

# Integrated Objective Bayesian Estimation and Hypothesis Testing

**José M. Bernardo**

Universitat de València, Spain

jose.m.bernardo@uv.es

**9th Valencia International Meeting on Bayesian Statistics**

**Benidorm (Alicante, Spain), June 8th, 2010**

# Summary

- (i) **Foundations**
- (ii) **Bayesian Inference Summaries**
- (iii) **Loss Functions**
- (iv) **Objective Bayesian Methods**
- (v) **Integrated Reference Analysis**
- (vi) **Basic References**

# Foundations

## Time to revise foundations?

- No obvious agreement on the appropriate Bayesian solution to even simple (textbook) stylized problems:

Best point estimate for the normal variance?

Inferences on the correlation coefficient of a bivariate normal?

Comparing two normal means or two binomial proportions?

Testing compatibility of the normal mean with a precise value?

- Let alone in problems within complex models with many parameters!

**Proposal:** Return to basics and use decision-theoretic machinery.

## Bayesian Inference Summaries

- Assume data  $\mathbf{z}$  have been generated as one random observation from  $\mathcal{M}_{\mathbf{z}} = \{p(\mathbf{z} | \boldsymbol{\theta}, \boldsymbol{\lambda}), \mathbf{z} \in \mathcal{Z}, \boldsymbol{\theta} \in \Theta, \boldsymbol{\lambda} \in \Lambda\}$ , where  $\boldsymbol{\theta}$  is the **vector of interest** and  $\boldsymbol{\lambda}$  a nuisance parameter vector.
- Let  $p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = p(\boldsymbol{\lambda} | \boldsymbol{\theta}) p(\boldsymbol{\theta})$  be the assumed joint prior.
- Given data  $\mathbf{z}$  and assuming model  $\mathcal{M}_{\mathbf{z}}$ , the **complete** solution to all inference questions about  $\boldsymbol{\theta}$  is contained in the **marginal posterior**  $p(\boldsymbol{\theta} | \mathbf{z})$ , derived by standard use of probability theory.
- Appreciation of  $p(\boldsymbol{\theta} | \mathbf{z})$  may be enhanced by providing both point and region **estimates** of the vector of interest  $\boldsymbol{\theta}$ , and by declaring whether or not some context-suggested specific value  $\boldsymbol{\theta}_0$  (or maybe a set of values  $\Theta_0$ ), is (are) **compatible** with the observed data  $\mathbf{z}$ . All of these provide useful (and often required) **summaries** of  $p(\boldsymbol{\theta} | \mathbf{z})$ .

## Decision-theoretic structure

- All these summaries may be framed as different **decision problems** which use precisely the same **loss function**  $\ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$  describing, as a function of the (unknown)  $(\boldsymbol{\theta}, \boldsymbol{\lambda})$  values which have generated the data, the loss to be suffered if, working with model  $\mathcal{M}_z$ , the value  $\boldsymbol{\theta}_0$  were used as a proxy for the unknown value of  $\boldsymbol{\theta}$ .
- The results dramatically depend on the choices made for both the prior and the loss functions but, given  $\mathbf{z}$ , only depend on those through the **expected loss**,  $\bar{\ell}(\boldsymbol{\theta}_0 | \mathbf{z}) = \int_{\Theta} \int_{\Lambda} \ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathbf{z}) d\boldsymbol{\theta} d\boldsymbol{\lambda}$ .
- As a function of  $\boldsymbol{\theta}_0 \in \Theta$ ,  $\bar{\ell}(\boldsymbol{\theta}_0 | \mathbf{z})$  is a measure of the **unacceptability** of all possible values of the vector of interest. This provides a **dual**, complementary information on all  $\boldsymbol{\theta}$  values (on a loss scale) to that provided by the posterior  $p(\boldsymbol{\theta} | \mathbf{z})$  (on a probability scale).

## □ Point estimation

To choose a point estimate for  $\theta$  is a decision problem where the action space is the class  $\Theta$  of all possible  $\theta$  values.

