Un Programa de Síntesis para la Enseñanza de la Estadística Matemática

José-Miguel Bernardo

Universitat de València

<jose.m.bernardo@uv.es>
http://www.uv.es/˜bernardo

27 Congreso Nacional de Estadística e I. O.

Lleida, 8 de Abril, 2003

Contenido

1. Introducción a la Teoría de Decisión

- *Axiomas de comportamiento racional*
  - Estructura y axiomas
    $\mathcal{A} = \{a_i, i \in I\}, \quad \Theta_i = \{\theta_{ij}, j \in J_i\},$
    Comparabilidad, Transitividad, Consistencia,...
  - Resultados básicos:
    $\exists \, \Pr(\theta_{ij} \,|\, a_i, D) > 0, \sum_j \Pr(\theta_{ij} \,|\, a_i, D) = 1$
    $\exists \, l(a_i, \theta_{ij}) \in \Re$
    Debe minimizarse la utilidad esperada
    $\bar{l}(a_i \,|\, D) = \sum_j l(a_i, \theta_{ij}) \Pr(\theta_{ij} \,|\, a_i, D)$
- *Concepto de probabilidad*
  - Probabilidad: Medida condicional racional de la incertidumbre
    $\Pr(A|D)$, función de dos argumentos
    Casos particulares: Concepciones clásica y frecuentista.

- *Distribución de probabilidad como expresión de incertidumbre*
  - Ejemplo: Examenes de respuesta múltiple

    Pregunta

    $\boxed{p_1}$ Respuesta 1

    $\boxed{p_2}$ Respuesta 2

    ....

    $\boxed{p_k}$ Respuesta k

    ¿Como calificar $\boldsymbol{p} = \{p_1, p_2, \ldots, p_k\}$, $\Sigma p_i = 1$ en función de la respuesta correcta, R_i?. ¿Puntuación $u(\boldsymbol{p}, R_i)$?
  - $u(\boldsymbol{p}, R_i) = A \log[\Pr(R_i)] + B, \; A > 0$

    $$u(\boldsymbol{p}, R_i) = 1 - \log(p_i)/\log(1/k), \quad p_i = \Pr(R_i)$$

    $$u(\{1, 0, \ldots, 0\}, R_1) = 1; \quad u(\{1/k, 1/k, \ldots, 1/k\}, R_i) = 0;$$

    $$u(\{0, p_2, \ldots, p_k\}, R_1) = -\infty$$
  - $k = 2$, $p = \Pr(\text{Respuesta correcta})$,

    $$u(p) = 1 - \log(p)/\log(1/2) = 1 + \log(p)/\log(2)$$

- *Funciones generales de pérdida*
  - Discrepancia $l[p_{\boldsymbol{x}}(\cdot), q_{\boldsymbol{x}}(\cdot)]$ entre dos distribuciones de probabilidad del vector aleatorio $\boldsymbol{x} \in \mathcal{X}$, con densidades $p_{\boldsymbol{x}}(\cdot)$ y $q_{\boldsymbol{x}}(\cdot)$.
  - Discrepancia intrínseca, $\delta(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)$

    $$M \equiv \{p(\boldsymbol{x} \,|\, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}, \quad p(\boldsymbol{x} \,|\, \boldsymbol{\theta}) > 0, \quad \int_{\mathcal{X}} p(\boldsymbol{x} \,|\, \boldsymbol{\theta})\, d\boldsymbol{x} = 1$$

    $$k(\boldsymbol{\theta}_j \,|\, \boldsymbol{\theta}_i) = \int_{\mathcal{X}} p(\boldsymbol{x} \,|\, \boldsymbol{\theta}_i) \log \left[\frac{p(\boldsymbol{x} \,|\, \boldsymbol{\theta}_i)}{p(\boldsymbol{x} \,|\, \boldsymbol{\theta}_j)}\right] d\boldsymbol{x}$$

    $$\delta\{p_{\boldsymbol{x}}(\cdot \,|\, \boldsymbol{\theta}_i),\, p_{\boldsymbol{x}}(\cdot \,|\, \boldsymbol{\theta}_j)\} = \delta(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) = \min\{k(\boldsymbol{\theta}_i \,|\, \boldsymbol{\theta}_j), k(\boldsymbol{\theta}_j \,|\, \boldsymbol{\theta}_i)\}$$
  - Aplicaciones:

    $\delta\{p_{\boldsymbol{x}}(\cdot),\, \hat{p}_{\boldsymbol{x}}(\cdot)\}$ (*aproximaciones*)

    $\lim_{n\to\infty} p_n(\boldsymbol{x}) = p(\boldsymbol{x}) \iff \lim_{n\to\infty} \delta\{p_n(.), p(.)\} = 0$
    (*convergencia intrínseca*)

    $l(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \delta\{p_{\boldsymbol{x}}(\cdot \,|\, \boldsymbol{\theta}),\, p_{\boldsymbol{x}}(\cdot \,|\, \hat{\boldsymbol{\theta}})\} = \delta(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ (*estimación*)

    $l(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \delta\{p_{\boldsymbol{x}}(\cdot \,|\, \boldsymbol{\theta}),\, p_{\boldsymbol{x}}(\cdot \,|\, \boldsymbol{\theta}_0)\} = \delta(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ (*contraste*)

2. Modelos Probabilísticos

- *Intercambiabilidad*

  - $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots\}$ "homogéneas" (orden no proporciona información).

