Inferencia Estadística

Tarea 3

18/09/2018

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. En este ejercicio corroborará mediante simulaciones la Ley de Grandes Números (LGN).

   a) Consulte su libro favorito de probabilidad y escriba la Ley Fuerte de Grandes Números (LFGN). Explíque el resultado con sus propias palabras.

   b) Simule una muestra $\{x_1, \ldots, x_n\}$ de una v.a. $Normal(\pi, \sqrt{2})$ de tamaño $n = 10^5$. Defina $y_m = \sum_{i=1}^{m} x_i/m$ y grafique esta cantidad. ¿Qué observa? ¿Cómo está esto relacionado con la LGN?

   c) Repita el procedimiento anterior 100 veces y grafique las y_m de cada iteración sobre una misma gráfica. ¿Qué observa?

   d) Repita los dos incisos anteriores para una distribución $Cauchy(\pi, \sqrt{2})$. ¿Qué observa?

2. Resuelva lo siguiente:

   a) Let $X \sim Exponencial(\beta)$. Encuentre $P(|X - \mu_X| \geq k\sigma_X)$ para $k > 1$. Compare esta probabilidad con la que obtiene de la desigualdad de Chebyshev.

   b) Sean $X_1, \ldots, X_n \sim Bernoulli(p)$ y $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Usando las desigualdades de Chebyshev y Hoeffding, acote $P(|\bar{X}_n - p| > \epsilon)$. Demuestre que para n grande la cota de Hoeffding es más pequeña que la cota de Chebyshev. ¿En qué beneficia esto?

3. Sean $X_1, \ldots, X_n \sim Bernoulli(p)$.

   a) Sea $\alpha > 0$ fijo y defina

   $$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{n}\right)}.$$

   Sea $\hat{p_n} = n^{-1} \sum_{i=1}^{n} X_i$. Defina $C_n = (\hat{p_n} - \epsilon_n, \hat{p_n} + \epsilon_n)$. Use la desigualdad de Hoeffding para demostrar que

   $$P(C_n \text{ contiene a } p) \geq 1 - \alpha.$$

   Diremos que C_n es un $(1-\alpha)$-intervalo de confianza para p. En la práctica, se trunca el intervalo de tal manera de que no vaya debajo del 0 o arriba del 1.

   b) Sea $\alpha = 0.05$ y $p = 0.4$. Mediante simulaciones, realice un estudio para ver que tan a menudo el intervalo de confianza contiene a p (la cobertura). Haga esto para $n = 10, 50, 100, 250, 500, 1000, 2500, 5000, 10000$. Grafique la cobertura contra n.

c) Grafique la longitud del intervalo contra n. Suponga que deseamos que la longitud del intervalo sea menor que 0.05. ¿Qué tan grande debe ser n?

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. La función debe tomar como parámetros al punto 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 lluvias. Ponga atención a los puntos de discontinuidad. ¿Qué observa?

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. Usando esta función, determine la gráfica Q-Q normal. ¿Qué observa?

c) Añada a la función anterior la opción de que grafique 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?

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?

e) ¿Los datos anteriores se distribuyen normalmente? Argumente.

6. En este ejercicio repasará la estimación de densidades.

a) Escriba una función en `R` que calcule el estimador de la densidad por el método de kerneles. La función deberá recibir al punto x donde se evalúa al estimador, al parámetro de suavidad h, al kernel que se utilizará en la estimación y al conjunto de datos.

b) Cargue en `R` al archivo "Tratamiento.csv", el cual contiene la duración de los períodos de tratamiento (en días) de los pacientes de control en un estudio de suicidio. Utilice la función del inciso anterior para estimar la densidad del conjunto de datos para $h = 20, 30, 60$. Grafique las densidades estimadas. ¿Cuál es el mejor valor para h? Argumente.

7. Cargue en `R` al conjunto de datos "Maíz.csv", el cual contiene el precio mensual de la tonelada de maíz y el precio de la tonelada de tortillas en USD. En este ejercicio tendrá que estimar los coeficientes de una regresión lineal simple.

a) Calcule de forma explícita la estimación de los coeficientes via mínimos cuadrado y ajuste la regresión correspondiente. Concluya.

b) Calcule de forma explícita la estimación de los coeficientes via regresión no-paramétrica tipo kernel (ver Nadaraya, E. A. (1964). "On Estimating Regression". Theory of Probability and its Applications. 9 (1): 141–2. doi:10.1137/1109020) y ajuste la regresión correspondiente. Concluya.

c) Compare ambos resultados. ¿Qué diferencias observa?

8. Demuestre que la fórmula de la densidad de la Beta integra 1.

Entrega: 02/10/2018.