PREMIOS JORGE JUAN 2013

PROBLEMA DE ESTADÍSTICA

Elegir entre la opción A y la B

OPCION A

Sea X una variable aleatoria cuya función de densidad es

$$f_\theta(x) = 1, \quad \theta - \frac{1}{2} \leq x \leq \theta + \frac{1}{2}, \quad \theta \in \mathbb{R}.$$

Consideremos una m.a.s. de X, $(X_1, ..., X_n)$ y el estadístico $S = \left(X_{(1)}, X_{(n)}\right)$

a) Demostrar que S es una estadístico suficiente para θ.

En la mayoría de textos de Inferencia Estadística, aparece el siguiente resultado: "Cualquier estimador de máxima verosimilitud (EMV) de θ debe ser función de cualquier estadístico suficiente para θ". En realidad, este resultado es cierto cuando el EMV de θ es único. Si no es único, se pueden encontrar EMV´s función del estadístico suficiente, pero pueden existir EMV´s que no lo sean.

b) Demostrar que cualquier combinación convexa de $X_{(n)} - \frac{1}{2}$ y $X_{(1)} + \frac{1}{2}$ es un EMV de θ. Por tanto, todos son función de S, pero no existe unicidad.

c) Demostrar que podemos elegir un EMV de θ que no sea función de S :

$$\widehat{\theta} = \left(X_{(n)} - \frac{1}{2}\right) + h(X_1, ..., X_n)\left(X_{(1)} - X_{(n)} + 1\right),$$

donde $h(x_1, ..., x_n)$ es una función no constante que debe cumplir ciertas condiciones. Poner un ejemplo.

- Un estadístico $S(X_1, ..., X_n)$ es suficiente para θ si, $\forall (x_1, ..., x_n) \in \mathbb{R}^n$, la función de densidad o probabilidad de $(X_1, ..., X_n)$ se puede factorizar
$$.f_\theta(x_1, ..., x_n) = b(x_1, ..., x_n)\, k(\theta, S(x_1, ..., x_n)).$$
- Un EMV de θ es un estadístico que maximiza la función de verosimilitud de θ, para cada realización de la m.a.s., siendo la función de verosimilitud
$$L_{(x_1,...,x_n)}(\theta) = f_\theta(x_1, ..., x_n),$$
para cada $(x_1, ..., x_n)$ en el soporte de $(X_1, ..., X_n)$.

OPCION B

a) Un conjunto de variables aleatorias $X_1, ..., X_n$ se dice *intercambiable* si la distribución conjunta del vector $(X_1, ..., X_n)$ es la misma que la de cualquier permutación de las n variables. Sean $X_1, ..., X_n$ variables aleatorias positivas e intercambiables. Probar que

$$E\left(\frac{X_1 + ... + X_k}{X_1 + ... + X_n}\right) = \frac{k}{n}, \qquad 1 \leq k \leq n.$$

b) Sea X una variable aleatoria con soporte $\{0, 1, 2, ...\}$. Probar que

$$\sum_{n=0}^{\infty} s^n P\left(X \leq n\right) = \frac{1}{1-s} P\left(s\right), \qquad |s| < 1,$$

donde $P\left(s\right) = \sum_{k=0}^{\infty} s^k P\left(X = k\right)$ es la *función generatriz de probabilidad* de X.

- Si $f\left(x_1, ..., x_n\right)$ es la función de densidad conjunta del vector aleatorio continuo $(X_1, ..., X_n)$, la esperanza de la variable aleatoria función del vector, $h\left(X_1, ..., X_n\right)$, (con $h : \mathbb{R}^n \to \mathbb{R}$) se define

$$E\left(h\left(X_1, ..., X_n\right)\right) := \int ... \int_{\mathbb{R}^n} h\left(x_1, ..., x_n\right) f\left(x_1, ..., x_n\right) dx_1 ... dx_n.$$

  Si el vector aleatorio es de tipo discreto y $f\left(x_1, ..., x_n\right)$ es su función de probabilidad conjunta,

$$E\left(h\left(X_1, ..., X_n\right)\right) := \sum_{\mathbb{R}^n} h\left(x_1, ..., x_n\right) f\left(x_1, ..., x_n\right).$$

SOLUCIONES OPCION A

a) La función de densidad de la m.a.s. se puede factorizar

$$f_\theta(x_1,...,x_n) = I_{\left]-\infty,\theta+\frac{1}{2}\right]}\left(x_{(n)}\right) I_{\left[\theta-\frac{1}{2},\infty\right[}\left(x_{(1)}\right),$$

para todo $(x_1,...,x_n) \in \mathbb{R}^n$, siendo

$$I_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases}$$

la función indicatriz de un conjunto $A \subset \mathbb{R}$ en nuestro caso.