**Definition 1** *The Bayes estimator  $\theta^*(z) = \arg \inf_{\theta_0 \in \Theta} \bar{\ell}(\theta_0 | z)$  is that which minimizes the posterior expected loss.*

- Conventional examples include the ubiquitous quadratic loss  $\ell\{\theta_0, (\theta, \lambda)\} = (\theta_0 - \theta)^t(\theta_0 - \theta)$ , which yields the **posterior mean** as the Bayes estimator, and the zero-one loss on a neighborhood of the true value, which yields the **posterior mode** as a limiting result.
- Bayes estimators with conventional loss functions are typically **not invariant** under one to one transformations. Thus, the Bayes estimator under quadratic loss of a variance is not the square of the Bayes estimator of the standard deviation. This is **rather difficult to explain** when one merely wishes to report an estimate of some quantity of interest.

## □ Region estimation

Bayesian region estimation is achieved by quoting posterior credible regions. To choose a  $q$ -credible region is a decision problem where the action space is the class of subsets of  $\Theta$  with posterior probability  $q$ .

**Definition 2** (Bernardo, 2005). A Bayes  $q$ -credible region  $\Theta_q^*(\mathbf{z})$  is a  $q$ -credible region where any value within the region has a smaller posterior expected loss than any value outside the region, so that  $\forall \theta_i \in \Theta_q^*(\mathbf{z}), \forall \theta_j \notin \Theta_q^*(\mathbf{z}), \bar{\ell}(\theta_i | \mathbf{z}) \leq \bar{\ell}(\theta_j | \mathbf{z})$ .

- The quadratic loss yields credible regions with those  $\theta$  values closest, in the Euclidean sense, to the posterior mean. A zero-one loss function leads to highest posterior density (HPD) credible regions.
- Conventional Bayes regions are often **not invariant**: HPD regions in one parameterization will not transform to HPD regions in another.

## □ Precise hypothesis testing

- Consider a value  $\boldsymbol{\theta}_0$  which deserves special consideration. Testing the hypothesis  $H_0 \equiv \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$  is as a decision problem where the action space  $\mathcal{A} = \{a_0, a_1\}$  contains only two elements: to accept ( $a_0$ ) or to reject ( $a_1$ ) the hypothesis  $H_0$ .

- Foundations require to specify the loss functions  $\ell_h\{a_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$  and  $\ell_h\{a_1, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$  measuring the consequences of accepting or rejecting  $H_0$  as a function of  $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ . The optimal action is to reject  $H_0$  iif

$$\int_{\Theta} \int_{\Lambda} [\ell_h\{a_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} - \ell_h\{a_1, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}] p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathbf{z}) d\boldsymbol{\theta} d\boldsymbol{\lambda} > 0.$$

- Hence, only  $\Delta\ell_h\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} = \ell_h\{a_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} - \ell_h\{a_1, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$ , which measures the **conditional advantage of rejecting**, must be specified.



- Without loss of generality, the function  $\Delta\ell_h$  may be written as

$$\Delta\ell_h\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} = \ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} - \ell_0$$

where (**precisely as in estimation**),  $\ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$  describes, as a function of  $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ , the non-negative loss to be suffered if  $\boldsymbol{\theta}_0$  were used as a proxy for  $\boldsymbol{\theta}$ , and the constant  $\ell_0 > 0$  describes (in the same loss units) the **context-dependent** non-negative advantage of accepting  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$  when it is true.

**Definition 3** (*Bernardo and Rueda, 2002*). *The Bayes test criterion to decide on the compatibility of  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$  with available data  $\boldsymbol{z}$  is to reject  $H_0 \equiv \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$  if (and only if),  $\bar{\ell}(\boldsymbol{\theta}_0 | \boldsymbol{z}) > \ell_0$ , where  $\ell_0$  is a context dependent positive constant.*

- The compound case may be analyzed by separately considering each of the values which make part of the compound hypothesis to test.

- Using a zero-one loss function, so that the loss advantage of rejecting  $\theta_0$  is equal to one whenever  $\theta \neq \theta_0$  and zero otherwise, leads to rejecting  $H_0$  if (and only if)  $\Pr(\theta = \theta_0 | \mathbf{z}) < p_0$  for some context-dependent  $p_0$ . Use of this loss requires the prior probability  $\Pr(\theta = \theta_0)$  to be *strictly positive*. If  $\theta$  is a continuous parameter this forces the use of a non-regular “sharp” prior, concentrating a positive probability mass at  $\theta_0$ , the solution early advocated by Jeffreys.

