CAPÍTULO 7. ESTIMACIÓN Y CONTRASTES

Para leer

Lee, Capítulo 4

Estimación puntual

Para los bayesianos, el problema de estimación es un problema de decisión. Asociada con cada estimador T hay una pérdida $L(T,\theta)$ que refleja la diferencia entre θ y T. Por ejemplo:

- $L(T,\theta) = (T-\theta)^2$, la pérdida cuadrática

- $L(T,\theta) = |T-\theta|$, la pérdida lineal absoluta

- $L(T,\theta) = \begin{cases} 0 & \text{si } T = \theta \\ 1 & \text{si } T \neq \theta \end{cases}$, la pérdida "todo o nada".

Definición 9 *El estimador Bayes* T^B *es una solución de*

$$T^B = \min_T E[L(T,\theta)]$$

Ejemplo 48 *Dada la pérdida cuadrática, ¿cuál es el estimador Bayes?*

$$\begin{aligned}
E[L(T,\theta)] &= \int (T-\theta)^2 f(\theta)\, d\theta \\
&= \int \left(T - E[\theta] + E[\theta] - \theta\right)^2 f(\theta)\, d\theta \\
&= \int \{(T-E[\theta])^2 + (E[\theta]-\theta)^2\} f(\theta)\, d\theta \\
&= V[\theta] + (T - E[\theta])^2
\end{aligned}$$

y entonces $T^B = E[\theta]$ *es el estimador Bayes.*

Ejemplo 49 *Con la pérdida lineal absoluta, tenemos*

$$\begin{aligned}
E[L(T,\theta)] &= \int |T-\theta| f(\theta)\, d\theta \\
&= \int_{-\infty}^{T} (T-\theta) f(\theta)\, d\theta + \\
&\quad + \int_{T}^{\infty} (\theta - T) f(\theta)\, d\theta \\
\frac{d}{dT} E[L(T|\theta)] &= (T-T)f(T) + \int_{-\infty}^{T} f(\theta)\, d\theta - \\
&\quad -(T-T)f(T) - \int_{T}^{\infty} f(\theta)\, d\theta \\
&= F(T) - (1 - F(T)) \\
&= 2F(T) - 1
\end{aligned}$$

Entonces, recordando que en el mínimo la derivada es igual a cero, tenemos $F(T^B) = 1/2$ *y el estimador Bayes es la mediana de la distribución de* θ.

Ejemplo 50 *Suponiendo que* θ *es discreta, con la pérdida "todo o nada" se tiene*

$$\begin{aligned} E[L(T,\theta)] &= \sum_{\theta \neq T} P(\theta) \\ &= P(\theta \neq T) \end{aligned}$$

y se minimiza la pérdida esperada eligiendo el estimador Bayes T^B *como la moda de la distribución de* θ*.*

Observación 21 *Esta pérdida no se puede utilizar con variables continuas porque* $P(\theta = T) = 0$ *si* θ *es continua y entonces, la pérdida esperada será 1 para cualquier elección de* T*.*

Intervalos

Se han visto intervalos de credibilidad anteriormente. Sigue la definición formal.

Definición 10 *Si $f(\theta|\mathbf{x})$ es una densidad a posteriori, se dice que (a,b) es* **un intervalo de credibilidad** *de* $100 \times (1-\alpha)\,\%$ *si*

$$P(a \leq \theta \leq b|\mathbf{x}) = 1 - \alpha$$

Ejemplo 51 $X|\mu \sim \mathcal{N}(\mu, 1)$. *Supongamos que $f(\mu) \propto 1$, entonces, $\mu|\mathbf{x} \sim \mathcal{N}(\bar{x}, 1/n)$ y algunos intervalos de credibilidad de 95 % son*

$$(-\infty, \bar{x} + 1{,}64/\sqrt{n}) \quad o \quad (\bar{x} - 1{,}64/\sqrt{n}, \infty) \quad o$$
$$(\bar{x} \pm 1{,}96/\sqrt{n})$$

Hay muchos intervalos de credibilidad. El más corto se llama **un intervalo de máxima densidad a posteriori** (MDP)

