Inferencia Estadística — 20 de diciembre de 2024 — primer parcial.

Instrucciones.

- Primera parte: TEORÍA. Ordenador apagado. Tienes hasta una hora para resolverla y entregar los folios con tu respuesta.
- Segunda parte: PROBLEMAS. En cuanto entregues la teoría, podrás encender el ordenador exclusivamente para usar R y acceder al campus virtual. No hay una "tarea" en el campus virtual para subir documentos. Has de entregar todas tus respuestas en papel: para un gráfico, esbózalo a mano dejando claro lo que quieras destacar; para código fuente, escríbelo con la mayor precisión posible. Has de entregar los folios en dos bloques separados: problema 1; problema 2.
- Justifica todas las respuestas.

Teoría (4 puntos).

1. A una inteligencia artificial se le planteó el siguiente problema:

   > **Pregunta:** *Sea X una variable aleatoria con densidad $f(x, \lambda) = \lambda * exp(-\lambda * (x - \lambda))\ I(x > \lambda)$. ¿Puedes calcular el estimador máximo-verosímil de λ?*

   Analiza si la respuesta es correcta y en caso contrario comenta los posibles fallos:

   > **Respuesta:** *Para calcular el estimador de máxima verosimilitud (EMV) del parámetro λ para una variable aleatoria X con densidad:*
   >
   > $$f(x|\lambda) = \lambda \exp(-\lambda(x - \lambda)) \quad (x > \lambda)$$
   >
   > *, seguimos estos pasos:*
   >
   > a) *Función de Verosimilitud*
   >
   > $$L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda(x_i - \lambda)) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n}(x_i - \lambda)\right)$$
   >
   > b) *Logaritmo de la Función de Verosimilitud*
   >
   > $$\log L(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n}(x_i - \lambda) = n \log \lambda - \lambda \left(\sum_{i=1}^{n} x_i - n\lambda\right)$$
   >
   > c) *Maximización de la Función de Verosimilitud*
   > *Derivamos respecto a λ y encontramos el valor que hace que la derivada sea cero:*
   >
   > $$\frac{d}{d\lambda} \log L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i + 2n\lambda = 0$$
   >
   > $$\frac{n}{\lambda} + 2n\lambda = \sum_{i=1}^{n} x_i \Longleftrightarrow \lambda(1 + 2\lambda) = \frac{1}{n}\sum_{i=1}^{n} x_i$$
   >
   > ***Pregunta:*** *¿Puedes resolver la ecuación $\lambda(1 + 2\lambda) = \frac{1}{n}\sum_{i=1}^{n} x_i$ en λ?*
   > ***Respuesta:*** *Para resolver la ecuación podemos reescribirla en términos de una ecuación cuadrática estándar.*
   >
   > $$2\lambda^2 + \lambda - \frac{1}{n}\sum_{i=1}^{n} x_i = 0$$

Ahora, tenemos una ecuación cuadrática en la forma estándar y obtenemos dos posibles soluciones:

$$\lambda = \frac{-1 \pm \sqrt{1 + \frac{8}{n}\sum_{i=1}^{n} x_i}}{4}$$

Dado que λ representa una cantidad física y debe ser positiva, seleccionaremos la solución positiva:

$$\lambda = \frac{-1 + \sqrt{1 + \frac{8}{n}\sum_{i=1}^{n} x_i}}{4}$$

2. Explica el procedimiento de *bootstrap* básico para calcular intervalos de confianza.

3. Define partición suficiente y minimal. Demuestra que la partición

$$\vec{x} \sim \vec{y} \iff \frac{L(\vec{x}, \theta)}{L(\vec{y}, \theta)} \text{ independiente de } \theta$$

es suficiente y minimal.

4. Define la región crítica de un contraste de hipótesis y el nivel de significación α.

Problema 1 (3 puntos). Cierta ceca fabrica monedas de forma que la probabilidad de que salga cara si tiramos al aire la moneda i-ésima es p_i. Es decir, distintas monedas tienen distinta probabilidad de cara. Cada p_i puede estimarse con mucha precisión, por lo que dada una muestra concreta de n monedas sus p_i se suponen conocidas ($i = 1, \ldots, n$). Las p_i provienen de una población beta $\beta(q, q)$, es decir, con ambos parámetros iguales. Se dispone de una realización muestral de tamaño $n = 10$, a saber, $(p_1, \ldots, p_{10})$ = `(0.449, 0.515, 0.432, 0.526, 0.433, 0.539, 0.560, 0.546, 0.476, 0.630)`.

a) Halla un estadístico suficiente minimal para q ($0{,}6$ puntos).
b) Halla una estimación de q por el método de máxima verosimilitud ($0{,}8$ puntos).
c) Halla una estimación de q por el método de los momentos ($0{,}8$ puntos).
d) Halla un intervalo de confianza para q a nivel $95\,\%$ ($0{,}8$ puntos).

Ten en cuenta que, si X sigue una distribución $\beta(p, q)$, entonces tiene función de densidad $f(x) = x^{p-1} \cdot (1-x)^{q-1}/B(p, q)$, donde B es la función **`beta`** en R; esperanza $E(X) = p/(p+q)$ y varianza $Var(X) = p \cdot q/[(p+q)^2 \cdot (p+q+1)]$.

Problema 2 (3 puntos). Se quiere estudiar el tiempo de vida (en años) de los dispositivos eléctricos de la marca Chispazo. Para ello se estudia una muestra de 100 de estos dispositivos eléctricos, recogiendo los tiempos de vida resultantes en el conjunto de datos 'Chispazo.RData'.

a) Basándote en los datos recogidos, estudia desde un punto de vista descriptivo si la distribución del tiempo de vida sigue aproximadamente una distribución de Weibull $W(\alpha, \lambda)$ con parámetro de forma $\alpha = 2$. Para ello, estima puntualmente el parámetro de escala λ de la distribución haciendo uso de su estimador máximo-verosímil.

b) Suponiendo que el tiempo de vida sigue una distribución de Weibull con parámetro de forma $\alpha = 2$, calcula un intervalo de confianza al nivel de confianza $0{,}99$ para el parámetro de escala λ basándote en la distribución asintótica de su estimador máximo-verosímil.

c) Se quiere realizar un nuevo estudio más preciso que consiga estimar λ por medio de un intervalo de confianza asintótico para λ al nivel de confianza $0{,}99$ que tenga una longitud esperada inferior a $0{,}1$ unidades. Suponiendo que el tiempo de vida sigue una distribución de Weibull con parámetro de forma $\alpha = 2$ y haciendo uso de la estimación de λ en el estudio preliminar, determina de manera aproximada el tamaño de muestra que se necesita considerar para lograr el objetivo.