Departamento de matemática y aplicaciones; Cure–Universidad de la República *Diplomа de Matemática*

ESTADÍSTICA

EJERCICIOS ENTREGABLES

§1. La función gamma de parámetros $\alpha > 0$ y $\beta > 0$, la cual notaremos $\Gamma(\alpha,\beta)$, tiene densidad dada por

$$f(x,\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \qquad x \in (0,\infty), \tag{1}$$

donde

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt. \tag{2}$$

(a) Hallar $E(X)$ y $Var(X)$ si $X \sim \Gamma(\alpha,\beta)$.
(b) Estimar los parámetros α y β por máxima verosimilitud.
(c) Programar el algoritmo en R.

§2. Consideremos una función ρ. Llamaremos M-estimador el que se obtiene de minimizar

$$\operatorname*{argmin}_\theta \sum_{i=1}^n \rho(x_i - \theta).$$

(a) Sea

$$h(x) = \begin{cases} \frac{1}{2}x^2 & si \quad |x| \le k \\ k|x| - \frac{1}{2}k^2 & si \quad |x| \ge k \end{cases} \tag{3}$$

Probar que $h(x)$ y $h'(x)$ son continuas.

(b) Sea $\psi = \rho'(x)$. Consideremos θ_0 el verdadero valor del parámetro y sea $\hat\theta_M$ el M-estimador. Pruebe que

$$\sqrt{n}(\hat\theta_M - \theta_0) = \frac{-\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(x_i - \theta_0)}{\frac{1}{n} \sum_{i=1}^n \psi'(x_i - \theta_0)}. \tag{4}$$

(c) Suponga que $E_{\theta_0}(\psi(x - \theta_0)) = 0$ y pruebe que

$$\sqrt{n}(\hat\theta_M - \theta_0) \to \mathcal{N}\left(0, \frac{E_{\theta_0}(\psi(x-\theta_0)^2)}{[E_{\theta_0}(\psi'(x-\theta_0))]^2}\right). \tag{5}$$

(d) Sean $X_1, \ldots, X_n$ iid con densidad $f(x-\theta)$ con f simétrica respecto a 0. Sea ρ la h definida en el apartado (a). Pruebe que $E_\theta(\psi(x-\theta)) = 0$.

(e) Si $f(x)$ es una densidad simétrica respecto a 0 y ρ una función simétrica, $\psi = \rho'$, pruebe que

$$\int \psi(x-\theta) f(x-\theta) dx = 0. \tag{6}$$

Pruebe que esto implica que si $X_1, \ldots, X_n$ son iid de $f(x-\theta)$ y $\hat\theta_n$ es el estimador obtenido de minimizar $\sum_i \rho(x_i - \theta)$, entonces $\hat\theta_n$ es asintóticamente normal con media igual a θ_0 (verdadero valor de θ).

(f) Aplique este resultado a la función h definida en el apartado (a).

§3. (a) Simular en R 1000 valores $X_1, X_2, \ldots, X_{1000}$ correspondientes a una v.a. geométrica (número de ensayos hasta obtener un éxito) de parámetro p (hacerlo para varios valores de p).

(b) A partir de los datos anteriores graficar la función de probabilidad y la función de distribución empírica.

(c) Dado $n \in \mathbb{N}$ sea $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Consideremos $n = 100$, 1000, 10000 y p un valor aleatorio entre 0 y 1. Realice 3000 simulaciones para cada n y para el p obtenido. Comparar S_n con una variable aleatoria exponencial de parámetro $\lambda = p$. Comparar valores, histogramas y función de distribución empírica.

