Inferencia Bayesiana

Manuel Mendoza R.

Departamento de Estadística
Instituto Tecnológico Autónomo de México

IV Escuela de Verano. Centro de Estadística Aplicada a Estudios Socioeconómicos.
Medellín, Colombia. Diciembre 5-7, 2011.

Contenido

- Proceso de aprendizaje
- Familias conjugadas
- Iniciales de referencia
- Problemas de Inferencia

Contenido

- Proceso de aprendizaje
- Familias conjugadas
- Iniciales de referencia
- Problemas de Inferencia

Análisis Bayesiano Paramétrico

- *Objetivo*: Describir la variable aleatoria X, con soporte $\mathscr{X}$ y función de probabilidad P(X | θ) totalmente conocida, excepto por valor del parámetro fijo, de dimensión finita, θ.

✓ Se cuenta con una muestra de observaciones $\underline{X}_{(n)}$ con función de probabilidad conjunta P($\underline{X}_{(n)}$ | θ).

✓ Antes de los datos $\underline{X}_{(n)}$, la información sobre θ se describe con la probabilidad inicial (*a priori*) P(θ).

Análisis Bayesiano Paramétrico

- ✓ Desde la perspectiva Bayesiana, el fenómeno bajo estudio se aborda con el *modelo conjunto* $P(\underline{X}_{(n)}, \theta)$ que describe la información disponible, tanto sobre los datos $\underline{X}_{(n)}$, como sobre el parámetro θ.

- ✓ La representación habitual de este modelo:

$$\boxed{P(\underline{X}_{(n)}, \theta) = P(\underline{X}_{(n)} \mid \theta)\, P(\theta)}$$

Análisis Bayesiano Paramétrico

- ✓ Las interpretaciones de la probabilidad en el modelo de muestreo

$$P(\underline{X}_{(n)} \mid \theta)$$

y en la distribución inicial

$$P(\theta)$$

son *distintas*.

En el primer caso describen *variabilidad* mientras que en segundo describen *incertidumbre*.

Análisis Bayesiano Paramétrico

- Las inferencias sobre el parámetro, una vez observados e incorporados los datos de la muestra, se realizan a partir de la distribución final (a posteriori) $P(\theta \mid \underline{X}_{(n)})$

$$\boxed{P(\theta \mid \underline{X}_{(n)}) \propto P(\underline{X}_{(n)} \mid \theta)\, P(\theta)}$$

Teorema de Bayes

Análisis Bayesiano Paramétrico

- ✓ De hecho, el Teorema de Bayes, o Fórmula de Bayes, establece que la distribución final (a posteriori) $P(\theta \mid \underline{X}_{(n)})$ satisface

$$\boxed{P(\theta \mid \underline{X}_{(n)}) = \frac{P(\underline{X}_{(n)} \mid \theta)\ P(\theta)}{P(\underline{X}_{(n)})}}$$

$P(\underline{X}_{(n)} \mid \theta)$ función de Verosimilitud

$P(\theta)$ distribución Inicial

$P(\underline{X}_{(n)})$ marginal de la muestra

Análisis Bayesiano Paramétrico

- ✓ Como función del parámetro θ,

$$\boxed{P(\theta \mid \underline{X}_{(n)}) \propto P(\underline{X}_{(n)} \mid \theta)\ P(\theta)}$$

Análisis Bayesiano Paramétrico

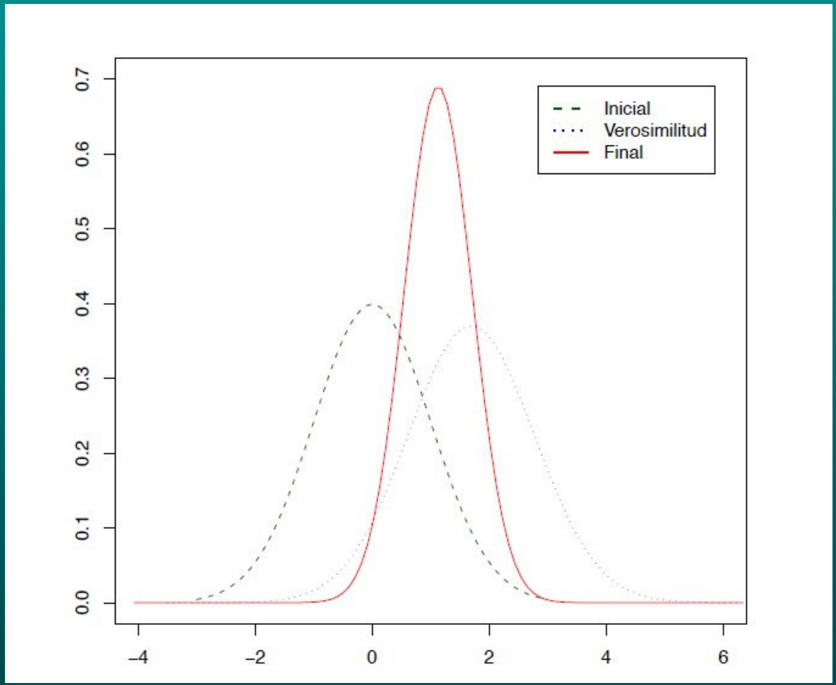


Diagrama Triplot

Análisis Bayesiano Paramétrico

- *Ejemplo 1a.* X variable aleatoria Normal con media μ y varianza σ^2 conocida (precisión $\tau = 1/\sigma^2$ conocida).

$$P(X \mid \theta) = N(X \mid \mu, \tau); \qquad \theta = \mu \qquad (\mu \in \Re)$$

- Si la inicial para μ es $P(\mu) = \text{Exp} (\mu \mid \lambda) \quad (\mu > 0)$

$$\boxed{P(\mu \mid \underline{X}_{(n)}) \propto P(\underline{X}_{(n)} \mid \mu) \; P(\mu) \qquad (\mu > 0)}$$

$\underline{X}_{(n)}$ muestra aleatoria de X.