    Distribución conjunta invariante ante permutaciones: observaciones intercambiables.

    Caso particular: muestras aleatorias

- *Teorema general de representación*

  - Si $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots\}$ es una sucesión intercambiable, entonces

    $\exists\, g(\cdot),\ \exists\, \boldsymbol{\theta} = \lim_{n\to\infty} g(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$,

    $\exists\, p(\boldsymbol{x} \,|\, \boldsymbol{\theta}) > 0,\ \int_{\mathcal{X}} p(\boldsymbol{x} \,|\, \boldsymbol{\theta})\, d\boldsymbol{x} = 1$,

    $\exists\, p(\boldsymbol{\theta}) > 0,\ \int_{\Theta} p(\boldsymbol{\theta})\, d\boldsymbol{\theta} = 1$,

    y cualquier subconjunto finito tiene una representación integral

    $p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \int_{\Theta} \prod_{i=1}^{n} p(\boldsymbol{x}_i \,|\, \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$.

- Si los datos son homogeneos (intercambiables), entonces:

  (i) Son una muestra aleatoria de algún *modelo probabilístico* $p(\boldsymbol{x} \mid \boldsymbol{\theta})$, cuyo *parámetro* $\boldsymbol{\theta}$ es el límite de una función de las observaciones,

  (ii) La información de que se dispone sobre el valor de $\boldsymbol{\theta}$ *debe* ser descrita mediante una distribución de probabilidad $p(\boldsymbol{\theta})$.

- Se trata de un *teorema de existencia*:

  Se sabe que $p(\boldsymbol{x} \mid \boldsymbol{\theta})$ y $p(\boldsymbol{\theta})$ existen, pero deben ser determinados.

- *Especificación de modelos*
  - Si $\boldsymbol{x} \in \{0, 1\}^k$, modelo multinomial (binomial para $k = 1$).
  - Si $\boldsymbol{x} \in \Re^k$, y existe simetría esférica, modelo multinormal (normal univariante para $k = 1$).
  - Si $\boldsymbol{x} \in \Re^k$, y $\{\overline{\boldsymbol{x}}, S, n\}$ es suficiente, modelo multinormal (normal univariante para $k = 1$).

- *Función de verosimilitud*

  - Datos D, cuya distribución depende de $\boldsymbol{\theta}$, $L(\boldsymbol{\theta} \mid D) = p(D \mid \boldsymbol{\theta})$.
    Si los datos son intercambiables, $L(\boldsymbol{\theta} \mid D) = \prod_{i=1}^{n} p(\boldsymbol{x}_i \mid \boldsymbol{\theta})$.

  - Estimador máximo-verosímil, $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(D) = \arg\max_{\Theta} L(\boldsymbol{\theta} \mid D)$
    Análisis de existencia y unicidad.

  - Condiciones de regularidad: Función de información de Fisher, $F(\boldsymbol{\theta})$
    $F(\boldsymbol{\theta})_{ij} = -\mathrm{E}_{\boldsymbol{x}|\boldsymbol{\theta}}[\frac{\partial^2}{\partial\theta_i\partial\theta_j} \log p(\boldsymbol{x} \mid \boldsymbol{\theta})]$

  - Aproximación cuadrática de la verosimilitud
    $L(\boldsymbol{\theta} \mid D) \propto \mathrm{N}_k(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}},\, [nF(\hat{\boldsymbol{\theta}})]^{-1})$

- *Suficiencia*

  - Criterio de factorización

  - Familia exponencial de distribuciones

3. El Paradigma Bayesiano

- *Proceso de aprendizaje*

  - $D =$ datos, $\quad \{p(D \,|\, \boldsymbol{\theta}), p(\boldsymbol{\theta})\}$,
    $$L(\boldsymbol{\theta} \,|\, D) = p(D \,|\, \boldsymbol{\theta})$$
    $$p(\boldsymbol{\theta} \,|\, D) \propto p(D \,|\, \boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

  - Caso particular: $D = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, muestra aleatoria
    $$L(\boldsymbol{\theta} \,|\, D) = \prod_{i=1}^{n} p(\boldsymbol{x}_i \,|\, \boldsymbol{\theta})$$
    $$p(\boldsymbol{x} \,|\, D) = \int_{\Theta} p(\boldsymbol{x} \,|\, \boldsymbol{\theta})\, p(\boldsymbol{\theta} \,|\, D)\, d\boldsymbol{\theta}$$

  - Parámetros marginales: $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\omega})$, $\quad p(\boldsymbol{\theta}) = p(\boldsymbol{\phi})\, p(\boldsymbol{\omega} \,|\, \boldsymbol{\phi})$
    $$p(\boldsymbol{\phi} \,|\, D) = \int_{\Omega} p(\boldsymbol{\phi}, \boldsymbol{\omega} \,|\, D)\, d\boldsymbol{\omega}$$
    $$p(\boldsymbol{\phi} \,|\, D) \propto L(\boldsymbol{\phi} \,|\, D)\, p(\boldsymbol{\phi})$$
    $$L(\boldsymbol{\phi} \,|\, D) = \int_{\Omega} L(\boldsymbol{\phi}, \boldsymbol{\omega} \,|\, D)\, p(\boldsymbol{\omega} \,|\, \boldsymbol{\phi})\, d\boldsymbol{\omega}$$

- *Estimación*

  - El "estimador" bayesiano de $\boldsymbol{\theta}$ es $p(\boldsymbol{\theta} \mid D)$.

