Inferencia Estadística: Proyecto Parte A

21 de junio de 2023

Universidad Nacional de Colombia

Mario Enrique Arrieta Prieto

Andrés Cadena Simons   Ander Steven Cristancho Sánchez
John Anderson Guarin Lopez   Juan David Carrascal Ibañez

Problema 1:

Distribución de un estimador máximo verosímil cuando no hay condiciones de regularidad.

Considere una muestra aleatoria $X_1, X_2, \cdots, X_n$ de una distribución $U[-\theta, \theta], \theta = k$, es decir, donde el verdadero valor del parámetro es igual al número del grupo. Se sabe (del taller 2) que $\widehat{\theta}_{MLE} = max\{-X^1, X^n\}$, sin embargo, como no se tienen condiciones de regularidad, este estimador no tiene una distribución asintótica normal (cuando es debidamente normalizado). Vamos a estudiar también la convergencia en distribución de la variable aleatoria $R_n = n(\theta - \widehat{\theta}_{MLE})$.

Solución:

a Encuentre la función de distribución exacta de $\widehat{\theta}_{MLE}$ y luego derívela para obtener la función de densidad correspondiente.
Esa distribución puede ser más fácilmente obtenida si primero prueban el siguiente lema: "Para una colección de números reales, $x_1, x_2, \cdots, x_n; max\{-x^{(1)}, x^{(n)}\} = max\{|x_1|, |x_2|, \cdots, |x_n|\}$**" (puede probarlo por inducción).**

Demostración. Por inducción:
Caso base n=2:
Note que $x^{(1)} y x^{(2)}$ son estadisticas de orden, por ende $x^{(1)} < x^{(2)}$, luego si $x^{(1)}$ fuera negativo, entonces $-x^{(1)} \geq 0$ y $-x^{(1)} = |x^{(1)}|$, además se tiene que $-x^{(1)} > -x^{(2)}$ luego $max\{-x^{(1)}, x^{(2)}\} = max\{|x_1|, |x_2|\}$
Hipótesis inductiva y paso inductivo:
Suponga que $max\{-x^{(1)}, x^{(n)}\} = max\{|x_1|, |x_2|, \cdots, |x_n|\}$, luego note que esto implica que $max\{-x^{(1)}, x^{(n+1)}\} = max\{|x_1|, |x_2|, \cdots, |x_n|\}$, ya que:
Suponga que $a = max\{|x_1|, |x_2|, \cdots, |x_n|\}$, luego note que usando el mismo razonamiento del caso base se puede demostrar que el $max\{a, |x_{n+1}|\} = max\{-x^{(1)}, x^{(n+1)}\}$.

∎

¿Qué distribución tiene $|X|$ si $X \sim U[-7,7]$?
Recordemos que la función generadora de momentos de $|X|$ se puede calcular como:

$$
\begin{aligned}
m_{|X|}(t) &= E[e^{t|X|}] \\
&= \int_{-\infty}^{\infty} e^{t|x|} * f_X(x)dx \\
&= 2\int_0^7 e^{tx} * \frac{1}{14}dx \\
&= \frac{1}{7}\int_0^7 e^{tx}dx \\
&= \int_0^7 e^{tx}\frac{1}{7}dx = m_Y(t)
\end{aligned}
$$

En donde $Y \sim U[0,7]$, luego como la función generadora de momentos identifica completamente una distribución, entonces al ser $m_{|X|}(t) = m_Y(t)$ significa que $|X|$ tiene la misma distribución que Y, así $|X| \sim U[0,7]$.

Halle la distribución de $\widehat{\theta}_{MLE}$ sabiendo que $\widehat{\theta}_{MLE} = max\{|X_1|, |X_2|, \cdots, |X_n|\}$ y $|X_1|, |X_2|, \cdots, |X_n|$ puede ser visto como una muestra aleatoria.
Suponga $A_1 = |X_1|$, luego se sabe que podemos realizar una estadística de orden sobre los A_i, por consiguiente tenemos que $\widehat{\theta}_{MLE} = A^{(n)}$, luego por anteriores resultados tenemos que:

$$
F_{Y^{(i)}}(y) = \sum_{j=i}^{n} [F_Y(y)]^j [1 - F_Y(y)]^{n-j}
$$

Ahora usando $\widehat{\theta}_{MLE} = A^{(n)}$ tenemos que:

$$
\begin{aligned}
F_{A^{(n)}}(a) &= \sum_{j=n}^{n} [F_A(a)]^j [1 - F_A(a)]^{n-j} \\
&= [F_A(a)]^n \\
&= [\frac{a}{7} * I_{[0,7)}(a) + 1 * I_{[7,\infty)}]^n \\
&= (\frac{a^n}{7^n}) * I_{[0,7)}(a) + 1 * I_{[7,\infty)}
\end{aligned}
$$

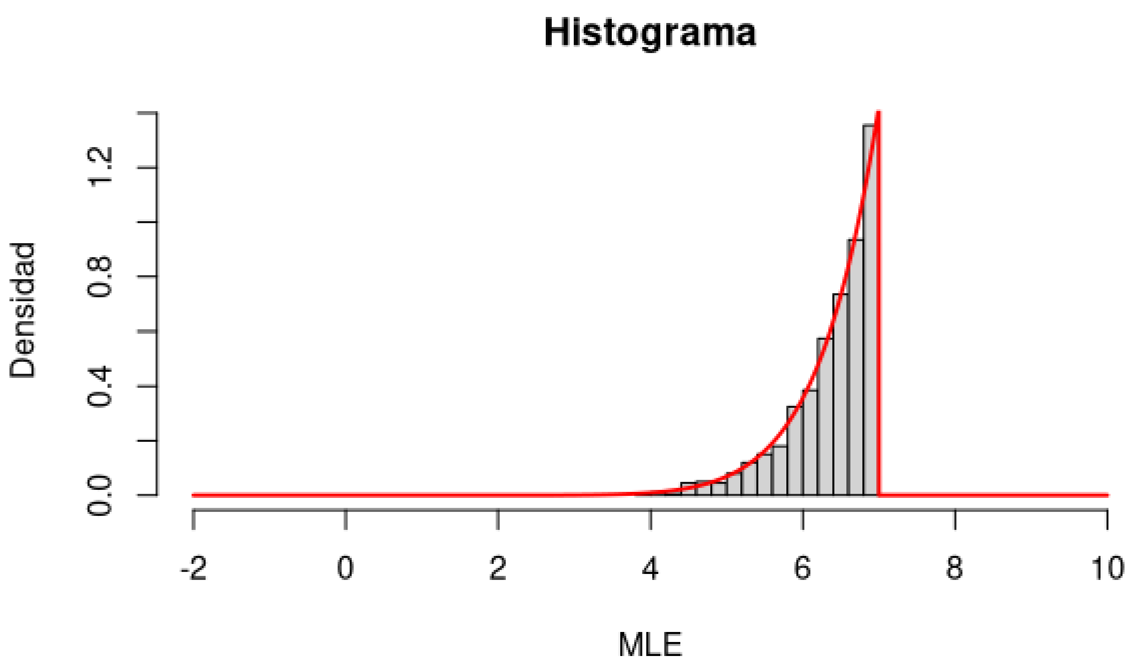
Luego si derivamos respecto a a tenemos que:

$$
\begin{aligned}
f_{A^{(n)}}(a) &= \frac{d}{da}(\frac{a^n}{7^n}) * I_{[0,7)}(a) + 1 * I_{[7,\infty)} \\
&= \frac{1}{7^n} * na^{n-1} * I_{[0,7)}(a)
\end{aligned}
$$

En conclusión:

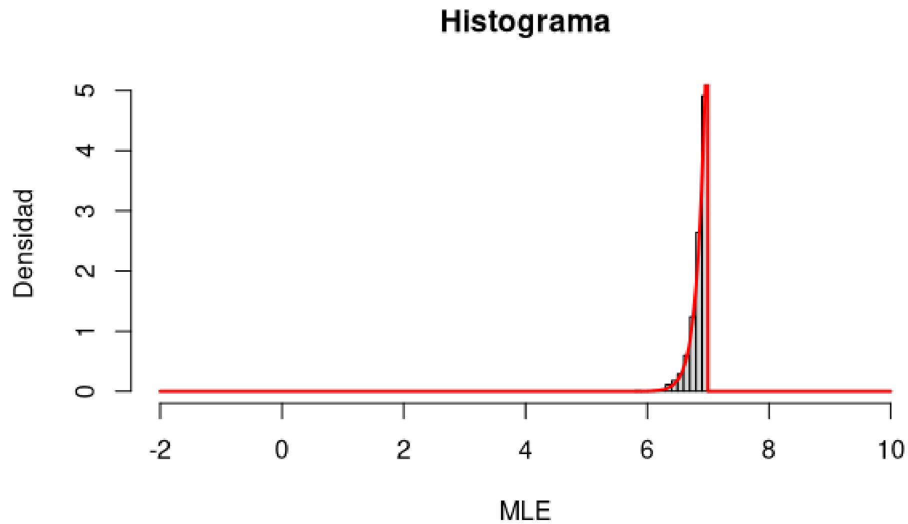
$$F_{\widehat{\theta}_{MLE}}(x) = (\frac{x^n}{7^n}) * I_{[0,7)}(x) + 1 * I_{[7,\infty)}$$
$$f_{\widehat{\theta}_{MLE}}(x) = \frac{1}{7^n} * nx^{n-1} * I_{[0,7)}(x)$$

b Genere $m = 1000$ simulaciones de una muestra aleatoria de tamaño $n = 10$ de la distribución $U[-\theta, \theta]$, $\theta = 7$, usando la función $runif$ de R o su equivalente en otro lenguaje de programación. Para cada una de las m muestras, calcule y almacene la estimación máximo-verosímil del parámetro. Haga un histograma de las m estimaciones y superponga sobre este histograma la función de densidad teórica de $\widehat{\theta}_{MLE}$.

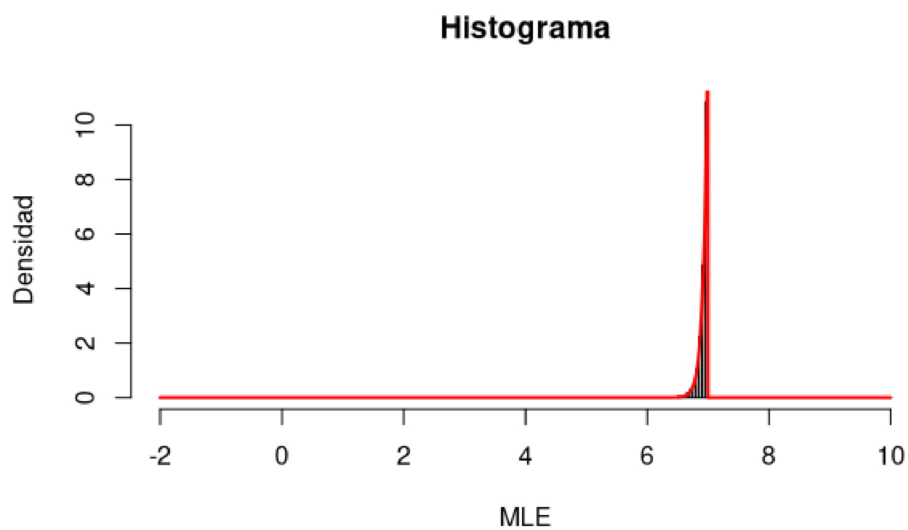


c Repita el procedimiento descrito en **b**, variando los tamaños de muestra por los valores $n = 50, 100, 200, 500, 1000$.

