1. En un problema de decisión estadístico con función de utilidad $u(a,\theta)$, considera la función de utilidad alternativa $u_A(a,\theta) = A\,u(a,\theta) + B(\theta)$, donde A es una constante positiva y $B(\theta)$ es una función acotada de θ.

Demuestra que $u(a,\theta)$ y $u_A(a,\theta)$ son equivalentes, en el sentido de que llevan a la misma decisión óptima.

Calculando directamente,

$$\mathbb{E}\left[u_A(a,\theta) \mid x \right] =$$

$$\mathbb{E}\left[A u(a,\theta) + B(\theta) \mid x \right] =$$

$$A\,\mathbb{E}\left[u(a,\theta) \mid x \right] + \mathbb{E}\left[B(\theta) \mid x \right].$$

donde $\mathbb{E}[B(\theta) \mid x]$ existe porque $B(\theta)$ es acotada.

Por lo tanto, como $A > 0$,

$$\max_{a \in \mathcal{A}} \mathbb{E}\left[u_A(a,\theta) \mid x \right] = \max_{a \in \mathcal{A}} \mathbb{E}\left[u(a,\theta) \mid x \right].$$

2. Demuestra que la función de puntaje cuadrática

$$s(\boldsymbol{q}, E_j) = A\left\{2q_j - \sum_{i=1}^{k} q_i^2\right\} + B \qquad (A > 0; j = 1, 2, \ldots, k)$$

es propia. (Nota que no es local.)

Para todo $l, j \in \{1, \ldots, k\}$,

$$\frac{\partial}{\partial q_l} s(q, E_j) = \frac{\partial}{\partial q_l}\left(A\left\{2q_j - \sum_{i=1}^{k} q_i^2\right\} + B\right)$$

$$= A\{2\delta_{jl} - 2q_l\}.$$

Entonces

$$\frac{\partial}{\partial q_l}\left(\sum_{j=1}^{k} s(q, E_j)\, p_j\right) =$$

$$\sum_{j=1}^{k} A\{2\delta_{jl} - 2q_l\} p_j =$$

$$2A\left(\sum_{j=1}^{k} \delta_{jl} p_j - \sum_{j=1}^{k} q_l p_j\right) =$$

$$2A(p_l - q_l).$$

$\therefore$ Igualando a 0 obtenemos $p_\ell = q_\ell$.

Mas aun,

$$\frac{\partial^2}{\partial q_m \partial q_\ell} \left(\sum_{j=1}^{k} s(q, E_j)\, p_j \right) =$$

$$\frac{\partial}{\partial q_m} \left(2A (p_\ell - q_\ell) \right) = -2A\, \delta_{\ell m}$$

Entonces

$$\frac{\partial^2}{\partial q^2} \left(\sum_{j=1}^{k} s(q, E_j)\, p_j \right) = -2A\, I_k$$

es definida negativa y por lo tanto el punto

critico $q = p$ es maximo.

3. Supón que se requiere una predicción puntual de la variable aleatoria (real y continua) $Y = X_{n+1}$ después de haber observado la muestra $\boldsymbol{x} = (x_1, \ldots, x_n)$. Plantea este problema como un problema de decisión estadístico usando la función de pérdida

$$L(\hat{y}, y) = k_1\,(y - \hat{y})\,\mathrm{I}_{(-\infty,y)}(\hat{y}) + k_2\,(\hat{y} - y)\,\mathrm{I}_{[y,\infty)}(\hat{y}),$$

donde $I_C(\cdot)$ denota a la función indicadora del conjunto C, y k_1 y k_2 son constantes positivas. Demuestra que el predictor óptimo $\hat{y}_*$ es el cuantil de orden $q = k_1/(k_1 + k_2)$ de la distribución predictiva $p(y|\boldsymbol{x})$.