Definición 11 *El intervalo MDP de* $100 \times (1-\alpha)\,\%$ *es el intervalo de forma*

$$C = \{\theta : f(\theta) \geq c(\alpha)\}$$

donde $c(\cdot)$ *es la constante más grande cumpliendo*

$$P(C) \geq 1 - \alpha$$

Ejemplo 52 *Volviendo al ejemplo 51, el intervalo MDP de 95 % es*

$$\bar{x} \pm 1{,}96/\sqrt{n}$$

Se puede aplicar la definición de un intervalo de credibilidad a densidades multivariantes $f(\boldsymbol{\theta}|\mathbf{x})$. En estas situaciones, se puede definir una región de credibilidad $\mathbf{C}$:

$$P(\boldsymbol{\theta} \in \mathbf{C}|\mathbf{x}) = 1 - \alpha.$$

Contrastes

Consideramos las hipótesis $\begin{array}{l} H_0 : \boldsymbol{\theta} \in \Theta_0 \\ H_1 : \boldsymbol{\theta} \in \Theta_1 \end{array}$, donde $\Theta_0 \cup \Theta_1 = \Theta$ y $\Theta_0 \cap \Theta_1 = \phi$.

Teóricamente es fácil distinguir entre las dos hipótesis; dados los datos, sólo se deben usar las probabilidades a posteriori. Dada una función de pérdida, se elige aceptar o rechazar H_0.

Ejemplo 53 *Dada la pérdida "todo o nada",*

$$L(H_0, \theta) = \begin{cases} 0 & \textit{si } H_0 \textit{ es verdadero} \\ 1 & \textit{si } H_1 \textit{ es verdadero} \end{cases}$$

elegimos H_0 si $P(H_0|\mathbf{x}) > P(H_1|\mathbf{x})$.

Ejemplo 54 *Supongamos que* $X|\theta \sim N(\theta, 1)$. *Queremos hacer el contraste:* $H_0 : \theta \leq 0$ *frente* $H_1 : \theta > 0$. *Si usamos una distribución inicial no informativa para* θ,

$$f(\theta) \propto 1,$$

tenemos $\theta|\mathbf{x} \sim \mathcal{N}\left(\bar{x}, \frac{1}{n}\right)$. *Entonces,*

$$\begin{aligned} P(H_0|\mathbf{x}) &= P(\theta \leq 0|\mathbf{x}) \\ &= P\left(\sqrt{n}(\theta - \bar{x}) \leq -\sqrt{n}\bar{x}|\mathbf{x}\right) \\ &= \Phi\left(-\sqrt{n}\bar{x}\right) \end{aligned}$$

donde $\Phi(\cdot)$ *es la función de distribución normal.*

Observación 22 *Esta probabilidad es igual al p valor clásico para el contraste* $H_0^c : \theta = 0$ *frente* $H_1 : \theta > 0$.

$$\begin{aligned} P(\bar{X} \geq \bar{x}|H_0) &= P\left(\sqrt{n}\bar{X} \geq \sqrt{n}\bar{x}|H_0\right) \\ &= 1 - \Phi\left(\sqrt{n}\bar{x}\right) \\ &= \Phi\left(-\sqrt{n}\bar{x}\right) \end{aligned}$$

La paradoja de Lindley/Jeffreys

Consideramos el contraste $H_0 : \theta = \theta_0$ frente la alternativa $H_1 : \theta \neq \theta_0$. En situaciones así, los resultados bayesianos pueden ser muy diferentes de los resultados clásicos.

Ejemplo 55 $X|\theta \sim \mathcal{N}(\theta, 1)$. *Hacemos el contraste* $H_0 : \theta = 0$ *frente* $H_1 : \theta \neq 0$.

Se definen las probabilidades a priori

$$f_0 = P(H_0) = 0{,}5 = P(H_1) = f_1$$

y se supone que $\theta|H_1 \sim \mathcal{N}(0, 1)$.