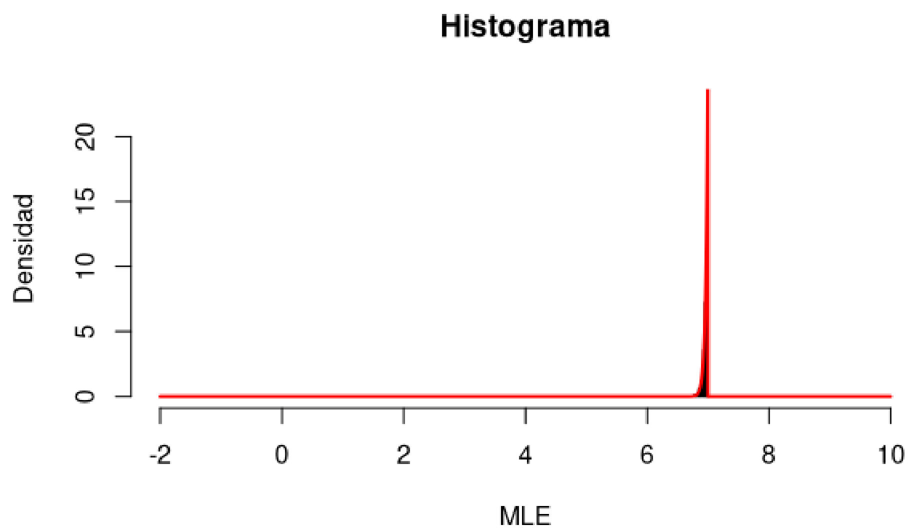
$n = 50$



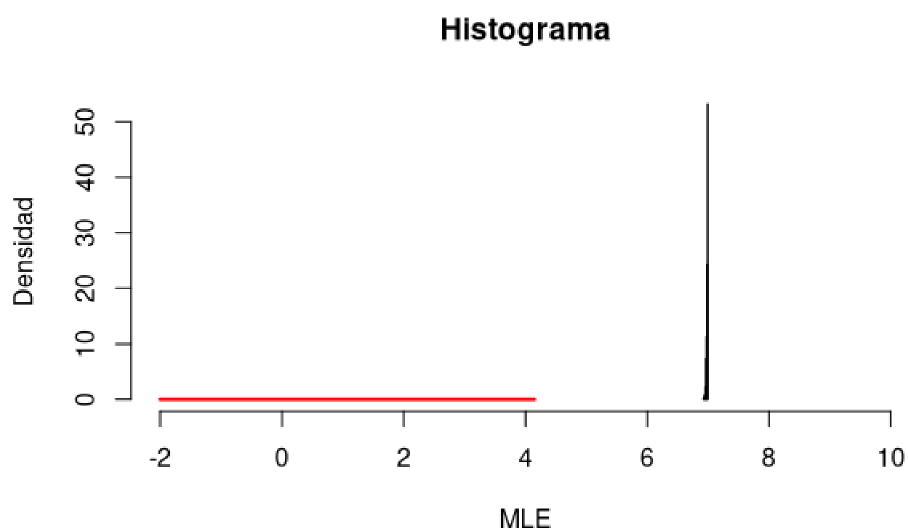
$n = 100$



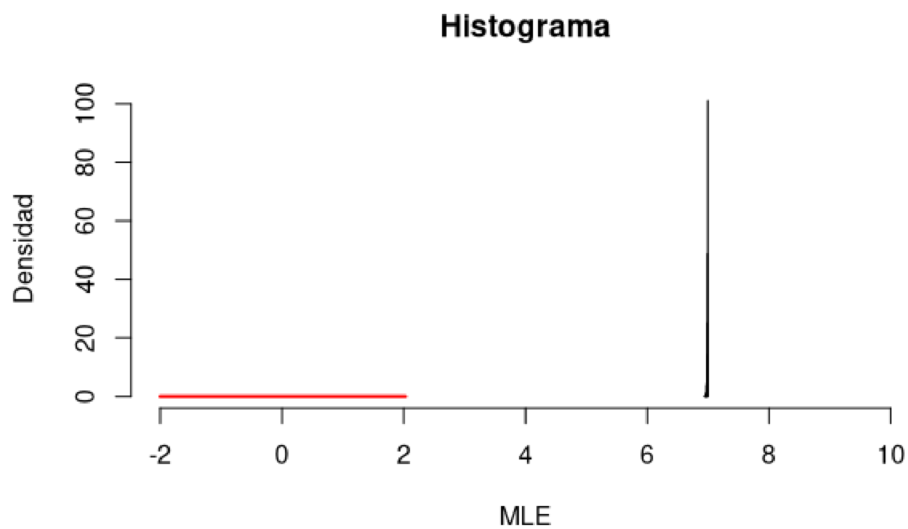
$n = 200$


Histograma
Densidad
0
5
10
15
20
-2
0
2
4
6
8
10
MLE

$n = 500$


Histograma
Densidad
0
10
20
30
40
50
-2
0
2
4
6
8
10
MLE

$n = 1000$



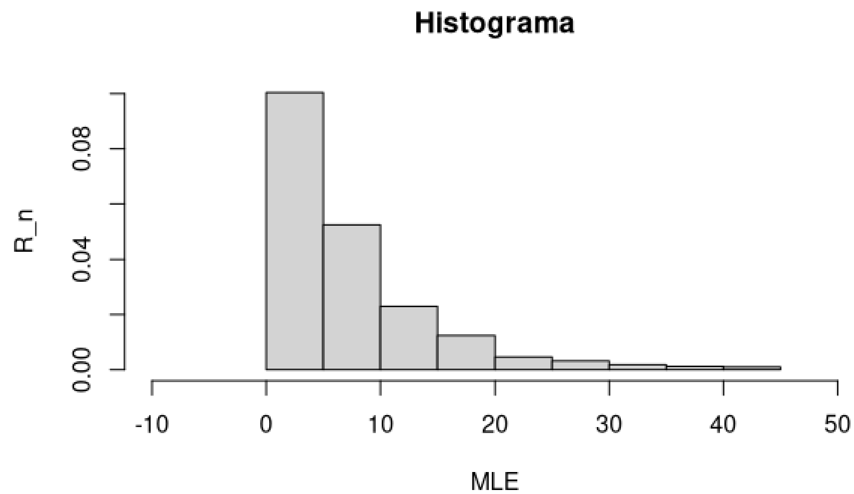
Ver Scrypt de R PuntoA.byc.R

d ¿Qué observa a medida que el tamaño de muestra aumenta? ¿En qué valor parecen concentrarse las realizaciones del estimador máximo-verosímil a medida que el tamaño de muestra aumenta? ¿Qué resultado de los vistos en clase explica ese fenómeno?

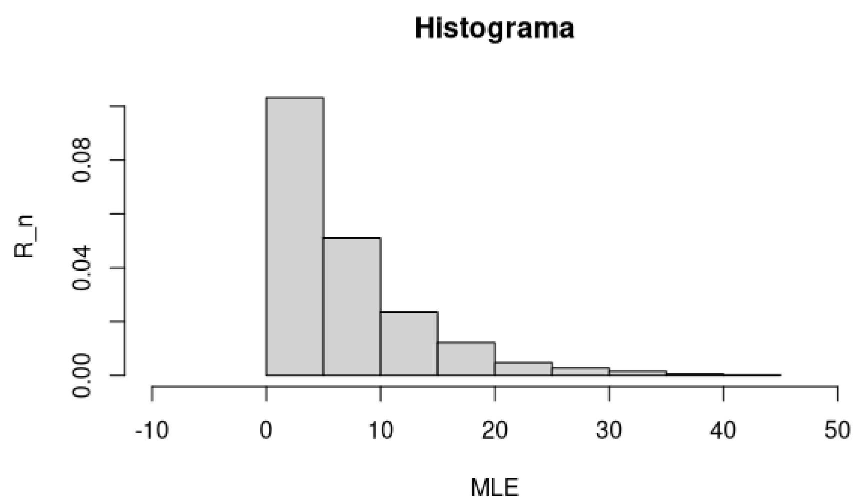
En medida de la que va aumentando el tamaño de la muestra (n) se ve como el estimador máximo-verosimil tiende a concentrarse y acercarse al valor real con el cuál obtuvimos nuestras realizaciones (en nuestro caso, 7). Por otro lado este acercamiento se puede explicar debido al gran comportamiento de lo que es el MLE y sus propiedades asintoticas respecto a la distribución uniforme, pues, se sabe que $A^{(n)}$ converge casi seguramente a θ.

e Genere nuevamente m muestras de tamaños $n = 10, 50, 100, 200, 500, 1000$; y calcule y almacene las m realizaciones de R_n para cada valor de n. Haga histogramas de los m valores obtenidos para cada n. ¿Observa que, a medida que n aumenta, dicho histograma se asemeja a alguna distribución conocida?

