Inferencia Estadística

Tarea 4

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. Resuelva lo siguiente:

   a) Sea $X \sim \mathit{Exponencial}(\beta)$. 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 $X_1, \ldots, X_n \sim \mathit{Bernoulli}(p)$ 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?

2. Sean $X_1, \ldots, X_n \sim \mathit{Bernoulli}(p)$.

   a) Sea $\alpha > 0$ fijo y defina
   $$\epsilon_n = \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}.$$
   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?

3. Considera el problema 5 de la Tarea 3. Usando la desigualdad de Dvoretzky-Kiefer-Wolfowitz, escriba una función en `R` que calcule y grafique una región de confianza para la función de distribución empírica. La función debe tomar como parámetros al conjunto de datos que se usan para contruir la función de distribución empírica.

4. En este ejercicio repasará la estimación de densidades.

   a) Escriba una función en `R` que estime una 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.

c) En el contexto de la estimación de densidades, escriba una función en `R` que determine el ancho de banda que optimiza al ISE. Grafique la densidad con ancho de banda óptimo para el conjunto de datos de “Tratamiento.csv”.

5. Considera el siguiente experimento en dos etapas: primero se escoge un punto X con distribución uniforme en $(0, 1)$; después se elige un punto Y con distribución uniforme en $(-X, X)$. El vector aleatorio (X, Y) representa el resultado del experimento. ¿Cuál es su densidad conjunta de (X, Y)? ¿Cuál es la densidad marginal de Y ? ¿Cuál es la densidad condicional de X dada Y?

6. 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 cuadrados 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?