability to detect spatial discontinuity. Computationally, the Poisson model reformulation maintains a conditional independence structure easing the computational burden and additional 'non-matched-for' covariate information can be accommodated. The methodology is demonstrated on perinatal mortality data derived in the North-West Thames region of the UK. Matching was according to gender and date-of-birth. A measure of social deprivation, Carstair's index, was also available and was included in the analysis. As anticipated social deprivation was found to be a highly significant predictor of perinatal mortality. The unadjusted analysis (when not accounting for Carstair's index) resulted in surface estimates which exhibited regions of atypical risk. When adjusted for Carstair's index, however, there was no evidence of residual spatial variation in perinatal risk.

An Adaptive Delayed Acceptance MCMC Algorithm for Massive Inverse Problems

Tiangang Cui
Univ. of Auckland, New Zealand
tcui001@aucklanduni.ac.nz

Colin Fox
University of Otago, New Zealand
fox@physics.otago.ac.nz

Mike O'Sullivan
University of Auckland, New Zealand
tm.osullivan@auckland.ac.nz

We aim at solving inverse problems by sample based inference, which requires generating samples from posterior distributions. For the 3D geothermal reservoir models we are focusing on, the posterior densities are usually have several thousands parameters and spend more than hours to calculate. This makes the standard Metropolis-Hastings (MH) algorithm being computationally very demanding. To speed-up the sampling, Christen and Fox (2005), introduced a delayed acceptance MH (DAMH) algorithm, in which an approximate posterior density is used in the first step to eliminate the rejected candidates in the standard MH algorithm. In the second step, the exact posterior density is only evaluated for the pre-accepted candidates to ensure the ergodicity of the algorithm.

Since our models are based on integrated finite volume method, one obvious way to construct an approximate posterior distribution is replacing the fine model by a model with coarse grid. However, this gives a very low acceptance rate in the DAMH algorithm. Fortunately, by employing the estimation of the discretization error between the coarse model and the fine model, Kaipio and Somersalo (2007) introduced an enhanced error model (EEM) that provides a modified posterior distribution based on the coarse model. The EEM shows a closer behavior to the exact posterior distribution, and hence can be potentially used as the approximate posterior density. However, constructing the EEM requires evaluating the coarse model and the fine model for a large number of input points *a priori*. To overcome this difficulty, we combine the DAMH algorithm, the EEM and techniques in the adaptive MCMC (e.g., Haario, Saksman and Tamminen, 2001, and Roberts and Rosenthal, 2001) together to give an adaptive DAMH (ADAMH) algorithm, which allows construction of the EEM from the posterior distribution adaptively. Sufficient conditions for the algorithm to converge to the target distribution are provided, and several adaptive approximations have also been designed under these conditions. On various test cases, the ADAMH algorithm and the adaptive approximations show good computational and statistical efficiencies. On a 3D geothermal reservoir model, we are able to run the algorithm for 7,000 iterations in about 20 days time, which originally would cost the standard MH algorithm about six months.

Dimension Reduction for Bayesian Analysis of High-Dimensional Computer Models Using Principal Variables

Jonathan A. Cumming and Michael Goldstein
Durham University, UK
j.a.cumming@durham.ac.uk michael.goldstein@durham.ac.uk

Many complex computer models produce a very high-dimensional collection of outputs. Investigation of these models typically requires some form of dimension reduction or screening to arrive at a point where a feasible uncertainty analysis can take place. We introduce the method of Principal Variables as a means of screening the outputs from a computer model, which identifies a subset of the model outputs that preserves the structure and information contained in the larger collection. Unlike many other methods of dimension reduction, a principal variable analysis identifies an informative subset of the original outputs, greatly simplifying subsequent Bayesian calibration and forecasting using the model, as assessments of discrepancy between the model and the physical system are restricted to the chosen subset of outputs. Additionally, we develop PV methods for performing such screening when the output variables have natural groupings, such as time series of different quantities or collections of quantities reported for different spatial locations. We also motivate the use of principal variables as a form of exploratory sensitivity analysis for the identification of "active" model inputs. We illustrate our methods with application to computer models from cosmology and hydrocarbon reservoir engineering.

Bayesian Inference and Model Selection for Models Involving Pair Copula Constructions

Claudia Czado
Technische Universität München, Garching, Germany
cczado@ma.tum.de

The pair-copula construction method can be used to build flexible multivariate distributions. This class includes drawable (D), canonical (C) and regular vines developed (see for example Kurowicka and Cooke (2006). The multivariate distribution is build by using only bivariate copulas, which can be identified as as specific conditional and unconditional bivariate margins. This flexible class is very useful for applications in finance and allows for non Gaussian dependency structures (see Aas *et al.* , 2009, and Czado, 2009). Estimation and model selection methods are studied in a Bayesian setting using Markov Chain Monte Carlo (MCMC) estimation. Often the model can be reduced if conditional independencies are present in the data. However the identification of conditional independencies in the data is a challenging problem and I present two methods on how to solve this problem. One involves reversible junp MCMC while the other involves latent indicators. The methods will be illustrated with two applications. This work is joint with A. Min, M. Smith and C. Almeida.

Distance-Based Probability Distributions on Set Partitions for Bayesian Nonparametric Models

David B. Dahl
Texas A&M University, USA
dahl@stat.tamu.edu

Ryan Day
University of the Pacific, USA
rday@pacific.edu

Jerry W. Tsai
University of the Pacific, USA
jtsai@pacific.edu

Clustering methods are either model-based or distance-based, but we propose a framework that is both. The Dirichlet process induces a partition distribution in which the probability an item is clustered with another is uniform across all items. We propose several distributions for partitions which are indexed by pairwise distances among the items being clustered. The first is defined algorithmically by a Markov chain. The second explicitly specifies a probability mass function incorporating pairwise distances. Both have the Dirichlet process-induced clustering distribution as a special case. A new class of Bayesian nonparametric models that utilizes these distance-based probability distributions is defined. Applying these methods to a model for protein structure prediction, we find that the methods incorporating distance information substantially improve predictive accuracy.

Averaged Likelihoods

Rozenn Dahyot
Trinity College Dublin, Ireland
Rozenn.Dahyot@tcd.ie

We consider the general problem of linear and non-linear least squares. We propose several posteriors of the latent variables with respect to the observations available. The difference between these posteriors is the definition of the random variables involved in the problem, and their associations with the observations. These modellings lead to different likelihoods that are expressed as an averaged likelihood computed on a set of likelihoods defined for each bootstrap or d-deleted jackknife replications of the set of observations. This new framework also explains the relation between the Hough Transform and the standard likelihood conventionally used in statistics.

Comparing and Assessing Discrepancy Measures for Bayesian Model Selection

Julián de la Horra
Universidad Autónoma de Madrid, Spain
julian.delahorra@uam.es

María-Teresa Rodríguez-Bernal
Universidad Complutense de Madrid, Spain
mayter@mat.ucm.es

In this work, we describe a method for Bayesian model selection in which a discrepancy measure is used to evaluate the discrepancy between the sample and the model. We use some discrepancy measures and we compare them through the posterior expected discrepancy. Once we have chosen one discrepancy measure, we use it for choosing the more appropiate Bayesian model. Finally, we have to decide if this model is a good representation for the data; for doing that, a large number of samples is simulated from that model for comparing the posterior expected discrepancies.

Bayesian Semiparametric Modelling of Volatility

Eleni-Ioanna Delatola and Jim E. Griffin
University of Kent, UK
ed62@kent.ac.uk J.E.Griffin-28@kent.ac.uk

In the last couple of decades, the area of Stochastic Volatility (SV) models has grown rapidly. SV models capture the unobserved time-varying volatility of the log-returns using stochastic processes. Distribution of the innovation of the level of returns is often assumed to be normal or of some other parametric form, which does not allow for some features of the data to be modeled. Kim, Shephard and Chib (1998) developed an offset mixture model that approximates the distribution in question. In our work, we extend it to a nonparametric mixture model and define efficient computational methods. To illustrate this method we analyze the stock returns of General Motors.

Bayesian Variable Selection in Cluster Analysis.

Vasiliki Dimitrakopoulou P. Brown and J. E. Griffin
University of Kent, UK
{vd29, Philip.J.Brown, J.E.Griffin-28}@kent.ac.uk

Over the last decade, technological advances in the measurement of gene expression have lead to data with high dimensionality. A very common characteristic of such data is the large number of variables with the number of observations being substantially smaller ($p >> n$). Two tasks are commonly addressed during

the analysis of high-dimensional data. One is to uncover the group structure of the observations and the other involves the identification of the important variables. The aim of our work is to deal with those two problems simultaneously. The clustering problem is approached using a multivariate normal mixture model. The selection of the discriminating covariates is handled by the introduction of an exclusion/inclusion latent vector which gets updated via stochastic search techniques.

Variable Selection for Classification of High Dimensional Data

Katarina Domijan
National University of Ireland Maynooth, Ireland
Katarina.Domijan@maths.nuim.ie

Simon Wilson
Trinity College Dublin, Ireland
swilson@tcd.ie

The problem of variable selection for a multi-category kernel classification model aimed at high dimensional data sets is considered. The reduction of the complexity of the model is set in a Bayesian decision theoretic framework. The decision space is the set of all possible subsets of variables, which is discrete and high dimensional. A utility function is chosen so that it measures the predictive ability of a sub-model and simultaneously penalizes the inclusion of variables. Searching through the decision space with the aim of identifying sparse sub-models that yield good classification rates translates to a stochastic optimization problem, where the optimal decision maximizes the expected utility function with respect to all the unknowns, including the model parameters and future data. Different stochastic optimization methods are considered.

Bayesian Analysis of a Complex Agricultural Field Trial

Margaret R. Donald
Queensland University of Technology, Australia
Margaret.Donald@student.qut.edu.au

Clair L.Alston
Queensland Univ. Tech, Australia
Clair.Alston@qut.edu.au

Rick R. Young
Tamworth Agric. Inst., Australia
RickYoung@ozemail.com.au

Kerrie L.Mengersen
Queensland University of Technology, Australia
Kerrie.Mengersen@qut.edu.au

We consider an agricultural dataset, collected to determine a cropping system for the Liverpool Plains, NSW, eastern Australia, which maximises water use for

grain production while minimising leakage below the crop root zone. The data were collected over three spatial dimensions and consist of neutron scattering measurements (a soil moisture surrogate), site treatment, site, row, column and depth of soil where the measurement was taken.

We adopted a Bayesian Conditional Autoregressive (CAR) model to account for the particularly local site correlation. Despite their ease of use and their being able to be easily modelled in freely available software, CAR models have not been much used in analyses of agricultural data.

CAR and other models allowing spatial-autocorrelation are compared, together with several ways of modelling the treatment effects. The final choice was a hierarchical model with several CAR spatial variance components and a fixed part which involved an errors-in-measurement component with interval-censored errors. Spline models with measurement error, rather than orthogonal polynomials, were preferred for the descriptions of the treatment effect since inclusion of a latent variable requires continuous recalculation of the bases. The Bayesian framework permits easy specification of these complex models.

Bayesian Analysis of Growth Curves Using Mixed Models Defined by Stochastic Differential Equations

Sophie Donnet
Université Paris-Dauphine, France
donnet@ceremade.dauphine.fr

Jean-Louis Foulley
INRA-Jouy, France

Adeline Samson
Université Paris Descartes, France

Growth curve data consist of repeated measurements of a continuous growth process over time in a population of individuals. These data are classically analyzed by nonlinear mixed models. More precisely, let $(y_{ij})_{1\leq i\leq n, 1\leq j\leq n_i}$ denote the data, where y_{ij} is the noisy measurement of the observed biological process for individual i at time t_{ij}. In classical mixed models, the process is modeled by a deterministic function, depending on individual random parameters. Formally, the classical nonlinear mixed model is defined as: $y_{ij} = f(\phi_i, t_{ij}) + \varepsilon_{ij}$, $\varepsilon_{ij} \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$; $\phi_i \sim$ i.i.d $\mathcal{N}_p(\mu, \Omega)$ with $(\phi_i)_{1\leq i\leq n}$ the p-vectors of individual parameter vectors. The ε_{ij} are the residual errors, assumed to be independently and identically normally distributed with null mean and variance σ^2. f is classically one of the four most famous parametric functions modeling growth curves, namely the logistic, the Gompertz, the Richards and the Weibull functions. Each of them can be written as the solution of an ordinary differential equation (ODE) describing the evolution of growth rate $\partial f(\phi, t)/\partial t = F(f, t, \phi)$, with $f(\phi, 0) = f_0(\phi)$.

However, the standard growth functions prescribe monotone increasing growth and can fail to model unexpected changes in growth rates. We propose to model these variations using stochastic differential equations (SDEs) that are deduced from the standard deterministic growth function by adding random variations to the growth dynamics. Growth curve is thus described by a random process, denoted (Z_t), solution of the SDE: $dZ_t = F(Z_t, t, \phi)dt + \Gamma(Z_t, \phi, \gamma^2)dW_t$, with $Z(t = 0) = Z_0(\phi)$ where W_t is a Brownian motion. $\Gamma(Z_t, \phi, \gamma^2)$ is the volatility function depending on the unknown parameter γ^2.

The nonlinear mixed model defined by an SDE is thus: $y_{ij} = Z_{t_{ij}}(\phi_i) + \varepsilon_{ij}$, $\varepsilon_{ij} \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$; $dZ_t(\phi_i) = F(Z_t, t, \phi_i)dt + \Gamma(Z_t, \phi, \gamma^2)dW_t)$, and $\phi_i \sim \mathcal{N}(\mu, \Omega)$. In the model, three fundamentally different noises are distinguished: the inter-subject variability Ω, the dynamic noise γ^2, reflecting the random fluctuations around the corresponding theoretical dynamic model, and the measurement noise σ^2 representing the uncorrelated part of the residual variability associated with assay or sampling errors. Many types of volatility functions can be proposed to extend an ODE into an SDE (*e.g.* constant, square root or polynomial volatility). This choice depends on several considerations and is discussed.

A Bayesian inference of the parameters of interest μ, Ω, σ^2 and the volatility γ^2 of these SDE mixed models is developed. In the case when the SDE has an explicit solution, we describe an easily implemented Gibbs algorithm. When the conditional distribution of the diffusion process has no explicit form, we propose to approximate it using the Euler-Maruyama scheme. Finally, we suggest to validate the SDE approach via criteria based on the predictive posterior distribution. We illustrate the efficiency of our method using the Gompertz function to model data on chicken growth, the modeling being improved by the SDE approach.

A Bayesian Decision-Theoretic Alternative to Standard Multiple-Comparisons Adjustments in Clinical Trials

David Draper
University of California, Santa Cruz, USA
draper@ams.ucsc.edu

Clinical trials are often designed using Neyman-Pearson (NP) logic, in which (a) a primary analytic goal is identified (e.g., a mean comparison between experimental and standard groups), (b) a significance level α is chosen, typically conventionally (e.g., 0.05) and without regard to real-world consequences of false-positive errors, (c) a non-null effect size is chosen, and (d) a sample size calculation is made to achieve power at least $(1 - \beta)$ against the effect size in (c) (with β also typically specified conventionally (e.g., 0.1 or 0.2) and without regard to real-world consequences of false-negative errors). Often secondary subgroup comparisons are ignored in this calculation and (in NP language) may be badly underpowered at analysis time. When a study is run with inadequate sample sizes for subgroup analyses, but strong and interesting subgroup effects are found that disappear after a standard inferential multiple-comparisons adjustment (an example is the randomized controlled trial of a vaccine against HIV published as rgp120 HIV Vaccine Study Group (2005): *Journal of Infectious Diseases*, **191**, 654–663), what conclusions should be drawn?

As noted above, in the NP approach false-negative errors are supposed to be taken care of by having done a power calculation at the time the experiment was designed, but this begs the question of what decision should be taken, now that this study has been conducted, about whether to run a new trial and/or give the vaccine to some of the subgroups now. When the problem is reformulated as a decision that properly weighs all of the real-world costs and benefits of both false-positive and false-negative errors, the multiple-comparisons analysis

(which is terrified of announcing that an effect is real when it's not, and has no built-in penalty for failing to announce an effect is real when it is) is sub-optimal. The point is that when the problem is really to make a decision, Bayesian decision-theoretic methods can lead to better choices than inferential methods that were not intended to be used in this way. In this talk I'll present a Bayesian decision analysis of this problem that indicates how clinical-trials practitioners should optimally trade off false-positive and false-negative errors at analysis time. The expected utility maximization with familiar (e.g., Gaussian, t) likelihoods cannot be performed analytically but is tractable numerically.

Approximate Bayesian Computation Using Indirect Inference

Christopher C. Drovandi
Queensland University of Technology, Australia
c.drovandi@qut.edu.au

Anthony N. Pettitt and Malcolm J. Faddy
Queensland University of Technology, Australia
a.pettitt@qut.edu.au m.faddy@qut.edu.au

In this poster we present a novel approach for developing summary statistics for use in approximate Bayesian computation (ABC) algorithms using indirect inference. ABC methods are useful for posterior inference in the presence of an intractable likelihood function. In the indirect inference approach to ABC the parameters of an auxiliary model fitted to the data become the summary statistics. The objective is to search for parameter values of the model of interest that produce similar simulated and observed auxiliary parameter estimates. Although applicable to any ABC technique, we embed this approach within a sequential Monte Carlo (SMC) algorithm that is completely adaptive and requires very little tuning. This SMC ABC algorithm determines the sequence of tolerances adaptively and makes use of an MCMC kernel whose proposal distribution is also generated dynamically. This methodological development was motivated by an application involving data on macroparasite population evolution modelled by a trivariate stochastic process for which there is no tractable likelihood function. The auxiliary model here is based on a Beta-Binomial distribution which incorporates covariates based on time of necropsy and initial larvae injection in both the proportion and overdisperion parameters. The data consist of these covariates as well as the response variable corresponding to mature parasite counts at particular autopsy times for 212 hosts The main objective of the analysis is to determine which parameters of the stochastic model are estimable from the observed data on mature parasite worms, why some parameters are estimated imprecisely and what additional data would produce more precise inferences.

Bayesian Model for Assessing Smoking Exposure During Pregnancy

Vanja M. Dukic
University of Chicago, Chicago, USA
vanja@uchicago.edu

Though studies of effects of prenatal exposure to cigarettes frequently acquire both self-report and biologic assays (such as cotinine levels in urine or blood) of maternal smoking, little attention has been paid to methods for combining information from both sources of exposure measurement. Both measures have their own sources of information and bias – single bioassay measure alone cannot reflect the intricate metabolic mechanism over time, while self-report is subject to reporting, topographic, and metabolic biases. This project presents a new Bayesian statistical model for smoking exposure, based on the combined biological and self-report information, capable of accounting for heterogeneity among women and during pregnancy.

Multivariate Bayesian Logistic Regression for Analysis of Clinical Trial Safety Issues

William DuMouchel
Phase Forward Lincoln Safety Group, USA
wdumouchel@phaseforward.com

This paper describes the method for a model-based analysis of clinical safety data called multivariate Bayesian logistic (MBLR). Parallel logistic regression models are fit to a set of medically related issues, or response variables, and the EB method allows information from the different issues to *borrow strength* from each other. The hierarchical Bayesian prior distribution has four variance components that govern shrinkage of coefficients across model terms, with special focus on treatment effects and treatment by covariate interactions. The method is especially suited to sparse response data, as often occurs when fine-grained adverse events are collected from subjects in trials sized more for efficacy than for safety investigations. A combined analysis of data from multiple trials can be performed and the method enables a search for vulnerable subgroups based on the covariates in the regression model. An example involving 10 medically related issues from a pool of 8 trials is presented.

On Eliciting Priors for Correlated Coefficients in Generalized Linear Models

Fadlalla G. Elfadaly and Paul H. Garthwaite
The Open University, UK
f.elfadaly@open.ac.uk p.h.garthwaite@open.ac.uk

An elicitation method for quantifying opinion about a generalized linear model was developed in Garthwaite and Al-Awadhi (2006). The relationships between predictors and the dependant variable were modeled by piecewise-linear functions or as factors. Regression coefficients were assumed to have a multivariate normal distribution but, to simplify the elicitation task, it was assumed that the regression coefficients for any one predictor were *a priori* independent of those associated with other predictors. The current work aims to relax these independence assumptions.

A method of elicitation is given in which the number of assessments is sufficient to estimate the full covariance matrix. Conditional assessments are elicited to relate beliefs about the different predictors. Conditions needed to guarantee positive definiteness of the covariance matrix have been investigated. A possible drawback of this proposed elicitation method is that the number of assessments can become uncomfortably large if many pairs of covariates are thought to be correlated.

Another method is introduced that models off-diagonal blocks of covariances using a small number of parameters. These parameters aim to capture the broad pattern of correlations between pairs of predictors. The parameters induce all the elements of the covariance matrix and, under mild conditions, the assessed matrix will be positive-definite. For $n > 2$ correlated vectors of coefficients, the expert is required to make only $n(n-1)/2$ additional assessments in order to estimate all the correlations between the vectors. Interactive graphical software is being developed and extended to implement these methods.

Sparsity-Inducing Priors: A Practical Comparison Across Different Types of Data

Barbara E Engelhardt and Matthew Stephens
University of Chicago, USA
engelhardt@uchicago.edu mstephens@uchicago.edu

A large number of sparsity-inducing priors have been introduced in the context of models for matrix factorization, including the exponential distribution, the automatic relevance determination prior, and the spike-slab prior. For the problem of low-dimensional matrix factorization where there are more explanatory variables than number of samples, we compare the performance of these priors on Gaussian data with different types of latent structure. In previous work, we applied matrix factorization models with a range of sparsity-inducing priors to genetic data with latent structure, and found a higher variability in outcome than we expected. We saw that much of the variability in outcome could be explained by the class of latent structure in these data and the type of prior used. In this work, we

generalize these observations by using both simulated and real biological data that conform to different models including a Markov model, a tree-structured model, and block-matrix and Toeplitz matrix covariance structures. We evaluate the performance of each of the sparsity-inducing priors on a number of levels, including consistency, robustness, and sparsity. We find that the performance of each prior varies among the different types of data, and that other model choices, including whether a mean term is used, the choice of an informative versus an uninformative prior, or the choice of the number of factors, appear to have a disproprotionately large impact on the results for some types of data.

Likelihood-Free Estimation of Model Evidence

Richard G. Everitt
University of Bristol, UK
richard.everitt@bristol.ac.uk

Xavier Didelot
University of Warwick, UK
X.Didelot@warwick.ac.uk

Adam M. Johansen
University of Warwick, UK
a.m.johansen@warwick.ac.uk

Dan Lawson
University of Bristol, UK
dan.lawson@bristol.ac.uk

Statistical methods of inference typically require the likelihood function to be computable in a reasonable amount of time. Likelihood-free methods (also called Approximate Bayesian Computation methods, ABC) are able to overcome this requirement, by evaluating the fit of a parameter value according to its capacity to generate data similar to the one observed. Likelihood-free methods have gained in efficiency and popularity in the past few years, following their integration with Monte-Carlo Markov Chain (MCMC) and Sequential Monte-Carlo (SMC) in order to better explore the parameter space. They have been mostly applied to the inference on the parameters of a given model, but can also be useful to compare models. Here we present a likelihood-free method of model selection, which is based on the independent estimation of the marginal likelihood of each model under study. Key advantages of this approach over previous techniques is that it can easily be integrated with SMC, a nd that it does not require to mix across models. We illustrate our method on a toy example where the relative merits of each model is known exactly, and on a realistic example from population genetics.

Bayesian Nonparametric Modeling in Space-Time Emission Tomography.

Mame-Diarra Fall
Université Paris Sud 11, France
fall@lss.supelec.fr

Eric Barat Claude Comtat
Commissariat à l'Energie Atomique, LIST, France
eric.barat@cea.fr claude.comtat@cea.fr

Thomas Dautremer Thierry Montagu
Commissariat à l'Energie Atomique, LIST, France
thomas.dautremer@cea.fr Thierry Montagut@cea.fr

Ali M. Djafari
CNRS/SUPELEC, France
djafari@lss.supelec.fr

Régine Trebossen
CE Atomique, France
regine.trebossen@cea.fr

We address the challenging problem of Positron Emission Tomographic (PET) continuous space-time reconstruction in the context of Poisson inverse problems. Namely, observations are discrete projections of detected random emission locations whose space-time probability distribution has to be estimated. We follow a nonparametric Bayesian approach where regularization of the inverse problem relies entirely on the nonparametric prior. We propose to model the random distribution of recorded events emission locations and arrival times using a dependent Pitman-Yor Mixture (PYM). Each component of the mixture is assumed separable in space and time for brain functional imaging and we use a Normal-Inverse Wishart model as base distribution for the marginalized spatial Pitman-Yor Process. We account for time dependency through a nested PYM of Pólya Trees. The resulting hierarchical nonparametric model allows inference on the so-called functional volumes which define regions of brain whose activity follows a particular kinetic. A key point of the approach is to deal with the infinite representations of the distributions without resorting to arbitrary truncation of models. Though based on an exchangeable Polya urn representation, we develop a conditional sampler for PYM using an update formula from Pitman (1996) and a slice sampling strategy from recent work of Kalli, Griffin and Walker (2009). This scheme is developped for the space-time PYM as for the nested PYM of Pólya Trees. The MCMC algorithm is thus able to generate draws from the posterior distribution of the space-time intensity in order to estimate desired functionals.

Bayesian Inference for Global Sensitivity Analysis and Calibration of Radiative Transfer Models

Marian Farah
University of California, USA
marian@ams.ucsc.edu

Athanasios Kottas
University of California, USA
thanos@ams.ucsc.edu

Robin D. Morris
USRA-RIACS, USA
rdm@riacs.edu

Global process models, typically implemented in complex computer programs, are widely used in geoscience and remote sensing. In particular, radiative transfer models (RTMs) simulate the interaction of sunlight with vegetation and are used, in conjunction with satellite measurements, to produce key data products associated with vegetation, such as Leaf Area Index (LAI). Bayesian statistics has been used to develop an array of tools to tackle sensitivity analysis and calibration of computationally expensive simulators through building an emulator, which is a statistical approximation of the computer model output. We present a framework that unifies the various Gaussian process based global sensitivity analysis tools and extends semi-Bayesian approaches to a fully Bayesian methodology. The approach is used to carry out sensitivity analysis of the Leaf-Canopy Model (LCM), a specific RTM with inputs that include leaf chemistry variables (e.g., chlorophyll, water fraction, protein) and canopy architecture variables (e.g., LAI and leaf angle distribution). The resulting inference for the main effects and sensitivity indices of the LCM inputs can be used to study the influence of each input and how uncertainty in the output is apportioned amongst the inputs. Moreover, we develop methods for LCM calibration to estimate the distributions of its inputs with particular emphasis on LAI. The modeling approach to calibration is based on combination of LCM simulator data with satellite reflectance measurements. Both types of data are available at 8 wavelengths, which are associated with MODIS (a key instrument aboard the Terra and Aqua satellites) spectral bands that are sensitive to vegetation.

Dynamic Bayesian Smooth Transition Auto-Regression Models

Alvaro Faria
The Open University, UK
a.e.faria@open.ac.uk

Smooth transition autoregressive (STAR) models are non-linear regime switching models that generalise threshold autoregressive (TAR) as well as self-exciting autoregression (SETAR) models by substituting their step transition with smooth transition functions. A STAR model can be seen as a mixture of AR models with mixing weights being smooth transition functions that assess the degree of nonlinearity in the time series. STAR models have been usually applied to the forecasting of econometric time series with inherent non-linear behavior such as

unemployment rates, industrial production and disaggregated electricity consumption and prices data as well as in the modelling of asymmetric business cycles. Despite their good performances in many of such applications they present inherent estimation problems that can jeopardise their practical appeal. In particular the estimation of the smoothing parameter is notoriously difficult requiring a large number of observations for an approximate grid search.

To address some of the estimation problems with STAR models, we investigate a class of dynamic Bayesian smooth transition autoregressive (DBSTAR) models. Those models, unlike existing Bayesian formulations for the STAR model based on numerical integration methods, estimate parameters in analytical closed form. Also, alternatively to the usual logistic and exponential transition functions, we investigate beta density forms for the smooth regime switching functions.

Dynamic Multiscale Spatio-Temporal Models for Areal Data

Marco A. R. Ferreira
University of Missouri, USA
ferreiram@missouri.edu

Scott H. Holan
University of Missouri, USA
holans@missouri.edu

Adelmo I. Bertolde
Univ. Fed. do Espírito Santo, Brazil
adelmo@cce.ufes.br

We introduce a new class of dynamic multiscale models for spatio-temporal processes arising from areal data. More specifically, we use nested geographical structures in order to decompose the data using an extension of the Kolaczyk-Huang multiscale decomposition. In our methodology, the original process is decomposed into multiscale coefficients which evolve through time following state-space equations. Our approach naturally accommodates data observed on irregular grids as well as heteroscedasticity. Moreover, we propose a multiscale spatio-temporal clustering algorithm that facilitates construction of the nested geographical multiscale structure. In addition, we propose a singular forward filter backward sampler for efficient Bayesian estimation. Our multiscale spatio-temporal methodology decomposes large data-analysis problems into many smaller pieces and thus leads to scalable and highly efficient computational procedures. Finally, we illustrate the utility and flexibility of our dynamic multiscale framework through two spatio-temporal applications. The first example examines agricultural production in Espírito Santo State Brazil whereas the second example considers mortality ratios in the state of Missouri.

Bayes in the Court

Manuela Figueira-Neves
Inst. Politéc. da Guarda, Portugal
mfigueira@ipg.pt

Fernando Rosado
Universidade de Lisboa, Portugal
fernandorosado@fc.ul.pt

In recent decades the use of statistical methods for legal purposes had large and rapid development and continues today. Statisticians are called to court

as experts or consultants. They evaluate and quantify the uncertainty. When a crime is committed different types of evidence may be found in the crime scene. The great concern is the assessment of the available evidence, found in the crime scene or in the possession of the suspect to measure the possibility of the suspect being in fact the criminal. Different approaches can be used in the evaluation of the evidence. The significance probabilities and Bayesian approach are the more used. The first one uses significant tests and has less followers and the Bayesian approach appears the most used in scientific works of forensic analysis. In fact, most works in Forensic Statistics are based and use Bayesian methodologies. Almost all literature refers to Lindley (1977) as being the pioneer in the use of statistical tools within the Forensic Science by describing a method to evaluate the evidence on the fragments of glass based on the use of the Bayesian theory. It was the first time the Bayesian Statistics was used in the context of Forensic Science. One aspect of the Bayesian approach is its formalization of subjective impressions as an ingredient of statistical analysis. One of the earliest discussions of the Bayesian approach to outliers is that of de Finetti (1961). He is primarily concerned with exploring basic attitudes rather than developing technique. Unfortunately, de Finetti does not work through the details of any particular special cases nor does he address the question of how to test whether any outliers are a posteriori evident in the data. One formulation for dealing with the problem of testing for the presence of outliers is very important both in the non-Bayesian and Bayesian point of view.

Inference for Gaussian Process Emulation of Oil Reservoir Simulation Codes

Maurizio Filippone
University of Glasgow, UK
maurizio@dcs.gla.ac.uk

Ben Calderhead and Mark Girolami
University of Glasgow, UK
bc@dcs.gla.ac.uk girolami@dcs.gla.ac.uk

Linah Mohamed and Mike Christie
Keriot-Watt University, UK
linah.mahgoub@pet.hw.ac.uk mike.christie@pet.hw.ac.uk

Computer models aim at building a description of a system with some degree of accuracy. Combining observations from the system with results from simulations through computer codes in artificial scenarios, can provide a way to gain some understanding of the process underlying the system.

In petroleum reservoir modeling, for example, it is important to calibrate computer models to fit real observations, as well as evaluate the uncertainty in both the model parameters and the predictions of future observations. The main problems in the oil industry are that the data is expensive to gather, so tends to be limited in extent, and the models, which are based on the solution of non-linear partial differential equations, are expensive to run thus limiting the use of some MCMC techniques that require large numbers of simulations.

We study the behaviour of a simple three parameter oil reservoir model originally created by Tavassoli, Carter and King (*SPE Journal*, 2004), using the approach proposed in Kennedy and O'Hagan (*JRSSB* 2001) In this method, the output of the real system is modeled as a combination of the output of the computer code and a term accounting for the model inadequacy, with both terms modeled using Gaussian Processes. Additionally, a simple i.i.d. noise model is considered to account for the noise in the measuring process. The reservoir model is known to have a non-trivial response with respect to the input parameters, due to the underlying nonlinear partial differential equations defining it. Advanced inference techniques are necessary to estimate the uncertainty in the model parameters along with the model inadequacy. We compare a variety of sampling techniques to obtain posterior distributions over the parameters of the model. In particular, we study the application of the recently proposed Riemannian Manifold Hamiltonian Monte Carlo (Girolami, Calderhead and Chin, 2009) to this problem.

Indian Buffet Epidemics: A Bayesian Approach to Modelling Heterogeneity

Ashley P. Ford and Gareth O. Roberts
Warwick University, UK
Ashley.Ford@warwick.ac.uk

The application of mathematical and computer models to the prediction of epidemics in real time is often lacking the crucial stage of statistical inference. A combination of statistical inference on over simplified models and ad-hoc parameter estimates for complex models is frequently used. Analysis of data from the the 1916 and 2009 influenza epidemics in the UK shows that simple stochastic epidemic models do not fit the observations and so predictions based on these models could be misleading. There is a clear need for techniques for inference on models which lie between the extremes of over simplification and too complex for inference.

Parameters of stochastic epidemic models are subject to both variation and uncertainty, in the case of a new strain of a disease this uncertainty is greater. The incorporation of prior information should be done using Bayesian techniques rather than the ad-hoc point estimates currently used.

The Indian Buffet Epidemic model has been developed to address the need for a model which is more suitable than assuming homogeneous mixing or an incorrect network model. The aim is to have a process which fits the heterogeneity and two or three parameters that measure the departure from homogenity.

The Indian Buffet Epidemic combines a bipartite network model with the Indian Buffet proces to provide a realistic model which is simple to define and simulate from and on which Bayesian inference is possible. The model assumes that there are a large number of potential classes, individuals belong to a subset of these classes. The classes might be households, schools, clubs, buses etcetera, an important feature of this class of models is that the classes do not need to be specified. Within each class infection occurs homogeneously and recovery is as in the basic SIS or SIR model. A significant difference between this approach and others is the use of the Indian Buffet process to define a prior over an unknown

and potentially infinite class membership matrix rather than trying to determine a priori a meaningful and useful set of classes.

This poster describes the model and develops an MCMC algorithm for deriving parameter estimates. An important aspect is the development of a new proposal distribution for large binary matrices. The algorithm is demonstrated on a range of simulated data from both the true model and other epidemic models and comparisons made between centered and non-centered representations for the augmented data.

Bayesian Methods for Analyzing Extrasolar Planet Observations

Eric B. Ford
University of Florida, USA
eford@astro.ufl.edu

M. Ghosh, P. Guo, A. V. Moorhead, B. E. Nelson, M. J. Payne, D.Veras and J. Wang
University of Florida, USA

The discovery of extrasolar planetary systems is providing insights into the formation of our solar system and humanity's place in the universe. Most known planets were detected indirectly by measuring periodic changes in the observable properties of the planet's host star. We present an overview of recent research on statistical methods for analyzing such exoplanet observations, including: (i) a Bayesian surrogate model for analyzing Doppler (radial velocity) and transit timing observations; (ii) Bayesian and quasi-Bayesian methods for population analysis (*e.g.*, distribution of orbital eccentricities), (iii) Differential Evolution Markov chain Monte Carlo methods for performing Bayesian parameter estimation based on self-consistent n-body integrations, and (iv) robust estimators for model comparison from posterior samples.

Invariant Priors for Variable Selection

Anabel Forte
Universitat de València, Spain
anabel.forte@uv.es

Susie Bayarri
Universitat de València, Spain
susie.bayarri@uv.es

James O. Berger
Duke University, USA
berger@stat.duke.edu

Gonzalo García-Donato
Univ. Castilla La Mancha, Spain
gonzalo.garciadonato@uclm.es

Berger (1985) proposed a "Robust prior for normal means", in an estimation scenario. This distribution has two important features which make it very interesting also as a model selection prior: having thick tails and being a proper density. In this paper we choose its hyper-parameters in an 'objective' way so as to achieve desirable properties as a convenient model selection priors. We also show the

relationship of this prior with other proposals in literature as hyper-g-priors from Liang et al (2008) and its close connection to Cauchy priors. A main advantage of these robust priors is that a particular choice of hyper-parameters not only achieve optimal properties but also produce close-form expressions for the Bayes factor making it very attractive for practical scenarios. We apply it to a couple small problems and to a large, but enumerable, problem

Area-Specific Decision Rules for Classifying Unusual Spatial Risks of Disease

Léa Fortunato
Imperial College London, London, UK
l.fortunato@imperial.ac.uk

Nicky Best and Sylvia Richardson
Imperial College London, London, UK
n.best@imperial.ac.uk sylvia.richardson@imperial.ac.uk

Small area disease mapping studies have become an established technique in spatial epidemiology to describe spatial patterns of diseases and to highlight unusual areas. Because most environmental risks are small, the power to detect unusual risks is limited. However, these limitations can be partially overcome by exploiting the full posterior distribution of the relative risk (RR) estimates. In practice, an area is classified as 'high risk' if the posterior probability $P(RR_i > \theta) > p_{cut}$, where θ and p_{cut} are the threshold and the probability cut-off, respectively. Richardson et al (2004) have shown that using this decision rule with $\theta = 1$ and p_{cut} between 0.7 and 0.8 gives good properties in terms of sensitivity and specificity in the case of moderate expected numbers (between 10 and 20). However, when the disease is rare and/or the expected numbers of cases are small, the decision rules are underpowered to detect excess risks (Richardson *et al.* 2004), suggesting the performance of a decision rule depends on the expected numbers.

We propose to use a decision rule with area-specific thresholds and probability cut-points, to take into account the variability of the expected numbers. The area-specific threshold θ is calculated using the posterior distribution of the RR under the null hypothesis of no variability in the risks in the study area. The probability cut-point is obtained from the false discovery rate, using a simulation-based method. This approach will be compared to the standard methods using the incidence of lung cancer in South East England.

Incorporating Cost in Bayesian Variable Selection, with Application to Cost-Effective Measurement of Quality of Health Care

Dimitris Fouskakis
National Technical University of Athens, Greece
fouskakis@math.ntua.gr

Ioannis Ntzoufras
Athens Univ. Econ. Busi, Greece
ntzoufras@aueb.gr

David Draper
Univ. California, Santa Cruz, USA
draper@ams.ucsc.edu

In the field of quality of health care measurement, one approach to assessing patient sickness at admission involves a logistic regression of mortality within 30 days of admission on a fairly large number of sickness indicators (on the order of 100) to construct a sickness scale, employing classical variable selection methods to find an "optimal" subset of 10–20 indicators. Such "benefit-only" methods ignore the considerable differences among the sickness indicators in cost of data collection, an issue that is crucial when admission sickness is used to drive programs (now implemented or under consideration in several countries, including the U.S. and U.K.) that attempt to identify substandard hospitals by comparing observed and expected mortality rates (given admission sickness). When both data-collection cost and accuracy of prediction of 30-day mortality are considered, a large variable-selection problem arises in which costly variables that do not predict well enough should be omitted from the final scale.

In this work (a) we develop a method for solving this problem based on posterior model odds, arising from a prior distribution that (1) accounts for the cost of each variable and (2) results in a set of posterior model probabilities which corresponds to a generalized cost-adjusted version of the Bayesian information criterion (BIC), and (b) we compare this method with a decision-theoretic cost-benefit approach based on maximizing expected utility. We use reversible-jump Markov chain Monte Carlo (RJMCMC) methods to search the model space, and we check the stability of our findings with two variants of the MCMC model composition (MC^3) algorithm. We find substantial agreement between the decision-theoretic and cost-adjusted-BIC methods; the latter provides a principled approach to performing a cost-benefit trade-off that avoids ambiguities in identification of an appropriate utility structure. Our cost-benefit approach results in a set of models with a noticeable reduction in cost and dimensionality, and only a minor decrease in predictive performance, when compared with models arising from benefit-only analyses.

Additionally to the above, the practical relevance of the selected variable subsets is ensured, by enforcing an overall limit on the total data collection cost of each subset: the search is conducted only among models whose cost does not exceed this budgetary restriction. Trying to implement usual model search algorithms, will frequently fail if the actual best model is outside the imposed cost limit and when collinear predictors with high predictive ability are present. The reason for this failure is the existence of multiple modes with movement paths that are forbidden due to the cost restriction. Therefore, a population based trans-dimensional reversible-jump Markov chain Monte Carlo algorithm (population RJMCMC) is developed, where ideas from the population-based MCMC and

simulated tempering algorithms are combined. Comparing the proposed technique with the simple RJMCMC we notice that population RJMCMC algorithm moves successfully and more efficiently between distinct neighborhoods of "good" models and achieves convergence faster.

All results are phrased in the language of health policy but apply with equal force to other quality assessment settings with dichotomous outcomes, such as the the study of retention rates in the workplace and the creation of cost-effective credit scores in business.

Bayesian Nonparametric Time Series Models for Complex Dynamical Phenomena

Emily B. Fox
Duke University, USA
fox@stat.duke.edu

Markov switching processes, such as hidden Markov models (HMMs) and switching linear dynamical systems (SLDSs), are often used to describe rich classes of dynamical phenomena. They describe complex temporal behavior via repeated returns to a set of simpler models: imagine, for example, a person alternating between walking, running and jumping behaviors, or a stock index switching between regimes of high and low volatility.

Traditional modeling approaches for Markov switching processes typically assume a fixed, pre-specified number of dynamical models. Here, in contrast, I develop Bayesian nonparametric approaches that define priors on an unbounded number of potential Markov models. In the talk, I specifically focus on discovering a set of latent dynamical behaviors that are shared among multiple time series. Using a beta process prior, my approach allows the size of the set and the sharing pattern to both be inferred from data. I develop an efficient Markov chain Monte Carlo inference method that is based on the Indian buffet process representation of the predictive distribution of the beta process. In particular, my approach uses the sum-product algorithm to efficiently compute Metropolis-Hastings acceptance probabilities, and explores new dynamical behaviors via birth/death proposals. I demonstrate promising results on the challenging application of segmenting visual motion capture data.

Combining Expert Judgement

Simon French
Manchester Business School, UK
simon.french@mbs.ac.uk

In a paper at Valencia 2, I distinguished three contexts in which one might wish to combine expert judgements of uncertainty: the expert problem; the group decision problem; and the textbook problem. Over the intervening years much has been written on the first two, which have the focus of a single decision

context, but little on the third, though the closely related field of meta-analysis has developed considerably. The textbook problem, however, seems to be gaining in importance since data and expert judgements can be made available over the web to be used by many different individuals to shape their own beliefs in many different contexts. Participatory approaches to societal risk analysis and decision making mean that many people of varying statistical and analytic sophistication may draw summaries of expert judgements into deliberations on a range of topics. There are a range of issues here that the Bayesian community need address before mistakes are made. I argue that we need a science of meta-analysis for expert judgements combined with some principles for reporting expert judgements.

Detection of Outliers and Interventions in INGARCH Time Series

Roland Fried
Technische Universität Dortmund, Germany
fried@statistik.tu-dortmund.de

Inoncent Agueusop
TU Dortmund Univ., Germany
agueusop@yahoo.fr

Björn Bornkamp
TU Dortmund Univ, Germany
bornkamp@statistik.tu-dortmund.de

Konstantinos Fokianos
University of Cyprus, Cyprus
fokianos@ucy.ac.cy

Katja Ickstadt
ickstadt@statistik.tu-dortmund.de
TU Dortmund Univ, Germany

An INGARCH(1,1) process for a count time series $\{Y_t\}$ is defined through the relationships $Y_t|\mathcal{F}_{t-1}^Y \sim \text{Poisson}(\lambda_t)$; $\lambda_t = \beta_0 + \beta_1 Y_{t-1} + \alpha_1 \lambda_{t-1}$, for $t \geq 1$, with $\beta_0, \beta_1, \alpha_1 > 0$ and $\alpha_1 + \beta_1 < 1$ for stationarity (Ferland *et al.* 2006). Here, the σ–field $\mathcal{F}_t^Y$ generated by $\{Y_0, \ldots, Y_t, \lambda_0\}$ represents the information up to time t. A first paper on intervention effects in INGARCH-models based on the concept of conditional likelihood is Fokianos and Fried (2010). They assume that instead of the "clean" INGARCH process $\{Y_t\}$ we observe a contaminated process $\{Z_t\}$, which includes the effect of an intervention at time τ, $Z_t|\mathcal{F}_{t-1} \sim \text{Poisson}(\kappa_t)$; $\kappa_t = \beta_0 + \beta_1 Z_{t-1} + \alpha_1 \kappa_{t-1} + \nu X_t$, for $t \geq 1$, where ν is the size of the intervention effect, $X_t = I(t = \tau)$ for a spiky outlier or $X_t = I(t \geq \tau)$ for a level shift, and $\mathcal{F}_t$ is the σ–field generated by $\{Z_0, \ldots, Z_t, \kappa_0\}$. A drawback of this model is that only intervention effects which influence all future observations through the dynamics of the process are considered. Pure additive outliers describing e.g. measurement errors are excluded, because the conditional likelihood is hard to obtain if Z_{t-1} is replaced by Y_{t-1} in the equation for κ_t. A Bayesian version of model above is $Z_t \sim \text{Pois}(\kappa_t) + \delta_t \text{Pois}(\omega) + I(t \geq \tau)\text{Pois}(\tilde{\omega})$, $\kappa_t = \beta_0 + \beta_1 Z_{t-1} + \alpha_1 \kappa_{t-1}$, where $\text{Pois}(\mu)$ denotes a Poisson distributed random variable with mean μ, and the δ_t are independent Bernoulli random variables with $P(\delta_t = 1) = \pi_t$ and τ is a random changepoint time. We use gamma priors for β_0, ω and $\tilde{\omega}$, a Dirichlet prior for (β_1, α_1) and a uniform distribution for τ. This model can be implemented in a straightforward manner in WinBugs.

Within the Bayesian framework, we can treat purely additive outliers influencing only single observations via $Z_t = Y_t + I(\delta_t = 1)\text{Pois}(\omega)$, $Y_t|\mathcal{F}_{t-1}^Y \sim$

Poisson(λ_t); $\lambda_t = \beta_0 + \beta_1 Y_{t-1} + \alpha_1 \lambda_{t-1}$, so that the outliers do not influence the dynamics of the process. We have implemented this modified model in R, using the same prior distributions as above. For simulating values from the posterior distribution we apply a componentwise Metropolis-Hastings algorithm.

We compare the approaches based on the models described via application to real and simulated data sets, either setting all π_t to the same small value, or generating the π_t from independent $Beta$-distributions.

A Bayesian Nonparametric Modeling Approach to Risk Assessment from Developmental Toxicity Studies

Kassandra Fronczyk and Athanasios Kottas
University of California, Santa Cruz, USA
kassief@ams.ucsc.edu thanos@ams.ucsc.edu

We develop a Bayesian nonparametric mixture modeling framework for replicated count responses in dose-response settings. We explore this methodology with applications in developmental toxicity studies, in which the primary objective is to determine the relationship between the level of exposure to a toxic chemical and the probability of a physiological or biochemical response, or death. Data from these experiments typically involve features that can not be captured by standard parametric approaches. To provide flexibility in the functional form of both the response distribution and the probability of positive response, the proposed mixture model is built from a dependent Dirichlet process prior, with the dependence of the mixing distributions governed by the dose level. The methodology is tested with a simulation study, which involves also comparison with semiparametric Bayesian approaches to highlight the practical utility of the dependent Dirichlet process nonparametric mixture model. Further illustration is provided through the analysis of data from a developmental toxicity study on the effects of 2,4,5-trichlorophenoxiacetic acid.

Bayesian Variable Selection for Random Intercept Modeling of Gaussian and non-Gaussian Data

Sylvia Frühwirth-Schnatter and Helga Wagner
Johannes Kepler Universität Linz, Austria
sylvia.fruehwirth-schnatter@jku.at helga.wagner@jku.at

The paper considers Bayesian variable selection for random intercept models both for Gaussian and non-Gaussian data and shows how this problem is closely related to the appropriate choice of the distribution of heterogeneity. While most papers view the choice of the distribution of heterogeneity only in the light of efficient estimation of all fixed parameters, we show that additionally individual shrinkage for the random effects toward 0 can be achieved through the appropriate selection of the prior of the random effects.

If, for instance, a Laplace rather than a normal prior is considered, we obtain a Bayesian Lasso random effects models which allows both smoothing and, additionally, individual shrinkage. In addition, we investigate alternatively shrinkage-smoothing priors like the spike-and-slab random effects model which is closely related to a finite mixtures of random effects model. We study spike-and-slab random effects models with both an absolutely continuous and a Dirac spike and provide details of MCMC estimation for all models.

Simulation studies comparing the various priors show that the spike-and-slab random effects model outperforms unimodal, non-Gaussian priors as far as correct classification of non-zero random effects is concerned and that there is surprisingly little difference between an absolutely continuous and a Dirac spike. The choice of appropriate component densities, however, is crucial and we were not able to identify a uniformly best distribution family. The paper concludes with an application to ANOVA for binomial data using a logit model with a random intercept.

Clustering and Characterizing Temporal Profiles of Gene Expression

Audrey Q. Fu
University of Cambridge, UK
qf205@cam.ac.uk

A. Krejci, B. Housden, B. Fischer, F. Bernard
S. Bray, S. Russell and S. Tavaré
University of Cambridge, UK

We aim to cluster a set of temporal profiles of gene expression and to characterize features of identified clusters. The temporal profiles in our study generally have small magnitude and large noise. Existing computer programs for clustering we have experimented with turn out to be inadequate: they tend to identify a large number of clusters with several potentially interesting clusters being very heterogeneous. A possible explanation for this type of results is that the variance structure defined in those programs may be too general for the type of data we have. To achieve a cleaner clustering, we propose a more detailed decomposition of total variance.

We further propose to describe the underlying process of gene expression changes as a stochastic process, summarized by stochastic differential equations (SDEs) with just a few parameters that have biological interpretations. The use of SDEs allows us to summarize clusters more concisely and effectively. To estimate the number of clusters as well as cluster-specific parameters, we use a Dirichlet-process-based Markov chain Monte Carlo (DPMCMC) method, as well as methods developed in Stephens (2000) to tackle label-switching. We apply this clustering scheme to a data set collected for thousands of genes from Drosophila at nearly 20 time points.

The Influence of the Criteria in the Bayesian AHP

Pilar Gargallo
Universidad de Zaragoza, Spain
altuzarr@unizar.es

A. Altuzarra J.M. Moreno and M. Salvador
Universidad de Zaragoza, Spain
altuzarra@unizar.es moreno@unizar.es salvador@unizar.es

This work proposes different methodologies for measuring the influence of a set of criteria on the final priorities of the Analytic Hierarchy Process (AHP) in a global context (a hierarchy). The priorities have been obtained by means of the Bayesian prioritization procedure of Altuzarra et al. (2007). Cross-validation methods have been used when measuring the influence. The methodology is illustrated by means of an empirical example.

Adaptive Optimal Scaling of Metropolis-Hastings Algorithms using the Robbins-Monro Process

Paul H. Garthwaite
Open University, UK
p.h.garthwaite@open.ac.uk

Yanan Fan and Scott A. Sisson
University of New South Wales, Australia
y.fan@unsw.edu.au Scott.Sisson@unsw.edu.au

We present an adaptive method for the automatic scaling of Random-Walk Metropolis-Hastings algorithms. Our method relies on the use of the Robbins-Monro search process, whose performance is determined by an unknown steplength constant. We give a very simple estimator of this constant for proposal distributions that are univariate or multivariate normal, together with detail for automating the algorithm. The effectiveness of the algorithm is demonstrated with both simulated and real data examples. The algorithm is a quick robust method for finding the scaling factor that yields a specified acceptance probability. One of its potential uses is in software developed for others to use, as it means the user does not need to tune the variance of the proposal distribution. But it could also form part of a tool in complex adaptive algorithms.

Uncovering Selection Bias in Case-Control Studies using Bayesian Poststratification

Sara Geneletti
London School of Economics, UK
s.geneletti@lse.ac.uk

Nicky Best Mireille Toledano
Imperial College London, London, UK
{n.best, m.toledano}@imperial.ac.uk

Paul Elliott and Sylvia Richardson
Imperial College London, London, UK
{p.elliott, sylvia.richardson}@imperial.ac.uk

Selection bias can present a serious problem for valid odds ratio estimation in case-control studies. These are especially sensitive to selection bias as the selection mechanism depends on the case/control status of the participants. If, in addition, the selection mechanism is also associated to the exposure, then selection bias can occur. As we typically have limited information on the distribution of the exposure, we are unable to adjust for selection bias.

In order to overcome this problem, we introduce a set of variables B such that first, B separates the exposure from the selection, and second, such that the distribution of B can be estimated from sources of data external to the study. In so doing we shift the selection bias from the exposure, whose distribution cannot be estimated without bias from the study, to B, whose distribution is estimated from data external to the study and thus potentially unbiased. This method is a type of post-stratification and arises naturally from the conditional independence assumptions used to find suitable candidates B. By using different sources of data to estimate the distribution of B we can investigate the sensitivity of the odds ratio to different selection processes. This aspect is the novelty and strength of this approach as it encourages us to think carefully about the populations of interest and the sources of bias.

We perform this sensitivity analysis on two case-control studies, one investigating the association between Hypospadias and exposure to hairspray and another investigating the association between Childhood acute lymphoblasitc leukaemia and exposure to electromagnetic fields.

Building Genuine Beliefs Into a Prior Distribution for the Covariance Matrix

Sarah E. Germain
Newcastle University, UK
s.e.germain@ncl.ac.uk

Malcolm Farrow and Richard J. Boys
Newcastle University, UK
malcolm.farrow@ncl.ac.uk richard.boys@ncl.ac.uk

In many analyses of multi–dimensional data, the dependence structure of the multivariate normal distribution is used to build relationships between variables, sometimes directly, but often indirectly through hierarchical structures. Constructing a prior distribution for the covariance matrix of this distribution is not straightforward because of the large number of parameters it can contain and the constraint that the matrix be positive definite, especially when the prior is intended to convey substantive information.

The inverse Wishart and generalised inverse Wishart priors are conjugate and therefore convenient choices. However, the paucity of hyperparameters in the former makes it inflexible, whilst the interpretation of the parameters in the latter is generally only helpful when an order can be imposed amongst the variables. Various non–conjugate alternatives have been proposed in which priors are placed on less constrained parameterisations of the covariance matrix, such as its matrix logarithm, but their utility is often limited if an intuitive understanding of the new parameters cannot be found.

We discuss how to construct a prior distribution for a positive definite matrix which does not impose any additional constraints, where the assessments required refer to interpretable quantities, and which is sufficiently flexible to allow specification of different degrees of certainty and prior associations for components of the matrix. We discuss a number of approaches, including the use of common and specific individual variation factors, and propose modifications to overcome disadvantages.

These ideas are illustrated through an example involving spatial data, where there is typically no natural ordering of the variables but where we may well have different degrees of association in our prior beliefs between different pairs of covariances.

Bayesian Variable Selection for Latent Class Models

Joyee Ghosh
University of North Carolina at Chapel Hill, USA
joyee@email.unc.edu

Amy H. Herring and Anna-Maria Siega-Riz
University of North Carolina at Chapel Hill, USA
amy_herring@unc.edu am_siegariz@unc.edu

In this work we develop a latent class model with class probabilities that depend on subject-specific covariates. One of our major goals is to identify important predictors of latent classes. We consider methodology that allows estimation of latent classes while allowing for variable selection uncertainty. We propose a Bayesian variable selection approach and implement a stochastic search Gibbs sampler for posterior computation to obtain model averaged estimates of quantities of interest such as marginal inclusion probabilities of predictors. Our methods are illustrated through simulation studies and application to data on weight gain during pregnancy, where it is of interest to identify important predictors of latent weight gain classes.

Bayesian Shape Restricted Regression with Multivariate Bernstein Polynomials

Sujit K. Ghosh and Jiangdian Wang
North Carolina State University, Raleigh, USA
ghosh@stat.ncsu.edu jwang8@ncsu.edu

One of the fundamental problems in statistics consists of determining the relationship between a response variable and a set of predictor variables through a regression function. Prior scientific knowledge is often available that suggests the regression function should have a certain shape (e.g. monotone, convex or concave etc.) but not necessarily a specific parametric form. Examples of this can be found in nearly all areas of applied statistics. In biomedical applications, dose response models are assumed to be non-decreasing with the possibility that the relationship is flat over certain (unknown) regions. In multivariate survival analysis, the survivor or cumulative hazard functions are coordinate?wise non-increasing.

Majority of the existing nonparametric methods that have been developed to address some of the above mentioned shape restricted curve estimation have at least one of the following limitations: (a) applicable only to monotone function of a single predictor, (b) interval estimates are typically unavailable, (c) naïve bootstrap methods can be inconsistent when the true curve has a flat portion, or (d) multivariate shape restrictions are not preserved. Multivariate Bernstein polynomials have been shown to be useful in imposing certain shape restrictions on regression functions. In this work, we make use of several attractive properties of Bernstein polynomials to develop a general class of shape restricted nonparametric regression models to overcome the above mentioned limitations. In particular, we develop a novel prior distribution for Bernstein coefficients that

allows the posterior estimates of the curve to adapt to a flat portion of the regression function. One major advantage of the proposed Bayesian procedure is that it allows us to compute interval estimates of the curve that also maintains the same shape restriction as the underlying curve. We demonstrate the effectiveness of our method through simulations and the analysis of several real data sets.

Bayesian Inference for Spartan Spatial Random Fields

Sujit K. Ghosh
North Carolina State University
ghosh@stat.ncsu.edu

Gentry White
The University of Queensland
gentry.white@uq.edu.au

Dionissios Hristopulos
Technical University of Crete
dionisi@mred.tuc.gr

We develop an innovative Bayesian methods for the statistical analysis of spatial variability based on Spartan Spatial Random fields (SSRFs). The SSRFs are inspired by methodology from statistical physics, where the distribution of fluctuations follows from the energy functional (the Hamiltonian), rather than a data-driven covariance matrix. The term Spartan denotes frugal need of free parameters (three), which have clear physical interpretations. The SSRF prior also is unique in spatial modelling as it has a conjugate prior for the parameters that allows for adaptive quadrature based numerical integration to directly evaluate the posterior without relying on Monte Carlo based methods, that can be time consuming and less accurate. The results are demonstrated to be robust in their ablity to recaputure true parameter values from simulated data. The resulting priors are then demonstrated on a real world example of Swiss rainfall data. Results show that the SSRF outperforms other geostatistical methods under several criteria.

Bayesian Inference for Generalized Stochastic Population Growth Models with Application to Aphids

Colin S. Gillespie and Andrew Golightly
Newcastle University, UK
c.gillespie@ncl.ac.uk a.golightly@ncl.ac.uk

A field study was conducted on the population numbers of cotton aphids (Aphis gossypii Glover). The study consisted of three irrigation levels (Low, Medium and High), three nitrogen fertility treatments (blanket nitrogen, variable nitrogen and no nitrogen) and three field blocks. At five weekly intervals the numbers of aphids were counted at each treatment combination. This gives a total of twenty-seven data sets. This paper explores parameter inference for a stochastic population growth model of aphids. It is believed that the death rate of the aphid population depends on the unobserved cumulative population size, whilst the birth depends on the current population size. The aim of this study is too

investigate how the treatment effects manifest themselves within the birth and death rates. Once interactions effects are considered, this involves fitting thirty-six parameters and estimating the unobserved cumulative aphid population. Markov chain Monte Carlo methods, coupled with a moment closure approximation are used to integrate over uncertainty associated with the unobserved component and estimate parameters. We highlight that blocking effects and interaction terms play a crucial role in understanding aphid dynamics.

Differential Geometric Structure in MCMC

Mark Girolami
University of Glasgow, UK
girolami@dcs.gla.ac.uk

Ben Calderhead
University of Glasgow, UK

Gareth Roberts
University of Warwick, UK

Markov Chain Monte Carlo (MCMC) theory and methodology has developed continually over the last sixty years and has had major impact on Bayesian statistical inference during this time. And yet there are still many open issues in MCMC methodology that restrict the possible classes of inferential problems which can be realistically addressed. This is especially the case in problems where target densities are very high dimensional and exhibit strong correlational structure as in spatial epidemiology for example. To address these issues in a theoretically systematic and practically convincing manner we turn to Differential Geometric arguments which have been employed to great effect in the study of statistical asymptotics, e.g. Efron, 1975; Kass, 1989; Amari, 1990, but have never been considered specifically for MCMC methodology.

In this talk we present methods for the design of proposal mechanisms and densities based on differential geometric principles which provide generalisations for Metropolis based methods, furthermore we will demonstrate their excellent statistical efficiency on complex inferential problems that have been particularly challenging for existing methodology. As an illustrative example let us consider the Metropolis Adjusted Langevin Algorithm (MALA). Tuning of the proposal is a well known difficulty in this particular method, as is the case with all non-Gibbs based MCMC. However, let us note that the space of probability distributions is defined by a Riemannian manifold with a metric tensor defined by the Fisher Information, Rao 1945. This being the case it is appropriate to define the Langevin diffusion, with invariant measure $\pi(\mathbf{x}), \mathbf{x} \in \mathbf{R}^n$, in local coordinates defined by the contravariant second-order metric tensor associated with the Fisher Information of $\pi(\mathbf{x})$ as $\mathbf{G}^{-1}(\mathbf{x}) = E\{\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x})\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x})^{\mathsf{T}}\}^{-1}$. The discrete Euler integrator which forms the basis of the proposal density follows as

$$\begin{aligned}\mathbf{x}_i &\leftarrow \mathbf{x}_i + \frac{\epsilon}{2}\left(\mathbf{G}^{-1}(\mathbf{x})\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x})\right)_i - \epsilon\sum_{j=1}^{n}\left(\mathbf{G}^{-1}(\mathbf{x})\frac{\partial \mathbf{G}(\mathbf{x})}{\partial \mathbf{x}_j}\mathbf{G}^{-1}(\mathbf{x})\right)_{ij} \\ &+ \frac{\epsilon}{2}\sum_{j=1}^{n}\left(\mathbf{G}^{-1}(\mathbf{x})\right)_{ij} Tr\left(\mathbf{G}^{-1}(\mathbf{x})\frac{\partial \mathbf{G}(\mathbf{x})}{\partial \mathbf{x}_j}\right) + \left(\sqrt{\epsilon\mathbf{G}^{-1}(\mathbf{x})}\mathbf{z}\right)_i\end{aligned}$$

For a density existing on a flat Riemann manifold with constant metric tensor this reduces further to the standard pre-conditioned MALA proposal $\mathbf{x} \leftarrow \mathbf{x} + \frac{\epsilon}{2}\mathbf{G}^{-1}\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}) + \sqrt{\epsilon\mathbf{G}^{-1}}\mathbf{z}$ where the pre-conditioning matrix $\mathbf{G}^{-1}$ is now defined correctly based on the expected information matrix of the target density. At the meeting extensive and thorough experimental evaluation on a range of complex and challenging inferential tasks will be presented to assess and highlight the potential of exploiting differential geometric structure in MCMC.

Approximations to Bayes Factors for Intrinsic Priors: Generalizing the BIC

Javier Girón
Universidad de Málaga, Spain
fj_giron@uma.es

Elías Moreno
Universidad de Granada, Spain
emoreno@ugr.es

Lina Martínez
Universidad de Málaga, Spain
mlmartinez@uma.es

The Schwarz approximation or BIC criterion is widely used for comparing different statistical models. Since its introduction in mainstream statistical practice, it has been recognized as a very good criterion when the number of parameters of the different competing models is small compared with the sample size. But, it has been criticised when the number of parameters is large, a situation which is frequently encountered in practice as exemplified, for instance, in the important problem of variable selection in regression.

Using tools from the intrinsic methodology, derived by the authors for the problem of variable selection in normal linear regression models, we propose an asymptotic approximation of the Bayes factor for intrinsic priors, which generalizes the BIC, and it is both simple and has very good asymptotic properties as a model selector criterion which, in turn, implies that in situations with very many parameters in the models, it performs much better than the BIC.

