INFERENCIA ESTADÍSTICA — 28 de mayo de 2025 — Parcial 1º

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

1. (2 puntos) Analiza la solución que da una I.A. sobre el E.M.V. de una Beta(a,1) a partir de una muestra de tamaño n. Comenta los errores que pueda tener, si te parece incorrecta, y calcula la función de densidad del E.M.V., si consideras que está bien.

`--------------------- INICIO DE LA SOLUCIÓN DE LA I.A. ---------------------`

PREGUNTA: Calcula el estimador máximo-verosímil del parámetro a de una Beta(a,1) a partir de una uestra de tamaño n.

1. Planteamiento del Problema

Sea $X \sim \text{Beta}(a, 1)$ con función de densidad:

$$f(x \mid a) = a x^{a-1}, \quad 0 < x < 1,$$

donde $a > 0$ es el parámetro a estimar. Dada una muestra aleatoria $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, calcularemos el estimador máximo verosímil (EMV) de a y determinaremos su distribución.

2. Estimador Máximo Verosímil (EMV)

2.1. Función de Verosimilitud

La función de verosimilitud para la muestra es:

$$L(a; \mathbf{X}) = \prod_{i=1}^{n} f(X_i \mid a) = \prod_{i=1}^{n} a X_i^{a-1} = a^n \left(\prod_{i=1}^{n} X_i \right)^{a-1}.$$

2.2. Log-verosimilitud

Tomamos el logaritmo natural para simplificar:

$$\ell(a; \mathbf{X}) = \ln L(a; \mathbf{X}) = n \ln a + (a-1) \sum_{i=1}^{n} \ln X_i.$$

2.3. Derivación del EMV

Derivamos respecto a a e igualamos a cero para encontrar el máximo:

$$\frac{d\ell}{da} = \frac{n}{a} + \sum_{i=1}^{n} \ln X_i = 0.$$

Resolviendo para a:

$$\frac{n}{a} = -\sum_{i=1}^{n} \ln X_i \implies a = -\frac{n}{\sum_{i=1}^{n} \ln X_i}.$$

Por tanto, el EMV de a es:

$$\hat{a}_{\mathrm{EMV}} = -\frac{n}{\sum_{i=1}^{n} \ln X_i}.$$

3. Distribución del Estimador $\hat{a}_{\mathbf{EMV}}$

3.1. Transformación de la Variable

Definimos $Y_i = -\ln X_i$. Como $X_i \sim \mathrm{Beta}(a, 1)$, la transformación Y_i sigue una distribución exponencial:

$$Y_i \sim \mathrm{Exp}(a), \quad \text{con } f_{Y_i}(y) = a e^{-ay}, \quad y > 0.$$

3.2. Suma de Variables Aleatorias

La suma $S = \sum_{i=1}^{n} Y_i = -\sum_{i=1}^{n} \ln X_i$ sigue una distribución gamma:

$$S \sim \mathrm{Gamma}(n, a), \quad \text{con } f_S(s) = \frac{a^n s^{n-1} e^{-as}}{\Gamma(n)}, \quad s > 0.$$

3.3. Distribución de $\hat{a}_{\mathbf{EMV}}$

El estimador puede escribirse como:

$$\hat{a}_{\mathrm{EMV}} = \frac{n}{S}, \quad \text{donde } S \sim \mathrm{Gamma}(n, a).$$

La variable $T = aS \sim \mathrm{Gamma}(n, 1)$, por lo que:

$$\hat{a}_{\mathrm{EMV}} = \frac{an}{T}, \quad \text{con } T \sim \mathrm{Gamma}(n, 1).$$

La distribución de $\hat{a}_{\text{EMV}}$ es una **inversa gamma** escalada:

$$\hat{a}_{\text{EMV}} \sim \text{Inv-Gamma}(n, an),$$

con función de densidad:

$$f_{\hat{a}}(x) = \frac{(an)^n}{\Gamma(n)} x^{-n-1} e^{-an/x}, \quad x > 0.$$

4. Propiedades del Estimador

4.1. Sesgo

El valor esperado de $\hat{a}_{\text{EMV}}$ es:

$$E[\hat{a}_{\text{EMV}}] = E\left[\frac{n}{S}\right] = \frac{an}{n-1}, \quad \text{para } n > 1.$$

Por tanto, el estimador es **sesgado** pero asintóticamente insesgado cuando $n \to \infty$.

