Probabilidad y estadística (M) - Práctica 10 - Estimadores y Test de hipótesis

1. Consideremos muestras aleatorias $X_1, ..., X_n$ para cada una de las siguientes distribuciones:

   i. exponencial de parámetro θ.

   ii. Poisson de parámetro θ.

   iii. con densidad
$$f(x,\theta) = \frac{1}{\theta} x^{(\frac{1}{\theta}-1)} \mathbb{I}_{[0,1]}(x), \quad \theta > 0.$$

   iv. geométrica de parámetro θ.

   (a) Encontrar en cada caso el estimador de máxima verosimilitud y el de momentos de θ.

   (b) En los dos primeros casos y para el estimador de máxima verosimilitud del tercero, decir si los estimadores obtenidos son insesgados o asintóticamente insesgados.

   (c) Decir si los estimadores obtenidos son consistentes. Justificar.

2. Sea $X_1, ..., X_n$ una muestra aleatoria de una distribución discreta Y a valores $y_1, ..., y_k$ con probabilidad $p_1, ..., p_k$ respectivamente.

   (a) Obtener el EMV para los valores $p_1, ..., p_k$ en el caso en que $p_1 > 0, ...p_k > 0$ y en la muestra se observaron todos los valores posibles de Y.

   (b) ¿Podría suceder que se observara el valor y_1 en la muestra y que p_1 sea 0?

   (c) Supóngase que y_1 no fue observado en la muestra, mientras que sí fueron observados los valores restantes. Obtener el EMV para los valores $p_1, ..., p_k$ en el caso en que $p_1 \geq 0, p_2 > 0, ..., p_k > 0$.

   (d) Concluir para el caso en que $p_1 \geq 0, p_2 \geq 0, ..., p_k \geq 0$ que los estimadores son
$$\tilde{p_j} = \frac{\sum_{i=1}^{n} \mathbb{I}_{y_j}(X_i)}{n} = \frac{\#\{i : X_i = y_j\}}{n} \quad \text{para } j = 1, ..., n.$$

3. 4. Sea $X_1, ..., X_n$ una muestra aleatoria de una distribución con densidad
$$f(x,\theta) = e^{-(x-\theta)} \mathbb{I}_{(0,+\infty)}(x)$$

   (a) Probar que $T = \min_{1 \leq i \leq n} X_i$ es el estimador de máxima verosimilitud de θ.

   (b) Calcular el estiamdor de momentos de θ.

   (c) Decir si los estimadores obtenidos son insesgados o asintóticamente insesgados, y consistentes. Justicar.

4. Se define el error cuadrático medio de un estimador $\hat{\theta}$ como $\mathrm{ECM}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$ Verificar que $\mathrm{ECM}(\hat{\theta}) = \mathbb{V}(\hat{\theta}) + (sesgo(\hat{\theta}))^2$, donde $sesgo(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$

5. Denotemos por μ al verdadero promedio de nivel de radiactividad (picocuries por litro). El valor 5 pCi/L se considera como la línea divisoria entre agua segura y no segura. ¿Qué se debería testear $H_0 : \mu = 5$ versus $H_A : \mu > 5$ ó $H_0 : \mu = 5$ versus $H_A : \mu < 5$? Explicar el razonamiento en términos de los errores de tipo I y tipo II.

6. Un fabricante de automóviles afirma que un 0km de determinado modelo tiene un rendimiento medio de, al menos, 12 km por litro de nafta. Un grupo de investigación de mercados, sospechando de tal afirmación, realiza un test. Se considera que la distancia recorrida por litro tiene distribución normal con desviación estándar $\sigma = 2$, y se toma una muestra aleatoria de 30 automóviles.

   (a) Hallar la región de rechazo del test para el nivel $\alpha = 0.02$.

   (b) Hallar la probabilidad de error de tipo II si el rendimiento real es de 11 km por litro.

   (c) Hallar la probabilidad de error de tipo I si el rendimiento real es de 12.5 km por litro. ¿Es más chica que 0.02?

7. Un artículo publicado en una revista norteamericana reportó que el tiempo medio libre por semana para los hombres de ese país es de 40 horas. Un investigador, sorprendido por lo alto de la cifra, decidió realizar un test al respecto. Para ello, eligió al azar una muestra aleatoria de 60 hombres en donde obtuvo una media de 37.8 horas libres por semana y un desvío estándar muestral de 12.2 horas. ¿Puede concluirse a nivel 0.05 que la información del artículo es falsa?

8. Se supone que 1 de cada 10 fumadores prefiere la marca A. Después de una campaña publicitaria en cierta región de ventas, se entrevistó a 200 fumadores para determinar la efectividad de la campaña. El resultado de esta encuesta mostró que 26 personas preferían la marca A.

   (a) ¿Indican estos datos, a nivel aproximado 0.05, un aumento en la preferencia por la marca A?

   (b) Calcular el p-valor

   (c) ¿Cuál es la probabilidad aproximada de decidir que la campaña publicitaria no fue efectiva, cuando en realidad la proporción de preferencia por la marca A después de la campaña es 0.15?

   (d) ¿Qué tamaño de muestra debería tomarse para que la probabilidad de c) fuese a lo sumo 0.05?