



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico 2

17 de junio de 2019

Probabilidad y Estadística

Integrante	LU	Correo electrónico
Capello, Bruno	007/007	bruno.icapello@gmail.com
De Sousa Bispo, Germán	359/12	german_nba11@hotmail.com
Serapio, Noelia	004/15	noeliaserapio@gmail.com



Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (++54 +11) 4576-3300

<http://www.exactas.uba.ar>

1. Ejercicio 1

Para la implementación de estos casos se utilizaron las funciones `mean()` y `sd()` de R, dado que ambas proveen la media muestral y el desvío estándar muestral. Se hicieron varias pruebas (a mano) contra las definiciones de media muestral y varianza muestral y en todos los casos el resultado fue igual.

1.1. Estimador de momentos

Sea $X_i \sim U(0, \theta)$

$$E(x_i) = \theta/2$$

Entonces:

$$\frac{\sum_{i=1}^n x_i}{n} = 2\hat{\theta}_n$$
$$\hat{\theta}_n = \frac{2}{n} \sum_{i=1}^n x_i$$

1.2. Estimador de máxima verosimilitud

Sea $X_i \sim U(0, \theta)$

$$L(X_1, X_2, \dots, X_n, \theta) = \frac{1}{\theta^n} \mathbb{1}_{0 \leq x_i \leq \theta, \forall x_i}(x_i)$$

$$L(X_1, X_2, \dots, X_n, \theta) = \frac{1}{\theta^n} \mathbb{1}_{\max(x_i) \leq \theta}(x_i)$$

Entonces:

$$\hat{\theta}_n = \max(x_i) \text{ con } i \leq n$$

2. Ejercicio 2

Para la implementación de este estimador, se utilizó la función `median()` que retorna la mediana de un vector.

3. Ejercicio 3

En este ejercicio, la muestra se genera aleatoriamente con la función `runif(15, min=0, max=b)`. Luego de eso se inserta esa muestra en nuestras funciones de estimación.

Dado que cada ejecución tiene una muestra aleatoria distinta, presentamos las estimaciones impresas por pantalla en nuestro código en R.

4. Ejercicio 4

Cada uno de los pasos para realizar este experimento están diferenciados con un comentario en nuestro código en R donde indica que inciso del ejercicio es.

5. Ejercicio 5

6. Ejercicio 6

7. Ejercicio 7

8. Ejercicio 8

9. Ejercicio 9

Para este ejercicio se hizo una pequeña adaptación de las simulaciones hechas para cada estimador en el ejercicio 5.

Estas modificaciones comprenden “ensuciar” la muestra aleatoria antes del calculo así como agregar el retorno de error cuadrático medio.

9.1. Cálculo de probabilidad de muestra contaminada

Dado que iteramos cada dato de nuestra muestra (de tamaño 15) y tiramos una moneda (hecho con una $Bi(1, 0,005)$ que es lo mismo que un $Ber(0,005)$) para ver si modificamos el dato o no.

Se puede pensar lo siguiente:

Sea X : cantidad de datos contaminados de la muestra.

Entonces $X \sim Bi(15, 0,005)$

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0)$$

Veamos primero:

$$P(X = 0) = \binom{15}{0} 0,005^0 (1 - 0,005)^{15-0}$$

$$P(X = 0) = (0,995)^{15}$$

$$P(X = 0) \simeq 0,9275$$

Entonces

$$P(X \geq 1) \simeq 1 - 0,9275 \simeq 0,0725$$

Por lo tanto, la probabilidad de que una muestra este contaminada es de 0,0725

9.2. Aproximaciones conseguidas

En este ejercicio, la muestra se genera aleatoriamente con la función $runif(15, min=0, max=b)$. Y encima también se le agrega un error con cierta probabilidad antes de insertar esa muestra en nuestras funciones de estimación.

Dado que cada ejecución tiene una muestra aleatoria distinta, presentamos las estimaciones impresas por pantalla en nuestro código en R.

9.3. Elección de estimador

Luego de ejecutar varias veces nuestro código para los tres estimadores, y comparando los resultados podemos notar ciertos patrones.

El sesgo, la varianza y el ECM del estimador de máxima verosimilitud siempre es ampliamente más grande que los demás estimadores, por lo que da la impresión de no ser el mejor estimador.

Por otro lado, el sesgo, varianza y ECM del estimador de momentos, a pesar de ser mucho mejor que lo obtenido para el de máxima verosimilitud, suele ser siempre un orden más grande que los obtenidos para el estimador que utiliza la mediana.

Analizando simplemente los dos estimadores, podemos ver que el de momentos depende de la media muestral (que es bastante propensa a modificaciones fuertes por un outlier). Por otro lado, el de máxima verosimilitud depende del máximo dato obtenido, que dado que nuestra muestra puede tener outliers más grandes que el resto de nuestros valores (por multiplicar por 100), es aún más propenso a estimar incorrectamente tomando ese dato.

Finalmente, la mediana es menos propensa a outliers, dado que, como nuestra probabilidad de agregar outliers es baja, la mayoría de los datos van a estar acumulados entre 0 y θ , por lo que es muy probable que nuestra mediana este en ese intervalo y que a su vez los datos sean uniformes. Esto hace que la presencia de outliers no afecte en gran medida a nuestro estimador.