Calculando directamente,

$$L^*_x(\hat{y}) = \int_{-\infty}^{\infty} L(\hat{y}, y)\, p(y|x)\, dy$$

$$= \int_{-\infty}^{\infty} \left[k_1 (y-\hat{y})\, I_{(-\infty,y)}(\hat{y}) + k_2(\hat{y}-y)\, I_{[y,\infty)}(\hat{y}) \right] p(y|x)\, dy$$

$$= \int_{-\infty}^{\infty} \left[k_1 (y-\hat{y})\, I_{(\hat{y},\infty)}(y) + k_2(\hat{y}-y)\, I_{(-\infty,\hat{y}]}(y) \right] p(y|x)\, dy$$

$$= \int_{\hat{y}}^{\infty} k_1 (y-\hat{y})\, p(y|x)\, dy + \int_{-\infty}^{\hat{y}} k_2(\hat{y}-y)\, p(y|x)\, dy$$

Entonces (derivando bajo el signo de la integral)

$$\frac{d}{d\hat{y}} L^*_x(\hat{y}) = -k_1 \int_{\hat{y}}^{\infty} p(y|x)\, dy + k_2 \int_{-\infty}^{\hat{y}} p(y|x)\, dy$$

$$= -k_1 + (k_1 + k_2) \int_{-\infty}^{\hat{y}} p(y|x)\, dy$$

donde la ultima igualdad es porque

$$\int_{\hat{y}_*}^{\infty} p(y|x)\,dy = \left(1 - \int_{-\infty}^{\hat{y}_*} p(y|x)\,dy\right)$$

Igualando a 0 y despejando obtenemos

$$\int_{-\infty}^{\hat{y}_*} p(y|x)\,dy = \frac{k_1}{k_1+k_2}.$$

Mas aun, por el TFC

$$\frac{d^2}{d\hat{y}^2} L_x^*(\hat{y}) = (k_1+k_2)\,p(\hat{y}|x) > 0$$

y por lo tanto $\hat{y}_*$ es minimo.

4. Supón que $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ es una muestra aleatoria de una distribución $p(x|\theta)$, con $\theta > 0$, y sea $p(\theta)$ la distribución inicial de θ. Estos elementos dan lugar a $p(\theta|\boldsymbol{x})$, la distribución final de θ.

Considera ahora la siguiente familia de funciones de pérdida para la estimación puntual de θ, vista como un problema de decisión estadístico:

$$L(\hat{\theta}, \theta) = \frac{(\hat{\theta} - \theta)^2}{\hat{\theta}^r \theta^s}; \qquad \text{donde } r, s = 0, 1, 2.$$

Encuentra el estimador bayesiano óptimo de θ.

$$L^*_x(\hat{\theta}) = \int_\Theta L(\hat{\theta}, \theta)\, p(\theta|x)\, d\theta$$

$$= \int_\Theta (\hat{\theta} - \theta)^2 \hat{\theta}^{-r} \theta^{-s}\, p(\theta|x)\, d\theta$$

Entonces, derivando bajo el signo de la integral

$$\frac{d}{d\hat{\theta}} L^*_x(\hat{\theta}) =$$

$$\int_\Theta \left[2(\hat{\theta} - \theta)\hat{\theta}^{-r} + (\hat{\theta} - \theta)^2 (-r)\hat{\theta}^{-r-1} \right] \theta^{-s}\, p(\theta|x)\, d\theta =$$

$$\int_\Theta \left[2\hat{\theta}^{-r+1} - 2\theta\hat{\theta}^{-r} - r\hat{\theta}^{-r+1} + 2r\hat{\theta}^{-r}\theta - r\theta^2\hat{\theta}^{-r-1} \right] \theta^{-s}\, p(\theta|x)\, d\theta =$$

$$\hat{\theta}^{-r-1} \int_\Theta \left[2\hat{\theta}^2 - 2\theta\hat{\theta} - r\hat{\theta}^2 + 2r\hat{\theta}\theta - r\theta^2 \right] \theta^{-s}\, p(\theta|x)\, d\theta$$

