Chapter 4

Estimadores Bayesianos y Minimax

4.1 Enfoque Bayesiano del problema de la estimación puntual

Consideremos nuevamente un problema estadístico de estimación paramétrico. Se observa un vector $\mathbf{X} = (X_1, ..., X_n)$, que puede ser, por ejemplo, aunque no necesariamente, una muestra aleatoria de cierta distribución) con densidad discreta o continua en la familia $f(\mathbf{x}, \boldsymbol{\theta})$, con $\boldsymbol{\theta} = (\theta_1, ..., \theta_p) \in \Theta \subset I\!R^p$.

El enfoque llamadado frecuentista que hemos estudiado no supone ningún conocimiento previo de $\boldsymbol{\theta}$. El enfoque bayesiano, por lo contrario, supone que se tiene alguna información previa sobre $\boldsymbol{\theta}$. Esta información está expresada por medio de una distribución sobre $\boldsymbol{\theta}$, denominada distribución *a priori.* Aquí supondremos que esta distribución a priori, indicada por τ, tiene una densidad $\gamma(\boldsymbol{\theta})$. La distribución a priori puede tener distintas interpretaciones según el problema. Se pueden dar las siguientes alternativas

- La distribución a priori está basada en experiencias previas similares.
- La distribución a priori expresa una creencia subjetiva.

El hecho de que el enfoque bayesiano considere una distribución de probabilidades sobre $\boldsymbol{\theta}$, supone tratar a $\boldsymbol{\theta}$ como una variable aleatoria, y por lo tanto a esta variable la denominaremos Θ para distinguirla del valor que toma $\boldsymbol{\theta}$. Esta notación puede llevar a confusión dado que también llamamos

Θ al conjunto de valores de $\boldsymbol{\theta}$. Sin embargo, por el contexto quedará claro el significado de este símbolo en cada caso.

Dado que consideramos ahora el valor del parámetro como el valor de una variable aleatoria, la interpretación de la familia de densidades $f(\mathbf{x}, \boldsymbol{\theta})$ en el enfoque bayesiano también cambia. En el enfoque bayesiano $f(\mathbf{x}, \boldsymbol{\theta})$ se interpreta como la distribución condicional de la muestra $\mathbf{X}$ dado que la variable Θ toma el valor $\boldsymbol{\theta}$.

Una vez observada la muestra $\mathbf{X}$ se puede calcular la distribución condicional de Θ dada $\mathbf{X}$. Esta distribución se denomina distribución *a posteriori* y está dada por

$$f(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\mathbf{x}, \boldsymbol{\theta})\gamma(\boldsymbol{\theta})}{\int ... \int f(\mathbf{x}, \mathbf{t})\gamma(\mathbf{t})d\mathbf{t}}. \tag{4.1}$$

En efecto el numerador de (4.1) corresponde a la densidad conjunta de $\mathbf{X}$ y Θ, y el denominador a la densidad marginal de $\mathbf{X}$.

Si la distribución de $\boldsymbol{\theta}$ fuese discreta, habría que reemplazar las integrales del denominador por las correspondientes sumatorias. En lo sucesivo, supondremos que las distribuciones de $\mathbf{X}$ y de $\boldsymbol{\theta}$ son continuas, pero el tratamiento en el caso discreto es similar.

Una de las ventajas del enfoque bayesiano es que se pueden definir en forma natural estimadores óptimos, sin necesidad de restricciones poco naturales como la de estimadores insesgados a la que debimos recurrir en el enfoque frecuentista. Para ver esto supongamos que queremos estimar $\lambda = q(\boldsymbol{\theta})$ y consideremos una función de pérdida $\ell(\boldsymbol{\theta}, d)$ que indica el costo de estimar $\lambda = q(\boldsymbol{\theta})$ utilizando del valor d. Supongamos que se tiene un estimador $\hat{\lambda} = \delta(\mathbf{x})$. Luego la pérdida será una variable aleatoria $\ell(\Theta, \delta(\mathbf{X}))$, y la pérdida esperada que llamaremos *riesgo de Bayes* está dada por

$$r(\delta, \tau) = E(\ell(\Theta, \delta(\mathbf{X}))), \tag{4.2}$$

donde aquí la esperanza se toma con respecto a la distribución conjunta de $\mathbf{X}$ y Θ. Por lo tanto, dada la distribución priori τ, un estimador óptimo será aquel que minimice $r(\delta, \tau)$. Este estimador se denomina *estimador de Bayes* correspondiente a la distribución a priori τ y será representado por δ_τ.