Some of its properties and the relation with the BIC are presented.

External Bayesian Analysis for Computer Simulators

Michael Goldstein
Durham University, UK
Michael.Goldstein@durham.ac.uk

Computer simulators o?er a powerful approach for studying complex physical systems. We consider their use in current practice and the role of external uncertainty in bridging the gap between the properties of the model and of the system. The interpretation of this uncertainty analysis raises questions about the role and meaning of the Bayesian approach. We summarise some theory which is helpful to clarify and amplify the role of external specifications of uncertainty, and illustrate some of the types of calculation suggested by this approach.

Bayesian Inference for Hybrid Discrete-Continuous Stochastic Kinetic Models

Andrew Golightly
Newcastle University, UK
a.golightly@ncl.ac.uk

Colin S. Gillespie
Newcastle University, UK
c.gillespie@ncl.ac.uk

We consider the problem of efficiently inferring the parameters in gene regulatory networks. Whilst it is possible to work with a discrete stochastic model for inference, computational cost can be prohibitive for networks of realistic size and complexity. By treating the numbers of molecules of biochemical species as continuous, a diffusion approximation can be used, and whilst this approach has been shown to work well for some networks, ignoring discreteness of low copy number species is unsatisfactory. Here, we consider a hybrid inference method, treating low copy number species as discrete and the remaining species numbers as continuous. The methodology uses a hybrid simulation scheme inside a recently proposed particle marginal Metropolis-Hastings (PMMH) scheme. We apply the scheme to a simple application and compare the output with a scheme for performing inference for the underlying discrete stochastic model.

Fast Bayesian Detection of Disease Clusters Using Generalized Linear Mixed Models

Virgilio Gómez-Rubio
Universidad de Castilla-La Mancha, Spain
Virgilio.Gomez@uclm.es

The detection of clusters of disease has usually been approached by means of ad-hoc tests to look for putative clusters and assess their significancy. Some recent works have approached this problem by using Generalised Linear Models. The advantage of these models is that they provide a flexible way of including disease clusters and more complex spatial and temporal trends as covariates.

We have explored the use of some Bayesian models to find clusters in space and time and have compared our results to the spatial scan statistic, which is widely used. In addition, we have used a model selection approach using the DIC in order to select the most significant disease cluster.

Given that fitting Bayesian models using MCMC can be very time consuming, we have used INLA to provide a computationally feasible way of exploring many different models. Finally, we have explored the use of other model selection criteria such as the Bayes Factor.

Approaching p-values by Means of Least Favorable Answers in the Multivariate Two-Sided Testing Problem

M. Ángel Gómez-Villegas and Beatriz González
Universidad Compluense de Madrid, Spain
ma.gv@mat.ucm.es beatrizg@mat.ucm.es

For iid observations from a multivariate distribution we revisit the issue of the apparent irreconcilability of frequentist and Bayesian answers for testing a point null. Theoretical results in Gómez-Villegas and González-Pérez (2008) are used. For each level of significance α, a value of the prior mass π_0 can be chosen, regardless of the sample size, to design least favorable answers with a similar behavior as the p-value. Furthermore, we illustrate how the p-value can be reconstructed by means of a system of such least favorable answers and well approximated in the crucial range $[0.01, 0.1]$. Asymptotic properties of the approximation are also studied.

Exact Simulation and Bayesian Inference for Jump-Diffusion Processes

Flávio B. Gonçalves and Gareth O. Roberts
University of Warwick, UK
F.B.Goncalves@warwick.ac.uk Gareth.O.Roberts@warwick.ac.uk

The last 10 years have seen a large increase in statistical methodology for diffusions, and computationally intensive Bayesian methods using data augmentation have been particulary prominent. This activity has been fuelled by existing and emerging applications in economics, biology, genetics, chemistry, physics and engineering. However diffusions have continuous sample paths so may natural continuous time phenomena require more general classes of models.

Jump-diffusions have considerable appeal as flexible families of stochastic models. Bayesian inference for jump-diffusion models motivates new methodological challenges, in particular requires the construction of novel simulation schemes for use within data augmentation algorithms and within discretely observed data.

In this work, we propose an algorithm for the simulation of conditional jump-diffusions and, based on that, propose a Bayesian methodology to perform likelihood-based inference for discretely observed jump-diffusions. The algorithm is exact in the sense that it involves no discretisation error and relies on a technique called *Retrospective Rejection Sampling*.

An extension of standard jump-diffusion models is also proposed allowing the jump rate itself to be stochastic. Such rate is assumed to evolve according to a diffusion process which may also depend on unknown parameters. We show that the algorithm proposed here can be extended to make inference in this new class of models.

The efficiency of our method is demonstrated with simulated data and a real example illustrates its applicability in financial economics. We analyse the exchange rate between the Brazilian currency Real and the US Dollar. This example also rises important and interesting issues regarding inference for such kind of data.

Analysis of a Prior Distribution of Resistance Compression of a Alkali-Activated Concrete

Marisol Gordillo
Universidad del Valle, Colombia
mgsuarez44@yahoo.com

J. Olaya, S. Delvasto and R.Mejía de Gutiérrez
Universidad del Valle, Colombia

This paper present the analysis of information that have been used to identify and analyze the function of a priori probability based on existing information on the alkali-activated concretes (CAA). From this perspective, the beta distribution is proposed from the results of the elicitation process in order to determine the parameters of a model analysis of variance. The application of the proposed approach is illustrated with an example that models the molar relation between Silicon and Aluminum (Si/Al) to evaluate the mechanical properties of an alkali-activated concrete. We propose this methodology because the data do not provide conclusively information due to the small sample size. The characteristic has been studied in various research (Spiegelhalter *et al.* 1999, Isabel Verdinelli, 2000, Richard J . Cruz *et al.* 2008). We propose an elicitation process to find the distribution of the variable resistance to compression. We will use the results of this distributions to find the posterior distribution. We will publish these result in a future article

A Mixture of Experts Latent Position Cluster Model

Isobel Claire Gormley and Thomas Brendan Murphy
University College Dublin, Ireland
claire.gormley@ucd.ie brendan.murphy@ucd.ie

Social network data represent the interactions between a group of social actors. Interactions between colleagues and friendship networks are typical examples of such data. The latent space model for social network data locates each actor in a network in a latent (social) space and models the probability of an interaction between two actors as a function of their locations. The latent position cluster model extends the latent space model to deal with network data in which clusters of actors exist – actor locations are drawn from a finite mixture model, each component of which represents a cluster of actors. A mixture of experts model builds on the structure of a mixture model by taking account of both observations and associated covariates when modeling a heterogeneous population. Herein, a mixture of experts extension of the latent position cluster model is developed. The mixture of experts framework allows covariates to enter the latent position cluster model in a number of ways, yielding different model interpretations. Estimates of the model parameters are derived in a Bayesian framework using a Markov Chain Monte Carlo algorithm. The algorithm is generally computationally expensive – surrogate proposal distributions which shadow the target distributions are derived, reducing the computational burden. The methodology is demonstrated through an illustrative example detailing relations between a group of lawyers in the USA.

Foundations of Probability and Quantum Theory

Philip Goyal
Perimeter Institute, Canada
pgoyal@perimeterinstitute.ca

Kevin H. Knuth
University at Albany, USA
drknuth@gmail.com

John Skilling
Max. Entropy Data Cons., Ireland
skilling@eircom.net

Careful attention to foundations gives a unified derivation of measure, probability and (as a remarkable bonus) quantum theory (see Goyal, Knuth and Skilling, *Phys. Rev.A*, 2010). In inference, composite hypotheses are *commutative* and *associative*: "θ_1 or θ_2" = "θ_2 or θ_1" and "θ_1 or (θ_2 or θ_3)" = "(θ_1 or θ_2) or θ_3". These symmetries *by themselves* require any scalar representation to be an additive *measure*: $m(\theta_1 \text{ or } \theta_2) = m(\theta_1) + m(\theta_2)$ (or a 1:1 encryption of linearity). The only freedom is rescaling $m' = \gamma m$ by a constant.

For *probability*, we seek $\Pr(\theta \mid \Theta)$ for θ in context Θ. The only freedom is rescaling of measure on θ, so $\Pr(\theta \mid \Theta) = \gamma(\Theta)m(\theta)$. Concatenation of inclusion "$A \in B$" and "$B \in C$" $\implies$ "$A \in C$" forces $\Pr(\theta \mid \Theta) = m(\theta)/m(\Theta)$. Bayes follows (with context understood) as $\Pr(\theta_1 \text{ or } \theta_2) = \Pr(\theta_1) + \Pr(\theta_2)$, $\Pr(\theta_1 \text{ and } \theta_2) = \Pr(\theta_1 \mid \theta_2)\ \Pr(\theta_2)$. It's simple!

In the quantum world, a possibly-small system can only be observed through interaction with a device (another system). *Two* numbers enter an interaction, whose readout is just *one* probability. A measurement is represented by a *pair* $\mathbf{x} = (x_1, x_2)$ instead of a single scalar, and there's no way of bootstrapping full knowledge. Measurements can be "in parallel" (as when beams interfere), where commutativity and associativity demand $\mathbf{x}(\theta_1 \text{ parallel } \theta_2) = \mathbf{x}(\theta_1) + \mathbf{x}(\theta_2)$ where $\mathbf{u} + \mathbf{v} = \{\{u_1 + v_1\}, \{u_2 + v_2\}\}$. Measurements can be concatenated "in series", where associativity demands $\mathbf{x}(\theta_1 \text{ series } \theta_2) = \mathbf{x}(\theta_1) \times \mathbf{x}(\theta_2)$, where $\mathbf{u} \times \mathbf{v} = \{\{u_1v_1 - u_2v_2\}, \{u_1v_2 + u_2v_1\}\}$. Thus pairs automatically behave as complex numbers.

These compelling symmetry arguments give a unified foundation of inference and physics. It's simple!

Optimization Under Unknown Constraints

Robert B. Gramacy
University of Cambridge, UK
bobby@statslab.cam.ac.uk

Herbert K. H. Lee
University of California, Santa Cruz, USA
herbie@ams.ucsc.edu

Optimization of complex functions, such as the output of computer simula- tors, is a diffcult task that has received much attention in the literature. A less studied problem is that of optimization under unknown constraints, *i.e.*, when the

simulator must be invoked both to determine the typical real-valued response and to determine if a constraint has been violated, either for physical or policy reasons. We develop a statistical approach based on Gaussian processes and Bayesian learning to both approximate the unknown function and estimate the probability of meeting the constraints. A new integrated improvement criterion is proposed to recognize that responses from inputs that violate the constraint may still be informative about the function, and thus could potentially be useful in the optimization. The new criterion is illustrated on synthetic data, and on a motivating optimization problem from health care policy.

Bayesian Inference for Categorical Survey Data

Erofili Grapsa and Jonathan Forster
University of Southampton, UK
eg2x07@soton.ac.uk J.J.Forster@soton.ac.uk

Bayesian inference for surveys and finite population quantities has been developed in a way that hierarchical Normal models can incorporate design features and work efficiently for continuous variables of interest. We are interested in categorical variables with more than two categories and we establish a methodology suitable for them. Different sampling designs are analysed and different models are suggested. Moreover, we extend the existing methods to account for less available information about the survey design. Although this kind of information is required for the sampling procedure, it is usually not provided in public use datasets due to confidentiallity or other issues. Thus, we are interested in cases where design information is available only for the sampled units. In particular, when design variables such as the size variable are not given for the non sampled units. Our approach is to model the unknown parts and try to incorporate the design features in order to make inference for finite population quantities like overall means or totals.

A Stochastic Frontier Model for Discrete Ordinal Outcomes: A Health Production Function

William Griffiths
The University of Melbourne, Australia
wegrig@unimelb.edu.au

Xiaohui Zhang and Xueyan Zhao
Monash University, Australia
{xiaohui.zhang, xueyan.zhao}@buseco.monash.edu.au

The traditional stochastic frontier model that is commonly used in the econometrics literature for measuring inefficiency in models with a continuous dependent variable is extended to accommodate output measured as a discrete ordinal outcome variable. Conditional on the inefficiency error, the assumptions of the ordered probit model are adopted for the log of output when it is observed only

as an ordered categorical variable. Bayesian estimation utilizing a Gibbs sampler with data augmentation is applied to a convenient re-parameterisation of the model. The model and estimation techniques are applied to a health production function using panel data from an Australian longitudinal survey. Demographic and socioeconomic characteristics are specified as inputs to health production, whereas production efficiency is made dependent on lifestyle factors, such as exercise level, alcohol consumption, smoking habits, and social networks. Posterior summary statistics are obtained for selected health status probabilities and efficiencies, and for the marginal effects of individual characteristics and lifestyle factors on health status probabilities and efficiencies.

Using Historical Placebo Information for the Design and Analysis of Clinical Trials in Multiple Sclerosis

Sandro Gsteiger
Novartis Pharma AG, Basel, Switzerland
sandro.gsteiger@novartis.com

B. Neuenschwander F. Mercier and H. Schmidli
Novartis Pharma AG, Basel, Switzerland

When planning a clinical trial, there is often relevant data on placebo treated patients available from previous studies [1, 2]. This information can be used to reduce the number of patients in the placebo group in the new trial and leads, therefore, to an ethically more satisfactory approach, since less patients in the new study receive an ineffective placebo treatment. In [3], a Bayesian meta-analytic-predictive framework for approximate normal endpoints was presented to derive a prior from historical data, which can be used in the analysis of a new trial.

We apply the meta-analytic-predictive approach to trials in multiple sclerosis, a chronic neurological disease associated with irreversible progression of physical disability. Endpoints are lesion counts based on magnetic resonance imaging, commonly described with the negative-binomial distribution to account for over-dispersion [4]. Study specific means are assumed exchangeable and special assumptions must be made to link the study specific overdispersion parameters

First, we investigate the case where individual level data from several in-house trials are available. A fully Bayesian model can be built and the available historical information is quantified using the predictive distribution of the mean counts in the planned trial given the historical data. Second, we consider the use of literature data. Here, only summary information on observed mean counts is given (but not on over-dispersion). The model for the individual observations must now be replaced by the sampling model for the reported parameter estimates. Finally, we study the combination of in-house trial and literature data.

Posterior Simulation in Countable Mixture Models for Large Datasets

Subha Guha
University of Missouri-Columbia, USA
GuhaSu@missouri.edu

Mixture models, or convex combinations of a countable number of probability distributions, offer an elegant framework for inference when the population of interest can be subdivided into latent clusters having random characteristics that are heterogeneous between, but homogeneous within, the clusters. Traditionally, the different kinds of mixture models have been motivated and analyzed from very different perspectives, and their common characteristics have not been fully appreciated. The inferential techniques developed for these models usually necessitate heavy computational burdens that make them difficult, if not impossible, to apply to the massive data sets increasingly encountered in real world studies.

This talk introduces a flexible class of models called generalized Pólya urn (GPU) processes. Many common mixture models, such as finite mixtures, hidden Markov models and Dirichlet processes, are obtained as special cases of GPU processes. Other important special cases include finite dimensional Dirichlet priors, infinite hidden Markov models, analysis of densities models, nested Chinese restaurant processes, hierarchical DP models, nonparametric density models, spatial Dirichlet processes, weighted mixtures of DP priors, and nested Dirichlet processes.

An investigation of the theoretical properties of GPU processes offers new insight into asymptotics that form the basis of cost-effective MCMC strategies for large datasets. These MCMC techniques have the advantage that they provide inferences from the posterior of interest, rather than an approximation, and are applicable to different mixture models. The versatility and impressive gains of the methodology are demonstrated by simulation studies and by a semiparametric Bayesian analysis of high-resolution comparative genomic hybridization data on lung cancer.

A Random-Effects Model for Survival analysis After Acute Myocardial Infarction

Alessandra Guglielmi
Politecnico di Milano, Italy
alessandra.guglielmi@polimi.it

Francesca Ieva Anna Maria Paganoni
Politecnico di Milano, Italy
francesca.ieva@fastwebnet.it anna.paganoni@polimi.it

Fabrizio Ruggeri
CNR-IMATI, Italy
fabrizio@mi.imati.cnr.it

The analysis of variations is an important area of interest in health services and outcomes research. It usually considers multilevel clustered data, with the aim of

identifying and quantifying variability across units (health care providers), and of exploring the links between processes (the hospital patterns) and outcomes (the patient mortality). In this paper we present a case-study, considering a Bayesian mixed-effects model for the survival probabilities of patients who suffered from Acute Myocardial Infarction with ST-segment Elevation (STEMI). We used clinical registers concerning a net of 23 hospitals in the Milan urban area, which was created in 2001 by Lombardia Region, in order to collect process indicators for patients with STEMI for the identification and development of new diagnostic, therapeutic and organizational strategies. The data collection MOMI^2 contains information concerning way of admission (by ambulance or not), demographic features (sex, age), severity of infarction, received therapy (thrombolytic or angioplasty), Symptom onset times, in-hospital times (first ECG times, Door to Balloon times), hospital organization (admission during On/Off hours) and clinical outcome (in-hospital survival). We fitted a generalized linear mixed model to predict the binary survival outcomes by means of relevant covariates, taking into account overdispersion induced by the grouping factor (the hospital each patient was admitted to). The dataset consists of 240 patients who were admitted into 17 hospitals after STEMI, but we used different MOMI^2 collections (of size 359 in total) to set hyperparameters informatively. A MCMC algorithm is needed to compute posterior distributions of parameters and predictive distributions of outcomes, as well as to use other diagnostic tools for goodness-of-fit analysis. The proposed model has been easily implemented in WinBUGS. Current work concerns a nonparametric prior for the random-effects parameters.

Bayesian Risk Management

Armin Haas
Potsdam-Institut für Klimafolgenforschung, Germany
armin.haas@pik-potsdam.de

Markus Fucik and Rainald Ötsch
Universität Potsdam, Germany
{markus.fucik, rainald.oetsch}@uni-potsdam.de

We suggest Bayesian Risk Management (BRM) as a universal concept for managing an unknowable future. BRM builds on tools and methods from Bayesian decision theory and statistics. A Bayesian risk manager is operational when she has a basic understanding of a decision situation, even if she lacks the data necessary for conventional risk management. We define two strands of BRM, the hardcore and the softcore version. Hardcore BRM uses Bayesian learning in the strict, mathematical sense, for updating probabilities. It is the method of choice when data are available that can be fed into Bayesian updating algorithms. Softcore BRM applies when data are not sufficient for hardcore BRM. It mobilises the judgment of experts and decision makers, e.g. via expert elicitations, stakeholder dialogues, or online (prediction) markets. We complement hardcore and softcore BRM by Bayesian Due Dilligence (BDD). BDD is a reflexive approach for scrutinising both a prioris and learning processes of a decision maker in light of the general setting he operates in.

Sparse Partial Factor Regression

P. Richard Hahn
Duke University, USA
hahn@stat.duke.edu

Carlos Carvalho
University of Chicago, USA
carlos.carvalho@chicagobooth.edu

Sayan Mukherjee
Duke University, USA
sayan@stat.duke.edu

We introduce a novel Bayesian linear regression model: sparse partial factor regression. †Building on earlier work on Bayesian factor regression models, we introduce modifications that cleanly address a long-recognized difficulty of factor analytic models – that, as Cox pointed out in 1968, there is no logical reason why the dependent variable should not be closely tied to the least important principal component." †By explicitly addressing this point, our model acquires three key advantages over common alternatives: improved out-of-sample prediction compared to ridge regression, g-prior regression, and standard factor models, the ability to naturally handle variable selection in the presence of highly correlated predictors, and inherent robustness to the choice of the number of factors. †We show that partial factor regression can be viewed as special cases of well known models, specifically a†regression model with a prior "centered" at a factor model, a†hierarchical generalization of Zellner's g-prior, and an errors-in-variables model which permits the "errors" to themselves be included in the regression. Our model clarifies, in the venerable setting of Bayesian MVN models, the intrinsic difficulty of estimating a sufficient subspace, showing that the main obstacle is an onerous nuisance parameter.

Bayesian Interactive Clustering for High Dimensional Data

Chao Han Scotland Leman and Leanna House
Virginia Tech, USA
{chaohan, leman, lhouse}@vt.edu

In light of advancements made in data collection techniques over the past two decades, data mining has become common practice to summarize large, high dimensional datasets, in hopes of discovering noteworthy data structures. However, one concern is that most data mining approaches rely upon strict criteria that may mask important data features. Bayesian Visual Analytics (BaVA) addresses this concern and enables experts to interact with the data and the feature discovery tools. In this paper, we use BaVA idea to enhance common high dimensional clustering techniques. For example, Mixture Probabilistic PCA (MPPCA), as developed by Tipping and Bishop (1999), has the potential to discover clusters while reducing the dimension of datasets. We extend MPPCA methods and create corresponding malleable, low dimensional visualizations of the data. Experts may include feedback into the cluster exploration a posteriori by manipulating the data points in the display directly. In turn, feature hidden by standard data mining methods may appear during the expert-guided exploration. We demonstrate the performance of our approach using both synthetic and real dataset and compare our algorithms with existing ones.

Penalized Regression via Orthant Normal Priors

Chris Hans
Ohio State University, USA
hans@stat.osu.edu

Motivated by penalized optimization approaches to variable selection and parameter estimation, this paper introduces a new class of prior distributions — the orthant normal distribution — for the regression coefficients in a Bayesian regression model. Parameter estimates based on penalized optimization are often interpreted as the mode of a Bayesian posterior distribution. We show that the orthant normal distribution is the prior that gives rise to the elastic net estimate and, in a limiting case, the lasso. By providing a complete characterization of this prior, we allow for model-based inference that moves beyond exclusive use of the posterior mode, including coherent Bayesian prediction and formal Bayesian model comparison. In contrast to penalized optimization procedures (where the penalty parameter is often selected via a potentially unstable cross validation), the Bayesian approach allows for uncertainty about these parameters to be included in the model, or, alternatively, allows the parameters to be selected via the method of maximum marginal likelihood. We show that the orthant normal distribution has a scale-mixture of normals representation, providing additional insight into the particular form of shrinkage employed by the elastic net. Posterior inference is achieved via MCMC.

This model-based approach to elastic net regression has the advantage that the basic model can be extended to accommodate more complex regression settings. Models can be built that include random effects to capture various covariance structures while at the same time inducing elastic-net-like shrinkage on the regression coefficients. We discuss approaches for incorporating prior information about dependence structure in the covariates that resemble Zellner's g-prior but that allow for lasso-like shrinkage.

Inferring Likelihoods and Climate System Characteristics from Climate Models and Multiple Tracers

Murali Haran
Pennsylvania State University, USA
mharan@stat.psu.edu

K.S. Bhat R. Tonkonojenkov and K. Keller
Pennsylvania State University, USA
{kgb130, romantonk, kkeller}@psu.edu

An important potential outcome of anthropogenic climate change is a possible collapse of the Atlantic meridional overturning circulation (AMOC). Assessing the risk of an AMOC collapse is of considerable interest since it may result in major temperature and precipitation changes and a shift in terrestrial ecosystems. One key source of uncertainty in AMOC predictions is uncertainty about background ocean vertical diffusivity (K_v), a key model parameter. K_v cannot be directly

observed but can be inferred by combining climate model output with observations on the oceans (so called "tracers"). In this work, we combine information from multiple tracers, each observed on a spatial grid. Our two stage approach emulates the computationally expensive climate model using a flexible hierarchical model to connect the tracers. We then infer K_v using our emulator and the observations via a Bayesian approach, accounting for observation error and model discrepancy. We utilize kernel mixing and matrix identities in our Gaussian process model to considerably reduce the computational burdens imposed by the large data sets. We find that our approach is flexible, reduces identifiability issues, and enables inference about K_v based on large data sets. We use the resulting inference about K_v to improve probabilistic projections of the AMOC.

Inferring Meta-Covariates in Classification Via Gibbs Sampling

KEITH J. HARRIS and MARK GIROLAMI
University of Glasgow, UK
keithh@dcs.gla.ac.uk girolami@dcs.gla.ac.uk

In this paper, we present a recently developed procedure that potentially improves the interpretability and classification of high dimensional data sets through the statistical coupling of a probit regression model with model based clustering. High dimensional data sets typically consist of several thousand covariates and a much smaller number of samples. Analyzing this data is statistically challenging, as the covariates are highly correlated, which results in unstable parameter estimates and inaccurate prediction. To alleviate this problem, we develop a statistical model that uses a small number of meta-covariates inferred from the data, rather than all the original covariates, to classify samples. The advantage of this approach over using a sparse classification model is that we can extract a much larger subset of covariates with essential predictive power and partition this subset into groups, within which the covariates are similar.

An overview of our procedure is as follows. By employing sufficient statistics from the allocations obtained from a Gaussian mixture model, we define "meta-covariates" and use them in a probit regression model, thereby attaining concise interpretation and accuracy. We accomplish this simultaneous inference task by an efficient Gibbs sampler initialized by an EM algorithm, where the joint distribution defined by our model rewards good performance at both classification and clustering. Similar ideas, from a non-Bayesian uncoupled two-step perspective have already been proposed. With our simultaneous procedure, the components are formed considering the correlation of the predictors with the response in addition to the correlations among the predictors. We will illustrate the wide applicability of our proposed methodology by demonstrating its performance on two different types of problems, identifying differentially expressed genes from DNA microarray data and classifying spatial patterns of brain activity from functional Magnetic Resonance Imaging data.

Fast Joint Posterior Modelling through Marginal Posterior Mixtures

JOHN HASLETT
Trinity College Dublin, Ireland
john.haslett@tcd.ie

ANDREW C. PARNELL
University College Dublin, Ireland
andrew.parnell@ucd.ie

We discuss the issue of creating a joint posterior distribution for a set of parameters when only marginal posteriors are available (or are reasonable to compute). More specifically, for data $\boldsymbol{x}$ and parameters $\boldsymbol{\theta}$ in $\mathbb{R}^n$, we require $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ from the marginal posterior $\pi(\theta_i|x_i)$. Through a simple adjustment of Bayes' theorem we can use $\pi(\theta_i|x_i)$ to inform the joint posterior, provided $\pi(\theta_i)$ and $\pi(\boldsymbol{\theta})$ (the marginal and joint priors, respectively) are, in some sense, compatible.

The technique can be further enhanced by treating $\pi(\theta_i|x_i)$ as a mixture of distributions conjugate to the joint prior. In simple cases where this joint prior is Gaussian, it is trivial and extremely quick to approximate any marginal posterior distribution as such a mixture. The computational savings to be had here are large, and allow for alternative techniques, such as importance sampling over indicator variables, to be used as alternatives to MCMC.

We apply this technique to two problems in palaeoclimatology (both described in Haslett et al 2006). The first involves long-tailed random walk smoothing of temporal climate histories ($c(t)$) created from pollen sediment cores where pollen is sampled at n layers $y_i,\ i = 1, \ldots, n$. The marginal posteriors $\pi(c_i|y_i)$ are easily obtained by other means, whereas the random walk gives flat marginal prior distributions $\pi(c_i)$. We obtain the joint prior $\pi(\boldsymbol{c}|\boldsymbol{y})$ in a two-stage process without resorting to more burdensome computational methods. The second problem involves spatial forward modelling of pollen changes given modern climate data (also known as response surface modelling; Huntley et al 1993). Here, the marginal posteriors are Gaussian surfaces with few hyper-parameters; they are relatively quick to create. The joint posterior surface then becomes a mixture of Gaussian processes. Again, the two-stage process dramatically decreases the computational burden, and allows for parallelisation. The models we propose have much in common with Rue et al (2009) and Holmstrom and Erasto (2002).

The technique seems widely applicable across the field of statistical modelling. We explore some of the extensions which may allow for higher dimensional models or more complex prior distributions.

Spatial Regression Using Kernel Averaged Predictors

MATTHEW J. HEATON and ALAN E. GELFAND
Duke University, USA
matt@stat.duke.edu alan@stat.duke.edu

Traditional spatial linear regression models assume the mean of the response is a linear combination of predictors measured at the same location as the response. In spatial applications, however, it seems plausible that neighboring predictors can also inform about the response. This article proposes using unobserved kernel averaged predictors in such regres- sions. The kernels are parametric introducing

additional parameters that are estimated with the data. Properties and challenges of using kernel averaged predictors within a regression model are detailed in the simple case of a univariate response and a single predictor. Additionally, extensions to multiple predictors and generalized linear models are discussed. The methods are demonstrated using a data set of dew duration and shrub density.

Bayesian Calibration of a Stochastic Computer Model of Mitochondrial DNA Population Dynamics Using Multiple Data Sources

Daniel A. Henderson
Newcastle University, UK
d.a.henderson@ncl.ac.uk

Richard J. Boys and Darren J. Wilkinson
Newcastle University, UK
richard.boys@ncl.ac.uk d.j.wilkinson@ncl.ac.uk

We consider the problem of parameter estimation for a stochastic kinetic computer model of mitochondrial DNA population dynamics. The model is an attempt to describe the hypothesised link between deletion accumulation and neuronal loss in the substantia nigra region of the human brain. As with many applications in the biological sciences, the data available to calibrate the model come from different sources and relate to two different aspects of the model; we have several independent sets of experimental data on both deletion mutation accumulation and neuron survival. Furthermore these data appear to provide somewhat conflicting information about the model parameters. We describe a modelling framework which allows us to synthesize this conflicting information and arrive at a consensus inference. In particular, random effects are incorporated into the model in order to account for between-individual heterogeneity which may be the source of the apparent conflict.

Bayes Factors for the Transmission Disequilibrium Test (TDT) for Trio Genome-wide Association Data

Eleisa Heron
Trinity College Dublin, Ireland
eaheron@tcd.ie

R. Anney L. Gallagher, and M. Gill
Trinity College Dublin, Ireland

Trio family data, consisting of genotypes for two parents and an affected child, for a vast number of genetic markers (single nucleotide polymorphisms (SNPs)) can sometimes be collected for genome-wide association (GWA) data analyses. These analyses aim to detect an association between a disease locus (or

a locus in linkage disequilibrium with the disease locus) and the particular disease phenotype under study. GWAS analyses for trio data are usually carried out using the transmission disequilibrium test (TDT). The TDT examines transmissions and non-transmissions of alleles from heterozygous parents to offspring, aiming to detect over- or under-transmission of alleles. We propose a Bayesian approach to examine associations for trio data using Bayes factors to compare models, ultimately resulting in posterior odds for a model of association for each individual marker.

Using simulated data, we show how the Bayes factor approach to the TDT is not affected by such factors as, for example, the minor allele frequency of the marker, unlike the conventional TDT p-values. Multiple testing issues that need to be considered for the p-value approach do not arise in the Bayesian context. Instead, meaningful prior information in the form of prior odds of association can be incorporated and we examine the influence of various priors. Trio data is often collected for early onset psychiatric diseases, such as attention deficit hyperactivity disorder (ADHD) and autism. We apply the Bayes factor approach to an ADHD trio data set and show how meaningful prior information can be incorporated from additional available case-control data. These results are compared with the standard TDT p-value approach.

Learning Structure in Directed Acyclic Graphs with Latent Variables

Ricardo Henao
Danmarks Tekniske Universitet
rhenao@binf.ku.dk

Ole Winther
Københavns Universitet, Denmark
owi@imm.dtu.dk

Recently, we proposed a method to perform inference on identifiable sparse factor models and directed acyclic graphs (DAGs) within the same framework exploiting their equivalence under certain conditions (Henao and Winther, *Adv. Neural Inform. Proc. Syst.* **22**, 2009). These two type models target different goals; the factor model infers the connectivity between a set of observed and latent variables while the DAG only allows to infer directed connections between the observed variables at hand. Nevertheless, a very well known issue in biology is that all the variables of interest cannot be measured or they are simply unknown. For this purpose, in this work we extend the DAG model by allowing also connections from a set of latent ones. This model is thus a mixed factor model and DAG in which heavy-tailed distributions for the latent factors ensure identifiability and parsimonious priors for the connectivity/mixing matrices reflect that we are interested on sparsely connected networks. Inference is done using MCMC, in particular we first learn a factor model and a set of variable orderings compatible with the DAG assumption, second we learn DAG models with latent variables for each of the ordering candidates in turn and finally, we use predictive densities to decide which model explains better the data among all candidates including the factor model. We evaluate our new model using both artificial and real world data taken from publicly available sources.

A Bayesian Model of Choice with Heterogeneity

R. Carter Hill
Louisiana State University, USA
eohill@lsu.edu

Ivan Jeliazkov
Univ. of California–Irvine, USA
ivan@uci.edu

Danny Weathers
Louisiana State University, USA
weathers@lsu.edu

To account for heterogeneity across households in a brand choice model we consider a hierarchical mixed effect setup which allows the random utility function to be separated into fixed and random components. We assume that household $i = 1, \ldots, N$, chooses among J brand alternatives on occasion $t = 1, \ldots, T_i$, where occasions may be irregularly spaced. Data on brand prices, advertising features and household demographics are available on each purchase occasion. The probability that household i chooses alternative j on occasion t is $\Pr(y_{it} = j) = F\left(\alpha_{ij} + x'_{ijt}\beta_i\right)$, where F is a link function, and α_{ij} is an alternative-specific constant that may be influenced by observed and unobserved household characteristics. Specifically, $\alpha_{ij} = \alpha_j + z'_i\delta_j + u_{ij}$, where α_j and δ_j are alternative-specific parameters, z_i is a vector of household characteristics that vary across household, but not across alternatives, and u_{ij} is a random heterogeneity term with mean 0 and variance $\sigma^2_{\alpha j}$.

The K variables x_{ijt} vary across alternative, individual and occasion. The coefficients are modeled as $\beta_{ik} = \bar{\beta}_k + w'_i\gamma + \nu_{ik}$, where $\bar{\beta}_k$ and γ are parameters, and w_i is a vector of household characteristics. The term ν_{ik} has mean 0 and variance η^2_{ik} which depends on a vector of household characteristics r_i (not including a constant term) and parameters ω_k and σ^2_k as

$$\eta^2_{ik} = \begin{cases} 0 & \text{for } k \in S \\ \sigma^2_k \exp(r'_i\omega_k) & \text{for } k \in \bar{S} \end{cases} .$$

We present tuned MCMC algorithms for posterior simulation with several data augmentation schemes that enable the analysis of models with Gaussian, Student's t and logistic link functions. We discuss the computation of marginal likelihoods and Bayes factors in order to address the uncertainty due to variable selection, distributional modeling, and the presence and extent of heterogeneity.

Joint Bayesian Inference for D-Vine Based Copula Models with GARCH Margins

Mathias Hofmann
Technische Universitaet Muenchen, Germany
hofmann@ma.tum.de

We consider a multivariate time-series model, where the margins follow univariate GARCH(1,1) models and the dependence is introduced by a pair copula construction (Kurowicka and Cooke (2006), Aas et. al (2009)), specified to be a D-vine.

We perform a joint estimation of the marginal and dependency parameters in a Bayesian context using MCMC methods. In a simulation study we consider the influence of model misspecification and the error introduced by the widely used 2-step estimation (e.g. Min et. al. (2009)), by means of the posterior predictive distribution of the VAR of an equally weighted portfolio. As a real data application, we estimate the model for Asian stock price index data. This is joint work with Claudia Czado.

Bayesian Animal Model Using Integrated Nested Laplace Approximations: a Wild House Sparrow Population Case Study

Anna-Marie Holand
Norges Teknisk-Naturvitenskapelige Universitet, Norway
anna.holand@bio.ntnu.no

Ingelin Steinsland and Henrik Jensen
Norges Teknisk-Naturvitenskapelige Universitet, Norway
ingelins@math.ntnu.no Henrik.Jensen@bio.ntnu.no

To obtain precise estimates of the additive genetic variances of various phenotypic characters is of fundamental importance to understand evolutionary change in animal and plant breeding as well as in natural populations. To estimate the additive genetic variance (and thus the heritability) of traits, biologist and breeders often use a general linear mixed model (additive genetic model) called an animal model. Previous Bayesian animal model approaches have used Markov chain Monte Carlo (MCMC). However, MCMC can be computationally demanding and hence time consuming. By using the fact that an animal model is a Gaussian Markov Random Field (GMRF), we can utilize the computational benefits of GMRFs. Recently, a new non-sampling based approximation method for making inference for GMRF models, INLA (integrated nested Laplace approximations), has been developed by Rue, Martino, and Chopin (2009). Using integrated nested Laplace approximation, it is possible to compute accurate approximations for the posterior marginals of interest. Here we use this approximation method in a Bayesian animal model to estimate the additive genetic variances and heritabilities of different fitness-related traits in natural house sparrow populations in Norway.

A Bayesian Latent Growth Model of Language Acquisition in Deaf Children

Christopher Holloman
The Ohio State University, USA
holloman@stat.osu.edu

S. Nittrouer C.Chapman, and B.Welling
The Ohio State University, USA

Language is a complex behavior acquired over a protracted period. Historically, language acquisition has been studied using tests of different aspects of language

such as vocabulary size, spoken language comprehension, and expressive ability. In contrast, the Early Development of Children with Hearing Loss study sought to measure language usage in a natural conversational environment. 205 children (118 with hearing loss) were tested with several language measures every six months from 12 to 48 months of age. We developed a Bayesian latent growth model that incorporates measures of different aspects of language, each with different functional forms over the course of language development, to characterize latent language ability. In addition, the model incorporated both static (e.g., gender and hearing loss status) and time-varying (e.g., configuration of prostheses and level of parental interaction) covariates allowing us to investigate which strategies are most effective for allowing deaf children to acquire language.

Bayesian Computation on Graphics Cards

Chris Holmes
University of Oxford, UK
cholmes@stats.ox.ac.uk

Arnaud Doucet
Univ. of British Columbia, Canada
arnaud@cs.ubc.ca

A. Lee M. B. Giles and C. Yau
University of Oxford, UK

Advances in computational methods and computing power have been instrumental in the uptake and development of Bayesian statistics in the last forty years. Recent trends in desktop computing offer the potential to make substantial further improvements.

We have been working on Monte Carlo methods for Bayesian computation run on graphics cards (GPUs). Graphics cards were originally designed to deliver real-time graphics rendering for computer games and other high-end graphics applications. However, there is an emerging literature on the use of GPUs for scientific computing. The advantages are clear. A typical graphics card has effectively around 250 parallel processors designed for fast arithmetic computation. The cards are cheap, dedicated, low maintenance, with low energy consumption and are able to plug directly into a standard desktop computer or laptop. For certain classes of scientific computing algorithms, GPUs offer the potential speed up of traditional massively parallel cluster-based computing at a fraction of the cost, power and time of uptake and programming effort.

We will review GPU architectures and the class of algorithms which are suited to GPU simulation; as well as those which are not suited. In particular, we discuss SIMD (Single Instruction Multiple Data) processing structures and how they relate to Monte Carlo methods. We have migrated a number of methods onto GPUs including population-based MCMC, sequential Monte Carlo samplers and particle filters with speed ups ranging from 35 to 500 fold over conventional single-threaded computation; reducing weeks worth of computing down to hours. We discuss the algorithmic structures, developments and hurdles which lead to such improvements. We illustrate these methods on a number of challenging examples in Bayesian statistical modelling.

Variable Selection for Bayesian Mixture Models

Chris Holmes and Chris Yau
University of Oxford, UK
cholmes@stats.ox.ac.uk yau@stats.ox.ac.uk

Bayesian mixture modelling provides a powerful tool for cluster analysis as it allows for a full characterisation of the uncertainty in the principle dependence structures in the data. Clustering is often used as an exploratory method to uncover hidden structure in data or to recover suspected structure. We have investigated the potential of variable selection priors to learn about the relative influence of explanatory variables within mixture models. Such models are appropriate when some, or perhaps many, of the recorded variables are *a priori* suspected to be irrelevant to the clustering task. We demonstrate a hierarchical approach that provides accurate indication of irrelevant variables and is able to quantify the relative relevance of those variables useful for clustering. We pay particular attention to efficient MCMC sampling schemes for inference within the model while allowing for an unknown number of mixture components and an unknown number of relevant variables.

Nonparametric Predictive Utility Inference

Brett Houlding
Trinity College Dublin, Ireland
houldinb@tcd.ie

Frank P. A. Coolen
Durham University, UK
rank.coolen@durham.ac.uk

Traditionally subjective expected utility theory reconstructs a DM's preferences over a collection of decisions by taking into account both the probability of decision outcome and the DM's relative preference for obtaining that outcome as measured by its utility value. However, implicit within this theory is the assumption that the DM knows their preferences, meaning they can assign an appropriate utility function for use.

Nevertheless, not for all situations is it appropriate to assume a known preference relation over outcomes. Indeed, it may be necessary for a DM to experience a novel outcome before being able to assign an appropriate utility. Such situations of utility uncertainty motivate Cyert and DeGroot's (1975) adaptive utility theory, where a utility function is only known up to the value of some uncertain parameter. Yet, despite the theory explicitly permitting a DM to remain uncommitted to a presumed known and correct utility function, previous use has required knowledge of a precise prior distribution concerning preferences.

Rather than assuming a precise prior for the uncertain utility parameter, interest here is in the use of Nonparametric Predictive Inference (NPI), which is a low structure technique arising as a result of the Hillís (1968) $A_{(n)}$ assumption. Given an ordered series of observations considered subject to pre-data exchangeability, NPI proceeds by assigning equal mass to the probability that a new observation falls within any of the intervals formed by the known ordered data, leading to the quantification of uncertainty through interval probability and a generalisation of the traditional approach.

Reinforcing Reification with Application to a Rainfall-Runoff Computer Model

Leanna House
Virginia Tech, USA
lhouse@vt.edu

Deterministic computer models or simulators are used regularly to assist researchers in understanding the behavior of complex physical systems when real-world observations are limited. However, simulators are often imperfect representations of physical systems and introduce a layer of uncertainty into model-based inferences that is hard to quantify. To formalize the use of expert judgement in assessing simulator uncertainty, Goldstein and Rougier (2008) propose a method called reification that decomposes the discrepancy between simulator predicitions and reality by an improved, hypothetical computer model known as a "reified" model. One criticism of reification is that validation is, at best, challenging; only expert critiques can validate the subjective judgements used to specify a reified model. For this paper, we created an artificial case study based on a well-used hydrological computer model and real data that allows us to validate reification using both data and expert judgment.

Simulation-Based Optimal Experimental Design with Polynomial Surrogates on Sparse Grids

Xun Huan and Youssef Marzouk
Massachusetts Institute of Technology, USA
xunhuan@mit.edu ymarz@mit.edu

The optimal selection of experimental conditions is essential to maximizing the value of data for inference and prediction, particularly in situations where experiments are time-consuming and expensive to conduct. We propose a general Bayesian framework for optimal experimental design with nonlinear simulation-based models. The formulation accounts for uncertainty in model parameters, experimental conditions, and observables. Straightforward Monte Carlo evaluation of the objective function—which reflects expected information gain (Kullback-Leibler divergence) from prior to posterior—is intractable when the likelihood is computationally intensive. Instead, we introduce polynomial chaos expansions to capture the dependence of observables on model parameters and on design conditions. Under suitable regularity conditions, these expansions converge exponentially fast. Since both the parameter space and the design space can be high-dimensional, dimension-adaptive sparse quadrature is used to construct the polynomials. Stochastic optimization methods are then used to maximize the expected utility.

While this approach is broadly applicable, we demonstrate it on a chemical kinetic system with strong nonlinearities. In particular, we estimate Arrhenius rate parameters in a combustion reaction mechanism from observations of autoignition. Results show multiple order-of-magnitude speedups in both experimental

design and parameter inference. Ongoing work focuses on more efficient stochastic optimization methods for maximizing expected utility in high dimensions and on the construction of sparse polynomial chaos representations adapted to inferential problems.

Using TPA for Bayesian Inference

Mark Huber
Claremont McKenna College, USA
mhuber@cmc.edu

Sarah Schott
Duke University, USA
schott@math.duke.edu

Finding the integrated likelihood of a model given the data requires the integration of a nonnegative function over the parameter space. Classical Monte Carlo methods for numerical integration require a bound or estimate of the variance in order to determine the quality of the output. The method called the product estimator does not require knowledge of the variance in order to produce a result of guaranteed quality, but requires a cooling schedule that must have certain strict properties. Finding a cooling schedule can be di?- cult, and finding an optimal cooling schedule is usually computationally out of reach. TPA is a method that solves this di?culty, creating an optimal cooling schedule automatically as it is run. This method has its own set of requirements; here it is shown how to meet these requirements for problems arising in Bayesian inference. This gives guaranteed accuracy for integrated likelihoods and posterior means of nonnegative parameters.

Semiparametric Bayesian Modeling of Spatio-Survival Data Under Cure Fraction

Sandra Hurtado and Dipak Dey
University of Connecticut, USA
shurtado@stat.uconn.edu dey@stat.uconn.edu

This paper proposes the use of Bayesian spatial modeling with cure rates to model time to event data. In recent times, spatial variability components, cure possibilities, and the changes over time of the survival curve have been of of paramount interest to researchers and public health decision makers. There has been research addressing each of these factors independently, but there is a need for a richer class of models to enable a comprehensive analysis of time to event data. We assume flexible semi-parametric baseline hazard functions with a grid defined by join-point parameters similar to the one used in join-point regression models. The hazard model is set up under the generalized proportional hazard framework (PH) in which we also include a covariance function. We integrate into the analysis the spatial correlation structure, in the form of county cancer

level frailties and the cure rates. Finally, we compare across a broad collection of rather high-dimensional hierarchical models using the log of the pseudo-marginal likelihood (LMPL). We apply our methodology to colon cancer survival times for patients diagnosed in the state of Connecticut between 1973-2004, and with follow up time until 2007. Although motivated by the National Cancer Institute, SEER data base, the proposed work offers useful contributions to general statistical theory and methodology in survival analysis, spatial statistics and cure rates modeling. In summary, this paper develops a Bayesian hierarchical model for capturing spatial heterogeneity and cure rates for right censored time to event data under a semiparametric proportional hazard framework. We obtain the usual posterior estimates, smoothed by counties level maps of spatial frailties and cure rates.

Bayesian Kernel Machines: The Third Way of Going Nonparametric

Ferenc Huszár
University of Cambridge, UK
fh277@mlg.cam.ac.uk

Bayesian nonparametrics – Bayesian inference involving infinite dimensional parameters – has become increasingly popular in the Bayesian statistics and machine learning communities over the past decades. The most direct way of obtaining Bayesian nonparametric models is to define a prior stochastic process over the infinite dimensional parameter explicitly, and derive the posterior process. Well known examples include Pólya-trees, Pitman-Yor and Gaussian process priors. Another technique of developing nonparametric Bayesian methods is through defining an exchangeable process in which the de Finetti measure becomes infinite dimensional. This is often achieved by deriving posterior predictive distributions in a finite parametric model and then taking the limit as the number of parameters tends to infinity. Examples of such a construction are the Chinese restaurant and Indian buffet processes. In this work we explore a third way of defining infinite dimensional Bayesian models: via the *kernel trick*. The kernel trick is a highly popular tool in machine learning. It is used to overcome limitations of a linear model by projecting observations $\boldsymbol{x}_i \in \mathcal{X}$ to a feature space $\mathcal{F}$ using a nonlinear mapping $\phi : \mathcal{X} \mapsto \mathcal{F}$. Then the linear algorithm is applied on the projected feature elements $\Phi_n = \phi(\boldsymbol{x}_n)$ instead of observations themselves. If relevant quantities are computed by combining inner products $\langle \Phi_n, \Phi_m \rangle$, and if the *kernel* $k(\boldsymbol{x}_n, \boldsymbol{x}_m) := \langle \Phi_n, \Phi_m \rangle$ can be evaluated effectively, then it is possible to use an infinite dimensional feature space $\mathcal{F}$, still with finite amount of computation. Recently the kernel trick has been used in conjunction with classical linear statistical models to obtain nonparametric versions of principal component analysis and independence tests. For Bayesians, the kernel trick offers a possibility to define tractable nonparametric models making use of infinite dimensional feature spaces. For example, it is possible to obtain Gaussian process regression by applying the kernel trick in the predictive distribution under a Bayesian linear regression model. Despite its popularity in machine learning, the kernel trick hasn't been applied in Bayesian methods other than linear regression, until now. The primary goal of

this work is to emphasise the importance of this tool for Bayesian nonparametrics. In particular, we consider applying the kernel trick in the Bayesian PCA model for unsupervised and semi-supervised learning problems. Under the PCA model, the observations are assumed to follow Gaussian distribution with unknown mean and covariance. We observe that if appropriate conjugate priors are placed on these parameters, the posterior predictive distribution is analytically tractable, and can be written in a Student-T form where the kernel trick can indeed be applied. In the talk I introduce the general kernel framework, and present how it can be applied to Bayesian models. The general procedure is illustrated on the well known Gaussian process regression and Bayesian PCA models. We will describe properties of the nonparametric objects involved and review connections to relevant related work, in particular to Hilbert space embedding of probability distributions. The practical potential of our methods is illustrated by applying them to real world novelty-detection and classification problems.

Predictive Hypothesis Identification

Marcus Hutter
Australian National University
marcus.hutter@anu.edu.au

There are many desirable properties any hypothesis identification algorithm ideally should satisfy: It should lead to good predictions, be broadly applicable, be analytically and/or computationally tractable, be defined and make sense also for non-i.i.d. and non-stationary data, be reparametrization and representation invariant, work for simple and composite hypotheses, work for classes containing nested and overlapping hypotheses, work in the estimation, testing, and model selection regime, reduce in special cases (approximately) to existing other methods, and others.

Arguably the most prevalent goal is to choose hypotheses (parameters/models) with good *predictive performance*. The poster and corresponding technical report http://arxiv.org/abs/0809.1270 present a unifying approach that addresses the problem head on, and show that the resulting principle nicely satisfies many of the other desirable properties. The developed Predictive Hypothesis Identification (PHI) Principle identifies hypotheses with predictive performance closest to Bayes-optimal prediction. This includes predictive point and interval estimation, simple and composite hypothesis testing, (mixture) model selection, and others as special cases. If prediction is the goal, but full Bayesian prediction not feasible or desirable, PHI allows to *identify* (estimate/test/select) the *hypothesis* (parameter/model/interval) that *predicts* best. What best is can depend on the problem setup: What distance function we use for comparison, how long we use the model compared to how much data we have at hand, and whether we continue to learn or not (batch↔offline).

For concrete instantiations we will recover well-known methods, variations thereof, and new ones. PHI nicely justifies, reconciles, and blends (a reparametrization invariant variation of) MAP, ML, MDL, and moment estimation. One particular feature of PHI is that it can genuinely deal with nested hypotheses.

Nonparametric Bayesian Networks

Katja Ickstadt
Technische Universität Dortmund University, Germany
ickstadt@statistik.tu-dortmund.de

B. Bornkamp M. Grzegorczyk J. Wieczorek
Technische Universität Dortmund University, Germany

M.R. Sheriff H. E. Grecco and E. Zamir
Max-Planck-Institut für Molekulare Physiologie, Dortmund, Germany

A convenient way of modelling complex interactions is by employing graphs or networks which correspond to conditional independence structures in an underlying statistical model. One main class of models in this regard are Bayesian networks, which have the drawback of making parametric assump- tions. Bayesian nonparametric mixture models o?er a possibility to overcome this limitation, but have hardly been used in combination with networks. This manuscript brigdes this gap by introducing nonparametric Bayesian network models. We review (parametric) Bayesian networks, in particular Gaussian Bayesian networks, from a Bayesian perspective as well as nonparametric Bayesian mixture models. Afterwards these two modelling approaches are combined into nonparametric Bayesian networks. The new models are com- pared both to Gaussian Bayesian networks and to mixture models in a simula- tion study, where it turns out that the nonparametric network models perform favorably in non Gaussian situations. The new models are also applied to an example from systems biology.

Autoregressions and Priors About Initial Growth Rates

Marek Jarociński
European Central Bank, Germany
marek.jarocinski@ecb.int

Albert Marcet
London School of Economics and CEPR, UK
a.marcet@lse.ac.uk

We discuss estimation of autoregressive models with a prior about initial growth rates of the modeled time series. This prior allows to specify prior beliefs about the behavior of time series in a natural way and it serves to replace arbitrary assumptions on initial conditions. To implement this prior we develop a technique for translating priors about observables into priors about coefficients. The posterior mean is attractive even from the frequentist point of view: it is often less biased than the OLS estimate and has better frequentist risk than bias corrected estimates. We apply our prior to some empirical studies from the literature and find that, compared with the flat prior, it makes a big difference for the estimated persistence of output responses to monetary policy shocks in a vector autoregression for the United States.

Bayesian Modeling of the Evolutionary Escape Response of HIV

Shane T. Jensen
The Wharton School, University of Pennsylvania, USA
stjensen@wharton.upenn.edu

Alexander Braunstein
The Wharton School, University of Pennsylvania, USA
braunsf@wharton.upenn.edu

Jon D. McAuliffe
The Wharton School, University of Pennsylvania, USA
jon@mcauliffe.com

A characteristic of viruses is their rapid evolution in response to detrimental challenges, leading to escape from therapeutic treatments. Our work is motivated by case studies of two different therapies targeted against the human immunodeficiency virus (HIV). The first study is a conventional drug therapy that attacks HIV virions before they reach human cells. The second study is an antisense gene therapy that is designed to attack the HIV genome after it has infected a human cell. It is important to characterize the HIV evolutionary response to therapy with a sophisticated statistical model that can lead to new therapeutic strategies for HIV.

Our Bayesian approach builds upon the coalescent evolutionary model, which allows for simultaneous modeling of mutation and recombination events. We expect spatial variation in the HIV evolutionary escape response, since specific local genome changes are needed to develop resistance to either of the two therapies under study. We address this issue with a piecewise-constant prior distribution that allows spatial heterogeneity of our mutation and recombination parameters while sharing some global information between different positions in the HIV genome. . Our model explicitly measures a therapy effect for both recombination and mutation through a hierarchical prior structure, linking rates between treatment and control sequences. Our Bayesian evolutionary model is estimated using a Markov chain Monte Carlo (MCMC) algorithm that samples from the posterior distribution of all unobserved parameters. The detection of biologically relevant and plausible signals in both therapy studies demonstrates the effectiveness of the method.

Bayesian Regression Models for Seasonal Forecast of Precipitation over Korea

Seongil Jo
Seoul National University, Korea
joseongil@gmail.com

Yaeji Lim
Seoul National University, Korea
yaeji@snu.ac.kr

Jaeyong Lee
Seoul National University, Korea
leejyc@gmail.com

Hee-Seok Oh
Seoul National University, Korea
heeseok.oh@gmail.com

Hyun-Suk Kang
Korea Meteorological Adm., Korea
hyunsuk306.kang@gmail.com

The model output statistics (MOS) technique integrates the two sources of information, the numerical data generated by a general circulation model (GCM) and the observed climate data, to improve the accuracy of the climate prediction. In this paper, we propose new MOS techniques based on Bayesian linear regression with three different Bayesian model (or variable) selection methods and apply the proposed MOS techniques to the seasonal forecasting of the precipitation in a region around Korea. The Bayesian paradigm incorporates the model uncertainty as an integral part of modeling in a natural way.

Object Oriented Bayesian Networks Designing for Simplicity and Integration

Sandra Johnson and Kerrie Mengersen
Queensland University of Technology, Australia
sandra.johnson@qut.edu.au k.mengersen@qut.edu.au

Bayesian network (BN) modelling facilitates the integration of information from diverse sources and the representation of multiple perspectives. However, combining several perspectives in one model can lead to large, unwieldy BNs which are difficult to maintain and understand. On the other hand, an over-simplified model may lead to an unrealistic representation of the problem of interest. The challenge therefore is to design an integrated model that accurately reflects the current information of this problem, yet which is simple to maintain, expand and refine.

BN modelling is typically an iterative process and we describe here a heuristic method, the Iterative Bayesian Network Development Cycle (IBNDC), for the development of BN models within a multi-field and multi-expert context. The IBNDC caters for object oriented BN (OOBN) modelling. OOBN modelling can be viewed as the next logical step in adaptive management modelling, which embraces iterative development. The IBNDC is conducive to parallel OOBN development, combining different aspects of a problem while keeping the resulting combined model uncomplicated.

The IBNDC and OOBN modelling will be discussed and illustrated within the context of a case study on the viability of the free ranging cheetah (Acinonyx jubatus) population in Namibia.

On the Consistency of Bayesian Model Selection Procedures

VALEN E. JOHNSON
M.D. Anderson Cancer Center, USA
vejohnson@mdanderson.org

Let $\mathbf{Y}_n = (y_1, \ldots, y_n)'$ denote a random vector, $\mathbf{X}_n$ an $n \times q$ matrix of real numbers, and $\boldsymbol{\beta}_q$ a $q \times 1$ regression vector. This poster addresses the selection of non-zero components of $\boldsymbol{\beta}_q$ when it is assumed that $\mathbf{Y}_n \sim N(\mathbf{X}_n\boldsymbol{\beta}_q, \sigma^2\mathbf{I}_n)$ and $q = O(n)$. Model selection is based on the calculation of posterior model probabilities using non-local prior densities on the regression coefficients for each possible model. The non-local prior densities used for model definition are obtained as products of independent normal moment priors and are called $pMOM$ prior densities. Under mild conditions on the matrix $(\mathbf{X}_n'\mathbf{X}_n)^{-1}$, I demonstrate that the use of these priors guarantees that the posterior probability of the true model approaches 1 as $n \to \infty$.

To fix notation, assume that a component of $\boldsymbol{\beta}$, say β_h, is excluded from the "true" model if its value is 0, and denote a model by $\mathbf{j} = \{j_1, \ldots, j_k\}$ $(1 \leq j_1 < \cdots < j_k \leq p)$ if and only if $\boldsymbol{\beta}_{j_1} \neq 0, \cdots, \boldsymbol{\beta}_{j_k} \neq 0$ and all other elements of $\boldsymbol{\beta}$ are 0. Let $\mathbf{X}_j$ denote the design matrix formed from the columns of $\mathbf{X}_n$ corresponding to model $\mathbf{j}$. Let $\mathbf{t}$ denote the true model.

Under model $\mathbf{k}$, the sampling density for the data $\mathbf{Y}_n$ and prior density for $\boldsymbol{\beta}_k$ are assumed to be $\mathbf{Y}_n \,|\, \mathbf{k}, \sigma^2 \sim N(\mathbf{X}_k\boldsymbol{\beta}_k, \sigma^2\mathbf{I}_k)$, and $p(\boldsymbol{\beta}_k \,|\, \tau, \sigma^2) = (2\pi)^{-|\mathbf{k}|/2}(\tau\sigma^2)^{-3|\mathbf{k}|/2}\left(\prod_{i=1}^{|\mathbf{k}|}\boldsymbol{\beta}_{ki}^2\right)\exp\left(-\frac{1}{2\tau\sigma^2}\boldsymbol{\beta}_k'\boldsymbol{\beta}_k\right)$. The prior density on the non-zero parameters, $p(\boldsymbol{\beta}_k \,|\, \tau, \sigma^2)$, is called a non-local prior density because it is identically 0 if any component of $\boldsymbol{\beta}_k$ is 0 (Johnson and Rossell, 2010).

Let $\lambda_1(\mathbf{B}) \geq \lambda_2(\mathbf{B}) \geq \cdots \geq \lambda_m(\mathbf{B}) > 0$ denote the eigenvalues of an arbitrary positive definite matrix $\mathbf{B}$ of rank m. With this notation, the main result follows:

Theorem: *Let $p(\mathbf{k})$ and $p(\mathbf{k} \,|\, \mathbf{y}_n)$ denote the prior and posterior probabilities assigned to model $\mathbf{k}$, and suppose $p(\mathbf{t}) > \epsilon > 0$. Assume that $q < bn$ for some $b < 1$ and that there exists an N for which the following conditions obtain for all $n > N$: (i) there exists $M, c > 0$ such that $\lambda_1(\mathbf{X}_n'\mathbf{X}_n) < nM$ and $\lambda_q(\mathbf{X}_n'\mathbf{X}_q) > nc$, and (ii) there exists $d > 0$ such that $|\boldsymbol{\beta}_m| \geq \delta$ for all $m \in \mathbf{t}$. Then $p(\mathbf{t} \,|\, \mathbf{y}_n) \xrightarrow{p} 1$.*

I explore this result in a small simulation study and contrast this finding with model selection procedures based on local prior densities on regression coefficients. Under the conditions specified in the theorem, I show that $p(\mathbf{t} \,|\, \mathbf{y}_n) \xrightarrow{p} 0$ when a variety of local prior densities are used in the model specification, including the intrinsic priors examined in, for example, Casella et al (2009).

Using a Bayesian Hierarchical Model for Tree-Ring Dating

Emma M. Jones
University of Sheffield, UK
stp08emj@sheffield.ac.uk

Caitlin E. Buck
University of Sheffield, UK

Clifford D. Litton
University of Nottingham, UK

Cathy Tyers
University of Sheffield, UK

Alex Bayliss
English Heritage, UK

Dendrochronology, or tree-ring dating, uses the annual growth of tree-rings to date timber samples. Variation in ring width is determined by variation in the climate. Trees within the same geographical region are exposed to the same climatic signal in each year, but the signal differs from year to year.

Dendrochronologists measure sequences of tree-ring widths with a view to dating samples by matching undated sequences to dated sequences known as 'master' chronologies. The tree-ring widths from undated timbers are measured and the data are processed to remove growth trend. The processed data are sequentially matched against one another, each match position is known as an offset; initially matching timbers from the same site or woodland and then matching average sequences from each site or woodland, known as 'site' chronologies, to master chronologies.

The hierarchical nature of the data leads to modelling the data using a Bayesian hierarchical model. The ring-width for tree j in year i is modelled as the sum of the climatic signal in year i and a random noise which is particular to a tree j in year i. This model can be extended to include climatic signals at varying geographic scales. A Gibbs sampler is used to produce posterior probabilities for a match at each offset. This methodology relies on careful prior specification of parameters at each level of the hierarchy. Data are currently being collated from trees of known age from several woods in the UK that will be used to provide informative prior knowledge.

Stochastic Differential Equation Models for Dynamic Force

Chaitanya Joshi Simon Wilson
Trinity College Dublin, Ireland
joshic@tcd.ie swilson@tcd.ie

Eugene O'Brien and Abdolrahim Taheri
University College Dublin, Ireland
eugene.obrien@ucd.ie rahim.taheri@ucd.ie

This paper describes a stochastic model for the dynamic forces exerted by a fleet of trucks onto a road surface using stochastic differential equations (SDEs). The progression of pavement damage is known to be strongly related to mean patterns of applied force which result from the dynamic interaction of vehicles with the road profile. Using stochastic differential equations allows this process to be

more succinctly captured and quantified. In addition to the truck-to-truck force variations, the natural inherent variation, often otherwise referred to as the ëerrorí in the model, is captured. The solution to an SDE is a stochastic diffusion process and often the explicit solution and the transition density for realistic SDE models are not known and must be approximated. Fitting model parameters to data is also difficult because the mathematical properties of SDEs cause all common fitting procedures to converge very slowly. A method for simulating an SDE model based on the quarter car model is described, and a model fitting method based on Bayesian inference approaches ñ GaMBA (Gaussian Modified Bridge Approximation) ñ is developed. It is shown that GaMBA is computationally more efficient than the standard Monte Carlo simulation approach to Bayesian model fitting for SDEs, and that it produces accurate inference for standard SDE models. The application of the SDE model is demonstrated on a mechanistic-empirical pavement damage model.

Inferring the Topology of a Non-linear Sparse Gene Regulatory Network Using Fully Bayesian Spline Regression

Miguel A. Juárez
University of Warwick, UK
m.a.juarez@warwick.ac.uk

Edward R. Morrissey and Nigel J. Burroughs
University of Warwick, UK
e.r.morrisey@warwick.ac.uk n.j.burroughs@warwick.ac.uk

We propose a semi-parametric Bayesian model, based on penalised splines, for the recovery of an interaction network topology from longitudinal data. Our motivation is inference of gene regulatory networks from low resolution time series (10-50 time points) arising from gene expression microarray measurements. Such data frequently precludes the use of simple, parametric linear autoregressive models; thus we advance a semi-parametric model able to capture non-linearities. On the other hand, such biological networks are known to be sparse, which we impose in the network by augmenting the data with parent indicators and providing these with either an overall or gene-wise hierarchical structure. We provide conditions for posterior propriety under a broad class of frequently used improper priors. Appropriate specification of the prior is crucial to control the flexibility of the splines, especially under circumstances of scarce data; thus we provide an informative, proper prior and carry out sensitivity analyses to guide its calibration. The posterior is analytically intractable and numerical methods are needed. A Metropolis-within-Gibbs sampler is proposed, with a novel Metropolis-Hastings step for sampling the topology and the spline coefficients simultaneously. We also construct a parametric, linear autoregressive model and use it for comparison purposes. These models are illustrated using synthetic and data drawn from gene interaction ODE models and an experimental data set comprising time course gene expression measurements from the *Arabidopsis thaliana* circadian clock.

