Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística
Primer semestre 2023

EYP2807/EYP280I - Métodos Bayesianos
Pauta Interrogación 1

Profesora: Ana María Araneda
Ayudante: Josefa Silva

1. Para cada una de las siguientes afirmaciones, indique si ella es verdadera o falsa. Si falsa, justifique, y si es verdadera, aporte un trozo de información adicional.

a) Bajo el enfoque Bayesiano, la predicción de nuevas observaciones debe realizarse en base a su distribución predictiva. **[1,0]**

Solución: *Falso. Debe realizarse en base a su distribución predictiva condicional, o a posteriori, puesto que, toda decisión debe considerar toda la información disponible.*

b) Para conocer la distribución predictiva a posteriori de una secuencia de variables aleatorias permutables, $x_{n+1}, \ldots, x_{n+m}$, solo se requiere conocer la distribución a posteriori del parámetro que define la mezcla, dada la muestra disponible, $x_1, \ldots, x_n$. **[1,0]**

Solución: *Falso. La distribución predictiva a posteriori de una secuencia $x_{n+1}, \ldots, x_{n+n}$ corresponde a*

$$p(x_{n+1}, \ldots, x_{n+m}|x_1, \ldots, x_n) = \int_\Theta \prod_{j=n+1}^{n+m} p(x_j|\theta)p(\theta|x)d\theta,$$

Luego, también se requiere conocer la verosimilitud de cada observación de la secuencia, $p(x_j|\theta)$.

c) Dado un problema de predicción, no existe una única secuencia suficiente ni una única secuencia suficiente minimal de estadísticos. **[1,0]**

Solución: *Verdadero. Toda función 1-1 de una secuencia suficiente también es suficiente. También se pueden agregar estadísticos y seguirá siendo suficiente. Para minimalidad, toda función 1-1 de una secuencia suficiente minimal también es suficiente minimal.*

d) En teoría de decisiones, el objetivo es realizar la acción, o tomar la decisión, que minimiza la funcion de pérdida. **[1,0]**

Solución: *<u>Falso</u>. El objetivo es realizar la acción, o tomar la decisión, que minimiza el valor esperado de la función de pérdida.*

e) En estimación puntual y bajo una función de pérdida cuadrática, el estimador de Bayes siempre corresponde a la media de las observaciones. **[1,0]**

Solución: *<u>Falso</u>. El estimador de Bayes bajo función de pérdida cuadrática corresponde a la media de la distribución a posteriori (de un parámetro o de una secuencia de observaciones futuras), la que puede o no corresponder a la media de las observaciones.*

f) Cuando la verosimilitud de las observaciones es Normal, la distribución a priori de su media es Normal, se considera una función de pérdida de valor absoluto, y al sobreestimar se incurre en la misma perdida que al subestimar, el estimador de Bayes de la media corresponde a un promedio ponderado entre el promedio muestral y la media a priori. **[1,0]**

Solución: *<u>Verdadero</u>. Para la función de pérdida descrita, el estimador de Bayes corresponde a la mediana a posteriori. En el modelo Normal-Normal, la distribución a posteriori es Normal donde la media es un promedio ponderado entre el promedio muestral y la media a priori. Dado que la distribución Normal es simétrica, la mediana a posteriori es igual a la media a posteriori, con lo que se cumple la afirmación realizada.*

[1,0] punto base

2. Sea $x_1, x_2, \ldots,$ una secuencia permutable de variables aleatorias con verosimilitud Weibull(α_0, β), con $\beta > 0$, y$\alpha_0 > 0$ conocido.

a) Explique, con sus propias palabras, lo que significa que una secuencia de estadísticos sea suficiente paramétrica. **[1,0]**

Solución: *Una secuencia es suficiente paramétrica si la distribución a posteriori del parámetro θ, dada la muestra, $x_1, \ldots, x_n$, y el estadístico de la secuencia, t_n, solo depende de este último.*

Esto significa que la distribución del parámetro deja de depender de la muestra si se conoce el valor del estadístico suficiente asociado a la secuencia.

Equivalentemente, significa que el parámetro es condicionalmente independiente de la muestra dado el estadístico suficiente asociado a la secuencia.

También significa que el estadístico contiene toda la información de la muestra necesaria para determinar la distribución a posteriori del parámetro.