Luego, la función de riesgo de la teoría frecuentista, $R(\delta, \boldsymbol{\theta})$, estará dada por

$$\begin{aligned} R(\delta, \boldsymbol{\theta}) &= E_{\boldsymbol{\theta}}(\ell(\boldsymbol{\theta}, \delta(\mathbf{X}))) \\ &= E(\ell(\Theta, \delta(\mathbf{X}))|\Theta = \boldsymbol{\theta}) = \int \ell(\boldsymbol{\theta}, \delta(\mathbf{x}))f(\mathbf{x}, \boldsymbol{\theta})d\mathbf{x}. \end{aligned} \tag{4.3}$$

Con lo cual,

$$r(\delta,\tau) = E_\tau\left(E(\ell(\Theta,\delta(\mathbf{X}))|\Theta)\right) = \int ... \int R(\delta,\boldsymbol{\theta})\gamma(\boldsymbol{\theta})d\boldsymbol{\theta}. \qquad (4.4)$$

Consideremos como caso particular la función de pérdida cuadrática, es decir,

$$\ell(\boldsymbol{\theta},d) = (q(\boldsymbol{\theta}) - d)^2.$$

En este caso, el estimador de Bayes será la función $\delta_\tau(X)$ que minimiza el error cuadrático medio

$$E((\delta(\mathbf{X}) - q(\Theta))^2)$$

y por lo tanto, de acuerdo a la teoría de esperanza condicional, éste será único y estará dado por

$$\delta_\tau(\mathbf{x}) = E(q(\Theta)|\mathbf{X}=\mathbf{x}) = \int ... \int q(\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta},$$

es decir, será la esperanza condicional de $q(\Theta)$ con respecto a la distribución a posteriori de Θ.

Ejemplo 1. Sea $\mathbf{X} = (X_1, ..., X_n)$ una muestra independiente de una distribución Bi$(\theta,1)$, y supongamos que la distribución a priori τ de θ sea una distribución $\beta(a,b)$, es decir, con una densidad

$$\gamma(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}I_{[0,1]}(\theta). \qquad (4.5)$$

Es conocido que esta distribución tiene la siguiente esperanza y varianza

$$E(\theta) = \frac{a}{a+b}, \qquad (4.6)$$

$$var(\theta) = \frac{ab}{(a+b)^2(a+b+1)} = \frac{E(\theta)\,(1-E(\theta))}{a+b+1}. \qquad (4.7)$$

Luego si se conoce la media y la varianza de la distribución a priori de Θ, se pueden determinar a y b. La fórmula (4.7) muestra que para un dado valor de la esperanza, la varianza depende de $a+b$, tendiendo a 0 cuando $a+b \to +\infty$.

La distribución de la muestra $X_1, X_2, ...X_n$ dado el valor de θ tiene una función de probabilidad puntual dada por

$$f(x_1,..x_n,\theta) = \theta^{\sum_{i=1}^n x_i}\,(1-\theta)^{n-\sum_{i=1}^n x_i}. \qquad (4.8)$$

Luego usando (4.1) se tiene que la distribución a posteriori de θ tiene una densidad dada por

$$f(\theta|x_1,...,x_n)=\frac{\theta^{\sum_{i=1}^n x_i+a-1}(1-\theta)^{n-\sum_{i=1}^n x_i+b-1}}{\int_0^1 t^{\sum_{i=1}^n x_i+a-1}(1-t)^{n-\sum_{i=1}^n x_i+b-1}dt}. \qquad (4.9)$$

Ahora bien, el denominador de esta expresión es igual a

$$\frac{\Gamma(n+a+b)}{\Gamma(a+\sum_{i=1}^n x_i)\Gamma(n-\sum_{i=1}^n x_i+b)}$$

por lo tanto, la distribución a posteriori de θ dado $\mathbf{X}=\mathbf{x}$ es $\beta(a+\sum_{i=1}^n x_i, n-\sum_{i=1}^n +b)$.

