DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico 2

20 de junio de 2019 Probabilidad y Estadística

Integrante	LU	Correo electrónico
Capello, Bruno	623/17	bruno.icapello@gmail.com
De Sousa Bispo, Germán	359/12	german_nba11@hotmail.com
Serapio, Noelia	871/03	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 max() de R, dado que ambas proveen la media muestral y el máximo muestral. Se hicieron varias pruebas (a mano) contra las definiciones de media muestral y en todos los casos el resultado fue igual.

1.1. Estimador de momentos

Sea X_i ~U(0, θ)

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

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 ~U(0, θ)

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

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

Entonces:

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

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. Para calcular el error tomamos módulo del valor del estimador menos el b.

Dado que cada ejecución tiene una muestra aleatoria distinta, presentamos las estimaciones y errores impresos por pantalla en nuestro código en R. Además agregamos aquí una tabla comparativa de un ejemplo de una estimación realizada con su respectivo error.

Estimador	Estimación	Error
Estimador de momentos	1.074311	0.07431135
Estimador de máxima verosimilitud	0.9652966	0.03470338
Estimador mediana	1.287941	0.2879408

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.

4.1. Varianza

Para calcular la varianza se utilizo la función sd() de R que devuelve el desvio estándar muestral, pues la fórmula para el calculo de la varianza muestral es:

$$\text{Varianza muestral} = S^2 \text{ donde S es el desvío estandar muestral}$$

4.2. Error cuadrático medio (ECM)

Para calcular el ECM se utilizó la siguiente fórmula:

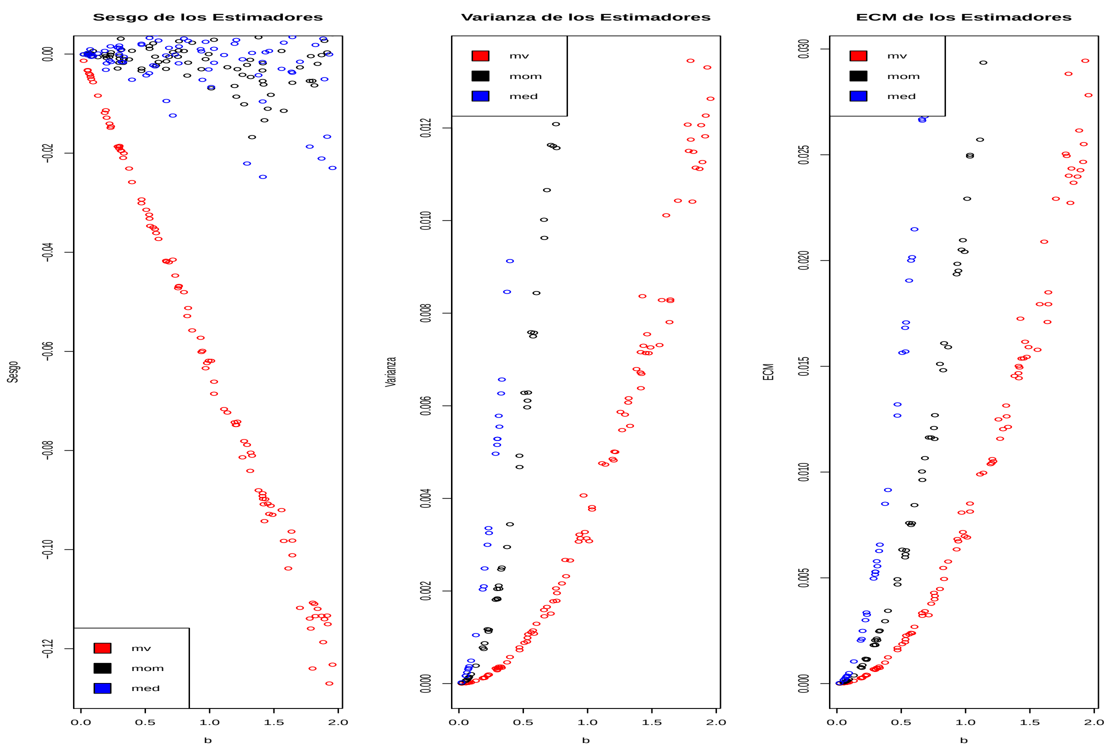
$$ECM_\theta(\hat{\theta}) = V_\theta(\hat{\theta}) + (sesgo_\theta(\hat{\theta}))^2$$

5. Ejercicio 5

Se realizaron las simulaciones de cada uno de los estimadores devolviendo el sesgo y la varianza. La cantidad de repeticiones en todos los casos fue 1000.