Análisis Bayesiano Paramétrico

- *Ejemplo 1a.* (*continuación*)

$$P(\mu \mid \underline{X}_{(n)}) \propto P(\underline{X}_{(n)} \mid \mu)\ P(\mu) \qquad (\mu > 0)$$

$$= N(\mu \mid \overline{X}, n\tau)\ \mathrm{Exp}(\mu \mid \lambda) \qquad (\mu > 0)$$

Análisis Bayesiano Paramétrico

- *Ejemplo 1a. (continuación)*

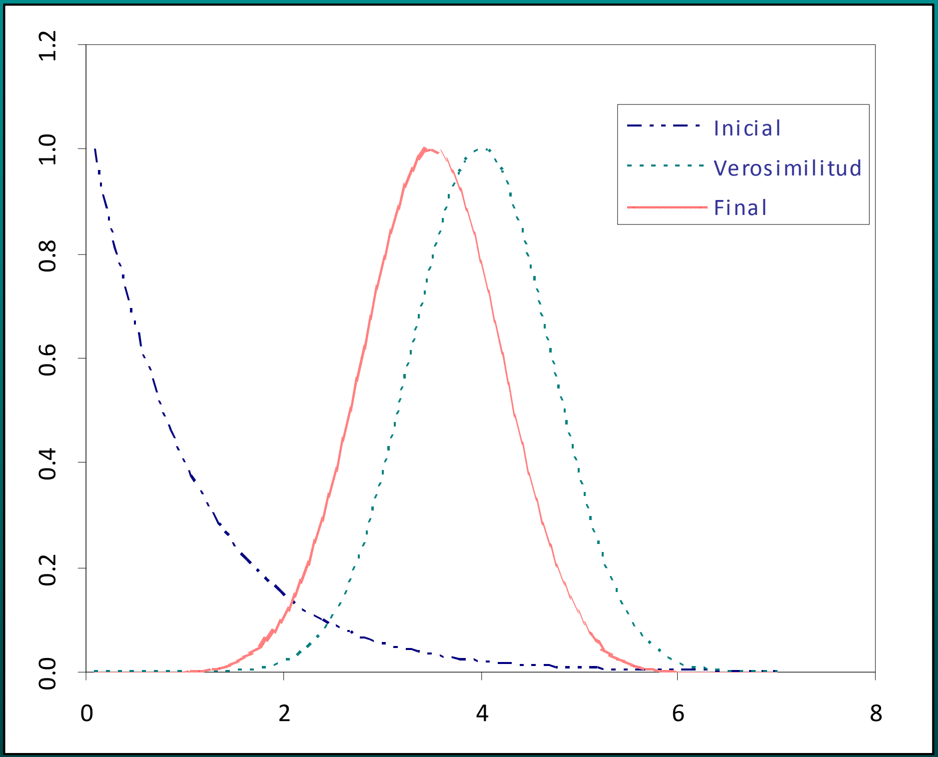
$\Rightarrow$

$$P(\mu \mid \underline{X}_{(n)}) \propto \begin{cases} \exp[-(n\tau/2)(\mu - m^*)^2] & \boxed{(\mu > 0)} \\ 0 & (\mu \leq 0) \end{cases}$$

$$\boxed{m^* = \overline{X} - \frac{\lambda}{n\,\tau}}$$

Distribución Normal Truncada

Análisis Bayesiano Paramétrico



Distribución Normal Truncada

Análisis Bayesiano Paramétrico

- *Ejemplo 1b.* X variable aleatoria Normal con media μ y varianza σ^2 conocida (precisión $\tau = 1/\sigma^2$).

$$P(X \mid \theta) = N(X \mid \mu, \tau); \qquad \theta = \mu \qquad (\mu \in \Re)$$

- Si la inicial para μ es $P(\mu) = N (\mu \mid m, \tau_0) \quad (\mu \in \Re)$

$$\boxed{P(\mu \mid \underline{X}_{(n)}) \propto P(\underline{X}_{(n)} \mid \mu) \; P(\mu)} \qquad (\mu \in \Re)$$

Análisis Bayesiano Paramétrico

- *Ejemplo 1b. (continuación)*

$$P(\mu \mid \underline{X}_{(n)}) \propto P(\underline{X}_{(n)} \mid \mu)\; P(\mu) \qquad (\mu > 0)$$

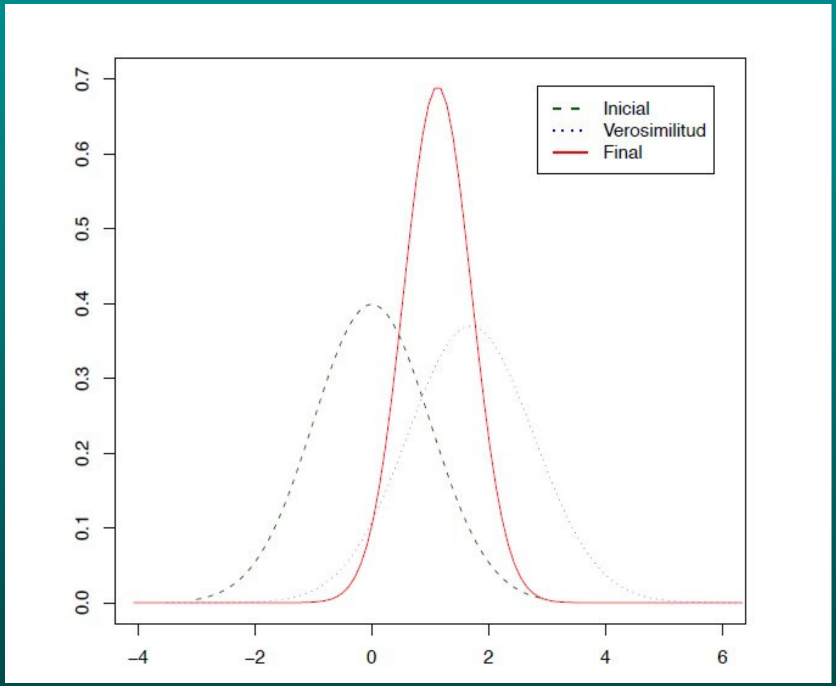
$\Rightarrow$

$$\boxed{P(\mu \mid \underline{X}_{(n)}) = N(\mu \mid m_X, \tau_X) \qquad (\mu > 0)}$$