Objective Bayesian Analysis of Spatially Correlated Data Including Measurement Error

Hannes Kazianka
Technische Universität Wien, Austria
hannes.kazianka@tuwien.ac.at

Juergen Pilz
Alpen Adria Universität Klagenfurt, Austria
juergen.pilz@uni-klu.ac.at

Berger *et al.* (2000) developed default priors for unknown regression parameters $\boldsymbol{\beta}$, variance parameter σ^2 and range parameter ϑ_1 of Gaussian random fields with isotropic covariance function. However, they did not take account of a possible nugget effect, a discontinuity at 0 in the covariance function that models the measurement error. De Oliveira (2007) was the first to calculate non-informative priors for the nugget effect, ϑ_2, but treated the range parameter as being known. In this work we extend the previously mentioned approaches and show how non-informative priors for $\boldsymbol{\Theta} = \left(\boldsymbol{\beta}, \sigma^2, \vartheta_1, \vartheta_2\right)$ can be derived. Following Paulo (2005), we compute the Jeffreys-rule prior, the independence Jeffreys prior and the reference prior and prove that for the correlation function models under consideration (spherical, power exponential, rational quadratic, Matern) only the reference prior leads to a proper posterior in case of the spherical model and in case of the Matern model with smoothness parameter $\kappa < 0.5$. Moreover, we examine which conditions the hyperparameter a has to fulfill for various choices of the marginal prior $p\left(\vartheta_1, \vartheta_2\right)$ so that priors of the form $p\left(\boldsymbol{\Theta}\right) \propto p\left(\vartheta_1, \vartheta_2\right) / \left(\sigma^2\right)^a$ lead to a proper posterior distribution. We discuss how the sample size and the sampling design affect the priors. In a comparative study we compute the frequentist coverage of 95% Bayesian credible intervals and find that the reference prior shows the best performance while the Jeffreys-rule prior is worst. Finally, we use the presented methodology for Bayesian copula-based spatial modeling (Kazianka and Pilz, 2009; Kazianka and Pilz, 2010) of extreme valued data and present the results obtained for different choices of priors.

Bayesian and Frequentist Parametric Predictions of a Tail Probability in an Industrial Reliability Context

Merlin Keller
EDF R&D and AgroParisTech, France
merlinkeller@gmail.com

E. Parent N. Bousquet and A. Pasanisi
AgroParisTech. France and *EDF R&D, France*

We consider the problem of predicting the tail probability of an unknown parametric distribution, in view of applications to industrial risk studies. In practice, the so-called 'plug-in' approach is most frequently used, wherein a point estimate such as the maximum likelihood estimate (MLE) is substituted to the unknown

value of the parameter in the predictive distribution. This approach can of course be criticized in that it fails to account for the uncertainty on the unknown parameter (Smith, 1998). Bayesian approaches offer an attractive alternative as they account naturally for such an uncertainty. Moreover, admissibility theory ensures that, for any given predictor and loss function, a Bayesian predictor can always be found that performs at least as well in terms of risk. However such a dominant estimator can be difficult to exhibit in practice, and the plug-in approach can yield lower risks than standard Bayes approaches, depending on the choice of prior distribution and loss function (Smith, 1998; Ren *et al*, 2004; Ren *et al*, 2004). Using models commonly employed for industrial reliability analysis, we compare the risks of the plug-in and Bayes predictors under several loss functions. Second order asymptotic approximations are derived for these risks, as in (Smith, 1998; Ren *et al*, 2004; Ren *et al*, 2004). These theoretical results are complemented by a simulation study where the same risks are evaluated numerically. Our results give useful indications for the choice of an efficient prediction procedure in the context of industrial risk studies, and emphasize the drawbacks of using 'off-the-shelf' Bayes methods, which in certain cases may not perform as well as general theory may seem to suggest.

Measurement Uncertainty of ELISA Concentration Estimates in Biochemistry

Katy Klauenberg
Physikalisch-Technische Bundesanstalt (PTB), Germany
Katy.Klauenberg@PTB.de

B. Ebert J. Voigt and C. Elster
Physikalisch-Technische Bundesanstalt (PTB), Germany

Immunoassays are biochemical tests applied to measure even very small amounts of substance using the highly specific bindings between an antibody and its antigen. Immunoassays thus have a wide range of applications, e.g. to detect the presence of an infection, of hormones or drugs. This work focuses on an Enzyme-Linked ImmunoSorbent Assay, called sandwich ELISA, which allows detection of antigens by ësandwichingí them between two antibodies and labelling one type of antibody with an enzyme to generate a detectable signal (e.g. fluorescence).

ELISAs typically involve a high number of protocol steps, each susceptible to perturbations. A recent publication (Noble *et al.* 2008) has highlighted the variability in concentration estimates in the scope of an international comparability study. Some laboratories estimated an average concentration twice as high as other laboratories. But little is published on the uncertainties of individual laboratory estimates.

We recommend a Bayesian framework to rigorously quantify the uncertainty in ELISA concentration estimates. This framework encompasses simultaneous calibration of a nonlinear model and estimation of the unknown concentration from a set of fluorescence measurements. Applying this approach to ELISA measurements conducted by the PTB (the German national metrology institute), we demonstrate coherent intra-laboratory concentration estimation. The same

framework will be applied including all data of the above mentioned comparability study.

For ELISA concentration estimates, applying the standard international guideline to express measurement uncertainty in metrology (the GUM) will inevitably lead to inadequate uncertainties and possibly misleading mean estimates due to the inherent nonlinearity of the calibration curve. For the same token standard frequentist approaches (using asymptotic normal approximations) will give inadequate uncertainties. For nonstandard (e.g. nonlinear) models, metrologists would implement a Monte Carlo based approach according to a recent supplement, called GUM S1. However, especially at low concentrations this approach leads to deviations in uncertainties as well as mean estimates compared to the Bayesian approach, because substantive prior knowledge (such as nonnegativity of concentrations) is disregarded.

Representation of Interactions in Protein Annotation

Jukka Kohonen
Helsingin Yliopisto, Finland
jukka.kohonen@helsinki.fi

Pairwise interactions between proteins are an important source of information for automatic protein annotation using machine learning methods. The representation of such interaction data, however, poses a challenge. Interactions are sometimes represented as features such that when proteins A and B interact, protein A is credited with a feature "interacts with B". Such a representation is inherently very sparse, leading to both computational and statistical challenges.

We discuss different ways of representing such pairwise interactions, and compare how the representation affects annotation results.

On the Bayesian Forecasting Algorithm under the Non-Stationary Binomial Distribution with the Hyper Parameter Estimation

Daiki Koizumi
Waseda University, Japan
dkoizumi@m.ieice.org

Tota Suko and Toshiyasu Matsushima
Waseda University, Japan
suko@matsu.mgmt.waseda.ac.jp toshimat@waseda.jp

A Bayesian forecasting algorithm under the non-stationary binomial distribution is discussed. The proposed algorithm guarantees the Bayes optimal forecasting under certain non-stationary parameter model of binomial distribution. This model can be regarded as a special case of the Simple Power Steady Model (Smith, 1979)

defined as the forecasting model under the non-stationary exponential family of distribution with a known hyper parameter, but this work assumes that the hyper parameter is unknown to be estimated. Some numerical calculation results about forecasting as well as hyper parameter estimation performances would be discussed after the Bayesian forecasting method with the unknown hyper parameter is formulated.

Modelling Via Normalisation for Bayesian Nonparametric Inference

Michalis Kolossiatis
Cyprus University of Technology
michaelk_cy@yahoo.com

Jim E. Griffin
University of Kent, UK
J.E.Griffin-28@kent.ac.uk

Mark F. J. Steel
University of Warwick, UK
M.F.Steel@stats.warwick.ac.uk

Bayesian nonparametric methods are particularly useful for flexible modelling of data. For grouped data, a successful, general model was proposed in Müller, Quintana and Rosner (2004). In this model the distribution of the data in each group is assumed to consist of a common component, shared by all distributions, and an idiosyncratic component. Dependence between the groups is induced by this common part. In this talk we describe a general way of constructing dependent, identically distributed random probability measures (RPMs), that can be used as the prior of the distributions of grouped data. These RPMs are constructed through normalising appropriate random measures. We illustrate this for two dependent random distributions, using gamma processes as the underlying measures. By normalising them, we get Dirichlet process priors. The proposed model is similar to the model of Müller, Quintana and Rosner (2004), with components shared by some of the groups. These shared components introduce dependence between the data groups. The model has nice theoretical properties and simple updating schemes.

Implementing this model is achieved using MCMC algorithms. Marginal simulation methods are available, using its Pólya-urn representation. In the case of two groups, an additional mix-split step is proposed, in which we propose splitting a cluster from the common component into two clusters in the idiosyncratic components or merging two clusters from the idiosyncratic components into one cluster in the common component. It is shown that this step improves mixing, both for our model and the model of Müller, Quintana and Rosner (2004). Conditional simulation approaches are also available, using the slice sampler of Griffin and Walker (2009). The proposed model, together with the model of Müller, Quintana and Rosner (2004) and a model with normalised inverse-Gaussian process priors are then embedded in a stochastic frontier setting and used to study the efficiency of some hospital firms.

Bayesian Semiparametric Modeling with Mixtures of Symmetric Distributions

Athanasios Kottas
University of California, Santa Cruz, USA
thanos@ams.ucsc.edu

Gilbert W. Fellingham
Brigham Young University, USA
gwf@byu.edu

We propose a semiparametric modeling approach for mixtures of symmetric distributions. The mixture model is built from a common symmetric density with different components arising through different location parameters. This structure results in identifiability for mixture components, which is a key feature of the model as it allows applications to settings where primary interest is inference for the subpopulations comprising the mixture. We focus on the two-component and three-component mixture settings and develop a Bayesian model formulation using parametric priors for the location parameters and for the mixture proportions, and a nonparametric prior probability model, based on scale uniform Dirichlet process mixtures, for the random symmetric density. The performance of the model is studied with a simulation experiment. Moreover, the methodology is applied to a mixture deconvolution problem from epidemiological research.

Bayesian Flexible Fodels for Censored Data

Milovan Krnjajić
National University of Ireland
milovan.krnjajic@nuigalway.ie

Athanasios Kottas
Univ. California, Santa Cruz, USA
thanos@soe.ucsc.edu

In regression applications the relationship of the response distribution with covariates may differ for higher and lower quantiles compared to average responses. In such cases, considering a set of quantiles of the response distribution can uncover important features and so provide a more comprehensive analysis than offered by standard mean regression. We develop a series of flexible Bayesian models for analysis of the quantiles of response distribution in additive regression setting and work with data sets that include censored observations typical for medical applications. In the first modelling approach we propose semiparametric models based on different parametric quantile regression function and a Dirichlet process (DP) mixture (including dependent DP) for the error distribution. We also develop a fully non-parametric model for quantiles of the response distribution using Gaussian process for the quantile function and retaining non-parametric priors for the error. The proposed models can uncover non-linearity in the data and non-standard features in the error distribution. Inference is carried out using a number of posterior MCMC simulation methods for the combination of DP mixtures and Gaussian processes. Before their application to real data the models were extensively validated and their robustness and performance assessed using simulated data set.

On Bayesian Estimation of the Long-Memory Parameter in the FEXP-Model for Gaussian Time Series

WILLEM KRUIJER and JUDITH ROUSSEAU
Université Paris Dauphine and CREST
{kruijer, rousseau}@ceremade.dauphine.fr

Given observations $X_1, \ldots, X_n$ from a zero mean Gaussian time series, we estimate its spectral density f_0 within the FEXP-model containing spectral densities of the form $|1 - e^{ix}|^{-2d} \exp\left\{\sum_{j=0}^{k} \theta_j \cos(jx)\right\}$.

The FEXP-part $\exp\{\sum_{j=0}^{k} \theta_j \cos(jx)\}$ is a nonparametric model for the short-memory behavior of the time series, whereas the factor $|1 - e^{ix}|^{-2d}$ models its long-memory behavior. We study the semi-parametric problem of estimating the long-memory parameter d, considering θ as a nuisance parameter. The true spectral density f_0 is assumed to be of the form (**??**), with long-memory parameter d_0 and a possibly infinite FEXP-expansion of Sobolev-regularity $\beta > \frac{1}{2}$. We show that the Poisson prior considered in (Rousseau, Liseo and Chopin, 2009) necessarily gives a suboptimal rate of $v_n = n^{-\frac{2\beta-1}{4\beta+2}}$. Under sieve prior $k \sim n^{\frac{1}{2\beta}}$ however, the posterior for d, the minimax rate $n^{-\frac{2\beta-1}{2\beta}}$ is found. Our result can be seen as the Bayesian counterpart of the result obtained by Moulines and Soulier (2003) in a frequentist setting .

Approximate Inference for the Loss-Calibrated Bayesian

SIMON LACOSTE-JULIEN and ZOUBIN GHAHRAMANI
University of Cambridge, UK
sl522@cam.ac.uk zoubin@eng.cam.ac.uk

Bayesian decision theory provides a well-defined theoretical framework for rational decision making under uncertainty. Its ingredients are a loss $L(\theta, a)$ for an action $a \in \mathcal{A}$, a prior $p(\theta)$ and an observation model $p(x|\theta)$. Even if we assume that our subjective beliefs about $p(x, \theta)$ have been well-specified, we usually need to resort to approximations in order to use them in practice. Despite the central role of the loss in the decision theory formulation, most prevalent approximation methods seem to focus on approximating the posterior $p(\theta|x)$ *with no consideration of the loss*. In this work, our main point is to bring back in focus the need to *calibrate* the approximation methods to the loss under consideration. This philosophy has already been widely applied in the frequentist statistics / machine learning literature, as for example with the use of *surrogate loss functions* (Bartlett *et al. JASA* 2006). In contrast, the "loss-calibrated" approximation approach seems to have been mainly limited in Bayesian statistics to simple settings and losses such as regression with quadratic loss or hypothesis testing with 0-1 loss. We provide several examples showing the limitation of disregarding the loss in standard approximate inference schemes and explore loss-calibrated alternatives.

Adaptive Reversible Jump MCMC and Applications to Model Selection and Averaging in Geophysical Remote Sensing

Marko Laine and Johanna Tamminen
Ilmatieteen Laitos, Finland
marko.laine@fmi.fi johanna.tamminen@fmi.fi

We present adaptive automatic reversible jump MCMC algorithm called AARJ. It is based on the adaptive Metropolis algorithm by Haario et al. (2001) and automatic reversible jump by Green (2003). The novel feature in our algorithm is the fact that it is fully automatic and easy to use and implement. We show how the AARJ algorithm can be implemented and used for model selection and averaging in problems in atmospheric remote sensing. The AARJ method itself is general and can be applied to several other model selection problems as well.

We demonstrate the AARJ method by applying it to the statistical inversion problem in satellite remote sensing of atmosphere by Envisat/GOMOS and EOS-Aura/OMI instruments. The retrieval methods for constituent profiles of minor gases such as Ozone and NO2 have to take aerosols into account. However, the correct way to model aerosol extinction depends on the unknown type of aerosols present. By model selection methods various aerosol models can be simultaneously included in the retrieval problem. The technique allows us to study the probability of different aerosol models and also to include the uncertainty related to aerosol model selection into the posterior distributions of other constituents retrieved simultaneously.

Bayesian Conditional Autoregressive Geometric Process Model for Range Data

Connie P. Y. Lam Jennifer S. K. Chan
The University of Sydney, Australia
{plam7745, jennifer.chan}@uni.sydney.edu.au

Philip L. H. Yu
University of Hong Kong, HK
plhyu@hkucc.hku.hk

Boris S. T. Choy
University of Sydney, Australia
boris.choy@sydney.edu.au

Extreme values theory suggests that range is an effective estimator of volatility. Range-based volatility model is better than the return-based model in predicting power. This paper proposes an extended and flexible model named Conditional Autoregressive Geometric Process Range (CARGPR) model which incorporates GP model of Lam (1988) with CARR model of Chou (2005). Log-t distribution is used for robustness consideration. To facilitate a simpler Gibbs sampling algorithm for the WinBUGS package and an outlier diagnostics, the t-distribution is expressed as a scale mixture of normal (SMN) distribution. We demonstrate the model with the range data of All Ordinaries Index (AORD) and show that the proposed CARGPR model is better than CARR model both in estimation and forecasting.

Estimation of Conditional Quantiles from Bivariate Histogram Data

Philippe Lambert
Université de Liège, Belgium
p.lambert@ulg.ac.be

Bivariate histogram data are a special case of interval censored data where each observed component is only known to take a value in one of the predefined intervals partitioning the support of the corresponding variable. Such data are often met in practice when contingency tables are used to summarize the joint observation of quantitative variables. These are particularly useful to limit the storage requirement of large databases or to ensure confidentiality.

We show how penalized B-splines combined with the composite link model can be used to estimate a bivariate density from histogram data. Two strategies are proposed: the first one is semi-parametric with flexible margins modeled using B-splines and a parametric copula for the dependence structure ; the second one is "nonparametric" and is based on Kronecker products of the marginal B-splines bases. Frequentist and Bayesian estimations are described, with credible regions readily available in the latter case from the generated MCMC chains. A large simulation study quantifies the performances of both methods under different dependence structures and varying strengths of dependence, sample sizes and amounts of grouping. It suggests that Schwarz's BIC is a good tool for classifying the competing models.

These density estimates are used to evaluate conditional quantiles in two applications in social and in medical sciences. An extension of the strategy to larger dimensions will be discussed.

Stability of the Gibbs Sampler for Hierarchical Generalized Linear Model

Krzysztof Latuszynski
University of Warwick, UK
latuch@gmail.com

Omiros Papaspiliopoulos
Universitat Pompeu Fabra, Spain
omiros.papaspiliopoulos@upf.edu

Natesh S. Pillai and Gareth O. Roberts
University of Warwick, UK
{N.Pillai, Gareth.O.Roberts}@warwick.ac.uk

The prolific success of MCMC in the early 90's was fuelled by samplers that appeared to have excellent convergence properties. A theoretical understanding of the statistical features of these models which lead to these convergence properties has not been available.

Validity of conclusions drawn from MCMC inference relies heavily on ergodic properties of the underlying samplers. In particular the Central Limit Theorem

can fail without geometrical ergodicity. The aims of this work are to provide general results on geometric ergodicity to underpin MCMC application in a wide class of hierarchical models.

We consider Bayesian hierarchical generalized linear models that are a fundamental tool in statistical modelling, and widely applicable due to their flexibility. Bayesian inference for these models involves exploration of the posterior distribution which is analytically intractable, but can be sampled easily using the Gibbs sampler. Precisely, we consider convergence of Gibbs samplers for the joint posterior distribution of parameters and missing data in the following model:

$$\begin{aligned} Y_{ij} &\overset{\text{iid}}{\sim} f_1(\cdot, L(X_i)), \qquad 1 \le j \le n,\ Y_{ij} \in \mathcal{Y}, \\ X_i &\overset{\text{iid}}{\sim} f_2(\cdot, \Theta), \qquad 1 \le i \le m, \\ \Theta &\sim \pi(d\theta), \end{aligned}$$

where $L : \mathbb{R} \mapsto \mathbb{R}$ is a link function, $f_1(\cdot, L(X_i))$ is a probability density function on $\mathcal{Y}$ with parameter $L(X_i)$ and $f_2(\cdot, \Theta)$ is a probability density function on $\mathbb{R}$ with parameter Θ.

We devise easily applicable criteria that allow to conclude on ergodic properties of the Gibbs sampler, and distinguish between uniform ergodicity, geometric ergodicity and nonergodicity by examination of the characteristics of the model, such as densities f_1, f_2 and the link function L. This work generalizes substantially over (Papaspiliopoulos and Roberts, 2008).

We consider a number of examples, including the binary and Poisson regression with different link functions.

A non Linear Regression Model Applied to Electricity Load Forecasting

Tristan Launay
Université de Nantes and EDF R&D, France
tristan.launay@edf.fr

Anne Philippe
Université de Nantes, France
anne.philippe@univ-nantes.fr

Sophie Lamarche
EDF R&D, France
sophie.lamarche@edf.fr

The EVENTAIL model is a non linear Gaussian model currently used to forecast the electricity load in France. We approach the problem from a non-informative prior Bayesian perspective. The electricity load can be decomposed as a sum of two components : the seasonalities and the weather dependant part.

The non-linearity of the model comes from the heating part of the electricity load. It is of the form $g \times (T-u)\mathbb{I}_{T<u}$ which relies on the temperature T and two unknown parameters: the heating gradient g and the heating threshold u above which the heating part of the electricity load is assumed to be null.

We derive the non informative reference prior for the parameters (g, u) of the heating part. We compare the posterior distributions coming from the reference priors to the ones induced by flat priors. The comparison is done using Markov chain simulation to approximate the posterior distributions. In addition, we applied the whole model, with both the seasonal (linear regression) and weather

dependant (non-linear) parts together, on a dataset to show the influence of priors on posterior and predictive distributions, where the priors were learnt on another similar dataset.

Bayesian Nonparametric Modelling of Risk Factors for Retinopathy of Prematurity

Dominic S. Lee
University of Canterbury, New Zealand
dominic.lee@canterbury.ac.nz

Brian A. Darlow
Christchurch Schc. of Medicine, NZ
brian.darlow@chmeds.ac.nz

Marina Zahari
University of Canterbury, NZ
mza19@student.canterbury.ac.nz

Retinopathy of prematurity (ROP) is an eye condition in preterm babies caused by anomalous development of blood vessels in the retina, which can result in scarring, retinal detachment and even blindness. Some risk factors that have been suggested include gestational age, birth weight, gender (Darlow, *et al.* 2005) and the level or variability (or both) of oxygen saturation in the blood (York, *et al.* 2004). In this study, Bayesian nonparametric modelling is used to discover how some of these factors affect the risk of ROP.

Randomized Phase II Trial Design with Bayesian Adaptive Randomization and Predictive Probability

J. Jack Lee
University of Texas M.D. Anderson Cancer Center, USA
jjlee@mdanderson.org

Guosheng Yin
University of Hong Kong, HK
gyin@hku.hk

Nan Chen
M.D. Anderson Cancer Cent, USA
snchen2@mdanderson.org

We propose a randomized phase II clinical trial design based on the Bayesian adaptive randomization scheme and predictive probability monitoring. Adaptive randomization assigns more patients to a more efficacious treatment arm based on comparing the posterior probabilities of the efficacy between different arms. We continuously monitor the trial using the predictive probability. The trial is terminated early when it is shown that one treatment is overwhelmingly superior to the others or that all the treatments are equivalent. By coupling adaptive randomization and predictive probability approaches, the trial can treat more patients with a more efficacious treatment and allow for early stopping whenever sufficient information is obtained to conclude treatment superiority or equivalence. An efficient computation algorithm is developed based on the expectation of the future sample. The operating characteristics of the design is evaluated via simulations. The design is efficient, flexible, and ethical. It controls both type I and type II errors and compares favorably with frequentist designs.

Criminal Career Trajectory Latent Class Estimation Using a Bayesian Hierarchical Dynamic Linear Model

Thomas J. Leininger C. Shane Reese
Brigham Young University, USA
tjleininger@byu.edu reese@stat.byu.edu

Bert O. Burraston and Michael Rutkowski
Brigham Young University, USA

Criminal career trajectories are a sociological construct for understanding criminal behavior over time, ironically called criminal "careers". General approaches to estimation of criminal career trajectories (CCT) seek classes of criminal behavior through estimation of the expected number of crimes as a function of age. Trajectories are then used to compare criminals and describe different classes of criminals. However, commonly used approaches are not flexible enough to model the contrasting career profiles observed in modern criminology. We propose a Bayesian hierarchical dynamic linear model which provides increased flexibility in CCT estimation by borrowing strength across criminals with similar criminal behavior, both in intensity and in criminal life span. An additional feature of our approach is the identification of latent class membership for individual criminals. We first consider the problem of a fixed number of latent classes (with unknown trajectory and membership), using previous literature to establish the number of classes, and evaluate the optimal number of latent classes through model selection. We then extend the fixed CCT class model to include an unknown number of CCT classes. The latter, more general model allows for the trajectory intensity and shape as well as the number of trajectories to be estimated. Finally, we demonstrate the model's ability to predict class membership for juveniles as a function of covariates such as socioeconomic status, providing sociologists with better knowledge of where to focus crime-prevention efforts. An analysis of the Cambridge male cohort data (West and Farrington, 1973) is used to illustrate the performance of our model.

Bayesian Visual Analytics (BaVA): A Formal Visual Updating Procedure

Scotland C. Leman Leanna L. House and Chris North
Virginia Tech, USA
leman@vt.edu lhouse@vt.edu north@cs.vt.edu

Massive data-sets generally contain information which is concentrated in very small regions of a very high dimensional space. In order to extract this information, it is necessary to draw on several different fields, with a variety of tools. We develop a new analytics framework, merging two areas of research, Bayesian Statistics and Visual Analytics, which we moniker: Bayesian Visual Analytics (BaVA). Mathematical and statistical disciplines rely on model based formulations, making use of structured parameterizations; whereas, simple visualization of high dimensional data has utility, but coherent organization of the display is

often a difficulty. In this new paradigm, we synthesize these tools, via Bayes' theorem, in order to make a cohesive, adjustable visualization. This process of adjustment, based on user feedback is what we call "sense-making". This talk will focus on presenting the BaVA process, the formal descriptions of cognitive and parametric feedback, and an some illuminating examples.

A Bayesian Application of Lasso in Covariance Network Selection

Alan Lenarcic and Edo Airoldi
Harvard University, USA
lenarcic@post.harvard.edu airoldi@fas.harvard.edu

Gaining popularity of the use of probalistic networks to model covariance interactions in biological, ecological, and economic data has led to reexamination of structure suggested in Dempster's 1973 "Covariance Selection", or sparse inverse covariance matrix, framework. Advances in LASSO linear model selection, Tibshirani's frequentist penalty selection method, led to the iterative "Glasso" or Graphical-Lasso algorithm for an estimate that seeks to identify an unknown network with covariance data. LASSO estimates have the advantage of being fast to compute and require very few input parameters, but the motivation for these parameters is questionable and the specification is difficult, and the "Glasso" estimator's biases and power to identify models can be improved. We created the "Two-Lasso" framework to take Bayesian advantage of Lasso algorithms; positing a prior that is a mixture distribution of LaPlacian densities leads to an EM framework introducing estimates for latent model participation and demonstrably consistent point-estimates are made through available Lasso maximization steps. Here we apply the Two-Lasso to the problem of covariance selection, both in networks that are entirely unknown, in networks where certain connections are specified but others might be learned, and situations where network connectivity is expected a-priori to have certain features, such as generation from an exponential model or a fully connected network. With simulations we verify improved selection performance, and apply G2Lasso to exploratory needs in astronomy, biological pathways, and financial crashes.

Model Selection in Probit Models with Non-Intrinsic Bayes Factors

Luis G. Leon-Novelo and George Casella
University of Florida, USA
luis@stat.ufl.edu casella@stat.ufl.edu

In this poster we describe a variable selection procedure for dichotomous responses. The candidate models are all probit models. We are interested in identifying the regression coefficients (and then the covariates) impacting the distribution of the response. The procedure searches for the model with the highest Intrinsic Bayes Factor.

The procedure is automatic in the sense that it does not depend on tuning parameters since the specification of the prior distribution of each model is automatic. It is an adaptation of the "Limiting Intrinsic Procedure" of Moreno *et al.* (1998) for the dichotomous response case. They define an "Intrinsic Bayes Factor" (IBF) that is a limit of Bayes Factors (with proper priors). The IBF is an extension of the Average Intrinsic Bayes Factor (Berger and Pericchi, 1996) when nested models are considered.

We assess, via simulation, the accuracy of the proposed procedure. Finally, we apply it to a gene expression data set where the response is whether a patient exhibits a certain disease. Since the number of candidate covariates is large, we perform a stochastic search through the space of all possible models to identify the ones with the highest IBF.

A Bayesian Semiparametric Approach to Causal Inference with Intermediate Variables

Fan Li
Duke University, USA
fli@stat.duke.edu

Scott Schwartz
Duke University, USA
sls37@stat.duke.edu

Fabrizia Mealli
Università di Firenze, Italy
mealli@ds.unifi.it

In causal inference studies, treatment comparisons often need to adjust for confounded post-treatment (intermediate) variables, such as partial compliance in randomized trials. Principal stratification (PS) is a popular framework to deal with such variables. Continuous intermediate variables introduce inferential challenges to the PS analysis. Existing methods either dichotomize the intermediate variable, or assume a fully parametric model for the joint distribution of the potential intermediate variables. However, the former is subject to information loss and the latter is often inadequate to capture complex density shapes. We propose a Bayesian semiparametric approach that consists of a parametric model for the outcomes and a Bayesian nonparametric model for the intermediate variables using Dirichlet process via a stick-breaking (SB) prior. The SB prior provides flexibility in modeling the possibly complex joint distribution of the potential intermediate outcomes. The Gibbs sampling based posterior inference is developed. We illustrate the method by two applications with partial compliance: one concerning in benchmark Efron and Feldman (1991) data, and one concerning the effect of a preventive intervention program on children's conduct disorder.

Flexible Modeling of Conditional Distributions using Smooth Mixtures of Asymmetric Student t Densities

Feng Li
Stockholms Universitet, Sweden
feng.li@stat.su.se

Mattias Villani
Sveriges Riksbank, Sweden
mattias.villani@riksbank.se

Robert Kohn
Univ. New South Wales, Australia
r.kohn@unsw.edu.au

A general model is proposed for flexibly estimating the density of a continuous response variable conditional on a possibly high-dimensional set of covariates. The model is a finite mixture of asymmetric student-t densities with covariate dependent mixture weights. The four parameters of the components, the mean, degrees of freedom, scale and skewness, are all modeled as functions of the covariates. Inference is Bayesian and the computation is carried out using Markov chain Monte Carlo simulation. To enable model parsimony, a variable selection prior is used in each set of covariates and among the covariates in the mixing weights. The model is used to analyze the distribution of daily stock market returns, and shown to more accurately forecast the distribution of returns than other widely used models for financial data.

A Bayesian Mixture Model for Detecting Unusual Time Trends in Small Area Estimates: Application to COPD Mortality in England

Guangquan Li
Imperial College London, UK
guang.li@imperial.ac.uk

Nicky Best Sylvia Richardson and Anna Hansell
Imperial College London, UK
{n.best, sylvia.richardson, a.hansell} @imperial.ac.uk

Space-time modeling of small area data is often used in epidemiology for mapping chronic disease rates and by government statistical agencies for producing local estimates of, for example, unemployment or crime rates. While the temporal changes in most local areas tend to resemble each other closely, some may exhibit unexpected changes over time, suggesting, e.g., the emergence of localized predictors/risk factor(s) or a change in diagnostic or treatment techniques or the impact of a new policy. Detection of areas with unusual temporal patterns is thus important. In this poster, we propose a novel Bayesian mixture model for short time series of small area data that provides estimates of both the common temporal trend and the area-specific temporal trends. For each area, the posterior probability of belonging to the area-specific versus the common trend is used to classify the local time trend as unusual or not. To examine the detection ability, we have constructed a comprehensive simulation study, in which the proposed model

shows consistently good performance in detecting various departure patterns seen in real world situations. For real data applications, where the common/unusual status of the areas is unknown, we have developed a simulation-based approach to estimate the false discovery rates (FDR) and thus to determine area-specific cutoff values for the classification. Standard methods (e.g., Benjamini&Hochberg (1995) and Storey (2002)) for FDR estimation are not applicable here as each area has its own alternative trend, which violates the identical assumption made in multiple testing. We have applied our model to a retrospective study of chronic obstructive pulmonary disease (COPD) mortality data in England (1990-1999) to assess the impact of government policy to make COPD a compensatable disease in 1992. Inspection of the identified unusual trends reveals some interesting stories of the corresponding areas, which will be detailed in the poster.

Bayesian Hierarchical Shrinkage Prior for Sparse Signals

Yingbo Li and Merlise A. Clyde
Duke University, USA
yl118@duke.edu clyde@stat.duke.edu

In this paper, we propose a new fully Bayes approach for estimating sparse signals from background Gaussian white noise. We construct a hierarchical shrinkage prior as mixtures of Cauchy densities, which in the limit leads to a Levy random field prior. The prior can also be considered as a mixture of a point mass at zero and a heavy-tailed density which permits it to adapt to different sparsity structures. By studying the tail property of its induced marginal likelihood, we prove it has bounded influence. This hierarchical shrinkage prior shrinks small values directly towards zero while keeping large signals almost unshrunk. The infinite divisibility of our prior leads to coherent prior specifications as the number of predictors increases. Based on simulation studies, under certain circumstances, our prior can achieve higher accuracy in terms of sum of squared error, comparing with some existing Bayesian model selection approaches such as Johnstone and Silverman's Empirical Bayes estimates.

Stitching Time to Analyze Nine: Hierarchical Modeling of Longitudinally Collected HIV-1 Phylogenies

Li-Jung Liang Robert Weiss Marc Suchard
University of California at Los Angeles, USA
{liangl, robweiss, msuchard}@ucla.edu

HIV-1 researchers are interested in the evolution of the HIV-1 virus over time within infected subjects. We have sets of HIV-1 molecular sequences from blood samples taken over time from nine subjects enrolled in the Multicenter AIDS cohort Study starting from the time that the subjects were first infected with the HIV-1 virus. Each blood draw at one time point contributes a sample of HIV-1 molecular sequences which can be assembled into a phylogenetic tree using the phylogenetic model that jointly infers alignment and phylogeny (Redelings and

Suchard, 2007). To describe how the indel processes of HIV-1 virus changes over time, we first fit individual phylogenetic models separately for the sequences from a patient at a single visit. Next, we fit a semi-parametric Bayesian hierarchical random effects regression model to combine the individual patient-visit indel analyses to better estimate the bivariate indel parameters of interest within and across patients. This model allows for patient-specific covariates and time effects on the indel parameters of interest and is computationally intensive. The key to this approach is that a dynamic iteratively reweighted Markov chain Monte Carlo algorithm conveniently recycles posterior samples from the individual analyses. In addition, we modify two Bayes factor algorithms (Verdinelli and Wasserman 1995; Basu and Chib 2003) to give us Bayes factors for our model.

Structured Prior Modelling for Bayesian Analysis of Complex Mixtures in Combinatorial Marker-Based Flow Cytometry Studies

Lin Lin Cliburn Chan and Mike West
Duke University, USA
{ll86, cliburn.chan, mike}@stat.duke.edu

Novel uses of automated flow cytometry technology for measuring levels of protein markers on thousands to millions of cells are promoting increasing need for relevant, customized Bayesian mixture modelling approaches in many areas of biomedical research and application. In studies of immune profiling in many biological areas, traditional flow cytometry measures relative levels of abundance of marker proteins using fluorescently labeled tags that identify specific markers by a single colour. One specific and important recent development in this area is the use of combinatorial marker assays in which each marker is targeted with a probe that is labeled with two or more fluorescent tags. Thus, use of K colours enables the identification of, in principle, up to (K choose m) subtypes of cells using m colours per type, representing potentially major advances in the ability to characterize variation in immune responses involving larger numbers of cell subtypes. We describe novel cl asses of Bayesian mixture models with highly structured priors that reflect the design and nature of resulting data in this setting of combinatorial encoding. We develop Bayesian computations using MCMC, discussing the challenging issue of identification and a novel, general approach to its solution, provide evaluation examples with simulated data and describe application in studies in cancer immunology.

Bayesian Partition Model Selection on High Dimensional Dependent Time series

Silvia Liverani
University of Bristol, UK
s.liverani@bristol.ac.uk

Jim Q. Smith
University of Warwick, UK
j.q.smith@warwick.ac.uk

One important recent domain where partition models are used is in the study of microarray models. Microarray experiments that measure gene expression of tens

of thousands of genes are now widespread and their shear size presents a challenge to any probability distribution guided clustering algorithm. More recently a large number of experiments have been performed that collect short longitudinal time courses - or time profiles - of microarrays. These profiles have been very useful in helping scientists to discover new genes in the various regulatory pathways in the studied organisms.

Because of its transparency one particularly successful methodology within this domain has been the use of MAP model selection on partitions of different clusters. MAP model selection using Gaussian regression models has been demonstrated to be an excellent tool for exploring vast partition spaces of clusters of thousands of short time series. Because of conjugacy, each partition can be scored algebraically using an explicit Bayes Factor score. This means that the score of each partition can be evaluated very quickly. It follows that fast local search algorithms can be derived. These algorithms explore the space and partition the space into co-expressing units exhibiting the same expected time series profiles.

However, this type of exploration is usually performed under the assumption that units lying in different clusters express independently of one another. In the context where the clusters form part of a mutual regulatory system this independence assumption is not tenable since regulation must logically entail dependence between the regulating unit in one cluster affecting the development of a regulated unit in a different cluster. In particular this means that no search can output hypotheses about how clusters might communicate with one another: one of the features of the process of most interest to the scientist.

In this presentation we demonstrate how the model space searched can be extended to include the dependence models that the scientist might hypothesise: in particular the excitation or inhibition of one cluster of genes on another. We represent the types of dependence deriving from regulation in a conjugate way so that the dependence models can be scored analogously to the widely-studied independence models.

The challenge we face is that our extended class of models is potentially orders of magnitude larger than the original space, which is itself enormous. Fortunately current scientific hypotheses posit sparse relationships between combinations of clusters. This makes it possible for us to design intelligent exploration algorithms so that some of the most promising regulatory models are traversed. The MAP score functions of the dependent models, whist being of closed form, are not as simple as those in the original space but it is nevertheless possible to adapt current propagation algorithms to provide exact or fast approximate scores for the dependent partition. We feed back the MAP dependence model to the scientist using a novel graph. Its semantics not only provides a formal description of the underlying explanatory probability model but also closely matches semantics familiar to the scientist. The paper is illustrated with an analysis of a topical data set.

Bayesian Inference in a Genome-Wide Association Kin-Cohort Studies

Charalambos Loizides
University of Oxford, UK
loizides@stats.ox.ac.uk

Christopher Yau and Chris Holmes
University of Oxford, UK
yau@stats.ox.ac.uk cholmes@stats.ox.ac.uk

Genome-wide association studies have fundamentally altered the study of genetics in recent years by allowing genome-wide screening of single nucleotide polymorphisms for association with a variety of common diseases with many significant findings. Most studies typically adopt a standard case-control design, involving collections of diseased and non-diseased individuals; or a family-based design, in which one studies the relatives of affected individuals. Here, we describe a Bayesian approach for inference in a hybrid study design known as the kin-cohort design, where the disease status for first degree relatives - siblings and parents - of case and control individuals are known but only the genotypes of case-control individuals are observed. We use hierarchical modelling where we assume a logistic regression model for the probability of disease and which also includes the missing genotypes of the first degree relatives. Assuming Mendelian inheritance these genotypes can be probabilistically imputed given the case-control genotypes and family member's disease status. This leads to a relative increase of the sample size and thus improves posterior inference of the parameters of interest. We develop exact Markov Chain Monte Carlo and approximate Expectation Maximisation methods for inference with this model and demonstrate that by retrospectively adopting the kin-cohort design, following an initial case-control study, the collection of additional phenotype data regarding disease status of relatives can increase power to detect disease associations and boost the information content of an existing case-control study at little extra cost.

Application of Gaussian Predictive Process to the Evaluations of a Volcanic Flow Computer Model

Danilo Lopes
Duke University, USA
dll11@stat.duke.edu

Our interest is the risk assessment of rare natural hazards, such as large volcanic pyroclastic flows. Since volcano eruptions are extremely rare events, we use a computer model to simulate natural conditions that are possible, but have not been observed yet, in order to help the risk assessment. A volcano simulator, however, is usually too computationally expensive to be used in the analysis. We then use a statistical approximation (emulator) to predict the output of this computer model at untried values of inputs. Gaussian process response surface is a technique commonly used in these applications, because it is fast and easy to

use in the analysis. However, there is a big disadvantage in the use of emulators: although a large data set is desirable in order to enrich the approximation, a large number of computer model runs represents an increase in the complexity of matrix decompositions of the Gaussian process during the Markov Chain Monte Carlo iterations. One part of our research pro ject is to explore and compare di?erent statistical approaches to the approximation of computer model evaluations, given some computational limits. Modified Predictive Process (Banerjee et al, 2008) is a class of model developed for similar problems in spatial statistics; this class of models does not involve complex matrix decom- positions, and has been showing some promising results in the solution of our problem. Other important part of our research pro ject is to deal with the spatial information in our computer model output: the output is a surface of maximum volcanic flow height over some geographical area. We know that the topology of the volcano area plays an important role in determining the shape of this surface, and we propose one way to include this information in our analysis.

Particle Learning for Sequential Bayesian Computation

Hedibert F. Lopes Carlos M. Carvalho
The University of Chicago, USA
{hlopes, carlos.carvalho}@chicagobooth.edu

Michael S. Johannes
Columbia University, USA
mj335@columbia.edu

Nicholas G. Polson
University of Chicago, USA
ngp@chicagobooth.edu

Particle learning provides a simulation-based approach to sequential Bayesian computation. To sample from a posterior distribution of interest we use an essential state vector together with a predictive and propagation rule to build a resampling-sampling framework. Predictive inference and sequential Bayes factors are a direct by-product. Our approach provides a simple yet powerful framework for the construction of sequential posterior sampling strategies for a variety of commonly used models.

Moving Average Smoothing for Spatio-Temporal Disease Mapping

Antonio López-Quílez
Universitat de València, Spain
antonio.lopez@uv.es

Paloma Botella
CEU-Cardenal Herrera, Spain
pbotella@uch.ceu.es

Miguel A. Martínez B.
CSISP, Valencia, Spain
martinez_mig@gva.es

Spatial modelling of disease occurrence is a very active research field in epidemiological applications. The use of Intrinsic Gaussian Markov Random Fields

(IGMRF), with a heterogeneous effect for every region, has been the usual procedure to model the underlying risk variability in many of these studies. Nevertheless, the correlation structure in an IGMRF is completely determined by the geographical structure of the lattice under study, therefore it is not possible to adapt the dependence structure of this prior distribution to the geographical pattern of the disease under study. Moving average provides an alternative class of spatial correlation structures different to IGMRF. In the same way that IGMRF generalizes random walk processes to the spatial domain, we will resort to moving average ideas in time series to induce spatial dependence in our context.

The main objective of this work will be to propose moving average smoothing of risks and extend its use for the spatio-temporal modelling of diseases. The model proposed will be formulated from a Bayesian perspective and Reversible Jump MCMC will be used to learn about the neighbouring range of dependence of the spatial pattern studied.

Rotating Stars and Revolving Planets: Bayesian Exploration of the Pulsating Sky

Thomas J. Loredo
Cornel l University, USA
loredo@astro.cornell.edu

I describe recent and ongoing work on development of Bayesian methods for exploring periodically varying phenomena in astronomy, in two areas: pulsars, and extrasolar planets (exoplanets). For pulsars, the methods address detec- tion and measurement of periodically varying signals in data consisting of photon arrival times, modeled as a non-homogeneous Poisson point processes. For exoplanets, the methods address detection and estimation of planetary orbits from data that measure the reflex motion "wobble" of a host star, and adaptive scheduling of observations.

Product Partition Models with Correlated Parameters

Rosangela H. Loschi
Universidade Federal de Minas Gerais, Brazil
loschi@est.ufmg.br

João V. D. Monteiro
University of Minnesota, USA
monte092@umn.edu

Renato M. Assunção
Univ. Fed. Minas Gerais, Brazil
assuncao@est.ufmg.br

In time series, Bayesian partition models aim at partitioning the entire observation period into disjoint temporal clusters. Each cluster is an aggregation of sequential observations and a simple model is adopted within each cluster. The main inferential problem is the estimation of the number and locations of the temporal clusters. The popularity of partition models is justified by its flexibility to analyze change point or clustering problems. However, its original formulation

assumes a common parameter indexing the distributions of the observations into the same temporal cluster. Furthermore, it also assumes independence among the common parameters associated with different temporal clusters. This approach may lead to an inaccurate identification of the number of clusters. We extend the well-known product partition model (PPM) for clustering analysis in the temporal context. We also assume independence among parameters in different temporal clusters, but contrary to what is assumed in the PPM, we consider that the observations in the same cluster have their distributions indexed by different parameters. Although different, the parameters are similar for observations within a given cluster. This is done by adopting a Gibbs distribution as the prior specification for the canonical parameters. As a result, the parameters within the same temporal cluster are correlated. One important advantage in allowing similar parameters within a temporal cluster is that, rather than having an unknown dimension, the dimension of the parameter vector is fixed and equal to the time series length. This facilitates the numerical procedures used to obtain the posterior distribution. We carried out several simulations and real dataset analyzes showing that our model provides better estimates for all parameters, including the number and position of the temporal clusters, even for situations favoring the PPM.

Association Tests that Accommodate Genotyping Uncertainty

Thomas A. Louis
Johns Hopkins Bloomberg School of Public Health, USA
tlouis@jhsph.edu

Qing Li
NIH-Genome Res. Inst., USA

Benilton S. Carvalho
CRUK Res. Inst., Cambridge, UK

Daniele Fallin Rafael A. Irizarry and Ingo Ruczinski
Johns Hopkins Bloomberg School of Public Health, USA

High-throughput single nucleotide polymorphism (SNP) arrays, typically used in genome-wide association studies with a trait of interest, provide estimates of genotypes for up to one million loci. Most genotype estimates are very accurate, but genotyping errors do occur and can influence test statistics, p-values and ranks. Some SNPs are harder to call than others due to probe properties and other technical/biological factors; uncertainties can be associated with features of interest. SNP and case-specific genotype posterior probabilities are available, but they are typically not used or used only informally, for example by setting aside the most uncertain calls. To improve on these approaches we take full advantage of Bayesian structuring and develop an analytic framework that accommodates genotype uncertainties in case-control studies. We show that the power of a score test (and statistical information more generally) is directly a function of the correlation of the genotype probabilities with the true genotypes. We demonstrate that compared to picking a single AA, AB or BB genotype or to setting aside diffcult calls, Bayesian structuring can substantially increase statistical information

for detecting a true association and for ranking SNPs whether the ranking be frequentist or optimal Bayes.

Clustering Posterior Densities-Application in Microarrays

Tanzy Love
University of Rochester, USA
tlove@bst.rochester.edu

Alicia Carriquiry
Iowa State University, USA
alicia@iastate.edu

In large time series or multiple treatment microarray experiments, we are interested in locating groups of genes that react together. Subject matter theory designates these groups as coregulated by the same biologic pathways. For example, genes responsible for photosynthetic processes may express together in an experiment covering time periods in light and darkness. The statistical problem is then clustering genes based on their expression values over multiple treatments. However, we don't have values for gene expression, rather replicated estimates for each treatment condition. The posterior mean of the gene expression estimates could be used in clustering genes. However, this ignores the uncertainty we have in the actual values of expression. To incorporate this uncertainty, we have modeled expression estimates using hierarchical models. This provides posterior probability distributions for quantities such as expression value and expression ratio for two treatments. We also can construct the joint posterior probability distribution of all expression values or all pairwise expression ratios for each gene. We use multiple sampling from the posterior distributions of gene expression vectors to cluster genes and estimate the uncertainty in this clustering.

How Should We Combine Expert Opinions: On Elicitation, Encoding, Priors or Posteriors?

Samantha J. Low-Choy
CRC National Plant Biosecurity, Australia
s.lowchoy@crcplantbiosecurity.com.au

Kerrie Mengersen
Queensland Univ. Tech., Australia
k.mengersen@qut.edu.au

Justine V. Murray
CSIRO Entomology, Australia
Justine.Murray@csiro.au

Judith Rousseau
CEREMADE, Université Dauphine, France
rousseau@ceremade.dauphine.fr

Within the Bayesian paradigm, expert knowledge is typically used to construct informative priors, with an emphasis on combination with empirical data to form posterior estimates. In pioneering research, however, the initial step of representing the current state of knowledge may deserve more attention, due to its pivotal role in helping focus scientific investigation and guide data collection. In this paper we concentrate on situations we have encountered in pioneering research

where it is worthwhile to invest considerable effort in both design and analysis of expert opinions. A recurring example from our experience involves discerning habitat preferences, in order to map potential spatial distribution, of rare and threatened species. Expert opinion is valuable in these contexts since empirical data is typically limited due to sparsity of patterns occurring across broad spatio-temporal extents. This problem arises both in landscape ecology where the aim is to conserve key species, and in biosecurity where the aim is to estimate risk that a pest species establishes.

Many mathematical methods for collating expert opinion have been developed. In this work we demonstrate the flexibility of a Bayesian statistical modelling framework, through careful consideration of how variability among and within experts enters into the model. Most commonly, expert opinions are collated once encoded into individual priors. It is also possible to aggregate posteriors resulting from an analysis informed by each expert individually via a form of Bayesian model averaging. When indirect elicitation methods are used then expert opinion may also be combined in intermediate steps, either by combining the elicited information or waiting until it is encoded to combine it. We compare these four methods of aggregating expert opinions using a case study on eliciting habitat preferences. This allows us to examine the choice of the level of aggregation from the perspective of the sources of variation addressed, and the summaries they provide. In addition, aggregation of expert opinion provides a useful analogy for aggregation of models more generally. We show that this aggregation can be flexibly accommodated within the Bayesian framework: by aggregating posteriors, priors, or inputs to priors.

Bayesian Spatial Quantile Regression: An Application to Conditional Birthweight Distributions

Kristian Lum and Alan Gelfand
Duke University, USA
kristian.lum@stat.duke.edu alan@stat.duke.edu

Quantile regression offers the investigator the opportunity to understand how the regression covariates affect not only the mean response but also the whole conditional response distribution, giving a more complete (and robust to outliers) analysis of the effects of covariates, e.g., a covariate may be more or less influential as we move across the range of quantile values There are various specifications that can be introduced to model residuals in quantile regression. An attractive, recently proposed approach assumes that the residuals are distributed asymmetric Laplace, parameterized by the quantile p and a scale parameter.

To date, little work has been done with quantile regression for spatially dependent data. Here we envision the regression varying not only with quantile but also across space, anticipating that locations near to each other will have similar regressions. In this setting, we present a spatial quantile model developed by taking advantage of the representation of the asymmetric Laplace distribution as a mixture of a Normal (here, a spatial Gaussian process) and latent independent identically distributed exponential random variables.

We investigate the correlation properties of the resulting error process model and show how to fit such models within an MCMC framework. We seek to apply

this model to a large set of birth weights and maternal covariates drawn from the North Carolina Detailed Birth Record database. Particular interest is in quantiles associated with low birth weight. The size of the dataset is on the order of 15,000 births; a large dataset is needed in order to learn about regressions over quantiles and over space. Given the computational difficulties that arise in spatial analyses of large data sets, we employ an extension of the predictive process model to enable computational tractability.

Using Expression Data in Genetic Association Studies: An Integrated Bayesian Approach to Determine Genetic Pathways

Sharon M. Lutz
Universität Bonn, Germany
smlutz@hsph.harvard.edu

Scott T. Weiss and Christoph Lange
Harvard University, USA
scott.weiss@channing.harvard.edu clange@hsph.harvard.edu

Following the paradigm that the biological road from a genetic locus to the disease phenotype of interest leads through expression data, the availability of genetic data, expression profiles and the disease phenotype should allow for the identification of genetic pathways. However, statistical analysis approaches that integrate all three of these types of data are not straightforward. Given an association between the marker locus and the expression profile, and an association between the marker locus and the disease phenotype, we want to conclude that the genetic association with the phenotype is solely attributable to the genetic association with the expression profile in order to establish the pathway from gene to disease. As we show in this manuscript, this question can not be addressed using standard statistical methodology for hypothesis testing. We propose a Bayesian approach that can assess the genetic association with the phenotype, in the presence of the genomic association. Using simulation studies, we verify that the our approach has the desired properties. The approach is illustrated by an application to a real data set.

On Bayesian Variable Selection Using Lasso and Related Methods

Anastasia Lykou
Lancaster University, UK
a.lykou@lancaster.ac.uk

Ioannis Ntzoufras
Athens University of Economics and Business, Greece
ntzoufras@aueb.gr

We propose a Bayesian method that accomplishes both shrinkage and variable selection by using the Lasso (Tibshirani, 1996) and related methods. The proposed

method takes advantages of the shrinkage methods, which can reduce efficiently the prediction error avoiding problems appearing when collinearity is present and indirectly implies which variables can be removed from the model. Similar approaches can be found in the literature by Park and Casella (2008) and Hans (2009).

Among the challenges is to choose the appropriate shrinkage parameter, which controls the shrinkage applied and the selected variables. Here we explore the sensitivity of the Bayes factor on this choice. We graphically represent this sensitivity on plots similar to the ones used in standard Lasso methods. We facilitate the univariate regressions and their Bayes factors (when compared to the null model) to tune and specify hyperpriors for the shrinkage parameter.

Bayesian Analyses in USA Sexually Violent Predator Civil Commitments

Diane Lytton
Private Practice, USA
dlytton@prodigy.net

The brief history of Bayesian analyses in USA sexually violent predator (SVP) civil commitments will be reviewed, and lead to examination of a large 10-year sample of SVP evaluations. The hypothesis is that most experts neglect Bayesian principles. SVP laws allow for incapacitation of sex offenders after completion of prison sentences. Over 4,000 SVPs have been civilly committed in 20 USA states; however, few have gained release. Wollert (2006, 2007, 2008) and colleagues (Waggoner *et al.* 2008; Wollert & Waggoner, 2009; Donaldson & Wollert, 2009), and others (e.g., Mossman, 2006), demonstrated that Bayes's Theorem leads to better-informed opinions, avoids confirmatory biases, and illusions of certainty. Interestingly, recognition that Bayesian analyses can further science in this important area, has been met with obfuscation, and illogical counter-arguments. Bayesian analyses can better inform two issues central to SVP evaluations: diagnosis (mental disorder) and risk to reoffend. Risk statements from the sample of SVP evaluations will be examined for accuracy of probabilistic estimates derived from risk instruments, from a Bayesian perspective. Following Wollert (2007), accuracy of diagnostic opinions will also be estimated. Trends over time will be examined for whether Bayesian analyses are being introduced, e.g., recognition of the importance of base rates (prior probabilities). Lastly, recent recommendations made by developers of a widely-used risk instrument will also be critiqued, by use of Bayesian logic.

Inference for the Generalized Rayleigh Distribution Based on Progressively Censored Data

Mohamed T. Madi and Mohamed Z. Raqab
UAE University, UAE *University of Jordan, Jordan*
mmadi@uaeu.ac.ae mraqab@ju.edu.jo

In this paper, and based on a progressive type II censored sample from the generalized Rayleigh distribution, we consider the problem of estimating the model

parameters and predicting the unobserved removed data. Maximum likelihood and Bayesian approaches are used to estimate the scale and shape parameters. The Gibbs and Metropolis samplers are used to predict the life lengths of the removed units in multiple stages of the progressively censored sample. An artificial and real data analyses have been performed for illustrative purposes, and a simulation study is carried out to compare the MLE's and Bayes approaches. It is observed the Bayes estimators perform better than the MLE's and are less sensitive to sample size and censoring schemes.

Bayesian Methodsn Pharmacovigilance

David Madigan
Columbia University, USA
madigan@stat.columbia.edu
Patrick Ryan and Shawn Simpson
GlaxoSmithKline, USA *Columbia University, USA*

Increasing scientific, regulatory and public scrutiny focuses on the obligation of the medical community, pharmaceutical industry and health authorities to ensure that marketed drugs have acceptable benefit-risk profiles. This is an intricate and ongoing process that begins with carefully designed randomized clinical trials prior to approval but continues after regulatory market authorization when the drug is in widespread clinical use. In the post-approval environment, surveillance schemes based on spontaneous reporting systems (SRS) represent a cornerstone for the early detection of drug hazards that are novel by virtue of their clinical nature, severity and/or frequency. SRS databases collect reports of adverse events made directly to the regulator or to the product manufacturer by clinicians or patientswithout regard to any assessment of causality. Such spontaneous report databases present a number of well-documented limitations such as under-reporting, over-reporting, and duplicate reporting. Furthermore, SRS databases have limited temporal information with regard to duration of exposure and the time order of exposure and condition.

Despite the limitations inherent in SRS-based pharmacovigilance, analytic methods for spontaneous report databases have attracted considerable attention in the last decade, and several different methods have become well established, both in commercial software products and in the medical literature. All of the more widely used methods compute measures of "disproportionality" for specific drug-condition pairs. That is, the methods quantify the extent to which a given condition is "disproportionally" reported with a given drug. We describe Bayesian shrinkage approaches that have been widely adopted in this context. More recent work has focused on Bayesian regression approaches that attempt to deal with certain types of confounding.

Newer data sources have emerged that overcome some of the SRS limitations but present methodological and logistical challenges of their own. Longitudinal observational databases (LODs) provide time-stamped patient-level medical information, such as periods of drug exposure and dates of diagnoses. Typical examples include medical claims databases and electronic health record systems.

The scale of some of these databases presents interesting computational challenges – the larger claims databases contain upwards of 50 million lives with up to 10 years of data per life. A nascent literature on signal detection in LODs now exists including adaptations of some of the Bayesian methods developed in the SRS context. We also consider one particular approach that draws on a method known in epidemiology as the self-controlled case series. We present a Bayesian analysis of this method and describe some generalizations.

Sensitivity to Prior Hyperparameters in Gaussian Bayesian Networks

Paloma Main
Univ. Complutense Madrid, Spain
pmain@mat.ucm.es

Miguel Gómez-Villegas,
Univ. Complutense Madrid, Spain
ma_ga@mat.ucm.es

Hilario Navarro
Univ. Complutense Madrid, Spain
hnavarro@ccia.uned.es

Rosario Susi
UNED, Spain
rsusi@estad.ucm.es)

Our focus is on learning Gaussian Bayesian networks(GBNs) from data. In GBNs the multivariate normal joint distribution can be alternatively specified by the normal regression models of each variable given its parents in a DAG (directed acyclic graph). In the latter representation the parameters are the mean vector, the regression coefficients and the corresponding conditional variances. The problem of bayesian learning in this context has been handled with different approximations, all of them concerning the use of different priors for the parameters considered. We work with the most usual prior given by the normal/inverse ,gamma form. In this setting we are interested in evaluating the effect of prior hyperparameters choice on the posterior distributions. The Kullback-Leibler divergence measure is used as a tool to define local sensitivity comparing the prior and posterior deviations. This method can be useful to decide the values to be chosen for the hyperparameters.

Multiset Model Selection

Dipayan Maiti and Scotland Leman
Virginia Polytechnic Institute and State University, USA
dipayanm@vt.edu leman@vt.edu

The Multiset Sampler has previously been deployed and developed for efficient sampling from complex stochastic processes. We extend the sampler and the surrounding theory to high dimensional model selection problems. In such problems efficient exploration of the model space becomes a challenge since independent and ad-hoc proposals might not be able to jointly propose multiple parameter sets which correctly explain a new proposed model. In order to overcome this we propose a multiset on the model space to enable efficient exploration of multiple model modes. The model selection framework is based on independent priors for

the parameters and model indicators on variables. While under this method we do not obtain typical Bayesian model averaged estimates for the parameters, we show that the multiset model averaged parameter estimate is a mixture a distribution from which the true Bayesian model probabilities and the model averaged parameter estimate can be obtained. We explore effects of model dimensionality, correlation in the predictor space and size of the multiset on computational complexity and exploration of the model space.

Bayesian Variable Selection in Normal Regression Models

Gertraud Malsiner-Walli and Helga Wagner
Johannes Kepler Universität Linz, Austria
gertraud.malsiner_walli@jku.at helga.wagner@jku.at

An important task in building a regression model is to decide which variables should be included as regressors. In a Bayesian approach variable selection is usually accomplished by MCMC methods with spike and slab priors on the effects subject to selection. In this presentation we compare different versions of spike and slab priors for variable selection in normal regression models. We consider priors, where the spike is a point mass at zero and specify the slab as independent normal distributions, as a g-prior or as a fractional prior. Variable selection under this type of prior requires to compute marginal likelihoods with integration over the parameters subject to selection.

A second type of priors specifies both the spike and the slab as continuous distributions, e.g. normal distributions (as in the SSVS appraoch) or scale mixtures of normal distributions. These priors allow a simpler MCMC algorithm where no marginal likelihood has to be computed. In a simulation study with different settings (independent or correlated regressors, different scales) the performance of these priors with respect to estimation and selection is investigated with a particular focus on sampling efficiency of the different MCMC implementations.

Dynamic Spatial Modelling in Inhomogeneous Force Fields

Ioanna Manolopoulou Thomas Kepler and Mike West
Duke University, USA
im30@stat.duke.edu kepler@duke.edu mw@stat.duke.edu

We present dynamic spatial modelling and computational methods for the analysis of collections of objects moving in an inhomogeneous force field. Core motivating examples come from single-cell systems biology, where multiple cells move in 3-D environments "driven" by underlying physio-chemical force fields. Models are based on underlying dynamic, state-space models for locations and directional velocities of each of a set of objects, combined with a latent force-field over the spatial domain that "drives" changes in velocities. Discrete time models generated from underlying diffusion processes are key. Locations are measured with noise over time, while velocities are latent. We develop novel models of the force fields using a Bayesian radial basis function approach to non-parametric regression. This

uses linear combinations of Gaussian kernels over the spatial domain to define a potential surface, with the force field given by the gradient of the potential in 3-D. This allows us to represent heterogeneity in the spatial structure of the force field in a flexible, adaptive Bayesian non-parametric regression framework, and define effective MCMC methods for model fitting. We exemplify the work with analysis of single cell data from immunofluorescence histology, using a movie from stained immune cells emerging into a lymph node follicle. Interest lies in characterizing the dynamic behaviour of the cell migration, which is key to understanding the mechanisms of the adaptive immune response.

Bayesian Clustering Based Multiple Hypotheses Testing

Juan-Miguel Marín and Teresa Rodríguez-Bernal
Universidad Carlos III de Madrid, Spain
jmmarin@est-econ.uc3m.es mayter@mat.ucm.es

Multiple hypothesis testing techniques are recently of great interest due to the application in massive data analysis related with microarray Data Analysis. Moreover, clustering techniques are among the most used procedures in this field. In this work we combine both methods, in the context of microarray data, in order to consider at the same time comparative studies of genes expression data undertaken with multiple conditions and different replicates. One advantage of this approach is that we can consider, in each hypothesis test, more than two groups comparisons.

Variable Selection with Gibbs Samplers and Zellner-Siow Priors

Miguel-Angel Martínez-Beneito
Centro de Investigaciones en Salud Pública, Valencia, Spain
martinez_mig@gva.es

Gonzalo García-Donato
Universidad de Castilla-La Mancha, Spain
gonzalo.garciadonato@uclm.es

We consider the problem of variable selection in normal linear models with a moderate to large number of regressors. We focus on developing ideas to produce Bayes factors, being easy to implement in existing Gibbs sampler-based software. The methodology produces a sample of models that, ideally, should have high integrated likelihood values. The priors used is an essential part of any testing scenario. In our setting, we impose to each single entertained model to follow the proposals of Zellner and Siow. That is, a constant prior is used for the common parameters while the new parameters are assigned a Cauchy distribution. We compare our methodology with other proposals for searching in large model spaces in real and in simulated data sets.

A Computationally Convenient Framework For Multivariate Disease Mapping

Miguel-Angel Martínez-Beneito
Centro de Investigaciones en Salud Pública, Valencia, Spain
martinez_mig@gva.es

Dongchu Sun
University of Missouri, USAi
sund@missouri.edu

Multivariate Disease Mapping has been an active research field during the last years. These models propose to share information among different diseases, and not only among neighboring places, in order to obtain reliable risk estimates able to depict the geographical pattern of a collection of diseases. Several models have been proposed in order to achieve this goal, most of them structuring dependence between locations and diseases by means of Kronecker products of covariance matrices. In these models specific matrix algebra routines, particularly fitted to deal with Kronecker products, are used in order to make the inference process affordable.

