Introducción al Aprendizaje Estadístico - Actividades de Clase - Unidad 3
Lucía Babino - Paula Spano

Modelo uniforme $\mathcal{U}(0, \theta)$

Sean $X_1, \ldots, X_n \sim \mathcal{U}(0, \theta)$ variables aleatorias independientes e idénticamente distribuidas (i.i.d.) con distribución uniforme en el intervalo $[0, \theta]$. Consideremos los siguientes estimadores de θ basados en la muestra :

$$\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n) = 2\overline{X}_n \ , \quad \widetilde{\theta}_n = \widetilde{\theta}_n(X_1, \ldots, X_n) = \max\{X_1, \ldots, X_n\}. \tag{1}$$

Ejercicio 1: Estimación

1. Implementar funciones `est1` y `est2` que tengan por argumento un conjunto de `datos` $(x_1, \ldots, x_n)$ y devuelvan el valor de la estimación $\widehat{\theta}_n(x_1, \ldots, x_n)$ y $\widetilde{\theta}_n(x_1, \ldots, x_n)$, para los estimadores definidos en (1), respectivamente.

   Calcular el valor de dichas estimaciones en cada uno de los siguientes conjuntos de datos:

   a) `1.17 1.75 0.28 2.56 2.36 0.36 1.82 0.24 1.17 1.86`

   b) `0.66 0.07 0.62 0.65 1.33 0.40 1.17 1.11 2.01 2.98`

   Simulación: En los siguientes ítems generaremos datos provenientes de variables con distribución uniforme en el intervalo $[0, 3]$. Es decir, trabajaremos con una muestra aleatoria $X_1, \ldots, X_n \sim \mathcal{U}(0, \theta)$ con $\theta = 3$.

2. Generar $n = 5$ datos correspondientes a una muestra aleatoria con distribución $\mathcal{U}(0, 3)$.

3. Calcular el valor de la estimación correspondiente a $\widehat{\theta}$ en base a los datos generados. (*Estamos omitiendo el subíndice n en $\widehat{\theta}$ para simplificar la notación, ya que n ahora está fijo*)

4. Repetir los dos pasos anteriores $m = 1000$ veces, obteniendo así m realizaciones de $\widehat{\theta}$ a las que llamaremos $\widehat{\theta}^1, \ldots, \widehat{\theta}^m$. Realizar un histograma y un boxplot con estos valores.

5. Repetir los ítems 2 a 4 para $n = 30$, $n = 50$ y $n = 100$ y graficar los cuatro histogramas juntos, uno debajo del otro y los cuatro boxplots, uno al lado del otro. ¿Qué puede decir de la distribución de $\widehat{\theta}_n$ a medida que n aumenta?

6. Repetir los ítems 2 a 5 para $\widetilde{\theta}_n$.

Ejercicio 2: Trade off sesgo-varianza

Utilizaremos las $m = 1000$ realizaciones $\widehat{\theta}^1, \ldots, \widehat{\theta}^m$ de $\widehat{\theta}$ y $\widetilde{\theta}^1, \ldots, \widetilde{\theta}^m$ de $\widetilde{\theta}$, obtenidas en base a las muestras de tamaño $n = 5$.

Sesgo:

7. Computar el *Sesgo empírico* de $\widehat{\theta}$ mediante la expresión

$$\left(\frac{1}{m} \sum_{i=1}^{m} \widehat{\theta}^i \right) - \theta$$

¿Qué cantidad está estimando el *Sesgo empírico*?

8. Repetir el ítem anterior para $\widetilde{\theta}$.

9. ¿Qué estimador tiene menor *Sesgo empírico*? ¿podría anticipar la respuesta en base a los histogramas y/o boxplots del ejercicio anterior?

Varianza:

10. Computar la *Varianza empírica* de $\widehat{\theta}$ mediante la expresión

$$\frac{1}{m} \sum_{i=1}^{m} \left(\widehat{\theta}^i - \overline{\widehat{\theta}} \right)^2$$

¿Qué cantidad está estimando la *Varianza empírica*?

11. Repetir el ítem anterior para $\widetilde{\theta}$.

12. ¿Qué estimador tiene menor *Varianza empírica*? ¿podría anticipar la respuesta en base a los histogramas y/o boxplots del ejercicio anterior?

Error Cuadrático Medio:

13. Computar el *Error cuadrático medio empírico* de $\widehat{\theta}$ mediante la expresión

$$EMSE = \frac{1}{m} \sum_{i=1}^{m} (\widehat{\theta}^i - \theta)^2$$

¿Qué cantidad está estimando el *Error cuadrático medio empírico*?

14. Repetir el ítem anterior para $\widetilde{\theta}$.

15. ¿Qué estimador tiene menor *Error cuadrático medio empírico*? ¿podría anticipar la respuesta en base a los histogramas y/o boxplots del ejercicio anterior?