(cualquiera de las tres últimas afirmaciones, o similares, dada en palabras, recibe puntaje)

b) Explique, con sus propias palabras, lo que significa que una secuencia de estadísticos sea suficiente minimal. [**1,0**]

Solución: *Una secuencia de estadísticos es suficiente minimal si contiene toda la información suficiente para determinar las distribuciones a posteriori (de un parámetro o de observaciones futuras de la secuencia) y no contiene información innecesaria. Esto se traduce en que, para todo otro estadístico suficiente, el estadístico suficiente minimal se puede obtener a partir de él.*

c) Encuentre una secuencia de estadísticos que sea suficiente minimal para β. Justifique sus pasos y argumentos. [**4,0**]

Solución: *Se utilizará teorema visto en clases.* [**0,5**]

Sean $(x_1, \ldots, x_n)$ *e* $(y_1, \ldots, y_n)$ *dos valores posibles para los elementos* $1, \ldots, n$ *de la secuencia permutable, y considere la razón:*

$$\begin{aligned}
\frac{p\left(x_1, \ldots, x_n | \beta\right)}{p\left(y_1, \ldots, y_n | \beta\right)} &= \frac{\prod_{i=1}^{n} \alpha_0 \beta (\beta x_i)^{\alpha_0 - 1} \cdot exp\left\{-(\beta x_i)^{\alpha_0}\right\} I_{\mathbb{R}^+}(x_i)}{\prod_{i=1}^{n} \alpha_0 \beta (\beta y_i)^{\alpha_0 - 1} \cdot exp\left\{-(\beta y_i)^{\alpha_0}\right\} I_{\mathbb{R}^+}(y_i)} \\
&= \frac{\prod_{i=1}^{n} \cancel{\alpha_0 \beta (\beta)^{\alpha_0 - 1}} \cdot (x_i)^{\alpha_0 - 1} \cdot exp\left\{-(\beta x_i)^{\alpha_0}\right\} I_{\mathbb{R}^+}(x_i)}{\prod_{i=1}^{n} \cancel{\alpha_0 \beta (\beta)^{\alpha_0 - 1}} \cdot (y_i)^{\alpha_0 - 1} \cdot exp\left\{-(\beta y_i)^{\alpha_0}\right\} I_{\mathbb{R}^+(y_i)}} \\
&= \frac{\prod_{i=1}^{n} \left(x_i^{\alpha_0 - 1} I_{\mathbb{R}_+}(x_i)\right)}{\prod_{i=1}^{n} \left(y_i^{\alpha_0 - 1} I_{\mathbb{R}_+}(y_i)\right)} \cdot exp\left\{-\sum_{i=1}^{n} (\beta x_i)^{\alpha_0} + \sum_{i=1}^{n} (\beta y_i)^{\alpha_0}\right\} \\
&= C_{x,y} \ \exp\left\{-\sum_{i=1}^{n} (\beta x_i)^{\alpha_0} + \sum_{i=1}^{n} (\beta y_i)^{\alpha_0}\right\} \\
&= C_{x,y} \ \exp\left\{-\beta^{\alpha_0} \sum_{i=1}^{n} x_i^{\alpha_0} + \beta^{\alpha_0} \sum_{i=1}^{n} y_i^{\alpha_0}\right\} \\
&= C_{x,y} \ \exp\left\{-\beta^{\alpha_0} \left(\sum_{i=1}^{n} x_i^{\alpha_0} - \sum_{i=1}^{n} y_i^{\alpha_0}\right)\right\}, \qquad \mathbf{[1{,}5]}
\end{aligned}$$

donde $C_{x,y}$ *no depende de* β.

La razón estudiada no depende de β ***si y sólo si***: [**0,5**]

$$\sum_{i=1}^{n} x_i^{\alpha_0} = \sum_{i=1}^{n} y_i^{\alpha_0} \quad \textbf{[1,0]}$$

Luego, $t(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i^{\alpha_0}$ *forma una secuencia suficiente minimal para* β*.* **[0,5]**
[1,0] punto base

3. Considere una secuencia permutable $x_1, x_2, \ldots$ con verosimilitud $Normal(\theta, \sigma^2)$, con σ conocido, y una distribución a priori $Normal(\mu, \tau^2)$ para θ, con μ y τ conocidos. Considere una muestra de esta secuencia, dada por $x = (x_1, \ldots, x_n)$. Denote la media y varianza a posteriori de θ por μ_p y τ_p^2, respectivamente.