$$m_X = \frac{(n\,\tau)\,\overline{X} + (\tau_0)\,m_0}{n\,\tau + \tau_0} \qquad \tau_X = n\,\tau + \tau_0$$

Distribución Normal

Análisis Bayesiano Paramétrico



Distribución Normal

Análisis Bayesiano Paramétrico

$$P(\underline{X}_{(n)} \mid \mu)$$

$$P(\mu) \ \text{Normal} \quad \Rightarrow \quad P(\mu \mid \underline{X}_{(n)}) \ \text{Normal}$$

Distribuciones Conjugadas

$$P(\underline{X}_{(n)} \mid \theta)$$

$$P(\theta) \in \mathscr{F}_{\phi} \quad \Rightarrow \quad P(\theta \mid \underline{X}_{(n)}) \in \mathscr{F}_{\phi}$$

Análisis Bayesiano Paramétrico

- *Ejemplo 2.* X variable aleatoria Normal con media μ y varianza σ^2 desconocidas las dos (precisión $\tau = 1/\sigma^2$).

$$P(X \mid \theta) = N(X \mid \mu, \tau); \qquad \theta = (\mu, \tau)$$

- La inicial para θ es

$$\begin{aligned} P(\theta) &= P(\mu, \tau) \\ &= P(\mu \mid \tau)\, P(\tau) \\ &= N (\mu \mid m, c\tau)\, \text{Gamma} (\tau \mid \alpha, \beta) \end{aligned}$$

Distribución *Normal-Gamma*($\mu, \tau \mid m, c, \alpha, \beta$)

Análisis Bayesiano Paramétrico

$$P(\theta \mid \underline{X}_{(n)}) \propto P(\underline{X}_{(n)} \mid \theta)\ P(\theta)$$

$$\begin{aligned} P(\theta \mid \underline{X}_{(n)}) &= P(\mu, \tau \mid \underline{X}_{(n)}) \\ &= P(\mu \mid \tau, \underline{X}_{(n)})\ P(\tau \mid \underline{X}_{(n)}) \\ &= N(\mu \mid m_X, c_X\, \tau)\ \text{Gamma}(\tau \mid \alpha_X, \beta_X) \end{aligned}$$

Distribuciones Conjugadas

$$(m, c, \alpha, \beta) \longrightarrow (m_X, c_X, \alpha_X, \beta_X)$$

Análisis Bayesiano Paramétrico

$$P(\tau \mid \underline{X}_{(n)}) = \text{Gamma}(\tau \mid \alpha_X, \beta_X)$$

$$P(\mu \mid \underline{X}_{(n)}) = \int P(\mu, \tau \mid \underline{X}_{(n)})\, d\tau$$

$$= \int N(\mu \mid m_X, \tau_X)\, \text{Gamma}(\tau \mid \alpha_X, \beta_X)\, d\tau$$

$$\boxed{P(\mu \mid \underline{X}_{(n)}) = \text{Stu}(\mu \mid m_X, \gamma_X, n\text{-}1)}$$

Análisis Bayesiano Paramétrico

- *Ejemplo 3.* X variable aleatoria Poisson con media λ.

$$P(X \mid \theta) = \text{Poisson}(X \mid \lambda); \qquad \theta = \lambda$$

- Si la inicial para λ es $P(\lambda) = \text{Gamma} (\lambda \mid \alpha, \beta)$

$$\boxed{P(\lambda \mid \underline{X}_{(n)}) \propto P(\underline{X}_{(n)} \mid \lambda) \; P(\lambda)}$$

Análisis Bayesiano Paramétrico

$\Rightarrow$ $\boxed{P(\lambda \mid \underline{X}_{(n)}) = \text{Gamma}(\lambda \mid \alpha_X, \beta_X)}$

Conjugada

Análisis Bayesiano Paramétrico

- ✓ En ocasiones es conveniente reportar una distribución final cuyas implicaciones provengan en mayor medida de los datos $\underline{X}_{(n)}$ y no de la inicial $P(\theta)$.

$$P(\theta \mid \underline{X}_{(n)}) \propto P(\underline{X}_{(n)} \mid \theta)\ \pi(\theta)$$

- ✓ A $\pi(\theta)$ se conoce como *no-informativa, mínimo-informativa* o de *referencia.*

Análisis Bayesiano Paramétrico

- ✓ Existen distintos métodos para producir distribuciones iniciales de referencia.
  - Laplace (principio de la razón insuficiente).
  - Jeffreys (la regla de Jeffreys).
  - Bernardo (finales de referencia).
  - Raiffa & Schlaifer (conjugadas no informativas).

Análisis Bayesiano Paramétrico

- ✓ El procedimiento a partir de distribuciones conjugadas es el más simple.