    Estimación puntual, $\tilde{\boldsymbol{\theta}}(D) \in \Theta$, con pérdida $l(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta})$

    $$\boldsymbol{\theta}^* = \arg\min_{\hat{\boldsymbol{\theta}} \in \Theta} \int l(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid D)\, d\boldsymbol{\theta}$$

    El estimador resultante depende de la función de pérdida.

  - Estimación por regiones, $C_p(D) \subset \Theta$

    Confianza p; $\quad \int_{C_p} p(\boldsymbol{\theta} \mid D)\, d\boldsymbol{\theta} = p$

    Existen infinitas soluciones.

    Regiones de máxima densidad en $\boldsymbol{\theta}$ (HPD)

- *Constraste de hipótesis*

  - Con función de pérdida $l(\boldsymbol{\theta}_0, \boldsymbol{\theta})$

    Criterio: Rechazar $H_0 \equiv \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$ sii

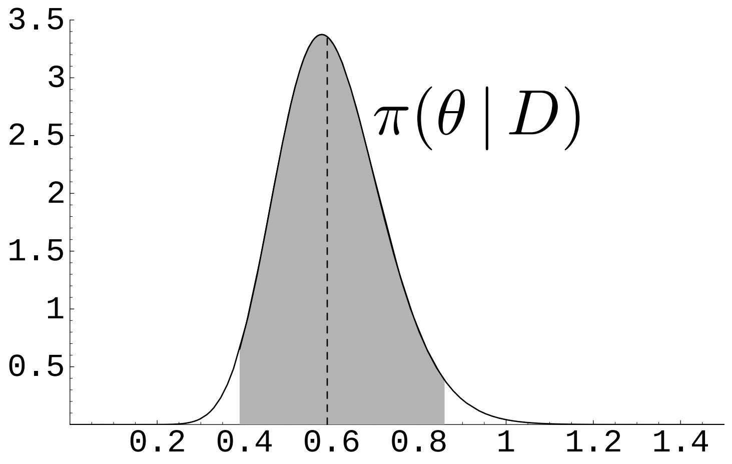
    $$\int l(\boldsymbol{\theta}_0, \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid D)\, d\boldsymbol{\theta} > l^*$$

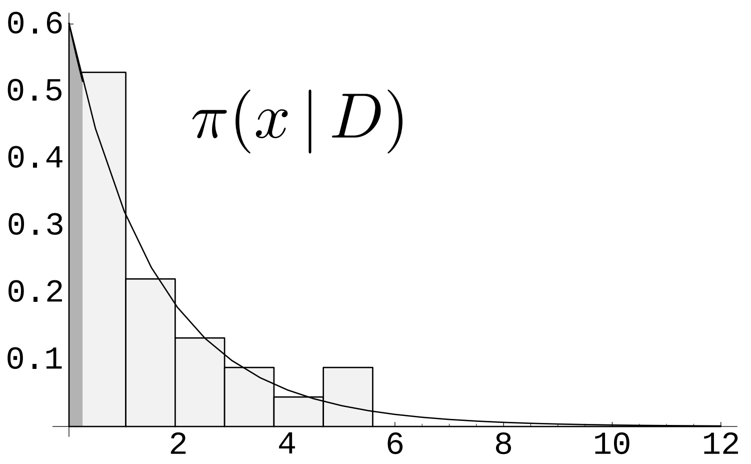
    l^* mide el valor de aceptar H_0 si es cierta.

    La región de rechazo depende de la función de pérdida.

- *Ejemplo: Datos exponenciales*
  - $p(x \mid \theta) = \theta e^{-\theta x}, \quad t = \sum_j x_j$ es suficiente
    $\pi(\theta) \propto \theta^{-1}, \quad \pi(\theta \mid D) = \text{Ga}(\theta \mid n, t)$
    $\pi(x \mid D) = n t^n (x + t)^{-(n+1)}$
  - Datos D: $n = 25$, $t = 41.57$, $1/\overline{x} = 0.601$

    $\Pr[0.39 < \theta < 0.86 \mid D] = 0.95$
    $\Pr[x < 0.25 \mid D] = 0.14$
    Estimador intrínseco, $\theta^* = 0.590$
  - Presentacion de resultados