Supongamos que la función de pérdida es cuadrática. Luego el estimador de Bayes, que indicaremos por $\delta_{a,b}$, está dado por $E(\Theta|\mathbf{X})$, y de acuerdo a (4.6) tendremos que

$$\delta_{a,b}=\frac{T+a}{a+b+n}=\frac{n}{n+a+b}\,\frac{T}{n}+\frac{a+b}{a+b+n}\,\frac{a}{a+b}, \qquad (4.10)$$

donde $T=\sum_{i=1}^n X_i$. Por lo tanto, el estimador de Bayes se comporta como un promedio ponderado de dos estimadores: el IMVU $\delta_1=T/n$ que no usa la información de la distribución a priori y $\delta_2=a/(a+b)$ que corresponde a la esperanza de la distribución a priori y que se usaría si no se hubiese observado la muestra. También vemos que el peso asignado a δ_2 tiende a 0 cuando el tamaño de la muestra n aumenta.

De acuerdo a (4.10), el estimador de Bayes correspondiente a una distribución a priori $\beta(a,b)$ puede interpretarse como el estimador frecuentista correspondiente a una muestra de tamaño $n+a+b$ con $\sum_{i=1}^n X_i+a$ éxitos.

Observación 1. En el ejemplo anterior hemos partido de una distribución a priori $\beta(a,b)$, y hemos obtenido que la distribución a posteriori también está en la misma familia, ya que es $\beta(a+\sum_{i=1}^n x_i, n-\sum_{i=1}^n x_i+b)$. Se dice entonces que la familia de distribuciones beta es la *conjugada* de la familia de distribuciones binomial.

Ejemplo 2. Sea $\mathbf{X}=(X_1,...,X_n)$ una muestra independiente de una distribución $\mathrm{N}(\theta,\sigma^2)$, con σ^2 conocido, y supongamos que la distribución a priori de θ sea $\mathrm{N}(\mu,\rho^2)$.

Luego la densidad de la muestra $\mathbf{X}=(X_1,...,X_n)$ dado θ está dada por

$$f(\mathbf{x},\theta)=\frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^n(x_i-\theta)^2\right), \qquad (4.11)$$

donde $\exp(x) = e^x$. La densidad de la distribución a priori está dada por

$$\gamma(\theta) = \frac{1}{(2\pi)^{1/2}\rho}\exp\left(\frac{-(\theta-\mu)^2}{2\rho^2}\right) \tag{4.12}$$

Luego multiplicando (4.11) y (4.12), desarrollando los cuadrados y haciendo algún manipuleo algebraico, se obtiene que distribución conjunta de $\mathbf{X}$ y Θ está dada por

$$f_{\mathbf{X},\Theta}(\mathbf{x},\theta) = C_1(\mathbf{x},\sigma^2,\mu,\rho^2)\exp\left(-\frac{\theta^2}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\rho^2}\right)+\theta\left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu}{\rho^2}\right)\right),$$

donde $C_1(\mathbf{x},\sigma^2,\mu,\rho^2)$ no depende de θ. Completando cuadrados, se obtiene

$$f_{\mathbf{X},\Theta}(\mathbf{x},\theta) = C_2(\mathbf{x},\sigma^2,\mu,\rho^2)\exp\left(-\frac{1}{2D}\left(\theta-D\left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu}{\rho^2}\right)\right)^2\right), \tag{4.13}$$

donde

$$D = \frac{1}{(n/\sigma^2)+(1/\rho^2)}. \tag{4.14}$$

Finalmente, usando (1) obtenemos

$$f(\theta|\mathbf{x}) = C_3(\mathbf{x},\sigma^2,\rho^2,\mu)\exp\left(-\frac{1}{2D}(\theta-D\left(\frac{\bar{x}}{\sigma^2}+\frac{\mu}{\rho^2}\right)\right)^2 \tag{4.15}$$

Luego, esta densidad, excepto una función que depende sólo de $\mathbf{x}$, corresponde a una distribución

$$\mathrm{N}\left(D\left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu}{\rho^2}\right),D\right). \tag{4.16}$$

Como se trata de la distribución condicional de Θ dado $\mathbf{X}=\mathbf{x}$, podemos considerar a C_3 como constante. Luego la distribución a posteriori de Θ está dada por (4.16).