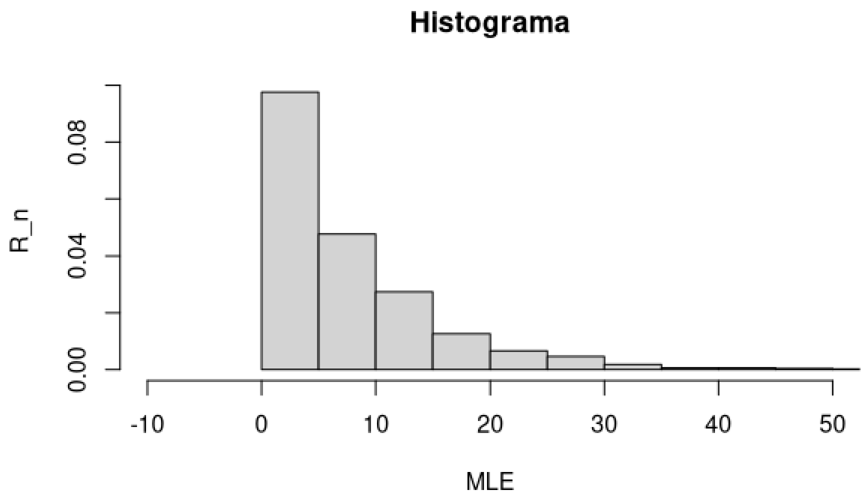
$$n = 50$$



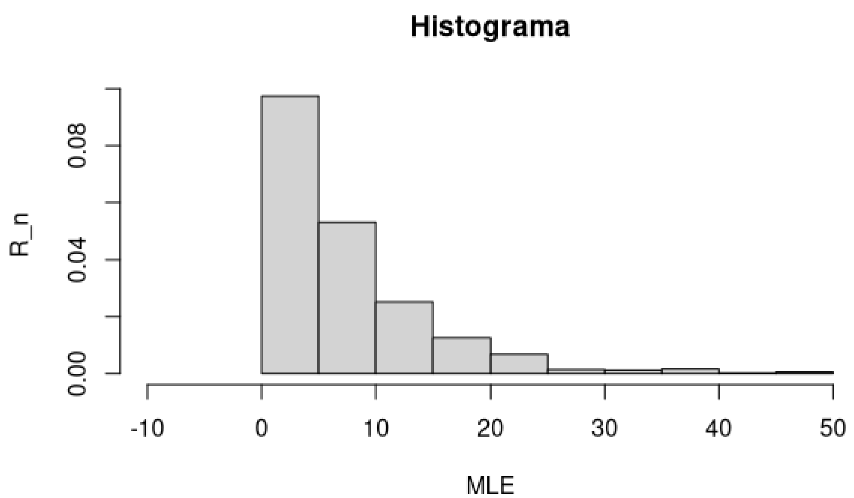
$$n = 100$$



$n = 200$

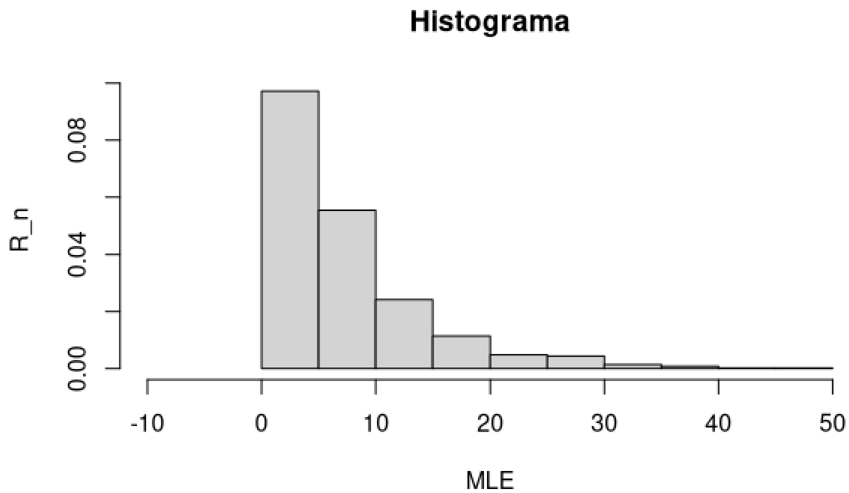


$n = 500$



Problema 1 continúa en la siguiente página. . .

$n = 1000$



Ver Scrypt de R PuntoA.e.R

f Pruebe formalmente su conjetura del apartado anterior, calculando el límite en distribución cuando $n \to \infty$:

$$R_n = n(\theta - \widehat{\theta}_{MLE}) \to^d ?$$

Problema 2:

Cálculo de una estimación máximo-verosímil cuando no hay una solución analítica.
Considere una muestra aleatoria $X_1, X_2, ..., X_n$ de una distribución $Logistica(\theta = 7, 1)$, es decir, donde el verdadero valor del parámetro de localización es igual al número del grupo. La función de densidad de este modelo es:

$$f_X(\theta, 1) = \frac{e^{-(x-\theta)}}{[1 + e^{-(x-\theta)}]^2}$$

Se sabe que, en este caso, el estimador ML no tiene una expresión analítica, pero es posible igual estudiar las propiedades del estimador máximo-verosímil haciendo uso de la simulación y de métodos numéricos, como la función optim de R.

Solución:

a. Justifique, de manera analítica, que los siguientes resultados son válidos para una variable aleatoria, X, con distribución $X \sim Logistica(\theta, 1)$.

I. $E[X] = \theta$.

Solución: Para hallar el valor esperado de X, primero debemos tener en cuenta que su función de densidad es la siguiente:

$$f_X(\theta, 1) = \frac{e^{-(x-\theta)}}{[1 + e^{-(x-\theta)}]^2}; \quad x, \theta \in \mathbb{R}$$

Esta función también puede quedar expresada de la siguiente forma:

$$f_X(\theta, 1) = \frac{1}{4}\text{sech}^2\left(\frac{x-\theta}{2}\right)$$

Antes de hallar el valor esperado, debemos demostrar la existencia del mismo. Por lo tanto, para demostrar la existencia, debemos resolver la siguiente operación:

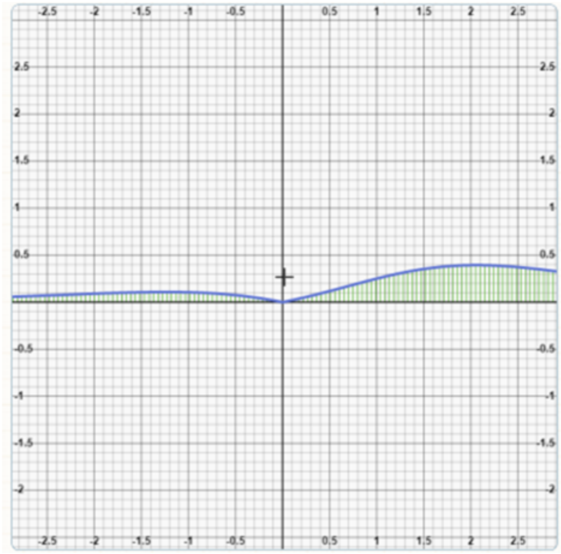
$$\int_{-\infty}^{\infty} |x| \cdot f_X(\theta, 1) \cdot dx$$

$$\int_{-\infty}^{\infty} \frac{|x|}{4} sech^2\left(\frac{x-\theta}{2}\right) \cdot dx$$

Ahora bien, gráficamente el área bajo la curva de la función no es divergente. Por lo tanto, es válido afirmar que:

$$\int_{-\infty}^{\infty} \frac{|x|}{4} sech^2\left(\frac{x-\theta}{2}\right) \cdot dx$$

Así, $E[X]$ existe.



Ahora bien, una vez demostrada la existencia, el valor esperado de X se halla de la siguiente manera:

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x \cdot f_X(\theta,1) \cdot dx \\
E[X] &= \int_{-\infty}^{\infty} \frac{x}{4} sech^2\left(\frac{x-\theta}{2}\right) \cdot dx \\
u &= \frac{x-\theta}{2} \\
du &= \frac{dx}{2} \\
E[X] &= \int_{-\infty}^{\infty} \left(\frac{2u+\theta}{4}\right) sech^2(u) \cdot 2du \\
E[X] &= \int_{-\infty}^{\infty} \left(\frac{2u+\theta}{2}\right) sech^2(u) \cdot du \\
E[X] &= \int_{-\infty}^{\infty} u \cdot sech^2 u \cdot du + \int_{-\infty}^{\infty} \frac{\theta}{2} \cdot sech^2 u \cdot du \\
a &= \int_{-\infty}^{\infty} u \cdot sech^2 u \cdot du \\
b &= \frac{\theta}{2} \cdot \int_{-\infty}^{\infty} sech^2(u) \\
a &= u \cdot tanh(u) - \int_{-\infty}^{\infty} tanh(u) \cdot du \\
a &= 0 \\
b &= \frac{\theta}{2}\left[tanh\left(\frac{x-\theta}{2}\right)\right]_{-\infty}^{\infty} \\
\lim_{x\to\infty} &= c \\
\lim_{x\to-\infty} &= -c \\
b &= \frac{\theta}{2}\left[tanh\left(\frac{x-\theta}{2}\right)\right]_{-c}^{c} \\
b &= \frac{\theta}{2}[1-(-1)] \\
b &= \frac{\theta}{2}[2] \\
b &= \theta \\
E[X] &= a+b \\
E[X] &= 0+\theta \\
E[X] &= \theta
\end{aligned}
$$

II. $Me(X) = \theta$.

Solución: Para hallar la mediana de X, primero vamos a dejar expresada la función de densidad de esta manera:

$$f_X(\theta, 1) = \frac{1}{4} sech^2 \left(\frac{x-\theta}{2} \right)$$

Ahora bien, la mediana de X se halla de la siguiente manera:

$$\frac{1}{2} = P(X < x_0) = \frac{1}{4} \cdot \int_{-\infty}^{x_0} sech^2 \left(\frac{x-\theta}{2} \right) dx$$

Tomando $u = \frac{x-\theta}{2}$ y $du = \frac{dx}{2}$

$$\begin{aligned}
\frac{1}{2} &= \frac{1}{4} \int_{-\infty}^{x_0} sech^2(u) \cdot 2du \\
&= \frac{1}{4} \int_{-\infty}^{x_0} sech^2(u) \cdot 2 \cdot du \\
&= \frac{1}{2} \int_{-\infty}^{x_0} sech^2(u) \cdot du \\
&= \frac{1}{2} \left[tanh(u) \right]_{-\infty}^{x_0} \\
&= \frac{1}{2} \left[tanh \left(\frac{x-\theta}{2} \right) \right]_{-\infty}^{x_0} \\
&= \frac{1}{2} \left[tanh \left(\frac{x_0-\theta}{2} \right) + 1 \right] \\
&= \frac{1}{2} tanh \left(\frac{x_0-\theta}{2} \right) + \frac{1}{2}
\end{aligned}$$

De esta manera,

$$\begin{aligned}
\frac{1}{2} \cdot tanh \left(\frac{x_0-\theta}{2} \right) &= 0 \\
tanh \left(\frac{x_0-\theta}{2} \right) &= 0 \\
\frac{x_0-\theta}{2} &= arctanh(0) \\
\frac{x_0-\theta}{2} &= 0 \\
x_0 - \theta &= 0 \\
x_0 &= \theta
\end{aligned}$$

Así, concluimos que $Me(\theta) = \theta$.

