Estadística e inferencia I

Primer Cuatrimestre — 2024

Primer Parcial

Apellido y nombre: ....

---

1. Considere muestras aleatorias de la distribución gaussiana inversa

$$f(x;\mu,\phi) = \frac{1}{\sqrt{2\pi x^3 \phi}} e^{-\frac{(x-\mu)^2}{2\phi x \mu^2}}.$$

(*a*) Encontrar los estimadores de momentos de μ y ϕ.

(*b*) Encontrar los estimadores de máxima verosimilitud de μ y ϕ.

(*c*) Decir si los estimadores de μ son insesgados o asintóticamente insesgados y calcular su error cuadrático medio.

(*d*) Decir si la siguiente afirmación es verdadera o falsa y justificar. "Dados dos estimadores de un parámetro θ, siempre conviene usar un estimador insesgado por sobre uno sesgado."

2. Sea Z la variable aleatoria con distribución *skew-normal* de parámetro $a = 10$. Consideremos la función $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = 2x - 1$. Sean $X \sim \mathcal{U}(0,1)$ y $Y := f(X) + Z$. [1]

(*a*) Fijemos $n = 100$. Tomar muestras $x_1, \ldots, x_n$ de X y $z_1, \ldots, z_n$ de Z y calcular $y_i = f(x_i) + z_i$, con $1 \le i \le n$. Realizar un ajuste lineal de la forma $Y = X\beta_1 + \beta_0 + \epsilon$ a partir de los samples $(x_1, y_1), \ldots, (x_n, y_n)$ generados.

(*b*) Repetir el paso anterior $m = 30$ veces para obtener m pares de coeficientes β_0 y β_1, es decir, m samples de $\hat{\beta}_0$ y m samples de $\hat{\beta}_1$. ¿Diríamos que estas distribuciones empíricas son normales? Hacer histogramas para comprobarlo.

3. Modelemos la cantidad de likes diarios que recibe un posteo en Instagram los 30 días posteriores a su publicación. Para eso, consideramos muestras $(x_1, y_1), \ldots, (x_n, y_n)$ de las variables aleatorias X, los días desde el posteo, que toma valores entre 0 y 30, y Y, la cantidad de likes por día, que es un proceso de Poisson que debe reflejar de manera razonable el fenómeno (por ejemplo, los primeros días se dan muchos likes y después van decayendo en el tiempo).

(*a*) Generar muestras sintéticas $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ para $n = 100$. Graficar y vs x.

(*b*) Proponer el Modelo Lineal Generalizado de Poisson y encontrar estimaciones puntuales para los parámetros de regresión β.

(*c*) Determinar el error estándar de cada parámetro de regresión.

4. En 1000 tiradas de moneda aparecen 540 caras y 460 cecas. Estamos interesados en testear la hipótesis H_0: es igual o más probable sacar ceca que cara.

(*a*) Plantear un test que tenga nivel aproximado $\alpha = 0.05$.

(*b*) Describir la región de rechazo. ¿Cuál es la decisión para el conjunto de datos obtenido?

(*c*) Encontrar el valor p aproximado para el conjunto de datos obtenido.

---

[1] Dado $a \in \mathbb{R}$, la variable aleatoria *skew-normal de parámetro a* es la que tiene pdf dada por $f(x) = 2\phi(x)\Phi(ax)$, donde ϕ y Φ son la pdf y la cdf de la normal estándar. Para usarla en Python, se puede importar usando `from scipy.stats import skewnorm` y luego se puede samplear con `skewnorm.rvs(a=10, size=100)`, para una muestra de tamaño 100 con $a = 10$.