Supongamos nuevamente que consideramos una función de pérdida cuadrática. El estimador de Bayes estará dado, en ese caso, por la esperanza condicional de Θ dado $\mathbf{X}$, y por lo tanto de acuerdo a (4.16) y (4.14) estará dado por

$$\delta_\tau(\mathbf{X}) = D\left(\frac{n\bar{X}}{\sigma^2}+\frac{\mu}{\rho^2}\right) = w\bar{X}+(1-w)\mu, \tag{4.17}$$

donde

$$w = \frac{n/\sigma^2}{(n/\sigma^2) + (1/\rho^2)}.$$

Por lo tanto, nuevamente, el estimador de Bayes es un promedio ponderado del estimador IMVU de la teoría frecuentista $\bar{X}$ y la media de la distribución a priori μ. Los pesos son inversamente proporcionales a las varianzas σ^2/n y ρ^2 de ambos estimadores. A medida que el tamaño de la muestra n crece, el peso del estimador basado en la información a priori tiende a 0. Es decir, a medida que el tamaño de la muestra crece, la información a priori tiene menos relevancia para determinar el estimador de Bayes.

Observación 2. En este ejemplo partimos de una distribución a priori en la familia $\mathrm{N}(\mu, \rho^2)$, y obtenemos que la distribución a posteriori está dada por (4.16), y por lo tanto está en la misma familia normal. Luego la familia de distribuciones conjugadas de la familia normal con varianza conocida es la familia normal.

Veamos algunas propiedades de los estimadores Bayes para funciones de pérdida arbitrarias.

Teorema 1. *Sea δ_τ un estimador Bayes respecto de la distribución a priori τ y supongamos que δ_τ es único, entonces δ_τ es admisible.*

DEMOSTRACIÓN. Supongamos que existe otro estimador δ^* tan bueno como δ_τ, es decir, $R(\delta^*, \boldsymbol{\theta}) \leq R(\delta_\tau, \boldsymbol{\theta})$ para todo $\boldsymbol{\theta} \in \Theta$. Integrando respecto a la distribución a priori de $\boldsymbol{\theta}$ en ambos miembros de la desigualdad, obtenemos $r(\delta^*, \tau) \leq r(\delta_\tau, \tau)$. Con lo cual, por la unicidad $\delta^* = \delta_\tau$.

Se puede obtener un resultado de admisibilidad para reglas Bayes sin pedir unicidad, siempre y cuando, Θ sea abierto, la distribución a priori tenga una densidad positiva para todo $\boldsymbol{\theta} \in \Theta$ y la función de riesgo $R(\delta, \boldsymbol{\theta})$ sea continua en $\boldsymbol{\theta}$ para todo estimador δ.

Hemos visto que en el caso de la pérdida cuadrática, el estimador Bayes podía obtenerse como la esperanza de la distribución a posteriori de Θ. El siguiente Teorema da una manera de obtener el estimador Bayes para el caso de otras funciones de pérdida.

Teorema 2. *Sea τ la distribución de Θ y $F_{\boldsymbol{\theta}}(\mathbf{x})$ la distribución condicional de $\mathbf{X}$ dado $\boldsymbol{\theta}$. Supongamos que se cumplen las siguientes condiciones para estimar $q(\boldsymbol{\theta})$ utilizando la pérdida $\ell(\boldsymbol{\theta}, d)$*

a) *Existe un estimador δ_0 con riesgo finito.*

b) *Para cada valor de* $\mathbf{x}$ *existe un valor, que indicaremos $\delta_\tau(\mathbf{x})$, que minimiza $E\left(\ell(\boldsymbol{\theta}, d)|\mathbf{X}=\mathbf{x}\right)$.*

Entonces, $\delta_\tau(\mathbf{x})$ es un estimador de Bayes respecto a τ.