This formulation (i) implies the use of **radically different** priors for hypothesis testing than those used for estimation, (ii) precludes the use of conventional, often improper, ‘noninformative’ priors, and (iii) may lead to the difficulties associated to Jeffreys-Lindley paradox.

- The quadratic loss function leads to rejecting a  $\theta_0$  value whenever its Euclidean distance to  $E[\theta | \mathbf{z}]$ , the posterior expectation of  $\theta$ , is sufficiently large.

- The use of continuous loss functions (such as the quadratic loss) permits the use in hypothesis testing of precisely the same priors that are used in estimation.
- With conventional losses the Bayes test criterion is **not invariant** under one-to-one transformations. Thus, if  $\phi(\boldsymbol{\theta})$  is a one-to-one transformation of  $\boldsymbol{\theta}$ , rejecting  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$  does not generally imply rejecting  $\phi(\boldsymbol{\theta}) = \phi(\boldsymbol{\theta}_0)$ .
- The threshold constant  $\ell_0$ , which controls whether or not an expected loss is too large, is part of the specification of the decision problem, and should be context-dependent. However a judicious choice of the loss function leads to calibrated expected losses, where the relevant threshold constant has an immediate, **operational** interpretation.

## Loss Functions

- A **dissimilarity measure**  $\delta\{p_{\mathbf{z}}, q_{\mathbf{z}}\}$  between two probability densities  $p_{\mathbf{z}}$  and  $q_{\mathbf{z}}$  for a random vector  $\mathbf{z} \in \mathcal{Z}$  should be
  - (i) non-negative, and zero if (and only if)  $p_{\mathbf{z}} = q_{\mathbf{z}}$  a.e.,
  - (ii) invariant under one-to-one transformations of  $\mathbf{z}$ ,
  - (iii) symmetric, so that  $\delta\{p_{\mathbf{z}}, q_{\mathbf{z}}\} = \delta\{q_{\mathbf{z}}, p_{\mathbf{z}}\}$ ,
  - (iv) defined for densities with strictly nested supports.

**Definition 4** The **intrinsic discrepancy**  $\delta\{p_1, p_2\}$  is

$$\delta\{p_1, p_2\} = \min [\kappa\{p_1 | p_2\}, \kappa\{p_2 | p_1\}]$$

where  $\kappa\{p_j | p_i\} = \int_{\mathcal{Z}_i} p_i(\mathbf{z}) \log[p_i(\mathbf{z})/p_j(\mathbf{z})] d\mathbf{z}$  is the (KL) divergence of  $p_j$  from  $p_i$ . The intrinsic discrepancy between  $p$  and a family  $\mathcal{F} = \{q_i, i \in I\}$  is the intrinsic discrepancy between  $p$  and the closest of them,  $\delta\{p, \mathcal{F}\} = \inf_{q \in \mathcal{F}} \delta\{p, q\}$ .

## The intrinsic loss function

**Definition 5** Consider  $\mathcal{M}_z = \{p(z | \boldsymbol{\theta}, \boldsymbol{\lambda}), z \in \mathcal{Z}, \boldsymbol{\theta} \in \Theta, \boldsymbol{\lambda} \in \Lambda\}$ . The *intrinsic loss* of using  $\boldsymbol{\theta}_0$  as a proxy for  $\boldsymbol{\theta}$  is the intrinsic discrepancy between the true model and the class of models with  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ ,  $\mathcal{M}_0 = \{p(z | \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0), z \in \mathcal{Z}, \boldsymbol{\lambda}_0 \in \Lambda\}$ ,

$$\ell_\delta\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) | \mathcal{M}_z\} = \inf_{\boldsymbol{\lambda}_0 \in \Lambda} \delta\{p_z(\cdot | \boldsymbol{\theta}, \boldsymbol{\lambda}), p_z(\cdot | \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0)\}.$$

### □ Invariance

- For any one-to-one reparameterization  $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\theta})$  and  $\boldsymbol{\psi} = \boldsymbol{\psi}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ ,

$$\ell_\delta\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) | \mathcal{M}_z\} = \ell_\delta\{\boldsymbol{\phi}_0, (\boldsymbol{\phi}, \boldsymbol{\psi}) | \mathcal{M}_z\}.$$

This yields invariant Bayes point and region estimators, and invariant Bayes hypothesis testing procedures.