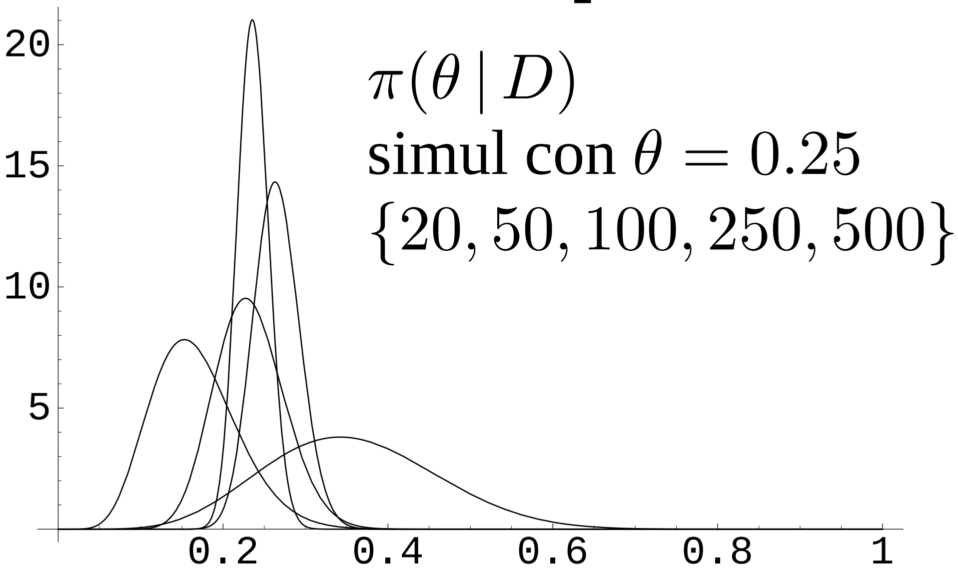
- *Comportamiento asintótico*

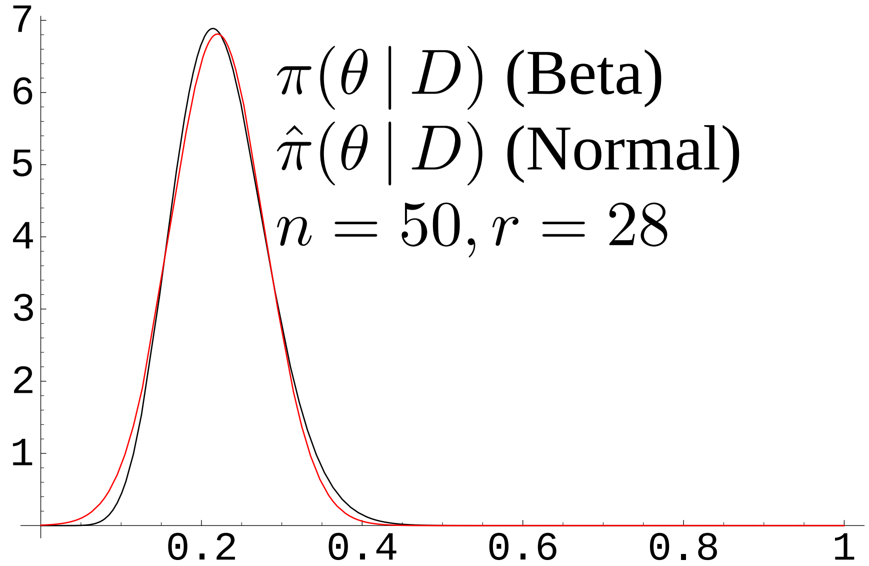
  - Caso discreto. $\Theta = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots\}$, $\boldsymbol{\theta}_t \in \Theta$

    $$\lim_{n\to\infty} \Pr[\boldsymbol{\theta} = \boldsymbol{\theta}_t \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] = 1$$
    $$\lim_{n\to\infty} \Pr[\boldsymbol{\theta} = \boldsymbol{\theta}_j \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] = 0, \quad \forall j \neq t$$

  - Caso continuo. $\Theta \subset \Re^k$ (n grande, $L(\boldsymbol{\theta} \,|\, D)$ regular)

    $$p(\boldsymbol{\theta} \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \approx N_k(\boldsymbol{\theta} \,|\, \hat{\boldsymbol{\theta}}, [nF(\hat{\boldsymbol{\theta}})]^{-1})$$
    $$F_{ij}(\boldsymbol{\theta}) = -\,\mathrm{E}_{\boldsymbol{x}\,|\,\boldsymbol{\theta}}\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log p(\boldsymbol{x} \,|\, \boldsymbol{\theta})\right]; \quad \text{(margs, conds)}$$





4. Métodos Bayesianos Objetivos

- *Análisis de referencia*
  - Información esperada

    $p(D \,|\, \theta), \quad \theta \in \Theta \subset \Re, \quad D \in \mathcal{D} \subset \Re^n$

    $p(\theta) \in \mathcal{P}$, distribuciones admisibles

    $I^\theta\{\mathcal{D}, p(\theta)\} = \mathrm{E}_{D\theta}\left[\log \frac{p(D,\theta)}{p(D)p(\theta)}\right]$ (Shannon)

  - Información desconocida

    $H_k\{p(\theta)\} = I^\theta\{\mathcal{D}^k, p(\theta)\}, \quad \pi_k(\theta) = \arg\sup_{p(\theta)\in\mathcal{P}} H_k\{p(\theta)\}$

    Cuando $k \to \infty$, $\pi_k(\theta)$ maximiza en $\mathcal{P}$ la info desconocida sobre θ.

  - Distribución final de referencia

    $\pi_k(\theta \,|\, D) \propto p(D \,|\, \theta)\, \pi_k(\theta)$

    $\pi(\theta \,|\, D) = \lim_{k\to\infty} \pi_k(\theta \,|\, D)$ (convergencia intrínseca)

  - Función inicial de referencia

    $\pi(\theta) > 0; \quad \forall D,\ \pi(\theta \,|\, D) \propto p(D \,|\, \theta)\, \pi(\theta)$

- Caso discreto finito, $\Theta = \{\theta_1, \ldots, \theta_m\}$

  $\lim_{k\to\infty} H_k\{p(\theta)\} = -\sum_{j=1}^{m} p_j \log p_j$

  $\pi(\theta)$ es la distribución de máxima entropia

  (uniforme sin restricciones adicionales)

- Caso contínuo, $\Theta \subset \Re$, $\hat{\theta}$ estimador consistente, asíntóticamente suficiente. Sin restricciones adicionales,

  $\pi(\theta) \propto \hat{p}(\theta \,|\, \hat{\theta})|_{\hat{\theta}=\theta}$ donde $\hat{p}(\theta \,|\, \hat{\theta})$

  es una aproximación asintótica cualquiera a la distribución final.