Distribuciones Conjugadas

$$P(\underline{X}_{(n)} \mid \theta)$$

$$P(\theta) \in \mathscr{F}_{\phi} \quad \Rightarrow \quad P(\theta \mid \underline{X}_{(n)}) \in \mathscr{F}_{\phi}$$

Análisis Bayesiano Paramétrico

$$\mathscr{F}_{\phi} = \left\{ P(\theta) \,\middle|\, P(\theta) = P(\theta \mid \underline{\phi});\ \ \underline{\phi} \in \Phi \subset \Re^{k} \right\}$$

$$P(\underline{X}_{(n)} \mid \theta)$$

$$\boxed{P(\theta) \in \mathscr{F}_{\phi} \quad \Rightarrow \quad P(\theta \mid \underline{X}_{(n)}) \in \mathscr{F}_{\phi}}$$

Análisis Bayesiano Paramétrico

$$P(\theta) \in \mathscr{F}_{\phi} \quad \Rightarrow \quad P(\theta) = P(\theta \mid \underline{\phi}) \quad \text{con} \quad \underline{\phi} \in \Phi$$

$$P(\theta \mid \underline{X}_{(n)}) \in \mathscr{F}_{\phi} \quad \Rightarrow \quad P(\theta \mid \underline{X}_{(n)}) = P(\theta \mid \underline{\phi}_X) \quad \text{con} \quad \underline{\phi}_X \in \Phi$$

$$\boxed{\underline{\phi} \quad \rightarrow \quad \underline{\phi}_X}$$

Análisis Bayesiano Paramétrico

$$\underline{\phi} \longrightarrow \underline{\phi}_X$$

$$\underline{\phi}_X = g(\underline{x}_{(n)}\ ,\ \underline{\phi}\ )$$

$$\boxed{\underline{\phi} \longrightarrow \underline{\phi}^{*} \in \overline{\mathscr{F}}_{\phi} \quad \Rightarrow \quad \underline{\phi}_X = g^{*}(\underline{x}_{(n)})}$$

Análisis Bayesiano Paramétrico

- ✓ Con frecuencia el hiper parámetro se hace tender a un punto en la frontera de Φ.

- ✓ *Ejemplo 1b*. X variable aleatoria Normal con media μ y varianza σ^2 conocida (precisión $\tau = 1/\sigma^2$).

  - Si la inicial para μ es $P(\mu) = N(\mu \mid m, \tau_0)$
  - $\phi = (m, \tau_0)$

Análisis Bayesiano Paramétrico

✓ *Ejemplo 1b*. X variable aleatoria Normal con media μ y varianza σ^2 conocida (precisión $\tau = 1/\sigma^2$).

- Si la inicial para μ es $P(\mu) = N(\mu \mid m, \tau_0)$
- $\phi = (m, \tau_0)$
- $m \to 0;\ \tau_0 \to 0$

Distribución mínimo informativa
límite de Conjugadas

- $\pi(\mu) \propto 1$ $\Rightarrow$

$$P(\mu \mid \underline{X}_{(n)}) = N(\mu \mid \bar{X}, n\tau)$$

Análisis Bayesiano Paramétrico

- ✓ *Ejemplo 3.* X variable aleatoria Poisson con media λ.

  - Si la inicial para λ es $P(\lambda) = \text{Gamma}(\lambda \mid \alpha, \beta)$
  - $\phi = (\alpha, \beta)$
  - $\alpha \to 0;\ \beta \to 0$
  - $\pi(\lambda) \propto \lambda^{-1}$ $\Rightarrow$

Distribución mínimo informativa
límite de Conjugadas

$$\boxed{P(\lambda \mid \underline{X}_{(n)}) = \text{Gamma}(\lambda \mid n\bar{X}, n)}$$

Análisis Bayesiano Paramétrico

- Problemas de Inferencia Paramétrica
  - ✓ Estimación puntual
  - ✓ Estimación por regiones
  - ✓ Contraste de hipótesis
  - ✓ Pronósticos

Análisis Bayesiano Paramétrico

5.2. Estimación puntual

Sea X una v.a. con f.d.p.g. $f(x \mid \theta)$, $\theta \in \Theta$, se desea estimar puntualmente a θ. La idea es proponer un valor de $\hat{\theta}$ como aproximación de el valor desconocido θ. Así, para expresar este problema como uno de decisión se define

$$D = \{d_{\hat{\theta}} \mid \hat{\theta} \in \Theta\}$$

donde $d_{\hat{\theta}}$ =estimar a θ con $\hat{\theta}$. Observe que en este caso el tamaño de D está determinada por la cardinalidad del conjunto Θ, por lo que también la representación gráfica del problema, mediante el árbol de decisión estará afectada por este conjunto. Sin embargo, es posible mostrar una rama *genérica* de este tal como se hace en la figura 5.2.

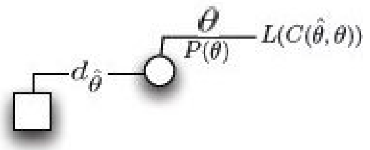


Figura 5.2: Rama típica del árbol de decisión para el problema de estimación puntual.

Análisis Bayesiano Paramétrico

$$\mathbb{E}\{L(d_{\hat{\theta}},\theta)\} = \int_{\Theta} L(d_{\hat{\theta}},\theta)P(\theta)d\theta = h(\hat{\theta}).$$

las consecuencias de una estimación $\hat{\theta}$ dependen de lo bien que se reproduzca el valor desconocido θ. De esta forma, resulta apropiado utilizar funciones de pérdida que dependan de la distancia entre $\hat{\theta}$ y θ, y en este sentido, que entre mayor sea dicha distancia mayor sea la pérdida. En particular, una opción es utilizar la función de pérdida cuadrática $L(\hat{\theta},\theta) = (\hat{\theta}-\theta)^2$, de modo que la solución se obtiene de