□ Reduction to sufficient statistics

- If  $\mathbf{t} = \mathbf{t}(\mathbf{z})$  is a sufficient statistic for model  $\mathcal{M}_{\mathbf{z}}$ , one may also work with marginal model  $\mathcal{M}_{\mathbf{t}} = \{p(\mathbf{t} | \boldsymbol{\theta}, \boldsymbol{\lambda}), \mathbf{t} \in \mathcal{T}, \boldsymbol{\theta} \in \Theta, \boldsymbol{\lambda} \in \Lambda\}$  since

$$\ell_{\delta}\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) | \mathcal{M}_{\mathbf{z}}\} = \ell_{\delta}\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) | \mathcal{M}_{\mathbf{t}}\}.$$

□ Additivity

- If data consist of a random sample  $\mathbf{z} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  from some model  $\mathcal{M}_{\mathbf{x}}$ , so that  $\mathcal{Z} = \mathcal{X}^n$ , and  $p(\mathbf{z} | \boldsymbol{\theta}, \boldsymbol{\lambda}) = \prod_{i=1}^n p(\mathbf{x}_i | \boldsymbol{\theta}, \boldsymbol{\lambda})$ ,

$$\ell_{\delta}\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) | \mathcal{M}_{\mathbf{z}}\} = n \ell_{\delta}\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) | \mathcal{M}_{\mathbf{x}}\}.$$

This considerably simplifies frequent computations.

## Objective Bayesian Methods

- The methods described may be used with **any** prior. However, an “objective” procedure, where the prior function is intended to describe a situation where there is no relevant information about the quantity of interest, is often required.
- **Objectivity** is an emotionally charged word, and it should be explicitly **qualified**. No statistical analysis is really objective (both the experimental design and the model have strong subjective inputs). However, frequentist procedures are branded as “objective” just because their conclusions are only conditional on the model assumed and the data obtained. Bayesian methods where the prior function is derived from the assumed model are objective in this **limited**, but precise sense.

## □ Development of objective priors

- Vast literature devoted to the formulation of objective priors.
- **Reference analysis**, (Bernardo, 1979; Berger and Bernardo, 1992; Berger, Bernardo and Sun, 2009), has been a popular approach.

Very general, easily computable one-parameter result:

**Theorem 1** *Let  $\mathbf{z}^{(k)} = \{\mathbf{z}_1, \dots, \mathbf{z}_k\}$  denote  $k$  conditionally independent observations from  $\mathcal{M}_z$ . For sufficiently large  $k$*

$$\pi_k(\theta) \propto \exp \{E_{\mathbf{z}^{(k)}|\theta}[\log p_h(\theta | \mathbf{z}^{(k)})]\}$$

*where  $p_h(\theta | \mathbf{z}^{(k)}) \propto \prod_{i=1}^k p(\mathbf{z}_i | \theta) h(\theta)$  is the posterior which corresponds to some arbitrarily chosen prior function  $h(\theta)$  which makes the posterior proper for any  $\mathbf{z}^{(k)}$ .*