III. $Mo(X) = \theta$.

Solución: Para hallar la moda de X, primero debemos tener en cuenta que su función de densidad es la siguiente:

$$f_X(\theta,1) = \frac{e^{-(x-\theta)}}{\left[1+e^{-(x-\theta)}\right]^2}; \quad x,\theta \in \mathbb{R}$$

Ahora bien, para hallar la moda podemos encontrarla sacando la primera derivada de la función e igualarla a 0. En valor hallado en x indica en dónde se alcanza el máximo de la función de densidad. Por lo tanto, tenemos lo siguiente:

$$\begin{aligned}
\frac{d}{dx}f_X(x,\theta) &= \frac{-e^{(-x-\theta)}\cdot\left[1+e^{-(x-\theta)}\right]^2 + 2e^{-(x-\theta)}\cdot\left[1+e^{-(x-\theta)}\right]\cdot e^{-(x-\theta)}}{\left[1+e^{-(x-\theta)}\right]^4}\\
\frac{d}{dx}\cdot f_X(x,\theta) &= \frac{\left[1+e^{-(x-\theta)}\right]\cdot(-e^{(-x-\theta)}\cdot\left[1+e^{-(x-\theta)}\right] + 2e^{-(x-\theta)}\cdot e^{-(x-\theta)})}{\left[1+e^{-(x-\theta)}\right]^4}\\
\frac{d}{dx}\cdot f_X(x,\theta) &= \frac{-e^{(-x-\theta)}\cdot\left[1+e^{-(x-\theta)}\right] + 2e^{-(x-\theta)}\cdot e^{-(x-\theta)}}{\left[1+e^{-(x-\theta)}\right]^3}\\
0 &= \frac{-e^{(-x-\theta)}\cdot\left[1+e^{-(x-\theta)}\right] + 2e^{-(x-\theta)}\cdot e^{-(x-\theta)}}{\left[1+e^{-(x-\theta)}\right]^3}
\end{aligned}$$

Teniendo en cuenta que $[1+e^{-(x-\theta)}]^3 > 0$, obtenemos que:

$$\begin{aligned}
0 &= -e^{(-x-\theta)}\cdot\left[1+e^{-(x-\theta)}\right] + 2e^{-2(x-\theta)}\\
&= -e^{(-x-\theta)} - e^{-2(x-\theta)} + 2e^{-2(x-\theta)}\\
&= -e^{(-x-\theta)} + e^{-2(x-\theta)}\\
&= e^{-2(x-\theta)}\\
&= e^{(-x-\theta)}
\end{aligned}$$

Luego,

$$\begin{aligned}
ln\left(e^{-2(x-\theta)}\right) &= ln\left(e^{(-x-\theta)}\right)\\
-2x+2\theta &= -x+\theta\\
x &= \theta
\end{aligned}$$

Como θ es un punto crítico en x se alcanza un máximo/mínimo en $f_X(\theta,1)$. Sin embargo, teniendo en cuenta la gráfica de $f_X(\theta,1)$, entonces θ es la moda de la distribución logística.

b. ¿Qué estimadores pueden proponerse para estimar por analogía a θ?

Solución: Teniendo en cuenta el método de estimación por analogía, así como las demostraciones realizadas en el punto anterior, se pueden designar los siguientes estimadores para θ:

- ($\bar{X}_n$): teniendo en cuenta que $E[X] = \theta$. Asimismo, por la ley fuerte de los grandes números $\bar{X}_n$ converge casi seguramente a θ entonces $E[\bar{X}_n] = \theta$.

- ($X^{(k)}$): con $X^{(k)}$ ubicado en la posición del Percentil 50 (P_{50}) en la muestra, teniendo en cuenta que $Me(X) = \theta$. Asimismo, si $X^{(k)}$ se ubica en el Percentil 50 de la muestra, entonces de forma asintótica $X^{(k)}$ converge casi seguramente a la mediana poblacional, la cual corresponde a θ.

En caso que necesitemos recurrir a unos estimadores más robustos y teniendo en cuenta que en la distribución logística se demostró que $E[X]$, $Me(X)$ y $Mo(X)$ son iguales; podemos tener en cuenta como estimadores de θ a las siguientes estadísticas:

- Promedio de los Cuartiles:

$$\left(\bar{q} = \frac{q_1 + q_3}{2}\right)$$

- Trimedia:

$$\left(TRI = \frac{p_{50} + \bar{q}}{2}\right)$$

- Media Intercuartílica:

$$\left(MID = \frac{x_{i_{q1+1}} + ... + x_{i_{q3-1}}}{n_i}\right)$$

c. Aunque el estimador ML no pueda encontrarse por vía analítica, justifique que dicho estimador existe y es único para este modelo (dado un conjunto de datos fijo).

Solución: Dado que el estimador ML no se puede hallar de forma analítica, para justificar que el estimador ML existe y es único vamos a tener que recurrir a uno de los teoremas vistos en clase (Teorema de Unicidad del MLE de Bickel). Por lo tanto, primero debemos hallar la función de verosimilitud $f_X(\theta, 1)$, la cual se encuentra de la siguiente manera:

$$\begin{aligned}
L(\theta|X) &= \prod_{i=1}^{n} f_X(x_i) \\
&= \prod_{i=1}^{n} \frac{e^{-(x_i-\theta)}}{[1+e^{-(x_i-\theta)}]^2} \\
&= \frac{e^{-(x_1-\theta)}}{[1+e^{-(x_1-\theta)}]^2} \cdot \frac{e^{-(x_2-\theta)}}{[1+e^{-(x_2-\theta)}]^2} \cdot ... \cdot \frac{e^{-(x_n-\theta)}}{[1+e^{-(x_n-\theta)}]^2} \\
&= \frac{e^{(n\theta - \sum_{i=1}^{n} x_i)}}{\prod_{i=1}^{n}[1+e^{-(x_1-\theta)}]^2}
\end{aligned}$$

Una vez hallada la función de verosimilitud, procedemos a hallar la función de log-verosimilitud.

$$l(\theta|X) = n\theta - \sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} ln\left(1 + e^{(\theta - x_i)}\right)$$

Ahora bien, para corroborar las condiciones del teorema, debemos encontrar la segunda derivada de la función de log-verosimlitud, con el fin de determinar si es dos veces diferenciable y estrictamente cóncava. Para ello, notemos lo siguiente:

$$\begin{aligned}
\frac{d}{d\theta} l(\theta|X) &= n - 2\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\left[ln(1 + e^{\theta - x_i})\right] \\
&= n - 2\sum_{i=1}^{n}\left[\frac{e^{\theta - x_i}}{(1 + e^{\theta - x_i})}\right] \\
&= n - 2\sum_{i=1}^{n}\left[\frac{e^{\theta - x_i}}{(1 + e^{\theta - x_i})}\right]
\end{aligned}$$

Igualando a 0 obtenemos:

$$\begin{aligned}
n - 2\sum_{i=1}^{n}\left[\frac{e^{\theta - x_i}}{(1 + e^{\theta - x_i})}\right] &= 0 \\
\sum_{i=1}^{n}\left[\frac{e^{\theta - x_i}}{(1 + e^{\theta - x_i})}\right] &= \frac{n}{2}(*)
\end{aligned}$$

Ahora bien observamos que la función de la izquierda es creciente pues

$$\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\left[\frac{e^{\theta - x_i}}{(1 + e^{\theta - x_i})}\right] = \sum_{i=1}^{n}\left[\frac{e^{\theta - x_i}}{(1 + e^{\theta - x_i})^2}\right] > 0$$

Por tanto, $\sum_{i=1}^{n}\left[\frac{e^{\theta - x_i}}{(1+e^{\theta - x_i})}\right]$ es una función de θ estrictamente creciente. Finalmente observamos que:

$$\begin{aligned}
\lim_{\theta\to-\infty}\sum_{i=1}^{n}\left[\frac{e^{\theta - x_i}}{(1 + e^{\theta - x_i})}\right] &= \sum_{i=1}^{n}\lim_{\theta\to-\infty}\left[\frac{e^{\theta - x_i}}{(1 + e^{\theta - x_i})}\right] = \sum_{i=1}^{n}\lim_{\theta\to-\infty}\left[\frac{1}{(e^{-\theta + x_i}) + 1}\right] = 0 \\
\lim_{\theta\to\infty}\sum_{i=1}^{n}\left[\frac{e^{\theta - x_i}}{(1 + e^{\theta - x_i})}\right] &= \sum_{i=1}^{n}\lim_{\theta\to\infty}\left[\frac{1}{(e^{-\theta + x_i}) + 1}\right] = \sum_{i=1}^{n} 1 = n
\end{aligned}$$

Así, la ecuación (*) tiene una única solución. Por otra parte la función $l(\theta|x) < 0$ paratodo θ y por tanto resulta ser estrictamente cóncava.

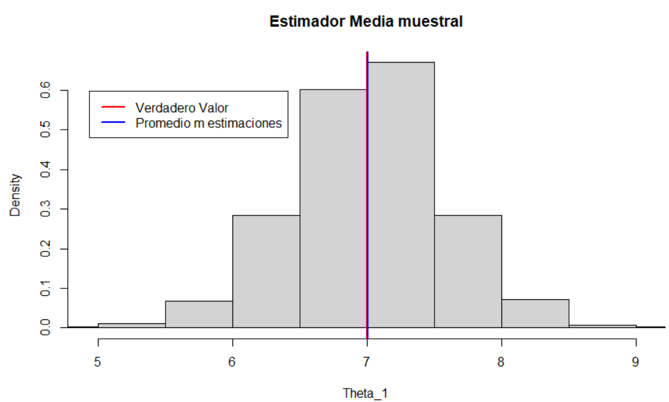
Así, de acuerdo a todo lo anterior, se cumplen las condiciones que demuestran el teorema de unicidad del MLE de Bickel. Por lo tanto, se concluye que el estimador $\hat{\theta}_{MLE}$ existe y es único; así mismo se puede encontrar resolviendo:

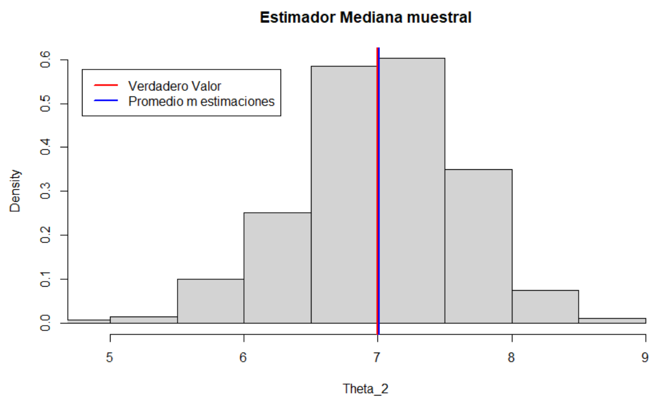
$$\triangledown l(\hat{\theta}_{MLE}|X) = 0$$

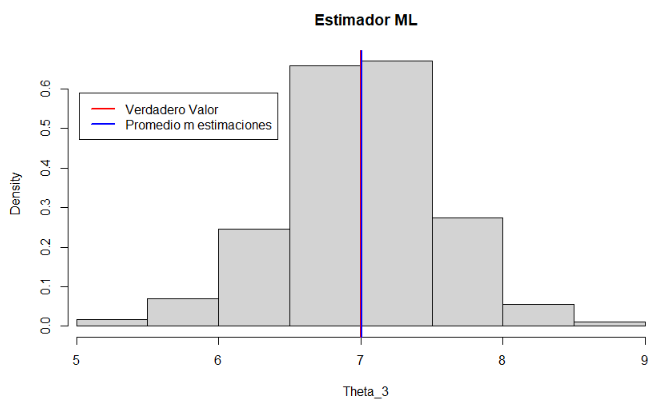
d. Genere $m = 1000$ simulaciones de una muestra aleatoria de tamaño $n = 10$ de la distribución $Logistica(\theta = 7, 1)$, usando la función rlogis de R o su equivalente en otro lenguaje de programación. Para cada una de las m muestras, calcule y almacene la estimación máximo-verosímil del parámetro, la media muestral y la mediana (como dos estimaciones por analogía). Haga un histograma de las m estimaciones ML, otro de las m estimaciones por analogía con la media muestral y otro de las m estimaciones por analogía con la mediana muestral. Añada una línea vertical en cada histograma para indicar el verdadero valor del parámetro y otra línea (de diferente color) para indicar el promedio de las m estimaciones en cada caso. Reporte en una tabla el promedio y la varianza obtenidos de las m estimaciones para este tamaño de muestra. ¿Hay indicios de que los tres estimadores sean insesgados? Establezca qué estimador es mejor en términos de los valores de error cuadrático medio (estimado) y añada el MSE estimado a la tabla para este tamaño de muestra.