DEMOSTRACIÓN. Sea $\delta(\mathbf{X})$ un estimador con riesgo Bayes finito. Luego, como la pérdida es nonegativa, $E\left(\ell(\boldsymbol{\theta}, \delta(X))|\mathbf{X}=\mathbf{x}\right)$ es finita para casi todo $\mathbf{x}$. Por lo tanto, tenemos

$$E\left(\ell\left(\boldsymbol{\theta}, \delta(\mathbf{x})\right)|\mathbf{X}=\mathbf{x}\right) \geq E\left(\ell\left(\boldsymbol{\theta}, \delta_\tau(\mathbf{x})\right)|\mathbf{X}=\mathbf{x}\right)$$

de donde, tomando esperanza respecto a la distribución marginal de $\mathbf{X}$, obtenemos $r(\delta, \tau) \geq r(\delta_\tau, \tau)$ y por lo tanto, δ_τ es un estimador Bayes.

Corolario *Sea τ una distribución a priori para $\boldsymbol{\theta}$ y supongamos que se cumplen las condiciones del Teorema 2.*

a) *Para la pérdida $\ell(\boldsymbol{\theta}, d) = w(\boldsymbol{\theta})(q(\boldsymbol{\theta}) - d)^2$, donde $w(\boldsymbol{\theta}) > 0$ y $E(w(\boldsymbol{\theta})) < \infty$, la regla Bayes δ_τ está dada por*

$$\delta_\tau(\mathbf{x}) = \frac{E\left(q(\boldsymbol{\theta})w(\boldsymbol{\theta})|\mathbf{X}=\mathbf{x}\right)}{E\left(w(\boldsymbol{\theta})|\mathbf{X}=\mathbf{x}\right)}$$

b) *Para la pérdida $\ell(\boldsymbol{\theta}, d) = |q(\boldsymbol{\theta}) - d|$, la regla Bayes $\delta_\tau(\mathbf{x})$ es la mediana de la distribución a posteriori de $q(\boldsymbol{\theta})$ condicional a* $\mathbf{x}$

c) *Para la pérdida $\ell(\boldsymbol{\theta}, d) = I_{|q(\boldsymbol{\theta})-d|>c}$, la regla Bayes δ_τ es el punto medio del intervalo $\mathcal{I}$ de longitud $2c$ que maximiza $P(q(\boldsymbol{\theta}) \in \mathcal{I}|\mathbf{X}=\mathbf{x})$*

4.2 Utilización de métodos bayesianos para resolver problemas frecuentistas

En esta sección vamos a mostrar como los resultados de la teoría bayesiana pueden ser útiles, aunque no se comparta ese punto de vista. Es decir, veremos que los resultados de esta teoría se pueden usar para resolver problemas que surgen de la teoría frecuentista.

Consideremos una muestra $\mathbf{X} = (X_1, ..., , X_n)$ con distribución conjunta $f(\mathbf{x}, \boldsymbol{\theta})$ donde el vector de parámetros $\boldsymbol{\theta} \in \Theta$. Supongamos que queremos estimar $\lambda = q(\boldsymbol{\theta})$ y que tenemos una función de pérdida $\ell(\boldsymbol{\theta}, d)$. En el enfoque frecuentista un estimador $\delta(\mathbf{X})$ de λ queda caracterizado por su función de riesgo

$$R(\delta, \boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(\ell(\boldsymbol{\theta}, \delta(\mathbf{X})) = \int \ell(\boldsymbol{\theta}, \delta(\mathbf{x}))f(\mathbf{x}, \boldsymbol{\theta})d\mathbf{x}. \quad (4.18)$$

Como θ es desconocido, lo ideal sería encontrar un estimador $\delta^*(\mathbf{X})$ tal que, dado cualquier otro estimador $\delta(\mathbf{x})$ se tuviese

$$R(\delta^*, \boldsymbol{\theta}) \leq R(\delta, \boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \in \Theta.$$

Como ya hemos visto al comienzo del curso estos estimadores no existen excepto en casos muy poco interesantes.