## □ Approximate reference priors

- Reference priors are derived for an **ordered** parameterization. Given  $\mathcal{M}_z = \{p(\mathbf{z} | \boldsymbol{\omega}), \mathbf{z} \in \mathcal{Z}, \boldsymbol{\omega} \in \Omega\}$  with  $m$  parameters, the reference prior with respect to  $\boldsymbol{\phi}(\boldsymbol{\omega}) = \{\phi_1, \dots, \phi_m\}$  is sequentially obtained as  $\pi(\boldsymbol{\phi}) = \pi(\phi_m | \phi_{m-1}, \dots, \phi_1) \times \dots \times \pi(\phi_2 | \phi_1) \pi(\phi_1)$ .
- One is often **simultaneously** interested in several functions of the parameters. Given  $\mathcal{M}_z = \{p(\mathbf{z} | \boldsymbol{\omega}), \mathbf{z} \in \mathcal{Z}, \boldsymbol{\omega} \in \Omega \subset \mathfrak{R}^m\}$  with  $m$  parameters, consider a set  $\boldsymbol{\theta}(\boldsymbol{\omega}) = \{\theta_1(\boldsymbol{\omega}), \dots, \theta_r(\boldsymbol{\omega})\}$  of  $r > 1$  functions of interest; Berger, Bernardo and Sun (work in progress) suggest a procedure to select a joint prior  $\pi_{\boldsymbol{\theta}}(\boldsymbol{\omega})$  whose corresponding marginal posteriors  $\{\pi_{\boldsymbol{\theta}}(\theta_i | \mathbf{z})\}_{i=1}^r$  will be close, for all possible data sets  $\mathbf{z} \in \mathcal{Z}$ , to the set of reference posteriors  $\{\pi(\theta_i | \mathbf{z})\}_{i=1}^r$  yielded by the set of reference priors  $\{\pi_{\theta_i}(\boldsymbol{\omega})\}_{i=1}^r$  derived under the assumption that each of the  $\theta_i$ 's is of interest.

**Definition 6** Consider model  $\mathcal{M}_{\mathbf{z}} = \{p(\mathbf{z} | \boldsymbol{\omega}), \mathbf{z} \in \mathcal{Z}, \boldsymbol{\omega} \in \boldsymbol{\Omega}\}$  and  $r > 1$  functions of interest,  $\{\theta_1(\boldsymbol{\omega}), \dots, \theta_r(\boldsymbol{\omega})\}$ . Let  $\{\pi_{\theta_i}(\boldsymbol{\omega})\}_{i=1}^r$  be the relevant reference priors, and  $\{\pi_{\theta_i}(\mathbf{z})\}_{i=1}^r$  and  $\{\pi(\theta_i | \mathbf{z})\}_{i=1}^r$  the corresponding prior predictives and marginal posteriors. Let  $\mathcal{F} = \{\pi(\boldsymbol{\omega} | \mathbf{a}), \mathbf{a} \in \mathcal{A}\}$  be a family of prior functions. For each  $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ , the best approximate joint reference prior within  $\mathcal{F}$  is that which *minimizes the average expected intrinsic loss*

$$d(\mathbf{a}) = \frac{1}{r} \sum_{i=1}^r \int_{\mathcal{Z}} \delta\{\pi_{\theta_i}(\cdot | \mathbf{z}), p_{\theta_i}(\cdot | \mathbf{z}, \mathbf{a})\} \pi_{\theta_i}(\mathbf{z}) d\mathbf{z}, \quad \mathbf{a} \in \mathcal{A}.$$

- **Example.** Use of the Dirichlet family in the  $m$ -multinomial model (with  $r = m + 1$  cells) yields  $\text{Di}(\boldsymbol{\theta} | 1/r, \dots, 1/r)$ , with important applications to sparse multinomial data and contingency tables.

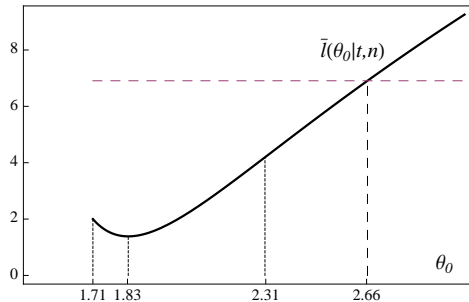
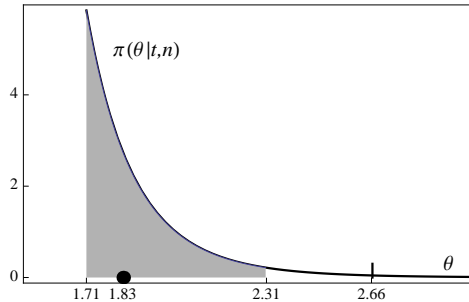
# Integrated Reference Analysis

- We suggest a systematic use of the **intrinsic loss function**, and an appropriate **joint reference prior**, for an integrated objective Bayesian solution to both estimation and hypothesis testing in **pure inference problems**.
- We have stressed foundations-like decision theoretic arguments, but a large collection of detailed, non-trivial examples prove that the procedures advocated lead to attractive, often novel solutions. Details in the paper version of this talk, and references therein.

