Estadística No Paramétrica

Tarea 4

Fecha: 6 de marzo de 2023

Fecha de entrega: miércoles 17 de marzo de 2023

1.

Residuos de aflatoxin en crema de cacahuate: En una preuba real, 12 lotes de crema de cacahuate tienen residuos de aflatoxin en partes por mil millones de 4.94, 5.06, 4.53, 5.07, 4.99, 5.16, 4.38, 4.43, 4.93, 4.72, 4.92 y 4.96.

- a) ¿Cuántas posibles muestras bootstrap hay?
- b) Calcular 50 muestras bootstrap, (fijar la semilla en 100) y encontrar la media de los correspondientes elementos de la muestra del conjunto de datos.
- c) Calcular la media de las medias bootstrap. Comparar con la media del conjunto de datos original.
- d) Encontrar el mínimo y el máximo de las 50 medias bootstrap. Este es un intervalo de confianza muy crudo de la media.

2.

Accidentes de aerolíneas: De acuerdo al U.S. National transportation safety Board, el número de accidentes aéreos por año de 1983 a 2006 fueron:

```
accidentes <- c(23,16,21, 24, 34, 30, 28, 24, 26, 18, 23, 23, 36, 37, 49, 50, 51,
                56, 46, 41, 54, 30, 40, 31)
```

- a) Para los datos de la muestra, calcular la media y su error estándar, y la mediana.
- b) Calcular muestras jackknife de la media y la mediana con estimadores de sus errores estándar. También calcular la mediana de los estimados de medianas
- c) Repetir (b) con $B = 1000$ muestras bootstrap (fijando la semilla en 1)
- d) Comparar (a),(b) y (c). ¿Cómo se comparan los resultados de cada método?

3.

Con los datos del ejercicio 1:

- a) Calcular un intervalo de confianza de 95% para la media usando la distribución t de Student estándar.
- b) Encontrar un intervalo de confianza de 95% usando el método percentil.
- c) Encontrar un intervalo de confianza de 95% usando el método BCa.
- d) Encontrar un intervalo de confianza de 95% usando el método percentil-t.
- e) ¿Cómo se comparan los intervalos? ¿En qué intervalos confían?

4.

Supongan que $\hat{\theta}$ es una estadística lineal de la forma

$$\hat{\theta} = \mu + \frac{1}{n}\sum_{i=1}^{n} \alpha(x_i)$$

Supongan que sabemos el valor de $\hat{\theta}$ para cada conjunta jackknife $\mathbf{x}_{-i}$, esto es, $T_n(\mathbf{x}_{-i}) = b_i$ para $i = 1, \ldots, n$

- a) Sea $\alpha_i = \alpha(X_i)$ y resuelvan el conjunto de n ecuaciones lineales

$$b_i = \mu + \sum_{j \neq i} \alpha_j/(n-1), \qquad i = 1, 2 \ldots, n$$

para $\alpha_1, \ldots, \alpha_n$.

- b) Luego, deduzcan el valor de $\hat{\theta}$ de cualquier conjunto de datos bootstrap $x_1^*, \ldots, x_n^*$.

5.

Mostrar que para $\hat{\theta} = \bar{x}$, $\hat{\theta}_{(\cdot)} - \hat{\theta} = 0$ y por lo tanto el estimado jackknife de sesgo es 0.

6.

Generar 100 muestras $X_1, \ldots, X_{20}$ de una población normal $\mathcal{N}(\theta, 1)$ con $\theta = 1$

- a) Para cada muestra calcular los estimados bootstrap y jackknife de varianza paea $\hat{\theta} = \bar{X}$ y calcular la media y desviación estándar de estos estimadores de varianza sobre las 100 muestras.
- b) Repetir (a) para la estadística $\bar{X}^2$ y comparar resultados. Den una explicación de lo que encontraron.

7.

Supongamos que X_i/θ se distribuyen independiente e identicamente distribuídas como χ_1^2 para $i = 1, \ldots, 20$. Realizar un pequeño estudio de simulación para comparara los siguientes intervalos para θ basados en $\bar{X}$, suponiendo que el verdadero valor de θ es uno:

- a) El intervalo exacto basado en $20 \cdot \hat{\theta}/\theta \sim \chi_{20}^2$.
- b) El intervalo estándar basado en $(\hat{\theta} - \theta)/\hat{se} \sim \mathcal{N}(0, 1)$, donde $\hat{se}$ es el estimador plugin del error estándar de la media.
- c) El intervalo bootstrap-t basado en $(\hat{\theta} - \theta)/\hat{se}$

Usar al menos 1000 muestras en la simulación, y para cada intervalo calcular la falta de cobertura en cada cola y la falta de cobertura total, así como la media y desviación estándar de la longitud del intervalo. Discutir los resultados.