Una alternativa es comparar los estimadores a través del máximo riesgo. Dado un estimador $\delta(\mathbf{X})$ de λ su máximo riesgo se define por

$$MR(\delta) = \sup_{\boldsymbol{\theta} \in \Theta} R(\delta, \boldsymbol{\theta}). \quad (4.19)$$

El criterio de comparar los estimadores por su máximo riesgo es pesimista, ya que actua como si el parámetro fuese a tomar el valor más desfavorable para el estimador. Un estimador óptimo de acuerdo a este criterio es un estimador δ^* tal que dado cualquier otro estimador δ se tiene

$$MR(\delta^*) \leq MR(\delta). \quad (4.20)$$

Definición 1. Un estimador satisfaciendo (4.20) se denomina *minimax.*

Vamos a ver como la teoría bayesiana nos ayuda a encontrar estimadores minimax. Para ello, consideremos una distribución a priori τ con densidad $\gamma(\theta)$. El correspondiente estimador de Bayes δ_τ verifica que, dado cualquier otro estimador δ, se tiene

$$r(\delta_\tau, \tau) \leq r(\delta, \tau). \quad (4.21)$$

Luego, de acuerdo a (4.4) se tendrá entonces que para cualquier estimador δ

$$\int R(\delta_\tau, \boldsymbol{\theta})\gamma(\boldsymbol{\theta})d\boldsymbol{\theta} \leq \int R(\delta, \boldsymbol{\theta})\gamma(\boldsymbol{\theta})d\boldsymbol{\theta}. \quad (4.22)$$

Sea $r_\tau = r(\delta_\tau, \tau)$, es decir, el mínimo riesgo de Bayes cuando la distribución a priori es τ.

Definición 2. Se dirá que una distribución a priori τ_0 es *menos favorable* si, para cualquier otra distribución τ, se tiene $r_\tau \leq r_{\tau_0}$.

Naturalmente uno se puede preguntar para qué distribuciones a priori τ el estimador Bayes δ_τ será minimax. Un procedimiento minimax, al minimizar el máximo riesgo, trata de comportarse lo mejor posible en la peor situación. Por lo tanto, uno puede esperar que el estimador minimax sea Bayes para la peor distribución posible que es la distribución menos favorable.

El siguiente Teorema nos permite usar la teoría bayesiana para encontrar estimadores minimax.

Teorema 1. *Supongamos que se tiene una distribución a priori τ_0 tal que el estimador de Bayes δ_{τ_0} tiene función de riesgo, $R(\delta_\tau, \boldsymbol{\theta})$, constante en $\boldsymbol{\theta}$. Entonces:*

a) *δ_{τ_0} es un estimador minimax,*

b) *si δ_{τ_0} es el único estimador Bayes respecto de τ_0, δ_{τ_0} es el único estimador minimax,*

c) *τ_0 es la distribución menos favorable.*

DEMOSTRACIÓN. Como el riesgo de δ_{τ_0} es constante se cumple que

$$r(\delta_{\tau_0}, \tau_0) = \int R(\delta_{\tau_0}, \boldsymbol{\theta})\gamma_0(\boldsymbol{\theta})d\boldsymbol{\theta} = \sup_{\boldsymbol{\theta}\in\Theta} R(\delta_{\tau_0}, \boldsymbol{\theta}). \tag{4.23}$$

a) Consideremos un estimador $\delta \neq \delta_{\tau_0}$, luego como

$$MR(\delta) = \sup_{\boldsymbol{\theta}\in\Theta} R(\delta, \boldsymbol{\theta}) \geq R(\delta, \boldsymbol{\theta})$$

tomando esperanza respecto a la distribución a priori τ_0 obtenemos

$$MR(\delta) = \sup_{\boldsymbol{\theta}\in\Theta} R(\delta, \boldsymbol{\theta}) \geq \int R(\delta, \boldsymbol{\theta})\gamma_0(\boldsymbol{\theta})d\boldsymbol{\theta} = r(\delta, \tau_0). \tag{4.24}$$