- Caso contínuo regular univariante, sin restricciones adicionales,

  $\pi(\theta) \propto i(\theta)^{1/2}$

  $i(\theta) = -\mathrm{E}_{D\,|\,\theta}\left[\frac{\partial^2}{\partial\theta^2}\log p(D\,|\,\theta)\right]$ (Jeffreys)

- Los problemas multivariantes pueden reducirse a la solución secuencial de una sucesión de problemas univariantes para obtener

  $\pi(\theta_k \,|\, \theta_1, \ldots, \theta_{k-1}) \times \cdots \times \pi(\theta_2 \,|\, \theta_1)\pi(\theta_1)$

- *Estimación puntual*

 - El verdadero "estimador" bayesiano objetivo de $\boldsymbol{\theta}$ es su densidad final de referencia $\pi(\boldsymbol{\theta} \mid D), \boldsymbol{\theta} \in \Theta \subset \Re^k$.

 - Pueden obtenerse estimadores puntuales $\boldsymbol{\theta}^*$ minimizando en Θ una pérdida esperada $\int l(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta})\, \pi(\boldsymbol{\theta} \mid D)\, d\boldsymbol{\theta}$

 Estimador intrínseco. La función de pérdida intrínseca

 $\delta(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) = \min\{k(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}_j), k(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_i)\}$

 $k(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_i) = \int_{\mathcal{X}} p(\boldsymbol{x} \mid \boldsymbol{\theta}_i) \log \left[\frac{p(\boldsymbol{x} \mid \boldsymbol{\theta}_i)}{p(\boldsymbol{x} \mid \boldsymbol{\theta}_j)}\right] d\boldsymbol{x}$

 da lugar a una pérdida esperada intrínseca

 $d(\tilde{\boldsymbol{\theta}} \mid D) = \int \delta(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta})\, \pi(\boldsymbol{\theta} \mid D)\, d\boldsymbol{\theta}$

 cuya minimización proporciona un único estimador invariante

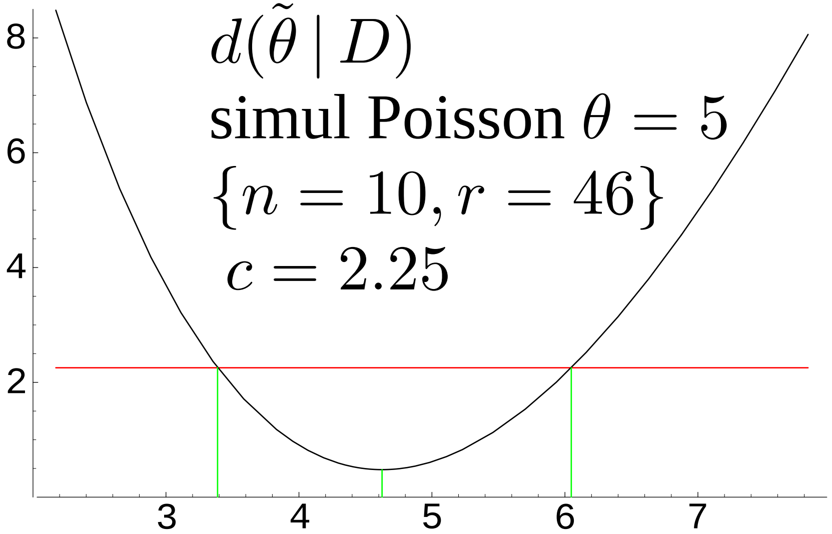
 $\boldsymbol{\theta}^* = \arg\min_{\tilde{\boldsymbol{\theta}} \in \Theta} d(\tilde{\boldsymbol{\theta}} \mid D)$

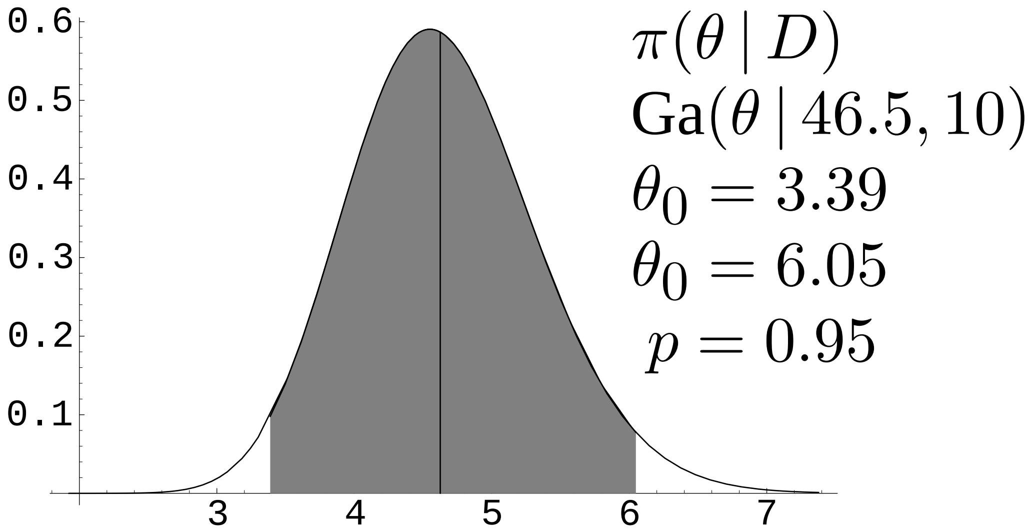
 el estimador intrínseco de referencia ($\forall$ dimension k).