Suponiendo que se observa la media de una muestra de tamaño n, *se quiere calcular las probabilidades a posteriori.*

En primer lugar

$$\begin{aligned}
\alpha_0 &= P(H_0|\bar{x}) \\
&\propto f_0 f(\bar{x}|\theta = 0) \\
&\propto \frac{1}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{n\bar{x}^2}{2}\right) \\
&= \frac{K}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{n\bar{x}^2}{2}\right)
\end{aligned}$$

para una constante $K = f(\bar{x})$. *También*

$$\begin{aligned}
f(\theta, H_1|\mathbf{x}) &\propto f_1 f(\bar{x}|\theta, H_1) f(\theta|H_1) \\
&\propto \frac{1}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{n(\bar{x}-\theta)^2}{2}\right) \times \\
&\quad \times \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{\theta^2}{2}\right) \\
&= \frac{K}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}} \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \times \\
&\quad \times \exp\left(-\frac{1}{2}\left[n(\bar{x}-\theta)^2 + \theta^2\right]\right)
\end{aligned}$$

donde K *es la misma constante.*

Entonces, se tiene

$$
\begin{aligned}
\alpha_1 &= P(H_1|\mathbf{x}) \\
&= \int f(\theta, H_1|\mathbf{x})\, d\theta \\
&= \int \frac{K}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}}\left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left[n(\bar{x}-\theta)^2+\theta^2\right]\right)\, d\theta \\
&= \frac{K}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}}\left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \times \\
&\quad \int \exp\left(-\frac{1}{2}\left[(n+1)\left(\theta-\frac{n\bar{x}}{n+1}\right)^2-\frac{n\bar{x}^2}{n+1}\right]\right)\, d\theta \\
&= \frac{K}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}}\frac{1}{\sqrt{n+1}}\exp\left(-\frac{n\bar{x}^2}{2(n+1)}\right)
\end{aligned}
$$

Recordando que $\alpha_0+\alpha_1=1$*, se tiene*

$$
K=\left(\frac{1}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{n\bar{x}^2}{2}\right)+\frac{1}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}}\frac{1}{\sqrt{n+1}}\exp\left(-\frac{n\bar{x}^2}{2(n+1)}\right)\right)^{-1}.
$$

Entonces

$$
\begin{aligned}
\alpha_0 &= \frac{\frac{1}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{n\bar{x}^2}{2}\right)}{\frac{1}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{n\bar{x}^2}{2}\right)+\frac{1}{2}\left(\frac{n}{2\pi}\right)^{\frac{1}{2}}\frac{1}{\sqrt{n+1}}\exp\left(-\frac{n\bar{x}^2}{2(n+1)}\right)} \\
&= \frac{\exp\left(-\frac{n\bar{x}^2}{2}\right)}{\exp\left(-\frac{n\bar{x}^2}{2}\right)+\frac{1}{\sqrt{n+1}}\exp\left(-\frac{n\bar{x}^2}{2(n+1)}\right)} \\
&= \frac{\exp\left(-\frac{n(n+1)\bar{x}^2}{2(n+1)}\right)}{\exp\left(-\frac{n(n+1)\bar{x}^2}{2(n+1)}\right)+\frac{1}{\sqrt{n+1}}\exp\left(-\frac{n\bar{x}^2}{2(n+1)}\right)} \\
&= \left\{1+\frac{1}{\sqrt{n+1}}\exp\left(\frac{n^2\bar{x}^2}{2(n+1)}\right)\right\}^{-1}
\end{aligned}
$$

Consideramos el caso $\bar{x} = 2/\sqrt{n} > 1{,}96/\sqrt{n}$. *Sabemos que si hubiéramos hecho un contraste clásico con un nivel de significación de 95 %, el resultado habría sido significativo, y habríamos rechazado la hipótesis* H_0.

$$
\begin{aligned}
\textit{Pero } \alpha_0 &= \left\{1+\frac{1}{\sqrt{n+1}}\exp\left(\frac{n}{n+1}\right)\right\}^{-1} \\
&\to 1 \quad \textit{cuando } n \to \infty.
\end{aligned}
$$

Una muestra que nos llega a rechazar H_0 con un contraste clásico nos proporciona una probabilidad a posteriori de H_0 que se acerca a 1 cuando el tamaño de la muestra es grande.

Esta paradoja se llama la paradoja de Lindley y Jeffreys.