Igualando a 0 obtenemos

$$0 = \int_{\Theta} 2\hat{\theta}^2 \theta^{-s} p(\theta|x)\, d\theta - \int_{\Theta} 2\hat{\theta}\theta^{-s+1} p(\theta|x)\, d\theta$$

$$- \int_{\Theta} r\hat{\theta}^2 \theta^{-s} p(\theta|x)\, d\theta + \int_{\Theta} 2r\hat{\theta}\, \theta^{-s+1} p(\theta|x)\, d\theta$$

$$- \int_{\Theta} r\theta^{-s+2} p(\theta|x)\, d\theta$$

$$= \hat{\theta}^2 (2-r) \int_{\Theta} \theta^{-s} p(\theta|x)\, d\theta +$$

$$\hat{\theta}(2r-2) \int_{\Theta} \theta^{-s+1} p(\theta|x)\, d\theta -$$

$$r \int_{\Theta} \theta^{-s+2} p(\theta|x)\, d\theta$$

$$= \hat{\theta}^2 (2-r)\, \mathbb{E}[\theta^{-s} | x] +$$

$$\hat{\theta}(2r-2)\, \mathbb{E}[\theta^{-s+1} | x] -$$

$$r\, \mathbb{E}[\theta^{-s+2} | x]$$

Resolviendo esta ecuación cuadrática en $\hat{\theta}$ obtenemos lo deseado.

5. *Invarianza.*

a) Calcula $\delta(\hat{\theta},\theta)$, $\Delta_\theta(\hat{\theta})$ y $\hat{\theta}_*$, así como $\delta_\mu(\hat{\mu},\mu)$, $\Delta_\mu(\hat{\mu})$ y $\hat{\mu}_*$, para el caso de la distribución exponencial (ver páginas 62 y 63 de las notas "Inferencia.pdf", correspondientes a la Sección 3 del temario). ¿Se cumple que $\hat{\mu}_* = \mu(\hat{\theta}_*)$?

(b) Calcula también $E[\theta|\mathbf{x}]$ y $E[\mu|\mathbf{x}]$ y nota que $E[\mu|\mathbf{x}] \neq \mu(E[\theta|\mathbf{x}])$.

(c) ¿Qué puedes decir de estos resultados?

a) Calculando directamente,

$$\delta(\hat{\theta},\theta) = \int_0^\infty p(x|\theta) \log\left\{\frac{p(x|\theta)}{p(x|\hat{\theta})}\right\} dx$$

$$= \int_0^\infty \theta e^{-\theta x} \log\left\{\frac{\theta e^{-\theta x}}{\hat{\theta} e^{-\hat{\theta} x}}\right\} dx$$

$$= \int_0^\infty \theta e^{-\theta x} \left(\log\left(\frac{\theta}{\hat{\theta}}\right) + (-\theta x + \hat{\theta} x)\right) dx$$

$$= \int_0^\infty \theta e^{-\theta x} \log\left(\frac{\theta}{\hat{\theta}}\right) dx + \int_0^\infty \theta e^{-\theta x} (-\theta + \hat{\theta}) x \, dx$$

$$= \log\left(\frac{\theta}{\hat{\theta}}\right) + (-\theta + \hat{\theta}) \frac{1}{\theta}$$

$$= \log\left(\frac{\theta}{\hat{\theta}}\right) + \frac{\hat{\theta}}{\theta} - 1 .$$

donde la penúltima igualdad se cumple porque

$$\int_0^\infty \theta e^{-\theta x}\, dx = 1 \quad \& \quad \int_0^\infty x\, e^{-\theta x}\, dx = \frac{1}{\theta^2}.$$

Luego,

$$\Delta_\theta(\hat{\theta}) = \mathbb{E}\left[\delta(\hat{\theta},\theta) \mid x\right]$$

$$= \mathbb{E}\left[\log\left(\frac{\theta}{\hat{\theta}}\right) + \frac{\hat{\theta}}{\theta} - 1 \mid x\right]$$

$$= \mathbb{E}\left[\log\theta \mid x\right] + \hat{\theta}\,\mathbb{E}\left[\frac{1}{\theta} \mid x\right] - \log\hat{\theta} - 1$$

Pero

$$\mathbb{E}\left[\frac{1}{\theta} \mid x\right] = \int_0^\infty \frac{1}{\theta}\, p(\theta \mid x)\, d\theta$$

$$= \int_0^\infty \frac{1}{\theta} \frac{b_x^{a_x}}{\Gamma(a_x)} \theta^{a_x - 1} e^{-b_x\theta}\, d\theta$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \int_0^\infty \theta^{a_x-2} e^{-b_x\theta}\, d\theta$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \frac{\Gamma(a_x - 1)}{b_x^{a_x-1}} = \frac{b_x}{a_x - 1}.$$