Así, el estadístico $S = \left(X_{(1)}, X_{(n)}\right)$ es suficiente para θ.

b) La función de verosimilitud, para cada realización $(x_1,...,x_n)$ de la m.a.s. es

$$L_{(x_1,...,x_n)}(\theta) = 1,$$

si $x_{(n)} - \frac{1}{2} \leq \theta \leq x_{(1)} + \frac{1}{2}$. Al ser constantemente igual a 1 en el intervalo $\left[x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2}\right]$, cualquier combinación convexa de $x_{(n)} - \frac{1}{2}$ y $x_{(1)} + \frac{1}{2}$ será un máximo de la función. Así,

$$\widehat{\theta}_\alpha = \alpha\left(X_{(n)} - \frac{1}{2}\right) + (1-\alpha)\left(X_{(1)} + \frac{1}{2}\right),$$

con $0 \leq \alpha \leq 1$ son EMV de θ.

c) Si $\widehat{\theta} = \left(X_{(n)} - \frac{1}{2}\right) + h(X_1,...,X_n)\left(X_{(1)} - X_{(n)} + 1\right)$ es EMV de θ, la función de verosimilitud en $\widehat{\theta}$ debe valer 1, luego, para toda realización $(x_1,...,x_n)$, $\widehat{\theta}(x_1,...,x_n)$ debe estar en el intervalo $\left[x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2}\right]$, es decir,

$$\left|\widehat{\theta} - x_{(n)} + \frac{1}{2}\right| \leq x_{(1)} - x_{(n)} + 1,$$

así, $0 \leq h(x_1,...,x_n) \leq 1$, para toda $(x_1,...,x_n)$.

Podemos tomar, por ejemplo, $h(x_1,...,x_n) = \cos^2 x_1$.

SOLUCIONES OPCION B

a) Dado que $E\left(\frac{X_1+...+X_n}{X_1+...+X_n}\right) = 1$, deducimos que

$$\sum_{i=1}^{n} E\left(\frac{X_i}{X_1+...+X_n}\right) = 1.$$

Al ser las variables aleatoria positivas, se tiene que

$$E\left(\frac{X_i}{X_1+...+X_n}\right) \geq 0, \text{ para todo } i = 1,...,n,$$

luego si la suma es uno, necesariamente todas son finitas. Veamos que todas son iguales, por lo que $E\left(\frac{X_i}{X_1+...+X_n}\right)=\frac{1}{n}$ y ya habremos llegado al resultado deseado.

Al ser intercambiables, lo haremos para X_1 y X_2. Supondremos el caso continuo, y si fuera discreto, basta cambiar integrales por sumas.

$$\begin{aligned} E\left(\frac{X_1}{X_1+...+X_n}\right) &= \int ... \int_{\mathbb{R}^n} \frac{x_1}{x_1+...+x_n} f\left(x_1,...,x_n\right) dx_1...dx_n \\ &= \int ... \int_{\mathbb{R}^n} x_1 g\left(x_1,...,x_n\right) dx_1...dx_n, \end{aligned}$$

siendo $g\left(x_1,...,x_n\right)=\frac{f(x_1,...,x_n)}{x_1+...+x_n}$ una función a su vez intercambiable. Así

$$E\left(\frac{X_1}{X_1+...+X_n}\right)=\int x_1\left[\int ... \int_{\mathbb{R}^{n-1}} g\left(x_1,...,x_n\right) dx_2...dx_n\right] dx_1.$$

Si $G\left(x_1\right):=\int ... \int_{\mathbb{R}^{n-1}} g\left(x_1,...,x_n\right) dx_2...dx_n$, tenemos que

$$E\left(\frac{X_1}{X_1+...+X_n}\right)=\int x_1 G\left(x_1\right) dx_1.$$

Claramente $G\left(x_1\right)=G\left(x_2\right)$, por ser $g\left(x_1,...,x_n\right)$ intercambiable, de manera que

$$E\left(\frac{X_1}{X_1+...+X_n}\right)=\int x_1 G\left(x_1\right) dx_1=\int x_2 G\left(x_2\right) dx_2=E\left(\frac{X_2}{X_1+...+X_n}\right).$$

b)

$$\sum_{n=0}^{\infty} s^n P\left(X \leq n\right)=\sum_{n=0}^{\infty}\left(\sum_{k=n}^{\infty} s^k\right) P\left(X=n\right)==\frac{1}{1-s}\sum_{n=0}^{\infty} s^n P\left(X=n\right).$$