Observación 23 *La elección de la varianza de θ en la distribución inicial es bastante importante pero el ejemplo demuestra que no tiene sentido usar niveles fijos de significación según crece n.*

Hipótesis nulos puntuales son poco razonables.

Factores Bayes

También es útil introducir otro concepto.

Supongamos que $f_0 = P(H_0)$ y $f_1 = P(H_1)$ y que $\alpha_0 = P(H_0|\mathbf{x})$ y $\alpha_1 = P(H_1|\mathbf{x})$.

Definición 12 *Se define*

$$B = \frac{\alpha_0/\alpha_1}{f_0/f_1} = \frac{\alpha_0 f_1}{\alpha_1 f_0}$$

el **factor Bayes** *a favor de H_0.*

Observación 24 *El factor Bayes representa las posibilidades (odds) a posteriori divididos por las posibilidades a priori. Nos informe de los cambios en nuestras creencias causados por los datos.*

Observación 25 *Es casi objetiva y parcialmente elimina la influencia de la distribución a priori.*

Ejemplo 56 *Supongamos el contraste simple* $H_0 : \theta = \theta_0$ *frente* $H_1 : \theta = \theta_1$. *Tenemos*

$$\begin{aligned}
\alpha_0 = P(H_0|\mathbf{x}) &= \frac{f_0 l(\theta_0|\mathbf{x})}{f_0 l(\theta_0|\mathbf{x}) + f_1 l(\theta_1|\mathbf{x})} \\
\alpha_1 = P(H_1|\mathbf{x}) &= \frac{f_1 l(\theta_1|\mathbf{x})}{f_0 l(\theta_0|\mathbf{x}) + f_1 l(\theta_1|\mathbf{x})}
\end{aligned}$$

Entonces el factor Bayes es

$$\begin{aligned}
B &= \frac{\alpha_0}{\alpha_1}\frac{f_1}{f_0} \\
&= \frac{f_0 l(\mathbf{x}|\theta_0)}{f_1 l(\mathbf{x}|\theta_1)}\frac{f_1}{f_0} \\
&= \frac{l(\theta_0|\mathbf{x})}{l(\theta_1|\mathbf{x})}
\end{aligned}$$

que coincide con la razón de verosimilutudes. Entonces, la distribución a priori no influye en el factor Bayes.

Ejemplo 57 *Se observa un dato de una distribución exponencial con densidad*

$$f(x|\lambda) = \lambda e^{-\lambda x}.$$

Se quiere contrastar $H_0 : \lambda = 6$ *frente* $H_1 : \lambda = 3$*. ¿Cuál es el factor Bayes?*

$$\begin{aligned} B &= \frac{l(\lambda = 6|x)}{l(\lambda = 3|x)} \\ &= \frac{6e^{-6x}}{3e^{-3x}} \\ &= 2e^{-3x} \end{aligned}$$

Suponiendo que la probabilidad a priori de H_0 *es* 0,25*, se puede demostrar que* $P(H_0|x) < 0{,}5$ *para cualquier valor de* x*.*

En primer lugar, hallamos el factor Bayes.

$$
\begin{aligned}
B &= \frac{P(H_0|x)}{P(H_1|x)}\frac{P(H_1)}{P(H_0)} \\
&= 3\frac{P(H_0|x)}{P(H_1|x)} \\
\frac{P(H_0|x)}{P(H_1|x)} &= \frac{2}{3}e^{-3x}
\end{aligned}
$$

y $P(H_0|x) \geq 1/2 \Rightarrow \frac{2}{3}e^{-3x} > 1$ *y entonces*

$$x < -\frac{1}{3}\log\frac{3}{2} < 0$$

que es imposible.

Observación 26 *El factor Bayes es consistente. Si* H_0 *es verdadero, entonces* $B \to \infty$ *cuando* $n \to \infty$ *y si* H_1 *es verdadero,* $B \to 0$ *cuando* $n \to \infty$.

