

Inferencia Estadística

Tarea 3 21/09/2024

Escriba de manera concisa y clara sus resultados, justificando los pasos necesarios. Serán descontados puntos de los ejercicios mal escritos y que contenga ecuaciones sin una estructura gramatical adecuada. Las conclusiones deben escribirse en el contexto del problema. Todos los programas y simulaciones tienen que realizarse en R.

1. Sean $\mu \in \mathbb{R}$ y $\beta > 0$. Considera la siguiente función

$$F(t) = \frac{1}{1 + e^{-(t-\mu)/\beta}} \quad t \in \mathbb{R}.$$

Justifica que F es una función de distribución. La función F anterior define a la denominada distribución Logística(μ, β). Calcula la función de riesgo (hazard) asociada a la distribución Logística(μ, β).

2. Sea $X \sim \text{Normal}(0, 1)$. Calcula los momentos pares e impares de X ; es decir, calcula $E(X^p)$, para $p = 2k$ y $p = 2k - 1$, con $k \in \mathbb{N}$. **Nota:** Conviene considerar la diferencia entre par e impar para facilitar la cuenta. Usa la función generadora de momentos e investiga el concepto de doble factorial.
3. En este ejercicio visualizaremos el Teorema de Moivre-Laplace (TML). Para $p = 0.1$ y $A = \{5, 10, 20, 50, 100, 500\}$, grafica lo siguiente:
 - a) Sobre la misma figura, grafica la función de masa $g(x)$ de una distribución Binomial(n, p) y una la función de densidad $f(x)$ de una distribución Normal(np, npq), para todo $n \in A$ (i.e. presenta las 6 figuras).
 - b) Haz lo mismo que en el inciso anterior pero ahora para las funciones de distribución acumuladas de las binomiales y normales anteriores.
 - c) ¿Cuál es la relación entre las figuras anteriores y el TML? ¿Cambia el resultado si uno toma $p = 0.5, 0.9$?
4. Una partícula se encuentra inicialmente en el origen de la recta real y se mueve en saltos de una unidad. Para cada salto, la probabilidad de que la partícula salte una unidad a la izquierda es p y la probabilidad de que salte una unidad a la derecha es $1 - p$. Denotemos por X_n a la posición de la partícula después de n unidades. Encuentre $E(X_n)$ y $\text{Var}(X_n)$. Esto se conoce como una caminata aleatoria en una dimensión.
5. El siguiente conjuntos de datos contiene mediciones del diámetro de un agave, medido en decímetros, en distintas localizaciones no cercanas.

23.37	21.87	24.41	21.27	23.33	15.20	24.21	27.52	15.48	27.19
25.05	20.40	21.05	28.83	22.90	18.00	17.55	25.92	23.64	28.96
23.02	17.32	30.74	26.73	17.22	22.81	20.78	23.17	21.60	22.37

- a) Escriba una función en R que calcule la función de distribución empírica para un conjunto de datos dado D . La función debe tomar como parámetros al valor x donde se evalúa y al conjunto de datos D . Utilizando esta función grafique la función de distribución empírica asociada al conjunto de datos de agave. Ponga atención a los puntos de discontinuidad. ¿Qué observa? **Nota:** Escriba la función mediante el algoritmo descrito en las notas de la clase; para este ejercicio no vale usar la funciones implementadas en R que hacen lo pedido.
 - b) Escriba una función en R que determine la gráfica Q-Q normal de un conjunto de datos. La función debe tomar como parámetro al conjunto de datos y deberá graficar contra el percentil estandarizado de la normal. Para poder comparar el ajuste más claramente, la función además deberá ajustar en rojo a la recta $sx + \bar{x}$ (s =desviación estándar muestral y \bar{x} =media muestral). Usando esta función, determine la gráfica Q-Q normal. ¿Qué observa? **Nota:** La misma del inciso a).
 - c) Añada a la función anteriores (función de distribución empírica y Q-Q normal) la opción de que grafiquen la banda de confianza, de cobertura $1 - \alpha$, basada en el estadístico de Kolmogorov-Smirnov. La función debe tomar como parámetros al conjunto de datos y el nivel de confianza $1 - \alpha$. Aplique esta función al conjunto de datos para un nivel de confianza $1 - \alpha = 0.95, 0.99$. ¿Qué observa? **Nota:** Recurra a las notas sobre las bandas de confianza de los gráficos Q-Q normales que se incluyeron en la clase 10; no vale usar la funciones implementadas en R que hacen lo pedido. No es necesario entender a detalle la prueba de Kolmogorov-Smirnov, en este punto solo consideraremos su aspecto operacional; al final del curso, en una de las exposiciones finales, se presentará la prueba con detalle.
 - d) Escriba una función en R que determine el gráfico de probabilidad normal. La función debe tomar como parámetro al conjunto de datos. ¿Qué observa? **Nota:** La misma del inciso a).
 - e) ¿Los datos anteriores se distribuyen normalmente? Argumente.
6. En este ejercicio se comprobará que tan buena es la aproximación dada por las reglas empíricas para algunas de las distribuciones estudiadas en la clase. Considerese las distribuciones $\text{Unif}(a = -3, b = 3)$, $\text{Normal}(0, 1)$, $\text{Exponencial}(2)$, $\text{Gamma}(\alpha = 2, \beta = 1)$, $\text{Gamma}(\alpha = 3, \beta = 1)$, $\text{Beta}(\alpha = 2, \beta = 2)$, $\text{Weibull}(\alpha = 4, \beta = 1)$ y $\text{Lognormal}(\mu = 3, \sigma = 2)$.
- a) Leer las reglas empíricas en https://en.wikipedia.org/wiki/68%E2%80%9395%E2%80%9399.7_rule
 - b) Para cada una de las distribuciones anteriores, haga una tabla que muestre las probabilidades contenidas en los intervalos $(\mu - k\sigma, \mu + k\sigma)$, para $k = 1, 2, 3$. Utilice las fórmulas de las medias y varianzas contenidas en las notas para determinar μ y σ en cada caso. Puede usar R para determinar las probabilidades pedidas.
 - c) En R, simule $n = 1000$ muestras de cada una de las distribuciones anteriores y calcule la media muestral \bar{x} y la varianza muestral s^2 como se mencionó en la clase. En cada caso, calcule la proporción de observaciones que quedan en los intervalos $(\bar{x} - ks, \bar{x} + ks)$, para $k = 1, 2, 3$. Reporte sus hallazgos en una tabla como la del inciso anterior. ¿Qué tanto se parecen la tabla de este inciso y la del anterior?

Entrega: 04/10/2024.