De donde,

$$\Delta_0(\hat{\theta}) =$$

$$\mathbb{E}\left[\log\theta \mid x\right] + \hat{\theta}\,\mathbb{E}\left[\frac{1}{\theta} \mid x\right] - \log\hat{\theta} - 1 =$$

$$\mathbb{E}\left[\log\theta \mid x\right] + \hat{\theta}\frac{b_x}{a_x - 1} - \log\hat{\theta} - 1 .$$

Luego,

$$\frac{d}{d\hat{\theta}}\Delta_0(\hat{\theta}) = \frac{b_x}{a_x - 1} - \frac{1}{\hat{\theta}}$$

Igualando a 0 y despejando obtenemos

$$\hat{\theta}_* = \frac{a_x - 1}{b_x} .$$

Mas aún, como $\dfrac{d^2}{d\hat{\theta}^2}\Delta_0(\hat{\theta}) = \dfrac{1}{\hat{\theta}^2} > 0$,

entonces $\hat{\theta}_*$ es mínimo.

Por otro lado, como

$$\delta(\hat{\theta},\theta) = \log\left(\frac{\theta}{\hat{\theta}}\right) + \frac{\hat{\theta}}{\theta} - 1.$$

Entonces,

$$\delta_\mu(\hat{\mu},\mu) = \log\left(\frac{\hat{\mu}}{\mu}\right) + \frac{\mu}{\hat{\mu}} - 1.$$

Además

$$p_\mu(\mu \mid x) = p(\theta(\mu) \mid x)\, |J_\theta(\mu)| \qquad \left(\theta(\mu) = \frac{1}{\mu}\right)$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \left(\frac{1}{\mu}\right)^{a_x - 1} e^{-b_x(1/\mu)} \left|-\frac{1}{\mu^2}\right|$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \mu^{-(a_x+1)} e^{-b_x(1/\mu)}$$

De donde,

$$\Delta_\mu(\hat{\mu}) = \mathbb{E}_{\hat{\mu}}\left[\delta_\mu(\hat{\mu},\mu) \mid x\right]$$

$$= \mathbb{E}\left[\log\left(\frac{\hat{\mu}}{\mu}\right) + \frac{\mu}{\hat{\mu}} - 1 \,\middle|\, x \right]$$

$$= \log \hat{\mu} - \mathbb{E}[\log \mu \mid x] + \frac{1}{\hat{\mu}} \mathbb{E}[\mu \mid x] - 1$$

Pero

$$\mathbb{E}[\mu \mid x] = \int_0^\infty \mu \, p_\mu(\mu \mid x) \, d\mu$$

$$= \int_0^\infty \mu \frac{b_x^{a_x}}{\Gamma(a_x)} \mu^{-(a_x+1)} e^{-b_x(1/\mu)} \, d\mu$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \int_0^\infty \mu^{-a_x} e^{-b_x(1/\mu)} \, d\mu$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \frac{\Gamma(a_x - 1)}{b_x^{a_x - 1}} = \frac{b_x}{a_x - 1} \quad \cdots (1)$$

De donde

$$\Delta_\mu(\hat{\mu}) = \log \hat{\mu} - \mathbb{E}[\log \mu \mid x] + \frac{1}{\hat{\mu}} \mathbb{E}[\mu \mid x] - 1$$

$$= \log \hat{\mu} - \mathbb{E}[\log \mu \mid x] + \frac{1}{\hat{\mu}} \frac{b_x}{a_x - 1} - 1$$

Luego, derivando obtenemos

$$\frac{d}{d\hat{\mu}} \Delta\mu(\hat{\mu}) = \frac{1}{\hat{\mu}} - \frac{1}{\hat{\mu}^2} \frac{b_x}{a_x - 1}.$$