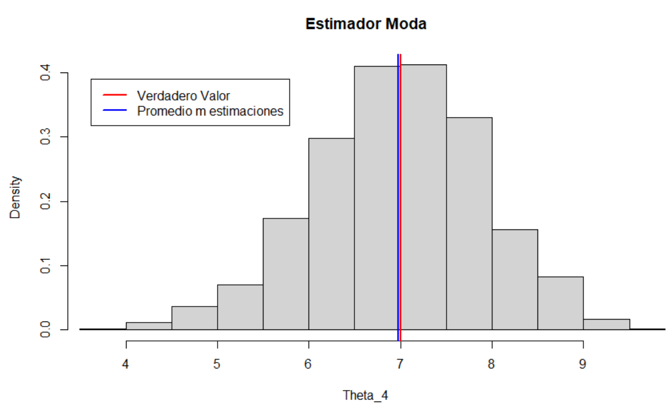
Solución: A continuación se muestran los histogramas de las $m = 1000$ muestras de tamaño $n = 10$. Para ello, se tomó como estimador $\hat{\theta_1}$ a la media muestral; como estimador $\hat{\theta_2}$ a la mediana muestral y finalmente $\hat{\theta}_{MLE}$ será el estimador máximo verosimil hallado en las m muestras.

Bonus: Asimismo se propondrá a $\hat{\theta}_4$ como un estimador que estará en términos de la moda. Para ello se utilizará la función MLV del paquete MODEEST de R. Teniendo en cuenta que la moda no puede ser estudiada de forma analítica como variable aleatoria y tampoco nos podemos remitir como el dato que "mas se repite", ya que la distribución logística es continua.









Tal y como se puede observar, $\hat{\theta}_1$, $\hat{\theta}_2$ y $\hat{\theta}_4$ se acercan al verdadero valor del parámetro θ (casi siendo el mismo), por lo que resultarían insesgados. En la teoría, estos parámetros son insesgados, teniendo en cuenta que la media, la mediana y la moda poblacional de una distribución logística son iguales. Respecto al estimador $\hat{\theta}_{MLE}$, aunque se acerque al valor del parámetro, el estimador tiene sesgo; aunque éste $\hat{\theta}_{MLE}$ es poco.

A continuación, en la siguiente tabla vamos a observar la media, la varianza y el error cuadrático medio (MSE) de los cuatro estimadores con $n = 10$.

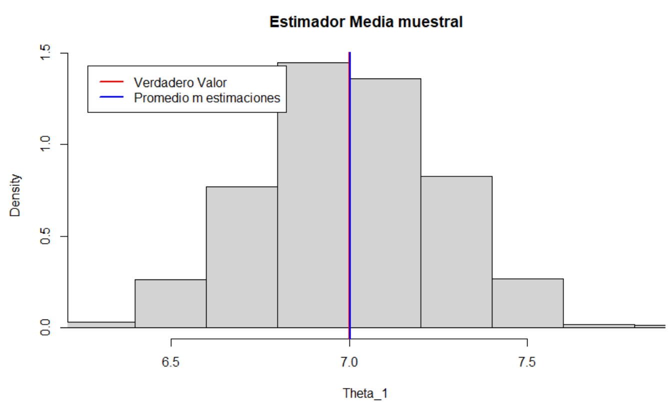
$n = 10$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta_1}$	7.008830	0.325760	0.325838
$\hat{\theta_2}$	7.006740	0.380914	0.380959
$\hat{\theta}_{MLE}$	7.004702	0.307750	0.307773
$\hat{\theta_4}$	6.976766	0.872241	0.872781

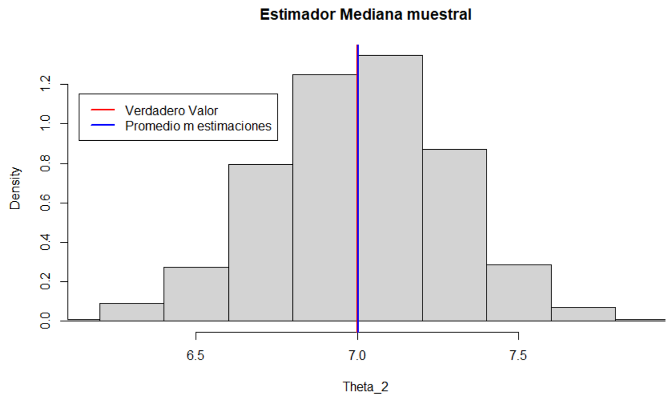
De acuerdo a los resultados obtenidos, se puede observar que, a pesar de que la media muestral ($\hat{\theta_1}$), la mediana muestral ($\hat{\theta_2}$) y la moda ($\hat{\theta_4}$) resultan ser estimadores insesgados. El mejor estimador, en términos de MSE, es el estimador máximo-verosimil ($\hat{\theta}_{MLE}$), debido a que éste es el estimador con menor error cuadrático medio, a pesar de que no se pueda hallar de manera analítica.

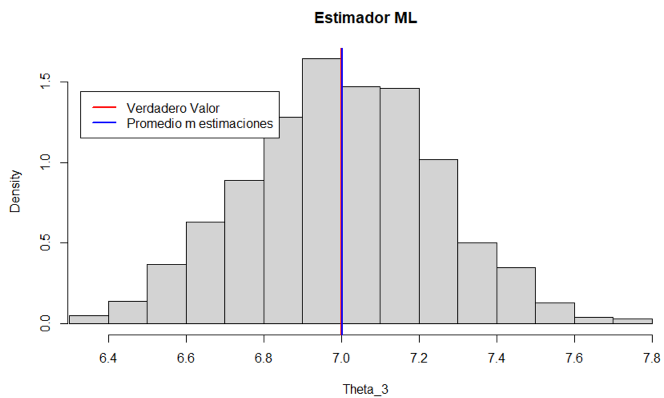
e. Repita el procedimiento descrito en d., variando los tamaños de muestra por los valores $n = 50, 100, 200, 500, 1000$. ¿Se mantiene la misma conclusión acerca de qué estimador es mejor en términos de error cuadrático medio (estimado)?

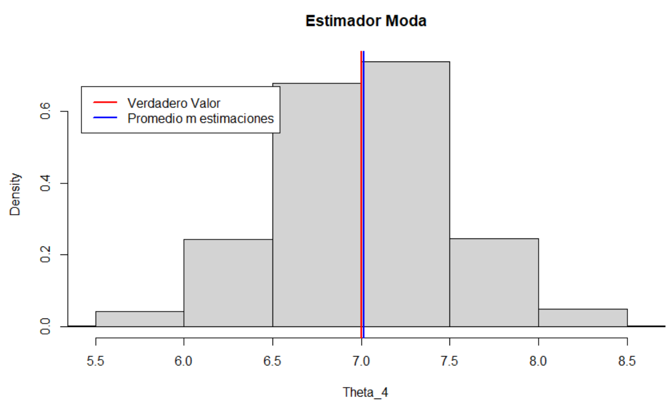
Solución: Para este caso, se va a variar el tamaño de muestra, con el fin de determinar si hay algún cambio respecto al mejor estimador para la distribución logística.

- Para un tamaño de muestra $n = 50$ se tiene la siguiente información:





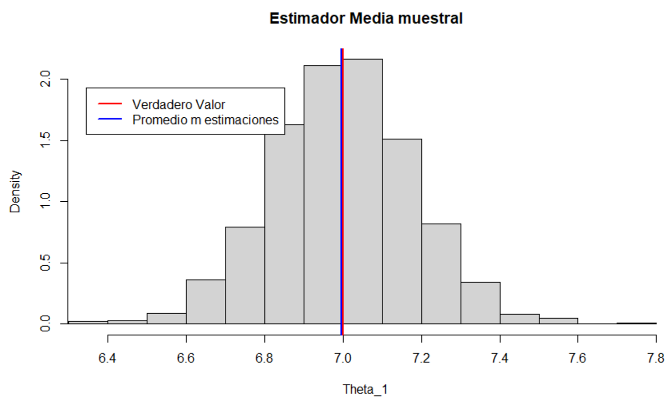


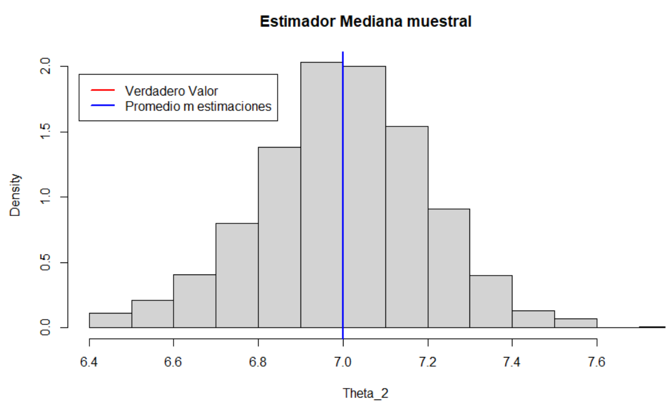


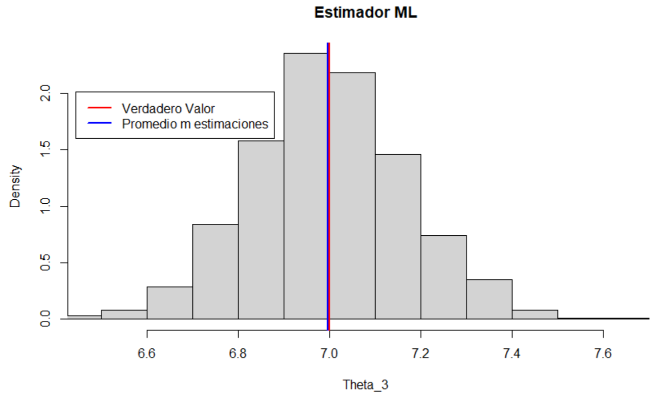
Como conclusión se obtiene que a pesar de que $\hat{\theta}_1$, $\hat{\theta}_2$ y $\hat{\theta}_4$ disminuyeron sus varianzas. Esto no les alcanza para ser los mejores estimadores, debido a que $\hat{\theta}_{MLE}$ se mantiene como el estimador con el menor error cuadrático medio, por ello, el estimador máximo verosimil sigue siendo el mejor estimador.