Como δ_{τ_0} es Bayes respecto de τ_0, se cumple que

$$r(\delta, \tau_0) \geq r(\delta_{\tau_0}, \tau_0). \tag{4.25}$$

Con lo cual, a partir de (4.23), (4.24) y (4.25) obtenemos

$$MR(\delta) \geq r(\delta_{\tau_0}, \tau_0) = \sup_{\boldsymbol{\theta}\in\Theta} R(\delta_{\tau_0}, \boldsymbol{\theta}) = MR(\delta_{\tau_0})$$

y por lo tanto, δ_{τ_0} es minimax.

b) Supongamos ahora que δ_{τ_0} es el único estimador Bayes, luego se cumple

$$r(\delta, \tau_0) > r(\delta_{\tau_0}, \tau_0). \tag{4.26}$$

Con lo cual, utilizando ahora (4.23), (4.24) y (4.26) obtenemos

$$MR(\delta) \geq r(\delta, \tau_0) > r(\delta_{\tau_0}, \tau_0) = MR(\delta_{\tau_0})$$

y por lo tanto, δ_{τ_0} es el único estimador minimax.

c) Sea τ otra distribución a priori y δ_τ el estimador Bayes respecto de τ. Luego, por ser δ_τ Bayes se cumple

$$r(\delta_\tau, \tau) \leq r(\delta_{\tau_0}, \tau). \tag{4.27}$$

Por otra parte, como el riesgo de δ_{τ_0} es constante se verifica

$$\begin{aligned} r(\delta_{\tau_0}, \tau) &= \int R(\delta_{\tau_0}, \boldsymbol{\theta})\gamma(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= \sup_{\boldsymbol{\theta}\in\Theta} R(\delta_{\tau_0}, \boldsymbol{\theta}) = r(\delta_{\tau_0}, \tau_0), \end{aligned} \tag{4.28}$$

Por lo tanto, (4.27) y (4.28) nos permiten concluir que

$$r(\delta_\tau, \tau) \leq r(\delta_{\tau_0}, \tau_0)$$

con lo cual, τ_0 es la distribución menos favorable.

Ejemplo 3. Consideremos el Ejemplo 1 de estimación bayesiana para la familia binomial, usando distribuciones a priori en la familia $\beta(a, b)$ y como fúnción de pérdida la función $\ell(\theta, d) = (\theta - d)^2$. Luego hemos visto que el único estimador de Bayes está dado por

$$\delta_{a,b} = \frac{T + a}{n + a + b},$$

con $T = \sum_{i=1}^n X_i$.

Si encontramos a y b tales que $R(\delta_{a,b}, \theta)$ es constante, ese estimador será minimax y la distribución a priori correspondiente será la distribución menos favorable. Como $E_\theta(T) = n\theta$ y $var(T) = n\theta(1 - \theta)$ se tiene

$$E_\theta(\delta_{a,b}) = \frac{n\theta + a}{n + a + b}, \tag{4.29}$$

y

$$var_\theta(\delta_{a,b}) = \frac{n\theta(1-\theta)}{(n+a+b)^2}, \tag{4.30}$$

Luego, usando (4.29) y (4.30) se deduce que

$$\begin{aligned} R(\delta_{a,b},\theta) &= E((\delta_{a,b}-\theta)^2) \\ &= var_\theta(\delta_{a,b}) + (\theta - E_\theta(\delta_{a,b}))^2 \\ &= \frac{n\theta(1-\theta)}{(n+a+b)^2} + \left(\theta - \frac{n\theta+a}{n+a+b}\right)^2 \\ &= \frac{n\theta(1-\theta) + (a+b)^2\theta^2 - 2a(a+b)\theta + a^2}{(n+a+b)^2} \\ &= \frac{(-n+(a+b)^2)\theta^2 + (n-2a(a+b))\theta + a^2}{(n+a+b)^2}. \end{aligned} \tag{4.31}$$

Para que (4.31) sea constante en θ, los coeficientes en θ y θ^2 del numerador deben ser 0. Por lo tanto, se debe cumplir

$$-n + (a+b)^2 = 0, \quad n - 2a(a+b) = 0$$

La solución de este sistema de ecuaciones es $a = b = \sqrt{n}/2$, y por lo tanto el estimador de Bayes correspondiente, que será minimax, estará dado por

$$\delta_{\text{mmax}} = \frac{T + (\sqrt{n}/2)}{n+\sqrt{n}}. \tag{4.32}$$

La correspondiente función de riesgo está dada por

$$R(\delta_{\text{mmax}},\theta) = \frac{n/4}{(n+\sqrt{n})^2} = \frac{1}{4(\sqrt{n}+1)^2}.$$

El Teorema 1 no nos permite obtener un estimador minimax en el caso de la media de una distribución normal. El siguiente Teorema resultará útil en esa situación.

