

Relación 2 de problemas

En esta hoja todos los estadísticos están basados en una muestra X_1, \dots, X_n de v.a.i.i.d. con $\mathbb{E}(X_i) = \mu$ y $\mathbb{V}(X_i) = \sigma^2$ (estas dos cantidades se supondrán finitas cuando aparezcan en los enunciados). La función de distribución de las X_i se denota por F y la de densidad por f . \mathbb{F}_n denota la función de distribución empírica. El símbolo \rightarrow denota convergencia cuando $n \rightarrow \infty$.

1. Se desea estimar el momento de orden 3, $\alpha_3 = \mathbb{E}(X^3)$, en una v.a. X con distribución exponencial de parámetro 2, es decir, la función de distribución de X es $F(t) = \mathbb{P}(X \leq t) = 1 - e^{-2t}$, para $t \geq 0$. Definir un estimador natural para α_3 y calcular su error cuadrático medio.

Indicación: Si $X \sim \exp(\lambda)$, entonces $\mathbb{E}(X^n) = \frac{n!}{\lambda^n}$ para todo entero positivo n .

2. Supongamos que la muestra tiene tamaño $n = 50$ y que la distribución de las X_i es una $N(4, 1)$ (i.e., normal con media $\mu = 4$ y desviación típica $\sigma = 1$).

(a) Obtener, utilizando la desigualdad de Chebichev, una cota superior para la probabilidad $\mathbb{P}\{|\bar{X} - 4| > 0.3\}$. Al utilizar la desigualdad de Chebichev no estamos usando el hecho de que la distribución de las observaciones es normal: es una desigualdad universal y proporciona una cota (no necesariamente muy ajustada) para $\mathbb{P}\{|\bar{X} - 4| > 0.3\}$, que solo depende de n y de σ .

(b) Calcula exactamente la probabilidad $\mathbb{P}\{|\bar{X} - 4| > 0.3\}$ utilizando el hecho de que las X_i tienen distribución $N(4, 1)$. Comparar el resultado con la cota obtenida en (a).

3. Utilizando R dibuja la función de densidad y la función de distribución de una v.a. con distribución beta de parámetros $a = 3$, $b = 6$. A continuación dibuja, sobrepuestas en cada uno de los gráficos, las aproximaciones a F y f obtenidas respectivamente mediante la función de distribución empírica y un estimador kernel.

Verificar empíricamente el grado de aproximación alcanzado, en las estimaciones de F y f , mediante un experimento de simulación basado en 200 muestras de tamaño 20. Es decir, considerando, por ejemplo, la estimación de F , se trata de simular 200 muestras de tamaño 20; para cada una de ellas evaluar el error (medido en la norma del supremo) cometido al aproximar F por \mathbb{F}_n . Por último, calcular el promedio de los 200 errores obtenidos. Análogamente para la estimación de f .

4. Denotemos por

$$C_n = \int_{\mathbb{R}} (\mathbb{F}_n(t) - F(t))^2 dF(t),$$

la llamada “discrepancia de Cramer-Von Mises” entre \mathbb{F}_n y F , . ¿Se verifica necesariamente $C_n \rightarrow 0$, c.s.?

Calcular la distribución asintótica de la sucesión $D_n = \sqrt{n}(\mathbb{F}_n(t) - F(t))$, para un valor fijo $t \in \mathbb{R}$.

5. Sea X una v.a. con distribución de Cauchy($\theta, a = 1$), cuya función de densidad

$$f(x; \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \quad x \in \mathbb{R},$$

depende del parámetro desconocido $\theta \in \mathbb{R}$. Comprobar que θ coincide con la mediana y la moda de X pero que la media $\mathbb{E}(X)$ no está definida.

Diseñar un experimento de simulación en R, tomando algún valor concreto de θ , orientado a comprobar cómo se comportan la mediana muestral y la media muestral como estimadores de θ : mientras la mediana muestral se acerca al verdadero valor de θ al aumentar n , la media muestral oscila fuertemente y no se acerca a θ aunque se aumente el tamaño muestral n .

6. Se extrae una muestra aleatoria de tamaño $n = 600$ de una v.a. cuya desviación típica es $\sigma = 3$. Calcular aproximadamente la probabilidad $\mathbb{P}\{|\bar{X} - \mu| < 0.1\}$.

7. Dada una muestra de v.a.i.i.d. de tamaño 100 de una distribución normal de media μ y desviación típica 1.5, determina aproximadamente la probabilidad de que la mediana muestral difiera de μ en menos que 0.1. ¿De qué tamaño habría que elegir la muestra para poder afirmar que con probabilidad 0.9, la mediana muestral difiere de μ en menos que 0.01? Resuelve el problema sustituyendo la mediana por la media y compara los resultados obtenidos.

8. Sea \hat{f}_n un estimador kernel de la densidad basado en un núcleo K que es una función de densidad con media finita. Comprobar que, en general, $\hat{f}_n(t)$ es un estimador sesgado de $f(t)$ en el sentido de que NO se tiene $\mathbb{E}(\hat{f}_n(t)) = f(t)$, para todo t y para toda densidad f .

9. Sea X una v.a. con distribución gamma de parámetros $\alpha = 2$ y $\lambda = 1$. Tomamos una muestra X_1, \dots, X_n de X . Queremos estimar la función de densidad f de X en el punto t , para lo cual utilizaremos un estimador de ventana móvil, $\hat{f}_n(t)$, con parámetro ventana $h = h_n$.

Indicación 1: Para $t > 0$ la densidad de una gamma de parámetros α y λ tiene la expresión

$$f(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t},$$

donde $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ y $\Gamma(n) = (n-1)!$ para todo $n \in \mathbb{N}$.

Indicación 2: Aproximar $\mathbb{P}\{X \in [t - h_n, t + h_n]\}$ por $2h_n f(t)$.

(a) Determinar $\mathbb{E}(\hat{f}_n(t))$ aproximadamente en términos de $f(t)$.

(b) Expresar la varianza de $\hat{f}_n(t)$ en términos de n y h . ¿Qué condición debe satisfacer h_n para que $\mathbb{V}(\hat{f}_n(t))$ tienda a 0 cuando $n \rightarrow \infty$?

(c) Utilizando la desigualdad de Chebyshev, determinar cómo de grande debe de ser n para que el error relativo

$$\frac{|\hat{f}_n(t) - f(t)|}{f(t)}$$

del estimador $\hat{f}_n(t)$ sea inferior a c con una probabilidad θ .

10. Sea F_4 la función de distribución empírica correspondiente a la muestra X_1, X_2, X_3, X_4 .

(a) Si se observa $x_1 = 1, x_2 = 3, x_3 = 5$ y $x_4 = 7$, determina el valor de $F_4^{-1}(1/2)$.

(b) Si se observa $x_1 = 1, x_2 = 3, x_3 = 5$ y $x_4 = 7$ y X^* es una variable aleatoria con distribución F_4 , calcula el valor esperado de X^* (condicionado a la muestra).

(c) Si X_1, X_2, X_3 y X_4 son vaaid con distribución uniforme en el conjunto $\{1, 2, 3, \dots, 10\}$, ¿cuál es la distribución de la variable aleatoria $4F_4(5)$?