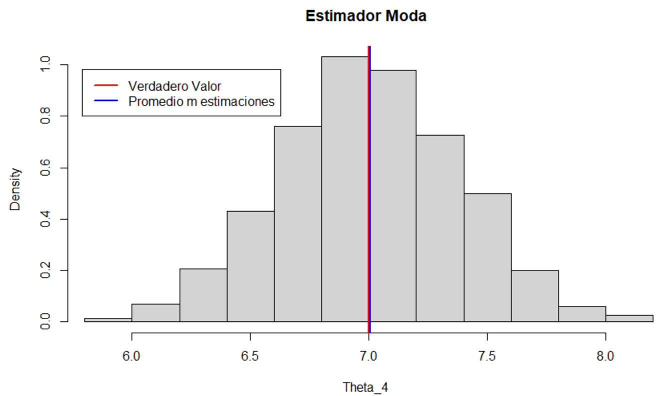
- Para un tamaño de muestra $n = 100$ se obtiene la siguiente información:

$n=50$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta}_1$	7.003596	0.064309	0.064322
$\hat{\theta}_2$	7.003701	0.078773	0.078787
$\hat{\theta}_{MLE}$	7.001648	0.060213	0.060215
$\hat{\theta}_4$	7.016578	0.238076	0.238351





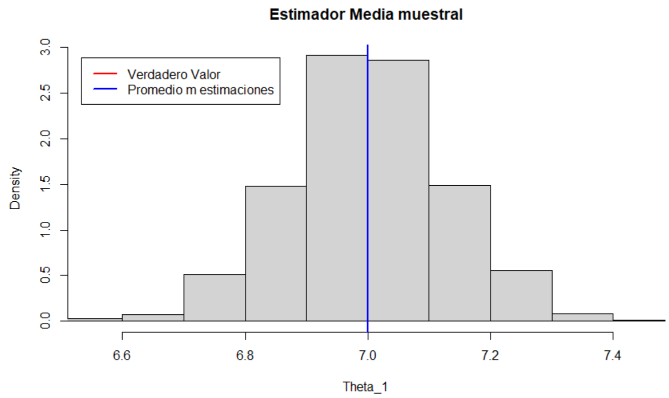


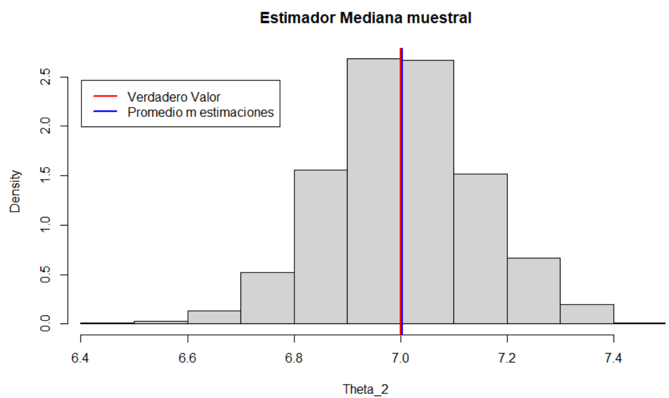


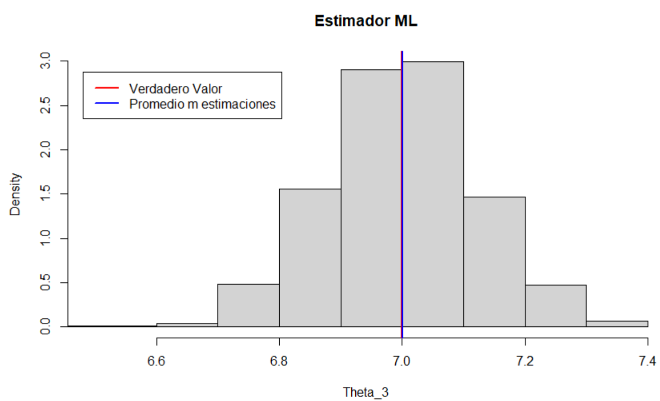
$n = 100$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta}_1$	6.996830	0.032553	0.032563
$\hat{\theta}_2$	7.000849	0.040235	0.040235
$\hat{\theta}_{MLE}$	6.996838	0.029503	0.029513
$\hat{\theta}_4$	7.005754	0.147019	0.147052

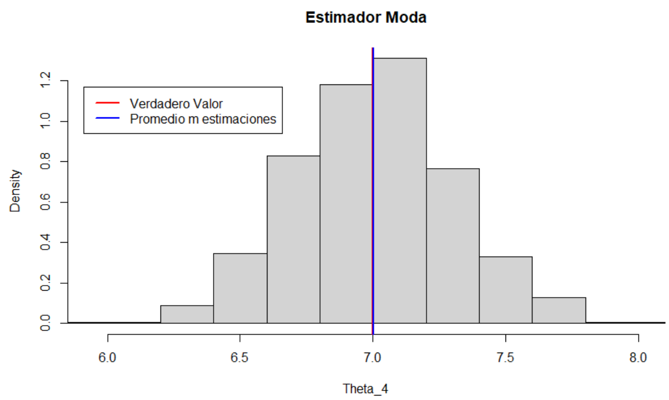
Como conclusión se obtiene que a pesar de que $\hat{\theta}_1$, $\hat{\theta}_2$ y $\hat{\theta}_4$ disminuyeron sus varianzas; esto no les alcanza para ser los mejores estimadores, debido a que $\hat{\theta}_{MLE}$ se mantiene como el estimador con el menor error cuadrático medio, por ello, el estimador máximo verosimil sigue siendo el mejor estimador.

- Para un tamaño de muestra $n = 200$ se obtiene la siguiente información:


Estimador Media muestral
Verdadero Valor
Promedio m estimaciones
Density
Theta_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
6.6
6.8
7.0
7.2
7.4


Estimador Mediana muestral
Verdadero Valor
Promedio m estimaciones
Density
Theta_2
0.0
0.5
1.0
1.5
2.0
2.5
6.4
6.6
6.8
7.0
7.2
7.4

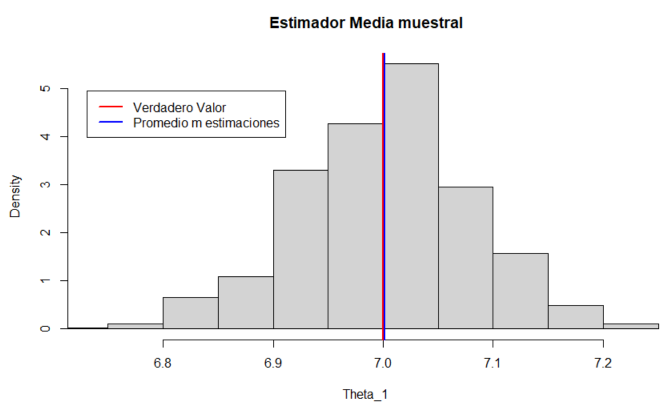

Estimador ML
Verdadero Valor
Promedio m estimaciones
Density
Theta_3
0.0
0.5
1.0
1.5
2.0
2.5
3.0
6.6
6.8
7.0
7.2
7.4

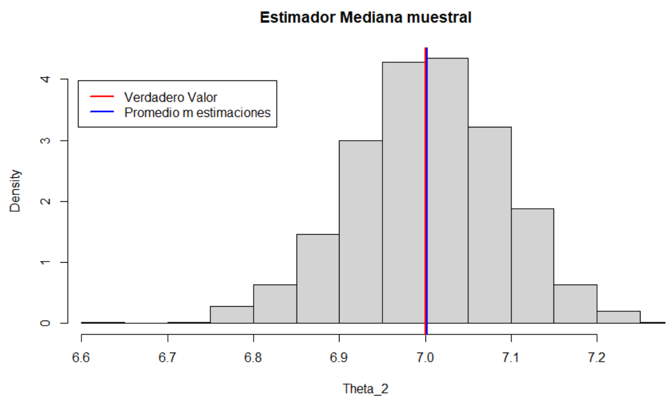


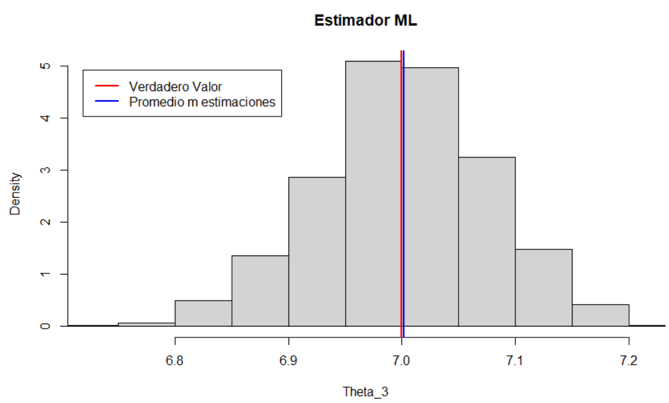
$n = 200$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta_1}$	7.000656	0.016440	0.016440
$\hat{\theta_2}$	7.003803	0.020115	0.020129
$\hat{\theta}_{MLE}$	7.001231	0.015225	0.015226
$\hat{\theta_4}$	7.001807	0.090814	0.090817

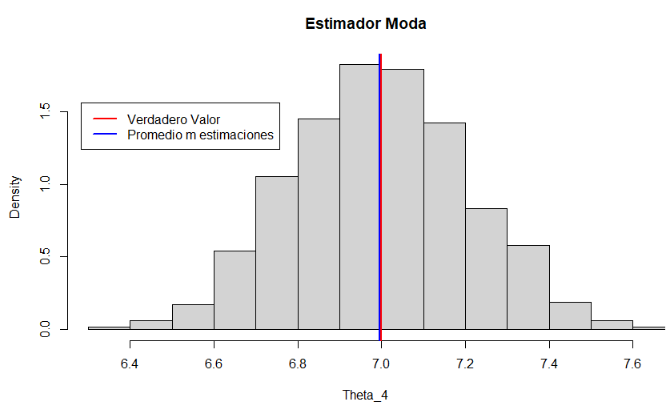
Como conclusión se obtiene que a pesar de que $\hat{\theta}_1$, $\hat{\theta}_2$ y $\hat{\theta}_4$ disminuyeron sus varianzas; esto no les alcanza para ser los mejores estimadores, debido a que $\hat{\theta}_{MLE}$ se mantiene como el estimador con el menor error cuadrático medio, por ello, el estimador máximo verosimil sigue siendo el mejor estimador. Sin embargo, la diferencia en MSE entre los estimadores cada vez se hace mas corta.