- *Estimación por regiones*
  - Las regiones creibles de confianza objetivas C_p verifican $\int_{C_p} \pi(\boldsymbol{\theta} \mid D)\, d\boldsymbol{\theta} = p$, donde $\pi(\boldsymbol{\theta} \mid D)$ es la posterior de referencia.
  - Regiones intrísecas de confianza. Las regiones de mínima pérdida esperada intrínseca

    $$d(\tilde{\boldsymbol{\theta}}_1 \mid D) < d(\tilde{\boldsymbol{\theta}}_2 \mid D) \quad \forall \boldsymbol{\theta}_1 \in C_p, \quad \forall \boldsymbol{\theta}_2 \notin C_p$$

    proporcionan soluciones únicas invariantes ($\forall$ dimension k).





- *Contrast de hipótesis puntuales*

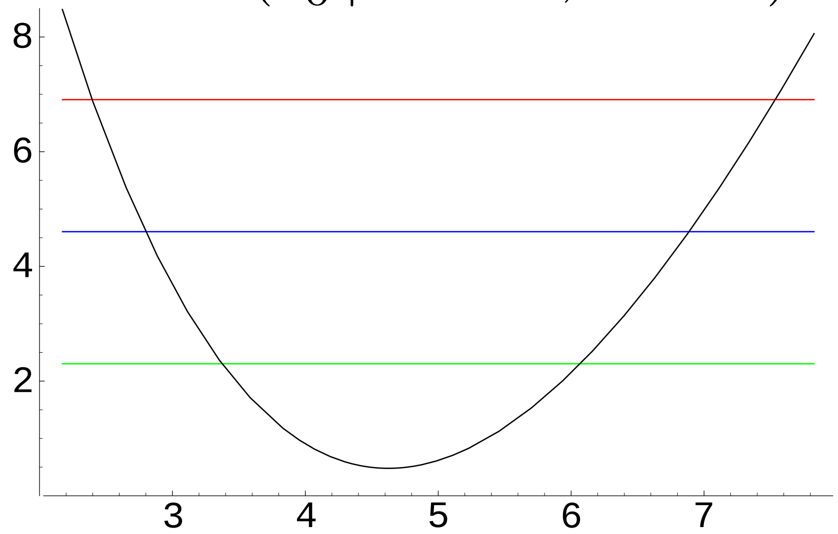
  - Para una función de pérdida, $l(\boldsymbol{\theta}_0, \boldsymbol{\theta})$,
    (i) calcular estadístico de contraste $t(D \,|\, \boldsymbol{\theta}_0) = \int l(\boldsymbol{\theta}_0, \boldsymbol{\theta})\, \pi(\boldsymbol{\theta} \,|\, D)\, d\boldsymbol{\theta}$
    (ii) rechazar $H_0 \equiv \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$ si $t(D \,|\, \boldsymbol{\theta}_0) > t^*$,
    donde t^* mide el valor de aceptar H_0 si es cierta.

  - La función de pérdida intrínseca, con estadístico de constraste
    $d(\boldsymbol{\theta}_0 \,|\, D) = \int \delta(\boldsymbol{\theta}_0, \boldsymbol{\theta})\, \pi(\boldsymbol{\theta} \,|\, D)\, d\boldsymbol{\theta}$
    permite calibrar el umbral d^* en términos de unidades de información:
    log del cociente de verosimilitudes medio a favor del modelo correcto
    (criterio bayesiano de referencia, BRC)

  - $\mathrm{E}[d(D \,|\, \boldsymbol{\theta}_0) \,|\, H_0] = 1.$
    $d^* = \log(10) = 2.3$ (suave)
    $d^* = \log(100) = 4.6$ (fuerte)
    $d^* = \log(1000) = 6.9$ (muy fuerte)

    Los valores $d^* = 2.5$ y $d^* = 5.0$
    corresponden a 2 y 3 desv. típicas
    en el contraste de la media normal.