4.2. Varianza

La varianza de $\hat{a}_{\text{EMV}}$ es:

$$\text{Var}(\hat{a}_{\text{EMV}}) = \frac{a^2 n^2}{(n-1)^2(n-2)}, \quad \text{para } n > 2.$$

5. Ejemplo Numérico

Supongamos $n = 5$ con $\mathbf{X} = (-1{,}2, 0{,}3, 0{,}9, -0{,}5, 1{,}1)$:

- **EMV**: $\hat{\theta}_{\text{EMV}} = \text{máx}(|-1{,}2|, |0{,}3|, |0{,}9|, |-0{,}5|, |1{,}1|) = 1{,}2.$
- **MM**: $\hat{\theta}_{\text{MM}} = \sqrt{\frac{3}{5}((-1{,}2)^2 + 0{,}3^2 + 0{,}9^2 + (-0{,}5)^2 + 1{,}1^2)} = \sqrt{\frac{3}{5} \times 3{,}8} \approx 1{,}51.$

6. Conclusiones

- Ambos estimadores son consistentes (convergen al verdadero θ cuando $n \to \infty$).
- El EMV es más intuitivo pero puede subestimar θ en muestras pequeñas.
- El estimador de momentos utiliza más información muestral pero puede ser menos robusto a outliers.

---------------------- FIN DE LA SOLUCIÓN DE LA I.A. -----------------------

2. (2 puntos) Explica en qué consiste el método de la función pivote para calcular intervalos de confianza y construye una función pivote para una variable aleatoria de tipo continuo cuya distribución dependa de un parámetro unidimensional θ. Aplica esa función para el caso de una distribución exponencial con parámetro λ.

PROBLEMA 1

La tensión diastólica (mm Hg) en cierta población de ancianos sanos sigue una distribución gausiana $N(80, \sigma)$. En una residencia geriátrica se han medido dichas tensiones y se han obtenido estos valores: (80, 82, 83, 85, 79, 78, 90, 92, 98, 100, 86, 91).

a) (1 punto) Halla el estimador y la estimación máximo-verosímiles para σ.

b) (1 punto) Halla un intervalo de confianza exacto para σ con $1 - \alpha = 0{,}9$.

c) (1 punto) Halla un intervalo de confianza *t-bootstrap* para σ con $1 - \alpha = 0{,}9$.

PROBLEMA 2

La empresa Neurovitalis está evaluando la efectividad de un nuevo programa de entrenamiento cognitivo para mejorar la memoria a corto plazo en adultos mayores. Para ello, se seleccionó aleatoriamente a 50 participantes y se midió su capacidad de memoria antes de iniciar el programa de entrenamiento cognitivo (Momento 1). Los mismos individuos participaron en el programa durante 8 semanas, y al finalizar se les volvió a medir la memoria (Momento 2). La variable de medida es una puntuación estandarizada en una escala continua de 0 a 100. Los datos obtenidos se encuentran en el fichero 'Neurovitalis.RData'. Basándote en los datos recogidos y asumiendo independencia entre individuos (pero no entre momentos de medición dentro de un mismo individuo), responde a las siguientes preguntas a nivel de significación $\alpha = 0{,}05$:

a) (1 punto) Estudia mediante un gráfico si la mejora en la capacidad de memoria sigue una distribución normal. Resalta algún aspecto descriptivo numérico para afianzar tu respuesta.

b) (1 punto) ¿Hay evidencias para afirmar que el entrenamiento cognitivo ayuda a mejorar la capacidad de memoria? Realiza un estudio descriptivo numérico y basa tus conclusiones en un intervalo de confianza al nivel de confianza 0,95.

c) (1 punto) Asumiendo distribución normal para la mejora en la capacidad de memoria y utilizando la estimación puntual de la desviación típica poblacional si fuese necesario, indica si el tamaño muestral actual es suficiente para garantizar que la amplitud esperada del intervalo de confianza al nivel de confianza 0,95 para la varianza de la mejora sea inferior a 5 unidades. En caso negativo, calcula el menor tamaño muestral necesario para lograr este objetivo.