$$\min_{\hat{\theta}\in\Theta} \mathbb{E}_{P(\theta)}\{L(d_{\hat{\theta}},\theta)\} = \min_{\hat{\theta}\in\Theta} \mathbb{E}_{P(\theta)}\{(\hat{\theta}-\theta)^2\}$$

donde $\mathbb{E}_{P(\theta)}\{(\hat{\theta}-\theta)^2\}$ se conoce como error cuadrático medio bayesiano, y que se puede desarrollar como

$$\begin{aligned}
\mathbb{E}_{P(\theta)}\{(\hat{\theta}-\theta)^2\} &= \mathbb{E}_{P(\theta)}\left\{(\hat{\theta}-\mathbb{E}(\theta)+\mathbb{E}(\theta)-\theta)^2\right\} \\
&= \mathbb{E}_{P(\theta)}\left\{(\hat{\theta}-\mathbb{E}(\theta))^2\right\} + 0 + \mathbb{E}_{P(\theta)}\left\{(\mathbb{E}(\theta)-\theta)^2\right\} \\
&= \left(\hat{\theta}-\mathbb{E}(\theta)\right)^2 + Var(\theta)
\end{aligned}$$

$$\boxed{\hat{\theta}_B = \mathbb{E}\{(\theta \mid x)\}}$$

Análisis Bayesiano Paramétrico

5.3. Estimación por regiones

Sea entonces X una v.a. con f.d.p.g. $P(x \mid \theta)$, $\theta \in \Theta$, se desea estimar θ por regiones. La idea es encontrar una región $A \subseteq \Theta$ que sea "lo más pequeña posible" y que tenga "buenas posibilidades" de incluir a θ.

Así, al igual que en los casos de contraste de hipótesis y estimación puntual es posible expresar este problema en términos de uno de decisión. En este caso $D = \{d_A \mid A \subseteq \Theta\}$, donde comúnmente la región A se restringe a un tipo que permita una interpretación útil en la práctica. Una rama *genérica* de este problema se exhibe en la figura 5.4.

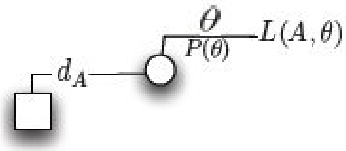


Figura 5.4: Rama típica del árbol de decisión para el problema de estimación por regiones.

Análisis Bayesiano Paramétrico

En particular si $\Theta \subseteq \Re$ y las regiones son intervalos, entonces el conjunto de decisiones resulta $D = \{d_{ab} \mid d_{ab} = [a,b] \subseteq \Theta\}$. Para este caso, una posible función de pérdida es la que tiene la forma

$$L([a,b],\theta) = \alpha g(A) + (1-\alpha)h(A,\theta)$$

con $g(A) = b - a,\ h(A,\theta) = I_{[a,b]^c}(\theta)$ y $\alpha \in (0,1)$. De donde resulta que

$$\begin{aligned}\mathbb{E}\left(L\left([a,b],\theta\right)\right) &= \alpha(b-a) + (1-\alpha)\left(1 - P(\theta \in A)\right)\\ &= \alpha(b-a) + (1-\alpha) - (1-\alpha)P(\theta \in A)\\ &= \alpha(b-a) + (1-\alpha)\left(F_\theta(b) - F_\theta(a)\right).\end{aligned}$$

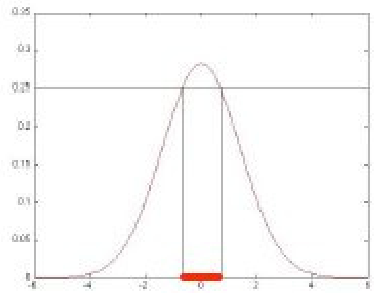
Análisis Bayesiano Paramétrico

Una simplificación a la que se recurre con frecuencia consiste en fijar $F_\theta(b) - F_\theta(a) = \alpha$, con lo que el problema se reduce a uno sin incertidumbre. Así, fijando $P(\theta \in A)$, el problema consiste en minimizar la longitud del intervalo.

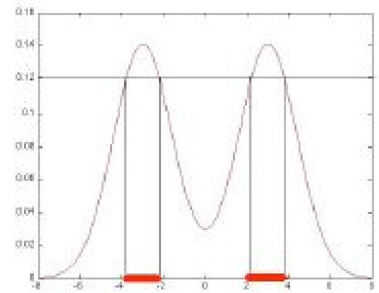
Ahora, es fácil concluir que para obtener la menor longitud de los intervalos es conveniente iniciar su construcción a partir de la imagen inversa de la(s) moda(s). De hecho, se puede probar que si se define una región $I \in \Theta$ tal que $P(\theta \in I) = 1 - \alpha$ y de manera que $P(\theta) > P(\theta') \ \forall \ \theta \in I$ y $\theta' \notin I$. Entonces, si A es cualquier otra región de Θ tal que $P(\theta \in A) = 1 - \alpha$ el área de A será al menos el de I.

Análisis Bayesiano Paramétrico

Una región con estas características, se conoce como región de máxima probabilidad o máxima densidad, y si bien en muchos casos se puede calcular analíticamente, en general se determina numéricamente con métodos como la bisección. En la figura 5.5 se muestran (en rojo) dos posibles formas que puede tomar un intervalo de máxima densidad.



(a)



(b)

Figura 5.5:
(a) Región de máxima densidad para una distrubución unimodal
(b) Región de máxima densidad para una distrubución multimodal

Análisis Bayesiano Paramétrico

5.1. Contraste de hipótesis

Sea X una v.a. con función de densidad de probabilidad generalizada (f.d.p.g.) $P(x \,|\, \theta)$, $\theta \in \Theta = \{\theta_0, \theta_1\}$ y $P(x \,|\, \theta)$ tiene distribución conocida. Se desea contrastar las hipótesis paramétricas simples $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$.