$d(\theta_0 \,|\, n = 10, r = 46)$

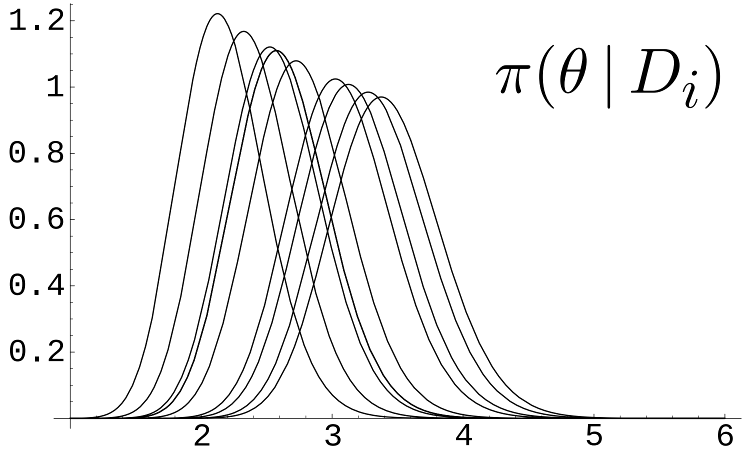


5. Calibrado de los Métodos Estadísticos

- *Simulación*
  - Determianr el q-cuantil $\boldsymbol{\theta}_q(D)$ de la distribución final de referencia
    $$\Pr[\boldsymbol{\theta} < \boldsymbol{\theta}_q \mid D] = \int_{\{\boldsymbol{\theta} < \boldsymbol{\theta}_q\}} \pi(\boldsymbol{\theta} \mid D)\, d\boldsymbol{\theta} = q$$
  - Generar k muestras simuladas de tamaño n,
    $\{D_1, \ldots, D_k\}$ del modelo analizado $p(D \mid \boldsymbol{\theta})$, para algún valor $\boldsymbol{\theta} = \boldsymbol{\theta}_t$.

    Calcular de las correspondientes distribuciones finales.

    Determinar la proporción q_{obs} de las k regiones $]-\infty, \boldsymbol{\theta}_q(D_i)]$
    que contienen a $\boldsymbol{\theta}_t$.
    Para k grande, $q_{obs} \approx q$.
  - $p(x \mid \theta) = \mathrm{Po}(x \mid \theta)$
    $\theta_t = 3$, $n = 20$
    con $k = 1000$ simul,
    $q = 0.90$, $q_{obs} = 0.91$



- *Distribuciones en el muestreo*

  - A partir de $p(D \,|\, \boldsymbol{\theta})$, $\boldsymbol{t} = \boldsymbol{t}(D)$, determinación de $p(\boldsymbol{t} \,|\, \boldsymbol{\theta})$.

    Métodos análiticos y cálculo por simulación.

- *Comportamiento esperado en muestreo repetido*

  - $D = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. Calibrado de regiones creibles en términos de un estadístico suficiente $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$.

    $\Pr[\boldsymbol{\theta} \leq \boldsymbol{\theta}_q(\boldsymbol{t}) \,|\, \boldsymbol{t}] = \int_{\boldsymbol{\theta} \leq \boldsymbol{\theta}_q(\boldsymbol{t})} \pi(\boldsymbol{\theta} \,|\, \boldsymbol{t}) \, d\boldsymbol{\theta} = q$

    $\Pr[\boldsymbol{\theta}_q(\boldsymbol{t}) \geq \boldsymbol{\theta} \,|\, \boldsymbol{\theta}] = \int_{\boldsymbol{\theta}_q(\boldsymbol{t}) \geq \boldsymbol{\theta}} p(\boldsymbol{t} \,|\, \boldsymbol{\theta}) \, d\boldsymbol{t} = \hat{q}_n$

  - Para toda distribución inicial $p(\boldsymbol{\theta})$, $\quad \hat{q}_n = q + O(\frac{1}{\sqrt{n}})$

    Para la inicial de referencia $\pi(\boldsymbol{\theta})$, $\quad \hat{q}_n = q + O(\frac{1}{n})$

- *Interpretación de las regiones creibles*

  - Cuando el estadístico suficiente $\boldsymbol{t} = \boldsymbol{t}(D)$ es un pivote, (su distribución exacta en el muestreo sólo depende de $\boldsymbol{\theta}$), frecuente doble interpretación *exacta* de las regiones $C_q(\boldsymbol{t}) \subset \Theta$:

    (i) Región creible. Subconjunto de Θ al que, dados los datos obtenidos $\boldsymbol{t}$, pertenece el verdadero valor $\boldsymbol{\theta}$ del parámetro con grado racional de creencia q (en una escala $[0, 1]$): $\int_{C_q(\boldsymbol{t})} \pi(\boldsymbol{\theta} \,|\, \boldsymbol{t})\, d\boldsymbol{\theta} = q$

    (ii) Región de confianza. Construcción de regiones $\{C_q(\boldsymbol{t}), \boldsymbol{t} \in \mathcal{T}\}$ que, en muestreo repetido de $p(\boldsymbol{t} \,|\, \boldsymbol{\theta})$, contienen al verdadero valor de $\boldsymbol{\theta}$ en una proporción $\hat{q}$ que tiende a q cuando el número de repeticiones tiende a infinito: $\int_{\{\boldsymbol{t};\, \boldsymbol{\theta} \in C_q(\boldsymbol{t})\}} p(\boldsymbol{t} \,|\, \boldsymbol{\theta})\, d\boldsymbol{t} = q$.

    Ejemplos: datos normales (t de Student), datos exponenciales,...

  - En general, se trata sólo de una aproximación asintótica ($\hat{q}_n \approx q$) con serios problemas en el caso de muestras extremas.