Igualando a 0 y despejando obtenemos

$$\hat{\mu}_* = \frac{b_x}{a_x - 1} = \mu(\hat{\theta}_*).$$

b) Calculando directamente,

$$\mathbb{E}[\theta | x] = \int_0^\infty \theta \, p(\theta | x) \, d\theta$$

$$= \int_0^\infty \theta \frac{b_x^{a_x}}{\Gamma(a_x)} \theta^{a_x - 1} e^{-b_x \theta} \, d\theta$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \int_0^\infty \theta^{a_x} e^{-b_x \theta} \, d\theta$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \frac{\Gamma(a_x + 1)}{b_x^{a_x + 1}} = \frac{a_x}{b_x}$$

Y ya habíamos visto en (1) que $\mathbb{E}[\mu | x] = \frac{b_x}{a_x - 1}$.

6. Sea $X_1, X_2, \ldots, X_n$ una muestra de variables aleatorias i.i.d. de una distribución exponencial $\text{Exp}(x|\theta)$, parametrizada de manera que $E(X|\theta) = 1/\theta$.

a) Utilizando una función de pérdida cuadrática con respecto a $1/\theta$, es decir,

$$L(\hat{\theta}, \theta) = \left(\hat{\theta}^{-1} - \theta^{-1}\right)^2$$

y suponiendo una distribución inicial conjugada para θ, encuentra el estimador bayesiano óptimo de θ.

b) Dado que θ es un parámetro de escala, podría ser más apropiado utilizar una función de pérdida que sea invariante ante cambios de escala. Encuentra el estimador bayesiano óptimo de θ, utilizando ahora la función de pérdida

$$L^*(\hat{\theta}, \theta) = \theta^2 \left(\hat{\theta}^{-1} - \theta^{-1}\right)^2$$

y suponiendo, al igual que antes, una distribución inicial conjugada para θ.

a) Como suponemos una distribución inicial conjugada para $\theta \implies \theta \sim \text{Gamma}(\theta | a, b)$ p.a. $a, b > 0$. Mas aún,

$$p(\theta | x) = \frac{b_x^{a_x}}{\Gamma(a_x)} \theta^{a_x - 1} e^{-b_x \theta}. \quad (\text{cf. p62 "Inferencia"})$$

con $a_x = a + n$, $b_x = b + \sum_{i=1}^n x_i$. Calculando directamente,

$$L_x^*(\hat{\theta}) = \int_0^\infty L(\hat{\theta}, \theta) \, p(\theta | x) \, d\theta$$

$$= \int_0^\infty (\hat{\theta}^{-1} - \theta^{-1})^2 \frac{b_x^{a_x}}{\Gamma(a_x)} \theta^{a_x - 1} e^{-b_x \theta} \, d\theta$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \int_0^\infty \left(\frac{1}{\theta^2} - \frac{2}{\hat{\theta} \theta} + \frac{1}{\hat{\theta}^2} \right) \theta^{a_x - 1} e^{-b_x \theta} \, d\theta$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \int_0^\infty \left(\theta^{a_x-3} e^{-b_x\theta} - \frac{2}{\hat\theta} \theta^{a_x-2} e^{-b_x\theta} + \frac{1}{\hat\theta^2} \theta^{a_x-1} e^{-b_x\theta} \right) d\theta$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \int_0^\infty \theta^{a_x-3} e^{-b_x\theta} \, d\theta$$

$$- \frac{b_x^{a_x}}{\Gamma(a_x)} \frac{2}{\hat\theta} \int_0^\infty \theta^{a_x-2} e^{-b_x\theta} \, d\theta$$

$$+ \frac{b_x^{a_x}}{\Gamma(a_x)} \frac{1}{\hat\theta^2} \int_0^\infty \theta^{a_x-1} e^{-b_x\theta} \, d\theta$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \frac{\Gamma(a_x-2)}{b_x^{a_x-2}}$$

$$- \frac{2}{\hat\theta} \frac{b_x^{a_x}}{\Gamma(a_x)} \frac{\Gamma(a_x-1)}{b_x^{a_x-1}} + \frac{1}{\hat\theta^2}$$

$$= \frac{b_x^2}{(a_x-1)(a_x-2)} - \frac{2}{\hat\theta} \frac{b_x}{a_x-1} + \frac{1}{\hat\theta^2}$$

Derivando respecto de $\hat\theta$ obtenemos

$$\frac{d}{d\hat\theta} L_x^*(\hat\theta) = \frac{2}{\hat\theta^2} \frac{b_x}{a_x-1} - \frac{2}{\hat\theta^3}$$

Igualando a 0 y despejando obtenemos $\hat\theta_* = \dfrac{a_x-1}{b_x}$.