El factor Bayes no elimina totalmente la influencia de la distribución a priori. Supongamos que H_0 y H_1 son compuestos y entonces

$$\begin{aligned} B &= \frac{P(H_0|\mathbf{x})}{P(H_1|\mathbf{x})}\frac{P(H_1)}{P(H_0)} \\ &= \frac{f(\mathbf{x}|H_0)}{f(\mathbf{x}|H_1)} \\ &= \frac{\int f(\mathbf{x}|H_0,\theta_0)f(\theta_0|H_0)\, d\theta_0}{\int f(\mathbf{x}|H_1,\theta_1)f(\theta_1|H_1)\, d\theta_1} \end{aligned}$$

donde $f(\theta_0|H_0)$ es la distribución a priori bajo la hipótesis H_0 y $f(\theta_1|H_1)$ es la distribución a priori bajo H_1.

Ejemplo 58 *Supongamos que $X|\lambda \sim \mathcal{E}(\lambda)$ como en el Ejemplo 57. Ahora se quiere contrastar $H_0 : \lambda = 6$ frente a $H_1 : \lambda \neq 6$. Sea la distribución a priori $\lambda|H_1 \sim \mathcal{E}(1/6)$.*

Suponiendo que se observa un dato x como anteriormente, se tiene

$$f(x|H_0) = 6e^{-6x}$$

y

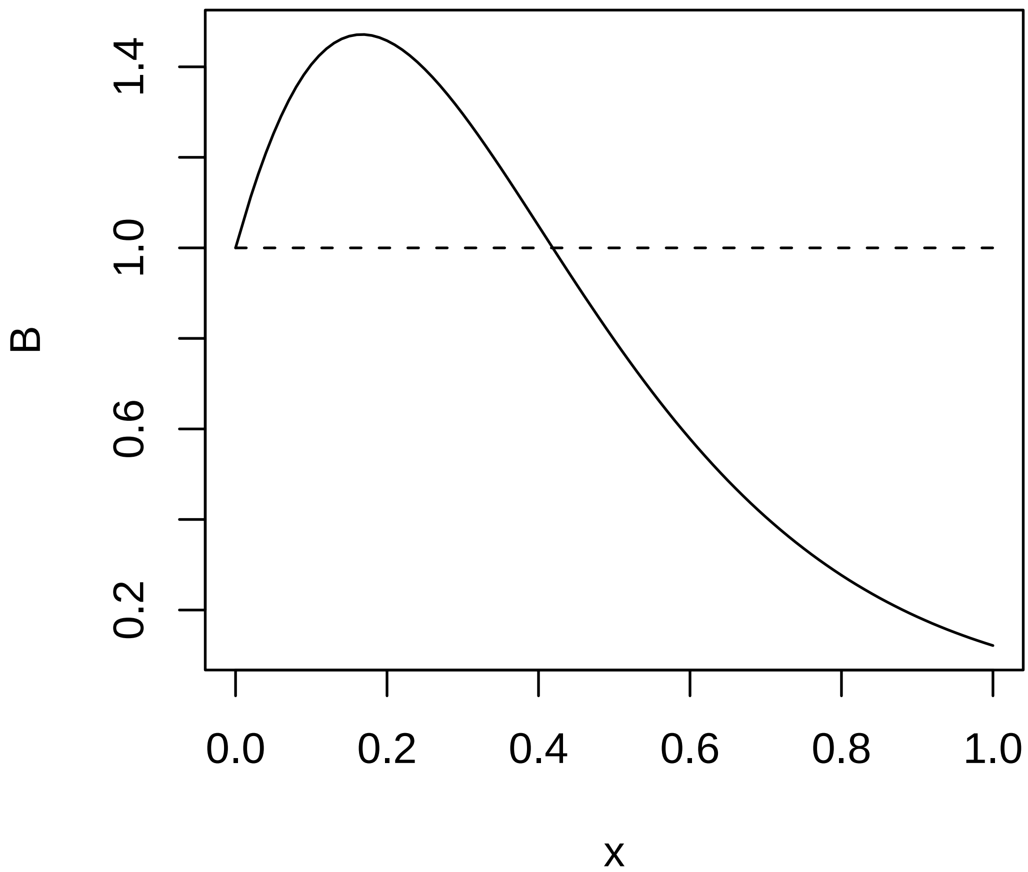
$$\begin{aligned} f(x|H_1) &= \int f(x|H_1, \lambda) f(\lambda|H_1)\, d\lambda \\ &= \int \lambda e^{-\lambda x} \frac{1}{6} e^{-\frac{1}{6}\lambda}\, d\lambda \\ &= \frac{1}{6} \int \lambda e^{-\left(x+\frac{1}{6}\right)\lambda}\, d\lambda \\ &= \frac{1}{6} \int \lambda^{2-1} e^{-\left(x+\frac{1}{6}\right)\lambda}\, d\lambda \\ &= \frac{1}{6} \frac{\Gamma(2)}{\left(x + \frac{1}{6}\right)^2} \\ &= \frac{6}{(6x+1)^2} \end{aligned}$$