a) Muestre que la distribución predictiva a posteriori de x_{n+1}, dada la muestra $x = (x_1, \ldots, x_n)$ corresponde a una distribución Normal de media μ_p y varianza $(\sigma^2 + \tau_p^2)$. **[3,0]**

Solución *Se tiene el siguiente modelo:*

$$\begin{aligned} x_1, x_2, \ldots | \theta &\sim Normal(\theta, \sigma^2) \\ \theta &\sim Normal(\mu, \tau^2) \\ \theta | x_1, \ldots, x_n &\sim Normal(\mu_p, \tau_p^2), \end{aligned}$$

con σ, μ, τ^2 *conocidos.*

Luego, la distribución predictiva a posteriori de x_{n+1}*, dada la muestra* $x_1, \ldots, x_n$ *corresponde a:*

$$
\begin{aligned}
p(x_{n+1}|x_1,\ldots,x_n) &= \int_{\mathbb{R}_+} p(x_{n+1}|\theta)\; p(\theta|x_1,\ldots,x_n)\; d\theta \\
&= \int_{\mathbb{R}^+} \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{1}{2\sigma^2}\left(x_{n+1}-\theta\right)^2\right\} \left(\frac{1}{2\pi\tau_p^2}\right)^{1/2} \exp\left\{-\frac{1}{2\tau_p^2}\left(\theta-\mu_p\right)^2\right\} d\theta \\
&= \int_{\mathbb{R}^+} \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{1}{2\sigma^2}\left(x_{n+1}^2-2x_{n+1}\theta+\theta^2\right)\right\} \times \\
&\qquad \left(\frac{1}{2\pi\tau_p^2}\right)^{1/2} \exp\left\{-\frac{1}{2\tau_p^2}\left(\theta^2-2\theta\mu_p+\mu_p^2\right)\right\} d\theta \\
&\propto \int_{\mathbb{R}^+} \exp\left\{-\frac{1}{2\sigma^2}\left(x_{n+1}^2-2x_{n+1}\theta+\theta^2\right)\right\} \exp\left\{-\frac{1}{2\tau_p^2}\left(\theta^2-2\theta\mu_p\right)\right\} d\theta \\
&\propto \int_{\mathbb{R}^+} \exp\left\{-\frac{1}{2\sigma^2}x_{n+1}^2-\frac{1}{2\sigma^2}\cdot(-2\theta)x_{n+1}-\frac{1}{2\sigma^2}\theta^2\right\} \exp\left\{-\frac{1}{2\tau_p^2}\theta^2-\frac{1}{2\tau_p^2}\cdot(-2\mu_p)\theta \right. \\
&= \exp\left\{-\frac{1}{2\sigma^2}x_{n+1}^2\right\} \int_{\mathbb{R}^+} \exp\left\{-\frac{1}{2}\theta^2\left(\frac{1}{\sigma^2}+\frac{1}{\tau_p^2}\right)-\frac{1}{2}(-2)\theta\left(\frac{x_{n+1}}{\sigma^2}+\frac{\mu_p}{\tau_p^2}\right)\right\} d\theta \\
&= \exp\left\{-\frac{1}{2\sigma^2}x_{n+1}^2\right\} \times \\
&\underbrace{\int_{\mathbb{R}^+} \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\tau_p^2}\right)\left[\theta^2-2\left(\frac{x_{n+1}}{\sigma^2}+\frac{\mu_p}{\tau_p^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\tau_p^2}\right)^{-1}\theta\right]\right\} d\theta}_{\text{Kernel de una Normal (**)}}
\end{aligned}
$$

(**) con media $\left(\frac{x_{n+1}}{\sigma^2}+\frac{\mu_p}{\tau_p^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\tau_p^2}\right)^{-1}$ *y varianza* $\left(\frac{1}{\sigma^2}+\frac{1}{\tau_p^2}\right)^{-1}$

Luego, para que la integral sea 1, falta agregar el término que complete el cuadrado:

$$
\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\tau_p^2}\right)\left(\frac{x_{n+1}}{\sigma^2}+\frac{\mu_p}{\tau_p^2}\right)^2\left(\frac{1}{\sigma^2}+\frac{1}{\tau_p^2}\right)^{-2}\right\}
$$

Recordar que se debe multiplicar y dividir por el término faltante.