In this work we propose a computationally convenient new approach to Multivariate Disease Mapping. Slight modifications of this new model have been shown to reproduce most of the proposals appeared at the literature. Nevertheless, our approach extends its computational benefits to all these models and enclose all them in a unique framework. Moreover our approach can be easily generalized to reproduce covariance structures beyond the Kronecker product of matrices. Last, inference for our model can be done in WinBUGS, therefore our approach yields a way to make this tool feasible for a wide collection of Multivariate Disease Mapping models.

Model introduction, exploration of theoretical details, implementation and applications on real datasets will be shown at this work.

Two Adaptive Rejection Sampling Methods for Sampling a Posteriori Probability Distributions

Luca Martino and Joaquín Míguez
Universidad Carlos III, Spain
luca@tsc.uc3m.es joaquin.miguez@uc3m.es

Monte Carlo methods are often necessary for the implementation of optimal Bayesian estimators. A fundamental technique that can be used to generate samples from virtually any target probability distribution is the so-called *rejection sampling* method, which generates candidate samples from a proposal distribution and then accepts them or not by testing the ratio of the target and proposal densities. The class of *adaptive* rejection sampling (ARS) algorithms (W. R. Gilks and P. Wild, *Applied Statistics*, 41, 2) is particularly interesting because they can achieve high acceptance rates. However, the standard ARS method can only be used with log-concave target densities.

Consider the problem of drawing samples from the posterior density $p(x|\mathbf{y})$, where X is the random variable of interest and $\mathbf{y}$ is a vector of observed data.

We investigate two different ARS schemes that can be used to draw exactly from a large family of probability density functions (pdf's), not necessarily log-concave. To be specific, the first method is based on the adaptive construction a piecewise-constant approximation $L_t(x)$ of the likelihood function $p(\mathbf{y}|x)$. The stepwise function $L_t(x)$ overbounds the likelihood, i.e., $L_t(x) \geq p(\mathbf{y}|x)\ \forall x$. Therefore, if $p(x)$ denotes the prior of X, then $L_t(x)p(x) \geq p(x|\mathbf{y})$ and the function $\pi(x) \propto L_t(x)p(x)$ is a suitable proposal density for rejection sampling. In order to use this method, one must be able to integrate the prior pdf $p(x)$ in finite intervals.

The second approach, is based on the *ratio of uniforms* (RoU) technique (L. Devroye, *Non-Uniform Random Variate Generation*, Springer 1986). The RoU method enables us to obtain a 2-dimensional region $\mathcal{A}$, in many cases bounded, such that drawing from $p(x|\mathbf{y})$ is equivalent to drawing *uniformly* from $\mathcal{A}$. We will describe an adaptive technique to obtain a collection of 2-dimensional triangular sets $\{\mathcal{T}_k\}_{k=1,\ldots,K}$, such that $\mathcal{A} \subseteq \cup_{k=1}^{K}\mathcal{T}_k$ and $\lim_{K\to\infty} \cup_{k=1}^{K}\mathcal{T}_k = \mathcal{A}$. Drawing uniformly from the sets $\mathcal{T}_k$ is straightforward and, as a consequence, we can efficiently draw from $\mathcal{A}$ by rejection sampling.

Both techniques are adaptive in the sense that every time a candidate sample is rejected, the approximation (either $L_t(x)$ or $\cup_k \mathcal{T}_k$) is improved. An illustrative numerical example will be shown.

Bayesian Inference Analysis for the Models in the Application Competing Risk

Camila B. Martins
Universidade de São Paulo, Brazil
cacamila_bertini@yahoo.com.br

Vera L. Tomazella and Adriano Polpo
Universidade Federal de São Carlos, Brazil
vera@ufscar.br polpo@ufscar.br

There are situations where various risk factors of failure are present, in the same time, in the life of system. For this reason we say that these factors are competing to cause the system failure. However, only one of these competitors is responsible for the system failure. The failure behavior of one system is, in most times, represented for its failure rate, which may be increasing, decreasing, remain constant or be combinations of these over time. Therefore it is desirable to use a probabilistic model that only with changes in the values of the parameters representing each of these situations. In this work we studied from the perspective of Bayesian reference analysis to the application of competitive risks under the Weibull model due to high flexibility of this model. The reference analysis is a method to produce Bayesian inferential statements which only depend on the assumed model and the available data (Bernardo, 1979). The goal is to find a specific joint reference prior function for all the unknown parameters of Weibull model in the application of competitive risks and a marginal reference posterior to the parameters of interest, which is always dominated by the observed data. The reference posterior distributions are obtained through the use of the Bayes theorem with the reference prior function that can be used to point estimates and

tests of hypotheses, providing a unified set of Bayesian objective solutions for our problem.

Bayesian Multi-Locus Model for Gene-Gene Interactions

Pekka Marttinen and Jukka Corander
Aalto-Universitetets, Finland and *Helsingin Yliopisto Finland*
pekka.marttinen@hut.fi jukka.corander@helsinki.fi

Since the introduction of large-scale genotyping methods that can be utilized in genome-wide association (GWA) studies for deciphering complex diseases, statistical genetics has been posed with a tremendous challenge of how to most appropriately analyze such data. Advanced model-based methods for genetic mapping of traits have been available for more than 10 years in animal and plant breeding. However, due to computational burden, GWA analyses have in practice been dominated by simple statistical tests concerned with a single marker locus at a time.

We introduce a novel Bayesian model-averaging method for association mapping which enables the detection of multiple loci and their interactions that influence a dichotomous phenotype of interest. The method is shown to perform consistently well in a simulation study when compared to widely used standard alternatives and its computational complexity is typically considerably smaller than that of a maximum likelihood based approach for testing gene-gene interactions.

Our results show that a full Bayesian multi-locus approach helps to remove some false positives and provides a better localization of the causal loci; however, our findings also illustrate that such an approach is somewhat sensitive to the prior distribution on the model structure.

Nonparametric Bayesian Density Estimation Using Polynomial Chaos Expansions

Youssef Marzouk
MIT Boston, USA
ymarz@mit.edu

Paul Boggs
Sandia National Laboratories, USA
ptboggs@sandia.gov

Polynomial chaos expansions are orthogonal polynomials of i.i.d. random variables. They provide a general and flexible *functional* representation of square-integrable random variables and processes. These expansions date back to Wiener, but have more recently emerged as a useful tool for the numerical solution of stochastic ordinary and partial differential equations—rooted not only in stochastic analysis, but in classical multivariate approximation theory.

We formulate a hierarchical Bayesian method for estimating polynomial chaos expansions from limited and noisy data. Likelihood evaluations rely on fast identification of polynomial roots via a Sturm-sequence polynomial solver. We introduce a reversible-jump Markov chain Monte Carlo scheme to simultaneously estimate the polynomial degree and the stochastic dimension of the expansion, along with the corresponding coefficients. This construction allows an infinitely

parametric representation of the underlying random quantities. Results include examples drawn from systems reliability; we also discuss extensions to multivariate problems.

Strategy for Modelling Non-Random Missing Data Mechanisms in Longitudinal Studies: Application to Income Data from the Millennium Cohort Study

Alexina J. Mason
Imperial College London, UK
a.mason05@imperial.ac.uk

Nicky Best
Imperial College London, UK
n.best@imperial.ac.uk

Sylvia Richardson
Imperial College London, UK

Ian Plewis
University of Manchester, UK

Longitudinal studies typically lose members over time and generally suffer from missing data. Bayesian full probability modelling provides a flexible approach for analysing such data, allowing a plausible model to be built which can then be adapted to carry out a range of sensitivity analyses. In this context, we propose a strategy for using Bayesian methods to carry out a 'statistically principled' investigation of data which contains missing covariates and missing responses, where we suspect that the latter are generated by an informative missing data mechanism.

The first part of this strategy entails constructing a 'base model' consisting of (i) a model of interest, (ii) a sub-model to impute the missing covariates, (iii) a sub-model to allow informative missingness in the response.

The second part involves running a series of sensitivity analyses to check the robustness of the conclusions from the base model, categorised as 'assumption sensitivities' and 'parameter sensitivities'. For the 'assumption sensitivities', we suggest running alternative models with key assumptions altered, including the model of interest error distribution, the model of interest response transform and the functional form of the response missingness model. By contrast, for the 'parameter sensitivities', the base model is rerun with the parameters associated with the informative missingness fixed to a range of plausible values. We implement our strategy, using the WinBUGS software, to investigate some typical research questions relating to the prediction of income, using data from the Millennium Cohort Study. A range of tables and plots that we found useful in the sensitivity analyses are presented.

OBANSoft, a Bayesian Software for Objective Bayesian Analysis: Its First Days

Asunción M. Mayoral M. Quesada
Universidad Miguel Hernández de Elche, Spain
asun.mayoral@umh.es quesada.mqm@gmail.com

Javier Morales and Xavier Barber
Universidad Miguel Hernández de Elche, Spain
j.morales@umh.es xbarber@umh.es

A new software named OBANSoft has been designed and implemented in order to provide students with a flexible tool to understand and experiment with basic concepts on Bayesian Statistics. Both perspectives, subjective and objective, are possible. From a complete data description (numeric and graphical), posterior and predictive analyses can be achieved, by taking into account any or none prior information, through conjugate or real non-informative prior distributions. This software has been programmed under Java, with R as the calculus engine. A really kindly interface facilitates its use. By now just the simplest models are programmed, but, OBANSoftís architecture facilitates integration of successive models, progressively more and more complex. The Catalog of Noninformative Priors by Yang and Berger (1996) is the base for non-informative models. When no references are provided for simulation, different algorithms will be tested, also making use of parallel programming, in order to improve efficiency in convergence for the MCMC chains. A demo presentation will be showed at a poster session. A distributable free version of OBANSoft will be provided to all interested participants through our website.

A Bayesian Factor Model for Gene Expression Detection on Oligonucleotide Microarrays

Vinicius D. Mayrink and Joseph E. Lucas
Duke University, USA
vinicius@stat.duke.edu joe@stat.duke.edu

A problem of scientific interest is to determine whether a gene is expressed in a probe set of a microarray. Different detection procedures have been proposed for different microarray platforms. A popular method for Affymetrix oligonucleotide arrays is part of the preprocessing algorithm MAS5. In short, presence-absence calls are assigned based on p-values obtained via one-sided Wilcoxon's signed rank test. This solution works relatively well in practice; however, other improved frameworks can be found. An example is the method Presence Absence calls with Negative Probe sets (PANP). Selected negative strand matching probe sets are assumed as controls and the empirical c.d.f. of their intensities is used to derive a cutoff value. Its simplicity and improved results make PANP an attractive method; however, defining thresholds based on an empirical c.d.f. seems an arbitrary decision. When several microarrays replicating the same experiment for different samples are available, th e pattern of expressions across samples is an important source of information for gene detection. All probes randomly alternate intensities across samples in a noise probe set, whereas, a consistent expression sequence is observed for every probe in a non-noise probe set. This result motivates a Bayesian Factor Model (BFM). The factor loadings reflect the strength of hybridization between the target and each probe. The factor scores describe the pattern of expression across samples. A mixture prior is attributed to the factor loadings vector; two Normal components are used, and one of them

is centered on zero with small variance. The detection call is based on the posterior probability that the loadings differ from zero. Methods are compared using simulated and real data sets, and the results indicate that BFM outperforms other procedures for detecting transcripts. The amount of data in microarrays represents a challenge for MCMC; therefore, strategies are considered for imp roving the computational performance.

Probabilistic Programming with Imperatively-Defined Factor Graphs

Andrew McCallum
University of Massachusetts Amherst, USA
mccallum@cs.umass.edu

Practitioners in computer vision, natural language processing, bioinformatics and other areas have achieved great empirical success using graphical models with repeated, relational structure. But as researchers explore increasingly complex structures, there has been increasing interest in new programming languages or toolkits that make it easier to implement such models in a flexible, scalable way. A key issue in these toolkits is how to define the templates of these repeated structure and tied parameters. Rather than using a declarative language, such as SQL or first-order logic, we advocate using an imperative language to express various aspects of model structure, inference, and learning. By combining the traditional, declarative statistical semantics of factor graphs with imperative definitions of their construction and operation, we allow the user to mix declarative and procedural domain knowledge, and also gain significant efficiencies. We have implemented such imperatively defined factor graphs in a system we call FACTORIE, a software library for an object-oriented, strongly-typed, functional language called Scala. I will introduce this probabilistic programming language and give several examples of its use.

Sequential Versus Full-Bayesian Regression Adjustment for the Propensity Score.

Lawrence C. McCandless
Simon Fraser University, Canada
mccandless@sfu.ca

Cutting feedback is a new idea in Bayesian computation for fitting complex models, and it has been implemented in BUGS. The idea is to estimate different model components separately rather than simultaneously, and this is achieved during posterior simulation by not updating from full conditional distributions. The advantage is that it prevents contamination between data sources in the face of model misspecification. We study the idea of cutting feedback in the context of Bayesian regression adjustment for the propensity score in epidemiology. We illustrate results using a data example from pharmacoepidemiology.

Latent Structure Models for Social Networks using Aggregated Relational Data

Tyler H. McCormick and Tian Zheng
Columbia University, USA
tyler@stat.columbia.edu tzheng@stat.columbia.edu

Social networks have become an increasingly common framework for understanding and explaining social phenomena. But despite an abundance of sophisticated models, social network research has yet to realize its full potential, in part because of the difficulty of collecting social network data. In contrast, Aggregated Relational Data, commonly collected as questions of the form "How many X's do you know?", measure network relationships indirectly and are easily incorporated into standard surveys. We propose a latent space model where the propensity of an individual to know members of a given alter group (people named Michael, for example) is independent given the positions of the individual and the group in a latent "social space." This framework is similar in spirit to previous latent space models proposed for networks (Hoff, Raftery and Handcock (2002), for example) but doesn't require that the entire network be observed. Using this framework, we derive evidence of social struc- ture in personal acquaintance networks, estimate homogeneity of groups, and estimate individual and population gregariousness. Our method makes information about more complicated network structure available to the multitude of researchers who cannot practically or fnancially collect data from the entire network.

Bayesian Inference in Augmented Designs for Assessing Principal Strata Causal Effects

Alessandra Mattei and Fabrizia Mealli
Università degli Studi di Firenze, Italy
mattei@ds.unifi.it mealli@ds.unifi.it

We introduce new methods to help identification and estimation of causal mechanisms (so-called direct and indirect effects) in the presence of confounded intermediate variables. In this setting the usually made assumption that the intermediate variables can be controlled and fixed by external intervention might be questionable. We deal with this confounding problem under the principal stratification framework (Frangakis and Rubin 2002, Biometrics 58(1), 21-29), focussing on Principal Causal Effects (PCEs). PCE analysis is challenging, due to the latent nature of principal strata. In order to ease identification and estimation of PCEs, we will investigate new augmented designs, where the treatment is randomized, and the mediating variable is not forced, but only randomly encouraged. The design will be feasible in some clinical and social experiments, when partial control of the intermediate variable can be conceived.

We will adopt a Bayesian approach for inference, which allows us to achieve valid estimates of quantities of interest and also properly account for our uncertainty about these quantities. This approach has the advantage of (a) gathering more information from the data, which may be useful for policy purposes; and

(b) exploiting the constraints of the implicit conditional distributions that are compared. In addition, crucial structural (behavioral) assumptions can be distinguished from functional assumptions and clearly discussed. Our framework permits clear assessment of assumptions and evaluation of their consequences, by means of sensitivity analysis, by using posterior predictive checks, and investigating the posterior distribution of weakly identified models.

Particle-Filter based Maximum Likelihood Ratio Test for Stochastic Volatility Models with leverage

Amaan Mehrabian
McGill University, Canada
mehrabian@math.mcgill.ca
David Stephens
McGill University, Canada
dstephens@math.mcgill.ca

We apply Sequential Monte-Carlo (SMC) methods for inference on Stochastic Volatility (SV) models with a leverage parameter. The leverage effect is the asymmetric phenomenon where there is more increase in the variance of a financial asset in price drops compared to price rises. We perform formal statistical tests to examine the presence of the leverage effect in various high- and medium-frequency financial time series.

Generally SV models can be represented as non-linear non-Gaussian state-space models where the asset prices and volatility variables form the observed and latent state variables, respectively. The likelihood for such models is often not available in closed form. Therefore, we use particle filter estimates of the likelihood to perform Maximum Likelihood Ratio (MLR) tests. We show, through extensive numerical simulations, that the distribution of the test statistic obtained by particle filter estimates converges to its true distribution when the number of data points and particles increases. We then apply this method to different financial time series data such as stock indices and foreign exchange (FX) rates.

In addition, we investigate the role that realized volatility can play in model assessment, and incorporate realized volatility into the volatility forecasts.

Some Thoughts on the Bayesian Robustness of Location-Scale Models

Manuel Mendoza
ITAM, Mexico
mendoza@itam.mx

Eduardo Gutiérrez-Peña
IIMAS-UNAM, Mexico
eduardo@stats.iimas.unam.mx

We review a number of results, widely discussed in the literature, concerning the Bayesian robustness of location-scale models. We underline some specific aspects which, in our opinion, deserve more attention, and illustrate our remarks by means of a simple case study.

What's the H in H-likelihood: A Holy Grail or An Achilles' Heel?

Xiao-Li Meng
Harvard University, USA
meng@stat.harvard.edu

"H-likelihood" refers to a likelihood function of both fixed parameters and random "unobservables," such as missing data and latent variables. H-likelihood inference is typically performed by maximizing over the unobservables via an adjusted profile H-likelihood, and carrying out a Fisher-information-like calculation for (predictive) variance estimation. The claimed advantage is its avoidance of all "bad" elements of Bayesian prediction, namely the need for prior specification and posterior integration. This talk attempts to provide an in-depth look into one of the most intriguing mysteries of modern statistics: why have the proponents of the H-likelihood method (Lee and Nelder, 1996, 2001, 2005, 2009) been so convinced of its merits when almost everyone else considers it invalid as a general method? The findings are somewhat intriguing themselves. On the one hand, H-likelihood turns out to be Bartlizable under easily verifiable conditions on the marginal distribution of the unobservables, and such conditions point to a transformation of unobservables that makes it possible to interpret one predictive distribution of the unobservables from three perspectives: Bayesian, Fiducial and Frequentist. On the other hand, the hope for such a Holy Grail in general is diminished by the fact that the log-H-likelihood surface cannot generally be summarized quadratically due to the lack of accumulation of information for unobservables, which seems to be the Achilles' Heel of the H-likelihood method.

Sequential Clustering and Anomaly Detection for Streaming Data

Daniel Merl
Lawrence Livermore National Laboratory, USA
dmerl@llnl.gov

The development and deployment of analytic tools for real-time surveillance of streaming data for purposes of characterizing normal behavior and identifying

annomolous behavior is a very challenging, open research problem. Increasingly, the trend has been to move away from signature based approaches, in which potentially interesting cases are identified by pattern matching using explicit descriptions of known threats. In computer security, for example, virus protection software, such as that produced by McAfee, employs this approach. The main criticism of signature-based methods is that they are inherently retrospective, and therefore unable to actively identify new behaviors.

A promising strategy for very flexible anomaly detection involves developing predictive models of observed behaviors in order to provide data-driven, quantitative measures of the degree of anomaly of newly observed behaviors. Having defined a set of features believed to be sufficient for distinguishing between normal and anomalous activity, statistical methods can then be used to uncover complex dependencies and correlations between observed features, and to identify statistically significant deviations from these predicted usage patterns.

In this poster, I will describe the application of sequentially learned Dirichlet process mixture models to several streaming data problems. Through the use of parallelized sequential inference techniques such as particle learning, it is possible to fit the DPM to incoming feature vectors in real time, thus enabling online clustering and anomaly detection via the estimated mixture model.

Adaptive Metropolis-Hastings-Within-Gibbs Algorithms using Copulae

Renate Meyer
University of Auckland, NZ
meyer@stat.auckland.ac.nz

Bo Cai
University of South Carolina, USA
bocai@mailbox.sc.edu

Different strategies have been proposed to improve mixing and convergence properties of Markov Chain Monte Carlo algorithms. These are mainly concerned with customizing the proposal density in the Metropolis-Hastings algorithm to the specific target density and require a detailed exploratory analysis of the stationary distribution and/or some preliminary experiments to determine an efficient proposal. Various Metropolis-Hastings algorithms have been suggested that make use of previously sampled states in defining an adaptive proposal density. Here we propose a general class of adaptive Metropolis-Hastings algorithms based on Metropolis-Hastings-within-Gibbs sampling. The proposal density is adapted by fitting elliptical copulae to sets of previously sampled points. Marginal distributions are estimated using local mixtures of triangular distributions as described in Cai *et al.* (2008) (B. Cai, B., R. Meyer, F. Perron, Metropolis-Hastings Algorithms with Adaptive Proposals, Statistics and Computing 18, 421-433, 2008). Using various different examples, we demonstrate the properties and efficiencies of these algorithms.

A Bayesian Spatial Model for Panel Time Series Data

Helio S. Migon Esther Salazar and L. C. Alves
Universidade Federal do Rio de Janeiro, Brazil
migon@im.ufrj.br

The focus of this poster will be on the development of spatio-temporal econometrics models for panel data, whose elements correspond to economic agents. The spatial dependence between agents' will be take into account via an exogenous economic distance, which will be incorporated both in the mean and variance structure of the model. In order to accomodate the presence of occasional outliers a t-Student regression model is also introduced. The sensitivity to alternative hyperparameters prior specifications will be evaluated. An illustrative application, using our proposed models, is presented to show how the joint movements in output growth across Brazilian industrial sectors depends on the similarity of sectors' technologies.

Solving a Class of Global Optimization Problems by Way of Bayesian Estimation Methods

Joaquín Míguez, Dan Crisan
Universidad Carlos III, Spain *Imperial College London, UK*
joaquin.miguez@uc3m.es d.crisan@imperial.ac.uk

Petar M. Djurić
Stony Brook University, USA
jdjuric@ece.sunysb.edu

The interplay between the fields of inference and optimization has often provided new insights and techniques in both areas. In this talk, we investigate the transformation of a class of optimization problems into equivalent maximum *a posteriori* (MAP) estimation problems. In particular, we aim at $\min_{x_{0:T}} C_T(x_{0:T})$, where C_T is a cost function with a high-dimensional argument $x_{0:T} = \{x_0, x_1, \ldots, x_T\} \in (\mathbb{R}^d)^{T+1}$. Assuming that C_t can be recursively decomposed into a series of costs $C_T(x_{0:T})$, $C_{T-1}(x_{0:T-1})$, ..., $C_0(x_0)$ with lower-dimensional supports, we show how to design dynamic state-space random systems such that any MAP estimate of the system state up to time T coincides with a global minimum of the cost C_T.

To be specific, let $\pi(x_0)$ denote a prior probability density function (pdf) of choice for x_0, let $\pi(x_t|x_{0:t-1})$ denote a conditional density of x_t given $x_{0:t-1}$, let $y_{1:T} = \{y_1, y_2, \ldots, y_T\}$ be known parameters of the cost function C_T (that play the role of observations) and let $\pi(y_t|x_{0:t}, y_{1:t-1})$ be a conditional pdf that yields the likelihood of $x_{0:t}$ given $y_{1:t}$. We show how to design state-space systems, described by $\pi(x_0)$, $\pi(x_t|x_{0:t-1})$ and $\pi(y_t|x_{0:t}, y_{1:t-1})$, $t = 1, \ldots, T$, that are *matched* to the cost function, so that $\hat{x}_{0:T} \in \arg\max_{x_{0:T} \in (\mathbb{R}^d)^{T+1}} \pi(x_{0:T}|y_{1:T})$ if, and only if, $\hat{x}_{0:T} \in \arg\min_{x_{0:T} \in (\mathbb{R}^d)^{T+1}} C_T(x_{0:T})$, where $\pi(x_{0:T}|y_{1:T})$ is the posterior pdf of $x_{0:T}$. Hence computing the Bayesian MAP estimator $\hat{x}_{0:T}$ is equivalent to solving the global optimization problem.

After this reformulation, we can apply a sequential Monte Carlo method to build a random grid approximation of the space of the unknowns that is dense

where the posterior probability mass is high and sparse elsewhere. A simple search over this random grid yields MAP estimates. We consider two specific algorithms that combine the standard particle filter with two search procedures and prove that their outputs converge almost surely to a global maximum of the posterior pdf $\pi(x_{0:T}|y_{1:T})$. Furthermore, we obtain an expression for the number of samples needed to attain a prescribed accuracy. Finally, we present an illustrative example in audio signal processing.

Non-Exchangeable Bayesian Nonparametric Latent Feature Models

Kurt T. Miller
University of California at Berkeley, USA
tadayuki@cs.berkeley.edu

Michael I. Jordan
University of California at Berkeley, USA
jordan@cs.berkeley.edu

Thomas L. Griffiths
University of California at Berkeley, USA
tom_griffiths@berkeley.edu

Bayesian nonparametric methods are based on prior distributions expressed as general stochastic processes. Due to the need to integrate over these priors at inference time, strong constraints such as exchangeability are often placed on the kinds of models that can be considered. Recently, two infinitely exchangeable nonparametric priors for latent feature models have been introduced–the Indian Buffet Process for which the De Finetti mixing distribution is Hjort's beta process, and the Infinite Gamma-Poisson Feature Model for which the De Finetti mixing distribution is the gamma process. These are priors over infinite binary matrices and infinite non-negative integer valued matrices, respectively, that allow us to perform nonparametric latent feature inference. Following on these developments, we aim to extend the range of Bayesian nonparametric latent feature modeling by presenting two non-exchangeable generalizations for each of these two models in which posterior inference is nearly as computationally efficient as in their exchangeable counterparts. Rather than having an exchangeable model in which the features for each of the objects are conditionally independent given the De Finetti mixing distribution, we draw the features for all objects jointly using a stochastic process that allows us to utilize prior knowledge about objects to infer better features. Our models are applicable to the general settings in which the dependencies between objects can be captured using either a rooted tree expressing how closely related objects are or a chain which expresses a linear relationship amongst the objects. We demonstrate the effectiveness of these models with an application to choice data.

Bayesian Model Selection for 2-Way ANOVA Models

Xiaoyi Min
University of Missouri, USA
xmin@mizzou.edu

Dongchu Sun
University of Missouri, USA
sund@missouri.edu

Two-way ANOVA models are considered under a class of Zellner's (1986) g-priors, and closed form expressions of Bayes factors among possible sub-models are derived. Under the inverse-gamma priors for g's, the exact convergence rates of Bayes factors under the corresponding models are calculated using some Laplace approximation. The consistency of Bayes factors is then considered under the situations when the number of replicates under each level is large and when the number of treatment levels is large.

Zero Variance Simulation: From Monte Carlo to Markov Chain Monte Carlo Simulation

Antonietta Mira
Univ. dell'Insubria, Italy and *Univ. della Svizzera Italiana, CH*
antonietta.mira@uninsubria.it and antonietta.mira@usi.ch

Reza Solgi
Swiss Finance Institute at the Università della Svizzera Italiana, Switzerland
reza.solgi@usi.ch

Variance reduction techniques are essential to Monte Carlo simulations; they can achieve the same level of precision in a shorter time and using a lower budget, or obtain more precise answer in a shorter time and at lower expenses. Several variance reduction techniques have been proposed since the beginning of popularity of Monte Carlo simulations (Ripley, 1987), and recently in the physics literature, Assaraf and Caffarel have proposed a novel approach to build a control variate that can reduce the variance of a Monte Carlo estimator extremely (Assaraf R. and Caffarel M., Zero-Variance principle for Monte Carlo algorithms, *Phys. Rev. Letters*, 1999). In this research we "translate" this technique from the language of physics to the language of statistics (and therefore make it available to more statisticians), and adapt the methodology to Markov chain Monte Carlo simulation. We also employ this technique in estimation of parameters of several statistical models (GLM, GARCH, mixture models, etc.) in a Bayesian setting. In these examples the proposed approach can reduce the variance of the estimators from hundreds to tens of thousands of times.

A Bayesian Estimation of the Residential Gas Demand on the Nonconvex Budget Set

Koji Miyawaki
National Institute for Environmental Studies, Japan
miyawaki.koji@nies.go.jp

Yasuhiro Omori
University of Tokyo, Japan
omori@e.u-tokyo.ac.jp

Akira Hibiki
Nat. Inst. Env. Studies, Japan
hibiki@nies.go.jp

The residential gas is often supplied under a decreasing block rate pricing, where the unit price discontinuously declines as the consumption amount is greater than the predetermined thresholds. A microeconomic theory shows that, under such a price schedule, a consumer's budget set is nonconvex.

The consumer's utility maximization problem on the nonconvex budget set derives the corresponding demand function involving the comparison of nonlinear indirect utilities. Previous literature focused on two-block decreasing block rate pricings, where there are two different unit prices, derived the demand function based on the utility maximization problem, and numerically solved the comparison of the indirect utilities to estimate the demand function. Because the residential gas is often supplied under multiple-block decreasing block rate pricings (three to six blocks in Japan), this article resolves the problem caused by the comparison of the indirect utilities by using the properties of convex functions (the Hermite-Hadamard integral inequality and the generalized mean inequality), and develops a Bayesian estimation method of the residential gas demand on the nonconvex budget set, where the number of blocks is more than two.

The proposed method is examined by a numerical simulation and is adopted to estimate the residential gas demand function in Japan.

Bayesian Inference for a Partial Order from Random Linear Extensions

Ometse S. Mogapi
University of Botswana
simimoga74@yahoo.co.uk

Tom A.B. Snijders
University of Oxford,UK
Tom.Snijders@nuffield.ox.ac.uk

Geoff K. Nicholls
University of Oxford, UK
nicholls@stats.ox.ac.uk

Bernard W. Silverman
University of Oxford, UK
bernard.silverman@stats.ox.ac.uk

David Johnson
University of Oxford, UK
david.johnson@spc.ox.ac.uk

Nicholas E. Karn
University of Southampton, UK
n.e.karn@soton.ac.uk

Witness lists are ordered lists of the signatories to historical legal documents called acta. We have a large collection of royal acta from twelfth century England. Witnesses generally signed in order of importance. However, it is not clear what order relations existed between the bishops who appear in the lists.

We represent the unknown true order relation in the data as a partial order, that is, as a transitively closed directed acyclic graph. Any particular witness list is modeled as a random linear extension of a suborder of the unknown order, and this determines the likelihood for a partial order. There is related work (by Berrenwinkel, Eriksson and Sturmfels in 2007 and Berrenwinkel and Sullivant in 2009) on maximum likelihood partial orders for conjunctive Bayesian networks. However, we know of no Bayesian inferential framework for partial orders.

We describe a range of prior distributions for partial orders, including priors with hierarchical structures designed to reflect likely social order relations. One class of models, related to k-dimensional random orders, extends latent-space model for social networks in a natural way. The k-dimensional random orders are marginally consistent for subsetting. We develop this class of distributions to incorporate prior knowledge.

We fit these models to the data using MCMC simulation of the posterior distribution of partial orders given observed linear extensions. Correct MCMC simulation of partial orders is not straightforward, because they are transitively closed.

Sparse Exponential Family Latent Variable Models

Shakir Mohamed
University of Cambridge, UK
sm694@eng.cam.ac.uk

Katherine Heller
University of Cambridge, UK
kah60@eng.cam.ac.uk

Zoubin Ghahramani
University of Cambridge, UK
zoubin@eng.cam.ac.uk

Recent efforts in high dimensional data modelling have highlighted the need for models that encode sparsity. In particular, latent factor models with sparsity have become relevant in situations where we believe that there are a number of underlying factors but that only a few factors are active and contribute to explaining any particular data point. One clear example where a sparse representation is applicable is in gene expression modelling. A gene's expression is influenced by the presence of a number of transcription factor proteins, and there exists a wide array of such transcription factors that may affect the expression of any set of genes. Here, the underlying biology is sparse, since an individual gene's activity may only be directly influenced by a small subset of the known transcription factors.

We consider sparsity in the framework of generalised latent variable models. These models are based on an exponential family likelihood and are generalised models in a manner analogous to GLMs for regression. We study these models using sparsity-favouring priors for the latent variables, exploring in depth continuous priors such as the Laplace or Exponential distribution, and "spike and slab" priors with a delta mass at zero. Continuous sparsity favouring priors allow for sparse learning but place no mass on zero itself and thus samples are never exactly zero. Spike and slab priors have the desirable property that posterior samples contain zeroes, but have thus far been relatively unexplored in the unsupervised setting.

We consider Bayesian models using a sampling approach to inference with a Laplace, Exponential and spike and slab prior. We compare these models to a commonly used optimisation approach based on L_1 norm regularisation. Experimental results show that the "spike and slab" model has the best performance on data reconstruction as measured using the predictive probability on held-out data as well as the root mean squared error. Evaluations are shown on both synthetic data generated from the model as well as three real world data sets consisting of human judgements of animal features, robot planning and SPECT images.

In compressed sensing and related areas, the idealised, but intractable optimisation criterion uses an L_0 norm to penalise the number of non-zero parameters. The spike and slab model can also be seen as placing a penalty on the non-zero parameters, and thus can be seen to enforce sparsity in a manner similar to an L_0 norm minimisation. The spike and slab model approach has the property that it is able to introduce sparsity, while not enforcing shrinkage on parameters where no shrinkage is necessary, as is the case with models which use the Laplace prior for example. This allows for accurate data reconstruction, while learning the appropriate sparsity pattern supported by the data. The results are extremely encouraging and suggests a much wider scope for the use of spike and slab models in Bayesian unsupervised learning settings.

Bayesian Spatial Prediction for Closed Skew Gaussian Random Field

Mohsen Mohammadzadeh
Tarbiat Modares University, Iran
mohsen_m@modares.ac.ir

Omid Karimi
Tarbiat Modares University, Iran

Most work in statistical spatial prediction assume the spatial data as realizations of a Gaussian random field. However, this assumption is hard to justify for a large number of applications. Sometimes, the distribution of data is skewed, closed skew normal (CSN) distribution can be used for modeling their skewness. CSN distribution is an extension of the multivariate skew normal distribution and has the advantage of being closed under marginalization and conditioning. Here, we generalize Bayesian prediction methods to closed skew Gaussian random variable to provide a Bayesian spatial predictor. A simulation study is performed to check validity of the proposed model. We compare the CSN model to Gaussian model to show improvement of mean-square error in spatial prediction by the CSN model. Also, the CSN model is applied to Bayesian spatial prediction on two real examples in Iran. The skewness of the data is captured by the CSN model and Mean-square error of cross-validation is improved by this model on the two examples.

Modelling the Hemodynamic Response in fMRI Using Gaussian Processes

Raquel Montes Diez and Alicia Quirós
Universidad Rey Juan Carlos, Spain
raquel.montes@urjc.es alicia.quiros@urjc.es

During the last few decades our knowledge of the human brain has developed significantly as a result of new neuroimaging techniques, such as functional magnetic resonance imaging (fMRI). By observing the relation between a stimulus paradigm (in an experiment) and the changes in blood flow and blood oxygenation in the brain (known as hemodynamics), fMRI provides a measure of brain activation. The change in the ratio of oxygenated to deoxygenated blood is described by the so-called hemodynamic response function (HRF).

Modeling the HRF in fMRI experiments is therefore an important aspect of the analysis of data in functional neuroimaging. This has been done in the past using parametric response functions, typically including the Poisson, gamma or Gaussian densities.

In this work, we consider the case in which the HRF is simply defined by a certain unknown function $z(\odot)$. General Gaussian Processes theory presents an attractive way of expressing prior beliefs about the function $z(\odot)$ and we show how, in this context, a combination of analytical methods may be used for making inference about the posterior predictive distribution of interest.

The authors present a comparison of their model with standard hemodynamic response kernels on simulated data, showing the ability of the former to better accommodate the shape of the HRF and its greater flexibility when compared with parametric approaches. Additionally, the authors perform a full analysis of data acquired during a real experiment.

Predicting Failures Due to Spaceborne Radiation: A Comparison of Parametric and Semiparametric Bayesian Isotonic Regression

Robin D. Morris
Research Institute for Advanced Computer Science, USA
rdm@riacs.edu

Marian Farah and Athanasios Kottas
University of California, Santa Cruz, USA
marian@ams.ucsc.edu thanos@soe.ucsc.edu

This work is concerned with the vulnerability of spaceborne microelectronics to single event upset (SEU), a change of state caused by electromagnetic radiation striking a sensitive node. To measure the susceptibility of a device to SEU, testing is conducted in a particle accelerator. The number of upsets depends on the linear energy transfer (LET) of the incident particles, the cross-section of interaction, and the fluence. The interaction cross-section is assumed to be monotonically increasing with LET. The prediction of the on-orbit upset rate, one of the main

goals of the particle accelerator experiments, is made by combining the device geometry and the cross-section vs. LET curve for the device with a model for the orbit-specific radiation environment. Standard practice in the device testing literature is to assume a parametric form for the cross-section vs. LET curve, using either a cumulative Weibull or cumulative Lognormal form. Neither of these has a strong physical justification. We compare this standard approach with a semiparametric isotonic regression method using a Poisson model for the upset counts and a Dirichlet process prior for the cross-section vs. LET curve. This allows the data to drive the shape of the cross-section vs. LET relationship. We study the characteristics of the resulting cross-section vs. LET curves and their influence on the predicted on-orbit upset rate distributions for a number of parts.

Small Area Estimation Using Skew-Normal Models

Fernando A. S. Moura and Valmária R. Silva
Universidade Federal do Rio de Janeiro, Brazil
fmoura@im.ufrj.br

In many applications of area level models, it is customary to assume that the sampling error variances are known and given by their respective estimates. As pointed out by Rao (2003), these assumptions may be quite restrictive in some applications, mainly when the sample sizes of the areas are small. Nevertheless, in many situations, it is possible to obtain the variance estimates of the direct estimator from the unit-level survey data. The main aim of this work is to propose two important connected extensions of the Fay-Heriot (1979) model that might be of practical and theoretical interests. The first extension allows for the sampling error to be non-symmetrically distributed. This is important for the case that the sample sizes in the areas are not large enough to rely on the Central limit theorem. We deal with this by assuming that the sample error is skew-normal distributed. The second extension proposes to jointly model the direct estimator and its respective variance estimator. Proceeding in this way, we manage to take into account all sources of uncertainties. We also discuss how to assign an informative prior to the shape parameter. We apply our propose model to a real data set and compare with the Normal model. An extra simulation study is also carried out and their results fully discussed.

Novel Bayesian Models and Inference for High-Resolution Lattice Data

Chiranjit Mukherjee and Mike West
Duke University, USA
chiranjit@stat.duke.edu mike@stat.duke.edu

Applied studies in multiple areas involving spatial systems increasingly challenge our modelling and computational abilities as spatial scales move to increasingly

high-resolution with parallel increase in complexity of dependency patterns. Motivated by a challenging problem of global CO emissions prediction based on high-resolution global satellite imagery ñ we have been led to explore a new class of flexible Gaussian Markov random field models for inhomogeneous/non-stationary spatial processes on a lattice. Traditional GMRF models like CAR and SAR specify spatial dependence via a constant spatial autoregression across the locations ñ they lack the ability to represent major patterns of spatial nonstationarities. Our new class of spatially-varying simultaneous autoregressions (SVSAR) model neatly extends this to a spatially varying autoregression coefficients model, introducing a second-stage SAR model as the spatial prior for the autoregressive parameters of the first stage (data) SAR model. This novel specification allows a smooth variation in the structure of the local dependencies among spatial outcomes. In essence, SVSAR is a spatial analogue of the time-varying autoregressive (TVAR) extension of traditional AR models in time series.

Our new model has a number of attractive properties arising from its basis in Markov random fields, like fast numeric computation for sparse matrices that enable the development of posterior computation even with large lattices. This potential is evidenced in our applied example in atmospheric chemistry, where focus is on inversion of satellite data to infer ground-level CO emissions from multiple candidate sources on a global scale.

Bayesian Model Selection on Inverse Correlation Matrices With Application to Sparse Networks

Nitai D. Mukhopadhyay
Virginia Commonwealth University, USA
ndmukhopadhy@vcu.edu

Sarat C. Dass
Michigan State University, USA
sdass@stt.msu.edu

Statistical inference for an inverse correlation matrix is challenging due to constraints placed on its entries which require that its inverse (a correlation matrix) have unit diagonal elements. The aim of this paper is to present a new parametrization for the inverse correlation matrix, in terms of the Cholesky decomposition, that is able to model these constraints explicitly. The Cholesky decomposition is also utilized in the development of a class of hierarchical correlation selection priors that allow for varying levels of sparsity. An explicit expression is obtained for the volume of the elicited priors. Bayesian inferential methodology is developed using a Reversible Jump Markov Chain Monte Carlo algorithm. Performance of the algorithm will be demonstrated through simulated dataset as well as gene expression dataset on Androgen pathway association.

Fast and Efficient Bayesian Semi-Parametric Curve-Fitting and Clustering in Massive Data

Sabyasachi Mukhopadhyay Sisir Roy
Indian Statistical Institute, India
sabyasachi_r@isical.ac.in sisir@isical.ac.in

Sourabh Bhattacharya
Indian Statistical Institute, India
sourabh@isical.ac.in

The problem of curve-fitting and clustering using Bayesian mixture models, treating the number of components as unknown, has received wide attention in the Bayesian statistical community. Among a number of available Bayesian methodologies specialised for the purpose, the approaches proposed in Escobar and West(1995) and Richardson and Green (1997) stand out. But in the case of massive data substantial computational challenges seem to blur the attractive theoretical advantages of such well-established Bayesian methodologies.

In this paper, based on the recently proposed methodology of Bhattacharya (2009) we propose a very fast and efficient curve-fitting and clustering methodology, the latter being based on a new approach to analysing posterior distributions of clusterings first proposed in Bhattacharya, Samanta, Dihidar and Ghosh (2009). We demonstrate theoretically as well as with extensive simulation studies, the significant advantages of our approach over the aforementioned established approaches, particularly in the case of massive data. We also illustrate our methodologies on a real, cosmological data set consisting of 96,307 bivariate observations and demonstrate that the approach of Escobar and West(1995) is infeasible in this example and the approach of Richardson and Green (1997), although implementable, is likely to be quite slow and inefficient.

Fast Methods of Spatio-Temporal Disease Mapping with INLA

Facundo Muñoz
Universitat de València, Spain
Facundo.Munoz@uv.es

Rubén Amorós
Universitat de València, Spain
ruasal@alumni.uv.es

Antonio López-Quílez
Universitat de València, Spain
Antonio.Lopez@uv.es

M.A. Martínez-Beneito
CSISP Valencia, Spain
martinez_mig@gva.es

Spatio-temporal disease mapping is an active research field. Health authorities are increasingly demanding rapid methods for fitting these kind of models. In particular, among the specific objectives of the EUROHEIS 2 project is the inclusion of spatio-temporal methods for disease mapping in the Rapid Inquiry Facility (RIF) and other GIS-based analytical tools. However, there is not a wide consensus on how to describe temporal and spatial evolution at the same time in a proper way. Although several spatio-temporal disease mapping techniques have

been proposed recently, the implementation of these methods is not always easy or adequate for a quick response tool.

We outline a general framework for spatio-temporal models, breaking them up into four stages: the probabilistic model for observations, the components of the linear predictor, the structures of the effects and the inference methodology. For each stage, the most commonly used alternatives are discussed, with special emphasis on the spatio-temporal interaction.

Some of the models can be fitted with INLA, a new tool for approximate bayesian inference which performs particularly fast. We present a comparison of the results with these models, using a dataset of lung cancer mortality in women in Valencia during a 20 years period.

MCMC in R

Duncan Murdoch
University of Western Ontario, Canada
murdoch@stats.uwo.ca

R (R Development Core Team, 2010) is an open source language and environment for statistical computing and graphics, based on the S language (Becker *et al.* 1988). It is available for free from `http://cran.r-project.org`, and in recent years has become increasingly popular. One of the attractive features of R is that it promotes the development of user-contributed "packages" of new functions and data. Several of these packages are designed to support Markov chain Monte Carlo (MCMC) algorithms. In this paper, I will briefly review simple MCMC programming in R, and survey the packages that support it.

Variational Bayes for the Latent Position Cluster Model in Social Network Analysis

Brendan Murphy and Michael Salter-Townshend
University College Dublin, Ireland
brendan.murphy@ucd.ie michael.salter-townshend@ucd.ie

Many recent approaches to modeling social networks have focussed on embedding the actors in a latent "social space". Links are more likely for actors that are close in social space than for actors that are distant in social space. In particular, the Latent Position Cluster Model (LPCM) allows for explicit modelling of the clustering that is exhibited in many network datasets.

However, inference for the LPCM model via Markov Chain Monte-Carlo is cumbersome and scaling of this model to large or even medium size networks with many interacting nodes is a challenge. Variational Bayesian methods offer one solution to this problem. An approximate, closed form posterior is formed, with unknown variational parameters. These parameters are tuned to minimize the Kullback-Leibler divergence between the approximate variational posterior and the true posterior, which known only up to proportionality. The variational Bayesian approach is shown to give a computationally efficient way of fitting the

LPCM. The approach is demonstrated on a number of data sets and it is shown to give a good fit in terms of correct cluster assignments and good posterior predictive probabilities for both links and non-links.

As the variational method is prone to convergence to local minima, good initialisation for the variational parameters is crucial. We develop an initialisation for the latent positions inspired by force-based algorithms such as the Fruchterman-Reingold method. Ours uses the latent position log-likelihood for the network links to yield starting positions in the social space that are compatible with the LPCM. We explore the impact of various initialisation procedures on our variational algorithm using simulated datasets.

Slice Sampling with Latent Gaussian Models

Iain Murray
University of Edinburgh, UK and *University of Toronto, Canada*
i.murray@ed.ac.uk

Ryan Prescott Adams
University of Toronto, Canada
rpa@cs.toronto.edu

Many probabilistic models incorporate multivariate Gaussian distributions to explain dependencies between observed variables. Gaussian process models and generalized linear mixed models are common examples. The posterior distribution over all unknowns, including parameters that specify the covariance matrix, is typically not available in closed form. Existing computational methods can be broadly split into two classes: deterministic approximations and simulation via Monte Carlo. In our current work we aim to make the sampling based approach easier to use.

Slice sampling (Neal, 2003) is a combination of an auxiliary variable sampler with an adaptive search procedure. The result is an easy-to-use Markov chain Monte Carlo algorithm which is robust to initial choices of step-size parameters. Agarwal and Gelfand (2005) demonstrated the utility of slice sampling for updating covariance parameters. In recent work we have supplemented this with *Elliptical Slice Sampling*, a slice-sampling variant that requires zero free parameters and is suitable for updating strongly coupled a-priori Gaussian variates given non-Gaussian observations. This procedure works well for many problems and is easy to apply.

The nature of a model's posterior distribution can vary significantly depending on how informative the observations are. Different parameterizations for the weak and strong data regimes may be necessary for sampling algorithms to work well. Christensen *et al.* (2006) have shown how it is possible to construct model parameterizations that automatically match the type of data available. We are currently investigating auxiliary variable models with this flavor that also allow slice sampling.

Bayesian Adaptive Optimal Design for Discriminating Models of Cognition

Jay Myung, Daniel Cavagnaro and Mark Pitt
Ohio State University, USA
myung.1@osu.edu cavagnaro.2@posu.edu pitt.2@osu.edu

Experimentation is fundamental to the advancement of science, whether one is interested in studying the neuronal basis of a sensory process in cognitive neuroscience or assessing the efficacy of a new drug in clinical trials. Adaptive design optimization, in which the information learned from each experiment is used to inform subsequent experiments, is a particularly attractive methodology because it can potentially reduce the time required for data collection while simultaneously increasing the informativeness of the knowledge learned in the experiment. More concretely, the problem to be solved in adaptive sequential design optimization for model discrimination is to identify an experimental design under which one can infer the underlying model, among a set of candidate models of interest, in the fewest possible steps. This problem is challenging because of the many, sometimes arbitrary, choices that must be made when designing an experiment. Nevertheless, it is generally possible to find a design that is optimal in a defined sense. In this paper, addressing the design optimization problem in discrimination of formal models of cognition, we apply a simulated-based Bayesian method that was recently introduced in statistics (Muller, Sanso & De Iorio, 2004). We use a utility function based on mutual information, and give three intuitive interpretations of the utility function in terms of Bayesian posterior estimates. Finally, we demonstrate the potential of adaptive design optimization for improving experimentation in psychology by implementing the method in experiments with simulated as well as human participants

Stochastic Volatility Model with Leverage and Asymmetrically Heavy-Tailed Error Using GH Skew Student's t-Distribution

Jouchi Nakajima
Duke University, USA
jouchi.nakajima@stat.duke.edu

Yasuhiro Omori
University of Tokyo, Japan
omori@e.u-tokyo.ac.jp

Bayesian analysis of a stochastic volatility model with a generalized hyperbolic (GH) skew Student's t-error distribution is described where we first consider an asymmetric heavy-tailness as well as leverage effects. An efficient Markov chain Monte Carlo estimation method is described exploiting a normal variance-mean mixture representation of the error distribution with an inverse gamma distribution as a mixing distribution. The proposed method is illustrated using simulated data, daily TOPIX and S&P500 stock returns. The model comparison for stock returns is conducted based on the marginal likelihood in the empirical study. The strong evidence of the leverage and asymmetric heavy-tailness is found in the stock returns. Further, the prior sensitivity analysis is conducted to investigate whether obtained results are robust with respect to the choice of the priors.

A Markov Chain Monte Carlo Implementation of the Bayesian Method of Moments for Linear Regression Models

Teruo Nakatsuma
Keio University, Japan
nakatuma@econ.keio.ac.jp

In this paper I first show a new interpretation of the post-data moment condition in the Bayesian method of moments (BMOM) and clarify the relationship between the classical GMM and the BMOM. Then I extend the BMOM to linear regression models with autocorrelated/heteroskedastic errors or Box-Cox transformed variables. Since analytical evaluation of the post-data statistics about parameters is not possible, I utilize a Markov chain Monte Carlo method to generate parameters from the post-data distribution and evaluate the post-data statistics numerically. A few empirical examples will be presented.

A Semiparametric Bayesian Approach to Extreme Value Estimation

Fernando F. Nascimento
Universidade Federal do Rio de Janeiro, Brazil
fernandofn@dme.ufrj.br

Dani Gamerman
UF do Rio de Janeiro, Brazil
dani@im.ufrj.br

Hedibert F. Lopes
University of Chicago, USA
hedibert.lopes@chicagobooth.edu

This work is concerned with extreme value density estimation. The generalized Pareto distribution (GPD) beyond a given threshold is combined with a nonparametric estimation approach above the threshold. This semiparametric setup is shown to generalize a few existing approaches and enables density estimation over the complete sample space. Estimation is performed via the Bayesian paradigm, which helps identify model components. Estimation of all model parameters, including the threshold and higher quantiles, and prediction for future observations are provided. Simulation studies suggests a few useful guidelines to evaluate the relevance of the proposed procedures. Models are then applied to environmental data sets. The paper is concluded with a few directions for future work.

Bayesian Estimation of ILI Incidence Using Non-Representative Data

Isabel Natário, and Lucília Carvalho
Universidade Nova de Lisboa, Portugal
icn@fct.unl.pt mlcarvalho@fc.ul.pt

In a changing world, the need for timely detection of contagious diseases outbreaks has resulted in the development of several working systems to do so, particularly

of systems relying on the investigation of symptoms rather than on confirmed diagnosis. Gripenet is a syndromic surveillance system of this kind, set up in some Portugal in 2005, for detection of influenza like illness (ILI) diseases, based on volunteer participation through the answer of weekly internet questionnaires on symptoms. Naturally, the nature of the data we get from this scheme does not allow its direct use to estimate ILI incidence, as we would like. In this work we propose a methodology based on state-space models, used before in animal population dynamics modeling, to overcome this data lack of randomness and to allow for incidence estimation. We consider the evolution in time of the process of the (unknown) population states of those that are ILI and non-ILI diseased in each week, running in parallel with the observational process of Gripenet answers about respondents ILI symptoms in the corresponding week, function of the unobserved states. This approach is flexible enough to decompose the population process into several subprocesses of interest, accommodate process variation and observational errors. The estimation through Bayesian techniques such as MCMC or sequential importance sampling with resampling come naturally to these models. Here we apply them both to the Portuguese Gripenet data from 2006, and we try further to compare the obtained ILI incidence estimates with the ones obtained with data from another existing symdromic scheme, a network of sentinel doctors reporting on this.

The Meta-Analytic-Predictive Approach to Deriving Priors from Historical Data in Clinical Trials

Beat Neuenschwander Gorana Capkun-Niggli
Novartis Pharma, Switzerland
{beat.neuenschwander, gorana.capkun-niggli}@novartis.com
Michael Branson David J. Spiegelhalter
Novartis Pharma, Switzerland *MRC, Biostatistics Unit, UK*

Historical information is always relevant for the design of a clinical trial, but one might also want to incorporate it as prior information in the analysis (Pocock, 1976). For setting up the prior distribution for the parameter $\theta^\star$ of the control group in a new trial based on historical estimates $Y_1, \ldots, Y_H$ for the control parameters $\theta_1, \ldots, \theta_H$ in H historical trials, we emphasize the importance of a hierarchical meta-analytic-predictive approach (Spiegelhalter, *et al.* 2004) and (Neuenschwander *et al.* 2010). Predictions require a relationship or model for past and future parameters. The assumption of fully exchangeable parameters, historical and new, forms the basis of the meta-analytic-predictive approach, but it can be extended to partially exchangeable parameters by adjusting for trial-specific covariates using meta-regression.

The predictive distribution for the new parameter $p(\theta^\star|Y_1, \ldots, Y_H)$ constitutes the prior for the new trial. An approximate *prior effective sample size* $n^\star$ for the new trial is derived. The prior effective sample size is bounded by the *prior maximum sample size* $n^\star_\infty = \sigma^2/\tau^2$, where σ and τ are the within- and between-trial standard deviations, respectively.

Results are presented for two trial designs: a Phase IV trial with 930 patients in 11 historical trials, and a Phase II Proof-of-Concept trial with 363 patients

in four historical trials. Between-trial variability was fairly small for the first and moderate to substantial for the second application, which resulted in prior effective sample sizes of 90 and 20, respectively.

Malnutrition Among Children under the Age of Five in the Democratic Republic of Congo: Do Geography Matters?

KANDALA NGIANGA-BAKWIN
University of Warwick, UK
n-b.kandala@warwick.ac.uk

There is interest in the extent to which childhood under-nutrition differs geographically across a least developing country like the DRC. In this case, the interest is strengthened by concern that environmental pollution associated with conflicts and displacement damages child health in parts of the country. Recent household survey, the 2007 Demographic and Health household survey of the DRC, has been used to investigate the impact of geographical locations on childhood nutritional status by mapping the residual net effect of malnutrition while accounting for important risk factors.

We use Bayesian geo-additive semi-parametric regression model to flexibly model the effects of selected socioeconomic covariates and spatial effects. Inference is fully Bayesian based on recent Markov chain Monte Carlo techniques. The novel empirical approach used is able to flexibly determine to what extent the substantial spatial pattern of child health and survival is driven by observable factors such as socioeconomic factors considered or be attributable to unmeasured factors such as conflicts, environmental, cultural factors.

Findings show that although childhood malnutrition were more pronounced in all provinces in the DRC, after accounting for the location effects, geographical differentiations were found to be quite significant: malnutrition is significantly higher in rural areas compared to urban centres and this difference is increasing over time. We offer a preliminary interpretation suggesting that these spatial trends may, at least partly, be attributed to unobserved factors such as environmental pollution, conflicts, deterioration of the social fabric and health care system in most of the provinces. Therefore, the search for an explanation for unobserved factors warrants further research with explicit measures exposure.

Robust Bayesian Models for ANOVA

GEORGE NICHOLSON and CHRIS C. HOLMES
University of Oxford, UK
nicholso@stats.ox.ac.uk cholmes@stats.ox.ac.uk

When faced with potentially outlier-contaminated data, a common method of achieving robust parametric inference is to lengthen the tails of the error distribution, replacing outlier-resistant distributions, such as the Gaussian, with outlier-prone distributions, such as the Student t. The use of outlier-prone likelihoods ensures that extreme observations have bounded or vanishing posterior influence.

We address certain aspects of the implementation of heavy-tailed modelling for scale-parameter estimation in the Bayes linear model. For simplicity, we restrict attention to the covariance structure induced by the simplest Bayes linear model: $\boldsymbol{y} = \sum_{r=1}^{m} \boldsymbol{Z}_{(r)}\boldsymbol{u}_{(r)} + \boldsymbol{\varepsilon}$; $\boldsymbol{u}_{(r)} \mid \sigma_{(r)} \sim \mathsf{N}_{n_{(r)}}(\boldsymbol{0},\ \boldsymbol{I}_{n_{(r)}}\sigma^2_{(r)})$. and $\boldsymbol{\varepsilon} \mid \sigma_{(0)} \sim \mathsf{N}_{n_{(0)}}(\boldsymbol{0},\ \boldsymbol{I}_{n_{(0)}}\sigma^2_{(0)})$, where the $\boldsymbol{u}_{(r)}$ and $\boldsymbol{\varepsilon}$ are vectors of nuisance parameters, and the $\boldsymbol{Z}_{(r)}$ represent known design matrices. The main analytical objective is perform inference on $\boldsymbol{\sigma} \equiv (\sigma_{(0)}, \ldots, \sigma_{(m)})$ by characterizing the posterior distribution, $p(\boldsymbol{\sigma} \mid \boldsymbol{y})$.

In many applications of the above model, it is not only the observation-specific effects, $\boldsymbol{\varepsilon}$, that potentially contain outliers, but also the $\boldsymbol{u}_{(r)}$ (corresponding to highly structured outlying units, such as individuals, plots, or batches). It is natural then to allow for outliers at these lower levels, by specifying heavy-tailed distributions for $\boldsymbol{u}_{(1)}, \ldots, \boldsymbol{u}_{(m)}$. We perform and compare inference under models with heavy-tailed distributions on some, none, or all of the effects.

The most common approach to Bayesian robust analysis involves replacing the Gaussian distribution with either (i) a Student-t distribution or (ii) a mixture of the Gaussian with a scale-inflated Gaussian (sometimes with priors placed and estimation performed on the degrees of freedom of the Student, or the mixing proportion of the Gaussian mixture). We examine and compare the properties of these, and other, heavy-tailed models, such as a Gaussian-Student mixture. Heavy-tailed distributions provide robustness to extreme observations, but sometimes at the expense of introducing a downwards bias in scale-parameter estimates. We examine univariate efficient score functions, $\partial \log[p(y \mid \sigma)]/\partial\sigma$, to gain insights into how various heavy-tailed likelihoods down-weight outliers and potentially induce bias. We argue that suitable choice of robust models can be guided by examination of the corresponding univariate score functions.

For post-fit model assessment, we introduce a diagnostic technique that quantifies the influence of each data point on parameter estimates (and thus to what extent outliers are down-weighted). The method decomposes the posterior score function, $\partial \log[p(\boldsymbol{\sigma} \mid \boldsymbol{y})]/\partial\sigma_{(r)}$, into datum-specific contributions. We apply these methods to a longitudinal study of metabolic profiles conducted on twins, in which ^{1}H-NMR spectroscopy was used to acquire metabonomic spectral data from blood and urine samples. In our implementation, heavy-tailed distributions are expressed as scale mixtures of normals, and MCMC is used to sample from the joint posterior, $p(\boldsymbol{\sigma} \mid \boldsymbol{y})$.

Bayesian Inference in Integer-Valued Self-Exciting Threshold Autoregressive Models SETINAR(2;p)

Raquel Nicolette
Universidade de Aveiro, Portugal
nicolette@ua.pt

Isabel Pereira and Manuel Scotto
Universidade de Aveiro, Portugal
isabel.pereira@ua.pt mscotto@ua.pt

We consider the class of self-exciting threshold integer-valued autoregressive models (SETINAR), driven by independent Poisson-distributed random variables. This

paper considers the parameters estimation problem for the p*th*-order self-exciting threshold integer-valued autoregressive model with two regimes, SETINAR(2;p). The Markov chain Monte Carlo (MCMC) algorithm developed by Neal and Subba Rao (2007) is used. In order to acess the MCMC algorithm inference capabilities a simulation study is carried out.

A Construction of Vectors of Dependent Dirichlet Processes

Bernardo Nipoti
Università di Pavia, Italy
bernardo.nipoti@unipv.it

We present the construction of two dependent Dirichlet processes (DPs) that we obtain by normalizing the increments of two dependent gamma completely random measures (CRMs). These are constructed as linear functionals of dependent Poisson random measures introduced in Griffiths & Milne (1978). If p_0, p_1 and p_2 are independent Dirichlet processes with suitable parameters, such a construction gives rise to dependent DPs, $\tilde{p}_1$ and $\tilde{p}_2$, with the property that $\tilde{p}_1$ is a mixture of p_0 and p_1 while $\tilde{p}_2$ is a mixture of p_0 and p_2. The vector of CRMs that we obtain is analiticaly tractable and the dependence structure between $\tilde{p}_1$ and $\tilde{p}_2$ can be described in terms of a parameter in $[0, 1]$. Given two ecxhangeable samples from $\tilde{p}_1$ and $\tilde{p}_2$, respectively, a posterior estimation of such a parameter suggests how distant we are from the two extreme situations of full exchangeability ($\tilde{p}_1 = \tilde{p}_2 = p_0$) and partial exchangeability ($\tilde{p}_1 = p_1$ and $\tilde{p}_2 = p_2$).
The vector $(\tilde{p}_1, \tilde{p}_2)$ may be also used to construct dependent mixture models. An extension of the well-known Blackwell-MacQueen sampling scheme allows one to implement a Gibbs sampler and achieve a full Bayesian analysis for clustering and density estimation. This approach also leads one to construct $k \geq 2$ dependent DPs and more general dependent random probability measures for Bayesian nonparametric inference.

Using Short-term Evidence to Predict Six-Month Outcomes in Clinical Trials of Signs and Symptoms in Rheumatoid Arthritis

Richard M Nixon
Novartis AG, Switzerland
richard.nixon@novartis.com

Objectives: A model is presented to generate a distribution for the probability of an ACR response at six months for a new treatment for rheumatoid arthritis given evidence from a one or three month clinical trial. Methods: The model is based on published evidence from 11 randomized controlled trials on existing treatments. A hierarchical logistic regression model is used to find the relationship between the proportion of patients achieving ACR20 and ACR50 at one and three months and the proportion at six months. We denote j as the treatment arm

within each clinical trial, and k the type of treatment used ($k = 1$ for MTX, $k = 2$ for biologic, $k = 3$ for biologic plus MTX); $t0$ as the earlier time point (either one month or three months depending on the model); and $t1$ the six-month time point. We then denote n_j as the number of patients in treatment arm j for which there is data available at both time points, and r_{jt} as the number of patients achieving the ACR response criteria at time point t. We assume that r_{jt0} and r_{jt1} are binomially distributed with parameters ϕ_{jt0} and ϕ_{jt1} representing the probabilities of response, respectively. ϕ_{jt1} is assumed to be dependent upon ϕ_{jt0}, in a logistic regression. To estimate the ACR response rate at time point $t1$ from the ACR response rate at time point $t0$, we fit the following statistical model: $r_{jt0} \sim \text{Bi}(n_j, \phi_{jt0})$; $r_{jt1} \sim \text{Bi}(n_j, \phi_{jt1})$; $\text{logit}(\phi_{jt1}) = \alpha + \text{logit}(\phi_{jt0})$; and $\alpha \sim \text{N}(0, \sigma^2)$.

Results: The model is assessed by Bayesian predictive P-values that demonstrate that the model fits the data well. Conclusions: The model can be used to predict the number of patients with an ACR response for proposed six-month clinical trials given data from clinical trials of one or three months duration.