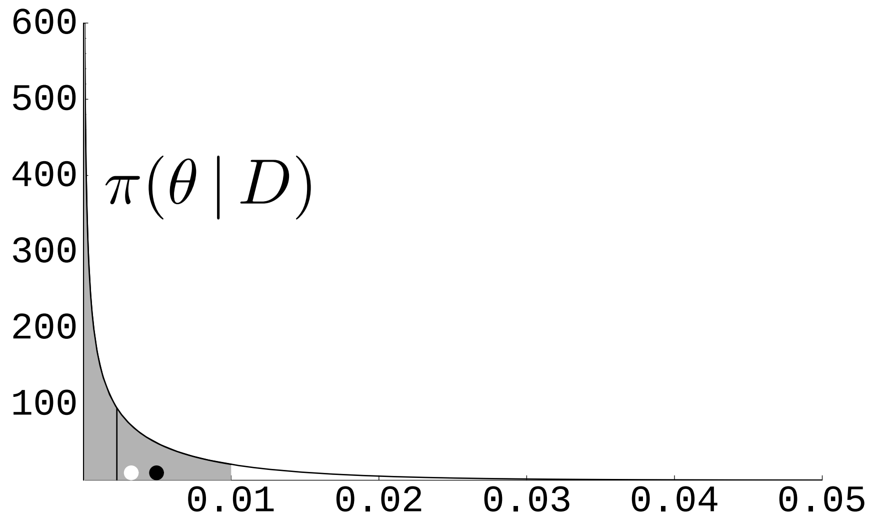
    Ejemplos: datos Binomiales, datos Poisson, datos Student

- *Datos binomiales extremos*

  - $p(x \mid \theta) = \theta^x (1-\theta)^{1-x}, x \in \{0,1\}, \quad \pi(\theta) = \mathrm{Be}(\theta \mid \frac{1}{2}, \frac{1}{2})$

    $D = (r, n), \quad \pi(\theta \mid D) = \mathrm{Be}(\theta \mid r + \frac{1}{2}, n - r + \frac{1}{2}) \quad n = 100, r = 0$

$\pi(\theta \mid D) = \mathrm{Be}(\theta \mid 0.5, 100.5)$
$\Pr(\theta > 0.05 \mid D) \approx 0.001$
$\Pr(\theta < 0.01 \mid D) \approx 0.844$
$\Pr(\theta < 0.0023 \mid D) \approx 0.5$
$\Pr(x = 1 \mid D) = \mathrm{E}[\theta \mid D] \approx 0.0050$
$\theta^*(D) \approx 0.0032$



  - Con muestras extremas $r = 0$, ó $r = n$, el estimador máximo verosimil $\hat{\theta} = r/n$ es totalmente inapropiado, $\hat{\theta} = 0$, ó $\hat{\theta} = 1$.
  - El estimador insesgado de $\psi = \theta^2$, $\tilde{\psi} = r(r-1)/\{n(n-1)\}$, puede producir resultados absurdos: si $r = 1$, $\tilde{\psi} = 0$
  - El buen comportamiento “en valor medio” no garantiza un buen comportamiento para datos concretos.

6. Algunas Conclusiones

- **Inevitabilidad** *lógica de la metodología Bayesiana*
  - Fundamentos basados en la teoría de la decisión.
  - Existencia de la inicial basada en la teoría de la probabilidad.
- *Resultados* **objetivos**, *directamente relevantes*
  - El análisis de referencia proporciona soluciones que sólo dependen del modelo aceptado y de los datos obtenidos.
  - La distribución final (en particular sus regiones de confianza) corresponden directamente a las preguntas formuladas por científicos y técnicos.
  - Con cualquier modelo, se obtienen resultados válidos para cualquier tamaño muestral y para cualquier muestra (el analisis es siempre condicional a los datos obtenidos, no a su comportamiento medio)
- *Facilidad de interpretación*
  - Graficas: distribucion final, estimatores, regiones creíbles, contrastes.

- *Reformulación de la enseñanza de la estadística matemática*
  - Elementos de teoría de decisión.
  - Intercambiabilidad, modelos y verosimilitud.
  - Métodos bayesianos objetivos.
  - Calibrado de los métodos estadísticos
- *Referencias básicas*

Libros

Bernardo, J. M. & Smith, A. F. M. (1994).
Bayesian Theory. Chichester: Wiley.

DeGroot, M. H. & Shervish, M. J. (2002).
Probability and Statistics. Reading, MA: Addison-Wesley.

Berger, J. O., Bernardo, J. M. & Sun, D. (2003).
Objective Bayesian Statistics (en preparación).

Artículos

On line: **http://www.uv.es/˜bernardo/publications.html**

Generales

Bernardo, J. M. (2001). Un programa de síntesis para la enseñanza de la estadística matemática contemporánea. *Rev. Real Acad. Ciencias* **95**, 87–105.

Bernardo, J. M. (2003). Bayesian Statistics. *Encyclopedia of Life Support Systems (EOLSS)*. Paris: UNESCO (en prensa).

Específicos

Bernardo, J. M. & Juárez, M. A. (2003). Intrinsic Estimation. *Bayesian Statistics 7*. Oxford: University Press, 465–475.

Bernardo, J. M. & Ramón, J. M. (1998). An introduction to Bayesian reference analysis. *J. Roy. Statist. Soc. D* **47**, 1–35.

Bernardo, J. M. & Rueda, R. (2002). Bayesian hypothesis testing: A reference approach. *Internat. Statist. Rev.* **70**, 351–372.

4th International Meeting on

OBJECTIVE BAYESIAN STATISTICS

Aussois, (Alpes Franceses)
15 – 20 Junio, 2003

www.ceremade.dauphine.fr/ xian/OBayes03.html
Prof. Christian P. Robert, <xian@ceremade.dauphine.fr>