Considere un problema de decisión donde los elementos del conjunto de eventos inciertos relevantes están dados por

$$E_0 = P(x \,|\, \theta_0) \text{ es el modelo que mejor representa la realidad y}$$
$$E_1 = P(x \,|\, \theta_1) \text{ es el modelo que mejor representa la realidad}$$

y en el que el conjunto de decisiones está dado por $D = \{d_0, d_1\}$ donde d_0 representa describir a X con $P(x \,|\, \theta_0)$ y d_1 describir a X con $P(x \,|\, \theta_1)$.

Análisis Bayesiano Paramétrico

Como ya se sabe, este problema puede ser representado gráficamente mediante el árbol de decisión, presentado en la figura 5.1.

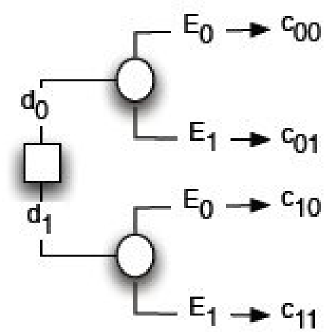


Figura 5.1: Árbol de decisión para el problema de contraste de hipotesis.

Así, la ocurrencia del evento E_0 implicaría que la hipóteis H_0 es verdadera y, por el contrario, si sucediera E_1, entonces H_1 sería correcta. De esta manera, contrastar las hipótesis H_0 v.s. H_1 implica elegir entre d_0 y d_1.

Análisis Bayesiano Paramétrico

$$
\begin{aligned}
d_1 \text{ es la solución de Bayes } &\iff \mathbb{E}\{L(d_1|E)\} > \mathbb{E}\{L(d_0|E)\} \\
&\iff P_0 l_{10} + P_1 l_{11} > P_0 l_{00} + P_1 l_{01} \\
&\iff (l_{01} - l_{11}) P_1 > (l_{10} - l_{00}) P_0 \\
&\iff k = \frac{(l_{01} - l_{11})}{(l_{10} - l_{00})} > \frac{P_0}{1 - P_0} \\
&\iff \frac{k}{1+k} > P_0
\end{aligned}
$$

		Naturaleza H_0	Naturaleza H_1
Decisión	d_0	Acierto c_{00}	ET2 c_{01}
Decisión	d_1	ET1 c_{10}	Acierto c_{11}

Análisis Bayesiano Paramétrico

$$\text{se rechaza } H_0 \iff \frac{l'_{01}}{l'_{10}} > \frac{P_0}{P_1}$$

Ahora, sea $x_{(n)}$ una m.a. de tamaño n de X. Entonces, utilizando la regla de Bayes

$$P(\theta_0|x_{(n)}) = \frac{P(x_{(n)}|\theta_0)P(\theta_0)}{P(x_{(n)})} \quad y \quad P(\theta_1|x_{(n)}) = \frac{P(x_{(n)}|\theta_1)P(\theta_1)}{P(x_{(n)})},$$

por lo que a posteriori

$$\text{se rechaza } H_0 \iff \frac{l'_{01}}{l'_{10}} > \frac{P(\theta_0|x_{(n)})}{P(\theta_1|x_{(n)})}.$$

Análisis Bayesiano Paramétrico

$$\frac{P(\theta_0|x_{(n)})}{P(\theta_1|x_{(n)})} = \frac{\dfrac{P(x_{(n)}|\theta_0)P(\theta_0)}{P(x_{(n)})}}{\dfrac{P(x_{(n)}|\theta_1)P(\theta_1)}{P(x_{(n)})}} = \frac{P(x_{(n)}|\theta_0)}{P(x_{(n)}|\theta_1)}\frac{P_0}{P_1}$$

lo que implica que

$$\boxed{\text{se rechaza } H_0 \Longleftrightarrow \boldsymbol{C} \equiv \frac{P_1 l'_{01}}{P_0 l'_{10}} > \frac{P(x_{(n)}|\theta_0)}{P(x_{(n)}|\theta_1)}}$$

Análisis Bayesiano Paramétrico

El hecho más destacado de este resultado, que como puede observarse, es totalmente general (no depende de las particulares hipótesis simples ni del modelo de los datos), es el que establece que la muestra $x_{(n)}$ interviene en la decisión sobre las hipótesis única y exclusivamente a través de *cociente de verosimilitudes*

$$\Lambda = \frac{P(x_{(n)}|\theta_0)}{P(x_{(n)}|\theta_1)},$$

pudiendo así establecer una regla de decisión $\delta : \mathfrak{X}_{(n)} \to D$ tal que

$$\delta(x_{(n)}) = \begin{cases} d_1 & \text{si } \boldsymbol{C} > \dfrac{P(x_{(n)}|\theta_0)}{P(x_{(n)}|\theta_1)} \\ d_0 & \text{e.o.c} \end{cases}$$

De hecho, si se recupera la idea Frecuentista de región de rechazo (para H_0), entonces el procedimiento Bayesiano establece que H_0 se rechaza si y sólo si $x_{(n)} \in \mathscr{C}$ donde

$$\mathscr{C} = \left\{x_{(n)} \in \mathfrak{X}_{(n)} \mid \boldsymbol{C} > \Lambda\right\}.$$

Análisis Bayesiano Paramétrico

5.4.1. Pronóstico puntual

Sea X una v.a. con f.d.p.g. $P(x \mid \theta)$, $\theta \in \Theta$, se desea pronosticar un valor $x_\star$ de una observación futura $x \in \mathscr{X}$.