Bayesian Spatial Smoothing over Complex Domains

Alessandro Ottavi
Norges Teknisk-Naturvitenskapelige Universitet, Norway
ottavi@math.ntnu.no

Spatial smoothing methods often assume the region of interest is $\Re^2$ but in practice we might have a river, lake or a mountain, which, at least "physically", prevents smoothing over the river, lake or mountain. Such geometrical "constraints" is not easy to deal with properly.

In this work, we try to not see these physical constraints as "annoying constraints" but rather reinterpret physical constraints into geometrical properties and then include the geometry into the model. To do this, we have to leave $\Re^2$ and define smoothing methods on a manifold in $\Re^3$. The essential idea, is to make use of the ideas by Lindgren and Rue (2007), which showed how Matérn fields on a manifold could be constructed using the corresponding stochastic partial differential equation (SPDE). Moreover, they also achieved a representation with Markov properties which ensure fast computations. We show how their approach could be extended to deal with spatial smoothing over complex domains which automatically respect and adjust for physical constraints.

This is joint work with H. Rue.

Bayesian Emulators for Multivariate Computer Models with Categorical Inputs

Antony M. Overstall and Dave C. Woods
University of Southampton, UK
amo105@soton.ac.uk D.Woods@soton.ac.uk

Computer models are mathematical representations of physical systems. They are used for real-world problems where it would be expensive, impossible or unethical to use a physical experiment. Many models are computationally expensive, taking considerable time to run. An emulator is a statistical meta-model whose purpose is to predict the output of a computer model at any particular set of inputs. Emulators allow fast prediction of the computer model, making practicable tasks such as sensitivity analysis, uncertainty analysis and model calibration.

A computer model is considered that has dynamic, multivariate output, and both continuous and categorical inputs. Examples of such models occur in many areas of science and include climate models and dispersion models. Our example is from emergency planning.

Challenges for such models include the definition and incorporation of appropriate distance metrics for the categorical variables (Qian *et al.* 2008) and implementing efficient methods for approximating the multivariate posterior predictive distribution. Several Bayesian methods for emulating the multivariate output are assessed, drawing on the recent work of Rougier (2007) and Conti and O'Hagan (2010).

Bayesian Emulators as a Tool for Comparing Ecosystem Models

Rachel Oxlade
University of Durham, UK
r.h.oxlade@durham.ac.uk

Complex computer simulators are widely used to forecast future events and to make judgements about the state of the system they model. However, due to our limited knowledge of the system and restrictions on the intricay of our model, such simulators often carry with them high degrees of error and uncertainty. This means that given more than one model of a particular system, they will not agree everywhere, and may even be highly contradictory. How we link these models to one another or to the real system is a difficult but important issue.

Due to the complexity of the models and the high dimensionality of the input space, it would be too costly to run the simulators at every input point, and so we are uncertain about the behaviour of the models themselves except for the limited data we have. Bayesian emulators are widely used to address this problem by providing an approximation to the simulator as well as quantifying our uncertainty at each point.

This poster introduces the notion of using emulators to compare two different models of the same system, with a view to tackling the problems mentioned above. HadOCC, the Hadley Centre Ocean Carbon Cycle model, is used as an example. For several physical process, the user can choose from a range of submodels, and so this provides a good first step to dealing with multiple simulators. HadOCC models plankton dynamics, and provides a way to estimate carbon exported to the deep ocean by the "biological pump".

Flexible Spatial Latent Variable Modeling for Combining Information Sources while Accounting for Systematic Errors in Proxies

Christopher Paciorek
Univ. of California at Berkeley and *Harvard School of Public Health, USA*
paciorek@stat.berkeley.edu

Environmental research increasingly uses high-dimensional remote sensing and numerical model output to help fill space-time gaps between traditional observations. Such output is often a noisy proxy for the process of interest. Thus we need to separate and assess the signal and noise (often called discrepancy) in the proxy, given sparse observations and complicated spatio-temporal dependence structure in both the discrepancy and the true process.

Building on recent statistical work that models the proxy as a second likelihood in a hierarchical Bayesian model, I develop a more flexible model that does not constrain the discrepancy to vary only at large spatial scales by using a Markov random field (MRF) approximation to a thin plate spline (Rue and Held, 2005, Gaussian Markov Random Fields). The spatial scale of the discrepancy is determined based on the concordance between the gold standard observations and proxy output. Unlike other spatial process representations, the model can capture small-scale discrepancy in a computationally efficient manner through sparse matrix manipulations, while also having more appealing statistical properties than standard CAR models.

I apply the approach to spatio-temporal mapping of fine particulate matter (PM) air pollution levels using satellite aerosol and atmospheric model output proxies. The results indicate little predictive improvement over modeling of the observations alone. The estimated discrepancies occur at a variety of spatial scales in the examples, with small-scale discrepancy particularly important in minimizing the contribution of the proxies to the prediction of PM. The results highlight the critical question of how best to make use of proxy information while minimizing the potential for proxy-induced bias, and I contrast the two-likelihood approach with simply regressing on the proxy.

Bayesian Local Contamination Models for Multivariate Outliers

Garritt Page and David Dunson
Duke University, USA
page@stat.duke.edu dunson@stat.duke.edu

It is not uncommon for there to exist observations that are quite unlike the majority in studies where data are generated from multiple locations or sources. Motivated by the application of establishing a reference value in an inter-laboratory setting when outlying labs are present, we propose a local contamination model that is able to accommodate unusual multivariate realizations in a flexible way. The proposed method models the process level of a hierarchical model using a mixture with a parametric component and a possibly nonparameteric contamination. Much of the flexibility in the methodology is achieved by allowing varying random subsets of the elements in the lab-specific mean vectors to be allocated to the contamination component. Computational methods are developed and the methodology is compared to three other possible approaches using a simulation study. We apply the proposed method to a NIST/NOAA sponsored inter-laboratory study which motivated the methodological development.

Bayesian Estimation of Bacterial Growth Curves using Neural Networks and Hierarchical Gompertz Models

Ana P. Palacios Juan-Miguel Marín
Universidad Carlos III de Madrid, Spain
appalaci@est-econ.uc3m.es jmmarin@est-econ.uc3m.es

Michael P. Wiper
Universidad Carlos III de Madrid, Spain
mwiper@est-econ.uc3m.es

Bacterial growth models are commonly used in food safety. Such models permit the prediction of microbial safety and the shelf life of perishable foods. Environmental factors such as temperature and acidity affect the process of growing, therefore these effects should be taken into account. In this study we develop two models to describe bacterial growth in different temperature and acidity conditions. Firstly, a hierarchical Gompertz regression model is proposed and secondly we consider a non-parametric approach based on the use of artificial neural networks (ANNs). The major advantages for using ANNs are the capacity to describe interacting effects of environmental factors and to fit the bacterial growth in a single step taking into account the controlling factors with no need of secondary models. Both models are fitted using `OpenBugs` (`Brugs`) and `R2Winbugs` which avoids the need for the selection of complex Metropolis Hastings samplers. Model selection is undertaken throughout using an appropriate version of the deviance information criterion. Both models are illustrated using experimental data from *Listeria monocytogenes* growth.

Model-Based Clustering of Categorical Time Series Using Finite Mixtures of Markov Chain Models with Logit Extension

Christoph Pamminger Sylvia Frühwirth-Schnatter
Johannes Kepler Universität Linz, Austria
{Sylvia.Fruehwirth-Schnatter, christoph.pamminge}@jku.at

Rudolf Winter-Ebmer
Johannes Kepler Univ., Austria
Rudolf.WinterEbmer@jku.at

Andrea Weber
RWI Essen, Germany
andrea.weber@rwi-essen.de

We present an approach for model-based clustering of categorical time series which also allows for the inclusion of time-constant covariates. Our clustering method is based on time-homogeneous first-order Markov chains where the individual transition probabilities are fixed to a group-specific transition matrix. It is therefore called Markov chain clustering and can be regarded as a finite mixture model for Markov chain models. Our new suggestion is an extension to this model where we use a multinomial logit model as a probabilistic (logit-type prior) model for the latent group indicator. This extended model now allows to include explanatory variables to further explain group membership. Estimation is carried out through Markov chain Monte Carlo including an auxiliary mixture sampler for the parameters of the multinomial logit model. It is a (full) Gibbs-type sampler that needs only draws from standard distributions. An application to a panel of Austrian wage mobility data will be presented which leads to an interesting segmentation of the Austrian labour market.

Bayesian Estimation of Pair Copula Constructions with Discrete Margins

Anastasios Panagiotelis and Claudia Czado
Technische Universität München, Garching, Germany
panagiot@ma.tum.de cczado@ma.tum.de

Models based on copulas have become increasingly popular as they enable the construction of multivariate distributions that flexibly capture non-linear dependencies while having fixed margins that can be either continuous, discrete or some combination of both. For an m-dimensional copula model with discrete margins, maximum likelihood estimation is infeasible, even for a moderate number of dimensions, since computing the probability mass function requires 2^m evaluations of the copula function. For Gaussian copulas, Pitt, Chan and Kohn (2006) propose a Bayesian solution to this problem by augmenting the likelihood with continuous latent variables and conducting inference using Markov chain Monte Carlo (MCMC). However, through an extensive simulation study, we show that estimates of quantities such as Kendall's tau, the K-function and the lambda -function are not robust to copula misspecification, even in the discrete case. As such, we develop an estimation approach similar to that of Pitt, Chan and Kohn (2006), but that can be applied to a more general class of copulas. In a simulated setting, we compare the efficiency of different sampling schemes, both

with regard to computational speed and convergence and mixing of the Markov chain. These schemes include a grouped move multigrid MCMC sampler (Liu and Sabatti (2000)) which improves mixing when copula and marginal parameters are estimated jointly. To demonstrate the potential of our approach we fit a highly flexible D-vine pair copula construction (Aas, Czado, Frigessi and Bakken (2009)) to medical, marketing and financial data.

Bayesian Profile Regression with an Application to the Study of Lung Cancer in a Large Cohort Study

Michail Papathomas John Molitor
Imperial College London, UK
m.papathomas@imperial.ac.uk john.molitor@imperial.ac.uk

Sylvia Richardson Elio Riboli and Paolo Vineis
Imperial College London, UK

Standard regression analyses are often plagued with problems encountered when one tries to make meaningful inference going beyond main effects, using datasets that contain dozens of variables that are potentially correlated. We propose a method that addresses these problems by using, as its basic unit of inference, a profile, formed from a sequence of covariate values. These covariate profiles are clustered into groups using the Dirichlet process, and are associated via a regression model to a relevant outcome. The Bayesian clustering aspect of the proposed modelling framework has a number of advantages over traditional clustering approaches in that it allows the number of groups to vary, allows comparison of arbitrary subgroups of the data, can incorporate a priori known structures, uncovers subgroups based on their association with an outcome of interest and fits the model as a unit, allowing an individualís outcome to influence cluster membership. Different variable selection approaches have been implemented and compared. Profile regression has been applied to a large cohort study in order to examine the effect of environmental carcinogens and explore possible gene-environment interactions.

Local Proper Scoring Rules for Discrete Outcome Spaces

Matthew F. Parry A. Philip Dawid
University of Cambridge, UK
M.Parry@statslab.cam.ac.uk A.P.Dawid@statslab.cam.ac.uk

Steffen L. Lauritzen
University of Oxford, UK
steffen@stats.ox.ac.uk

In Bayesian decision theory, where the action is to quote a distribution Q for a variable $X \in \mathcal{X}$, a *scoring rule*, assigning loss $S(x, Q)$ to the subsequent observation $X = x$, is *proper* if the choice $Q = P$ is a Bayes act when $X \sim P$. It is *local* if it depends on the behaviour of the distribution Q only at x and at outcomes in $\mathcal{X}$ that are in some sense close to x.

For a discrete sample space, we show that the relation "y is close to x" must be symmetric, and so defines an undirected graph G on the outcome space; and any undirected graph can be used for this purpose. Any local proper scoring rule $S(x, Q)$ is generated by the log score and a collection of functions $\{\phi_K\}$, one for each clique K of G, where ϕ_K is a convex homogeneous function of degree 1 whose arguments are $\{q(x) : x \in K\}$. Excluding the log score term, any such scoring rule is independent of the normalization of $q(x)$. These characterisations and properties parallel similar results for continuous sample spaces, where "locality" is defined as dependence on a finite number of derivatives of the density at x. Besag's pseudolikelihood is a local proper scoring rule in the above sense. This allows us to place pseudolikelihood in a broader theoretical context.

Bayesian Geostatistical Modeling with Informative Sampling Locations

Debdeep Pati
Duke University, USA
dp55@stat.duke.edu

Brian J. Reich
North Carolina State Univ., USA
reich@stat.ncsu.edu

David B. Dunson
Duke University, USA
dunson@stat.duke.edu

We consider geostatistical models that allow the locations at which data are collected to be informative about the outcomes. For example, in monitoring of air pollution, one may place more monitors at locations believed to have a high value of ozone or another pollutant, while in studying distribution of animal species one may systematically look in locations thought to commonly contain the species of interest. Motivated by Diggle et al. (2009), we propose a novel Bayesian geostatistical joint model, which models the locations using a log Gaussian Cox process, while modeling the outcomes conditionally on the locations as Gaussian with a Gaussian process spatial random effect and adjustment for the location intensity process. To our knowledge, we are the first to develop a Bayesian approach to

the informative locations problem in geostatistical modeling. Our primary contribution lies in studying the theoretical properties of the model. In particular, it is not obvious that the data contain information about the informativeness of the sampling locations, and one may wonder to what extent the prior is driving the results even in large samples. We address this concern by proving posterior propriety under an improper prior on the parameter controlling the degree of informative sampling. In addition, we show that the density of the locations and mean function of the outcome process can be estimated consistently under mild assumptions. This later result extends recent work showing posterior consistency in Gaussian process regression models (Choi & Schervish, 2007; Choi, 2007). The methods are illustrated through a variety of simulation examples and applied to the Ozone data in the Eastern US and also to a species abundance data.

Bayesian Stochastic Volatility Model for Leverage Effect

Shelton Peiris Jennifer S.K. Chan and Derek Feng
The University of Sydney, Australia
plam7745@uni.sydney.edu.au jennifer.chan@sydney.edu.au

This paper investigates modeling alternatives for the leverage effect in stochastic volatility model. Modeling alternatives encompass adding the lag-one return and/or other related variables as covariates and allowing for their asymmetric effects using regime switching models. For model robustness, several heavy-tailed and skewed distributions including Student-t and Variance Gamma distributions are considered and compared to the standard Normal distribution. Proposed models are demonstrated using the daily exchange rate returns of the Australian dollar to various currencies and the daily stock market index returns of various international stock markets. WinBUGS software package is used to implement the models. Lastlym model performance is evaluated through the Deviance Information Criterion, as well as the accuracy of the model forecasts.

Detection of Additive Outliers in Integer-Valued AutoRegressive Models

Isabel Pereira and M. Eduarda Silva
Universidade de Aveiro, Portugal *Universidade do Porto, Portugal*
isabel.pereira@ua.pt mesilva@fep.up.pt

Outliers are commonly encountered in time series data analysis. However the existence of additive outliers is often ignored and their impact is overlooked. In this paper the problem of detecting additive outliers in Integer-valued AutoRegressive models is considered. We use MCMC methods in order to detect aberrant observations in INAR(1) processes, particularly to estimate the probability of its occurrence at each time point. The proposed methodology is illustrated using some generated examples.

Limiting the Shrinkage for the Exceptional: A solution to the "Clemente Problem"

María E. Pérez and Luis R. Pericchi
Universidad de Puerto Rico, Rio Piedras, Puerto Rico
meglee@uprrp.edu luarpr@gmail.com

Modern Statistics is made of the sensible combination of direct evidence (the data directly relevant or the "individual data") and indirect evidence (the data and knowledge indirectly relevant or the "group data"). The admissible procedures are a combination of the two sources of information, and the advance of technology is making indirect evidence more substantial and ubiquitous. It has been pointed out however, that in "borrowing strength" a fundamental problem of statistics is to treat in a drastically different way, exceptional cases, namely those cases that do not adapt to the central "aurea mediocritas". This is what has been recently coined as "the Clemente problem" (Efron 2009). In this poster we propose that the problem is caused by the simultaneous use of square loss function and conjugate priors which is the usual procedure. We call their use "light penalty" because square loss is too lenient a penalty and conjugate priors have too light tails. We propose in their place to use heavy penalties, in the form of losses that penalizes more severely huge errors, or priors of heavy tails which make more probable the exceptional. We put forward concrete heavy loss functions, empirical (robust) Bayes and full robust Bayes practical procedures. As an striking bonus we find an example where Robust Bayes can beat the James-Stein estimator in its own field, decreasing the mean square error of prediction.

Meta-Analysis in Genome-Wide Association Studies

Matti Pirinen Garrett Hellenthal
University of Oxford, UK
matti.pirinen@iki.fi

Chris C. A. Spencer and Peter Donnelly
University of Oxford, UK

Contemporary genome-wide association studies (GWAs) consider genotype data on thousands of individuals at up to a million polymorphic loci. A central question is to quantify the statistical evidence about connections between the genetic variants and the phenotypes of interest. In the past few years GWAs have identified many loci that show convincing statistical association with complex human diseases, and the expectation is that these signals will lead us closer to the actual biological mechanisms behind the phenotypes.

Typically a large sample of individuals is needed to establish convincing evidence of association, and therefore it has become common to increase power by meta-analysis (*i.e.*, combining the association evidence from several different studies on the same phenotype).

We introduce a joint model for genotype data from multiple GWAs using hierarchical Bayesian modeling. This approach provides a flexible framework for

implementing the idea that different populations are expected to show similar – but not necessarily the same – size of genetic effects.

Our implementation applies an MCMC algorithm to find the joint posterior distribution of the parameters and importance sampling to compute the marginal likelihoods. Despite of the use of sampling algorithms the method is readily applicable to the genome-wide data sets that are encountered in contemporary human genetics. Examples with real data are presented and compared with other approaches to meta-analysis in case-control study design.

Bayesian Analysis of Partial Observability in Bivariate Probit Models

Dale J. Poirier
University of California, Irvine, USA
dpoirier@uci.edu

This paper considers a partially observed bivariate probit model where only a single discrete outcome is observed indicating whether both binary outcomes are *one*. Although this model was introduced thirty years ago, no Bayesian analysis has been performed. Areas of application have included accounting banking, economics of education, development, foreign direct investment, industrial relations, labor, law and economics, mortgage lending, political science, public economics, and sociology. The parameters of such a model are locally, but not globally identified. There is a *labeling* problem. A simple zero restriction on one of the coefficients in the univariate bivariate probit equations is sufficient to give identification, and such restrictions are often plausible. Covariates which affect one choice, but not the other, are particularly valuable.

Given this information loss, an informed Bayesian analysis is an attractive option for supplementing sample information. This paper describes a prior family that should be attractive in many situations and it depends on only four easily interpreted prior hyperparameters. Because identification in such models is often weak, what is needed is an informative prior with some public appeal, not a *noninformative* prior.

Inferential Implications of Over-Parameterization: A Case Study in Incomplete Categorical Data

Frederico Z. Poleto
Universidade de São Paulo
fpoleto@ime.usp.br

Carlos D. Paulino
Univ. Técnica de Lisboa, Portugal
dpaulino@math.ist.utl.pt

Geert Molenberghs
Katholieke Univ. Leuven, Belgium
geert.molenberghs@uhasselt.be

Julio M. Singer
Universidade de São Paulo, Brazil
jmsinger@ime.usp.br

The use of non-identifiable models has been spreading across many branches of statistics. In this work, we scrutinize Bayesian and classical analyses in the

incomplete categorical response setting. We show that the subjective parts of each approach can influence results in non-trivial ways. Specifically, we illustrate that prior distributions commonly regarded as slightly or non-informative may actually be too informative for non-identifiable parameters even when they appear in the likelihood. We also show that the choice of more parsimonious over-parameterized models, usually adopted in classical analyses merely for computational simplicity, may affect inferences in unforeseen ways. These effects remain for large sample sizes (e.g., from 4,000 to 4,000,000), suggesting that the prior distribution and the model should be more carefully examined than usual for these specialized models. When contrasting Bayesian and classical approaches we note that the posterior inferences dependence on the prior does not disappear asymptotically and that the respective credible and uncertainty intervals are close only for extremely diffuse priors and very large sample sizes.

Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction

Nicholas G. Polson
University of Chicago, USA
ngp@chicagobooth.edu

James G. Scott
University of Texas, USA
james.scott@mccombs.utexas.edu

We use Lévy processes to generate joint prior distributions for a location parameter $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ as p grows large. This approach, which generalizes normal scale-mixture priors to an infinite-dimensional setting, has a number of connections with mathematical finance and Bayesian nonparametrics. We argue that it provides an intuitive framework for generating new regularization penalties and shrinkage rules; for performing asymptotic analysis on existing models; and for simplifying proofs of some classic results on normal scale mixtures.

A Continuum of Inductive Methods for Discrete Universes of Uncertain Constitution and Cardinality

Stephan Poppe and Pierre-Yves Bourguignon
Max-Planck Gesellschaft, Germany
{stephan.poppe, pierre-yves.bourguignon}@mis.mpg.de

Carnap's continuum of inductive methods has set a milestone in the axiomatization of Bayesian inference, see Carnap(1952), providing the usage of Dirichlet priors with a robust foundation in the multinomial setting. It has nonetheless been repeatedly criticized on the ground that a.s. the data will ultimately confirm the universal hypothesis "all considered outcomes occur". In particular, it has been felt by several that inferences drawn within this framework are overly sensitive to the choice of the state space, restricting their relevance when data are sparse, or even more importantly, when little is known in advance about the state space itself. Capable of accomodating an infinite and a priori unknown universe, the Dirichlet process and extensions of it have been considered as an appealing alternative meeting those concerns. For instance, the Pitman process has been given

an inductive axiomatization by Zabell, see Zabell (2005), extending arguments which originally lead to Carnap's approach -namely Johnson's sufficientness postulate, *i.e.* the predictive probability of an outcome depends exclusively on the number of its past occurrences and the sample size. This latter assumption however entails that the number of different outcomes is almost surely unbounded, with the unpleasant consequence to prevent the induction of the actual cardinality of the sample space -a task sometimes termed "the sample species problem".

We propose here a framework that borrows much from previous results by Hintikka, Niiniluoto and Kuipers see Hintika & Niiniluoto(1980) and Kuipers (1978), who made use of a weaker version of the sufficientness postulate for extending the Carnap's continuum. The key relaxation here is to allow predictive probabilities to depend also on the number of different outcomes seen so far. Combining their continuum of inductive methods with the powerful results pertaining to partially exchangeable random partitions, see Pitman(1995), we reconstruct a continuum of inductive methods based on Hintikka-Niiniluoto-Kuipers' system, which encompasses Pitman-Zabell's. It is noteworthy that this is merely an implementation of ideas already present in Kuipers' program. This new continuum comes along with an appropriate set of prior parameters. The continuum's de Finetti-type representation involves an adequate multinomial model that appears in Pitman(1995). While generalized multinomial models proposed earlier to deal with the species sampling problem fell short of providing consistent inferences for the frequence of unseen outcomes, the Bayesian model associated with the present continuum overcomes this limitation. A particular choice of method from this continuum is further investigated, yielding a tractable inference calculus for which closed-form expressions and a straightforward stochastic representation are given.

Assessing Robustness Issues in Microarray Data Analysis

Alezandra Posekany
Universität für Bodenkultur Wien, Austria
alexandra.posekany@boku.ac.at

K. Felsenstein and P. Sykacek
Technische Universität Wien, Austria
klaus.felsenstein@tuwien.ac.at peter.sykacek@boku.ac.at

Microarray data analysis is a prominent application field for Bayesian modelling. Microarray assays are either analysed independently for important genetic players contributing to a biological process (Zucknick *et al.* (2003)), or in an integrated analysis considering its genomic context (Broet *et al.* (2004), Lucas *et al.* (2009), Sykacek *et al.* (2007)).

In spite of evidence that Gaussian noise might be an over-optimistic choice (Hardin and Wilson (2009)), statistical analyses of microarrays often rely on this assumption. This apparent controversy underlines the importance of a detailed investigation of this issue. We propose to this end extending Gottardo *et al.* (2006) and modelling individual transcription profiles by mixing a linear model, representing genetic activity and a constant, representing inactive genes. A finite set of a Gaussian distribution and certain t-distributions with suitably chosen

degrees of freedom defines the likelihood functions (Berger (1994)) which we assess via Bayes factors obtained from a hybrid Monte Carlo sampler.

Simulations lead us to carefully tuned hyper-parameters for minimising model sensitivity. All chains were furthermore analysed with CODA (Plummer *et al.* (2006)) for convergence. Analysing a wide range of microarray experiments, covering different experimental paradigms, model organisms and measurement platforms, we got the arresting result that *in all data sets* a t-distribution with small degrees of freedom is the noise model with the largest posterior probability. The importance of this finding is backed up by the observation that the choice of error model considerably influences the biological conclusions drawn from the analyses. We can therefore conclude that microarray residuals are heavy-tailed and ignoring this can have adverse effects on biological interpretations of microarray data.

Summary Statistics for Approximate Bayesian Computation

Dennis Prangle
Lancaster University, UK
d.prangle@lancaster.ac.uk

Paul Fearnhead
Lancaster University, UK

Approximate Bayesian Computation (ABC) methods are a family of algorithms for 'likelihood-free' Bayesian inference. The domain of use is models where numerical evaluation of the likelihood is impossible or impractical, but from which data can easily be simulated. For example, over the last decade ABC has allowed investigation of realistic but previously intractable models in population genetics. Other applications include infectious disease epidemiology and missing data models.

ABC operates by simulating data X_{sim} from the model of interest for many parameter values θ and constructing an approximation to the posterior from those θ values for which the associated X_{sim} closely matches the observations X_{obs}. Algorithms have been proposed which implement this idea within the frameworks of rejection and importance sampling, Markov Chain Monte Carlo and Sequential Monte Carlo. A key insight in past research is that to achieve practical acceptance rates, 'closeness of match' should be judged by some norm $||S(X_{\mathsf{sim}})-S(X_{\mathsf{obs}})||$ where $S(.)$ are low dimensional *summary statistics* of a data set. However the problem of how to choose S well is an open question in the literature.

This work presents a novel methodology to generate efficient summary statistics for ABC. Theoretical considerations suggests choosing S as a vector of parameter predictors, and we propose a computationally cheap method of constructing these by performing linear regression on a large set of simulated training data.

Multinomial Cluster Analysis of Electoral Data

Xavier Puig and Josep Ginebra
Universitat Politècnica de Catalunya, Spain
xavier.puig@upc.edu josep.ginebra@upc.edu

This work is motivated by the analysis of the results in Barcelona on the last five elections to the Parliament of Catalonia. The diversity of political parties that are represented in the catalan parliament leads to a very rich political map; Barcelona is organized in 248 small areas each training somewhere between 500 and 15,000 eligible voters. For each one of these areas we have the results of these elections categorized in either 8 or 9 groups depending on the elections. The first aim is to help develop tools to validate and compare Bayesian hierarchical and non-hierarchical multinomial mixture models in the context of the cluster analysis of electoral data. These tools rely mainly on the creative use of graphics, but it also uses permutation tests to check wether mixture models capture all the spatial dependence in the data or not. Non-hierarchical models end up being discarded because they do not capture the variability of data appropriately.

Through the use of Bayesian multinomial mixture models voting patterns are identified and described, the areas are allocated to each pattern of vote and the time evolution of these patterns and of the cluster structure is explored.

Forecasting Traffic Flow series: A Comparison of Dynamic Bayesian Approaches.

Catriona M. Queen
The Open University, UK
C.Queen@open.ac.uk

Casper J. Albers
Rijksuniversiteit Groningen, NL
c.j.albers@rug.nl

A Linear Multiregression Dynamic Model (LMDM) is a dynamic Bayesian network for multivariate time series in which the time series are represented by nodes on a directed acyclic graph (DAG). The DAG then breaks the multivariate time series model into univariate components, each of which is a Bayesian dynamic linear model.

An LMDM has been shown to be promising for forecasting multivariate time series of traffic flows. Traffic flow is measured as being the count of vehicles passing a particular data collection site in the road. As such, the traffic flow at a single site for a particular time period each week can be modelled by a Poisson distribution. The LMDM, however, assumes normality of the traffic flow series. Of course, when traffic flow is high, this assumption of normality is not a problem. But when flows are low, as they generally are in the middle of the night, for example, a model suitable for Poisson time series may perform better.

One way in which univariate dynamic linear models can be adapted for Poisson time series is by introducing a variance law into the model. This paper will introduce the same idea into the LMDM. The model's performance for forecasting traffic flows will be investigated using traffic flow data at a motorway junction in the UK.

Bayesian and Non-Bayesian Variable Selection for Parametric AFT Models in High Dimensions

Hasinur Rahaman Khan and J. Ewart H. Shaw
University of Warwick, UK
M.H.Rahaman-Khan@warwick.ac.uk Ewart.Shaw@warwick.ac.uk

Many alternative penalized regression approaches to simultaneous variable selection and coefficient estimation from very high-dimensional datasets have been suggested in recent years. Very few such approaches have been adapted for use with high-dimensional and low-sample size survival data. Among standard variable selection methods that have been shown both to have good predictive accuracy and to be computationally efficient is the elastic net penalization approach. Here we examine Bayesian and other adaptations of the elastic net applied to log-normal accelerated failure time (AFT) models. Performance is assessed through simulation studies and through analysis of microarray data.

A Nonparametric Bayesian Approach for Supervised Classification Based on pairs of covariates

Sandra Ramos
Politécnico do Porto, Portugal
sfr@isep.ipp.pt

Antónia Amaral-Turkman
Universidade de Lisboa, Portugal
antonia.turkman@fc.ul.pt

Marília Antunes
Universidade de Lisboa, Portugal
marilia.antunes@fc.ul.pt

In a previous work (Ramos, 2010) , we introduced a Bayesian optimal screening method (BOSc) for supervised classification, based on the observation of pairs of covariates, $\mathbf{X}$. To build the optimal classification region, a bivariate normal model is assumed for the population distribution of $\mathbf{X}$ in each group. However, the performance of the classifier is compromised when the data violate this assumption. In this work, we introduce a Bayesian nonparametric approach to provide more flexible and general classifiers. The approach makes use of multivariate mixtures of finite Polya trees prior (Hanson *JASA* 2006) to estimate the predictive densities of X in each group. Real and simulated data are analyzed and results are compared with those obtained using the parametric approach (BOSc). Finally, we make use of the versatility of the nonparametric Bayesian framework to build HPD intervals for the operational characteristics of the classification region.

Bayesian Integrated Genomics

Paola M.V. Rancoita
IDSIA, Switzerland
paola@idsia.ch

Marcus Hutter
Australia National University
marcus.hutter@anu.edu.au

Francesco Bertoni
IOSI, Switzerland
frbertoni@mac.com

Ivo Kwee
IOSI, Switzerland
ivo@idsia.ch

Tumors are largely related to chromosomal lesions. One type of aberration is the change of DNA copy number (CN) in one or more regions of the genome (the normal CN is two because we have two copies of each chromosome). Other genomic alterations can be displayed by unusual long stretches of homozygous single nucleotide polymorphisms (SNPs) in regions with normal CN (called copy-neutral loss of heterozygosity, LOH, regions). A SNP is a base-pair location in the genome that have two alleles and it can be classified as either homozygous, if its two copies consist of equal alleles, or heterozygous, otherwise. SNP-microarrays are able to measure simultaneously both homozygous status and CN at thousands SNPs. In this way, several types of genomic alterations can be observed and combined for a better identification of the events occurred.

We present the progress/findings of a 4-year project, which started in 2006, devoted to developing an integrated Bayesian analysis of genomic microarray data. We propose a Bayesian piecewise constant regression, to infer the genomic profile of aberrations occurred (normal state, copy-neutral LOH, high amplification, gain, loss of one copy, loss of two copies), taking into account all the possible influences in the microarray detection of the homozygous status, resulting from an altered CN level. We call our method genomic Bayesian piecewise constant regression (gBPCR). The novelty of gBPCR consists in modeling the complex biological relationship between these two types of genomic data, which allows us to better identify the genomic alterations occurred.

The comparison with two well-known methods for LOH estimation on artificial data showed that gBPCR outperformed the others on data with high noise. Moreover, some CN alterations identified by gBPCR on real data have been biologically validated.

Bayesian Approaches to the Analysis of Spectral Metabolic Profiling Data

Mattias Rantalainen
University of Oxford, UK
rantalai@stats.ox.ac.uk

George Nicholson and Chris Holmes
University of Oxford, UK
nicholso@stats.ox.ac.uk cholmes@stats.ox.ac.uk

Metabolic profiling using Nuclear Magnetic Resonance (NMR) spectroscopy is a well-established methodology for quantitative measurement of metabolite abun-

dances in biological samples. NMR spectroscopy enables concurrent measurement of hundreds of metabolites in a sample, represented as peaks in a spectral profile.

In recent years metabolic profiling has gained interest for applications in genomic epidemiology, where changes in the metabolic profile in response to a pathophysiological state are studied in an epidemiological context. Metabolic profiling in genomic epidemiology has the potential to enable novel biomarker discovery as well as providing a tool for disease diagnostics in humans. A valuable property of metabolic profiling is that samples can be collected with minimally invasive procedures, making the technology potentially applicable in a wide range of areas of both basic and clinical research.

Here we describe how Bayesian methods can be used for analysis of spectral data originating from NMR spectroscopy. Modelling of spectral NMR-based metabolic profiling data poses numerous statistical challenges including spectral alignment, peak extraction, normalisation, handling of structured noise and the presence of outlying observations. Improved inference of the components of variation in metabolic profiling data can improve the information recovery from the spectral profiles, provide means for biomarker discovery and improve the understanding of metabolic responses relating to disease.

Guiding Bayesian Model Choice within ABC

Oliver Ratmann
Duke University, USA
oliver.ratmann@duke.edu

Gareth Peters
Univ. New South Wales, Australia

Sylvia Richardson
Imperial College London, UK

Approximate Bayesian Computation (ABC) is a developing area of statistics that is motivated by the need to fit, assess and compare models that have a well-defined underlying generative process, but for which classical statistical procedures of estimation and goodness-of-fit encounter formidable computational challenges. ABC circumvents likelihood calculations by comparing model simulations with the data in terms of summary statistics.

To aid Bayesian model choice in this context, methods for model criticism and for model comparison have been introduced. We place these tools within the context of available Bayesian methodology and highlight their limitations and unique opportunities to Bayesian analysis.

We explored the application of sequential Bayesian tools that comprise a training and testing phase, such as (approximate) posterior or partial predictive densities. The extra volatility induced by simulating data sets typically offsets any gains in precision during the training phase, and we argue for using so-called simultaneous methods (Ratmann *et al.* 2009). ABC enables us to check several aspects of the data, and thereby opens up multi-dimensional avenues towards more accurate model choice (Ratmann *et al.* 2010). We discuss in detail the power of multivariate predictive error densities for model criticism.

We demonstrate the use of all tools on simple examples and real-world examples in finance, network modeling and virus evolution/ecology.

Multiscale Bayesian Reconstruction of Binary Permeability Fields from Sparse Static and Dynamic Data

Jaideep Ray
Sandia National Labs, USA
jairay@somnet.sandia.gov

Sean McKenna
Sandia National Labss, USA
samcken@sandia.gov

Bart van BloemenWaander
Sandia National Labs, USA
bartv@sandia.gov

Youssef Marzouk
MIT Boston, USA
ymarz@mit.edu

Binary fields are a useful representation of many heterogeneous porous media. In many practical situations a detailed measurement of the binary media is infeasible, and direct upscaling poses a challenge. Upscaled properties may instead be inferred from indirect measurements at the macroscale. Using a link function to connect disparate scales together, one may generate realizations of the microscale binary field, conditioned on all available data.

We present an Bayesian method for performing a stochastic multiscale reconstruction of binary field representations of heterogeneous porous media. We model a 2D porous medium as a matrix with high permeability inclusions whose proportion varies in space. We use truncated Gaussian random fields as a flexible geometrical model of the inclusions. We develop a subgrid model to upscale binary permeability fields to a coarse (block) scale effective permeability. The bivariate Gaussian kernels are anisotropic to replicate the directional nature of the permeability tensor. The inputs into the model are the proportion of inclusions at the block scale, the aspect ratio and the axial size of the kernels and the two modal permeabilities; the model returns permeability values in orthogonal directions as outputs.

We employ this subgrid model in an inversion process, where the proportionality and geometry of the latent binary field is inferred from sparse, macroscale observations of permeability and tracer recovery data. The variogram of the proportionality field is assumed known, and the field is represented by its Karhunen-Loève (KL) expansion. The coefficients of the KL modes, and the geometrical parameters of the Gaussian kernel are the objects of inference and we develop a joint posterior distribution from noisy static (macroscale proportionality) and dynamic (tracer breakthrough) data.

We subject the reconstructed binary field to posterior predictive tests by generating posterior realizations of the binary field, simulating tracer recovery tests and comparing them with a predetermined "ground truth". We also investigate the sensitivity of the posterior distribution to different combinations of observational data. The correlation between tracer breakthrough times and the posterior distribution of the KL mode coefficients is analyzed to evaluate the ability of individual tracer recoveries to inform on various spatial scales.

Estimating Risks of Identification Disclosure in Partially Synthetic Data

Jerome P. Reiter
Duke University, USA
jerry@stat.duke.edu

Robin Mitra
University of Southampton, UK
R.‘‘@soton.ac.uk

To limit disclosures, statistical agencies and other data disseminators can release partially synthetic, public use microdata sets. These comprise the units originally surveyed, but some collected values, for example sensitive values at high risk of disclosure or values of key identifiers, are replaced with multiple draws from statistical models. Because the original records are on the file, there remain risks of identifications. In this paper, we describe how to evaluate identification disclosure risks in partially synthetic data, accounting for released information from the multiple datasets, the model used to generate synthetic values, and the approach used to select values to synthesize. We illustrate the computations using the Survey of Youths in Custody.

Bayesian Adjusted R^2 for the Meta-Analytic Evaluation of Time-to-Event Surrogate Endpoints in Clinical Trials

Lindsay A. Renfro
Baylor University, USA
lindsay_renfro@baylor.edu

Bradley P. Carlin
University of Minnesota, USA
brad@biostat.umn.edu

Surrogate endpoints are often desired in clinical trials when the primary endpoint is costly to obtain, difficult to measure, or requires a long period of follow-up to observe. Many have noted the benefits of multi-trial evaluation of potential surrogates, namely increased sample size and generalizability to future trials, but few have discussed performing these evaluations in a both meta-analytic and Bayesian context. Moreover, these existing approaches are limited to summarized trial data or normal endpoints. Burzykowski et al. (2001) introduced a two-stage model for the evaluation of both trial-level and patient-level surrogacy in the setting of time-to-event endpoints, but their maximum likelihood approach often suffered from numerical problems once perfect estimation of treatment effects and common baseline hazards across trials were no longer assumed. We propose bringing the advantages of the two-stage hierarchical setting into a Bayesian framework, where we may naturally borrow strength across trials in the evaluation of potential surrogate survival endpoints with these more realistic assumptions. Mixed models are used to capture the association of the treatment effects on time-to-event endpoints at the trial level, while the bivariate nature of the true and surrogate endpoints at the individual level is captured by copulas. Posterior distributions on measures of trial-level surrogacy may then be used for comparison of potential surrogates or decision-making regarding the candidacy of a specific endpoint. We perform a simulation study to assess the sensitivity of trial-level surrogacy measures to chosen prior distributions as well as data characteristics including number of trials, trial size, and amount of censoring. Furthermore, we present a Bayesian evaluation of the surrogacy of popular secondary endpoints (TTR, DFS) for overall survival in a meta-analysis of adjuvant therapy trials in colon cancer (Sargent *et al.* 2005).

Correlated GMRF Priors and INLA

Andrea Riebler
Universität Zürich, Switzerland
andrea.riebler@ifspm.uzh.ch

Leonhard Held
Universität Zürich, Switzerland
leonhard.held@ifspm.uzh.ch

Håvard Rue
Norges Tek. Natur. Univ., Norway
Havard.Rue@math.ntnu.no

Birgit Schrödle
Universität Zürich, Switzerland
birgit.schroedle@ifspm.uzh.ch

In public health research, longitudinal or spatio-temporal health outcomes are commonplace. For full Bayesian inference of such data, Gaussian Markov random field (GMRF) priors are extensively used. However, if multiple observations are available for each unit and time-point, standard GMRF priors do not account for a possible dependence between the outcomes. For example in the analysis of multinomial logistic regression, unknown non-parametric trends of individual (relative) risk probabilities are usually independently smoothed although a joint reference category is used. A correlated GMRF model is more appropriate here. We will present a novel framework based on correlated GMRF priors that combines different choices for the correlation structure and different choices for the GMRF model using a Kronecker product formulation. This framework provides a unique and flexible approach for multivariate modelling of dependent (count) data. Algorithmic routines can be based on either Markov chain Monte Carlo or integrated nested Laplace approximations (INLA) (Rue, Martino and Chopin (*JRSSB* 2009), INLA is particularly attractive for this framework because it is already programmed in a modular fashion, so that a wide range of latent GMRF models (e.g. random walks, models for seasonal variation, models with user-defined structure matrix, etc.) can be correlated as components of structured additive regression models. The methodology will be illustrated with applications to rates with several time scales as well as to space-time counts.

Modelling the Sustainability of Nuclear Power

Jesus Rios and Simon French
Manchester Business School, UK
jesus.rios@mbs.ac.uk simon.french@mbs.ac.uk

Growing concerns about environmental, economic and social issues associated with energy production including climate change, the depletion of hydrocarbon resources, declining public trust in science and technology and increasing energy prices have led the UK government to a reappraisal of its energy policy and of individual power generation technologies. We present here the modelling issues facing a UK project to evaluate the sustainability of nuclear power within a variety of energy mixes and indicate how a full Bayesian multi-attribute decision model integrated with scenario analysis techniques may be formulated to do so. Our underpinning problem is complex and large, with consequences extending 50 years and longer into the future and the analysis needs to support the perspectives of many stakeholders. Our modeling will allows us to explore and assess the

consequences over the next 50 years associated with interventions in electricity generation capacity taking into account relevant technical, economic, environmental, social and governance-related criteria as well as the associated uncertainties. For the value part of the problem, a model has been developed to compute the environmental, economical and technical consequences associated to every intervention in the electricity system. We link these measurable consequences with the attributes we identify from stakeholder. We use multi-attribute values and utilities to represent each stakeholderís preferences. We use scenario analysis to deal with the uncertainty part of the problem, using subjective probabilities to represent uncertainties within each scenario. However, the final objective of the analysis is the identification of actions that are robust across the all scenarios and in which all stakeholder may agree. Our approach integrates then multi-attribute value models with scenario analysis, allowing for Bayesian modeling with each scenario. Finally, we discuss how our evaluation framework may be used to support a debate among multiple stakeholders and decision makers on the sustainability of nuclear power.

On Resolving the Savage-Dickey Paradox

Christian P. Robert
Université Paris-Dauphine, France
xian@ceremade.dauphine.fr

Jean-Michel Marin
Université Montpellier 2, France
jean-michel.marin@math.univ-montp2.fr

The Savage-Dickey ratio is a specialised representation of the Bayes factor

$$B_{01} = \frac{\int \pi_0(\omega_0) f_0(x|\omega_0)\, \mathrm{d}\omega_0}{\int \pi_1(\omega_1) f_1(x|\omega_1)\, \mathrm{d}\omega_1} = \frac{m_0(x)}{m_1(x)}\,.$$

(Dickey, 1971, *Ann. Statist.*, O'Hagan and Forster, 2004) that allows for a functional plugging approximation of this quantity as

$$B_{01} = \frac{\pi_1(\theta_0|x)}{\pi_1(\theta_0)}\,,$$

with the obvious notations $\pi_1(\theta) = \int \pi_1(\theta, \psi)\mathrm{d}\psi$, and $\pi_1(\theta|x) = \int \pi_1(\theta, \psi|x)\mathrm{d}\psi$. Dickey's identity thus reduces the Bayes factor to the ratio of the posterior over the prior marginal densities of θ under the alternative model only, when taken at the tested value θ_0. However, it is based upon the assumption that the conditional prior density of ψ under the alternative model, given $\theta = \theta_0$, $\pi_1(\psi|\theta_0)$, is equal to the prior density under the null hypothesis, $\pi_0(\psi)$, $\pi_1(\psi|\theta_0) = \pi_0(\psi)$.

We demonstrate here (and the arXiv document 0910.1452) that this identity is a generic approximation method instead of a special case imposing the constraint on the prior distributions, while incidentally clarifying the measure-theoretic bases

of the method. We show in particular that Dickey's identity can be produced by simulating (directly or by MCMC) the pseudo-posterior

$$\tilde{\pi}_1(\theta, \psi|x) = \frac{\pi_0(\psi)\pi_1(\theta)f(x|\theta,\psi)}{\tilde{m}_1(x)}\,,$$

where $\tilde{m}_1(x)$ is the normalising constant. In this representation, the Monte Carlo approximation

$$\widehat{B_{01}} = \frac{1}{T}\frac{\sum_t \tilde{\pi}_1(\theta_0|x, z^{(t)}, \psi^{(t)})}{\pi_1(\theta_0)} \times \frac{1}{T}\sum_{t=1}^{T}\frac{\pi_0(\bar{\psi}^{(t)})}{\pi_1(\bar{\psi}^{(t)}|\bar{\theta}^{(t)})}$$

is an unbiased estimator of the Bayes factor. We thus provide a general framework for approximating the Bayes factor, which is furthermore unrelated with the earlier approach of Verdinelli and Wasserman (*JASA*, 1995).

A Bayesian Destructive Weighted Poisson Cure Rate Model and an Application to Cutaneous Melanoma Data

Josemar Rodrigues
Univ. Federal de São Carlos, Brazil
vjosemar@ufscar.br

Mário de Castro
Univ. de São Paulo, Brazil
mcastro@icmc.usp.br

Vicente G. Cancho
Univ. de São Paulo, Brazil
garibay@icmc.usp.br

N.Balakrishnan
McMaster University, Canada
bala@univmail.cis.macmaster.ca

In this paper we propose a new Bayesian flexible cure rate survival model which generalizes the stochastic model formulated by Klebanov *et al.* (1993). This model has much in common with the destructive model formulated by Rodrigues *et al.* (2009). In our approach the number of lesions, or altered cells, accumulated follows a compound weighted Poisson distribution. This model is more flexible in terms of dispersion than the promotion time cure model. Moreover, it gives an interesting and realistic interpretation of the biological mechanism of the occurrence of the event of interest as it includes a destructive process of tumor cells after an initial treatment or the capacity of an individual exposed to irradition to repair altered cells that results in cancer induction. In other words, what is recorded is only the damaged portion of the original number of altered cells not eliminated by the treatment or repaired by the repair system of an individual. Markov Chain Monte Carlo (MCMC) methods are used to develop a Bayesian procedure for the proposed model. Also, some discussions on the model selection and an illustration with a cutaneous melanoma data set described in recent paper (Rodrigues *et al.* 2009) are considered.

Dynamic Blockmodels for Social Networks through Bayesian Nonparametric Mixtures

Abel Rodriguez
University of California, Santa Cruz, USA
abel@soe.ucsc.edu

Blockmodels, which explain relationships among actors through their membership to unobserved groups/factions, are a popular tool in social network analysis. This poster develops a novel class of dynamic blockmodels using Bayesian nonparametric mixtures; the approach allows us to automatically estimate the number of groups as well as to identify changes in group membership. The models are illustrated using simulations as well as two real datasets.

Reference Priors for Prediction.

Lizbeth Román
Universitat de València, Spain
lizbeth.roman@uv.es

Objective Bayesian methods need priors which only depend on the assumed model and the quantity of interest. A general solution is provided by *reference analysis* (Bernardo 1979, Berger and Bernardo 1992, Bernardo 2005, Berger, Bernardo and Sun 2009), when the quantity of interest is some function of the parameters. However, it is not clear how to proceed when the quantity of interest is not an explicit function of the parameters but some function t of future observations. For prediction of a future observation x, Berger and Bernardo (1992) suggested one might use the conditional mean of the assumed model as the quantity of interest. Their proposal may be extended to consider as the parameter of interest some function $\phi(\boldsymbol{\theta}) = g[p_x(\cdot|\boldsymbol{\theta})]$ of the assumed model, where g can be any estimator of the function to be predicted such as the mean $\mathrm{E}(t|\boldsymbol{\theta})$, the mode $\mathrm{Mo}(t|\boldsymbol{\theta})$ or the median $\mathrm{Me}(t|\boldsymbol{\theta})$.

On the other hand, one way to *discriminate* among alternative priors is to analyze the coverage properties of the resulted posterior predictives. The use of coverage properties for discriminating among priors has a long history (Welch and Peers, 1963; Berger and Bernardo 1989; Datta and Sweeting, 2005).

In our work, we propose the use of the median $\mathrm{Me}(t|\boldsymbol{\theta})$ as the more appropriate choice of the parameter of interest for prediction, and justify this choice by the adequate coverage properties of the resulting posterior predictive distributions. The theory is illustrated with both one-parameter and two-parameter models. Besides, a generalization of the result by Jeffreys (1961) in prediction within location-scale models is provided, analyzing the probability that a new observation lies within the minimum and the maximum observations in a random sample of arbitrary size $n \geq 2$.

Comparison of Fully Bayesian and Empirical Bayes approaches for joint Bayesian model selection and estimation of sinusoids via reversible jump MCMC

Alireza Roodaki Julien Bect and Gilles Fleury
SUPELEC, Gif-sur-Yvette, France
{alireza.roodaki, julien.bect, gilles.fleury}@supelec.fr

This work addresses the sensitivity of the algorithm proposed by Andrieu and Doucet (IEEE T. Signal Proces., 47(10), 1999), for the joint Bayesian model selection and estimation of sinusoids in white Gaussian noise, to the values of a certain hyperparameter claimed to be weakly influential in the original paper. On the basis of extensive numerical experiments, we argue on the contrary that the value of this hyperparameter (the scale parameter of the expected signal-to-noise ratio) has a strong influence on 1) the mixing rate of the Markov chain and 2) the posterior distribution of the number of components. Fully Bayesian and Empirical Bayes methods are proposed and compared for estimating an appropriate value for this hyperparameter. In the Fully Bayesian approach, a weakly informative proper prior is assigned over that hyperparameter. In the Empirical Bayes approach, marginal likelihood maximization is performed by means of an importance sampling-based Monte Carlo EM (MCEM) algorithm. The pros and cons of each method are discussed on the basis of numerical experiments conducted with several sample sizes and signal-to-noise ratios.

Non-local Priors for Variable Selection

David Rossel
Instituto de Investigación Biomédica, Barcelona, Spain
david.rossell@irbbarcelona.org

Valen Johnson
M.D. Anderson Cancer Cen., USA
vejohnson@mdandersonn.org

Donatello Tellesca
UC Los Angeles, USA
dtelesca@ucla.edu

Bayesian model selection is an appealing choice to perform variable selection, as Bayes factors penalize model complexity and therefore favor parsimony. However, the penalty imposed by Bayes factors derived from most currently popular approaches is relatively mild. This results in unbalanced rates of accumulation of evidence, *i.e.* as the sample size grows, Bayes factors increase at an exponential rate for variables which should be included in the model, but they decrease only at an inverse square root rate for variables which should be excluded. In practice this means that, for small up to moderately large sample sizes, it is usually hard to obtain conclusive evidence for excluding any one variable.

We propose a framework based on non-local priors, an intuitively appealing alternative which specifies a better *a priori* separation of the model space, and therefore favors the exclusion of unnecessary variables. As shown in Johnson and Rossell (*JRSSB* 2010), our framework achieves better balanced convergence rates and a highly improved performance in a wide variety of scenarios. For instance, in

sequential clinical trials there are substantial savings in sample sizes, as ineffective treatments are discarded earlier. In high-dimensional setups such as prediction or functional data analysis, the strong model complexity penalty imposed by non-local priors results in improved inference. Importantly in such high-dimensional setups, our approach is computationally efficient, as we provide closed-form expressions and approximations for linear models and generalized linear models, respectively.

Asymptotic Behaviour of the Posterior Distribution in Mixture Models with too many Components

Judith Rousseau
Université Paris Dauphine, France
rousseau@ceremade.dauphine.fr

Kerrie Mengersen
Queensland UT, Australia
k.mengersen@qut.edu.au

In this work we study the asymptotic behaviour of the posterior mean estimate of the parameters when there are too many parameters. In this case it is well known that the maximum likelihood does not have a stable behaviour since the true distribution corresponds to an infinite number of different parameters (non identifiability). Interestingly we prove that the posterior mean estimate can have a much more stable behaviour. More specifically, consider a mixture model in the form :

$$p(x|\theta) = \sum_{i=1}^{k} p_i g(x|\gamma_i) \quad \gamma_i \in \Gamma \quad \sum_i p_i = 1$$

and assume that the true distribution corresponds to a mixture model in the above form with k_0 components and $k_0 < k$. We prove that if the dimension of $\gamma > 2$ the posterior distribution *empties the extra components when the number of observations increases* i.e.

$$P^{\pi}\left[\sum_{j=k_0+1}^{k} p_j > \epsilon | X^n\right] = o_p(1),$$

which implies the consistence of the Bayesian estimate. On the contrary if the dimension of the γ_j's is equal to 1 then the posterior distribution forces the extra groups to merge with *existing true groups* forbidding small weights and when $d = 2$ no specific asymptotic result is to be expected unless some other constraints in the prior are added.

This result can be explained by the fact that even for estimation, integration acts as penalization . This results implies that in the case where the dimension of the parameters γ_j is larger than 2 and the number of components is too large, Bayesian estimates of the parameters still have nice interpretations and have a stable behaviour.

Reference Prior for the Shape Parameter of the Exponential Power Distribution

Raúl Rueda and Eduardo Gutiérrez-Peña
IIMAS-UNAM, Mexico
pinky@sigma.iimas.unam.mx eduardo@stats.iimas.unam.mx

The exponential power family of distributions (Box & Tiao, 1962), also known as "generalized Gaussian", has played an important role in the study of Bayesian robustness and has been used in applied fields such as engineering.

We derived the reference prior for the shape parameter of the exponential power distribution using the algorithm proposed by Berger and Bernardo (1992), and compare it with other default priors proposed in the literature for this model.

Kullback-Leibler Divergence to Pool Prior Ppinions

María-Jesús Rufo
Universidad de Extremadura, Spain
mrufo@unex.es

Jacinto Martín and Carlos Pérez
Universidad de Extremadura, Spain
jrmartin@unex.es carper@unex.es

This work provides a general Bayesian approach to analyze a class of distributions when several sources of prior information are available from experts. Each prior belief is elicited as a conjugate prior distribution and all of them are merged through logarithmic pooling. An inconvenience when using this method is determining the vector of weights. In this context, several authors have provided different interpretations for this vector (see, for example, Abbas, A. E., A Kullback-Leibler View of Linear and Log-Linear Pools, Decision Analysis, 6 (1), 2009, and references therein).

Bousquet (Diagnostic of prior-data agreement in applied Bayesian analysis, 35 (9), 2009) proposed a criterion that can be used as a calibration tool when subjective priors are used. Overall, an expert's calibration is made by comparing the expert's prior distribution with the posterior distribution for a non informative prior distribution.

Based on the previous paper, a novel general method is proposed. It uses an expected Kullback-Leibler divergence to obtain the vector of weights in the combined prior distribution. Specifically, the combined prior distribution that minimizes the expectation with respect to the corresponding predictive prior distribution is found. Therefore, a pooled prior distribution is achieved for which the expected calibration is the best one. Besides, a direct implementation for all considered distributions is possible.

Next, the posterior opinion for each expert is calibrated by analyzing Kullback-Leibler divergences that involve the posterior distribution used in the previous method. This procedure leads to analytical solutions. Finally, some illustrative examples are presented.

Hyper-g Priors for Generalized Linear Models

Daniel Sabanés and Leonhard Held
University of Zurich, Switzerland
{daniel.sabanesbove, leonhard.held}@ifspm.uzh.ch

Hyper-g priors have been proposed by Liang et al. (2008, *JASA*, **103**: 410–423) for Bayesian variable selection in the normal linear model. Their approach avoids the choice of a fixed covariance factor g required for the conventional g-prior while maintaining computational tractability through the specific form of the prior on g. Recently, we have used the hyper-g prior for Fractional Polynomial (FP) modelling (Sabanés Bové and Held 2010, *Statistics and Computing*, to appear): Our approach combines variable selection with automatic parametric modelling of additive effects. The use of a Markov chain Monte Carlo algorithm for the exploration of the model space constitutes a substantial improvement to widely used stepwise procedures for the selection of FP models. The R-package `bfp` (available at `http://r-forge.r-project.org/projects/bfp`) implements the methodology. Here we present an extension of the hyper-g prior to the class of generalized linear models. The key for efficient model space exploration is a well-behaving numerical approximation, since the marginal likelihood cannot be computed analytically in non-Gaussian models. To this end we use an integrated Laplace approximation, which was proposed by Rue et al. (2009, *JRSS-B*, **71**: 319–392). In a second step, a tuning-free Metropolis-Hastings sampler provides posterior parameter samples in a given model. The resulting MCMC approximation of the marginal likelihood (Chib and Jeliazkov 2001, *JASA*, **96**: 270–281) allows assessing the accuracy of the numerical approximation. The prior on the regression coefficients is more flexible than the Liang et al. (2008) proposal as any prior on g can be used. We illustrate the methodolgy in the FP modelling context with data sets previously analysed in the literature.

Bayesian Classification of Partially Observed Outbreaks using Time-Series Data

Cosmin Safta Jaideep Ray
Sandia National Laboratories, Livermore, CA, USA
{csafta,jairay}@sandia.gov

Karen Cheng and David Crary
Applied Research Associates, Inc., Arlington, VA, USA
{kcheng,dcrary}@ara.com

In many cases sufficient/suitable laboratory facilities may not be present to identify the etiologic agent of an outbreak in a timely manner. This can result in underestimation of the scope of the problem and lead to disastrous outcomes. We propose a novel statistical method to identify the etiologic agent/pathogen of an ongoing outbreak based solely on a time series of observed morbidity. Simple statistical models were developed based on incubation and visit delay distributions that captures the dynamics of ILI epidemics. The unknown parameters in

the models, *i.e.*, the number of infected people, time of infection, dosage, and a measure of the visit delay, are obtained through Bayesian inference using the observed time-series. Joint posterior distributions for the model parameters for all competing etiological agents are developed using MCMC. The classification problem for the competing statistical models presents some significant challenges, since observations are typically noisy, as they include background morbidity. Further, there exists a stochastic visit delay between the appearance of symptoms and their reporting; thus, the reported values from the recent past are merely a lower bound on the actual levels of morbidity. MLE and Bayes factors are used to choose the most probable agent. This methodologies are tested for a simulated inhalational anthrax attack which is corrupted with a stochastic visit delay and background morbidity. Inhalational anthrax and plague are proposed as competing etiological agents. Our results indicate that it may be possible to perform the classification correctly with about one week's worth of data; this may include data collected before the identification of anomalous morbidity levels which would typically trigger a classification exercise. Our method is very sensitive to the completeness of the set of competing hypotheses. It is sufficiently accurate to evaluate the demand on a medical system, but should not be used to determine the medical treatments since very different pathogens may have similar epidemiological signatures.

Variable Selection and Functional Form Uncertainty in Cross-Country Growth Regressions

Tim Salimans and Richard Paap
Erasmus Universiteit Rotterdam, The Netherlands
salimans@ese.eur.nl paap@ese.eur.nl

Regression analyses of cross-country economic growth data are complicated by two main forms of model uncertainty: the uncertainty in selecting explanatory variables and the uncertainty in specifying the functional form of the regression function. Most discussions in the literature address these problems independently, yet a joint treatment is essential. We perform this joint treatment by extending the linear model to allow for multiple-regime parameter heterogeneity of the type suggested by new growth theory, while addressing the variable selection problem by means of Bayesian model averaging. Controlling for variable selection uncertainty, we confirm the evidence in favor of new growth theory presented in several earlier studies. However, controlling for functional form uncertainty, we find that many of the explanatory variables identified in the literature do not robustly explain economic growth.

The Persistence of Abnormal Returns at Industry and Firm Levels: A Bayesian Approach

Manuel Salvador Pilar Gargallo
Universidad de Zaragoza, Spain
salvador@unizar.es pigarga@unizar.es

José Luis Gallizo and Ramó Saladrigues
Universitar de Lleida, Spain
gallizo@aegern.udl.es ramon.saladrigues@aegern.udl.cat

In this paper we carry out a statistical analysis of the abnormal returns in a sample of firms of the EU. To that aim, we use a Bayesian dynamic model that let us to decompose the annual evolution of the firm returns in a permanent, a transitory and an idiosyncratic component also distinguishing between industry and firm effects. We analyse homogeneity patterns taking into account the sector and country of the firm and using Bayesian models comparison tools.

Bayesian Methods for Discontinuity Detection in Climate Model Predictions

Khachik Sargsyan Cosmin Safta
Bert Debusschere and Habib Najm
Sandia National Laboratories, Livermore, CA, USA
{ksargsy,csafta,bjdebus,hnnajm}@sandia.gov

Discontinuity detection is an important component in many fields, *e.g.*, climate change research, image recognition, or digital signal processing. However, current methods for discontinuity or edge detection are often (a) restricted to one- or two-dimensional settings, (b) require uniformly spaced, and typically quite dense data collection, and (c) lead to a deterministic, fixed answer without quantifying the confidence in the result. Further, in predictive model analysis, response surface methods for surrogate modeling are strongly challenged by nonlinear or discontinuous output data, since global, smooth-basis methods require high order expansions and exhibit Gibbs phenomena. While domain refinement methods can reduce the impact of nonlinearities and jumps, they often require prohibitively expensive sampling in each subdomain.

We propose a probabilistic, Bayesian framework of discontinuity detection that parameterizes and infers the discontinuity location together with its associated uncertainties. Namely, adaptive Markov Chain Monte Carlo sampling is used to obtain joint distributions for the discontinuity location parameters, effectively leading to a distribution over all possible discontinuity curves. The methodology can be generalized to multiple dimensions and is robust with respect to limited and arbitrarily distributed data. Discontinuity detection then enables effective domain decompositions in the construction of response surfaces for systems with discontinuous behavior so that smooth approximations can be used on both sides of the discontinuities, allowing uncertainty propagation from input parameters to the output distributions using spectral methods for each side of the discontinuity.

Moreover, since the Bayesian approach leads to the full posterior distribution of the discontinuity location, the expectation of spectral expansions with respect to this distribution leads to a compact representation of the output of interest, thus effectively accomplishing the uncertainty propagation.

We will illustrate our methodology on the example of the Meridional Overturning Circulation (MOC) in the Atlantic Ocean. It is known that the maximum stream function of the MOC exhibits discontinuity across a curve in the space of two uncertain parameters: the climate sensitivity and the CO_2 forcing rate. The proposed methods prove very efficient in this context compared to other discontinuity detection algorithms, since uncertainty quantification in climate models is challenged by the sparsity of the available model realizations given the high computational cost of the model runs.

Bayesian Approach to Population PKPD Modelling: Advantages and Drawbacks

Marina Savelieva
Novartis Pharma, Basel, Switzerland
marina.savelieva@novartis.com
Ivan Demin and Beat Neuenschwander
Novartis Pharma, Basel, Switzerland

Population Pharmacokinetic-Pharmacodynamic (PKPD) modelling is becoming increasingly important in the drug development to support the dose finding and labeling throughout all development phases. The approach is complex, since compartment models described in terms of Ordinary Differential Equations (ODEs) are coupled with the Non-Linear Mixed Effects (NLME) models. Population PKPD models allow, for example, to relate a dose given to a patient to a response of interest and provide estimates of the population and individual means as well as between- and within- patient variability. The maximum likelihood method for parameter estimates is commonly used in this situation. To allow for evaluation of the likelihood function in the presence of random effects, appro- ximations to the maximum likelihood function are employed. Moreover, for technical and computational reasons, a so-called two-step analysis (instead of the simultaneous approach) to population PKPD data is performed, where the PK parameters are first estimated and then these estimates are plugged into the PD analysis. This approach, while computationally feasible, ignores the uncertainty about PK parameters, resulting in underestimation of the PD parameters variability. Moreover, to construct better calibrated confidence intervals, a computationally-intensive bootstrapping procedure becomes necessary. In such a context, the Bayesian approach is comparable computationally and, in addition, allows for a better assessment and propagation of uncertainty from the PK to the PD model. However, when a complex PKPD model (e.g. one described by a system of non-linear ODEs) is used to model real clinical data, additional challenges (prior choice, MCMC convergence diagnostics, etc) arise. Different ways of handling sequential PKPD were discussed in Zhang, Beal and Sheiner (J Pharmacokinet Pharmacodyn, 2003), and in a Bayesian context in Lunn *et al.* (J Pharmacokinet Pharmacodyn, 2009).

