Estadística e inferencia I
Segundo Cuatrimestre — 2024

Práctica 1: Estimadores

Modelos no paramétricos

1. Se analizó una muestra de 12 piezas de pan blanco de cierta marca y se determinó el porcentaje de carbohidratos contenido en cada una de las piezas, obteniéndose los siguientes valores:

76.93,	76.88,	77.07,	76.68,	76.39,	75.09,
77.67,	76.88,	78.15,	76.50,	77.16,	76.42.

(a) Estimar el promedio del porcentaje de carbohidratos contenido en las piezas de pan de esta marca.

(b) Estimar la mediana del porcentaje de carbohidratos.

(c) Estimar la proporción de piezas de pan de esta marca cuyo contenido de carbohidratos no excede el 76.5%.

2. Generar 100 observaciones de una distribución $\mathcal{N}(0,1)$ y calcular un intervalo de confianza del 95% para la función de distribución acumulada F.

Repetir esto 1000 veces y calcular la cuántas veces el intervalo de confianza contiene la verdadera función de distribución.

3. Repetir el ejercicio anterior para la distribución de Cauchy y la exponencial.

4. En 1975 se llevó a cabo un experimento para ver si la siembra de nubes producía lluvia. Se sembraron 26 nubes con nitrato de plata y otras 26 no fueron sembradas. En `https://dasl.datadescription.com/datafile/cloud-seeding/` se encuentra el dataset con la cantidad de lluvia que generó cada nube en cada uno de los dos grupos. Llamamos θ a la diferencia entre la precipitación media de los dos grupos.

(a) Estimar θ

(b) Estimar el error estándar de $\hat{\theta}$ y producir un intervalo de confianza de 95%

Modelos paramétricos

5. Considere muestras aleatorias de cada una de las siguientes distribuciones:

i. normal de parámetros μ y σ^2;

ii. exponencial de parámetro λ;

iii. Poisson de parámetro λ;

iv. con PDF que depende de un parámetro $\theta \in (0,1)$;

$$f(x;\theta) = \frac{1}{\theta} x^{\left(\frac{1}{\theta}-1\right)} I_{[0,1]}(x), \quad \text{donde } I_{[0,1]}(x) = \begin{cases} 1 & \text{si } 0 \le x \le 1, \\ 0 & \text{caso contrario;} \end{cases}$$

v. geométrica de parámetro p;

vi. gamma de parámetros α y λ;

vii. uniforme $\mathcal{U}[0,\theta]$.

(*a*) Hallar en cada caso el estimador de momentos de los parámetros

(*b*) Hallar en cada caso salvo vi el estimador de máxima verosimilitud de los parámetros (salvo vi)

(*c*) Para i,ii,iii y vii decir si los estimadores son insesgados o asintóticamente insesgados.

(*d*) Calcular el ECM de los estimadores de θ en vii. Comparando uno con otro, ¿cuál de los dos estimadores usarías?

6. Con cierto instrumento se realizan 20 mediciones de una magnitud física μ. Cada observación es de la forma $X = \mu + \varepsilon$, donde ε es el error de medición (aleatorio). Se obtienen los siguientes datos:

25.11,	25.02,	25.16,	24.98,	24.83,	25.05,	24.94,	25.04,	24.99,	24.96,
25.03,	24.97,	24.93,	25.12,	25.01,	25.12,	24.90,	24.98,	25.10,	24.96

Suponiendo que los errores de medición tienen distribución normal con media cero y varianza 0.01, estimar μ. ¿Cuál es la varianza del estimador de μ?

7. Se supone que la longitud en milímetros de cierto tipo de eje tiene una distribución normal con desvío estándar $\sigma = 0.05$. Se toma una muestra de 20 ejes y se observa que la longitud media de los ejes es de 52.3.

(*a*) Hallar un intervalo de confianza para la verdadera longitud media de nivel 0.99.

(*b*) ¿Qué tamaño debe tener la muestra para que la longitud de un intervalo de nivel 0.99 sea a lo sumo 0.03?

8. La nota de una prueba de aptitud sigue una distribución normal. Una muestra aleatoria de nueve alumnos de la ciudad arroja los siguientes resultados: 5, 8.1, 7.9, 3.3, 4.5, 6.2, 6.9, 7.5, 9.1.

(*a*) Hallar un intervalo de confianza para la nota media de los alumnos de la ciudad.

(*b*) La nota media de todos los alumnos de la ciudad en ese mismo examen el año anterior es 7.50. ¿Le parece que hay motivos para afirmar que la nota media de los alumnos ha cambiado con respecto al año anterior?

9. Sean $X_1, \dots, X_n$ vaiid distribuidas como una Bernoulli p para cierto $p \in (0,1)$ y $n \in \mathbb{N}$. Fijado $\alpha > 0$, definimos

$$\hat{p}_n = \frac{1}{n-1}\sum_{i=1}^{n-1} X_i, \qquad \varepsilon_n = \left(\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)\right)^{1/2},$$

y, finalmente, $C_n = (\hat{p}_n - \varepsilon_n, \hat{p}_n + \varepsilon_n)$.

(*a*) Sean $\alpha = 0.05$ y $p = 0.4$. Simular en la computadora para ver con qué frecuencia el intervalo C_n contiene a p. Hacer esto para varios valores de n entre 1 y 10000. Graficar la cobertura, es decir la proporción de veces que el intervalo contiene a p, versus n.

(*b*) Graficar la longitud del intervalo versus n. Supongamos que queremos que la longitud del intervalo no sea más grande que 0.05. ¿Cuán grande debería ser n?