Teorema 2. *Supongamos que $\delta(\mathbf{X})$ sea un estimador tal que*

(i) $R(\delta,\boldsymbol{\theta}) = C \ \forall \boldsymbol{\theta} \in \Theta$,

(ii) existe una sucesión de distribuciones a priori τ_k tales que

$$\lim_{k\to\infty} r(\delta_{\tau_k},\tau_k) = C.$$

Entonces δ es minimax.
DEMOSTRACIÓN: Sea δ' otro estimador para $q(\boldsymbol{\theta})$. Se cumple entonces que

$$\sup_{\boldsymbol{\theta}} R(\delta', \boldsymbol{\theta}) \geq \int R(\delta', \boldsymbol{\theta}) \gamma_k(\boldsymbol{\theta}) d\boldsymbol{\theta} = r(\delta', \tau_k) \geq r(\delta_{\tau_k}, \tau_k). \tag{4.33}$$

Con lo cual, tomando límite en ambos miembros de (4.33), y usando (ii) se obtiene

$$MR(\delta') = \sup_{\boldsymbol{\theta}} R(\delta', \boldsymbol{\theta}) \geq C = MR(\delta),$$

y por lo tanto, δ es minimax.

Ejemplo 4. Consideremos una muestra aleatoria $\mathbf{X} = (X_1, ..., X_n)$ de una distribución $\mathrm{N}(\theta, \sigma^2)$, donde σ^2 conocida. El estimador $\delta(\mathbf{X}) = \bar{X}$ tiene como función de riesgo $R(\delta, \theta) = \sigma^2/n$, y por lo tanto se cumple la condición (i) del Teorema 2. Por otro lado, consideremos una sucesión de distribuciones a priori τ_k=N$(0, \rho_k^2)$ con $\rho_k^2 \to +\infty$. Usando la función de pérdida cuadrática, de acuerdo a lo visto en el ejemplo 2, los estimadores de Bayes son

$$\delta_{\tau_k} = w_k \bar{X},$$

donde

$$w_k = \frac{n/\sigma^2}{(n/\sigma^2) + (1/\rho_k^2)}. \tag{4.34}$$

Es fácil ver que

$$\lim_{k \to \infty} w_k = 1 \tag{4.35}$$

y que

$$\lim_{k \to \infty} \rho_k^2 \, (1 - w_k)^2 = \lim_{k \to \infty} \rho_k^2 \frac{1/\rho_k^4}{((n/\sigma^2) + (1/\rho_k^2))^2} = 0 \tag{4.36}$$

Por otro lado, se tiene

$$R(\delta_{\tau_k}, \theta) = var_\theta(\delta_{\tau_k}) + (\theta - E_\theta(\delta_{\tau_k}))^2 = w_k^2 \frac{\sigma^2}{n} + (1 - w_k)^2 \theta^2. \tag{4.37}$$

Luego

$$r(\delta_{\tau_k}, \tau_k) = E_{\tau_k}(R(\delta_{\tau_k}, \boldsymbol{\theta})) = w_k^2 \frac{\sigma^2}{n} + (1 - w_k)^2 \rho_k^2.$$

Con lo cual, usando (4.35) y (4.36) se concluye que

$$\lim_{k \to \infty} r(\delta_{\tau_k}, \tau_k) = \frac{\sigma^2}{n}$$

Por lo tanto se cumple la condición (ii) del Teorema 2, y el estimador $\delta(\mathbf{X}) = \bar{X}$ es minimax. El Teorema 2 no nos permite obtener la unicidad del estimador minimax.