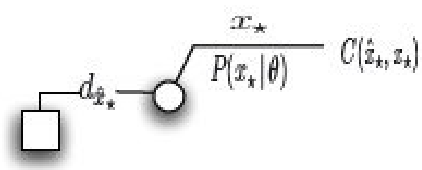
Así, se trata de elegir una $\hat{x}_\star$ como anticipación del valor $x_\star$ que efectivamente producirá el fenómeno cuando sea observado. Por tanto, en este problema es posible definir $D = \{d_x \mid x \in \mathscr{X}\}$. Ahora, las consecuencias de un pronóstico particular $\hat{x}_\star$ dependen de lo bien que este reproduzca al valor futuro $x_\star$. Así, las funciones de pérdida apropiadas, como en el caso de estimación puntual, en general dependen de alguna forma de la distancia entre $\hat{x}_\star$ y $x_\star$, y son tales que entre mayor sea la distancia, asignen mayor pérdida.

Análisis Bayesiano Paramétrico

- Si θ es conocido.

  En este escenario, una reprsentación gráfica del problema de resulta en la figura 5.6, donde puede observarse que este problema tiene exactamente la misma estructura que el problema de estimación puntual.

Figura 5.6: Rama típica del árbol de decisión para el problema de pronostico puntual con θ conocido.



Y en particular, si se utiliza $L\left(C(\hat{x}_\star, x_\star)\right) = a(\hat{x}_\star - x_\star)^2$ donde a es una constante positiva, resulta que a priori la solución está determinada por

$$\hat{x}_{\star_B} = \begin{cases} \mathbb{E}_{P(x_\star \mid \theta)}(x) & \text{si } \mathbb{E}_{P(x_\star \mid \theta)}(x) \in \mathscr{X} \\ \text{el valor más cercano a } \mathbb{E}_{P(x_\star \mid \theta)}(x) & \text{e.o.c.} \end{cases}$$

Análisis Bayesiano Paramétrico

- Si θ es desconocido

En este caso, a diferencia del caso en el que θ es conocido, la utilidad esperada de la opción d_x no puede calcularse con respecto a $P(x \mid \theta)$ puesto que θ es desconocido. De hecho, ocurre que θ siendo desconocido introduce otro factor de incertidumbre, y entonces si tanto $x_\star$ como θ son desconocidos la distribución de probabilidad que debe asignar el tomador de decisiones es necesariamente de la forma de una conjunta $P(x_\star, \theta)$ y ya no la de una condicional $P(x_\star \mid \theta)$. De esta manera, el problema tiene asociado un árbol con una estructura como el de la figura 5.7. En esta figura, se hace evidente que existen en el problema dos fuentes de incertidumbre $x_\star$ y θ. Sin embargo, es interesante observar que la función de pérdida involucra a $x_\star$ pero no a θ.

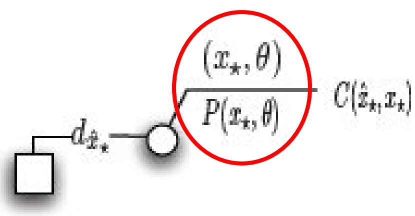


Figura 5.7: Rama típica del árbol de decisión para el problema de pronostico puntual con θ desconocido.

Así, la solución de Bayes estará dada por $\hat{x}_{\star_B} = \underset{x}{\operatorname{argmin}}\, \mathbb{E}_{P(x_\star,\theta)}\{L(\hat{x}_\star, x_\star)\}$.

Análisis Bayesiano Paramétrico

$$
\begin{aligned}
\mathbb{E}_{P(x_\star,\theta)}\{L(\hat{x}_\star, x_\star)\} &= \int_{\mathscr{X}} \int_{\Theta} L(\hat{x}_\star, x_\star) P(x_\star, \theta) d\theta dx_\star \\
&= \int_{\mathscr{X}} L(\hat{x}_\star, x_\star) \left\{ \int_{\Theta} P(x_\star, \theta) d\theta \right\} dx_\star \\
&= \int_{\mathscr{X}} L(\hat{x}_\star, x_\star) \left\{ \int_{\Theta} P(x_\star \,|\, \theta) P(\theta) d\theta \right\} dx_\star \\
&= \int_{\mathscr{X}} L(\hat{x}_\star, x_\star) P(x_\star) dx_\star \\
&= \mathbb{E}_{P(x_\star)}\{L(\hat{x}_\star, x_\star)\} \\
&= g(\hat{x}_\star)
\end{aligned}
$$

Análisis Bayesiano Paramétrico

En otras palabras, el hecho de que la función de pérdida no dependa de θ permite expresar esta pérdida esperada con una formulación alternativa en donde efectivamente el único factor de incertidumbre es $x_\star$ como en el caso en que el parámetro es conocido. La diferencia, sin embargo es que el modelo que se utiliza en aquel caso $P(x_\star \,|\, \theta)$ es ahora remplazado por $P(x_\star)$ que se relaciona con el primero a través de la expresión

$$P(x_\star) = \int_\Theta P(x_\star \,|\, \theta) P(\theta) d\theta.$$

Distribución Predictiva

Análisis Bayesiano Paramétrico

5.4.2. Pronóstico por regiones

Sea X una v.a. con f.d.p.g. $P(x \mid \theta)$, $\theta \in \Theta$, se desea pronosticar por regiones un valor $x_\star$ de una observación futura de $x \in \mathscr{X}$.

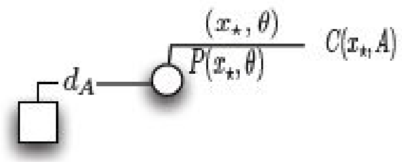


Figura 5.8: Rama típica del árbol de decisión para el problema de pronostico por regiones.

De la misma forma que ocurre con el pronóstico puntual, el problema de pronóstico por regiones resulta ser completamente análogo a su contraparte de estimación por regiones. Aquí, el espacio de opciones es $D = \{d_A \mid A \subseteq \mathscr{X}\}$, con $\mathscr{X}$ el soporte de X, y la distribución de probabilidades relevante es la predictiva para x, sea con θ desconocida (a priori o a posteriori) o con θ conocida.