- Para un tamaño de muestra $n = 500$ se obtiene la siguiente información:





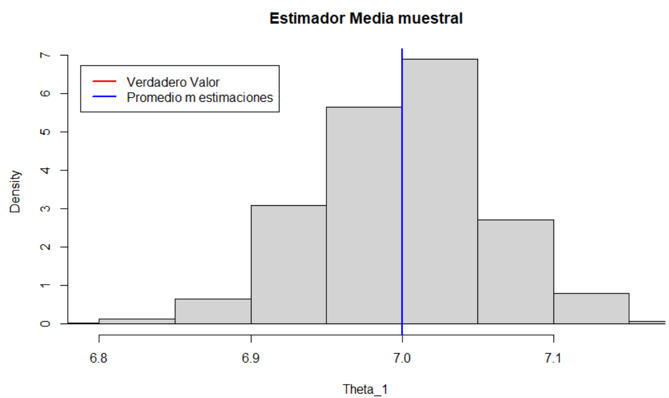


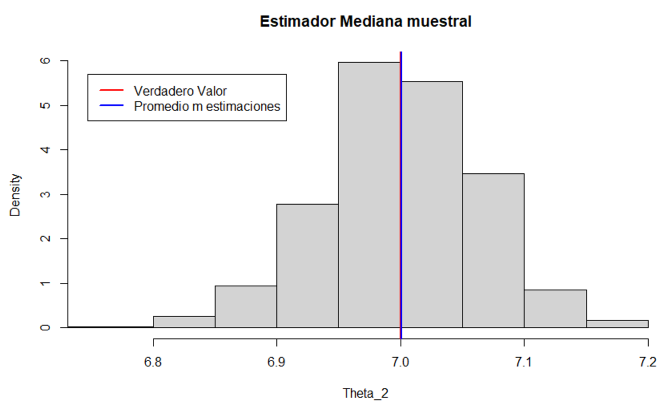


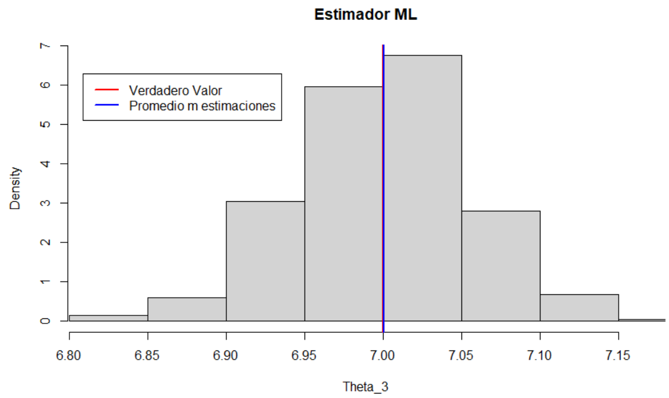
En este caso, los estimadores $\hat{\theta_1}$ y $\hat{\theta_2}$ acortan distancia en términos de MSE a $\hat{\theta}_{MLE}$. Mientras que $\hat{\theta}_4$ no resulta siendo un buen estimador, a pesar de ser insesgado. Por lo tanto, $\hat{\theta}_{MLE}$ se mantiene como el mejor estimador para este tamaño de muestra.

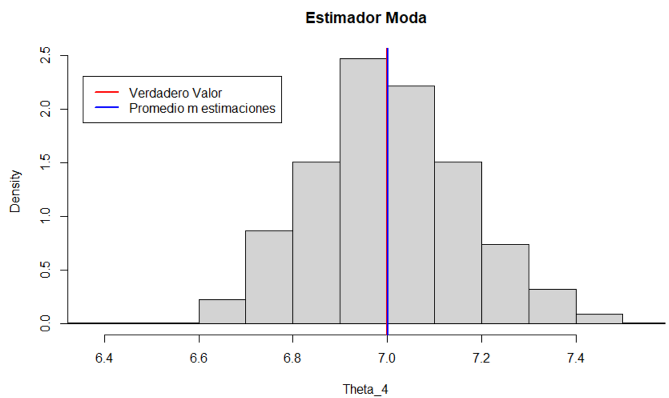
- Para un tamaño de muestra $n = 1000$ se obtiene la siguiente información:

$n = 500$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta_1}$	7.001410	0.006206	0.006208
$\hat{\theta_2}$	7.002744	0.007973	0.007980
$\hat{\theta}_{MLE}$	7.001506	0.005763	0.005766
$\hat{\theta}_4$	6.996429	0.045546	0.045558









$n = 1000$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta}_1$	7.000195	0.003358	0.003359
$\hat{\theta}_2$	7.000563	0.004113	0.004113
$\hat{\theta}_{MLE}$	7.000587	0.003083	0.003083
$\hat{\theta}_4$	7.001018	0.027033	0.027034

Como conclusión, se mantiene la misma tendencia que con un tamaño de muestra $n = 500$. Asimismo, el mejor estimador sigue siendo $\hat{\theta}_{MLE}$, dado que éste tiene menor MSE que los demás estimadores. Sin embargo, también podemos recurrir a $\hat{\theta}_1$ como un buen estimador, debido a que la diferencia con el estimador máximo verosimil es muy corta.

f. ¿A qué convergen en distribución las siguientes secuencias?

- $V_n = \sqrt{n}\left(\hat{\theta}_{MLE} - \theta\right) \xrightarrow{d} ?$

Solución: teniendo en cuenta que $\hat{\theta}_{MLE}$ se encuentra bajo condiciones de regularidad. Teniendo que cuenta que en la teoría el estimador se encuentra bajo condiciones de regularidad. Entonces, por el Teorema de la Distribución Asintótica del MLE, se cumple que:

$$V_n = \sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} N(0, I^{-1})$$

Ahora bien, para encontrar el valor de la varianza, debemos primero encontrar la información de Fisher para la distribución logística. Con este dato se sabrá hacia donde converge en distribución V_n. Para ello, se utilizará la siguiente fórmula para hallar la información de Fisher:

$$I(\theta) = E_\theta\left[-\frac{\partial^2}{\partial\theta^2} ln(f_X(X, \theta))\right]$$

Luego, la información de Fisher se halla de la siguiente forma:

$$\begin{gathered}
f_X(x,\theta) = \frac{e^{-(x-\theta)}}{[1+e^{-(x-\theta)}]^2} \\
ln(f_X(x,\theta)) = -(x-\theta) - 2ln(1+e^{-(x-\theta)}) \\
\frac{\partial}{\partial\theta} ln(f_X(x,\theta)) = 1 - \frac{2e^{(\theta-x)}}{1+e^{(\theta-x)}} \\
\frac{\partial}{\partial\theta} ln(f_X(x,\theta)) = \frac{1+e^{(\theta-x)}-2e^{(\theta-x)}}{1+e^{(\theta-x)}} \\
\frac{\partial}{\partial\theta} ln(f_X(x,\theta)) = \frac{1-e^{(\theta-x)}}{1+e^{(\theta-x)}} \\
\frac{\partial^2}{\partial\theta^2} ln(f_X(x,\theta)) = \frac{-e^{(\theta-x)}[1+e^{(\theta-x)}]-e^{(\theta-x)}[1-e^{(\theta-x)}]}{[1+e^{(\theta-x)}]^2} \\
\frac{\partial^2}{\partial\theta^2} ln(f_X(x,\theta)) = \frac{-e^{(\theta-x)}}{[1+e^{(\theta-x)}]^2} \\
-\frac{\partial^2}{\partial\theta^2} ln(f_X(x,\theta)) = \frac{e^{(\theta-x)}}{[1+e^{(\theta-x)}]^2} \\
-\frac{\partial^2}{\partial\theta^2} ln(f_X(x,\theta)) = f_X(x,\theta)
\end{gathered}$$

Como se puede observar, al realizar el negativo de la segunda derivada del logaritmo de la función de densidad, nos da como resultado la misma función de densidad. Por lo tanto, tenemos lo siguiente:

$$\begin{gathered}
I(\theta) = E[f_X(x,\theta)] \\
I(\theta) = \int_{-\infty}^{\infty} f_X \cdot f_X \cdot dx \\
I(\theta) = \int_{-\infty}^{\infty} f_X^2 \cdot dx \\
I(\theta) = \int_{-\infty}^{\infty} \frac{e^{2(\theta-x)}}{[1+e^{(\theta-x)}]^4} \cdot dx \\
I(\theta) = \int_{-\infty}^{\infty} \frac{1}{16} sech^4(\tfrac{x-\theta}{2}) \cdot dx \\
I(\theta) = [-\frac{tanh(\frac{x-\theta}{2})(tanh(\frac{x-\theta}{2})-3)}{24}]_{-\infty}^{\infty}
\end{gathered}$$

Evaluando la integral en $-\infty$ e ∞ obtenemos que $I(\theta) = \frac{1}{6}$. Por lo tanto $I^{-1} = 6$.

Finalmente, se concluye que V_n converge en distribución a una Normal con $\mu = 0$ y $\sigma^2 = 6$.

- $W_n = \sqrt{n}(Me - \theta) \xrightarrow{d}$?

Solución: Teniendo en cuenta que la Mediana se comporta como una estadística de orden, ya que ésta es igual al percentil 50; asimismo, la distribución de una variable aleatoria da como resultado que $P(X < Me) = \frac{1}{2}$. Finalmente, teniendo en cuenta que la distribución logística es continua alrededor de la mediana, se tiene que:

$$\sqrt{n}(X^{(k)} - \xi_p) \underset{\rightarrow}{d} N(0, \frac{P(1-P)}{f_X^2(\xi_p)})$$

Ahora bien, reemplazando por la secuencia inicial tenemos que:

$$X^{(k)} = Me$$
$$\xi_p = \theta$$
$$P = \frac{1}{2}$$
$$f_X^2(\xi_p) = f_X^2(\theta)$$

Ahora bien, al evaluar $f_X^2(\theta)$ en la distribución logística, tenemos lo siguiente:

$$f_X^2(\theta) = (\frac{e^{-(\theta-\theta)}}{[1+e^{-(\theta-\theta)}]^2})^2$$
$$f_X^2(\theta) = (\frac{e^0}{[1+e^0]^2})^2$$
$$f_X^2(\theta) = (\frac{1}{[1+1]^2})^2$$
$$f_X^2(\theta) = (\frac{1}{[2]^2})^2$$
$$f_X^2(\theta) = (\frac{1}{4})^2$$
$$f_X^2(\theta) = \frac{1}{16}$$

Luego, una vez hallado $f_X^2(\theta)$ tenemos lo siguiente:

$$\sqrt{n}(Me - \theta) \underset{\rightarrow}{d} N(0, \frac{\frac{1}{2}(1-\frac{1}{2})}{\frac{1}{16}})$$
$$\sqrt{n}(Me - \theta) \underset{\rightarrow}{d} N(0, \frac{\frac{1}{4}}{\frac{1}{16}})$$
$$\sqrt{n}(Me - \theta) \underset{\rightarrow}{d} N(0, 4)$$

Por lo tanto, se concluye que W_n converge en distribución a una Normal con $\mu = 0$ y $\sigma^2 = 4$.

