Estadística I
Grado en Matemáticas, UAM, 2015-2016

Hoja 4. Estimadores. Método de momentos y máxima verosimilitud

Sesgo y eficiencia de estimadores

1. Comprueba que si T_1 y T_2 son estimadores insesgados de un parámetro de θ, entonces, para todo $\lambda \in (0,1)$, $Z = \lambda T_1 + (1-\lambda)T_2$ es estimador insesgado de θ.

2. Sean T_1 y T_2 dos estimadores insesgados de un parámetro θ que actúan sobre muestras de tamaño n. Formamos el estimador U (que actúa sobre muestras $(X_1, \dots, X_{2n})$ de tamaño $2n$):

$$U(X_1, \dots, X_{2n}) = \frac{1}{2}\big(T_1(X_1, \dots, X_n) + T_2(X_{n+1}, \dots, X_{2n})\big)\,.$$

Comprueba que $\mathbf{V}_\theta(U) = \frac{1}{4}\big(\mathbf{V}_\theta(T_1) + \mathbf{V}_\theta(T_2)\big)$.

3. Tenemos una variable X que toma los valores $\{-1, 0, +1\}$ con probabilidades respectivas

-1	0	$+1$
$(2+\theta)/4$	$\theta/4$	$(2-2\theta)/4$

para $\theta \in [0,1]$.

a) Consideremos el estadístico N_1 dado por

$$N_1 = h_1(X_1, X_2, \dots, X_n)\,,$$

donde $h_1(x_1, x_2, \dots, x_n) =$ número de $\{x_j = -1\}$. Observa que $N_1 \sim \text{BIN}(n, (2+\theta)/4)$. Comprueba que el estadístico

$$T_1 = \frac{4}{n}N_1 - 2$$

es un estimador insesgado de θ.

b) Consideremos el estadístico N_2 dado por

$$N_2 = h_2(X_1, X_2, \dots, X_n)\,,$$

donde $h_2(x_1, x_2, \dots, x_n) =$ número de $\{x_j = 0\}$. Observa que $N_2 \sim \text{BIN}(n, \theta/4)$. Comprueba que el estadístico

$$T_2 = \frac{4}{n}N_2$$

es un estimador insesgado de θ. ¿Cuál de los estimadores, T_1 o T_2, es más eficiente?

c) Supongamos que la muestra $(x_1, x_2, \dots, x_{100})$ tiene las siguientes frecuencias:

-1	0	1
60	16	24

¿Cuál es la estimación de θ si usamos T_1? ¿Y si usamos T_2?

4. La variable X tiene una distribución dada por dos parámetros $\delta > 0$ y $\lambda > 0$, de manera que

$$X = \delta + Y\,, \quad \text{donde } Y \sim \text{EXP}(\lambda)\,.$$

Los parámetros δ y λ son desconocidos.

a) Sea $m_n = \text{mín}(X_1, X_2, \dots, X_n)$. Comprueba que

$$\mathbf{E}(m_n) = \delta + \frac{1}{n\lambda} \qquad \text{y que} \qquad \mathbf{E}(\overline{X}) = \delta + \frac{1}{\lambda}\,.$$

b) Comprueba que

$$T_1 = \frac{n}{n-1}(\overline{X} - m_n)$$

es un estimador insesgado de $1/\lambda$ y que

$$T_2 = \frac{n}{n-1}(m_n - \overline{X}/n)$$

es un estimador insesgado de δ.

5. Sea $(X_1, X_2, \ldots, X_n)$ una muestra aleatoria de una variable $X \sim \textsc{exp}(\lambda)$. Queremos estimar el parámetro $\theta = 1/\lambda$.

a) Compara la eficiencia para estimar θ de los estadísticos insesgados

$$T_1 = \overline{X} \quad \text{y} \quad T_2 = n \,\text{mín}(X_1, X_2, \ldots, X_n)\,.$$

b) Sea ahora T_3 el estadístico

$$U_\omega = \omega \sum_{j=1}^{n} X_j\,.$$

donde ω es una constante, $\omega \neq 0$. Observa que U_ω es estimador insesgado de θ si y sólo si $\omega = 1/n$. Halla ω para que el ECM sea mínimo.