Entonces el factor Bayes es

$$B = (6x + 1)^2 e^{-6x}.$$

Supongamos ahora que las probabilidades a priori son $P(H_0) = P(H_1) = 0{,}5$*. ¿Para cuáles valores de* x *es más probable* H_0 *a posteriori?*

La probabilidad a posteriori de H_0 *es mayor que* 0,5 *si* $B > 1$*. Construimos un gráfico del valor de* B *frente a* x*.*



El modelo H_0 es más probable a posteriori si $x <$ 0,4188 a 4 plazas decimales.

¿Cuál es el máximo valor posible de $P(H_0|x)$?

La probabilidad de H_0 es máxima cuando el factor Bayes es lo más grande posible. Calculamos el máximo del factor Bayes como función de x.

$$
\begin{aligned}
B &= (6x+1)^2 e^{-6x} \\
\log B &= 2\log(6x+1) - 6x \\
\frac{d}{dx}\log B &= \frac{2}{6x+1} - 6 \\
0 &= \frac{2}{6\hat{x}+1} - 6 \\
36\hat{x} &= 8 \\
\hat{x} &= \frac{2}{9}
\end{aligned}
$$

y en este caso, el valor de B es

$$\hat{B} = \left(6 \times \frac{2}{9} + 1\right)^2 e^{-6\times\frac{2}{9}} = 1{,}43514$$

Recordamos que

$$
\begin{aligned}
\widehat{B} &= \frac{P(H_0|\widehat{x})}{P(H_1|\widehat{x})}\frac{P(H_1)}{P(H_0)} \\
1{,}43514 &= \frac{P(H_0|\widehat{x})}{1-P(H_0|\widehat{x})} \\
P(H_0|x) &= \frac{1{,}43514}{1+1{,}43514} \\
&\approx 0{,}5893
\end{aligned}
$$

es el máximo valor posible de la probabilidad a posteriori.

Problemas y Generalizaciones

Si usamos distribuciones a priori impropias para los parámetros, puede que el factor Bayes no exista.

Volviendo a la situación de la transparencia , supongamos que $f(\theta_0|H_0)$ y $f(\theta_1|H_1)$ son impropias, por ejemplo

$$f(\theta_i|H_i) = c_i g_i(\theta_i)$$

para algunas constantes c_i indefinidas.

Luego

$$\begin{aligned} B &= \frac{\int f(\mathbf{x}|H_0,\theta_0)f(\theta_0|H_0)\,d\theta_0}{\int f(\mathbf{x}|H_1,\theta_1)f(\theta_1|H_1)\,d\theta_1} \\ &= \frac{c_1}{c_2}\frac{\int f(\mathbf{x}|H_0,\theta_0)g_0(\theta_0)\,d\theta_0}{\int f(\mathbf{x}|H_1,\theta_1)g_1(\theta_1)\,d\theta_1} \end{aligned}$$

que depende de la razón de las constantes indefinidas.

Hay algunas alternativas

- factores Bayes fraccionales (O'Hagan, A. Bayesian Inference, Edward Arnold, 1995)

- factores Bayes intrínsicos (Berger J. y Perrichi L. The Intrinsic Bayes Factor for linear models. En Bayesian Statistics V, eds Bernardo et al, O.U.P., 23 – 42.)

Los dos métodos utilizan partes de los datos para crear una distribución inicial propia.