Por lo tanto,

$$
\begin{aligned}
p(x_{n+1}|x_1,\dots,x_n) &\propto \exp\left\{-\frac{1}{2\sigma^2}x_{n+1}^2+\frac{1}{2}\left(\frac{x_{n+1}}{\sigma^2}+\frac{\mu_p}{\tau_p^2}\right)^2\left(\frac{1}{\sigma^2}+\frac{1}{\tau_p^2}\right)^{-1}\right\}\\
&\propto \exp\left\{-\frac{1}{2\sigma^2}x_{n+1}^2+\frac{1}{2}\left(\frac{x_{n+1}}{\sigma^4}+2\frac{\mu_p}{\sigma^2\tau_p^2}x_{n+1}+\frac{\mu_p^2}{\tau_p^4}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\tau_p^2}\right)^{-1}\right\}\\
&= \exp\left\{-\frac{1}{2\sigma^2}x_{n+1}^2+\frac{1}{2}\left(\frac{x_{n+1}^2}{\sigma^4}+2\frac{\mu_p}{\sigma^2\tau_p^2}x_{n+1}+\frac{\mu_p^2}{\tau_p^4}\right)\left(\frac{\sigma^2\tau_p^2}{\sigma^2+\tau_p^2}\right)\right\}\\
&\propto \exp\left\{-\frac{1}{2}x_{n+1}^2\left(\frac{1}{\sigma^2}-\frac{\cancel{\sigma^2}\tau_p^2}{\sigma^2+\tau_p^2}\cdot\frac{1}{\sigma^4}\right)-\frac{1}{2}\cdot(-2)\left(\frac{\cancel{\sigma^2\tau_p^2}}{\sigma^2+\tau_p^2}\right)\frac{\mu_p}{\cancel{\sigma^2\tau_p^2}}x_{n+1}\right\}\\
&= \exp\left\{-\frac{1}{2\sigma^2}x_{n+1}^2\left(1-\frac{\tau_p^2}{\sigma^2+\tau_p^2}\right)-\frac{1}{2}\cdot(-2)\frac{\mu_p}{\sigma^2+\tau_p^2}x_{n+1}\right\}\\
&= \exp\left\{-\frac{1}{2\sigma^2}x_{n+1}^2\left(\frac{\sigma^2+\tau_p^2-\tau_p^2}{\sigma^2+\tau_p^2}\right)-\frac{1}{2}\cdot(-2)\frac{\mu_p}{\sigma^2+\tau_p^2}x_{n+1}\right\}\\
&= \exp\left\{-\frac{1}{2\sigma^2}x_{n+1}^2\left(\frac{\sigma^2}{\sigma^2+\tau_p^2}\right)-\frac{1}{2}\cdot(-2)\frac{\mu_p}{\sigma^2+\tau_p^2}x_{n+1}\right\}\\
&= \exp\left\{-\frac{1}{2(\sigma^2+\tau_p^2)}x_{n+1}^2-\frac{1}{2}\cdot(-2)\frac{\mu_p}{\sigma^2+\tau_p^2}x_{n+1}\right\}\\
&= \exp\left\{-\frac{1}{2(\sigma^2+\tau_p^2)}\left(x_{n+1}^2-2x_{n+1}\frac{\mu_p}{\sigma^2+\tau_p^2}(\sigma^2+\tau_p^2)\right)\right\}\\
&= \exp\left\{-\frac{1}{2(\sigma^2+\tau_p^2)}\left(x_{n+1}^2-2x_{n+1}\mu_p\right)\right\},
\end{aligned}
$$

que corresponde al kernel de una distribución Normal, de media μ_p y varianza $(\sigma^2+\tau_p^2)$, que es lo que se pide demostrar.

b) La presencia de mercurio en los peces comestibles por el ser humano es un factor de riesgo para la salud. La OMS ha establecido como límite una concentración de 5 ppm (partes por millón). Para estudiar la concentración de mercurio en róbalos de lago, se cuenta con una muestra de 10 de estos peces, escogidos de manera aleatoria, cuya concentración promedio de mercurio resulto ser de 1,23 ppm.