However, those examples were restricted to relatively simple models, with solutions available in closed form. In the present work, we implement a Population PKPD model described by a system of non- linear ODEs in a Bayesian Framework using the BUGSModelLibrary from Metrum Institute. A two-step PKPD approach is considered, and the uncertainty about PK parameters is taken into account when fitting the PD model. We address the following questions: what is the best way to relate PK and PD models, how do we use PK and PD data, and how do we integrate our confidence about PK and PD models into this process? Moreover, sensitivity of the model to the prior choice for individual parameters as well as variance-covariance matrices is discussed.

Using Integrated Nested Laplace Approximation for Modeling Spatial Healthcare Utilization

Erik Sauleau
Université de Strasbourg, France
ea.sauleau@unistra.fr

Valentina Mameli and Monica Musio
Università degli studi di Cagliari, Italy
vmameli@unica.it mmusio@unica.it

In recent years, spatial and spatio-temporal modeling have become an important area of research in many fields (epidemiology, environmental studies, disease mapping). In this work we propose different spatial models to study hospital recruitment, including some potentially explicative variables. Interest is on the distribution per geographical unit of the ratio between the number of patients living in this geographical unit i, say O_i, and the expected number of patients, E_i in the same unit. Models considered are within the framework of Bayesian Latent Gaussian models (Fahrmeir and Tutz 2001). Our response variable O_i is assumed to follow a binomial distribution, with logit link, whose parameters are the population at risk N_i in the geographical unit i and the corresponding relative risk π_i. The structured additive predictor η_i accounts for effects of various covariates in an additive way: $\eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}(\boldsymbol{u}_{ji}) + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \epsilon_i$. Here, the $f^{(j)}(\cdot)$s are unknown functions of the covariates $\boldsymbol{u}$ (including also a spatial effect) the β_ks represent the linear effect of covariates $\boldsymbol{z}$ and the ϵ_is are unstructured terms. To approximate posterior marginals, which not available in closed form, we use integrated nested Laplace approximations (INLA) (Rue, Martino and Chopin,2009, JRSSB), recently proposed for approximate Bayesian inference in latent Gaussian models. INLA has the advantage to give very accurate approximations and to be faster than McMC methods when the number of parameters does not exceed 6 (as it is in our case). Model comparisons are assesses using DIC criterion (Spiegelhalter *et al.* 2002, JRSSB).

Testing Variance Components in Multilevel Linear Models using Approximate Bayes Factors

Benjamin R. Saville
Vanderbilt University School of Medicine, USA
b.saville@vanderbilt.edu

Amy H. Herring
Univ. of NC at Chapel Hill, USA

Jay S. Kaufman
McGill University, USA

Racial/ethnic disparities in birthweight are a large source of differential morbidity and mortality worldwide and have remained largely unexplained in epidemiologic models. We assess the impact of maternal ancestry and census tract residence on infant birth weights in New York City and the modifying effects of race and nativity by incorporating random coefficients in a multilevel linear model. Evaluating the significance of these predictors involves the test of whether the variances of the random coefficients are equal to zero. This is problematic because the null hypothesis lies on the boundary of the parameter space. We extend an approach for the linear mixed model to multilevel models by scaling the random coefficients to the residual variance and introducing parameters that control the relative contribution of the random coefficients. After integrating over the random coefficients and variance components, the resulting integrals needed to calculate the Bayes factor can be efficiently approximated with Laplace's method

Cross-Entropy and Sequential Monte Carlo Methods for Large Scale Bayesian Model Choice Problems

Christian Schäfer
ENSAE, CREST, France
christian.schafer@ensae.fr

We propose two methods to assess the mode and the marginal distribution of the posterior of large Bayesian variable selection problems. Typically, for large problems, the modes of the posterior are poorly connected and Markov chains easily get stuck when exploring the solution space. Therefore, we refrain from pure Makov chain Monte Carlo techniques and rather suggest an adapted version of Rubinstein's cross-entropy method to find the most likely model. Further, we provide a static sequential Monte Carlo algorithm which produces reasonable approximations to the marginal distributions of the posterior.

Raw versions of both the cross-entropy method and sequential Monte Carlo are easily implemented using vectors of independent binary variables as proposal distributions of the posterior. However, the true key to efficient algorithms are more complex binary parametric families that take the dependencies between components of the posterior into account. Our main contribution is the construction of such binary parametric families which also allow to be quickly estimated and easily sampled from. Extensive computational studies on real life data with a few hundred covariates prove the superiority of these binary families when incorporated in cross-entropy or sequential Monte Carlo methods.

Modelling Multivariate Counts Varying Continuously in Space

Alexandra M. Schmidt
Universidade Federal do Rio de Janeiro, Brazil
alex@im.ufrj.br
Marco A. Rodríguez
Université du Québec à Trois-Rivières, Canada
marco.rodriguez@uqtr.ca

We discuss models for multivariate counts observed at fixed spatial locations of a region of interest. Our approach is based on a continuous mixture of independent Poisson distributions. The mixing component is able to capture correlation among components of the observed vector and across space through the use of a linear model of coregionalization. We introduce here the use of covariates to allow for possible non-stationarity of the covariance structure of the mixing component. We analyze joint spatial variation of counts of four fish species abundant in Lake Saint Pierre, Quebec, Canada. Models allowing the covariance structure of the spatial random effects to depend on a covariate, geodetic lake depth, showed improved fit relative to stationary models.

A Multivariate Hierarchical Poisson Approach to Construct Candidate Lists of Sites for Possible Improvements

Kristian Schmidt and Alicia Carriquiry
Iowa State University, USA
kristian@iastate.edu alicia@iastate.edu

Improving traffic safety is a priority of Departments of Transportation nationwide. Because every agency typically faces budgetary constraints, methods to identify those sites with the highest risk potential and that may respond to interventions are of special importance. New and efficient ways to produce lists of candidate sites can help save money and lives and could increase safety for everybody. The initial objective of this work is to develop a reliable and accurate approach that can be used to construct candidate lists of intersections for future improvement. To do so, the methodology must take into account not only the estimated expected crash frequency (or crash rate) but also the uncertainties associated with that estimate. This work focuses on the development of an approach for ranking intersections in the State of Iowa. In so doing, we start from the approach proposed by Miaou and Song (2005) and extend it in several ways. Using Receiver Operating Characteristic curves, the performance of these methods is compared to a naïve ranking of raw crash numbers. Special difficulties are introduced by the dimension of the data set, since approximately 180,000 intersections in the state of Iowa are to be analyzed using this model. As not every crash falls into the same category with regard to severity, we then develop an approach to model different crash types simultaneously using a Bayesian hierarchical model based on Multivariate Poisson distributions and Gamma priors.

Combining Evidence for Cost Effectiveness Analysis An application in Rheumatoid Arthritis

Susanne Schmitz
Trinity College Dublin, Ireland
schmitzs@tcd.ie

Roisin Adams
NC Pharmacoeconomics, Ireland
RAdams@stjames.ie

Cathal Walsh
Trinity College Dublin, Ireland
Cathal.Walsh@tcd.ie

A cost-effectiveness analysis (CEA) compares the costs of an intervention with its effectiveness and therefore gives a guideline of whether to introduce the new intervention or not. Bayesian methods provide a natural framework to decision making in this area where combining evidence from different sources and estimation of uncertainty are indispensable. A probabilistic sensitivity analysis is included in the NICE (National Institute for Clinical Excellence, England) guidelines, and can be effectively implemented in the Bayesian framework. Useful outcomes such as cost-effectiveness acceptability curves are obtained easily as well. The National Centre for Pharmacoeconomics (NCPE) in Ireland reviews submissions from pharmaceutical companies to determine whether their treatments are cost-effective and advises the government in the decision to fund the product. Many of these submissions now include mixed treatment comparisons (MTCs) in the absence of head to head trials. In this work we examine how the model choice in a MTC impacts the results of the cost-effectiveness analysis. The drugs examined include four biologic agents used in the treatment of rheumatoid arthritis, adalimumab, golimumab, etanercept and infliximab. A Bayesian MTC model estimates all pair wise odds ratios. Since methotrexate (MTX) is sometimes given in one or several treatment arms of the studies the model is extended to a meta-regression by including a regression parameter for MTX. We estimate the treatment effect in four different models that differ in the choice of random and fixed effects for the parameter of treatment effect and for the regression parameter and compare the outcomes of cost-effectiveness analyses based on the different estimates.

A Bayesian Nonparametric Test for Logistic Distribution and Odds Ratio Estimation without Dichotomizing

Angela Schoergendorfer Adam J. Branscum
University of Kentucky
angela.sch@uky.edu abran3@email.uky.edu

Timothy E. Hanson
University of Minnesota
hans0058@umn.edu

With continuous response data, it is common to create a dichotomous outcome for logistic regression analysis by specifying a threshold for positivity. Fitting a linear regression via least-squares to the non-dichotomized response assuming a

logistic error distribution has been shown to yield more efficient estimators of odds ratios than traditional logistic regression. However, odds ratios are not constant across covariate values if the assumption of a logistic distribution for the residuals is violated, leading to biased inferences. We expand upon the idea of regression of a non-dichotomized response and develop a novel test for logistic distribution based on a Polya tree model. Bayes factors are calculated using the Savage-Dickey ratio for testing the null hypothesis of logistic distribution of residuals versus a nonparametric generalization. We propose an empirical Bayes approach which is computationally efficient and present methods for nonparametric risk estimation in the case that the hypothesis of a logistic error distribution is rejected.

Speeding Up the Product Estimator using Random Temperatures

Sarah Schott
Duke University, USA
schott@math.duke.edu

Mark Huber
Claremont McKenna College, USA
mhuber@cmc.edu

An important quantity in Bayesian model selection is the integrated likelihood, also known as the evidence, marginal likelihood, and the normalizing constant of the posterior. By use of an auxilliary variable, the problem of approximating the integrated likelihood to a specified accuracy can be reduced to estimating the ratio between the measure of a set B and the measure of a smaller set B' contained in B. The classic Monte Carlo method generates random variates from B and counts how many fall into B'. But to guarantee the relative accuracy of this method requires a number of samples equal to the ratio itself, which is problematic as for many problems the size of B' is exponentially small in the dimension when compared to the size of B. To overcome this difficulty, the product estimator introduces a sequence of nested subsets between B and B', indexed by a parameter β.

In the new approach presented here, an idea drawn from nested sampling is applied to choose these interpolating subsets adaptively in a fashion that is nearly optimal. The result is a fast Monte Carlo method that does not rely on variance estimates and whose output can be analyzed exactly without recourse to asymptotic approximations. To illustrate the method, we use it to approximate the integrated likelihood for a hierarchical model for a data set testing pump reliability.

Bayesian Inference for Spatio-Temporal Models using INLA and MCMC: A Case Study on Animal disease Counts from Switzerland

Birgit Schrödle
Universität Zürich, Switzerland
birgit.schroedle@ifspm.uzh.ch

Leonhard Held and Andrea Riebler
Universität Zürich, Switzerland
{leonhard.held, andrea.riebler}@ifspm.uzh.ch

Spatio-temporal models (Knorr-Held (2000), *Statistics in Medicine*, **19**: 2555-2567) are usually formulated in a hierarchical Bayesian framework and Markov chain Monte Carlo (MCMC) techniques are used for parameter estimation. However, this approach might be slow and difficult to apply from a user's point of view due to highly correlated samples and the inherently large Monte-Carlo error present in MCMC estimates. An alternative method which was proposed recently (Rue, Martino & Chopin (2009), *JRSS-B*, **71**: 319-392) uses integrated nested Laplace approximations (INLA) to obtain posterior marginals of the parameters of interest. This method runs remarkably fast concerning computational time and provides very accurate approximations for a wide range of models.

In the presented case study INLA is used for Bayesian inference in spatio-temporal disease mapping models applied to case reporting data provided by the Swiss federal veterinary office (Schrödle & Held (2009), available as a *Technical Report*, www.biostat.uzh.ch/research/manuscripts/schroedle_held_09.pdf.zip).

Since case reporting is supposed to be biased the objective of this analysis is to detect unusual spatial and temporal trends which hint at factors influencing the reporting behaviour. A second intention of this application is to compare the parameter estimates obtained by INLA vs. MCMC. Features of the two approaches concerning spatio-temporal disease mapping are pointed out. Furthermore, the model choice criteria given as standard output by the INLA routine are discussed briefly.

Robust Bayesian Shrinkage in Sparse Signal-Extraction Problems

James G. Scott
University of Texas at Austin, USA
james.scott@mccombs.utexas.edu

This talk is about how to construct Bayes estimators that reconcile the two primary goals of sparse signal extraction: shrink the noise, don't shrink the signals. Rather than study optimality results for a specific estimator, we will instead focus on general principles that can suggest wide classes of estimators that work well for different problems.

I will describe two theoretical results along these lines. The first is a representation theorem that relates the posterior mean to the score function of the marginal

distribution implied by a certain combination of prior and sampling model. This theorem is related to classic results of Stein and Masreliez, and gives qualitative insight into some aspects of robust Bayesian analysis in sparse settings.

The second result relates certain properties of a shrinkage prior to the efficiency of the resulting Bayes estimator in reconstructing the true sampling density under Kullback-Leibler loss. This provides guidance as to what kinds of priors yield good predictions in sparse problems.

Together, these results build up a picture of what a good "sparsity prior" should look like. I will provide a few examples to contrast priors that meet these criteria with those that don't, thereby demonstrating that the theoretical results explored here have important practical implications. I will also make some connections with recent research on Bayesian methods for multiple testing and sparse multiple regression.

Assessing Model Discrepancy for Simulators of Complex Systems

Allan H. Seheult
Durham University, UK
a.h.seheult@durham.ac.uk

Deterministic mathematical models of complex physical systems, such as those for hydrocarbon reservoirs, galaxy formation and climate, are usually implemented as a computer code, called a "simulator".

We consider inference and diagnostics for the inevitable difference between the simulator and the physical system it purports to represent. This difference, termed "model discrepancy", is regarded as a random quantity defined with respect to the notion of a best input to the computer model. Thus, if $f(x)$ is the value of the simulator value at an input x, there is postulated a special simulator input x^* such that the relationship between the real process y and the simulator value $f(x^*)$ at x^* is $y = f(x^*) + d$, where $f(\cdot)$, x^* and the model discrepancy d are all regarded as random quantities which are assumed to be uncorrelated or independent. This stochastic formulation provides a formal basis for inference about the distribution of d.

It is widely accepted that the uncertainties associated with both calibrating a mathematical model to observations on a physical system and forecasting for a physical system should take account of model discrepancy. However, model discrepancy specification is challenging, drawing upon expert knowledge of the physical system and any crucial simplifications assumed in the mathematical model.

Various methods are explored (expert, informal, likelihood, full Bayes and Bayes linear) for assessing model discrepancy such as estimating a parameterised version of a specification of the distribution of d.

We focus on inference for model discrepancy when it has been partially specified and diagnostics when it has been fully specified. However, there is not always a clear distinction, and both may arise in a particular study.

Methods are illustrated on a model for galaxy formation.

Biodiversity Estimation: Unraveling the T-cell Receptor Repertoire in the Body's Cellular Ecosystem

Nuno Sepúlveda
I. Gulbenkian Ciência, Portugal
nunosep@gmail.com

Vitor Sousa
M.D. Anderson Cancer C., USA
vsousa@mdanderson.org

Michele Guindani
University of New Mexico, USA
michele@stat.unm.edu

Peter Mueller
M.D. Anderson Cancer C., USA
pmueller@mdanderson.org

Carlos Daniel Paulino
Univ. Técnica de Lisboa, Portugal
dpaulino@math.ist.utl.pt

Jprge Carneiro
I. Gulbenkian Ciência, Portugal
jcarneir@igc.gulbenkian.pt

T-cell diversity estimation is essential to understand how body's immune system efficiently works to fight harmful pathogens without reacting against self components. Until now this statistical issue has been mainly tackled through a "standard" method based on the assumption that each T-cell "species" is equally represented in the body. However, several experimental data on T-cell diversity invariably show a great number of singletons and just a few highly abundant T-cell species, a pattern in complete disagreement with the above assumption. Therefore, current T-cell diversity estimates need to be revised. Recently we redeployed the Poisson abundance models from Ecology to infer on T-cell diversity and the respective clonal size distribution. Based on maximum likelihood estimation, several comands were created in R language and included in a new package called PAM. However, no confidence intervals for T-cell diversity (or other related parametric functions) are available in the package, since one needs to invoke asymptotic results that are questionable in most data sets. Therefore, the analysis of T-cell diversity by frequentist methods is limited to point estimates and goodness-of-fit tests for the models. Bayesian methods via stochastic simulation are then a good alternative to get a more broader analysis of T-cell diversity. Using previously published data on regulatory and effector T-cell diversity and non-informative priors, different stochastic simulations methods (MCMC, ABC and Sampling-Resampling) are applied to determine T-cell diversity posterior distributions implied by different Poisson abundance models. Non-parametric Bayes methods are also used and compared to their parametric counterparts. Finally, practical issues will be discussed for future implementation of Bayes methods in the package PAM.

Exact Inference for Discretely Observed Diffusions

Giorgos Sermaidis
Lancaster University, UK
g.sermaidis@lancs.ac.uk

The aim of this work is to perform Bayesian inference for discretely observed diffusions, the challenge being that the transition density of the process is typically

unavailable. Existing methods rely on augmenting the data with the missing paths since there exists an analytical expression for the complete-data likelihood. Such implementations require a rather fine discretization of the imputed path leading to slow convergent and computationally expensive algorithms.

Our method is based on exact simulation of diffusions (Beskos *et al.* 2006) and has the advantage that there is no discretization error. We present a Gibbs sampler for sampling from the posterior distribution of the parameters and discuss how to increase its efficiency using reparametrizations of the augmentation scheme (Papaspiliopoulos *et al.* 2007).

Reference Priors, Information Inequalities, and Hellinger Distance

Arkady Shemyakin
University of St. Thomas, Minnesota, USA
a9shemyakin@stthomas.edu

A new approach is suggested for construction of non-informative priors in the absence of Fisher information. A prior (which can be dubbed *Hellinger prior*) is defined in terms of the limiting behavior of the Hellinger distance between two adjacent points of the parametric space. Motivation for this construction is suggested by information inequalities of Borovkov and Sakhanienko (1982) and Shemyakin (1992). The behavior of *Hellinger priors* is compared for the non-regular class of distributions defined by Ghosal and Samanta (1997) and some cases outside of this class to that of the reference priors suggested by Berger, Bernardo, and Sun (2009), which are obtained by maximizing missing information defined with the help of Kullback-Leibler divergence. Several examples demonstrate that for the non-regular class of distributions *Hellinger priors* often coincide with the reference priors. Outside of non-regular class the latter does not seem to be true and two priors may exhibit close but somewhat different behaviors. Examples of uniform (with support depending in parameter) and asymmetric triangular distributions are considered following Berger, Bernardo, and Sun (2009).

Optimal Scaling of the Random Walk Metropolis: General Criteria for the 0.234 Acceptance Rule

Chris Sherlock
Lancaster University, UK
c.sherlock@lancaster.ac.uk

Scaling of proposals for Metropolis algorithms is an important practical problem in MCMC implementation. Criteria for scaling based on empirical acceptance rates of algorithms have been found to work consistently well across a broad range of problems. Essentially, proposal jump sizes are increased when acceptance rates are high and decreased when rates are low. Analyses of the random walk Metropolis for theoretically accessible classes of high dimensional targets has shown that in

many cases the optimal acceptance rate is approximately 0.234, but that there are exceptions. We present three general and related sets of criteria, any of which, if satisfied, ensures that the limiting optimal acceptance rate is 0.234. The criteria essentially require that as $d \to \infty$ the norm of the gradient, and a particular scalar measure of the curvature of the log-target both show little relative variation across the target, and that the local curvature of the target is never too eccentric. Further, it is much easier to check a particular set of criteria than to prove from scratch that the optimal acceptance rate is 0.234; The criteria are shown to hold for targets arising from a p-th order Markov process, and spherically symmetric targets where the log-density is a polynomial function of the radius.

Regularized Decompositions for Sparse MAP Estimation in Factor Analysis

Minghui Shi and David B. Dunson
Duke University, USA
ms193@stat.duke.edu duson@stat.duke.edu

It has become routine to encounter massive dimensional data sets in a rich variety of applications ranging from genomics to marketing. Large p, small n and large p, large n statistical problems have become a focus in the statistical community, but there is still a pressing need for new methods that scale to really large settings involving hundreds of thousands to millions of variables. For those kinds of problems, there is clearly a necessity of dimensionality reduction and strong prior information that the high-dimensional data can be sparsely characterized. A promising tool for addressing such problems is sparse Bayesian latent factor regression modeling. We propose a method that scales to much higher dimensions through conducting sparse maximum a posteriori (MAP) estimation in carefully-structured Bayesian factor models. In particular, we present a penalized matrix decomposition for the loading matrix using lasso (elastic net). Motivated by the singular value decomposition of the loading matrix, we use a decomposition formula for the loading matrix Λ as $\Lambda = \sum_{i=1}^{p} \lambda_i \mathbf{u}_i$, where λ_i are nonnegative real values and u_i are $p \times 1$ column vectors. Of particular interest are the uses of L_1 penalty (the exponential prior) on $\{\lambda_i\}_{i=1}^{p}$, which yields an automatic selection of the number of latent factors and on $\{\mathbf{u}_i\}_{i=1}^{p}$, which provides sparse vectors in the estimated loading matrix. An EM algorithm for maximizing the joint posterior distribution is proposed to produce Bayes estimates. As an illustration, this method is applied to both real and simulated data.

Modelling the Proportion of Burned area in Portuguese Forest Fires

Giovani L. Silva
Univ. Técnica de Lisboa, Portugal
gsilva@math.ist.utl.pt

Maria-Inés Dias
Universidade de Evora, Portugal
misd@uevora.pt

Forest fires are increasing as a result of climate change, and they are affecting larger areas and becoming more severe in several regions of the world. In the last

decade forest fires became a serious problem especially in Portugal due to different issues such as climatic characteristics and nature of Portuguese forest. In order to analyse forest fire data, we model the proportion of burned forest area by using hierarchical spatio-temporal models. Our goal is to identify 'hotspot' regions and trends over time through modelling and mapping of burned forest area proportion by regions such as districts or municipalities. The related models account for (i) spatial correlation (ii) nonlinear temporal effects, using both hierarchical spatial models and B-spline smoothing from a Bayesian perspective. An analysis of spatial burned area proportions in Portugal from 1980 to 2006 by municipalities motivates and illustrates the methods developed.

Comparison between Bayesian and Classical Methodologies for Discovering Transcription-Factor Targets from RIP-chip Experiments

Carina Silva-Fortes
Universidade de Lisboa, Portugal
carina.silva@estesl.ipl.pt

Maria-Antónia Amaral-Turkman and Lisete Sousa
Universidade de Lisboa, Portugal
antonia.turkman@fc.ul.pt lmsousa@fc.ul.pt

RIP-Chip (RNA immunoprecipitation on chip) refers to the immunoprecipitation of RNPs (RiboNucleoProteins) from cell extracts and the subsequent microarray analysis of associated RNA molecules and identifies DNA targets of DNA-binding proteins. An important remark is that RIP-chip experiments are not performed in the same way as the DNA expression experiments. In this case the immunoprecipitated and wild samples are processed in different ways and hybridized separately, leading to microarrays with expression levels very different. In a common DNA expression experiment we would expect few genes with differences on expression levels (1%) and on RIP-chip we would expect differences between 5%-10%. The main problem is that there are not specific methods to select features on this kind of experiments, thus the methods available for detecting differentially expressed (DE) genes are usually applied to these data. We used RIP-chip data from Gama-Carvalho *et al.* (2006) experiment to identify 2AF65 associated with postsplicing mRNP complexes. The goal is not to select DE genes but enriched RNA in the immunoprecipitaded samples. In this work we compare Bayesian (BAM, B statistic, empirical Bayes statistic, modT and ibmT) and classical (Rank Products, Fold Change, SAM and t-statistic of Welsh) methods for selecting enriched mRNA in the immunoprecipitaded samples (equivalent to select up-regulated genes). We conclude that among the methods for detecting DE genes, the most appropriate for this experiment are those based on Bayesian methodology.

Gaussian Process Structural Equation Models with Latent Variables

Ricardo Silva
University College London, UK
ricardo@stats.ucl.ac.uk

Robert B. Gramacy
University of Cambridge, UK
rbobby@statslab.cam.ac.uk

In a variety of disciplines such as social sciences, psychology, medicine and economics, the recorded data are considered to be noisy measurements of latent variables connected by some causal structure. This corresponds to a family of graphical models known as the structural equation model with latent variables. While linear non-Gaussian variants have been well-studied, inference in nonparametric structural equation models is still underdeveloped. We introduce a sparse Gaussian process parameterization that defines a non-linear structure connecting latent variables, unlike common formulations of Gaussian process latent variable models. In such a formulation, each latent variable is generated by a function of its parents in the graph plus an error term. Observed variables are linear combinations of latent variables with additive noise. Each function linking latent variables is given a modified Gaussian process prior that depends on a fixed but unknown set of support points, instead of the whole sample. In the literature, such support points are traditionally fit by maximum likelihood. Instead, we show how some convenient priors can be applied in order to infer, from data, a distribution over support points. An efficient Markov chain Monte Carlo procedure follows from this sparse parameterization, which is then evaluated empirically regarding its stability with respect to different initial points, and mixing behavior. Also importantly, we show how identifiability plays a role in such non-linear models. Finally, we evaluate the predictive ability of the model compared against the current practice.

Forecasting in INBL(1,0,1,1) Model

Nélia Silva and Isabel Pereira
Universidade de Aveiro, Portugal
neliasilva@ua.pt isabel.pereira@ua.pt

The class of non-negative integer-valued bilinear process was introduced by Doukhan et al.(2006) and was extended by Drost et al.(2008) to the class of superdiagonal INBL(p,q,m,n). In this paper parameter estimation is addressed considering the simple non-negative integer-valued bilinear process INBL(1,0,1,1). We consider the problem of forecasting integer-valued time series modelled by the INBL(1,0,1,1) process. Here Bayesian methodology is used to obtain point predictions as well as confidence intervals for future values of the process. The predictions thus obtained are compared with their classic counterparts. The proposed approaches are illustrated with a simulation study.

A Bayesian Self-Controlled Method for Drug Safety Surveillance in Large-Scale Longitudinal Data

Shawn E. Simpson and David Madigan
Columbia University, USA
shawn@stat.columbia.edu madigan@stat.columbia.edu

Increasing scientific, regulatory and public scrutiny focuses on the obligation of the medical community to ensure that marketed drugs have acceptable benefit-risk profiles. This is an intricate process that begins with careful pre-approval studies but continues after regulatory approval when a drug is in widespread clinical use. Much current attention focuses on drug safety surveillance in large-scale electronic health record databases and in medical claims databases.

Most existing methods for surveillance rely on statistics based on the many 2x2 tables formed by all possible drug-condition pairs. Such analyses ignore drug-drug interactions and have the potential to provide misleading results.

We provide a new approach for post-marketing surveillance in large-scale longitudinal health records. This approach is based on the self-controlled case series (SCCS) method [Farrington, 1995], which implicitly controls for fixed baseline covariates since each individual acts as their own control. We extend the model to include large numbers of potential time-varying confounders such as other drugs. In addition, our Bayesian version of SCCS can deal with high dimensionality and provides a sparse solution via a Laplacian prior. We present details of the model and the optimization procedure, as well as empirical results for drugs and AEs that have been previously linked.

Reconstructing Population Histories from Single-Nucleotide Polymorphism Data

Jukka Sirén
University of Helsinki, Finland
jukka.p.siren@helsinki.fi

Pekka Marttinen
Aalto University, Finland
pekka.marttinen

ukka Corander
University of Helsinki, Finland
jukka.corander@helsinki.fi

Population genetics encompasses a strong theoretical and applied research tradition on the multiple demographic processes that shape genetic variation present within a species. When several distinct populations exist in the current generation, it is often natural to consider the pattern of their divergence from a single ancestral population in terms of a binary tree structure. Inference about such population histories based on molecular data has been an intensive research topic in the most recent years. The most common approaches use coalescent theory to model genealogies of individuals sampled from the current populations. However, the computational complexity associated with the need to trace all possible genealogies limit their applications to moderately small data sets. In this work we propose a novel fully Bayesian method for inferring population histories from

unlinked single-nucleotide polymorphisms, which is applicable also to datasets harboring large numbers of individuals from distinct populations. We assume that the genetic variation between the populations is due to genetic drift and model its effect to the allele frequencies with the Balding-Nichols model. A combination of analytical, numerical and Monte Carlo integration techniques are utilized for the inferences.

Some Comments on "Likelihood-Free" Samplers and Models

Scott Sisson Yanan Fan and Gareth Peters
University of New South Wales, Australia
Scott.Sisson@unsw.edu.au

In recent years, applications of Bayesian inference using models with computationally intractable likelihood functions have become increasingly commonplace. These have been motivated by a desire to fit realistic models to data, rather than simplify the models for the sake of computational convenience. Rather than model with the usual posterior

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta), \tag{1}$$

for $\theta \in \Theta, y \in \mathcal{Y}$, as $\pi(y|\theta)$ may not be evaluated pointwise, the target posterior is first reduced to a distribution on a vector of statistics $s_y = f(y) \in \mathcal{S}$, giving

$$\pi(\theta|s_y) \propto \pi(s_y|\theta)\pi(\theta). \tag{2}$$

If s_y are sufficient statistics for θ then equations (1) and (2) are equivalent. The model is then augmented by an artificial dataset, $s_x \in \mathcal{S}$, resulting in the model

$$\pi(\theta, s_x|s_y) \propto \pi(s_y|s_x)\pi(s_x|\theta)\pi(\theta), \tag{3}$$

where $\pi(s_y|s_x)$ is a distribution that places high density on regions where $s_y \approx s_x$. Interest is typically in

$$\pi_{LF}(\theta|s_y) \propto \int_{\mathcal{S}} \pi(\theta, s_x|s_y) ds_x \tag{4}$$

as an approximation to the true posterior (1). If $\pi(s_y|s_x) = 1$ for $s_y = s_x$ and 0 elsewhere, and if s_y are sufficient statistics for θ, then equation (4) reduces to equation (1). Through the form of (3) or (4), simulation methods are available to simulate from either $\pi(\theta, s_x|s_y)$ or $\pi_{LF}(\theta|s_y)$, with the latter based on Monte Carlo integration via (4).

In this paper, we demonstrate that:

(i) Samplers targeting either $\pi(\theta, s_x|s_y)$ or $\pi_{LF}(\theta|s_y)$ are exactly equivalent. This implies that "likelihood-free" samplers are one of the few cases where one may substitute a Monte Carlo estimate of the likelihood in (say) a Markov Chain Monte Carlo sampler, and not generate biased results.

(ii) While the form of the distribution $\pi(s_y|s_x)$ has to date been considered arbitrary, more accurate inference (for a given computational burden) can be achieved if $\pi(s_y|s_x)$ closely describes the distribution of the statistics, $s \in \mathcal{S}$.

Foundations of Computational Inference

John Skilling
Maximum Entropy Data Consultants, Ireland
skilling@eircom.net

Too many important problems remain beyond Bayesian computation. Rather than congratulating success, we should discover and cure our algorithms' flaws. Bayes' theorem

$$\pi(\theta)\, L(\theta) = Z\, P(\theta)$$

gets posterior P and evidence Z from prior π and likelihood L. Z logically precedes P, and is at least as important.

Inference is governed by the information content

$$H = \int P(\theta)\, \log\left(\frac{P(\theta)}{\pi(\theta)}\right)\, d\theta$$

which uniquely invariant relation quantifies the compression. Although π necessarily supports P, P need not support π.

Inference is inherently asymmetric. Symmetrical detailed balance is inappropriate for compression, which should be systematic. A single step cannot efficiently compress more than $O(1)$, so compression must be iterative.

Dimension is not fundamental. Nothing in probability calculus imposes topology or dimensionality on the hypothesis space, whose elements can be arbitrarily re-ordered. Dimension merely encodes adjacency relationships that may be useful in exploration, but could be changed.

As for volume, point samples occupy a set of measure zero in the hypothesis space, and have no natural connection with volume. Samples from π often miss the bulk of P, thereby under-estimating Z.

Samples drawn from P (harmonic mean and Chib), are blind to small-scale structure in L, which risks over-estimating Z.

Sampling must be from a measure (i.e. a distribution or density). The prior is a measure on θ, but likelihood is only a function. Hence iteration must step through pseudo-likelihood functions $\lambda(\theta)$ modulating π. Annealing uses $\lambda = L^\beta$ to compress π to P, but fails for phase transitions. The only pseudo-likelihoods that are impervious to phase transitions are constant, with compression achieved by progressive cutoff in likelihood. That uniquely robust scheme is nested sampling.

Nested sampling compresses systematically, with compression ratios known by construction. If the user explores correctly, nested sampling's results are guaranteed unbiassed convergence in mean square provided only that Z and H are bounded.

Nested sampling is properly Bayesian, with proven estimates of the uncertainty that necessarily accompanies random sampling. It makes no assumptions about dimensionality or continuity, and can process any likelihood.

Bayesian Prediction of Non-Fish Bycatch in Commercial Fisheries using Hierarchical Models

Murray H. Smith and Suze J. Baird
National Institute of Water and Atmospheric Research, New Zealand
mh.smith@niwa.co.nz

Commercial fishing interacts with many species other than the fish species targeted by the fishers. Fishing in the New Zealand Exclusive Economic Zone is regulated and restrictions are in place in fisheries where rare and endangered marine mammal and seabird species are caught. Estimates of numbers of captures of these species are required and a fully Bayesian hierarchical prediction method has been developed that has been applied to trawl fisheries. Capture data is collected by scientific observers on selected commercial fishing vessels and commercial fishers are required to make trawl by trawl data returns, which do not include bycatch data. The observer data is often not very representative of the commercial data. The model, which is fitted to the observer data using WinBUGS, uses a negative binomial distribution for the counts, includes covariates, and includes effects for each year and for each observed vessel within each year that are hierarchical. Samples from the predictive distribution of the total numbers of captures are made by applying the MCMC samples from the posterior distribution to the commercial data for the unobserved trawls. The model means that the predicted counts include adjustments for systematic differences between the observed and unobserved trawls and differences in capture rates per trawl between vessels are incorporated as additional uncertainty. Results from the application of the method to predicting NZ sea lions captures in the Sub-Antarctic squid trawl fishery are presented. The use, in recent years, of devices in nets which allow most sea lions to escape required some additional modelling.

How Much Can Change at a Change Point? A Note on Modelling Change Points in Non-Homogeneous Poisson Processes

Paulo Soares
Universidade Técnica de Lisboa, Portugal
psoares@math.ist.utl.pt

Jorge A. Achcar
Univ. Estadual de Maringá, Brazil
jorge.achcar@pq.cnpq.br

Carlos D. Paulino
Univ. Técnica de Lisboa, Portugal
dpaulino@math.ist.utl.pt

The non-homogeneous Poisson process (NHPP) plays an important role in the analysis of events occurring in time such as failures of repairable systems. To model the occurrence of interventions on those systems, several change point models have been proposed that rely on the choice of particular parametric models from the NHPP class of models, such as the Musa-Okumoto or the Power Law processes.

In this work we discuss a potentially serious inadequacy of those change point models when an arbitrary change is allowed at a single change point. In this situation the models are close to non-identifiability and so they can pose severe difficulties on a MCMC analysis of the posterior distribution or, even worst, lead to misleading results strongly dominated by the prior distribution. After the discussion of the problem, a tentative remedy is presented and illustrated.

Simulation-Based Optimal Design using a Response Variance Criterion

Antti Solonen
Lappeenranta University of Technology, Finland
solonen@lut.fi

Heikki Haario
Lappeenranta Univ. Tech., Finland

Marko Laine
Finnish Meteorological Institute

Classical optimal design criteria suffer from two major flaws. First, they are routinely used in nonlinear problems by linearizing the model around a point estimate. Second, classical design methods are unavailable in ill-posed estimation problems, where the data does not contain enough information to identify all parameters. Bayesian optimal design can, in principle, solve these problems. A wide usage of Bayesian design methods is hindered by the lack of efficient and robust implementations that allow routine application in practical problems.

We point out a concrete recipe for implementing Bayesian optimal design. The approach utilizes an existing sample from the parameter posterior, nowadays routinely produced by efficient adaptive MCMC samplers. The proposed design criterion is based on searching for designs where the variance in model response is large (where a new measurement is able to discriminate different model parameters). The numerical sampling approach allows one to start the optimization of experiments at an early stage, before the estimation problem becomes well-posed and classical methods start to work. This potentially results in large reduction in the number of experiments needed to obtain a desired level of accuracy in the parameter estimates.

Bayesian Data Assimilation for Dynamical Systems

Elaine T. Spiller
Marquette University, USA
elaine.spiller@marquette.edu

There is a common trade off one makes in choosing between different filtering methods to employ for data assimilation: nonlinearity vs. dimension. Sequential Monte-Carlo (SMC) techniques allow one to approximate posterior distributions of hidden states at an observation instant without making the assumption of linearity on the dynamic model. However, new techniques need to be developed in order for nonlinear filters to be useful for high-dimensional nonlinear problems.

One compelling strategy to overcome this dilemma involves using a SMC, which naturally handles nonlinearity, in conjunction with dimensional reduction of the model variables in order to identify directions of expanding dynamics. Of specific interest are problems where the data being assimilated is Lagrangian in nature as is the case in many atmospheric and oceanic problems.

Neutral-Data Comparisons for Bayesian testing

Dan J. Spitzner
University of Virginia, USA
dan.spitzner@virginia.edu

A novel approach to evidence assessment in Bayesian testing is proposed as a "neutral-data comparison", which modifies the Bayes factor by reformulating its baseline of comparison to reflect data that are "neutral" between the hypotheses of interest. Suppose the data $\boldsymbol{Y}$ are observed, and it is of interest to test the null hypothesis H_0 against a general alternative, H_1. Whereas the Bayes factor is $BF = \{P[H_0|\boldsymbol{Y}]/(1-P[H_0|\boldsymbol{Y}])\}/\{\rho_0/(1-\rho_0)\}$, the corresponding neutral-data comparison is

$$NDC \;=\; \frac{P[H_0|\boldsymbol{Y}]/(1-P[H_0|\boldsymbol{Y}])}{P[H_0|\tilde{\boldsymbol{Y}}]/(1-P[H_0|\tilde{\boldsymbol{Y}}])}\,,$$

in which $\tilde{\boldsymbol{Y}}$ is an imaginary set of data that is believed to exhibit evidence neither for H_0 nor H_1. The posterior probability, $P[H_0|\tilde{\boldsymbol{Y}}]$, is interpreted as a "phantom" prior null probability that is free of such dimensionality mismatches between H_0 and H_1 that give rise to the Jeffreys-Lindley paradox. While the nature of $\tilde{\boldsymbol{Y}}$ is ultimately an elicitation of prior belief, a precise "objective" characterization is proposed, which follows from high-dimensional analysis within a certain standardized context.

An important advantage of NDC over BF is that it is drastically less sensitive to modifications of the prior, and is thus quite suitable for problems involving degenerate H_0 with vague priors on H_1. Moreover, NDC retains sufficient sensitivity to respond to specialized aspects of the prior such as hierarchical structures intended to induce shrinkage within parameter groupings.

The methodology is demonstrated on several existing data sets, emphasizing in particular how neutral-data comparisons provide an informative supplement to the typical estimation-oriented approach to analysis-of-variance. Neutral-data comparisons are showcased for such purposes as analysis under a mixture model, analysis of partitions, analysis-of-variance with a Poisson response, and multiple testing of adverse-event data. Calculations are shown to be readily accessible by available Markov chain Monte Carlo algorithms.

Evidence and Stopping Rules in Sequential Trials

Jan Sprenger
Universiteit van Tilburg, The Netherlands
j.sprenger@uvt.nl

Bayesians and frequentists are divided by fundamental disagreements on the inferential role of stopping rules. While both sides agree that not every stopping rule

is not equally suited for conducting a specific experiment (e.g. in cost management), the Bayesian's committal to the Likelihood Principle forces her to regard stopping rules as evidentially irrelevant. Frequentists argue, on the other hand, that we can tweak posterior probabilities and Bayes factors by modifying the stopping rules; methods of inference where the latter are neglected will lead to biased conclusions.

Since both camps adhere to different statistical methodologies, it seems hard to find a common ground for settling the dispute. The paper argues, however, that the problem is a special case of testing a hypothesis at a fixed level where a random choice between different experiments is made, or where the value of a nuisance parameter is unknown. In both cases, sticking to fixed-level testing leads to incoherence (M. Schervish, J. Kadane and T. Seidenfeld (2003): "Measures of incoherence: How not to gamble if you must", in J. Bernardo *et al.* (eds.): Bayesian Statistics 7, pp. 385-402, Oxford: OUP). It is argued that frequentist decision rules which sensitively depend on the actually used stopping rule lead to the same kind of incoherence. Finally, the paper explores to which extent the frequentist's insistence on the relevance of stopping rules can be rationalized from a Bayesian perspective.

Efficient Bayesian Inference for Mixture Models

Vassilios Stathopoulos and Mark Girolami
Department of Computing Science, University of Glasgow, UK
stathv@dcs.gla.ac.uk girolami@dcs.gla.ac.uk

Bayesian inference for Mixture Models is a long standing problem in the statistical literature. While the use of Gibb's sampler is the most prominent approach for mixture models, it can exhibit strong random walk behaviour which in effect reduces the effective sample size. Efficient Monte Carlo algorithms, such as the Hamiltonian (or Hybrid) Monte Carlo (HMC) and the Riemanian Manifold Hamiltonian Monte Carlo (RMHMC) (Girolami, M. etal., 2009), overcome the problem of random walks in the expense of calculating higher order derivatives of the joint log likelihood function. Although first order derivatives for mixture models are usually easily computed analytically, there is no closed form solution for the Fisher Information matrix and it's partial derivatives required by RMHMC.

In this work we study the application of efficient Monte Carlo algorithms for Bayesian inference of mixtures of Gaussian densities. We start by deriving a metric tensor for mixtures of Gaussians based on the Integrated Squared Error (ISE) metric and we propose its use as opposed to the true Fisher Information. We also study the use of approximations to the true Fisher Information, such as the observed Fisher matrix. Even though, the ISE metric tensor can be calculated in closed form, doing so for multivariate data it is an algebraically tedious process. For multivariate data we study the application of Gaussian Processes to approximate the true joint log likelihood. For HMC and RMHMC the Gaussian Process can be used as a "guide" Hamiltonian as in (Rasmussen, 2003). In that way only the higher order derivatives of the predictive mean function of the Gaussian process are required, which for most kernel functions are easily computed in closed form.

Bayesian Models for Variable Selection that Incorporate Biological Information

Francesco C. Stingo and Marina Vannucci
Rice University, USA
fcs1@rice.edu marina@rice.edu

Variable selection has been the focus of much research in recent years. Bayesian methods have found many successful applications, particularly in situations where the amount of measured variables can be much greater than the number of observations. One such example is the analysis of genomics data. In this paper we first review Bayesian variable selection methods for linear settings, including regression and classification models. We focus in particular on recent prior constructions that have been used for the analysis of genomic data and briefly describe two novel applications that integrate di?erent sources of biological information into the analysis of experimental data. Next, we address variable selection for a di?erent modeling context, i.e. mixture models. We address both clustering and discriminant analysis settings and conclude with an application to gene expression data for patients affected by leukemia.

Robust Bayesian Functional Mixed Effects Modeling

Foteini Strimenopoulou and Philip J. Brown
University of Kent, UK
fs54@kent.ac.uk Philip.J.Brown@kent.ac.uk

A general functional mixed effects model that inherits the flexibility of linear mixed effects models in handling complex designs and correlation structures is considered. Through the use of wavelet decompositions we allow functions of arbitrary form. Therefore, this method yields nonparametric estimates for the fixed and random effects functions. Priors are assigned to all quantities of the model and a fully Bayesian approach is pursued.

In contrast with the methodology (normality assumptions) proposed in the literature, here we employ a more robust model setting. We use a semi-heavy tailed distribution for the wavelet coefficients of the fixed effects, random effects and residual errors. This distribution is the normal inverse Gaussian which is a continuous probability distribution and is defined as the normal variance-mean mixture where the mixing density is the inverse Gaussian distribution. It has four parameters which control the mean, scale, shape and asymmetry of the distribution. Limiting cases of the NIG is the normal and the Cauchy distribution. In the non-Cauchy limit, the tails exhibit a combined algebraic and exponential decay, and the exponential component is controlled by the values of two of the parameters, namely the shape and asymmetry parameter.

We chose to assess the model fitting using two simulated data sets; one is generated from a normal distribution, and the other one from a Cauchy distribution, for fixed effects, random effects and errors. Also, we compare the fitting of three different models; the double exponential, the normal and the normal inverse Gaussian.

Inference in Queueing Networks with Missing Data

Charles Sutton and Michael I. Jordan
University of Edinburgh, UK *University of California, Berkeley, USA*
csutton@inf.ed.ac.uk jordan@stat.berkeley.edu

Modern Web services, such as those at Google, Yahoo!, and Amazon, handle billions of requests per day, serving large numbers of simultaneous users on clusters of thousands of computers. Despite the computational demands on these services, they also operate under strict performance requirements. It is extremely important to minimise the amount of time required for a response to be returned to the user, because even small delays, such as 100 ms, are sufficient to cause a measurable decrease in business. A primary factor that determines the response time of a Web service is the rate at which requests arrive. It is therefore of great practical interest to predict the latency of a Web service from the request rate. This is essentially a regression problem, but standard regression techniques fall short, because a primary goal of modelling a computer system is extrapolation to high request rates. Additionally, analysis is complicated by the fact that the data is incomplete, because recording detailed information about every request to a heavily used system can require unacceptable overhead. The system is modelled by a network of queues, in which one queue models each of the individual computers that make up a large-scale system. This model naturally achieves the first goal—the ability to extrapolate—because queueing models predict the explosion in response time under high request rates in a way that is often reasonable for real systems. However, the second goal—the ability to handle missing data—is more elusive. Indeed, we are unaware of previous work that concerns inference in networks of queues in the face of missing data.

In this work, to handle incomplete observations, we propose a novel Bayesian perspective on queueing models, in which we view the queueing network as a structured probabilistic model over the arrival and departure time of each request, and the arrival and departure times that are not observed are treated as hidden variables. Underlying this viewpoint is the observation that a queueing model defines a deterministic transformation from a set of independent variables, namely, the service times, to the arrival and departure times, which can be observed. This perspective is general enough to handle fairly advanced types of queueing models, including general service distributions, multiprocessor queues, and processor-sharing queues.

With this viewpoint in hand, we can sample from the posterior distribution over missing data and model parameters using Markov chain Monte Carlo. The development of the sampler presents significant algorithmic challenges, for two reasons. First, the local conditional distribution over a single departure can have many singularities, making it difficult to design a Metropolis-Hastings proposal. Second, a change to a single departure time can affect arbitrarily many later departures, by keeping the queue busy.

We evaluate our framework on data generated by a recently-proposed benchmark for modern Web applications. We also present a simple technique for selection among two nested queueing networks.

Approximate Joint Statistical Inference for Large Spatial Datasets

James Sweeney
Trinity College Dublin, Dublin, Ireland
sweeneja@tcd.ie

We propose an approximate sequential approach for inferring the correlation matrix in large multivariate spatial regression problems. This enables the decomposition of the computationally intensive, mul- tivariate, "joint" problem, into a set of independent univariate prob- lems with possible correlation structure inferred sequentially. Omis- sion of correlation structure(where not appropriate) in potential mod- els will lead to increased uncertainty in the degree of confidence at the reconstruction stage of an associated inverse problem. The results from the proposed sequential approach are compared to those obtained using a (correct) full joint approach through the com- parison of bias and predictive properties for simulated and palaeocli- mate data. Inference procedures used are Empirical Bayes(EB) based where the hyperparameters governing a given model are considered as unknown fixed constants

Bayesian Estimation of a Time-Varying Correlation Matrix using the Wishart Process

Aline Tabet Arnaud Doucet and Paul Gustafson
University of British Columbia, Canada
{a.tabet, arnaud, gustaf}@stat.ubc.ca

Multivariate longitudinal data arise in many research settings. Such data are characterized by multiple response variables measured over a period of time. Many statistical methods developed to analyze multivariate longitudinal data focus on the efficient estimation of the regression coefficients to study the mean structure evolution over time as a function of covariate information. However, in many cases, the actual covariance structure, which encodes the variation, co-variation and temporal dynamics of the response variables, may be of interest. For example, in financial applications, the primary aim is to model the volatility and the co-volatility of financial variables. Typically, in this setting, the covariance matrix is very complex because there exists an association between the elements of the multidimensional response at each time point, as well as over time. Furthermore, this association is often time-varying. The estimation of a combination of fixed and time-varying parameters is required. In this work, we propose using the Wishart auto-regressive process and an efficient particle Markov chain Monte Carlo algorithm to estimate time-varying covariance matrices. The particle Markov chain Monte Carlo algorithm provides a flexible computational framework that allows the sampling of fixed parameters using traditional Markov chain Monte Carlo and sampling the time-varying parameters using sequential Monte Carlo, while ensuring that samples are drawn from the same invariant posterior distribution of interest. We present result from simulation studies as well as an application to an actual financial data data set.

Dynamic Point Process Modeling with a DDP

Matthew A. Taddy
Chicago Booth, USA
taddy@chicagobooth.edu

Bayesian nonparametric methods have been previously employed in flexible modeling for spatial data that may be viewed as the realization of a Poisson point process. When these events occur over time, it is sometimes possible to incorporate temporal dependence by viewing the data as a single realization of a point process which occurs in both space and time. However, in discrete time settings—when each time point corresponds to a set of spatially distributed events—it is more appropriate to model the data as arising from multiple point processes that have an intensity function which is changing in time. This is made possible by employing a novel version of the Dependent Dirichlet process as a prior for the time-indexed normalized process intensities. Details of the development of the model and prior measure will be provided, along with an example involving crime event data.

Novel Advanced Mathematical and Statistical Methods for Understanding Climate (NOVAC)

Johanna Tamminen Marko Laine
Finnish Meteorological Institute, Finland
johanna.tamminen@fmi.fi marko.laine@fmi.fi

Heikki Haario Antti Solonen
Lappeenranta University of Technology, Finland
heikki.haario@lut.fi antti.solonen@lut.fi

Erkki Oja
Aalto University, Finland
jerkki.oja@tkk.fi

Heikki Järvinen
Finnish Meteorological Institute
heikki.jarvinen@fmi.fi

Climate models contain closure parameters which can act as effective "tuning handles" of the simulated climate. These appear in physical parameterization schemes where unresolved variables are expressed by predefined parameters rather than being explicitly modeled. In the current climate model tuning process, best expert knowledge is used to define the optimal closure parameter values, based on observations, process studies, large eddy simulations, etc. In fact, closure parameters span a low-dimensional space, and the parameter probability densities should be objectively estimated simultaneously for all relevant parameters. The authors have just started a project called "Novel advanced mathematical and statistical methods for understanding climate" (NOVAC, 2010–2013) which is funded by the Academy of Finland. Several research problems addressed during the project are discussed here. The uncertainties of climate model closure parameters are estimated to improve understanding of reliability of climate predictions. We focus on the ECHAM5 model closure parameter distribution, and study the impacts on the

reliability of future climate predictions. The methodology is, however, generic and applicable in any multi-scale problem with similar closure parameters. Efficient Markov chain Monte Carlo (MCMC) sampling techniques are developed to tackle computationally challenging problems. MCMC simulations is an attractive tool for solving complex inverse problems. However, they are computationally very expensive and only efficient and maximally optimized MCMC techniques make the approach realistic in practice. We develop new tools based on adaptive algorithms, multiple computational grids, parallel chains as well as methods based on early rejection. Also, novel statistical methods are developed to analyze very large observed and modeled data sets involved in climate research. For compression and for finding underlying spatio-temporal factors, methods such as empirical orthogonal functions (principal components) and related techniques are commonly used. The objective here is to develop advanced statistical data mining methods that surpass these well-known techniques, to more effectively detect the most significant multidimensional signals from the data. This enables us to find efficient cost function criteria for the application of the sampling methods for closure parameter estimation.

A Bayesian Mixture Model for Partitioning Zero-Inflated Continuous Feature Vectors

Jing Tang
VTT Tech. Res. Center, Finland
jing.tang@vtt.fi

Jukka Kohonen
Helsingin Yliopisto Finland
ukka.kohonen@helsinki.fi

Jukka Corander
Helsingin Yliopisto Finland
jukka.corander@helsinki.fi

Matej Oresic
VTT Tech. Res. Center, Finland
matej.oresic@vtt.fi

Bioinformatics research often involves the analysis of continuous feature data that contain an excessive proportion of zeros. Such so-called zero-inflated data often arise in biological or ecological experiments in which a massive number of features are measured simultaneously. The word inflation is used to emphasize that the probability of a feature going undetected exceeds that from well-defined distributions, such as Poisson or Gaussian distributions. One of many examples bearing such a zero inflation characteristic can be found in LC-MS/MS proteomic analysis, or GCGC-TOF metabolomic profiling. A common aim of statistical analysis of zero-inflated data is to determine the similarity between the observations in terms of their feature patterns. In this article, we describe a model-based Bayesian clustering model for partitioning such data. A key step in the proposed modeling method is the construction of a two-fold data generating process that approximates reality. Accordingly we derive an analytical expression for the posterior probability of a given partition solution. We avoid using typical Markov chain Monte Carlo (MCMC) methods and utilize a much more efficient stochastic search algorithm for finding the posterior optimal partition. Experiments on simulated data show a significant improvement for clustering compared to discrete models and Gaussian mixture models.

A Bayesian Hierarchical Model for Identifying Epitopes in Peptide Microarray Data

Luca Tardella
Università di Roma La Sapienza, Italy
luca.tardella@uniroma1.it

Arima Serena
Univ. di Roma La Sapienza, Italy
serena.arima@uniroma1.it

Valentina Pecora
Policlinico A. Gemelli, Roma, Italy
valentina.pecora@hotmail.it

Peptide microarray immunoassay (MIA) is a novel technology that enables researchers to map a large number of epitopes, providing information regarding the relationship between antibody response and clinical sensitivity. Detection of antibody recognized peptides and epitope regions presents new challenges for statistical analysis mainly due to the interior spot dependence: in fact, each spot (peptide) corresponds to a sequence of amino acids partially overlapping with their neighbors. We propose a flexible Bayesian Hierarchical model which allows to detect recognized peptides and bound epitope regions in a single framework, taking into account the dependence between probes. Parameter estimation is carried out using Markov chain Monte Carlo and inference is based on the joint distribution of the parameters. Recognized peptides are detected using posterior probabilities computed from the joint posterior distribution and epitopes regions specified as sequences of recognized peptides. Suitable thresholds for these posterior probabilities are calibrated through an estimate of the false discovery rate. Model extensions for embedding different patients and experimental conditions have been introduced and criteria to compare bounding regions have been proposed. We present a simulation study, which shows the inferential power of the proposed model compared to other competing models in terms of epitope detection. Model criticism is also addressed via posterior predictive checks. Inference is illustrated on novel peptide microarray data of ovalbumin protein belonging to a study about IgE and IgG_4 mediated allergies.

An Adaptive Sequential Monte Carlo Method

Benjamin M. Taylor and Paul Fearnhead
Lancaster University, UK
b.taylor1@lancaster.ac.uk p.fearnhead@lancaster.ac.uk

Particle filtering methodology is usually applied to non-linear dynamic state space models, but it is possible to adapt the methods to work in static, fixed dataset, scenarios. This strategy is advantageous because it combines the best aspects of sequential importance sampling and MCMC . By propagating a swarm of particles through a sequence of tempered targets, SMC algorithms are less likely to become trapped in local posterior modes than MCMC, and since it is not necessary to evaluate the full likelihood at each step of the algorithm, a considerable computational saving can also be made. The class of SMC algorithms of interest in this poster presentation are those that employ MCMC kernels for particle dynamics,

such as the method suggested by Chopin (2002). SMC has two further advantages in this case: firstly, posterior moments are easily computed using the particle set, these can be used to inform the shape of the proposal density; secondly, the kernel move does not have to be ergodic, which means that rather than using a fixed kernel, adaptive MCMC may be employed with a greater degree of freedom than is usual. Our adaptive SMC algorithm uses information from the particles to select optimal MCMC tuning parameters, but moreover, allows automatic choice between proposal densities.

Our strategy has proved highly successful in simulation studies. We consider the problem of Bayesian mixture analysis, where the posterior is invariant to permutations in the parameter vector. Our method is able to select the permutation that best describes the posterior covariance, which further improves MCMC mixing. We investigate the choice between simple random walk proposals and mixture proposals based on the idea of Liu and West (2001). The results from these studies show that the additional complexity of our algorithm is justified, with performance comparing well to fixed, approximately optimally scaled counterparts. We present the results of our simulation studies, provide theoretical guidance on implementation and under mild regularity assumptions, prove convergence of our algorithm to the optimal choice of scaling and proposal.

Characterizing Uncertainty of Future Climate Change Projections Using Hierarchical Bayesian Models

Claudia Tebaldi
University of British Columbia, Canada and *Climate Central Inc., USA*
c.tebaldi@ubc.ca
Richard L. Smith
UNC, Chapel Hill, USA
rls@email.unc.edu
Bruno Sansó
Univ. California, Santa Cruz, USA
bruno@ams.ucsc.edu

The use of projections from ensembles of climate models to characterize future climate change at regional scales has become the most widely adopted framework, as opposed to what was standard practice until just a few years ago when single model's projections constituted the basis for arguing about future changes and their impacts. It is believed that comparing and synthesizing multiple models' simulations is key to quantify a best estimate of the future changes and its uncertainty. In the last few years there has been an explosion of literature in the climate change science where mostly heuristic methods of synthesizing multiple models' output have been proposed, and the statistical literature is showing more involvement by our community as well of late. In this paper we give a brief overview of the main streams of research in this area and then focus on our recent work, through which we have proposed the framework of hierarchical Bayesian models to combine information from model simulations and observations, in order to derive posterior probabilities of temperature and precipitation change at regional scales.

Hierarchical Bayesian Nonparametric Models for Language and Text

Yee Whye Teh
University College London, UK
ywteh@gatsby.ucl.ac.uk

Frank Wood
Columbia University, USA
fwood@stat.columbia.edu

Jan Gasthaus
University College London, UK
j.gasthaus@gatsby.ucl.ac.uk

Cédric Archambeau
Xerox Research Centre, France
cedric.archambeau@xrce.xerox.com

Lancelot James
Hong Kong Univ. of Sci. and Tech.
lancelot@ust.hk

Statistical models of languages (as sequences of words) and text (as sequences of characters) are important components of natural language processing systems. Language and text exhibit a number of interesting statistical characteristics making them good candidates for sophisticated Bayesian modeling. In particular both exhibit long range, non-trivial dependencies as well as power-law properties. Producing good models of data with these characteristics requires models with very large numbers of parameters. The amount of data available relative to the number of parameters is usually small, making Bayesian solutions which integrate out parameters attractive. We model sequences of words and characters using high order Markov models, where the transition probabilities are given a hierarchical Pitman-Yor process prior. The hierarchy allows for sharing of statistical strength within the model, while Pitman-Yor processes allow modeling of the power-law propertes of language and text. We developed efficient Gibbs sampling and particle filtering inference algorithms for the models. In large scale applications We show that these lead to state-of-the-art language modeling and text compression results, surpassing those of existing approaches. We have also extended the models to be truly nonparametric, where the order of the Markov models is infinite. Note that such models are no longer Markov. To make inference in the model computationally efficient, we used coagulation/fragmentation identities of Pitman-Yor processes to marginalize out unnecessary parameters. Inference in the resulting models scales linearly with the lengths of sequences, making very large scale applications with millions of symbols possible.

Statistical Regular Sub-Pavings for Multivariate Density Estimation

Gloria Teng Jennifer Harlow
Dominic Lee and Raazesh Sainudiin
University of Canterbury, New Zealand
{gat41,j ah217}@student.canterbury.ac.nz
{d.lee, r.sainudiin}@math.canterbury.ac.nz

We present an efficient, adaptive statistical data-structure that is sufficient for density estimation of massive, multi-dimensional data. This data-structure is

formalized as statistical regular sub-pavings (SRSPs). A regular sub-paving (RSP) (Jaulin et al. (2001), *Applied Interval Analysis*, Springer-Verlag London) or n-tree (Samet (1990), *The Design and Analysis of Spatial Data Structures*, Addison-Wesley Longman Publishing Co., Inc., Boston) is an ordered binary tree that recursively bisects a root box $\mathbb{X} \subset \mathbb{R}^d$ along the first longest side. SRSP augments RSP by mutably caching the recursively computable sufficient statistics of the data. Our proposed adaptive histogram estimate is constructed from a discrete time Markov chain over the space of all SRSPs. The Markov chain is proposed by a randomized priority queue based on statistically equivalent blocks that adaptively prunes and grows the SRSP tree. The stationary distribution of the Markov chain is then the posterior density over the space of all possible histograms. In this poster we will present the SRSP data-structure in the context of density estimation and provide simulation results for several d-dimensional distributions where $1 < d < 1000$.

Stochastic Volatility Models Including Open, Close, High and Low Prices

Enrique ter Horst
Euromed Management, France
enriqueterhorst@gmail.com

Abel Rodríguez
Univ. California Santa Cruz, USA
abel@soe.ucsc.edu

Henryk Gzyl
IESA, Venezuela
henryk.gzyl@gmail.com

Germán Molina
Tudor Hedge Fund, UK
german@germanmolina.com

Mounting empirical evidence suggests that the observed extreme prices within a trading period can provide valuable information about the volatility of the process within that period. In this paper we define a class of stochastic volatility models that uses opening and closing prices along with the minimum and maximum prices within a trading period to infer the dynamics underlying the volatility process of asset prices and compares it with similar models that have been previously presented in the literature. The paper also discusses sequential Monte Carlo algorithms to fit this class of models and illustrates its features using both a simulation study and data form the S&P500 index.

Marginally Specified Hierarchical Models for Relational Data

Andrew C. Thomas
Carnegie Mellon University, USA
acthomas@stat.cmu.edu

Joseph K. Blitzstein
Harvard University, USA
blitzstein@stat.harvard.edu

We present a unified approach to modelling dyadic relational data, namely that seen in social, biological and technological networks, without restriction to the binary format. The approach involves three principles: considering the marginal specification of any edge as the fundamental unit, embedding as much dependence as possible in latent structural forms, and using distributional forms that favour

high-throughput computational methods for their solution. We show that this approach allows for an extremely flexible and generalizable way of describing the structural properties of relational systems; namely, we offer alternate explanations for two approaches popular in the networks literature, the "small-world" and "scale-free" mechanisms, and demonstrate the ability of marginal hierarchical modelling to expand beyond them.

A Goodness of Fit Test of the Normal Distributions Against a Dirichlet Process Mixture Alternative

Surya T Tokdar
Duke University, USA
tokdar@stat.duke.edu

While testing the goodness-of-fit of a parametric family has a long and rich history in the frequentist literature, a comprehensive Bayesian treatment is yet to develop. Half of the difficulty lies in carefully specifying the alternative distribution, the other half lies in computing the Bayes factor. The Polýa tree mixture alternative of Berger and Guglielmi (2001 JASA vol. 96 pp 174–184), though computationally tractable, is somewhat unappealing due its inherent lack of smoothness and difficulty in generalizing to multiple dimensions. We explore an alternative based on Dirichlet process mixtures that overcomes both these problems. The mixture perspective offers a well interpreted extension of the null model and is easy to compute with. Our main contribution lies in a careful construction of the Dirichlet process mixture model that appropriately embeds the null model and justifies the use of a common, non-informative prior on the parameter that indexes this embedding map.