Derivando de nuevo y evaluando en $\hat\theta_*$ vemos que es min.

b) Procediendo de la misma manera que en el inciso anterior es fácil ver que

$$L_x^*(\hat{\theta}) = \int_0^\infty L(\hat{\theta},\theta)\, p(\theta|x)\, d\theta$$

$$= \int_0^\infty \theta^2 (\hat{\theta}^{-1} - \theta^{-1})^2 \frac{b_x^{a_x}}{\Gamma(a_x)} \theta^{a_x-1} e^{-b_x\theta}\, d\theta$$

$$= \frac{b_x^{a_x}}{\Gamma(a_x)} \int_0^\infty \theta^{a_x-1} e^{-b_x\theta}\, d\theta$$

$$- \frac{b_x^{a_x}}{\Gamma(a_x)} \frac{2}{\hat{\theta}} \int_0^\infty \theta^{a_x} e^{-b_x\theta}\, d\theta$$

$$+ \frac{b_x^{a_x}}{\Gamma(a_x)} \frac{1}{\hat{\theta}^2} \int_0^\infty \theta^{a_x+1} e^{-b_x\theta}\, d\theta$$

$$= 1 - \frac{2}{\hat{\theta}} \frac{b_x^{a_x}}{\Gamma(a_x)} \frac{\Gamma(a_x+1)}{b_x^{a_x+1}} + \frac{1}{\hat{\theta}^2} \frac{b_x^{a_x}}{\Gamma(a_x)} \frac{\Gamma(a_x+2)}{b_x^{a_x+2}}$$

$$= 1 - \frac{2}{\hat{\theta}} \frac{a_x}{b_x} + \frac{1}{\hat{\theta}^2} \frac{(a_x+1)a_x}{b_x^2}$$

Derivando

$$\frac{d}{d\hat{\theta}} L_x^*(\hat{\theta}) = \frac{2}{\hat{\theta}^2} \frac{a_x}{b_x} - \frac{2}{\hat{\theta}^3} \frac{(a_x+1)a_x}{b_x^2}.$$

Igualando a 0 y despejando obtenemos $\hat{\theta}_x = \frac{a_x+1}{b_x}$.

Derivando de nuevo y evaluando en $\hat{\theta}_x$ vemos que es min.

c) Recordemos que $a_x = a+n$. Por lo tanto,

$$\frac{\frac{a_x-1}{b_x}}{\frac{a_x+1}{b_x}} = \frac{a+n-1}{a+n+1} \xrightarrow{n\to\infty} 1$$

En este sentido, asintoticamente son igual de buenos.

7. Sea $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ una muestra aleatoria de observaciones independientes con distribución desconocida $f(x)$. Supón que $f(x)$ sólo puede ser uno de los siguientes modelos paramétricos:

$$\mathcal{M}_1 : f(x) = p_1(x|\theta_1),\;\; \theta_1 \in \Theta_1;\; \text{con distribución inicial } \; p_1(\theta_1), \text{propia};$$

$$\mathcal{M}_2 : f(x) = p_2(x|\theta_2),\;\; \theta_2 \in \Theta_2;\; \text{con distribución inicial } \; p_2(\theta_2),\; \text{propia}.$$

Supón también que, *a priori*, se considera que estos dos modelos son igualmente probables. El problema consiste en seleccionar uno de ellos.[1]

Sea μ una variable dicotómica que indica cuál de los dos modelos es el correcto. Entonces podemos escribir

$$f(x) = p(x|\theta_\mu, \mu) = \begin{cases} p_1(x|\theta_1) & \text{si } \mu = 1 \\ p_2(x|\theta_2) & \text{si } \mu = 2 \end{cases}$$

Dado que el valor de μ es desconocido, podemos considerarlo un parámetro más en este modelo extendido.