Referencias

Bibliografía

Box, G.E.P & Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Reading: Addison Wesley.

Congdon, P. (2001). Bayesian Statistical Modelling. Chichester: Wiley

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (1995). *Bayesian Data Analysis*. London: Chapman & Hall.

Mendoza, M. y Regueiro P. (2011). *Introducción al Análisis Estadístico Bayesiano*. Documentos de Trabajo. Departamento de Estadística, Instituto Tecnológico Autónomo de México. DE-A11.2. ITAM, México.

Referencias

Bibliografía

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Second edition. New York: Springer Verlag.

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

Box, G.E.P & Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Reading: Addison Wesley.

Congdon, P. (2001). Bayesian Statistical Modelling. Chichester: Wiley

De Groot, M.H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

De Groot, M.H. (1988). *Probabilidad y Estadística*. México: Addison Wesley Iberoamericana.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (1995). *Bayesian Data Analysis*. London: Chapman & Hall.

Lindley, D.V. (1965). *An Introduction to Probability and Statistics from a Bayesian Viewpoint*. Vol 2. *Inference*. Cambridge: Cambridge University Press.

Lindley, D.V. (1985). *Making Decisions*. Second edition. London: Wiley.

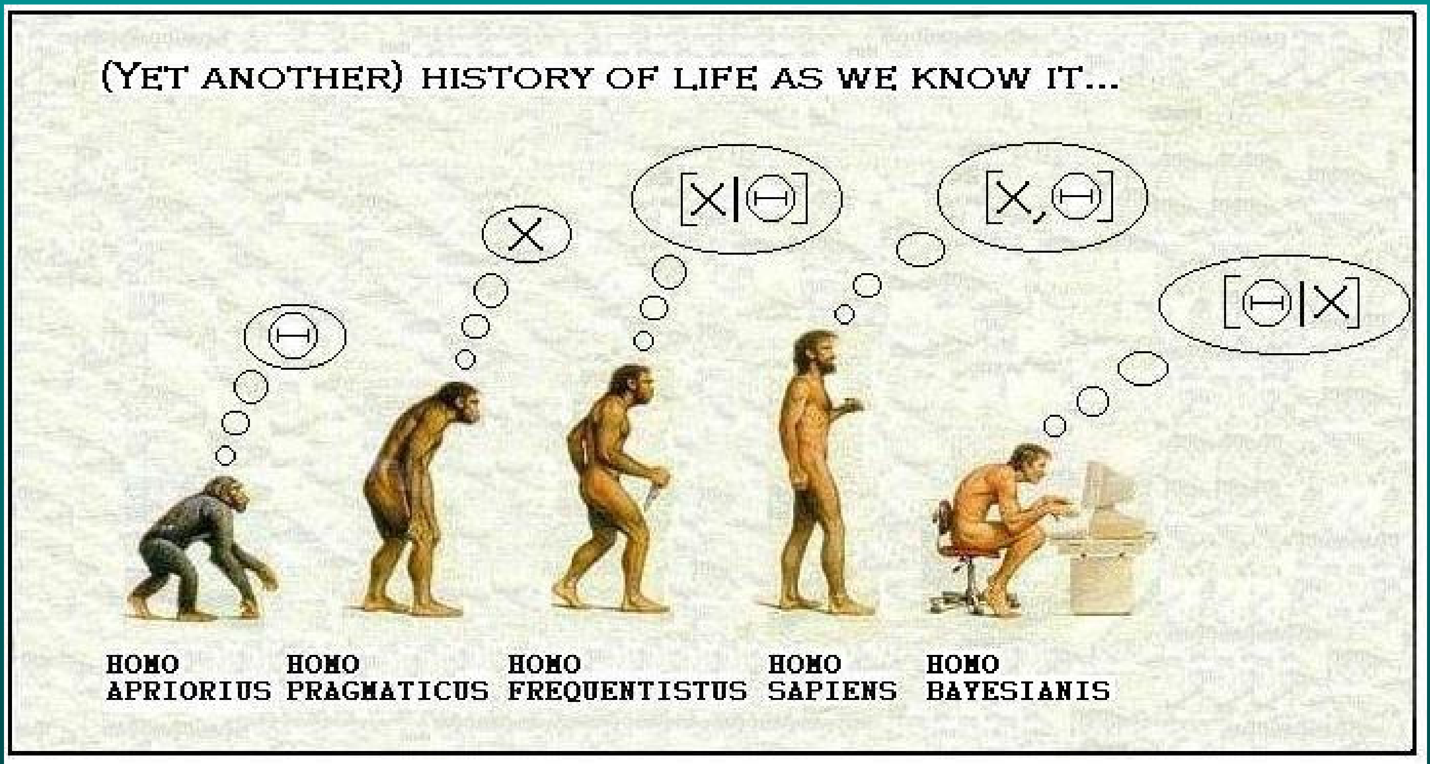
Mignon, H.S. and Gamerman, D. (1999). *Statistical Inference: An Integrated Approach*. London: Arnold.

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics*. Vol 2b. *Bayesian Inference*. Cambridge: Edward Arnold.

Press, S.J. (1989). *Bayesian Statistics. Principles, Models and Applications*. New York: Wiley.

Robert, C.P. (2001). *The Bayesian Choice*. Second edition. New York: Springer Verlag.

Resumen



(YET ANOTHER) HISTORY OF LIFE AS WE KNOW IT...
[Θ]
[X]
[X|Θ]
[X,Θ]
[Θ|X]
HOMO APRIORIUS
HOMO PRAGMATICUS
HOMO FREQUENTISTUS
HOMO SAPIENS
HOMO BAYESIANIS