Reference Analysis and Precise Hypothesis Testing of the Hardy-Weinberg Equilibrium

Vera L. Tomazella
Universidade Federal de São Carlos
vera@power.ufscar.br

José M. Bernardo
Universitat de València, Spain
jose.m.bernardo@uv.es

Many of the problems which traditionally have been formulated in terms of hypothesis testing are really complex decision problems on model choice, whose appropriate solution naturally depends on the structure of the problem. In this paper a probability model for the formation of genotypes from two alleles is given and expressed in terms of two parameters, α and β, with $\alpha = 0$ corresponding to Hardy-Weinberg equilibrium (Lindley, *Valencia* 3, 1988), a particular scientific hypothesis of genetical importance. In some genetical applications the proportion of A alleles is known fairly precisely before sampling, and special attention is given to this particular case, which corresponds to the case where β is known. The relevant precise hypothesis testing problem is considered from a decision-theoretic viewpoint, where the null hypothesis is rejected if the null model is expected to

be too far from the true model in the logarithmic divergence (Kullback-Leibler) sense. The results are illustrated using examples with data previously analyzed in the literature

Software Failure Prediction using Metrics Information via Gaussian Process Models.

Nuria Torrado Michael Wiper and Rosa Lillo
Universidad Carlos III de Madrid, Spain
{nuria.torrado, michael.wiper, rosaelvira.lillo}@uc3m.es

Software reliability is one of a number of aspects of computer software which can be taken into consideration when determining the quality of the software in terms of the probability of not encountering any bugs over a specified period of time. Research on how to predict software failures accurately is of great practical importance. A large number of models have been proposed to address this topic, but a few incorporate some significant metrics data observed in software testing.

We consider the prediction of times between software failures or numbers of failures in a given time when it is assumed that the software is possibly imperfectly repaired after each failure and that software metrics information for each version of the software is available. Our model is a hierarchical non-parametric regression model based on exponential inter-failure times or Poisson failure counts where the rates are modeled as Gaussian processes with software metrics data such as lines of code, complexity measures or even previous failure times are used as inputs. We undertake fully Bayesian inference based on MCMC and illustrate our approach with real software failure data. Also, we make comparisons between our model and the classical models of software reliability based on nonhomogeneous Poisson processes.

A Temporal Bayesian Model with Discrete response: An Application in Epidemiology

Francisco Torres-Avilés
Universidad de Santiago de Chile, Chile
francisco.torres@usach.cl

Nelly González and Elena Carrasco
Universidad de Santiago de Chile, Chile
nelly_gonzalez112@yahoo.es ecarrasc.carrasco@gmail.com

This work presents a Bayesian methodology for modeling discrete data considering a temporal structure of dependence. Descriptive temporal studies developed in order to estimate the incidence rate of Type 1 Diabetes Mellitus in Chile have motivated the proposal of an adequate model to explore and determine the existence of seasonality and trend, both implicit in the temporal behavior of the available information. A temporal Bayesian generalized linear model with autoregressive errors and lags in the observations is proposed. Environmental measures, such as pollution, are incorporated as covariates looking for an explanation for the growing incidence rate of Diabetes Mellitus in the Metropolitan Region of Chile, between 2000 and 2008.

Credible Intervals that Should Impress Frequentists

Frank Tuyl
University of Newcastle, Australia
frank.tuyl@newcastle.edu.au

In the context of Bayesian estimation, the aim of this article is to prescribe a consensus interval, i.e. a generally agreed on method of interval calculation given a posterior. The latter would typically be based on a consensus prior, i.e. a generally agreed on noninformative prior. The proposed interval is not new: given a posterior, the lower and upper limits are found by ensuring they have the same likelihood function value; however, it seems that its importance with respect to frequentist coverage was not considered by e.g. Box & Tiao (1973). A more interesting view of this interval is as the highest posterior density (HPD) interval for the function of the parameter of interest for which the chosen prior is uniform, so that the name UHPD seems appropriate: Uniform prior based HPD. While generally not necessary, for practical purposes this uniform prior could be considered to be defined on $[0, 1]$, for extreme data leading to one-sided intervals when required.

In the presence of nuisance parameters, it is proposed to choose an "appropriate" likelihood of the parameter of interest. Based on an examination of various examples, it appears that there is no single type of likelihood that is suitable for all such models. For the usual quantities of interest following from the 2×2 contingency table, such as the relative risk θ_1/θ_2, it appears that the simple profile likelihood is suitable for this purpose. Based on uniform priors for θ_1 and θ_2, resulting intervals have the advantage of nominal mean coverage in (θ_1, θ_2), reasonable minimum coverage, and convergence to the usual sample estimates when $\alpha \to 0$.

Bayesian Adaptive Designs of Clinical Trials. Example of CIGB-300 in Phase I Studies.

Carmen Valenzuela,
Centro de Ingeniería Genética y Biotecnología, Cuba
carmen.valenzuela@cigb.edu.cu

Elizeth García Angela Tuero Pedro López
Idania Baladron Lidia González Idrian García
Boris Acevedo and Silvio Perea
Centro de Ingeniería Genética y Biotecnología, Cuba
{elizeth.garcia, angela.tuero, lopez.saura}@cigb.edu.cu
{idania.baladron,lidia.gonzalez,idrian.garcia}@cigb.edu.cu
{boris.acevedo, silvio.perea}@cigb.edu.cu

In recent years, the use of adaptive design methods in clinical research and development based on accrued data has become very popular due to its flexibility and efficiency. Phase I clinical trials are an essential step in the development of anticancer drugs and the use of adaptive methods has become a powerful tools to

increase the success rate (of only 5% in oncology). This work presents multiple-adaptive designs of CIGB-300 (anticancer peptide) in some phase I cancer studies in Cervical Intraepithelial Neoplasia and other solid tumors as example of use of all available information in the process of taking decision as strategy of development, using Bayesian analysis as the more natural way to do it.

Functional Data Analysis with Latent Variables

Angelika van der Linde
University of Bremen, Germany
avdl@math.uni-bremen.de

In functional data analysis (FDA) it is of interest to generalize techniques of multivariate analysis like regression, principal components analysis (PCA) or canonical correlation analysis (CCA) from vectors to functions, which are often observed with noise. While there are Bayesian methods for multivariate analysis, the estimation of functions is often handled in a nonparametric way, and a combination of these approaches to develop a *nonparametric* FDA is difficult. *Parametric* two stage procedures by which functions are individually fitted within a basis and the corresponding vectors of coefficients subsequently subjected to PCA or CCA require that sufficient data is available to reconstruct each single curve.

In the proposed Bayesian approach to FDA partial curves and non-Gaussian error schemes can be handled. To this end two tools are combined: (i) a special (Demmler-Reinsch like) basis of interpolation splines to flexibly represent functions; (ii) latent variable models for probabilistic PCA or CCA of the corresponding coefficients, in which covariance structures are estimated only in their dominating parts. The number of latent variables and the number of necessary basis functions (acting as a smoothing parameter) are treated as model parameters for which point estimates are obtained. Bayesian inference is based on a variational algorithm such that computations are straight forward and fast corresponding to an idea of FDA as a toolbox for explorative data analysis. The performance of the approach is illustrated with synthetic and real data sets.

Metropolis Hastings with Partially Collapsed Gibbs Samplers with Application in High-Energy Astrophysics

David A. van Dyk
Univ. California, Irvine, USA
dvd@ics.uci.edu

Taeyoung Park
University of Pittsburgh, USA
tpark@pitt.edu

The popularity of the Gibbs sampler stems from its simplicity and power to effectively generate samples from a high-dimensional probability distribution. It can sometimes, however, be very slow to convergence, especially when it is used to fit highly structured complex models. The recently proposed Partially Collapsed Gibbs (PCG) sampler offers a new strategy for improving the convergence characteristics of a Gibbs sampler (van Dyk and Park, 2008, *JASA*, Park and van

Dyk, 2009, *J Comp Graph Statist*). A PCG sampler achieves faster convergence by reducing the conditioning in some or all of the component draws of its parent Gibbs sampler. Although this strategy can significantly improve convergence, it must be implemented with care to be sure that the desired stationary distribution is preserved. In some cases the set of conditional distributions sampled in a PCG sampler may be functionally incompatible and permuting the order of the draws can change the stationary distribution of the chain.

As in an ordinary Gibbs sampler, we sometimes find that one or more of the conditional draws of a PCG sampler is not available in closed form and we may consider implementing such draws with the help of the Metropolis-Hastings sampler. Doing so, however, must be done with care. Even in a simple two-step Metropolis within Gibbs sampler, reducing the conditioning in one draw in an effort to form a Metropolis within PCG sampler may alter the stationary distribution of the chain. This can happen even when the PCG sampler would work perfectly well if all the conditional updates were available without resorting to Metropolis updates. In this poster we illustrates the difficulties that may arise when using Metropolis-Hastings updates within a PCG sampler and develop a general strategy for using such updates while maintaining the target stationary distribution. Our proposed computational methods are illustrated in an example from High-Energy Astrophysics that involves sampling from the complex distributions that arise when uncertainty in instrument calibration is quantified via a Bayesian posterior distribution.

The Variational Bayes Method for Inverse Regression Problems with Application to the Palaeoclimate Reconstruction

Richa Vatsa and Simon Wilson
Trinity College Dublin, Ireland
vatsabr@tcd.ie swilson@tcd.ie

The palaeoclimate reconstruction problem is formulated as an example of an inverse regression problem. In the reconstruction problem, past climate is inferred using pollen data. The inference on climate is carried out in two stages. At the first stage, a regression model of pollen response to climate is built using modern data on pollen. At the second stage, the knowledge of responses of pollen is used to infer past climate from data on ancient pollen prevalence.

We implement a Bayesian approach to the reconstruction problem. The inference problem is highly multivariate and so computationally challenging. We present Variational Bayes approximation for the inference on past climate. The Variational Bayes (VB) Method, a functional approximation is quick and easy to implement. The VB method permits more dimensions of climate, and fit the data with more types of pollen taxa, than has been possible with MCMC. However this is at the expense of approximating independence assumptions. We explore the consequence of these approximations on the inference.

The pollen data exhibit zero-inflated behaviour and should be modeled with a zero-inflated distribution. The VB method is restricted to posterior distributions that factorize in a certain ways and standard zero-inflated models do not fit into

that category. We discuss ways in which the VB method can be adapted to work in these situations. We also discuss the application of multiscale modelling with the VB method for a quick and its successful approximation for the model with many unknown climates variables.

Bayesian Spatial Model Selection.

Ciro Velasco and Scotland Leman
Virginia Tech, USA
cvelasco@vt.edu leman@vt.edu

Spatial statistical analysis is used for studying data coming from an area in which observations over a region present a degree of dependence with their neighbors. The possibility of accounting for such dependence is the main purpose the spatial analysis. It is believed that the dependence between a pair of observations is appropriately modeled using some of the covariogram functions: Exponential, Guassian, etc. This time we are concerned on Spatial Model Selection. The Response is supposed to be related to a collection of regressors through linear function, but the relation could be stronger for some regressors. Thus the goal is to find those regressors. If the number of regressors is large, then the number of possible models is large as well. Geweke (1994) proposed a model selection procedure, based on a point mass mixed model prior of the parameters. Model selection combined with Spatial Analysis lead us to an interesting problem: how to select the best model(s) within subregions of a region of study. This will raise several difficulties: Find the best way to split the region into subregions that could make sense for further analysis and also for its interpretation, and Select a model taking into account the dependence that those subregions would entail. It is assumed that the presence of a given regressor into the selected model within a subregion would conform to the dependence structure in the whole region. It would mean that if a particular regressor is important in the selected model in a subregion but not in any other sub-region across the region, this would be because in that region a particular event is taking place that causes the regressor to be important, because the cause-effect its neighboring regions are inducing.

Recent Advances on Bayesian inference for $P(X > Y)$

Laura Ventura
Università di Padova, Italy
ventura@stat.unipd.it

Walter Racugno
Università di Cagliari, Italy
racugno@unica.it

This contribution deals with the statistical problem of evaluating $R = P(X > Y)$, where X and Y are two independent random variables. Bayesian parametric inference about R has been discussed under various distributional assumptions on $X \sim f(x|\theta_x)$ and $Y \sim f(y|\theta_y)$ (Kotz *et al.* 2003), with $\theta = (\theta_x, \theta_y) \in \Theta \subset \mathbb{R}^p$, $p \geq 2$. The classical way to perform Bayesian inference is to derive the posterior density of R, using transformation rules. This approach requires both elicitation of a prior on the complete parameter θ and numerical integration in order to

derive the marginal distribution of R. In this paper, we discuss and apply recent advances in Bayesian inference, based on higher-order asymptotics and pseudo-likelihoods and related matching priors, to perform accurate inference on the parameter of interest R only (Ventura *et al.* 2009). The proposed approach has the advantages of avoiding the elicitation on the nuisance parameters and the computation of multidimensional integrals. These advantages are illustrated assuming different distributions within the exponential family.

Quantity R appears in many statistical problems, including quality control, engineering statistics, medical statistics and biostatistics, among others. For example, in a clinical study, X is the response of a control group, Y the response of a treatment group and R measures the effectiveness of the treatment. Alternatively, for diagnostic tests used to distinguish between diseased and non-diseased patients, the area under the ROC curve, based on the sensitivity and the complement to specificity at different cut-off points of the range of possible test values, is equal to R. In this paper, the accuracy of the proposed methodology is illustrated both by numerical studies and by real-life data concerning results of two different diagnostic tests.

Testing Hardy-Weinberg Equilibrium: An Objective Bayesian Analysis

Sergio Venturini
Università Bocconi, Italy
sergio.venturini@unibocconi.it

Guido Consonni
Università di Pavia, Italy
guido.consonni@unipv.it

Elías Moreno
Universidad de Granada, Spain
emoreno@ugr.es

We analyze the general (multiallelic) Hardy-Weinberg equilibrium problem from an objective Bayesian testing standpoint. We argue that for small or moderate sample sizes the answer is rather sensitive to the prior chosen, and this suggests to carry out a Bayesian sensitivity analysis to the prior. This objective is achieved through the identification of a class of priors specifically designed for this testing problem. In this paper we consider the class of intrinsic priors under the full model, indexed by a tuning quantity, the training sample size. These priors are objective, satisfy Savage's continuity condition and have proved to behave extremely well for many statistical testing problems. We compute the posterior probability of the Hardy-Weinberg equilibrium model for the class of intrinsic priors, and thus provide a range of plausible answers. If our decision of rejecting the null hypothesis does not change as the intrinsic prior varies over this class, we conclude that our analysis is robust. Illustrations on simulated and real data are provided.

Visualising the Input Space of a Galaxy Formation Simulation

Ian Vernon Michael Goldstein and Richard Bower
Durham University, UK
{i.r.vernon, michael.goldstein, r.g.bower}@durham.ac.uk

The question of whether there exists large quantities of Dark Matter in our Universe is one of the most important problems in modern cosmology. This project deals with a complex model of the Universe known as Galform, developed by the ICC group, at Durham University. This model simulates the creation and evolution of approximately 1 million galaxies from the beginning of the Universe until the current day, a process which is very sensitive to the presence of Dark Matter. A major problem that the cosmologists face is that Galform requires the specification of a large number of input parameters in order to run. The outputs of Galform can be compared to available observational data, and the general goal of the project is to identify which input parameter specifications will give rise to acceptable matches between model output and observed data, given the many types of uncertainty present in such a situation. As the model is slow to run, and the input space large, this is a very difficult task. We have solved this problem using general techniques related to the Bayesian treatment of uncertainty for computer models (see the MUCM website:
http://mucm.group.shef.ac.uk/index.html). These techniques are centred around the use of emulators: fast stochastic approximations to the full Galform model. Visualising the results of such an analysis is a non-trivial task. The acceptable region of input space is a complex shape in high dimension. Although the emulators are fast to evaluate, they still cannot give detailed coverage of the full volume. We have therefore developed fast emulation techniques specifically targeted at producing lower dimensional visualisations of higher dimensional objects, leading to novel, dynamic 2- and 3-dimensional projections of the acceptable input region. These visualisation techniques allow full exploitation of the emulators, and provide the cosmologists with vital physical insight into the behaviour of the Galform model.

Hierarchical Bayesian Models for Manpower Planning

Leonel Vicente
Instituto Politécnico de Lisboa, Leiria, Portugal
lsv@estg.ipl.pt

Kamil Feridun Turkman
Universidade de Lisboa, Portugal
kfturkman@fc.ul.pt

Manpower Planning is a useful tool for human resource management in large organizations. Effective management requires reliable mathematical models, so that the effects of management action can be predicted, and the uncertainty in these predictions quantified. Classical Manpower Planning models are analytical time-discrete push and pull models. Typically, stocks form a multidimensional counting

process and except the wastages, flows are assumed to be fixed proportions representing the stationary flow of people between stocks in different grades. In this work, we aim to make inference on the time dynamics of probability structure of this multidimensional markov chain, contrary to the usual study of expected values. The objective is to treat the chain as non-homogeneous and capture the time dynamics by describing it through a fully Bayesian hierarchical model. The Likelihood can be written and using MCMC techniques, the predictive distribution of the future stocks and flows can be simulated leading the way for inference based on this predictive distribution. The prediction as well as model errors then can be quantified with the usual methods of Bayesian analysis. We apply these models to study the manpower structure of the Faculty of Sciences of Lisbon for a pilot analysis before applying to larger structures. Using the computer package Winbugs, we demonstrate how such a model can be used for exploring impact of different management actions, such as recruitment and promotions.
Finally, using adaptive strategies, we simulate the possible policy changes for attaining a desired structure.

Posterior Modal Estimation in Elliptical Measurement Error Models

Ignacio Vidal
Universidad de Talca, Chile
ividal@inst-mat.utalca.cl

Heleno Bolfarini
Universidade de São Paulo, Brasil
hbolfar@ime.usp.br

The main object of this work is to discuss Bayes estimation of the regression coefficient in the dependent elliptically distributed simple regression model with measurement errors. The posterior distribution for the line parameters is obtained in a closed form, considering informative prior distributions for the error precision parameters. Investigation of the posterior mode of the regression coefficient indicates the possibility of at most two real modes being more likely, specially in large samples, situations with only one mode. Precision of the modal estimators are studied by deriving the Hessian matrix. Connections with results in the literature are reported.

A Bayesian Hierarchical ^{13}Carbon-Octanoate Breath Test to Assess Gastric Emptying

Georgios Vlasakakis Leslie J.C. Bluck
MRC Human Nutrition Research, UK
{georgios.vlasakakis, les.bluck}@mrc-hnr.cam.ac.uk

Sarah J.Jackson
MRC Human Nutrition Res., UK
sarah.casey@mrc-hnr.cam.ac.uk

Adrian Mander
MRC Biostatistics Unit, UK
adrian.mander@mrc-bsu.cam.ac.uk

The ^{13}C-Octanoate acid Breath Test (^{13}COBT) was first introduced as an indirect, low cost and ethically acceptable method to measure gastric emptying in man. Due to its non-radioactive nature the test is suitable for repetitive studies

as well as in investigations with pregnant women and children. The octanoate, in common with all medium chain fatty acids, passes unchanged through the stomach enters the small intestine where it is rapidly absorbed and then transported into the liver where it undergoes an immediate oxidation into CO_2 and water. Subsequently, the only limiting factor for its metabolism is the time interval that it actually resides in the stomach. We adopted a Bayesian Hierarchical Model to interpret breath measurements obtained during a ^{13}COBT and we performed the statistical analysis using the computer package WinBUGS. We adopted a hierarchical setting where poorly defined parameters associated with a delayed gastric emptying were able to borrow strength from global distributions and we compared the results with their Non-linear Least Squares (NLS) counterparts. NLS method failed to return acceptable estimates in cases where gastric emptying was slow whereas the Bayesian model was capable to correct model's failures and data inconsistencies. Direct comparison of the two parameters describing gastric emptying (t_{lag}, $t_{1/2}$) revealed a strong correlation between the two methods. Despite our large dataset (n=164), Bayesian modelling was fast and successful for all subjects. We would recommend that the WinBUGS package be adopted routinely for the estimation of gastric emptying by ^{13}C-Octanoate Breath Test.

Sampling from the Posterior- MCMC, Importance resampling or Numerical integration?

Dina Vrousai
Trinity College Dublin, Ireland
vrousair@tcd.ie

John Haslett
Trinity College Dublin, Ireland
John.Haslett@tcd.ie

Many methods and algorithms have been developed to sample from the posterior distribution. Importance resampling (IR) and particularly Markov Chain Monte Carlo (MCMC) methods are widely used for this purpose. Sampling from the posterior using these methods doesn't require the knowledge of the normalizing constant. Another alternative is to compute the normalizing constant and then to sample from the posterior. This can be very computationally demanding, especially in high dimensional problems.

We are using an R package, lately released, which implements multidimensional integration algorithms, only for Riemann integrals (unit hypercube). The aim is to compare the special characteristics of these three methods (IR, MCMC, Numerical integration) using an application on blood lactate data. We are using Kriging with Gaussian processes to model these data. We then compare the posterior distributions for our model obtained using these three different methods (MCMC, IR and Numerical integration).

Do Rare Genetic Variants Lead to Greater Disease risk? Inferring the Relationship between Allele Frequency and Relative Risk using Replicated Disease Loci

Damjan Vukcevic
University of Oxford, UK
vukcevic@well.ox.ac.uk

Chris Holmes and Peter Donnelly
University of Oxford, UK
cholmes@stats.ox.ac.uk peter.donnelly@well.ox.ac.uk

Genome-wide association studies (GWAS) have become a major tool for investigating the genetic basis of common complex human diseases. These case-control studies collect DNA from typically thousands of individuals and read their genetic types at a large number of loci, now often up to a million markers, along the genome. By comparing diseased and healthy individuals, genetic mutations which are associated with disease risk can be discovered.

A major question of interest is whether the frequency of a mutation in the population is informative of its effect on disease risk. In particular, it is hypothesised that rare variants are more likely to lead to higher risk. Here we tackle this question using Bayesian inference on a database of GWAS discoveries from a range of diseases.

The analysis is challenging due to an ascertainment effect: the power of observing a significant association in a study depends on both the frequency and the effect size of the genetic variant. By making suitable prior assumptions about study designs based on the information in the database and modelling the ascertainment process, we show how to take this bias into account. We then conduct a sensitivity analysis across a range of plausible models and priors.

This work helps to shed light on the major question above, particularly in determining to what extent our current discoveries provide evidence in favour of any particular relationship. It can also inform future studies, by providing more suitable priors to use in their analyses.

Nested Dirichlet Process with Shared Mixture Components

Radhakrishna Vuppala and Abel Rodriguez
University of California, Santa Cruz, USA
vrk@soe.ucsc.edu abel@soe.ucsc.edu

We consider clustering problems for nested samples where the goal is to cluster within and between different levels of the nesting hierarchy. We build on existing Bayesian nonparametric models by combining the structure of the hierarchical Dirichlet process (HDP), which allows sharing of clusters across groups, and that of the nested Dirichlet process (NDP), which allows for multilevel clustering. The resulting construction induces a joint distribution on the population parameters that generalizes the Chinese Restaurant process, which we dub the "multi group Chinese Restaurant Franchise" (mCRF). We consider various algorithms for posterior inference, including a blocked Gibbs sampler and a variational approximation. The model is illustrated through applications in information retrieval and genetics.

A New Generalization of the Dirichlet Process with Application to Mixture Models

Sara Wade
Università Bocconi, Italy
sara.wade@phd.unibocconi.it

Silvia Mongelluzzo and Sonia Petrone
Università Bocconi, Italy
{silvia.mongelluzzo, sonia.petrone}@phd.unibocconi.it

The precision parameter α plays an important role in the Dirichlet Process. When assigning a Dirichlet Process prior to a random probability over $\mathcal{R}^k$ this can be quite restrictive in the sense that the random marginal probabilities are Dirichlet Processes all with the same parameter α. In essence, this says that we believe that X is distributed according to the marginal of P_0 (P_{0X}) just as strongly as we believe that Y is distributed according to P_{0Y}, and if we take a random sample from P_X and P_Y, we believe the number of clusters will be the same for X and Y. Many generalizations of the Dirichlet Process already exist, but they usually require intensive computations. The aim of this paper to consider a generalization of the Dirichlet Process so that our prior for a random probability measure over $\mathcal{R}^k, k > 1$, is more flexible yet still conjugate. In particular, we construct the process from the fact that the joint prob! ability can be reparametrized in terms of the product of the conditional and the marginal, and we give Dirichlet Process priors to each of these terms. The resulting joint distribution can also be obtained from the stick-breaking representation of the marginal and conditional. This joint distribution is no longer a stick-breaking prior, but still maintains desirable properties such as conjugacy. An interesting consequence of the prior is the clustering structure; it creates global clustering of the marginal variable and local clustering of the condition variables. We show that the process can also be constructed from a generalized Polya urn scheme, examine properties of the process, and consider applications to mixture models.

Robust Bayesian Analysis of Stochastic Volatility Model with Generalised-t Distribution via Scale Mixtures

Joanna J. J. Wang
The University of Sydney Australia
jwan4640@mail.usyd.edu.au

S.T. Boris Choy and Jennifer S.K. Chan
The University of Sydney, Australia
boris.choy@sydney.edu.au jchan@maths.usyd.edu.au

In modelling financial return time series and time-varying volatility, the Gaussian and Student-t distributions are widely used for the innovation terms in stochastic volatility (SV) models. In many real applications, it is found that the robust Student-t outperforms the Gaussian distribution because many financial data are heavy-tailed. However, other distributions such as the Laplace and generalised

error distributions are also common in SV modelling. Therefore, this paper proposes the use of the generalised-t (GT) distribution whose special cases are the Gaussian, Student-t, Laplace and generalised error distributions. Since the GT distribution is a member of the scale mixture of uniform (SMU) family of distribution, we handle the GT distribution via its SMU representation in the SV modelling. We show that such a SMU representation can substantially simplify the Gibbs sampler for Bayesian simulation-based computation and can provide a means for identifying outlying returns and volatilities. In an empirical study, we adopt a GT-SV model to fit the daily return data of the exchange rate of the Australian dollar to three other currencies and the exchange rate to the American dollar is used as a covariate in the return equation. Model implementation relies on Bayesian Markov chain Monte Carlo (MCMC) algorithms using the WinBUGS package.

Network Routing in Counterterrorism: An Adversarial Risk Analysis Framework

Shouqiang Wang
Duke University, USA
sw55@duke.edu

David Banks
Duke University, USA
banks@stat.duke.edu

This paper applies an adversarial risk analysis (ARA) framework to the problem of selecting a route through a network in which an opponent chooses links to block. The motivating application is convoy routing across a road network when there may be improvised explosive devices and imperfect intelligence about their locations.

The ARA framework attempts to combine statistical risk analysis with game theory in ways that support practical decision-making. Formally, the decision maker will conduct a "mirroring" argument to model the opponent's decision-making process based on his/her subjective probability assessment about the uncertainty of the payoff functions. The fixed-point of such an mirroring argument, a pair of probability distributions over the decision maker's and the opponent's action spaces, will provide the decision-maker with a subjective prediction of the opponent's behavior, based on which the decision-maker should optimize his/her own strategies. In the context of counterterrorism network routing, we prove the existence of such an ARA fixed-point solution. To implement this solution concept, we use a simulation-based algorithm of the Round Robin type. We also show a sufficient condition that guarantees the convergence of such an algorithm.

Bayesian Dynamic Item Response model

Xiaojing Wang James O. Berger
Duke University, USA
xiaojing.wang@duke.edu berger@stat.duke.edu

Donald S. Burdick
MetaMetrics, Inc., USA
DBurdick@lexile.com

Item response models enjoy popularity in large-scale educational and measurement testing, many studies have contributed to its development and applications in static time domain. However,when there are repeated observations of individuals over the time,a dynamic structure should be considered for the latent trait of persons' ability. In the paper,a new class of state space models for time series dichotomous response data is proposed, which not only shows serial dependence in the data,but also allows effects of the person's ability growth to be varying in different time lapse. A novel Markov Chain Monte Carlo algorithm is proposed for statistical inference, which incorporates isotonic constraints on the latent trait of the ability without extra burden for computation. Simulation studies show the proposed models are reasonable. Finally, our models are applied to the 2 years reading test data conducted by MetaMetrics, Inc.. The application results show that the proposed models provide an appealing approach for measuring and monitoring persons' reading ability growth.

Sparse Process Classification via the Gaussian Copula

Fabian L. Wauthier and Michael I. Jordan
University of California, Berkeley, USA
flw@cs.berkeley.edu jordan@cs.berkeley.edu

Gaussian process classifiers (GPCs) provide a Bayesian approach to nonparametric classification that have the key advantage of providing predictive class probabilities. We have investigated a GPC-based approach to improve the accuracy of conditional probabilities estimated in sparsely populated regions of an input space for the purpose of protein structure modeling. Our approach selectively shrinks conditional probability estimates towards uniform values: estimators in sparsely sampled regions should experience stronger shrinkage than in densely sampled regions. We induce this shrinkage by replacing the Gaussian process underlying the GPC with a stochastic process that has sparse marginals and for which inference is still tractable. We embed the stochastic process in a classification model and analyze it in a simple scenario to argue that selective shrinkage occurs provided the stochastic process marginals are sufficiently sparse. Since predictive distributions in the new classification model are intractable to compute, we approximate them by Laplace's method. For a given set of training data we estimate hyperparameters by maximizing a regularized approximate marginal likelihood. Our experiments compare the sparse process classifier (SPC) on prediction tasks involving 17 different amino acids against a GPC model that is similarly estimated by maximizing a regularized approximate marginal likelihood. The results show

that when restricting evaluation to test data from a predefined sparse region of input space, the SPC almost always outperforms the GPC, sometimes by up to 21%, and degrades by at most 1% on two amino acids. On average, SPC outperforms GPC by about 5.5%. When evaluating the algorithms on test data from the entire input space, SPC almost never degrades by more than 1% compared to GPC.

An Encompassing Prior Generalization of the Savage-Dickey Density Ratio

Ruud Wetzels
Universiteit van Amsterdam, The Netherlands
wetzels.ruud@gmail.com

Raoul P.P.P. Grasman and Eric-Jan Wagenmakers
Universiteit van Amsterdam, The Netherlands
R.P.P.P.Grasman@uva.nl EJ.Wagenmakers@gmail.com

Hoijtink, Klugkist, and colleagues proposed an encompassing prior (EP) approach to facilitate Bayesian model selection for nested models with inequality constraints. In this approach, samples are drawn from the prior and posterior distributions of an encompassing model that contains an inequality restricted version as a special case. The Bayes factor in favor of the inequality restriction then simplifies to the ratio of the proportions of posterior and prior samples consistent with the inequality restriction. To date, this formalism has been applied almost exclusively to models with inequality or about equality constraints. Here it is shown that the EP approach naturally extends to exact equality constraints by considering the ratio of the heights for the posterior and prior distributions at the point that is subject to test (i.e., the Savage-Dickey density ratio). The EP approach generalizes the Savage-Dickey ratio method, and can accommodate both inequality and exact equality constraints. Unfortunately, the EP approach to exact equality constraints is vulnerable to the Borel-Kolmogorov paradox. We conclude that the general EP approach is an elegant and computationally efficient procedure to calculate Bayes factors for nested models, but that users should be aware of its limitations and underlying assumptions.

Improved Approximate Sum-Product Inference Using Multiplicative Error Bounds

Ydo Wexler and Christopher Meek
Microsoft Research, USA
ydowexler@gmail.com meek@microsoft.com

We describe the Multiplicative Approximation Scheme (MAS) for approxi- mate inference in multiplicative models. We apply this scheme to develop the DynaDecomp approximation algorithm. This algorithm can be used to obtain bounded

approximations for various types of sum-product-product problems including the computation of the log probability of evidence and the log- partition function. We demonstrate that this algorithm yields bounded ap- proximations superior to existing methods using a variety of large graphical models.

Bayesian Hidden Markov Models for the Analysis of Extracellular Recordings

Nicole White
Queensland UT, Australia
nm.white@qut.edu.au

Helen Johnson
Queensland UT, Australia
h.johnson@qut.edu.au

Peter Silburn
St. Andrew's Hospital , Australia
p.silburn@neurosciencesqld.com.au

Kerrie Mengersen
Queensland UT, Australia
k.mengersen@qut.edu.au

The reliable detection and classification of spikes in extracellular recordings is fundamental to studies in clinical neuroscience. However, of the techniques currently used in practice, many are highly supervised, are based on "hard" classification and treat the steps of detection and classification as independent processes. Consequently, the quality of classification is in part dependent on the chosen method of detection. To this end, there is a need to develop models that combine these steps into a single, data-driven algorithm. Furthermore, there is a need for strategies that properly manage the uncertainty in detection and classification.

We consider Bayesian hidden Markov models (HMMs) as generative models for the analysis of multiple action potentials in extracellular recordings. In particular, we focus on data with low signal-to-noise ratios, a complicating factor in reliable spike detection, and irregular firing activity or "bursting". The latter feature is explored in two settings; a fixed multiple changepoint and hierarchical Dirichlet process model. All models are estimated by MCMC, with posterior predictive checks for identified spikes by neuron. Motivating this work is the analysis of microelectrode recordings of the subthalamic nucleus, a surgical target for Deep Brain Stimulation in the treatment of Parkinson's disease. By exploring these data, we aim to contribute to better understanding of neural behaviour in PD, including both the effects of stimulation and execution of intraoperative patient tasks.

Parameter Inference for Stochastic Kinetic Models of Bacterial Gene Regulation: A Bayesian Approach to Systems Biology

Darren J. Wilkinson
Newcastle University, UK
d.j.wilkinson@newcastle.ac.uk

Bacteria are single-celled organisms which often display heterogeneous behaviour, even among populations of genetically identical cells in uniform environmental

conditions. Markov process models arising from the theory of stochastic chemical kinetics are often used to understand the genetic regulation of the behaviour of individual bacterial cells. However, such models often contain uncertain parameters which need to be estimated from experimental data. Parameter estimation for complex high-dimensional Markov process models using diverse, partial, noisy and poorly calibrated time-course experimental data is a challenging inferential problem, but a computationally intensive Bayesian approach turns out to be effective. The utility and added-value of the approach is demonstrated in the context of a stochastic model of a key cellular decision made by the gram-positive bacterium Bacil lus subtilis, using quantitative data from single-cell fluorescence microscopy and flow cytometry experiments.

Decision Support for Policy Makers using the Bayesian Analysis of Computer Models

Daniel Williamson
Durham University, UK
daniel.williamson@durham.ac.uk

Michael Goldstein
Durham University, UK
michael.goldstein@durham.ac.u

Policy makers increasingly rely on computer models to aid policy judgements for complex systems. The climate system, for example, is extremely complicated and its reaction to changes in radiative forcing through CO_2 emissions can only be explored using models. Bayesian methods for making inferences about the physical system that combine information from computer simulators and observations of the system have become increasingly well studied. We apply some of these methods to the policy problem where the decisions to be made are inputs to the computer model. Particular features of our methodologies include: the provision of Bayesian decision support for the policy problem when it is known that policy may be adapted in reaction to future observations of the complex system; and careful integration of the knowledge that our computer simulators will evolve and improve over time, which may affect downstream strategies and, hence, current policy.

Our methods also allow research investment questions to be explored in the context of the wider policy problem. For example, the question of whether or not an improved version of a computer simulator should be build and how much it should be run can be addressed as part of the policy problem. We present our ideas motivated by the problem of making policy in order to address the possibility of rapid climate change.

Dependent Beta Processes

Sinead Williamson
University of Cambridge, UK
saw56@cam.ac.uk

Peter Orbanz and Zoubin Ghahramani
University of Cambridge, UK
po245@cam.ac.uk zoubin@cam.ac.uk

The beta process, a pure-jump independent increments process introduced by (N. L. Hjort, 1990), has proven a flexible nonparametric prior over cumulative hazard functions in survival and event history modeling. More recently, beta processes have proven useful in defining sparse nonparametric latent feature models. In such models, the pattern of sparse features is given by a draw from an Indian Buffet Process (IBP, T. L. Griffiths, Z. Ghahramani, 2005), which is a distribution over binary matrices with an unbounded number of columns. The IBP can be constructed using samples from a beta process (R. Thibaux, M. I. Jordan, 2007), in an exact analogy to the construction of the Chinese Restaurant Process from samples from a Dirichlet Process.

In survival applications, we may expect the cumulative hazard function to vary with some covariate - for example dose of treatment, or age of subjects. In latent variable models, we might expect the distribution over latent features to vary with some covariate - in document modeling applications the popularity of certain topics may wax and wane with time, or in a collaborative filtering application movie preferences might depend on age. The dependent Indian Buffet Process (dIBP, S. Williamson, P. Orbanz, Z. Ghahramani, 2010) provides a framework for addressing the latter, but is restricted to beta process with a fixed concentration parameter, and does not represent samples from the beta processes explicitly.

We present the dependent beta process (DBP), a dependent nonparametric process (S. Williamson, P. Orbanz, Z. Ghahramani, 2010) with beta process marginals. We describe a Gibbs sampler for sampling from the dependent beta process, and demonstrate its efficacy in a survival context.

Bayes Linear Kinematics in the Design of Experiments

Kevin J. Wilson and Malcolm Farrow
Newcastle University, UK
k.j.wilson@ncl.ac.uk malcolm.farrow@ncl.ac.uk

The choice of the design of an experiment, including both selection of design points and choice of sample size, can be viewed as a decision problem. We wish to choose the design which maximises the prior expectation of a utility function which depends on both costs of the experiment and benefits from the information gained. The latter may be realised through a second decision to be made after the experiment. Solving the problem requires the evaluation of this expectation for each candidate design involving summation or integration over all possible outcomes for each design. With non-Gaussian models, where posterior evaluations would typically involve intensive numerical methods such as Markov Chain

Monte Carlo, evaluation of the conditional expectation of the utility, given an outcome, becomes computationally demanding and so solving such design problems becomes difficult for all but fairly simple cases. Bayes linear kinematics (Golstein and Shaw, 2004) offers an alternative approach. It is the Bayes linear analogue of probability kinematics (Jeffrey 1965). It offers a method for propagating changes in belief about some quantities through to others within a Bayes linear structure, for example when the changes result from observing related non-Gaussian variables. We adopt a conjugate relationship between observables and parameters and then update beliefs about other quantities using Bayes linear kinematics. Applying this approach to the design problem greatly reduces the computational burden and the problem can be solved without the need for intensive numerical methods. The method is illustrated using two examples.

Model Prior Choice and Multiplicity Correction in Bayesian Model and Variable Selection

Melanie A. Wilson
Duke University, USA
maw27@stat.duke.edu

Edwin S. Iversen and Merlise A. Clyde
Duke University, USA
iversen@stat.duke.edu clyde@stat.duke.edu

The Bayesian approach to model selection allows for uncertainty in both model specific parameters and in the models themselves. Much of the recent Bayesian model uncertainty literature has focused on defining these prior distributions in an objective manner, providing conditions under which Bayes factors lead to the correct model selection, particularly in the situation where the number of variables, p, increases with the sample size, n. This is certainly the case in our motivation; the biological application of genetic association studies involving single nucleotide polymorphisms. While the most common approach to this problem has been to apply a marginal test to all genetic markers, we employ analytical strategies that improve upon these marginal methods by modeling the outcome variable as a function of a multivariate genetic profile using Bayesian variable selection. In doing so, we must be able to perform variable selection on a large number of correlated covariates within studies involving modest sample sizes. In this context we characterize commonly used prior distributions on the model space and investigate their implicit multiplicity correction properties first in the extreme case where the model includes an increasing number of redundant covariates and then under the case of full rank design matrices. We provide conditions on the asymptotic (in n and p) behavior of the model space prior to achieve consistent selection of the global hypothesis of at least one associated variable in the analysis in terms of the global posterior probabilities (0-1 loss). In particular, under the assumption that the null model is true, we show that the commonly used Uniform prior on the model space leads to inconsistent selection of the global hypothesis (the posterior probability of at least one association goes to 1) when the rank of the design matrix is finite. In the full rank case, we also show inconsistency when p goes to infinity faster than $\sqrt{n}$. Alternatively, we show that for any model space prior such that the global prior odds increases at a rate slower than $\sqrt{n}$ we have consistent selection of the global hypothesis in terms of posterior probabilities.

Bayesian Estimation of the Number of Unknown Species: Incorporating a Model for the Discovery Process

Simon P. Wilson Brett Houlding
Trinity College Dublin, Ireland
swilson@tcd.ie houldinb@tcd.ie

Mark Costello
University of Auckland, New Zealand
m.costello@auckland.ac.nz

We consider the problem of estimating unknown species number within an ecosystem. The usual approach estimates the number of distinct species S, with unknown abundances $N_1, \ldots, N_S$, by considering a sample of n individuals in which there are $R \leq S$ distinct species with abundances $n_1, \ldots, n_R$. Nevertheless, records of species discovery do not usually contain anything directly relating to abundance, so when faced with estimating global numbers of unknown species, we face the more difficult problem where neither the sample size n, nor the sample abundances $n_1, \ldots, n_R$, are available.

Here we assume that only the number of sampled species R and their first observation times $t_1, \ldots, t_R$ are available. To perform estimation of S using such data, the first observation times are linked to unknown sample abundances through a latent Gaussian 'effort' process. Furthermore, information concerning this process is assumed available by the annual number of distinct author surnames in manuscripts published on species discovery. Our approach is then to use information on the number of annual manuscripts published to perform inference on the parameters of the latent process. This in turn allows inference on the total sample size n. A series of constraints relating sample size with known first time of species observation and the number of samples taken between species discoveries, is then used to perform inference on the latter. Finally, with estimates of the number of samples taken between species discoveries, inference is performed on the unknown species abundances by utilizing traditional sampling theory approaches to species estimation.

Source Separation for Multi-Spectral Image Data with Gaussian Mixture Priors, with Application to the Cosmic Microwave Background

Simon P. Wilson
Trinity College Dublin, Ireland
simon.wilson@tcd.ie

Ercan Kuruoğlu
CNR Pisa, Italy
ercan.kuruoglu@isti.cnr.it

Jiwon Yoon
Trinity College Dublin, Ireland
yoonj@tcd.ie

Alicia Quirós
Univ. Rey Juan Carlos, Spain
alicia.quiros@urjc.es

Source separation is a common task in signal processing and is often analogous to factor analysis. In this work we look at a factor analysis model for source separation of multi-spectral image data where prior information about the sources

is quantified as a Gaussian mixture model with an unknown number of factors. Markov chain Monte Carlo techniques for model parameter estimation are used. The development of this methodology is motivated by the need to bring an efficient solution to the separation of components in the microwave radiation maps to be obtained by the satellite mission Planck which has the objective of uncovering cosmic microwave background radiation. The proposed algorithm successfully incorporates a rich variety of prior information available to us in this problem in contrast to most of the previous work that assumes completely blind separation of the sources. Results on realistic simulations of Planck maps and on WMAP 5th year results are shown. The technique suggested is easily applicable to other source separation applications by modifying some of the priors.

The computational challenges of this application are large. Multivariate prior mixture models, that incorporate spatial smoothness of sources and dependencies between them, considerably complicate implementation. In addition, Planck data consist of 9 images of order 10^7 pixels each. We explore various functional approximation approaches to computing marginal posterior distributions, and compare performance with the best MCMC algorithms that we have been able to implement.

Bayesian Nonparametric Modeling of the INAR(1) process

Michael P. Wiper
Univ. Carlos III de Madrid, Spain
michael.wiper@uc3m.es

Ana Justel
Univ. Autónoma de Madrid, Spain
ana.justel@uam.es

The INAR(1) process is defined by $x_t = \alpha \circ x_{t-1} + \epsilon_t$ where $x_t, \epsilon_t \in \{0, 1, 2, \ldots\}$ and $0 \leq \alpha \leq 1$. We assume a binomial thinning variable, that is $\alpha \circ X = \sum_{i=1}^{X} Y_j$ where Y_j is a Bernoulli variable with parameter α and attempt to model the error terms, ϵ_t, nonparametrically. Two possibilities are considered. Firstly, we assume that the errors are generated from an unknown distribution $\epsilon_t \sim F$ where a Dirichlet process prior is placed on F and secondly, we consider a Dirichlet process mixture model based on continuous mixtures of Poisson distributions. It is shown via both real and simulated examples that inference for the first model produces unreasonable solutions whereas the Dirichlet process mixture model provides reasonable estimates for the error distribution.

The Posterior Predictive Information Criterion

Andrew J Womack
Washington University in St. Louis, USA
ajwomack@wustl.edu

This paper introduces a new information theoretic model selection criterion based on posterior predictive densities. The motivation is to provide an objective Bayesian model selection criterion that allows the use of improper priors while not requiring the investigator to select separate priors to perform different inferential tasks. The criterion is a Kullback-Leibler correction to Aitkin's Posterior

Bayes Factor that compensates for overfitting. In contrast to existing information theoretic model selection criteria, because this criterion is based solely on predictive densities, it is not dependent on the choice of parameters of model focus. In the example of the linear model using the modified Jeffreys' prior, the method provides: asymptotic control of Type I error in null hypothesis testing; consistency in non-nested hypothesis testing; and - in contrast to intrinsic Bayesian methods - fully consistent tests in the case when the number of parameters grows as a fraction of the number of observations. As the method removes model selection from a strict ordering, it is more conservative than existing methods. This is reflected in the model averaging weights obtained from the criterion. Unlike marginal distributions, which are difficult to acquire through MCMC methods, the predictive densities are easily obtained through sampling. This allows one to quickly compute the model difference for two models using only the posterior distributions of the parameters of the models and the sampling distributions. Taken together, these results suggest that this criterion has the potential to become an everyday, widely applied tool of Bayesian inference.

A Bayesian Weighted Approach for Clustering of Spatial and/or Temporal Data

Darren Wraith Florence Forbes, and Senan Doyle
INRIA Grenoble Rhone-Alpes, France
darren.wraith@inrialpes.fr

We propose an extension for clustering approaches using a hidden Markov random field which provides for greater flexibility in cases where there are multiple sources of information, different distributional forms and/or groups which are under represented in the data. The approach extends the standard hidden Markov random field approach by including an additional missing variable which can convey information specific to the source of information and/or spatio-temporal domain. Prior information on the missing variable can be used to better identify specific groups of interest. We apply the model to the problem of fusing the output of multiple Magnetic Resonance sequences to robustly and accurately segment brain lesions for multiple sclerosis and stroke, with promising results.

Bayesian Methods of Estimating Population Size under Type I Censoring

Chang Xu and Dongchu Sun
University of Missouri, USA
cx3z9@mail.missouri.edu sund@missouri.edu

In many situations it is of interest to estimate the population size N. Generally, consider a system containing N faults and each fault could cause a failure. The failure time of each fault follows an i.i.d. exponential distribution with unknown mean. The observations are censored at a fixed time T (Type I) or at the time

when a fixed number of failures occurred (Type II). In this study we consider the Type I censoring case and derived several object priors using different approaches. We also conducted a simulation to compare the estimators of N under various priors.

Nonparametric Bayes Stochastically Ordered Latent Class Models

Hongxia Yang David Dunson and Sean O'Brien
Duke University, USA
hy35@duke.edu dunson@stat.duke.edu obrie027@mc.duke.edu

Latent class models (LCMs) are used increasingly for addressing a broad variety of problems, including sparse modeling of multivariate and longitudinal data, model-based clustering, and flexible inferences on predictor effects. Typical frequentist LCMs require estimation of a single finite number of classes, which does not increase with the sample size, and have a well-known sensitivity to parametric assumptions on the distributions within a class. Bayesian nonparametric methods have been developed to allow an infinite number of classes in the general population, with the number represented in a sample increasing with sample size. However, Bayes methods relying on Markov chain Monte Carlo sampling encounter a challenging label ambiguity problem, which makes it difficult to perform inferences on class-specific quantities. In this article, we propose a new nonparametric Bayes model that allows predictors to flexibly impact the allocation to latent classes, while limiting sensitivity to parametric assumptions and label switching problems by allowing class-specific distributions to be unknown subject to a stochastic ordering constraint. An efficient MCMC algorithm is developed for posterior computation. The methods are validated using simulation studies and applied to the problem of ranking medical procedures in terms of the distribution of patient morbidity.

Novel Algorithms for Bayesian Hierarchical Clustering

Ester Yen and Chun-Fu Chung
Institute of Information Management, Taiwan
ester_yen@yahoo.com

Web image clustering on WWW become popular in recent years. People want to get information from lots of web image easily, so we need a good way to find what is the user really want. However, how to determine the number of clusters is a problem. It is almost impossible to predict the number of clusters. We then employ unsupervised learning to solve this problem, and hierarchical clustering is the most common method. Bayesian Hierarchical Clustering (BHC) uses probability model to select clusters to merge, so the clustering benchmark are probability (i.e. 0 1) no matter what kind of dataset. It overcomes the limitations of distance-based hierarchically clustering: All ranges are the same and we may see how good those two clusters should be merged. BHC is good; however, it

is time inefficient to decide which two clusters should be merged. Because of its high complexity, we may say that BHC is inappropriate for the large dataset clustering. We proposed two ways to improve BHC's efficiency. In this paper we describe how BHC and then introduce two novel algorithms to improve BHC's efficiency. We also use image dataset to test their operational efficiency and clustering result. We have some discussions in the end.

We shows these two algorithms we proposed did speed up the operation of BHC by reduce repetition operators from clusters effectively and did reach a reasonable time for large data clustering by BHC. Our result shows the operation time between BHC and our two novel algorithms are 13.5:1 and 25:1.

Decision Theoretic Bayesian Nonparametric Inference for the Molecular Characterisation and Stratification of Colorectal Cancer using Genome-Wide Microarrays

Christopher Yau and Christopher C. Holmes
University of Oxford, UK
yau@stats.ox.ac.uk cholmes@stats.ox.ac.uk

With advances in genomic technologies, the molecular profiling of cancer genomes is increasingly seen as an important weapon in the fight against cancer. In particular, clinicians foresee future clinics in which genomic technologies help provide targeted therapeutic regimes tailored to the specific genetic makeup of patients (personalised medicine).

Through on-going collaborations with medical geneticists we have developed novel Bayesian statistical methodologies tailored to tackle key obstacles that stand in the way of molecular profiling of cancer genomes using genome-wide genotyping arrays. In particular, we present a Bayesian nonparametric approach to the problem of inferring structural variation whereby stretches of DNA are duplicated or deleted within heterogeneous cancer samples.

The models are designed to capture the full complexity of the structural variations observed in cancer as well as genetic heterogeneity within the tumour. The model is based on a Bayesian Hidden Markov model (HMM) where the state conditional sampling density (or likelihood) follows a mixture model; and due to the complexity of the noise process we include a nonparametric mixture of Dirichlet process (MDP) model.

We also incorporate formal methods for decision making into our models under a Bayesian decision theoretic approach using a class of loss functions which incorporate errors on the breakpoints (the locations of DNA deletion-duplication events) which amounts to a loss on the state transitions of the (hidden) state sequence. We show that reporting under this particular loss (or utility) function better reflects the reality of how the models are used in practice than conventional methods based on the most probable (Viterbi) state sequence or set of marginally most probable states.

Offline and Online Expectation Maximization via Forward Smoothing in Changepoint Systems

Sinan Yildirim Sumeetpal S. Singh
University of Cambridge, UK
sy276@cam.ac.uk sss40@cam.ac.uk

Arnaud Doucet
University of British Columbia, Canada
arnaud@cs.ubc.ca

In this work, we develop both offline and online expectation-maximization (EM) algorithms for estimating the parameters in a changepoint system, as described in [Fearnhead, P. and Vasileiou, D. (2009)]. A changepoint system is a hidden semi-Markov system where governing model numbers between successive changepoint times constitute a semi-Markov chain. A batch Monte-Carlo EM algorithm for a specific changepoint system was developed in [Caron, F., Gottardo, R. and Doucet, A. (2007)], where the E-step is approximated by Monte-Carlo sampling using filtering densities. We show that; once we have updated our filtering density for the current time, using the idea of forward-only smoothing in [Caron, F., Gottardo, R. and Doucet, A. (2007)] E-step can be performed recursively in time, without storing previous filtering densities. This results in a significant memory saving. Moreover, it enables an online version of the EM algorithm, which can be useful for long data sets. Our online EM algorithm can be considered as an alternative to recursive maximum likelihood methods for parameter estimation in changepoint systems [Azimi, M., Nasipoulos, P. and Ward, R.K. (2005)] and more generally in hidden semi-Markov models [Azimi, M., Nasipoulos, P. and Ward, R.K. (2005)]. We show the performances of our EM algorithms for two different changepoint systems.

Efficient Gibbs Sampler Approaches for Reconstructing Multidimensional NMR Spectra

Ji Won Yoon and Simon P. Wilson
Trinity College Dublin, Ireland
yoonj@tcd.ie swilson@tcd.ie

Multidimensional Nuclear Magnetic Resonance (NMR) is useful to determine the structure of bio-molecules but a serious limitation for the promising multidimensional NMR spectroscopy is to take long time required to record a set of multidimensional NMR experiment. The minimal measurement time of an N-dimensional NMR experiment increases as the number of dimension increases. This makes higher multidimensional NMR spectroscopy intractable in practice. For these reasons, several investigators have been trying to speed up these measurements by more efficient approaches. Many approaches based on the demands are traced back to the concept of accordion spectroscopy. One of the recent approaches to address this problem is Projection Reconstruction NMR (PR-NMR), which

is developed by R. Freeman. PR-NMR is a new technique to generate multi-dimensional NMR spectra, which have discrete features that are relatively sparsely distributed in space. A small number of projections from lower dimensional NMR spectra are used to reconstruct the multi-dimensional NMR spectra. We propose several efficient algorithms which employ Gibbs sampler to reconstruct close NMR spectra to ground truth. The statistical method generates samples in Bayesian scheme. Our proposed algorithms are tested on a set of six projections derived from the three dimensional 700MHz HNCO spectrum of a protein HasA.

A Bayesian Analysis for Biomarker-Adaptive Threshold Design

Kenichi Yoshimura
Kyoto University Hospital, Japan
keyoshim@kuhp.kyoto-u.ac.jp

Difficulty in prediction of efficacy and toxicity are hallmarks of most anticancer therapies. Predictive markers are defined as factors that are associated with response or resistance to a particular therapy. However, reliable predictive markers are often unavailable in designing phase. Jiang *et al.* (2007) proposed biomarker-adaptive threshold design which combines a test for overall treatment effect in all randomly assigned patients with the development and internal validation of a cut point for a pre-specified biomarker of the sensitive subpopulation as classifier. In this study, we incorporate Bayesian analysis into their design. Simulation results showed that efficiency was improved by the incorporation of prior information into the design, and posterior probabilities are straightforward to interpret the results. Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. J Natl Cancer Inst 2007;99:1036ñ43.

Asymptotic Model Selection and Identifiability of Directed Tree models with Hidden Variables

Piotr Zwiernik
Warwick University, UK
p.w.zwiernik@warwick.ac.uk

The standard Bayesian Information Criterion (BIC) is derived under some regularity conditions which are not always satisfied by the graphical models with hidden variables. In this paper we derive the BIC score for Bayesian networks in the case when the data is binary and the underlying graph is a rooted tree and all the inner nodes represent hidden variables. These models are for example widely used in modeling evolution in the phylogenetic data.

The simplest example is the naive Bayes model with n observable and one hidden variable. In this case the correct asymptotic approximation for the marginal likelihood was given by (Rusakov, Geiger 2005). Our work provides a direct generalization of this result for more general tree structures.

We use our geometric insight to propose a convenient parameterization of the models which allows for a better understanding of the asymptotic and identifiability issues. The main analytical tool used in this paper is the connection between asymptotic approximation of Laplace integrals and the real log-canonical threshold. This link has been recently brought to statistical community by (Watanabe 2009).