6. Sea T un estimador de un parámetro θ. Supongamos que $\mathbf{E}_\theta(T) = \alpha\theta$ y que $\mathbf{V}_\theta(T) = \beta\theta^2$, donde α y β son dos constantes fijas.

Halla r (en función de α y β) para que $\text{ECM}_\theta(rT)$ sea mínimo. Aplícalo al caso en el que $X \sim \textsc{Unif}[0, a]$ y al estimador $T = \text{máx}(X_1, \ldots, X_n)$ de a.

Método de momentos y máxima verosimilitud

7. Calcula los estimadores por máxima verosimilitud para a) parámetro λ de $X \sim \textsc{poiss}(\lambda)$; b) parámetro p de $X \sim \textsc{bin}(N, p)$, con N dado.

8. Una variable aleatoria X tiene función de densidad

$$f(x; \theta) = \frac{2}{\theta^2}\, x \quad \text{para } x \in [0, \theta],$$

y $f(x; \theta) = 0$ si $x \notin [0, \theta]$. El parámetro θ es positivo. Halla el estimador M_n de máxima verosimilitud de θ para una muestra de tamaño n.

9. Sea $X_1, X_2, \ldots, X_n$ muestra aleatoria de la función de densidad que depende del parámetro $\theta \in \Theta = (0, +\infty)$ dada por

$$f(x; \theta) = \begin{cases} \theta/x^2, & \text{si } x > \theta\,, \\ 0, & \text{si } x \leq \theta\,. \end{cases}$$

Obtén los estimadores de θ por máxima verosimilitud y por el método de momentos.

10. Calcula el estimador por máxima verosimilitud del parámetro $\theta \in [0, 1]$ para la variable X que toma tres valores $-1, 0, +1$ con probabilidades respectivas

-1	0	$+1$
$(2+\theta)/4$	$\theta/4$	$(2-2\theta)/4$

11. Sea X una variable normal de media μ y varianza conocida $\sigma^2 = 1$. El espacio de parámetros (para μ) es el intervalo $[-1, 1]$. Dada una muestra $(x_1, \ldots, x_n)$ de X, determina el estimador de máxima verosimilitud de μ.

12. Sea $\Theta = \{0, 1\}$ y considérense las dos funciones de densidad (con soporte en $(0, 1)$) alternativas dadas, para $x \in (0, 1)$, por

$$f(x; 0) = 1\,, \qquad f(x; 1) = 1/(2\sqrt{x})\,.$$

Determina el estimador de máxima verosimilitud del valor de $\theta \in \{0, 1\}$.

13. Sea $X \sim \text{UNIF}(a, b)$, con $a < b$. Determina los estimadores de a y b por máxima verosimilitud y por el método de momentos.

14. Considérese, para $\theta > 0$, la función de densidad dada

$$f(x;\theta) = \begin{cases} 0, & \text{si } x \leq 0\,, \\ 1 - e^{-\theta}, & \text{si } x \in (0,1)\,, \\ \theta e^{-\theta x}, & \text{si } x \geq 1\,. \end{cases}$$

Sea T el estadístico:

$$T(X_1, \ldots, X_n) = \frac{1}{n} \# \big\{ X_j \in (0,1); 1 \leq j \leq n \big\}\,,$$

que registra la proporción de la muestra que está por debajo de 1. Y sea Y el estadístico

$$Z(X_1, \ldots, X_n) = \frac{1}{n} \sum_{1 \leq j \leq n:\, X_j \geq 1} X_j\,,$$

que registra el promedio respecto de n de aquellos valores de la muestra por encima de 1.

a) Comprueba que $\mathbf{E}_\theta(T) = 1 - e^{-\theta}$ y que $\mathbf{E}_\theta(Z) = e^{-\theta}\big(1 + 1/\theta\big)$.

b) Comprueba que el estimador máximo verosímil M de θ viene dado implícitamente por la ecuación:

$$T\Big(\frac{1}{e^M - 1} - \frac{1}{M}\Big) = Z - \frac{1}{M}\,.$$

15. Un arquero (con nula experiencia) dispara n veces a una diana de radio θ (desconocido). En cada lanzamiento logra darle al disco, pero en lugares completamente aleatorios cuyas distancias al centro del disco son $r_1, r_2, \ldots, r_n$. Determina el estimador de máxima verosimilitud del radio del disco.