## □ Estimation of the normal variance

- The intrinsic (invariant) point estimator of the normal standard deviation is  $\sigma^* \approx \frac{n}{n-1} s$ . Hence,  $\sigma^{2*} \approx \frac{n}{n-1} \frac{ns^2}{n-1}$ , larger than both the mle  $s^2$  and the unbiased estimator  $ns^2/(n-1)$ .

□ Uniform model  $\text{Un}(x \mid 0, \theta)$



$$\ell_\delta\{\theta_0, \theta \mid \mathcal{M}_z\} = n \begin{cases} \log(\theta_0/\theta), & \text{if } \theta_0 \geq \theta, \\ \log(\theta/\theta_0), & \text{if } \theta_0 \leq \theta. \end{cases}$$

$$\pi(\theta) = \theta^{-1}, \quad \mathbf{z} = \{x_1, \dots, x_n\},$$

$$t = \max\{x_1, \dots, x_n\}, \quad \pi(\theta \mid \mathbf{z}) = n t^n \theta^{-(n+1)}$$

The  $q$ -quantile is  $\theta_q = t(1 - q)^{-1/n}$ ;

Exact probability matching.

$$\theta^* = t 2^{1/n} \text{ (posterior median)}$$

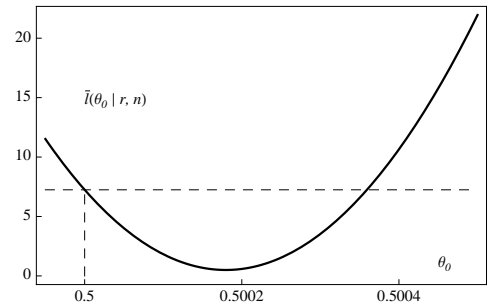
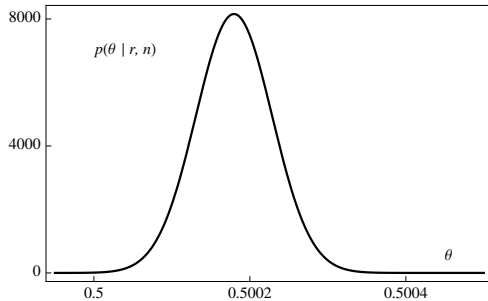
$$E[\bar{\ell}_\delta(\theta_0 \mid t, n) \mid \theta] = (\theta/\theta_0)^n - n \log(\theta/\theta_0);$$

this is equal to 1 if  $\theta = \theta_0$ ,

and increases with  $n$  otherwise.

- **Simulation:**  $n = 10$  with  $\theta = 2$  which yielded  $t = 1.71$ ;  
 $\theta^* = 1.83$ ,  $\Pr[t < \theta < 2.31 \mid \mathbf{z}] = 0.95$ ,  $\bar{\ell}_\delta(2.66 \mid \mathbf{z}) = \log 1000$ .

□ Extra Sensory Power (ESP) testing



Jahn, Dunne and Nelson (1987)  
 Binomial data. Test  $H_0 \equiv \{\theta = 1/2\}$   
 with  $n = 104,490,000$  and  $r = 52,263,471$ .

For any sensible **continuous** prior  $p(\theta)$ ,  
 $p(\theta | \mathbf{z}) \approx N(\theta | m_{\mathbf{z}}, s_{\mathbf{z}})$ ,  
 with  $m_{\mathbf{z}} = (r + 1/2)/(n + 1) = 0.50018$ ,  
 $s_{\mathbf{z}} = [m_{\mathbf{z}}(1 - m_{\mathbf{z}})/(n + 2)]^{1/2} = 0.000049$ .

$$\bar{\ell}(\theta_0 | \mathbf{z}) \approx \frac{n}{2} \log[1 + \frac{1}{n}(1 + t_{\mathbf{z}}(\theta_0)^2)],$$

$$t_{\mathbf{z}}(\theta_0) = (\theta_0 - m_{\mathbf{z}})/s_{\mathbf{z}}, \quad t_{\mathbf{z}}(1/2) = 3.672.$$

$$\bar{\ell}(\theta_0 | \mathbf{z}) = 7.24 = \log 1400: \text{Reject } H_0$$

- **Jeffreys-Lindley paradox:** With any “sharp” prior,  $\Pr[\theta = 1/2] = p_0$ ,  $\Pr[\theta = 1/2 | \mathbf{z}] > p_0$  (Jefferys, 1990) suggesting data **support**  $H_0$  !!!