Participants

(by alphabetic order of of family name)

Abellán, **Juan-José**
CSISP València (Spain) <abellan_jua@gva.es>

Abrams, **Keith**
University of Leicester (UK) <keith.abrams@le.ac.UK>

Adams, **Ryan**
University of Toronto (Canada) <rpadams@alum.mit.edu>

Adhikari, **Kaustubh**
Harvard University (USA) <kadhikar.hsph@gmail.com>

Akerkar, **Rupali**
Norwegian University of Science and Technology (Norway) <akerkar@math.ntnu.no>

Aktekin, **Tevfik**
The George Washington University (USA) <taktekin@gwu.edu>

Alkema, **Leontine**
National University of Singapore (Singapore) <alkema@nus.edu.sg>

Almeida, **Carlos**
Technische Universität München (Germany) <calmeida@ma.tum.de>

Alston, **Claire**
Queensland University of Technology (Australia) <a.ahlberg@qut.edu.au>

Altuzarra, **Alfredo**
Universidad de Zaragoza (Spain) <altuzarr@unizar.es >

Amaral-Turkman, **Maria-Antonia**
Universidade of Lisboa (Portugal) <maturkman@fc.ul.pt>

Amorós Salvador, **Rubén**
Universitat de València (Spain) <ruasal@alumni.uv.es>

Anacleto Jr, **Osvaldo**
The Open University (UK) <o.anacleto-junior@open.ac.uk>

Andrade, **António**
IST (Portugal) <andrade@civil.ist.utl.pt>

Andrianakis, **Yiannis**
National Oceangraphy Centre (UK) <I.Andrianakis@soton.ac.uk>

Antoniano-Villalobos, **Isadora**
University of Kent (UK) <ia57@kent.ac.uk>

Antunes, **Marilia**
Universidade of Lisboa (Portugal) <marilia.antunes@fc.ul.pt>

Arbel, **Julyan**
Université Paris Dauphine (France) <julyan.arbel@ensae.fr>

ArellanoVvalle, **Reinaldo**
Pontificia Universidad Católica de Chile (Chile) <reivalle@mat.puc.cl>

Argiento, **Raffaele**
CNR-IMATI (Italy) <raffaele@mi.imati.cnr.it >

Armero, **Carmen**
Universitat de València (Spain) <carmen.armero@uv.es>

Arnold, **Richard**
Victoria University of Wellington (New Zealand) <richard.arnold@msor.vuw.ac.nz>

Aslett, **Louis**
Trinity College Dublin (Ireland) <aslettl@tcd.ie>

Asseburg, **Christian**
University of Eastern Finland (Finland) <Christian.Asseburg@uef.fi>

Azadi, **Namman**
Lancaster University (UK) <n.azadi@lancaster.ac.uk>

Azevedo, **Caio**
Univ. Estadual de Campinas (Brazil) <cnaber@ime.unicamp.br >

Bacallado, **Sergio**
Stanford University (USA) <sergiobacallado@gmail.com>

Banks, **David**
Duke University (USA) <banks@stat.duke.edu>

Barber, **Xavier**
Universidad Miguel Hernández, Elche (Spain) <xbarber@umh.es>

Barney, **Bradley**
Texas A&M University (USA) <bjbstat@stat.tamu.edu>

Barrantes-Reynolds, **Ramiro**
University of Vermont (USA) <ramiro.barrantes@uvm.edu>

Barry, **Paul**
Trinity College Dublin (Ireland) <barrypb@scss.tcd.ie>

Bayarri, **Susie**
Universitat de València (Spain) <susie.bayarri@uv.es>

Beamonte, **M Asuncion**
Universidad de Zaragoza (Spain) <asunbea@unizar.es>

Becker, **Jeremie**
Oxford University (UK) <becker@stats.ox.ac.uk>

Beffy, **Magali**
CREST INSEE (France) <magali.poncon@polytechnique.org>

Berger, **James**
Duke University (USA) <berger@stat.duke.edu>

Bermúdez, **José D.**
Universitat de València (Spain) <bermudez@uv.es>

Bernardo, **Jose-Miguel**
Universidad de València (Spain) <jose.m.bernardo@uv.es>

Berrett, **Candace**
The Ohio State University (USA) <cberrett@stat.osu.edu>

Berry, **Robert**
Sandia National Laboratories, California (USA) <rdberry@sandia.gov>

Bhattacharya, **Anirban**
Duke University, USA (USA) <ab179@stat.duke.edu>

Bhattacharya, **Arnab**
Trinity College Dublin (Ireland) <bhattaca@tcd.ie>

Bickis, **Mikelis**
University of Saskatchewan (Canada) <bickis@snoopy.usask.ca>

Bolstad, **William**
University of Waikato (New Zealand) <bolstad@waikato.ac.nz>

Bornkamp, **Bjoern**
Novartis (Switzerland) <bornkamp@statistik.tu-dortmund.de>

Bornn, **Luke**
University of British Columbia (Canada) <l.bornn@stat.ubc.ca>

Bortz, **David**
University of Colorado (USA) <dmbortz@colorado.edu>

Botella Rocamora, **Paloma**
Universidad CEU-Cardenal Herrera (Spain) <paloma.botella@gmail.com>

Bottolo, **Leonardo**
Imperial College London (UK) <lbottolo@imperial.ac.uk>

Bourguignon, **Pierre-Yves**
Max-Planck Gesellschaft (Germany) <bourguig@mis.mpg.de>

Bouriga, **Mathilde**
Université Paris-Dauphine (France) <mathilde.bouriga@hotmail.fr >

Bousquet, **Nicolas**
EDF R&D (France) <nicolas.bousquet@edf.fr>

Boys, **Richard**
Newcastle University (UK) <richard.boys@ncl.ac.uk>

Brown, **Philip**
University of Kent (UK) <pjb8@kent.ac.uk>

Brynjarsdottir, **Jenny**
The Ohio State University (USA) <jennyb@stat.osu.edu>

Byrne, **Simon**
University of Cambridge (UK) <s.byrne@statslab.cam.ac.uk>

Cabras, **Stefano**
Università di Cagliari (Italy) <s.cabras@unica.it>

Caglar, **Toros**
The George Washington University (USA) <toros@gwu.edu>

Calder, **Catherine**
The Ohio State University (USA) <calder@stat.osu.edu>

Calderhead, **Ben**
University of Glasgow (Scotland) <bc@dcs.gla.ac.uk>

Cano, **Javier**
Universidad Rey Juan Carlos (Spain) <javier.cano@urjc.es>

Cano, **Juan Antonio**
Universidad de Murcia (Spain) <jacano@um.es >

Carlin, **Brad**
University of Minnesota (USA) <brad@biostat.umn.edu>

Caron, **Francois**
INRIA (France) <francois.caron@inria.fr>

Carta, **Alessandro**
University of Warwick (UK) <a.carta@warwick.ac.uk>

Carvalho, **Carlos M.**
The University of Chicago (USA) <carlos.carvalho@chicagobooth.edu>

Cassese, **Alberto**
Università di Firenze (Italia) <cassese@ds.unifi.it>

Castellanos, **Maria-Eugenia**
Universidad Rey Juan Carlos (Spain) <maria.castellanos@urjc.es>

Clarke, **Bertrand**
University of Miami (USA) <bclarke2@med.miami.edu>

Clyde, **Merlise**
Duke University (USA) <clyde@stat.duke.edu >

Commeau, **Natalie**
AgroParisTech (France) <natalie.commeau@agroparistech.fr>

Conesa, **David**
Universitat de València (Spain) <david.v.conesa@uv.es>

Consonni, **Guido**
Universidtà di Pavia (Italy) <guido.consonni@unipv.it>

Corander, **Jukka**
Helsingin Yliopisto (Finland) <jukka.corander@helsinki.fi>

Cornebise, **Julien**
University of British Columbia (Canada) <cornebis@cs.ubc.ca>

Corradi, **Fabio**
Università di Firenze (Italy) <corradi@ds.unifi.it>

Correia, **Elisete**
Univ. Trás-os-Montes e Alto Douro (Portugal) <ecorreia@utad.pt>

Cosma, **Ioana**
University of Cambridge (UK) <ioana@statslab.cam.ac.uk>

Costain, **Deborah**
Lancaster University (Lancaster) <d.costain@lancaster.ac.uk>

Cui, **Tiangang**
The University of Auckland (New Zealand) <tcui001@aucklanduni.ac.nz>

Cule, **Madeleine**
University of Oxford (UK) <cule@stats.ox.ac.uk>

Cumming, **Jonathan**
Durham University (UK) <j.a.cumming@durham.ac.uk>

Czado, **Claudia**
Technische Universität München (Germany) <cczado@ma.tum.de>

Chan, **Jennifer S.K.**
The University of Sydney (Australia) <jchan@maths.usyd.edu.au>

Chang, **Rui**
University of California, San Diego (USA) <chang.rui@hotmail.com>

Chen, **Qian**
University of Sydney (Australia) <siva1103@hotmail.com>

Chen, **Zhen**
NICHD/NIH (USA) <chenzhe@mail.nih.gov>

Cheng, **Lu**
Helsingin Yliopisto (Finland) <lu.cheng@helsinki.fi>

Chikobvu, **Delson**
University of Free State (South Africa) <chikobvu@ufs.ac.za>

Chopin, **Nicolas**
ENSAE-CREST (France) <nicolas.chopin@ensae.fr>

Choy, **Boris**
The University of Sydney (Australia) <boris.choy@econ.usyd.edu.au>

Christen, **Andrés**
CIMAT (Mexico) <jac@cimat.mx>

Christensen, **Anette Luther**
Aarhus University Hospital (Denmark) <anluc@rn.dk>

Dahl, **David**
Texas A&M University (USA) <dahl@stat.tamu.edu>

Daiki, **Koizumi**
Waseda Univeristy (Japan) <dkoizumi@m.ieice.org>

Datta, **Gauri Sankar**
University of Georgia (USA) <gaurisdatta@gmail.com>

Dawid, **Philip**
University of Cambridge (UK) <apd@statslab.cam.ac.uk>

De la Horra, **Julian**
Universidad Autonoma de Madrid (Spain) <julian.delahorra@uam.es>

Delatola, **Eleni-Ioanna**
University of Kent (UK) <ed62@kent.ac.uk>

Delesen, **Heinz**
Bayer Schering Pharma AG (Germany) <heinz.delesen@bayerhealthcare.com>

Desmond, **Tony**
University of Guelph (Canada) <tdesmond@uoguelph.ca>

Dimitrakopoulou, **Vasiliki**
University of Kent (UK) <vd29@kent.ac.uk>

Djeundje Biatat, **Viani Aime**
Heriot Watt University (Scotland, UK) <vad5@hw.ac.uk>

Djuric, **Petar**
Stony Brook University (USA) <djuric@ece.sunysb.edu>

Domijan, **Katarina**
NUI Maynooth (Ireland) <katarina.domijan@maths.nuim.ie>

Donald, **Margaret**
Queensland University of Technology (Australia) <Margaret.Donald@student.qut.edu.au>

Donnet, **Sophie**
Université Paris Dauphine (France) <donnet@ceremade.dauphine.fr>

Draper, **David**
University of California, Santa Cruz (USA) <draper@ams.ucsc.edu>

Drovandi, **Christopher**
Queensland University of Technology (Australia) <c.drovandi@qut.edu.au>

Dukic, **Vanja**
University of Chicago (USA) <vanja@uchicago.edu>

DuMouchel, **William**
Phase Forward Lincoln Safety Group (USA) <dumouchel@comcast.net>

Dunlop, **Jonathan**
Schlumberger (UK) <jdunlop1@slb.com>

Dunson, **David**
Duke University (USA) <dunson@stat.duke.edu>

Elfadaly, **Fadlalla**
The Open University (UK) <f.elfadaly@open.ac.uk>

Engelhardt Martin, **Barbara**
University of Chicago (USA) <engelhardt@uchicago.edu>

Enjalric, **Sophie**
INRA (France) <sophie.enjalric@agroparistech.fr>

Everitt, **Richard**
University of Bristol (UK) <richard.g.everitt@gmail.com>

Fall, **Mame Diarra**
University Paris Sud 11 (France) <fall@lss.supelec.fr >

Farah, **Marian**
University of California, Santa Cruz (USA) <marian@ams.ucsc.edu>

Faria, **Alvaro**
The Open University (UK) <a.e.faria@open.ac.uk>

Farrow, **Malcolm**
Newcastle University (UK) <malcolm.farrow@newcastle.ac.uk>

Fawcett, **Lee**
Newcastle University (UK) <lee.fawcett@ncl.ac.UK>

Feng, **Li**
Stockholms Universitet (Sweden) <feng.li@stat.su.se>

Fermandez, **Vanesa Maribel**
Universidad de Buenos Aires (Argentina) <vanesamaribel77@gmail.com>

Ferreira, **Marco**
University of Missouri (USA) <ferreiram@missouri.edu>

Fienberg, **Stephen**
Carnegie Mellon University (USA) <fienberg@stat.cmu.edu>

Figueira Neves, **Manuela**
Instituto Politécnico da Guarda (Portugal) <mfi gueira@ipg.pt >

Filippone, **Maurizio**
University of Glasgow (UK) <maurizio@dcs.gla.ac.uk>

Fischer, **Martina**
German Cancer Research Centre (Germany) <martina.fischer@dkfz.de>

Ford, **Ashley**
Warwick University (UK) <a.p.ford@warwick.ac.uk>

Ford, **Eric**
University of Florida (USA) <ericbford@gmail.com>

Forte Deltell, **Anabel**
Universitat de València (Spain) <anabel.forte@uv.es>

Fouskakis, **Dimitris**
National Technical University of Athens (Greece) <fouskakis@math.ntua.gr>

Fox, **Colin**
University of Otago (New Zealand) <fox@physic.otago.ac.nz>

Fox, **Emily**
Duke University (USA) <fox@stat.duke.edu>

French, **Simon**
Manchester Business School (UK) <simon.french@mbs.ac.uk>

Fried, **Roland**
Technische Universität Dortmund (Germany) <fried@statistik.tu-dortmund.de>

Fronczyk, **Kassie**
University of California Santa Cruz (USA) <kassief@ams.ucsc.edu>

Frühwirth-Schnatter, **Sylvia**
Johannes Kepler Universität Linz (Austria) <sylvia.fruehwirth-schnatter@jku.at>

Fu, **Audrey**
University of Cambridge (UK) <audreyqyfu@gmail.com>

Gamerman, **Dani**
Universidade Federal do Rio de Janeiro (Brazil) <dani@im.ufrj.br >

Garcia-Donato, **Gonzalo**
Universidad de Castilla-La Mancha (Spain) <gonzalo.garciadonato@uclm.es>

Gargallo, **Pilar**
Universidad de Zaragoza (Spain) <pigarza@unizar.es>

Garthwaite, **Paul**
Open University (UK) <p.h.garthwaite@open.ac.uk>

Gelfand, **Alan**
Duke University (USA) <alan@stat.duke.edu>

Geneletti, **Sara**
London School of Economics and Political Science (UK) <s.geneletti@lse.ac.uk>

George, **Ed**
University of Pennsylvania (USA) <edgeorge@wharton.upenn.edu>

Germain, **Sarah**
Newcastle University (UK) <s.e.germain@ncl.ac.uk>

Ghahramani, **Zoubin**
University of Cambridge (UK) <zoubin@eng.cam.ac.uk>

Ghosh, **Joyee**
UNC Chapel Hill (USA) <joyee@email.unc.edu>

Ghosh, **Sujit**
North Carolina State University (USA) <ghosh@stat.ncsu.edu>

Giannoulatou, **Eleni**
University of Oxford (UK) <giannoul@stats.ox.ac.uk>

Giardina, **Federica**
Swiss Tropical and Public Health Institute (Switzerland) <federica.giardina@unibas.ch>

Gillespie, **Colin**
Newcastle University (UK) <c.gillespie@ncl.ac.uk>

Girolami, **Mark**
University of Glasgow (UK) <girolami@dcs.gla.ac.uk>

Girón, **Francisco Javier**
Universidad de Málaga (Spain) <fj_giron@uma.es>

Goldstein, **Michael**
Durham University (UK) <michael.goldstein@durham.ac.uk>

Golightly, **Andrew**
Newcastle University (UK) <a.golightly@ncl.ac.uk>

Gómez-Rubio, **Virgilio**
Universidad de Castilla-La Mancha (Spain) <virgilio.gomez@uclm.es>

Gómez-Villegas, **Miguel A.**
Universidad Complutense de Madrid (Spain) <ma.gv@mat.ucm.es>

Goncalves, **Flavio**
University of Warwick (UK) <F.B.Goncalves@warwick.ac.uk>

González Pérez, **Beatriz**
Universidad Complutense de Madrid (Spain) <beatrizg@mat.ucm.es>

Gordillo Suárez, **Marisol**
Universidad del Valle, Cali (Colombia) <mgsuarez44@yahoo.com >

Gormley, **Claire**
University College Dublin (Ireland) <claire.gormley@ucd.ie>

Gorst-Rasmussen, **Anders**
Aalborg University (Denmark) <gorst@math.aau.dk>

Gosoniu, **Dominic**
Swiss TPH (Switzerland) <dominic.gosoniu@unibas.ch>

Gosoniu, **Laura**
Swiss TPH (Switzerland) <laura.gosoniu@unibas.ch>

Gramacy, **Robert**
University of Cambridge (UK) <rbgramacy@gmail.com>

Grapsa, **Erofili**
University of Southampton (UK) <eg2x07@soton.ac.uk>

Green, **Peter**
University of Bristol (UK) <P.J.Green@bristol.ac.uk>

Griffin, **Jim**
University of Kent (UK) <J.E.Griffin-28@kent.ac.uk>

Griffiths, **William**
University of Melbourne (Australia) <wegrif@unimelb.edu.au>

Gsteiger, **Sandro**
Novartis Pharma AG (Switzerland) <sandro.gsteiger@novartis.com>

Guglielmi, **Alessandra**
Politecnico di Milano (Italy) <alessandra.guglielmi@polimi.it>

Guha, **Subharup**
University of Missouri-Columbia (USA) <GuhaSu@missouri.edu>

Gutiérrez - Peña, **Eduardo**
IIMAS, UNAM (Mexico) <eduardo@sigma.iimas.unam.mx>

Haas, **Armin**
Institut Klimafolgenfors. (Germany) <armin.haas@pik-potsdam.de >

Hahn, **Richard**
Duke University (USA) <hahn@stat.duke.edu>

Han, **Chao**
Virginia Tech (USA) <chaohan@vt.edu>

Han, **Minkyu**
Seoul National University (South Korea) <crysnal2@snu.ac.kr>

Hans, **Chris**
Ohio State University (USA) <hans@stat.osu.edu>

Haran, **Murali**
Pennsylvania State University (USA) <mharan@stat.psu.edu>

Harris, **Keith**
University of Glasgow (UK) <keithh@dcs.gla.ac.uk>

Heaton, **Matthew**
Duke University (USA) <matt@stat.duke.edu>

Heaton, **Tim**
University of Sheffield (UK) <t.heaton@shef.ac.uk>

Helland, **Inge S.**
Universitetet i Oslo (Norway) <ingeh@math.uio.no>

Heller, **Katherine**
University of Cambridge (UK) <kheller@gmail.com>

Henao, **Ricardo**
Danmarks Tekniske Universitet (Denmark) <rh@imm.dtu.dk>

Henderson, **Daniel**
Newcastle University (UK) <d.a.henderson@ncl.ac.uk>

Hermansen, **Gudmund**
Universitetet i Oslo (Norway) <gudmund.hermansen@gmail.com>

Heron, **Eleisa**
Trinity College Dublin (Ireland) <eaheron@tcd.ie>

Hill, **Carter**
Louisiana State University (USA) <eohill@lsu.edu>

Hill, **Jennifer**
New York University (USA) <jennifer.hill@nyu.edu>

Hofmann, **Mathias**
Technischen Universität München (Germany) <hofmann@ma.tum.de>

Holand, **Anna Marie**
Norges Teknisk-Natur. Univ. (Norway) <anna.holand@bio.ntnu.no>

Holmes, **Christopher**
University of Oxford (UK) <cholmes@stats.ox.ac.uk>

Holloman, **Christopher**
The Ohio State University (USA) <holloman@stat.osu.edu>

Houlding, **Brett**
Trinity College Dublin (Ireland) <brett.houlding@tcd.ie>

House, **Leanna**
Virginia Tech (USA) <lhouse@vt.edu>

Hu, **Jianhua**
MD Anderson Cancer Center (USA) <jhu@mdanderson.org>

Huan, **Xun**
Massachusetts Institute of Technology (USA) <xunhuan@mit.edu>

Huber, **Mark**
Claremont McKenna College (USA) <mhuber@cmc.edu>

Hurtado Rua, **Sandra M**
University of Connecticut (USA) <shurtado@stat.uconn.edu>

Huszar, **Ferenc**
University of Cambridge (UK) <fh277@cam.ac.uk>

Ickstadt, **Katja**
Technische Universität Dortmund (Germany) <ickstadt@statistik.uni-dortmund.de>

Inoue, **Lurdes**
University of Washington (USA) <linoue@uw.edu>

Jääskinen, **Väinö**
Helsingin Yliopisto (Finland) <vainoj@gmail.com>

Jacob, **Pierre**
ENSAE-CREST (France) <pierre.jacob@ensae.fr>

Jarocinski, **Marek**
European Central Bank (Germany) <marek.jarocinski@gmail.com>

Jensen, **Shane**
University of Pennsylvania (USA) <stjensen@wharton.upenn.edu>

Jo, **Seongil**
Seoul National Univeristy (Korea, Republic of) <joseongil@gmail.com>

Johnson, **Sandra**
Queensland University of Technology (Australia) <j.watt@qut.edu.au>

Johnson, **Valen**
MD Anderson Cancer Center (USA) <vejohnson@mdanderson.org>

Jones, **Emma**
University of Sheffield (UK) <stp08emj@shef.ac.uk>

Jordan, **Michael**
University of California, Berkeley (USA) <jordan@stat.berkeley.edu>

Joseph, **Maria**
Iowa State University (USA) <josepall@iastate.edu>

Joshi, **Chaitanya**
Trinity College Dublin (Ireland) <joshic@scss.tcd.ie>

Juárez, **Miguel**
University of Warwick (UK) <m.a.juarez@warwick.ac.uk>

Juergens, **Verena**
Swiss Tropical and Public Health Institute (Switzerland) <verena.juergens@unibas.ch>

Julien, **Cornebise**
University of British Columbia (Canada) <cornebis@cs.ubc.ca>

Justel, **Ana**
Universidad Autonoma de Madrid (Spain) <ana.justel@uam.es>

Keller, **Merlin**
EDF R&D (France) <merlinkeller@gmail.com>

Khatab, **Khaled**
Aachen University (Germany) <kkhatab@ukaachen.de>

Klauenberg, **Katy**
PTB-Berlin (Germany) <katy.klauenberg@ptb.de>

Kohn, **Robert**
University of New South wales (Australia) <r.kohn@unsw.edu.au>

Kokonen, **Jukka**
Helsingin Yliopisto (Finland) <jukka.kohonen@helsinki.fi>

Kolossiatis, **Michalis**
Cyprus Univ. of Technology (Cyprus) <Michalis.Kolossiatis@cut.ac.cy >

Kottas, **Athanasios**
University of California, Santa Cruz (USA) <thanos@ams.ucsc.edu>

Krnjajic, **Milovan**
National University of Ireland, Galway (Ireland) <milovan.krnjajic@nuigalway.ie>

Kruijer, **Willem**
Université Paris Dauphine (France) <kruijer@ceremade.dauphine.fr>

La Rocca, **Luca**
University of Modena and Reggio Emilia (Italy) <luca.larocca@unimore.it>

Lacoste-Julien, **Simon**
University of Cambridge (UK) <sl522@cam.ac.uk>

Laine, **Marko**
Ilmatieteen Laitos (Finland) <marko.laine@fmi.fi>

Lam, **Pui Yu**
The University of Sydney (Australia) <plam7745@uni.sydney.edu.au>

Lambert, **Philippe**
University of Liege (Belgium) <p.lambert@ulg.ac.be>

Lasscock, **Ben**
Boronia Capital (Australia) <Ben.Lasscock@boroniacapital.com.au>

Latuszynski, **Krzysztof**
University of Warwick (UK) <latuch@gmail.com>

LAUNAY, **Tristan**
EDF R&D (France) <tristan.launay@edf.fr>

Lee, **Anthony**
University of Oxford (UK) <lee@stats.ox.ac.uk>

Lee, **Dominic**
University of Canterbury (New Zealand) <dominic.lee@canterbury.ac.nz>

Lee, **Herbie**
University of California, Santa Cruz (USA) <herbie@ams.ucsc.edu>

Lee, **Jiun-Kae Jack**
MD Anderson Cancer Center (USA) <jjlee@mdanderson.org>

Leininger, **Thomas**
Brigham Young University (USA) <tjleininger@byu.edu>

Leman, **Scotland**
Virginia Tech (USA) <leman@vt.edu>

Lenarcic, **Alan**
University of Harvard (USA) <lenarcic@post.harvard.edu>

Leon Novelo, **Luis Gonzalo**
University of Florida (USA) <leonnovelo@hotmail.com>

Li, **Fan**
Duke University (USA) <fli@stat.duke.edu>

Li, **Guangquan**
Imperial College London (UK) <guang.li@imperial.ac.uk>

Li, **Yingbo**
Duke University (USA) <yl118@duke.edu>

Liang, **Li-Jung**
University of California, Los Angeles (USA) <liangl@ucla.edu>

Lin, **Lin**
Duke University (USA) <ll86@stat.duke.edu>

Liverani, **Silvia**
University of Bristol (UK) <s.liverani@bristol.ac.uk>

Loizides, **Charalambos**
University of Oxford (UK) <loizides@stats.ox.ac.uk>

Lopes, **Danilo**
Duke University (USA) <dll11@stat.duke.edu>

Lopes, **Hedibert**
The University of Chicago (USA) <hlopes@ChicagoBooth.edu>

Lopez-Quilez, **Antonio**
Universitat de València (Spain) <antonio.lopez@uv.es>

Loredo, **Thomas**
Cornell University (USA) <loredo@astro.cornell.edu>

Loschi, **Rosangela Helena**
Universidade Federal de Minas Gerais (Brazil) <loschi@est.ufmg.br>

Louis, **Thomas**
Johns Hopkins Bloomberg SPH (USA) <tlouis@jhsph.edu>

Love, **Tanzy**
Univeristy of Rochester (USA) <tanzy@cmu.edu>

Low Choy, **Samantha**
Queensland University of Technology (Australia) <s.lowchoy@qut.edu.au>

Lum, **Kristian**
Duke University (USA) <kristian.lum@gmail.com>

Lutz, **Sharon**
Harvard University/ University of Bonn (USA) <sharon.m.lutz@gmail.com>

Lykou, **Anastasia**
Lancaster University (UK) <a.lykou@lancaster.ac.uk>

Lytton, **Diane**
Private Practice, Forensic Psychologist (USA) <dlytton@prodigy.net>

Madi, **Mohamed Tahar**
UAE University (United Arab Emirates) <mmadi@uaeu.ac.ae>

Madigan, **David**
Columbia University (USA) <madigan@stat.columbia.edu>

Main, **Paloma**
Universidad Complutense de Madrid (Spain) <pmain@mat.ucm.es>

Maiti, **Dipayan**
Virginia Tech (USA) <dipayan.maiti@gmail.com >

Makov, **Udi**
University of Haifa (Israel) <makov@stat.haifa.ac.il>

Malsiner Walli, **Gertraud**
Johannes Kepler University Linz (Austria) <gertraud.malsiner_walli@jku.at>

Mallick, **Bani**
Texas A&M (USA) <bmallick@stat.tamu.edu>

Manolopoulou, **Ioanna**
Duke University (USA) <im30@stat.duke.edu>

Mansfield, **Keith**
Oxford University Press (UK) <keith.mansfield@oup.com>

Marín, **Juan Miguel**
Universidad Carlos III (Spain) <jmmarin@est-econ.uc3m.es>

Martínez, **María Lina**
Universidad de Málaga (Spain) <mlmartinez@uma.es>

Martinez-Beneito, **Miguel A.**
CSISP València (Spain) <miguebeneito@gmail.com>

Martino, **Luca**
Universidad Carlos III de Madrid (Spain) <luca@tsc.uc3m.es>

Marttinen, **Pekka**
Aalto-Universitetets (Finland) <pekka.marttinen@tkk.fi>

Marzouk, **Youssef**
Massachusetts Institute of Technology (USA) <ymarz@mit.edu>

Mason, **Alexina**
Imperial College London (UK) <a.mason05@imperial.ac.uk>

Matano, **Alessandro**
Novartis Pharma (Swizerland) <alessandro.matano@novartis.com>

Mattei, **Alessandra**
Univeristà di Firenze (Italy) <mattei@ds.unifi.it>

Mayrink, **Vinicius**
Duke University (USA) <vdinizm@gmail.com>

McAlinn, **Kenichiro**
Keio University (Japan) <kenmcalinn@gmail.com>

McCallum, **Andrew**
University of Massachusetts (USA) <mccallum@cs.umass.edu>

McCandless, **Lawrence**
Simon Fraser University (Canada) <mccandless@sfu.ca >

McCormick, **Tyler**
Columbia University (USA) <tyler@stat.columbia.edu>

Mealli, **Fabrizia**
Università di Firenze (Italy) <mealli@ds.unifi.it>

Meek, **Chris**
Microsoft Research (USA) <meek@microsoft.com>

Mena, **Ramsés H**
IIMAS-UNAM (Mexico) <ramses@sigma.iimas.unam.mx>

Mendoza, **Manuel**
ITAM (Mexico) <mendoza@itam.mx>

Meng, **Xiao-Li**
Harvard University (USA) <meng@stat.harvard.edu>

Merl, **Daniel**
Lawrence Livermore National Laboratory (USA) <merl1@llnl.gov>

Meyer, **Renate**
University of Auckland (New Zealand) <meyer@stat.auckland.ac.nz>

Migon, **Helio**
Universidade Federal do Rio de Janeiro (Brazil) <migon@im.ufrj.br>

Míguez, **Joaquín**
Universidad Carlos III de Madrid (Spain) <joaquin.miguez@uc3m.es>

Miller, **Kurt**
University of California at Berkeley (USA) <tadayuki@cs.berkeley.edu>

Min, **Xiaoyi**
University of Missouri (USA) <xm4y4@mail.missouri.edu>

Mira, **Antonietta**
Università dell'Insubria (Italy) <antonietta.mira@uninsubria.it>

Mitra, **Robin**
University of Southampton (UK) <R.Mitra@soton.ac.uk>

Miyawaki, **Koji**
National Institute for Environmental Studies (Japan) <miyawaki.koji@nies.go.jp>

Modenese, **Mirko**
EURAC Bozen (Italy) <mirko.modenese@eurac.edu>

Mohamed, **Shakir**
University of Cambridge (UK) <sm694@cam.ac.uk>

Mohammadzadeh, **Mohsen**
Tarbiat Modares University (Iran) <mohsen_m@modares.ac.ir>

Mongelluzzo, **Silvia**
Bocconi University (Italy) <silvia_monge@virgilio.it>

Moreno, **Elias**
Universidad de Granada (Spain) <emoreno@ugr.es>

Moura, **Fernando Antônio da Silva**
Universadade Federal Rio de Janeiro (Brazil) <fmoura@im.ufrj.br>

Mueller, **Peter**
MD Anderson Cancer Center (USA) <pmueller@mdanderson.org>

Mukherjee, **Chiranjit**
Duke University (USA) <chiranjit@stat.duke.edu>

Mukhopadhyay, **Nitai**
Virginia Commonwealth University (USA) <ndmukhopadhy@vcu.edu>

Mukhopadhyay, **Sabyasachi**
Indian Statistical Institute (India) <sabstat123@gmail.com>

Muñoz, **Facundo**
Universitat de València (Spain) <facundo.munoz@uv.es>

Murdoch, **Duncan**
University of Western Ontario (Canada) <510>

Murphy, **Thomas Brendan**
University College Dublin (Ireland) <brendan.murphy@ucd.ie>

Murray, **Iain**
University of Toronto (Canada) <murray@cs.toronto.edu>

Musio, **Monica**
Università di Cagliari (Italy) <mmusio@unica.it>

Myung, **Jay**
Ohio State University (USA) <myung.1@osu.edu >

Najm, **Habib**
Sandia National Labs (USA) <hnnajm@sandia.gov>

Nakajima, **Jochi**
Duke University (USA) <jouchi.nakajima@stat.duke.edu>

Nakatsuma, **Teruo**
Keio University (Japan) <nakatuma@econ.keio.ac.jp>

Nascimento, **Fernando**
Universidade Federal do Piauí (Brazil) <fernandofn@ufpi.edu.br >

Natário, **Isabel**
Universidade Nova de Lisboa (Portugal) <icn@fct.unl.pt>

Navarro, **Hilario**
UNED (Spain) <hnavarro@ccia.uned.es>

Neuenschwander, **Beat**
Novartis Pharma (Switzerland) <beat.neuenschwander@novartis.com>

Nicolette, **Raquel**
Universidade de Aveiro (Portugal) <nicolette@ua.pt>

Nicholson, **George**
University of Oxford (UK) <nicholso@stats.ox.ac.uk>

Nicholls, **Geoff**
Oxford University (UK) <nicholls@stats.ox.ac.uk>

Nipoti, **Bernardo**
Università di Pavia (Italy) <bernardo.nipoti@unipv.it>

Nixon, **Richard**
Novartis (Switzerland) <richard.nixon@novartis.com>

Omori, **Yasuhiro**
University of Tokyo (Japan) <omori@e.u-tokyo.ac.jp>

Orbanz, **Peter**
University of Cambridge (UK) <p.orbanz@eng.cam.ac.uk>

Ortiz, **Antonio**
University of Kent (UK) <aao33@kent.ac.uk>

Ottavi, **Alessandro**
Norges Teknisk-Natur. Univ (Norway) <ottavi@math.ntnu.no>

Overstall, **Antony**
University of Southampton (UK) <amo105@soton.ac.uk>

Oxlade, **Rachel**
Durham University (UK) <r.h.oxlade@durham.ac.uk>

Paciorek, **Christopher**
Univ. of California, Berkeley (USA) <paciorek@stat.berkeley.edu>

Page, **Garritt**
Duke University (USA) <page@stat.duke.edu>

Palacios, **Ana-Paula**
Universidad Carlos III de Madrid (Spain) <appalaci@est-econ.uc3m.es>

Pamminger, **Christoph**
Johannes Kepler University Linz (Austria) <christoph.pamminger@jku.at>

Panagiotelis, **Anastasios**
Technische Universität München (Germany) <panagiot@ma.tum.de>

Papathomas, **Michail**
Coventry University and Imperial College London (UK) <m.papathomas@imperial.ac.uk>

Parent, **Eric**
AgroParisTech/ENGREF (France) <eric.parent@agroparistech.fr>

Parnell, **Andrew**
University College Dublin (Ireland) <andrew.parnell@ucd.ie>

Pasanisi, **Alberto**
EDF R&D (France) <alberto.pasanisi@edf.fr>

Pati, **Debdeep**
Duke University (USA) <dp55@stat.duke.edu>

Paulino, **Carlos Daniel**
Universidade Técnica de Lisboa (Portugal) <dpaulino@math.ist.utl.pt>

Paulo, **Rui**
Universidade Técnica de Lisboa (Portugal) <rui@iseg.utl.pt>

Peltola, **Tomi**
Aalto University (Finland) <tomi.peltola@tkk.fi>

Pennino, **Maria-Grazia**
Instituto Español de Oceanografía (Spain) <grazia.pennino@mu.ieo.es>

Peña, **Daniel**
Universidad Carlos III (Spain) <dpena@est-econ.uc3m.es>

Pereira, **Isabel**
Universidade de Aveiro (Portugal) <isabel.pereira@ua.pt>

Pérez Hernández, **Maria-Eglée**
Universidad de Puerto Rico Rio Piedras (Puerto Rico) <meglee@uprrp.edu>

Pericchi, **Luis**
Universidad de Puerto Rico Rio Piedras (Puerto Rico) <luarpr@gmail.com>

Perpiñán Fabuel, **Hèctor**
Universitat de València (Spain) <hecperfa@alumni.uv.es>

Petrone, **Sonia**
Bocconi University (Italy) <sonia.petrone@unibocconi.it>

Philippe, **Anne**
Université de Nantes (France) <anne.philippe@univ-nantes.fr>

Pirinen, **Matti**
Wellcome Trust Centre for Human Genetics, Oxford (UK) <matti.pirinen@iki.fi>

Pitt, **Michael**
University of Warwick (UK) <m.pitt@warwick.ac.uk>

Poirier, **Dale**
University of California, Irvine (USA) <dpoirier@uci.edu>

Poleto, **Frederico**
IUniversidade de São Paulo, (Brazil) <fpoleto@ime.usp.br>

Polson, **Nicholas**
University of Chicago, USA (USA) <University of Chicago, USA>

Poppe, **Stephan**
Max-Planck Gesellschaft (Germany) <poppe@mis.mpg.de>

Posekany, **Alexandra**
Univ. Bodenkultur Wien (Austria) <alexandra.posekany@gmail.com >

Prangle, **Dennis**
Lancaster University (UK) <d.prangle@lancaster.ac.uk>

Pruenster, **Igor**
Università di Torino (Italy) <igor@econ.unito.it>

Puig, **Xavier**
Universitat Politècnica de Catalunya (Spain) <xavier.puig@upc.edu>

Qin, **Zhaohui**
University of Michigan (USA) <qin@umich.edu>

Queen, **Catriona**
The Open University (UK) <C.Queen@open.ac.uk>

Quinn, **Anthony**
Trinity College Dublin (Ireland) <aquinn@tcd.ie >

Quirós, **Alicia**
Universidad Rey Juan Carlos (Spain) <alicia.quiros@urjc.es>

Racugno, **Walter**
Università di Cagliari (Italy) <racugno@unica.it>

Rahaman Khan, **Hasinur**
University of Warwick (UK) <m.h.rahaman-khan@warwick.ac.uk>

Ramos, **Sandra**
Poliécnico do Porto (Portugal) <sfr@isep.ipp.pt>

Rancoita, **Paola**
IDSIA (Switzerland) <paola@idsia.ch>

Rantalainen, **Mattias**
University of Oxford (UK) <rantalai@stats.ox.ac.uk>

Ratmann, **Oliver**
Duke University (USA) <oliver.ratmann@duke.edu>

Ray, **Jaideep**
Sandia National Labs (USA) <jairay@sandia.gov>

Reese, **Shane**
Brigham Young University (USA) <reese@stat.byu.edu>

Renfro, **Lindsay**
Baylor University (USA) <lindsay_renfro@baylor.edu>

Ricciardi, **Federico**
Università di Firenze (Italia) <ricciardi@ds.unifi.it>

Richardson, **Thomas**
University of Washington (USA) <thomasr@uw.edu>

Riebler, **Andrea**
Universität Zürich (Switzerland) <andrea.riebler@ifspm.uzh.ch>

Riedel, **Nadine**
Swiss Tropical and Public Health Institute (Switzerland) <n.riedel@unibas.ch>

Ringer, **Maurice**
Schlumberger (UK) <mringer@slb.com>

Rios, **Jesus**
Manchester Business School (UK) <jesus.rios@mbs.ac.uk>

Robert, **Christian**
Université Paris-Dauphine (France) <xian@ceremade.dauphine,fr>

Roberts, **Gareth**
University of Warwick (UK) <Gareth.O.Roberts@warwick.ac.uk>

Rodrigues, **Josemar**
Federal University of São Carlos (Brazil) <vjosemar@ufscar.br >

Rodriguez, **Abel**
University of California, Santa Cruz (USA) <abel@soe.ucsc.edu>

Rodríguez, **Marco A.**
Université du Québec (Canada) <marco.rodriguez@uqtr.ca>

Rodríguez Bernal, **María Teresa**
Universidad Complutense de Madrid (Spain) <mayter@mat.ucm.es>

Román, **Lizbeth**
Universitat de València (Spain) <lizbeth.roman@uv.es>

Roodaki, **Alireza**
SUPELEC (France) <alireza.roodaki@supelec.fr >

Rosenthal, **Jeffrey**
University of Toronto (Canada) <jeff@math.toronto.edu>

Roskelley, **Kenneth**
Mississippi State University (USA) <KRoskelley@cobilan.msstate.edu>

Rossell, **David**
IRB Barcelona (Spain) <david.rossell@irbbarcelona.org>

Rotondi, **Renata**
CNR-IMATI (Italy) <reni@mi.imati.cnr.it>

Rousseau, **Judith**
Université Paris Dauphine (France) <rousseau@ceremade.dauphine.fr>

Rueda, **Raul**
IIMAS, UNAM (Mexico) <pinky@sigma.iimas.unam.mx>

Rufo, **Maria-Jesús**
Universidad de Extremadura (Spain) <mrufo@unex.es>

Ruggeri, **Fabrizio**
CNR IMATI (Italy) <fabrizio@mi.imati.cnr.it>

Rumisha, **Susan Fred**
Swiss Tropical and Public Health Institute (Switzerland) <susan.rumisha@unibas.ch>

Sabanés, **Daniel**
Universität Zürich (Switzerland) <daniel.sabanesbove@ifspm.uzh.ch>

Safta, **Cosmin**
Sandia National Labs (USA) <csafta@gandia.gov>

Sahu, **Sujit**
University of Southampton (UK) <sujitsid@yahoo.co.uk>

Salazar Mendoza, **Isabel**
Universidad Complutense de Madrid (Spain) <isalazar@vet.ucm.es>

Salimans, **Tim**
Erasmus Universiteit Rotterdam (The Netherlands) <salimans@ese.eur.nl>

Salvador, **Manuel**
Universidad de Zaragoza (Spain) <salvador@unizar.es>

Sargsyan, **Khachik**
Sandia National Labs (USA) <ksargsy@sandia.gov>

Sauleau, **Erik-A.**
University of Strasbourg (France) <ea.sauleau@unistra.fr>

Savelieva, **Marina**
Novartis (Switzerland) <marina.savelieva@novartis.com>

Saville, **Ben**
Vanderbilt University School of Medicine (USA) <b.saville@vanderbilt.edu>

Scott, **James**
University of Texas at Austin (USA) <james.scott@mccombs.utexas.edu>

Scott, **Steven**
Google (USA) <stevescott@google.com>

Schafer, **Christian**
CREST (France) <christian.schafer@ensae.fr>

Schmidt, **Alexandra**
Universidade Federal do Rio de Janeiro (Brazil) <alex@im.ufrj.br>

Schmidt, **Kristian**
Iowa State University (USA) <kristian@iastate.edu>

Schmitz, **Susanne**
Trinity College Dublin (Ireland) <schmitzs@tcd.ie>

Schoergendorfer, **Angela**
University of Kentucky (USA) <angela.sch@uky.edu >

Schott, **Sarah**
Duke University (USA) <schott@math.duke.edu>

Schroedle, **Birgit**
Universität Zürich (Switzerland) <birgit.schroedle@ifspm.uzh.ch>

Segurado, **Ricardo**
Trinity College Dublin (Ireland) <rsegurdo@tcd.ie>

Seheult, **Allan**
Durham University (UK) <a.h.seheult@durham.ac.uk>

Sepulveda, **Nuno**
Instituto Gulbenkian de Ciência (Portugal) <nunosep@gmail.com>

Sermaidis, **Giorgos**
Lancaster University (UK) <linuxfever@yahoo.gr>

Shakir, **Mohamed**
University of Cambridge (UK) <sm694@cam.ac.uk>

Shemyakin, **Arkady**
University of St. Thomas (USA) <a9shemyakin@stthomas.edu>

Sherlock, **Chris**
Lancaster University (UK) <c.sherlock@lancaster.ac.uk>

Shi, **Minghui**
Duke University (USA) <ms193@stat.duke.edu>

Silva, **Giovani**
Universidade Técnica de Lisboa (Portugal) <gsilva@math.ist.utl.pt>

Silva, **Nélia**
Universidade de Aveiro (Portugal) <neliasilva@ua.pt>

Silva, **Ricardo**
University College London (UK) <ricardo@stats.ucl.ac.uk>

Silva-Fortes, **Carina**
Instituto Politécnico de Lisboa (Portugal) <carina.silva@estesl.ipl.pt>

Simpson, **Shawn**
Columbia University (USA) <shawn@stat.columbia.edu>

Sirén, **Jukka**
Helsingin Yliopisto (Finland) <jukka.p.siren@helsinki.fi>

Sisson, **Scott**
University of New South Wales (Australia) <scott.sisson@unsw.eu.au>

Skilling, **John**
Maximum Entropy Data Consultants (Ireland) <skilling@eircom.net>

Smith, **Adrian FM**
DG Science and Research (UK) <afmsmith@gmail.com>

Smith, **James**
University of Warwick (UK) <J.Q.Smith@warwick.ac.uk>

Smith, **Murray**
Natl. Inst. Water and Atmospheric Res. (New Zealand) <mh.smith@niwa.co.nz>

Soares, **Paulo**
Universidade Técnica de Lisboa (Portugal) <psoares@math.ist.utl.pt>

Solonen, **Antti**
Lappeenranta University of Technology (Finland) <antti.solonen@gmail.com>

Solgi, **Reza**
Swiss Finance Institute at the Università della Svizzera Italiana (Switzerland) <reza.solgi@usi.ch>

Sousa, **Lisete**
Universidade de Lisboa (Portugal) <lmsousa@fc.ul.pt>

Soyer, **Refik**
George Washington University (USA) <soyer@gwu.edu>

Spiller, **Elaine**
Marquette University (USA) <elaine.spiller@marquette.edu>

Spitzner, **Dan**
University of Virginia (USA) <dan.spitzner@virginia.edu>

Spoeck, **Gunter**
University of Klagenfurt (Austria) <gunter.spoeck@uni-klu.ac.at>

Sprenger, **Jan**
Tilburg Center for Logic and Philosophy of Science (Netherlands) <j.sprenger@uvt.nl>

Stathopoulos, **Vassilios**
University of Glasgow (UK) <stathv@dcs.gla.ac.uk>

Steel, **Mark**
University of Warwick (UK) <M.F.Steel@stats.warwick.ac.uk>

Strickert, **Donald**
PepsiCo/FritoLay (USA) <don.strickert@fritolay.com>

Strimenopoulou, **Foteini**
University of Kent (UK) <fs54@kent.ac.uk>

Sun, **Dongchu**
University of Missouri (USA) <sund@missouri.edu>

Susi, **Rosario**
Universidad Complutense de Madrid (Spain) <rsusi@estad.ucm.es>

Sutton, **Charles**
University of Edinburgh (UK) <csutton@inf.ed.ac.uk>

Svensson, **Lennart**
Chalmers University of Technology (Sweden) <lennart.svensson@chalmers.se>

Sweeney, **James**
Trinity College Dublin (Ireland) <sweeneja@tcd.ie>

Tabet, **Aline**
University of British Columbia (Canada) <a.tabet@stat.ubc.ca>

Taddy, **Matthew**
Chicago Booth (USA) <taddy@chicagobooth.edu>

Takada, **Ryota**
Keio University (Japan) <ryota1987@gmail.com>

Tamminen, **Johanna**
Finnish Meteorological Institute (Finland) <Johanna.Tamminen@fmi.fi>

Tang, **Jing**
VTT Technical Research Center (Finland) <jing.tang@vtt.fi>

Taylor, **Benjamin**
Lancaster University (UK) <b.taylor1@lancaster.ac.uk>

Tebaldi, **Claudia**
UBC/Climate Central (US) <claudia.tebaldi@gmail.com>

Teh, **Yee Whye**
University College London (UK) <ywteh@gatsby.ucl.ac.uk>

Teng, **Gloria**
University of Canterbury (New Zealand) <gat41@uclive.ac.nz>

ter Horst, **Enrique**
Euromed Management (France) <terhorstgomez@hotmail.com>

Thomas, **Andrew**
Carnegie Mellon University (USA) <acthomas@stat.cmu.edu >

Tokdar, **Surya T.**
Duke University (USA) <tokdar@stat.duke.edu>

Tomazella, **Vera L.**
Universidade Federal de São Carlos (Brazil) <vera@ufscar.br>

Torrado, **Nuria**
Universidad Carlos III de Madrid (Spain) <nuria.torrado@uc3m.es >

Torres, **Francisco**
Universidad de Santiago de Chile (Chile) <francisco.torres@usach.cl>

Turkman, **Kamil Feridun**
Universidade de Lisboa (Portugal) <kfturkman@fc.ul.pt>

Tuyl, **Frank**
The University of Newcastle (Australia) <frank.tuyl@newcastle.edu.au>

Unger, **Sigrun**
Bayer Schering Pharma AG (Germany) <sigrun.unger@bayerhealthcare.com>

Valenzuela, **Carmen**
Centro Ing. Genética y Biotecnología (Cuba) <carmen.valenzuela@cigb.edu.cut>

van der Linde, **Angelika**
University of Bremen (Germany) <avdl@math.uni-bremen.de>

van Dyk, **David A.**
University of California, Irvine (USA) <dvd@ics.uci.edu>

Vannucci, **Marina**
Rice University (USA) <marina@rice.edu>

Varini, **Elisa**
CNR-IMATI (Italy) <elisa@mi.imati.cnr.it>

Vatsa, **Richa**
Trinity College Dublin (Ireland) <vatsar@tcd.ie>

Ventura, **Laura**
Università di Padova (Italy) <ventura@stat.unipd.it>

Venturini, **Sergio**
Universitá Bocconi (Italy) <sergio.venturini@unibocconi.it >

Vernon, **Ian**
Durham University (UK) <i.r.vernon@durham.ac.uk>

Veronese, **Piero**
Università Bocconi (Italy) <piero.veronese@unibocconi.it>

Vidal, **Ignacio**
Universidad de Talca (Chile) <ividal@utalca.cl>

Villani, **Mattias**
Sveriges Riksbank and Stockholms Univ. (Sweden) <mattias.villani@gmail.com>

Vistisen, **Jan**
H. Lundbeck (Denmark) <jbv@lundbeck.com>

Vlachos, **Pantelis**
Merck Serono (Switzerland) <pantelis.vlachos@merckserono.net>

Vlasakakis, **Georgios**
University of Cambridge (UK) <georgios.vlasakakis@mrc-hnr.cam.ac.uk>

Vounatsou, **Penelope**
Swiss TPH (Switzerland) <penelope.vounatsou@unibas.ch>

Vrousai, **Rountina**
Trinity College Dublin (Ireland) <vrousair@tcd.ie>

Vukcevic, **Damjan**
University of Oxford (UK) <vukcevic@well.ox.ac.uk>

Vuppala, **Radhakrishna**
Univ. California Santa Cruz (USA) <v_vrky@yahoo.com>

Wade, **Sara**
Universita Bocconi (Italy) <sara.wade@phd.unibocconi.it>

Wagenmakers, **Eric-Jan**
Universiteit van Amsterdam (The Netherlands) <EJ.Wagenmakers@gmail.com>

Wago, **Hajime**
Kyoto Sangyo University (Japan) <wago@ism.ac.jp>

Walshaw, **David**
Newcastle University (UK) <david.walshaw@ncl.ac.uk>

Wang, **Jing**
University of Michigan (USA) <jingjw@umich.edu>

Wang, **Joanna (Jia Jia)**
The University of Sydney (Australia) <jwang43@gmail.com>

Wang, **Shouqiang**
Duke University (USA) <sw55@duke.edu>

Wang, **Xiaojing**
Duke University (USA) <xiaojing.wang@duke.edu>

Wauthier, **Fabian**
Univ. California at Berkeley (USA) <flw@berkeley.edu>

West, **Mike**
Duke University (USA) <mw@stat.duke.edu>

Wetzels, **Ruud**
Universiteit van Amsterdam (The Netherlands) <wetzels.ruud@gmail.com>

White, **Gentry**
The University of Queensland (Australia) <gentry.white@uq.edu.au>

White, **Nicole**
Queensland University of Technology (Australia) <a.ahlberg@qut.edu.au>

Wilkinson, **Darren James**
Newcastle University (UK) <d.j.wilkinson@ncl.ac.uk>

Wilson, **Kevin**
Newcastle University (UK) <k.j.wilson@ncl.ac.uk>

Wilson, **Melanie**
Duke University (USA) <melanie.wilson@duke.edu>

Wilson, **Simon**
Trinity College Dublin (Ireland) <simon.wilson@tcd.ie>

Williamson, **Daniel**
Durham University (UK) <daniel.williamson@durham.ac.uk>

Williamson, **Sinead**
University of Cambridge (UK) <saw56@cam.ac.uk>

Winther, **Ole**
Københavns Universitet (Denmark) <owi@imm.dtu.dk>

Wiper, **Michael**
Universidad Carlos III de Madrid (Spain) <michael.wiper@uc3m.es>

Wolpert, **Robert**
Duke University (USA) <wolpert@stat.duke.edu>

Womack, **Andrew**
Washington University in St. Louis (USA) <ajwomack@math.wustl.edu>

Wraith, **Darren**
INRIA Alpes (France) <darren.wraith@inrialpes.fr>

Xifara, **Tatiana**
Lancaster University (UK) <t.xifara@lancaster.ac.uk>

Xu, **Chang**
University of Missouri (USA) <cx3z9@mail.missouri.edu>

Yang, **Hongxia**
Duke University (USA) <hy35@duke.edu>

Yau, **Christopher**
University of Oxford (UnitedKingdom) <yau@stats.ox.ac.uk>

Yildirim, **Sinan**
University of Cambridge (UK) <sy276@cam.ac.uk>

Yoon, **Jiwon**
Trinity College Dublin (Ireland) <jiwon.yoon@scss.tcd.ie>

Zucknick, **Manuela**
German Cancer Research Centre (Germany) <m.zucknick@dkfz-heidelberg.de>

Zwiernik, **Piotr**
University of Warwick (UK) <p.w.zwiernik@warwick.ac.uk>

Notes

Bayesian Singalong

The Bayesian Singalong Book

Bradley P. Carlin (editor)
University of Minnesota
brad@biostat.umn.edu

At the biennial international meetings on Bayesian statistics in Valencia, Spain and other beachfront locations as now selected by the International Society for Bayesian Analysis (ISBA), one of the most popular features (after the sun and the free wine) is the "cabaret" performance, which traditionally takes place on the last night following the conference dinner. Acts over the years have included jugglers, magicians, jokesters, and even the occasional male striptease (the now-infamous "Full Monty Carlo"). Still, the cornerstone of the cabaret has always been the singing of new and often humorous Bayes-related lyrics to popular songs, a practice dating to the landmark work of Box (1979; reprinted herein).

In this abbreviated collection, we present a few of the songs that have been performed at Bayesian cabarets over the years, emphasizing those that might make good singalong material for the audience at the Valencia 9 cabaret. For more complete information, visit **www.biostat.umn.edu/~brad/cabaret.html**.

CONTENTS

The Songs

Bayeseamus Igitur

Words: A.P. Dawid[1]
Music: medieval student drinking song
First performance: Valencia 3

V1: Let's have more fun while we can
Until the whole world's Bayesian!
Since our prior expectation
Of posterior location
In the limit's six feet down.

V2: We'll drink a toast to former days
When everyone loved Thomas Bayes.
Be he now in Heaven or El
Lobo's den let's wish him well:
Here's to what his Theorem says!

V3: If it's inference you desire
Make the toast, "Long live the prior!"
Set it up and hit it later
With the model and the data –
That's how Dennis could retire.

V4: Frequentists may spit and curse
But they're in for something worse:
To the depths of Hell so ample
May they take repeated sample
Theory with them in the hearse!

V5: Now Adrian will lead the cheer
And we'll toast "Valencia!"
Where Jose's inspired ambition
Sets the Bayesian position:
He's the reason why we're here.

V6: And so we'll raise another glass:
"More Bayesian Valencias!"
But Morrie's gaining weight in practice
For our last toast, which is in fact is
"To the days Bayes rules at last!"

[1]being a loose translation, from the original Latin of Mammitzsch et al. (1987)

Bayesian

Words: B. Carlin
Music: Olivia Newton-John ("Physical")
First performance: ISBA 2008 (Hamilton Island, Australia)

V1: I know they taught you all about that Fisher man,
and how to get the MLE,
I wanna get you to see the light – you know what I mean,

I poked a little fun at Neyman and Pearson, contrasting them
with deFinetti,
There's nothin' left to talk about, 'less it's – subjectivity!

Chorus: Let's get Bayesian, Bayesian
I wanna get Bayesian, let's get into Bayesian,
Show me your posterior, posterior – show me your posterior,

Oh, let's get Bayesian, Bayesian
I wanna get Bayesian, let's get into Bayesian,
Let me see your DIC, DIC – let me see your DIC!

V2: I've been patient, I've been good, searchin' for your prior feelings,

You gotta know there's no holdin' back – you know what I mean,

I'll help you understand those convergence plots, and every model
hierarchy,
Your bimodality's bringin' out the analyst in me!

Chorus: (repeat)

Solo: (guitar? / chorus changes here still...)

Chorus: (repeat)

Ending: Oh, let's get animal, animal
– I wanna get animal, let's get into animal
Show me your posterior, posterior – show me your posterior....
Show me your posterior – show me your posterior....

Bayesian Believer

Words: B. Carlin
Music: Neil Diamond/The Monkees/Smashmouth ("I'm a Believer")
First performance: Valencia 7

Intro: (key/guitar lick)

V1: I thought inference was just a fairy tale,

Confused by stats and probability,

Frequentist approaches (doo-doot doo-doot)

made no sense to me (doo-doot doo-doot)

Summarizing evidence by p?!

Chorus: Then I saw Tom Bayes – Now I'm a believer,

Without a trace – of doubt in my mind,

[I'm a] Bayesian (ooooh) – Oh, I'm a believer –

I couldn't p now if I tried!

V2: I thought likelihood was just the only thing,
Turn the crank and get the MLE,
What's the use of thinking (doo-doot doo-doot)
Disconnect your brain (doo-doot doo-doot)
Play along and minimize the pain...

Chorus: (repeat)
Solo: (keys/guitar)

V3: `SAS` was out to get me (doo-doot doo-doot)
(partial) – that's the way it seemed (doo-doot doo-doot)

Fixed effects and forced normality...

Chorus: (repeat 2x w/assorted hollering and out!)

Bayesian Prison Blues

Words: B. Carlin
Music: Johnny Cash ("Folsom Prison Blues")[2]
First performance: Valencia 8

Intro: "Ciao, Io sono Giovanni Soldi..." (guitar lead)

V1: I hear the plane a comin' – it's flyin' toward the beach,
And I've got lots of sunshine, but the water's out of reach,

I'm stuck in Bayesian Prison – and time keeps draggin' on...

But that airplane keeps a'movin' – on down to Ben-i-dorm!

V2: When I was back in grad school – my professor told me, please,
Always use a Bayes factor – don't ever play with p's,
But that alpha-spending function really caught my eye....
Now I'm lost and incoherent – and locked in here to die!

Solo: (guitar lead – Sooey!)

V3: I can see those Bayesians standin' 'round some late-night poster session,
Drinkin' wine and talkin' 'bout some multi-stage regression,
Well I know I had it comin' – I know I can't be free...
But they're samplin' my posterior – and that's what tortures me!

Solo: (guitar lead – Hidee-ho!)

V4: Well if they freed me from this prison, that would be my lucky day,
I'd hit the conference dinner, then join the cabaret,
Far from Bayesian Prison – that's where I long to be...
You could watch me runnin' WinBUGS – and check my D-I-C!

Outtro:(guitar lick!)

[2]English translation of Intro: "Hello, I'm Johnny Cash..."

Bayesian Wonderland

Words: H. Ashih and R.A. Reutter
Music: D. Smith and F. Bernard ("Winter Wonderland")
First performance: Valencia 6

V1: Glasses clink, are you listenin'?
Have a drink, [the] wine is glistenin'!
A beautiful sight, we're tipsy tonight,
Stumblin' through our Bayesian Wonderland.

V2: It's a bad situation,
you get a fault of segmentation,
A long sleepless night, your program's not right,
Strugglin' with the Bayesian paradigm.

Bridge 1: In the theory we can build a sampler
With the jumps reversible in time,
But in practice it's not quite that simple,
So conjugate analysis is fine!

V3: P-val's stink, where's your prior?
It can't be flat, or you're a liar,
Ask what is known, not what is shown,
To specify our Bayesian Wonderland!

(potential solo break, over V1 and V2)

Bridge 2: In the theory we can build a sampler,
With convergence surely guaranteed,
But beware of autocorrelations,
Or it will take forever to succeed!

V4: When it runs, ain't it thrillin',
To the last iteration,
It frolics and plays, throughout n-space,
Walkin' in a Bayesian Wonderland.

Ending: Random walkin' in a Bayesian Wonderland!

Bayesians in the Night

Words: B. Natvig and M. DeGroot[3]
Music: Kaenpfert/Singleton/Snyder ("Strangers in the Night")
First performance: Valencia 3

V1:
Bayesians in the night
with exchangeable glances
Assessing in the night
the prior chances
We'd be sharing risks
before the night was through.

V2:
Something in your prior
was so exciting
Something in your data
was so inviting
Something in my model
told me I must have you.

Bridge:
Bayesians in the night
two statisticians
We were Bayesians in the night
Then came the moment when we walked down to the sea
Under a fault tree
Our likelihoods were close together
and Sir Ronald lost his final feather

V3:
And ever since that night
we've been adherents
Leaders of the fight
to have coherence
It turned out all right
for Bayesians in the night.

[3]original version by Natvig (1986) replaces the second part of the bridge with:
"Until the moment that we'd said our first hello
Little did we know
Love was just a glance away
and Sir Ronald never thought that way."

I.S.B.A.

Words: J. Wakefield, D. Stephens, and B. Carlin
Music: The Village People ("Y.M.C.A.")
First performance: Valencia 5; Second performance: ISBA 2000 (Crete)

V1: Bayesians – won't you listen to me,
I said, Bayesians – find out what you can be,
So just come on – to the I.S.B.A.,
It will boost your career today!

Bayesians – do you want something more,
I said, Bayesians – is your research a bore,
Then just come on – to the I.S.B.A.,
Because they will take you anyway!

Chorus 1: It's fun to be in the I.S.B.A – it's fun to be in the I.S.B.A!
You can grease a few palms – go hunting for jobs,
You can suck up to all the knobs!

It's fun to be in the I.S.B.A – it's fun to be in the I.S.B.A!
You can work on your tan – you can swim in the sea,
You can hang out with Arnie Zee!

V2: ISBA – so the newsletter's late,
But at ISBA – the food is just great,
And though we don't know – where the meetings will be,
I'm sure they will work it out finally.

ISBA – the location is fine,
And at ISBA – there is lots of free wine,
We've got a journal – and though it's fully online,
We still can't seem to stay on time!

Chorus 2: It's fun to be in the I.S.B.A – it's fun to be in the I.S.B.A!
You can drink a few beers – go hunting for jobs,
You can suck up to all the knobs!

It's fun to be in the I.S.B.A – it's fun to be in the I.S.B.A!
You can work on your tan – you can go back to bed,
You can hang out with Arnie Zed!

(Repeat Chorus 1, and out)

Imagine

Words: B. Carlin and R. McCulloch
Music: John Lennon ("Imagine")
First performance: Valencia 4

V1: Imagine you're a Bayesian–
It's easy if you try,
You just adopt a prior,
And the data updates π.
Statistics is so simple
With subjective probabilityyyyy – ah-ah! ah ah...

V2: Now imagine you're a frequentist,
Worrying about what might have been,
Spending your whole lifetime
Analyzing data you've never seen.
And if you want an interval,
You'll need a pivotal quantityyyyy – ah-ah! ah ah...

Chorus: You may say I sound like a textbook –
But I'm not the only one:
Every four years we all get together,
To talk, drink beer, and lie in the sun.

V3: We used to sweat computation –
But MCMC took care of that,
And if you want elicitation,
Then Kadane et al. is where it's at.
And Jose and Jim talk reference priors –
Building on work by Jeffreyyyyys – ah-ah! ah ah...

Chorus: You may say, "He must've flunked out at Berkeley,"
But you stick around and see,
All the misguided will someday join us –
And then the world will *finally* be free!

An MCMC Saga

Words: J. Rosenthal
Music: Elvis Presley ("Jailhouse Rock")
First performance: Valencia 7

```
Verse 1:
Had some data ready to inspect,
   I modeled the relation as a random effect,
The number of parameters just grew and grew,
   I had to get some help from you-know-who!

Chorus:
Run run -- Markov chain run,
  Programming you was fun, but I'll be happy when you're done!

Verse 2:
I coded up a simple MCMC,
   To do all the difficult work for me,
The sampler went funny and refused to mix,
   Caused me a problem that I couldn't fix!

Chorus:  (repeat)

Solo 1:  Harmonica (one verse)

Verse 3:
It seemed that my posterior required more,
   A super-duper sampler it couldn't ignore,
With Langevin, and tempering, a hybrid chain,
   I had to tweak it again and again!

Chorus:  (repeat)

Solo 2:  Piano (two verses)

Verse 4:
I knew that my algorithm was no joke,
   When my computer started spewing smoke,
My plan wasn't working so I had to sub,
   I drowned my MC sorrows at the local pub!


Chorus:  (repeat twice and out)
```

Prior

Words: M. Glickman
Music: Shocking Blue ("Venus")
First performance: MCMSki II (2008; Bormio, Italy)

```
Verse 1:
I had some extra information -- didn't know where it should go,

  A method to express this knowledge -- is what I don't know...

Chorus:
I've got it -- yeah baby, I've got it!

  Well, I'm the thesis -- I'm the prior that you require!

  Well, I'm the thesis -- I'm the prior that you require!

(Guitar solo -- one verse)

Verse 2:
I had myself a complex model -- I didn't know how to constrain,

  I tried to estimate the unknowns -- my attempts were in vain!

Chorus:  (repeat)

Break

Half-verse:  "Ahhhhh....."

Chorus:  (repeat)

Break

Intro guitar riff to end
```

The Simulator

Words: M. Huber
Music: Kenny Rogers ("The Gambler")
First performance: ISBA 2008 (Hamilton Island, Australia)

V1: On a warm summer's evening on a plane bound for ISBA,
I met up with the simulator; we were both too tired to sleep,
So we took turns a-staring out the window at the darkness,
'Til the boredom overtook us, and he began to speak.

V2: He said,
''Son, I've made a lifetime out of reading people's data,
And knowing what the numbers tell, by the way the series lies,
So if you don't mind my saying, I can see you're out of models,
For a taste of your data, I'll give you some advice."

V3: So I handed him my laptop, and he downloaded my last file,
Then he bummed a thumbdrive, and I watched its blinking light,
And the night got deathly quiet, and his face lost all expression,
Said, "If you're gonna run the chains, boy,
ya gotta learn to do it right.

Chorus:
You got to know when to propose 'em, know when to reject 'em,
Know when to stop a chain, and know when to run,
You never find your error, while you're still collecting samples,
There'll be time enough for error bars, when the chains are done.

V4: Every runner knows that the secret to chain burn-in,
Is knowing what to throw away, and knowing what to keep,
Because every run is perfect, and every run is worthless,
And the best that you can hope for is results you can repeat.

V5: And when he'd finished speaking,
he turned back towards the window,
Closed down the black ThinkPad, and faded off to sleep,
And somewhere in the darkness, his chain it reached convergence,
But in his final words I found a trick that I could keep.

Chorus: (repeat and out!)

Statistician

Words: R. Weiss and B. Carlin
Music: Roy Orbison ("Oh, Pretty Woman")
First performance: ISBA 2008 (Hamilton Island, Australia)

```
VERSE 1:
       A              F#m                          A
Statistician, walking down the street -- Statistician,
                                                          F#m
       D                   E              the kind I like to meet,
Statistician -- I don't believe you, you're not the truth,
                                  (Riff 2, 4x)
No one computes as good as you -- Mercy!

VERSE 2:
Statistician, come on have some fun, Statistician, be a Bayesian,
Statistician -- Your tests are lovely as can be,
But not significant like me -- Rrrrowrrrr!.....

BRIDGE:
Dm            G7          C             Am
Statistician, simulate -- Statistician, iterate,
Dm            G7              C
Statistician, run it late for me
Dm            G7                C             Am
Statistician, yeah yeah yeah -- Statistician, think my way
Dm            G7             C    A
Statistician, integrate with me....
            F#m   Dm          E
'Cause your data, it looks so right,
A            F#m   Dm          E  (Riff 2, 4x)
Come with me baby, run BUGS toni -- i -- ight!

V3:    A            F#m                A             F#m
Statistician, don't maximize -- Statistician, try me on for size,
       D             E
Statistician -- Don't walk away, hey...
E
OK... your p is less than oh-five, OK...
I guess I'll go on home, it's late,
There'll be tomorrow night, but wait -- What do I see?....

Is she drifting back to me?...Yeah, she's converging back to me...
               A
Oh, oh, Statistician!
```

There's No Theorem Like Bayes' Theorem

Words: G.E.P. Box
Music: Irving Berlin ("There's no Business like Show Business")
First performance: Valencia 1

V1: The model, the data you can't wait to see
The theta, beta, sigma, and the rho
The Normal, the Poisson, the Cauchy, the t
The need to specify what you don't know
The likelihood for data you acquire
The perspicacious choosing of the prior

Refrain: There's no theorem like Bayes' theorem
Like no theorem we know
Everything about it is appealing
Everything about it is a wow
Let out all that a priori feeling
You've been concealing right up to now!

There's no people like Bayes people
All odd balls from the urn
The other day you thought that you had got it straight
Take my advice and don't celebrate
A paradox by Lindley could arrive quite late
Another Stone to unturn!

Refrain: There's no theorem like Bayes' theorem
Like no theorem we know
You can lose forever that perplexed look
If you start to study it right now
Even more enthralling than a sex book
You'll find that textbook by Box and Tiao!

There's no dogma like Bayes' dogma
It's great knowing you're right
We know of a fiducialist who knew the lot
We thought at first he had hit the spot
But after three more seminars we lost the plot
We just could not see the light!

Refrain: There's no theorem like Bayes' theorem
Like no theorem we know

Fisher felt its use was quite restricted
Except in making family plans for mice
But there, he said, for pinning down a zygote
I'd give it my vote and not think twice!

There're no answers like Bayes' answers
Transparent, clear and precise
Stein's conundrums you can solve without a blink
Best estimators in half a wink
You can even understand what makes 'em shrink
Their properties are so nice!

V2: There's Raiffa and Schlaifer, Mosteller & Pratt
There's Geisser, Zellner, Novick, Hill and Tiao
And these all are people who know what they're at
They represent Statistics' finest flower
And tho' on nothing else they could agree
With us they'd join and sing in harmony!

Refrain: There's no theorem like Bayes' theorem
Like no theorem we know
Just recall what Pearson said to Neyman
Emerging from a region of type B
"It's difficult explaining to the Lehmann;
I fear it lacks Bayes' simplicity!"

There's no haters like Bayes' haters
They spit when they see a prior
Be careful when you offer your posterior
They'll try to kick it right through the door
But turn the other cheek if it is not too sore
Of error they may yet tire!

Refrain: There's no theorem like Bayes' theorem
Like no theorem we know
Critics carp at Bayes's hesitation
Claiming that his doubts on what he'd done
Led to late posthumous publication
We will explain that to everyone:

When Bayes got up to Heaven
He asked for an interview
Jehovah quickly told him he had got it right
Bayes popped down earthwards at dead of night
His spectre ceded Richard Price the copyright
It's very strange but it's true!!

Voodoo Bayesian Child

Words: E. Green
Music: Willie Dixon/Muddy Waters ("Hoochie Coochie Man")
First performance: Valencia 9?

Verse 1:
Harold Jeffreys told my mother -- 'fore I was born,
You got a Bayesian comin' -- gonna be a son of a gun,
Gonna make all them frequentists -- stand up and shout,
Then the world gonna know -- what Bayes all about!

Chorus:
But you know I'm here -- everybody knows I'm here,
Well I'm the voodoo Bayesian child -- everybody knows I'm here!

Verse 2:
Got a prior and a likelihood -- got a posterior too,
Got a full inference system -- I'm gonna slip it to you,
Gonna make all my clients -- jump up and shout,
Cause they'll finally understand -- what I'm talking 'bout!

Chorus: (repeat)

Verse 3:
Now we got our own journal -- and our own society,
I think there's a secret handshake -- but no one's shown it to me,
We can out-drink all others -- and still function at dawn,
With a hundred versions of WinBUGS -- all our analytical
problems are gone!

Chorus: (repeat)