§4. El siguiente ejercicio trata sobre modelos lineales. Supongamos que deseamos modelizar una variable Y como una combinación de $p-1$ variables X_1, X_2, …, X_{p-1}. Supongamos que tenemos n observaciones, entonces

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i, \qquad i = 1, 2, \ldots, n. \tag{7}$$

Matricialmente, el modelo queda

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1,p-1} \\ 1 & X_{21} & \cdots & X_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \tag{8}$$

(a) Pruebe que si a y β son vectores y si $\frac{d}{d\beta} = \left(\frac{d}{d\beta_i}\right)$, entonces

- $\dfrac{d(\beta' a)}{d\beta} = a.$
- $\dfrac{d(\beta' A\beta)}{d\beta} = 2A\beta.$

(b) Halle el estimador de β mediante el método de mínimos cuadrados, o sea

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \varepsilon_i^2 = \underset{\beta}{\operatorname{argmin}}\ \varepsilon'\varepsilon. \tag{9}$$

(c) Si consideramos que $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I)$, halle los estimadores de β y σ^2 mediante máxima verosimilitud.

(d) Consideremos nuevamente el modelo lineal $Y = X\beta + \epsilon$. Pero ahora $Var(\epsilon) = \sigma^2 V$ con V una matriz definida positiva nxn (no se pide normalidad). Halle los estimadores de β y σ^2. Pruebe que el estimador de β es insesgado y halle su varianza.

(e) Aplique la parte (d) al caso particular donde V es una matriz diagonal de ponderadores (no necesariamente iguales).

§5. Supongamos que $X_1, \ldots, X_n \sim f$ es una población iid con- $X_i \in \mathbb{R}^d$. Consideremos una sucesión $k_n \in \mathbb{Z}$ y $k_n > 0$ con $k_n \to \infty$. Dado $x \in \mathbb{R}^d$, sea $H_n(x) = \|x_i - x\|^{k_n}$. Aquí, x^{k_n} representa el k-ésimo estadístico de orden, respecto a la norma $\|\cdot\|$. O sea tomamos la distancia al k-ésimo vecino más cercano. Definimos el estimador de vecinos más cercanos de f

como

$$f_n(x) = \frac{k}{nH_n^d(x)\lambda(B_1)} = \frac{k}{n\lambda(B_{H_n}(x))}. \tag{10}$$

Observación. $\lambda(B_1)$ es la medida de la bola de radio 1.

(a) Usando el hecho que si $X \sim F$ entonces $F_X(x^k) \sim \beta(k, n-k+1)$, probar que si

$$Z_n(x) = P(B_{H_n(x)}) = \int\limits_{B_{H_n(x)}} f(t)dt, \tag{11}$$

entonces $Z_n \sim \beta(k, n-k+1)$.

(b) Pruebe que $Z_n(x) \xrightarrow{p} 0$ si $k_n \to \infty$ y $\frac{k_n}{n} \to 0$.

(c) Suponiendo que x es un punto de continuidad de f y que $f(x) > 0$, pruebe que $H_n(x) \xrightarrow{p} 0$ si $k_n \to \infty$ y $\frac{k_n}{n} \to 0$.

(d) Pruebe que $\dfrac{P(B_{H_n(x)})}{P_n(B_{H_n(x)})} \xrightarrow{p} 1$.

(e) Usando el teorema de diferenciación que expresa que

$$\lim_{r\to 0} \frac{1}{\lambda(B_r)} \int_{B_r} f(t)dt \to f(x), \tag{12}$$

si x es un punto de continuidad de f, pruebe que $f_n(x) \xrightarrow{p} f(x)$.

§6. (a) Implemente en R el estimador de densidad y de regresión por núcleos. Programe los siguientes núcleos:

(i) Uniforme.
(ii) Triangular.
(iii) Epanechnikov.
(iv) Quartic.
(v) Triweight.
(vi) Gaussiano.
(vii) Coseno.

Para el caso de la densidad multidimensional, utilice las transformaciones correspondientes. Se pide que la implementación permita estimar con un vector de puntos la densidad o la regresión y que también grafique la solución.

(b) Implemente los intervalos de confianza asintóticos de los estimadores.

§7. Implemente en R los métodos de validación cruzada vistos en el curso.