(a) Encuentra $p(\boldsymbol{x}|\mu)$ y, por lo tanto, la función de verosimilitud de μ;

(b) Encuentra la distribución final de μ.

(c) Supón ahora que se desea contrastar las hipótesis

$$H_1 : \mu = 1 \;\text{ vs }\; H_2 : \mu = 2$$

con base en la función de utilidad

$$u(\text{“Elegir } H_i\text{”}, \text{“}H_j \text{ es cierta”}) = I(i = j), \quad i, j = 1, 2,$$

donde $I(\cdot)$ denota la función indicadora. ¿Cuál es la regla de decisión para este problema?

a) Calculando directamente,

$$p(x|\mu) = \int_{\Theta_\mu} p(x, \theta_\mu | \mu)\, d\theta_\mu$$

$$= \int_{\Theta_\mu} p(x|\theta_\mu, \mu)\, p(\theta_\mu|\mu)\, d\theta_\mu$$

$$= \int_{\Theta_\mu} p_\mu(x|\theta_\mu)\, p_\mu(\theta_\mu)\, d\theta_\mu$$

b) Claramente,

$$p(x) = p(x|\mu=1)\,p(\mu=1) + p(x|\mu=2)\,p(\mu=2)$$

$$= \frac{1}{2} \left(p(x|\mu=1) + p(x|\mu=2) \right)$$

pues $p(\mu=i) = 1/2$, $i=1,2$. Entonces,

$$p(\mu|x) = \frac{p(x|\mu)\, p(\mu)}{p(x)}$$

$$= \frac{p(x|\mu)}{p(x|\mu=1) + p(x|\mu=2)}.$$

c) Calculando directamente

$$\bar{u}(a_1) = \underbrace{u(a_1, E_1)}_{1} \Pr(E_1|x) + \underbrace{u(a_1, E_2)}_{0} \Pr(E_2|x)$$

$$= \Pr(E_1|x) = p(\mu=1|x)$$

$$= \frac{p(x|\mu=1)}{p(x|\mu=1) + p(x|\mu=2)}.$$

$\therefore$ Rechazamos H_1 sii $p(x|\mu=1) < p(x|\mu=2)$.

8. El problema de "clasificación supervisada" es un problema de decisión estadístico. Supongamos que tenemos una muestra de n observaciones (X_1, G_1), $(X_2, G_2), \ldots, (X_n, G_n)$, donde cada $G_i \in \{1, 2, \ldots, K\}$ indica que la observación $X_i \in \mathbb{R}_+$ pertenece a uno de K grupos distintos. Supongamos ahora que

$$p(x|G = g) = \text{Gamma}(x|\alpha_g, \beta_g) \qquad (g = 1, 2, \ldots, K)$$

y que

$$\Pr(G = g) = 1/K,$$

donde $\alpha_g > 0$ y $\beta_g > 0$ son parámetros conocidos.

El problema de clasificación consiste en lo siguiente: dada una nueva observación $X_* = x_*$, determinar de cuál de los K grupos proviene. Desde el punto de vista estadístico, el problema es entonces predecir el valor de G_* asociado con X_*.

(a) Plantea y resuelve este problema como un problema de decisión, usando la función de pérdida

$$L(g_*, g) = 0 \qquad (\text{si } g_* = g)$$

$$L(g_*, g) = 1 \qquad (\text{si } g_* \neq g).$$

(b) ¿Cómo se resolvería este problema si los valores de los parámetros $\beta_1, \beta_2, \ldots, \beta_K$ fueran desconocidos?

a) Calculando directamente

$$L_x^*(g_*) = \sum_{g=1}^{K} L(g_*, g)\, p(g|x)$$

$$= \sum_{g \neq g_*} p(g|x)$$

$$= 1 - p(g_*|x).$$

Entonces, $\hat{g}_* = \underset{g_*}{\text{argmax}}\; p(g_*|x) = \text{moda}\; p(g_*|x)$

De hecho, en este caso $p(g_*|x) \propto p(x|g_*) \underbrace{p(g_*)}_{1/K}$

$\propto p(x|g_*) = \text{Gamma}(x|\alpha_{g_*}, \beta_{g_*}).$