□ More sophisticated examples

- **Two sample problems: Equality of two normal means.**

$$\bar{\ell}(H_0 | \mathbf{z}) \approx n \log[1 + \frac{1}{2n}(1 + t^2)], \quad t = \sqrt{n}(\bar{x} - \bar{y})/(s/\sqrt{2}).$$

- **Trinomial data: Testing for Hardy-Weinberg equilibrium.**

$$\bar{\ell}(H_0 | \mathbf{z}) \approx \int_{\mathcal{A}} \ell_{\delta}\{H_0, (\alpha_1, \alpha_2)\} \pi(\alpha_1, \alpha_2 | \mathbf{z}) d\alpha_1 d\alpha_2,$$

$$\text{where } \ell_{\delta}\{H_0, (\alpha_1, \alpha_2)\} = n \theta(\alpha_1, \alpha_2),$$

$\theta(\alpha_1, \alpha_2)$  is the KL distance of  $H_0$  from  $\text{Tri}(r_1, r_2, r_3 | \alpha_1, \alpha_2)$  and

$$\pi(\alpha_1, \alpha_2 | \mathbf{z}) = \text{Di}[\alpha_1, \alpha_2 | r_1 + 1/3, r_2 + 1/3, r_3 + 1/3].$$

- **Contingency tables: Testing for independence.**

Data  $\mathbf{z} = \{\{n_{11}, \dots, n_{1b}\}, \dots, \{n_{a1}, \dots, n_{ab}\}\}$ ,  $k = a \times b$ ,

$$\bar{\ell}(H_0 | \mathbf{z}) \approx \int_{\Theta} n \phi(\boldsymbol{\theta}) \pi(\boldsymbol{\theta} | \mathbf{z}) d\boldsymbol{\theta}, \quad \phi(\boldsymbol{\theta}) = \sum_{i=1}^a \sum_{j=1}^b \theta_{ij} \log \left[ \frac{\theta_{ij}}{\alpha_i \beta_j} \right],$$

where  $\alpha_i = \sum_{j=1}^b \theta_{ij}$  and  $\beta_j = \sum_{i=1}^a \theta_{ij}$  are the marginals, and

$$\pi(\boldsymbol{\theta} | \mathbf{z}) = \text{Di}_{k-1}(\boldsymbol{\theta} | n_{11} + 1/k, \dots, n_{ab} + 1/k).$$

## Basic References

(In chronological order)

- Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B* **41**, 113–147 (with discussion).
- Berger, J. O. and Bernardo, J. M. (1992). On the development of reference priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) Oxford: University Press, 35–60 (with discussion).
- Bernardo, J. M. (1997). Noninformative priors do not exist *J. Statist. Planning and Inference* **65**, 159–189 (with discussion).
- Bernardo, J. M. and Rueda, R. (2002). Bayesian hypothesis testing: A reference approach. *Internat. Statist. Rev.* **70**, 351–372.

- Bernardo, J. M. (2005a). Reference analysis. *Bayesian Thinking: Modeling and Computation, Handbook of Statistics* **25** (Dey, D. K. and Rao, C. R., eds). Amsterdam: Elsevier, 17–90.
- Bernardo, J. M. (2005b). Intrinsic credible regions: An objective Bayesian approach to interval estimation. *Test* **14**, 317–384 (with discussion).
- Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis* **1**, 385–402 and 457–464, (with discussion).
- Bernardo, J. M. (2007). Objective Bayesian point and region estimation in location-scale models. *Sort* **31**, 3–44, (with discussion).
- Berger, J. O., Bernardo, J. M. and Sun, D. (2009). The formal definition of reference priors. *Ann. Statist.* **37**, 905–938.
- Berger, J. O., Bernardo, J. M. and Sun, D. (2010). Reference priors for discrete parameters. *J. Amer. Statist. Assoc.* (under revision).