Utilice el resultado en el apartado anterior para encontrar la distribución predictiva a posteriori de la concentración de mercurio en un nuevo róbalo escogido de manera aleatoria desde el mismo lago que los peces de la muestra. Asuma el modelo Normal - Normal, con una distribución a priori de media 5 ppm y coeficiente de variación 2 %. Asuma además que la desviación estándar de la concentración de mercurio en estos peces es de 0,08 ppm. **[1,0]**

Solución: *Se tiene que:*

- Límite $= 5ppm$
- $n = 10$
- $\bar{x} = 1{,}23ppm$
- $\mu = 5ppm$
- $\tau/\mu = 0{,}02$
- $\sigma = 0{,}08ppm$
- $\tau = 0{,}02 \cdot 5 = 0{,}1$ **[0,2]**

Luego,

$$\mu_p = w \cdot \bar{x} + (1 - w)\mu,$$

donde,

$$w = \frac{\tau^2}{\tau^2 + \sigma^2/n} = \frac{0{,}1^2}{0{,}1^2 + 0{,}08^2/10} = 0{,}94. \qquad \textbf{[0,3]}$$

Por lo tanto,

$$\begin{aligned} \mu_p &= 0{,}94 \cdot 1{,}23 + 0{,}06 \cdot 5 = 1{,}46 \\ \tau_p^2 &= \frac{\sigma^2\tau^2}{n\tau^2 + \sigma^2} = \frac{0{,}08^2 \cdot 0{,}1^2}{10 \cdot 0{,}1^2 + 0{,}08^2} = 0{,}0006. \qquad \textbf{[0,3]} \end{aligned}$$

Por consiguiente:

$$x_{n+1}|x_1, \ldots, x_n \sim Normal(1{,}46, (0{,}0006 + 0{,}08^2)) \qquad \textbf{[0,2]}$$

c) Se desea obtener una predicción de la concentración de mercurio en este nuevo róbalo. Para ello, asuma una función de pérdida de valor absoluto. Proponga valores para las constantes de crecimiento de la perdida, considerando la peligrosidad de las altas concentraciones de mercurio. Entregue la predicción pedida en base a las constantes que propuso. ¿Cree usted que este nuevo róbalo presenta riesgo para la salud de las personas debido a su concentración de mercurio? Explique. **[2,0]**

Solución: *Se tiene que la función de pérdida de valor absoluto es:*

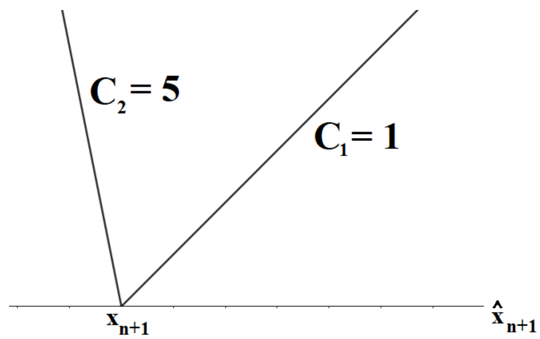
$$l(a, \theta) = c_1(a - \theta)I_{\theta \leq a(a)} + c_2(\theta - a)I_{\theta > a(a)}$$

con $c_1 > 0$ *y* c_2*, donde el estimador de Bayes corresponde al cuantil, a, tal que:*

$$P(\theta \leq a) = \frac{c_2}{c_1 + c_2}$$

Como se debe considerar la peligrosidad de las altas concentraciones de mercurio, en el contexto de función de pérdida, subestimar es más grave que sobreestimar, por lo que la condición que se debe tomar es: $c_2 > c_1$. [**0,5**]

Ahora, considerando $c_2 = 5$ *y* $c_1 = 1$ *(arbitrarios, que cumplan* $c_2 > c_1$*), se tiene:*



Luego, el estimador de Bayes de una nueva observación corresponde al percentil $\frac{5}{5+1} = \frac{5}{6} = 0,83$ [**0,4**] *de la distribución* $Normal(1{,}46, (0{,}0006 + 0{,}08^2))$: [**0,4**]

$$\begin{aligned} q_{0,83} &= 1{,}46 + z_{0,83} \cdot \sqrt{(0{,}0006 + 0{,}08^2)} \\ &= 1{,}46 + 0{,}95 \cdot \sqrt{(0{,}0006 + 0{,}08^2)} \\ &= 1{,}54 \; ppm \qquad \textbf{[0,4]} \end{aligned}$$

Si se realiza con el comando $qnorm(0{,}83, 1{,}46, \sqrt{(0{,}0006 + 0{,}08^2)}$ ***obtiene*** [**0,3**] *en vez de* [**0,4**].

La predicción es menor a 5 ppm, lo que sugiere que el nuevo róbalo no presenta riesgo para la salud de las personas debido a su concentración de mercurio. [**0,3**]

[1,0] punto base