6. Ejercicio 6

6.1. Graficos



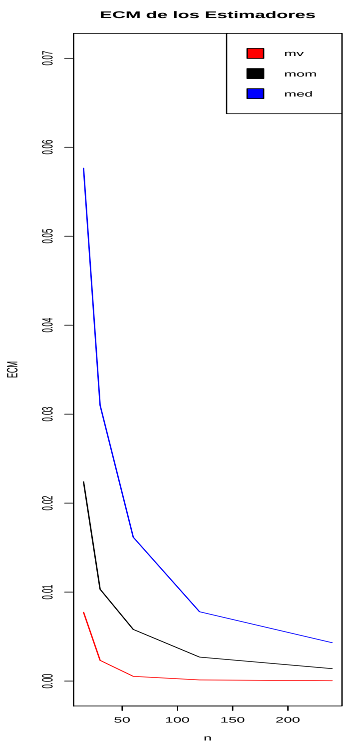
6.2. Observaciones y elección de estimador

Viendo los graficos podemos ver que :
Con respecto al sesgo, a medida que aumenta b, los estimadores de momentos y mediana mantienen poco sesgo pero el de maxima verosimilitud en cambio va aumentando.
Con respecto a la varianza y el ECM los graficos son muy parecidos, a medida que aumenta b, aumenta tanto la varianza como el ECM .

En este caso elegimos el estimador de maxima verosimilitud, ya que los estimadores de momentos y mediana como tienen poco sesgo van a tener mucha varianza a medida que aumenta b. En cambio el de máxima verosimilitud mantiene algo de sesgo y varianza por lo que parece más estable.

7. Ejercicio 7

7.1. Graficos



7.2. Observaciones y elección de estimador

Podemos observar para b = 1, que a medida que aumenta el n, el ECM que tienen los estimadores disminuye y que el estimador de maxima verosimilitud es el que menos error comete, por ende decidimos elegir el estimador de maxima verosimilitud en este caso.

7.3. Consistencia a partir de los Gráficos

Sea (X_i) iid que dependen de θ y sea $\hat{\theta_n}$ un estimador de θ si pasa :

$$1 : E_\theta(\hat{\theta_n}) \rightarrow \ \theta \ cuando \ n \rightarrow \infty$$

$$2 : V_\theta(\hat{\theta_n}) \rightarrow \ 0 \ cuando \ n \rightarrow \infty$$

entonces es consistente el estimador.

Viendo el grafico como

$$ECM(\hat{\theta_n}) \rightarrow 0 \ cuando \ n \rightarrow \infty \ y \ el \ ECM = sesgo^2 + varianza$$

podemos decir que son consistentes.

8. Ejercicio 8

8.1. Cálculo de los estimadores

Sea la siguiente muestra : [0.917, 0.247, 0.384, 0.530, 0.798, 0.912, 0.096, 0.684, 0.394, 20.1, 0.769, 0.137, 0.352, 0.332, 0.670]

Usando los estimadores calculados en el ejercicio 1 dan lo siguiente:
Maxima verosimilitud da 20.1
Momentos da 3.642933
Mediana da 1.06

8.2. Observaciones

Lo que podemos observar es que esta muestra dada tiene un solo punto outlier muy grande comparado con los demas puntos, lo que afecta mucho en este caso a la estimación de maxima verosimilitud y afecta en menor medida a la de momentos. Por eso en este caso elegimos el de la mediana que es la menos afectada y da una buena estimacion.

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. Además agregamos aquí una tabla comparativa de un ejemplo de una estimación realizada.

Estimador	Sesgo	Varianza	Error Cuadrático Medio
Estimador de momentos	0.4883285	4.036921	4.275386
Estimador de máxima verosimilitud	3.502678	227.5533	239.8221
Estimador mediana	0.004252857	0.05841771	0.0584358

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.