- $U_n = \sqrt{n}(\bar{X}_n - \theta) \underset{\rightarrow}{d}$?

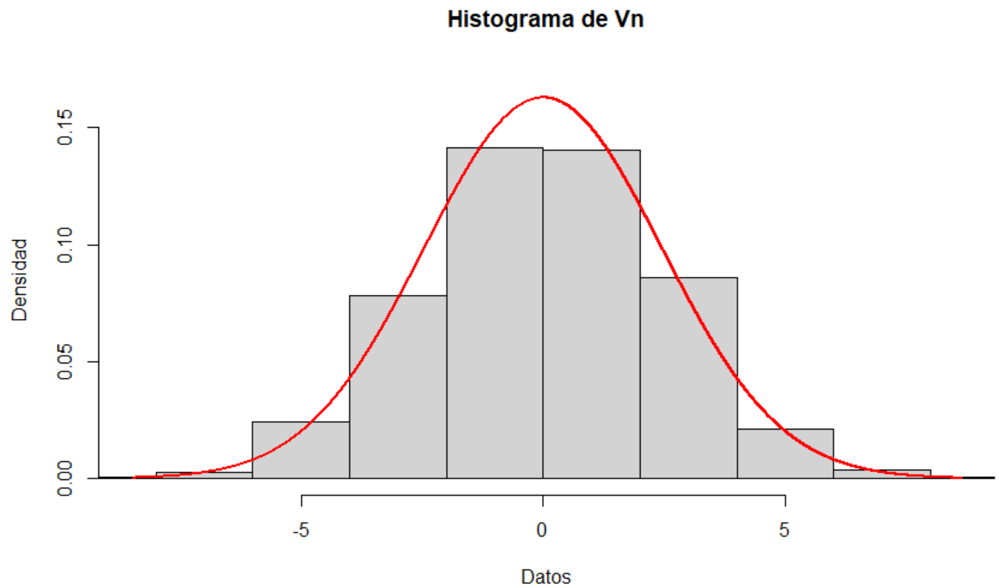
Solución: Finalmente, tenemos como última secuencia a $\bar{X}_n$. Esta convergencia es más sencilla, debido a que solo tenemos que recurrir al Teorema Central del Límite para determinar que:

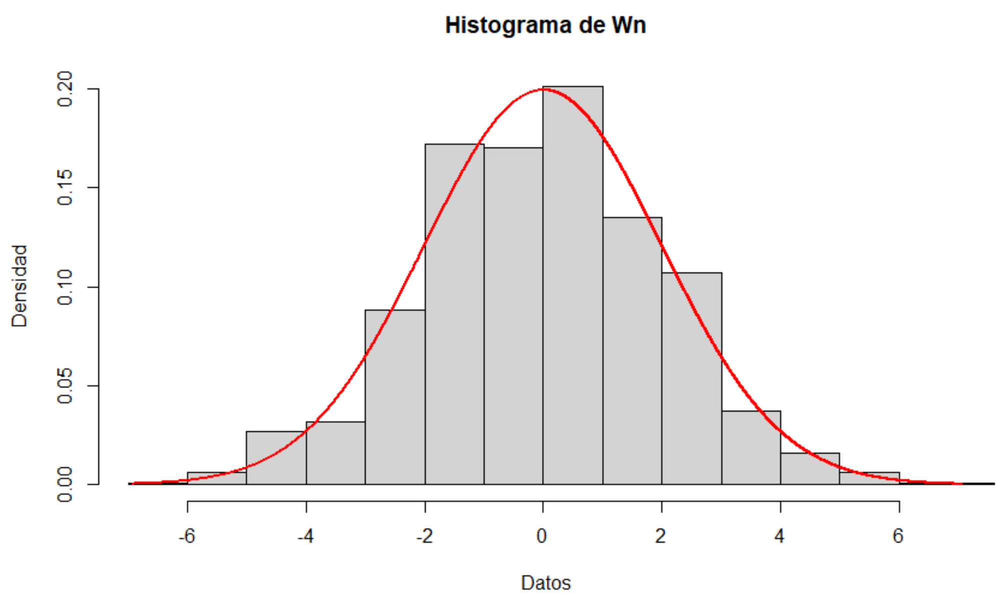
$$\sqrt{n}(\bar{X}_n - \theta) \underset{\rightarrow}{d} N(0, 1)$$

Por lo tanto se concluye que U_n converge en distribución a una Normal estándar.

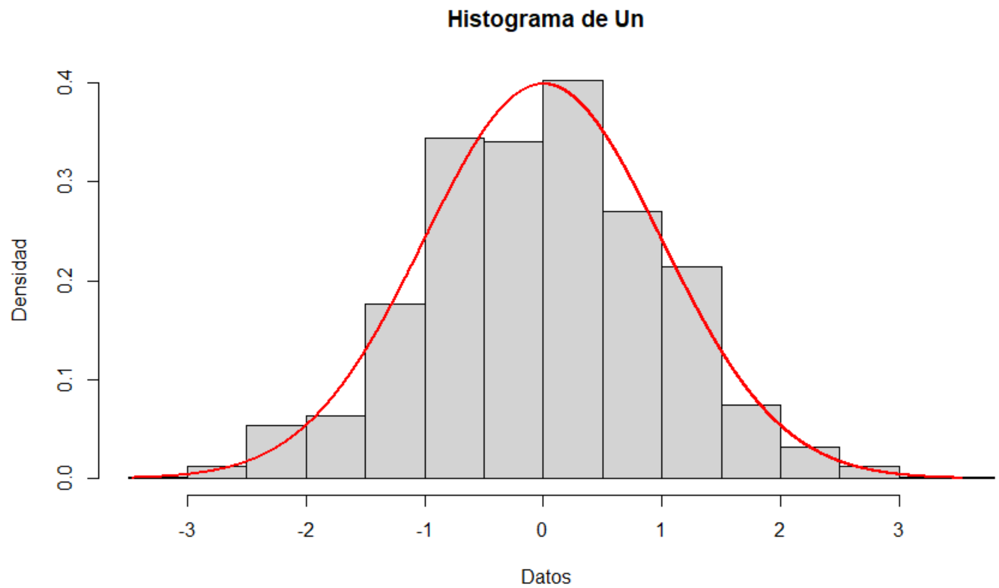
g. Repitiendo el procedimiento de generar m muestras de tamaño $n = 1000$, calcule m realizaciones de V_{1000}, de W_{1000} y de U_{1000}, realice un histograma de cada una de ellas y superponga las distribuciones que encontró en el punto anterior. ¿Se ve que el resultado de la convergencia es correcto? ¿Cuál de los tres estimadores es asintóticamente más eficiente?

Solución: a continuación se muestran los histogramas de cada una de las realizaciones de V_n, W_n y U_n.





Problema 2 continúa en la siguiente página. . .



Tal y como se puede observar al graficar las secuencias se concluye que el resultado de la convergencia es correcto en las tres secuencias.

Ahora bien, para determinar cuál de los tres estimadores es asintóticamente más eficiente, debemos tener en cuenta el valor de sus varianzas al momento de realizar la convergencia en distribución. Estas varianzas son las siguientes:

$$\sigma^2_{\hat{\theta}_{MLE}} = 6$$
$$\sigma^2_{Me} = 4$$
$$\sigma^2_{\bar{X}_n} = 1$$

Luego, al tener las varianzas, vamos a compararlas para determinar cuál de los tres estimadores es más eficiente asintóticamente. Para ello, vamos a recurrir a la fórmula del ARE.

Primero vamos a comparar $\hat{\theta}_{MLE}$ con Me para determinar cuál de los dos estimadores es mas eficiente:

$$ARE(\hat{\theta}_{MLE}, Me) = \frac{6}{4}$$
$$ARE(\hat{\theta}_{MLE}, Me) = 1{,}5$$

Como se puede observar el $ARE(\hat{\theta}_{MLE}, Me) > 1$. Por lo tanto, el estimador Me es más eficiente asintóticamente que $\hat{\theta}_{MLE}$.

Luego, vamos a comparar Me con $\bar{X}_n$ para determinar cuál de los dos estimadores es mas eficiente:

$$ARE(Me, \bar{X}_n) = \tfrac{2}{1}$$
$$ARE(Me, \bar{X}_n) = 2$$

Finalmente, como el $ARE(Me, \bar{X}_n) > 1$ se concluye que el estimador asintóticamente más eficiente es $\bar{X}_n$.

h. **Bonus [+0.2]**. Incluya la Moda como estimador por analogía y compárelo con los otros tres estimadores en los apartados d. y e. La moda no puede ser estudiada como variable aleatoria de manera analítica ya que no tiene una dependencia funcional clara de la muestra aleatoria. Además, como las variables son continuas, no se puede calcular la moda como "el dato que más se repite", y es necesario usar la fórmula de la moda para datos agrupados. Explore la función mlv del paquete modeest en R y explore los métodos que allí se proponen (use el método 'naive' como uno de ellos).

Solución: Para este bonus, remitirse a los puntos **d** y **e**.

Problema 3:

Solución:

a. Construya el primer estimador por método de los momentos, $\hat{\theta}_{MOM,1}$, usando el primer momento informativo de la distribución exponencial.

   Solución. Para esto, tengamos en cuenta que θ, en este caso, representa a $1/\lambda$. El primer momento de la distribución exponencial es θ, por otra parte, el primer momento de la distribución muestral es $\bar{X}$.

   Al igualarlos obtenemos que:

$$\bar{X} = \hat{\theta}_{MOM,1}$$

b. Construya $\hat{\theta}_{MOM,2}$ y $\hat{\theta}_{MOM,3}$ usando los momentos informativos de orden 2 y 3 respectivamente.

   Solución. Para construir estos dos estimadores, tengamos en cuenta lo siguiente:

$$E[X^2] = 2 \cdot \theta^2 \qquad \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$E[X^3] = 6 \cdot \theta^3 \qquad \hat{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n} X_i^3$$

   Ahora, para hallar $\hat{\theta}_{MOM,2}$ y $\hat{\theta}_{MOM,3}$ respectivamente, igualamos $E[X^2]$ con $\hat{\mu}_2$ y $E[X^3]$ con $\hat{\mu}_3$, luego despejamos θ en ambas ecuaciones, y obtendremos así a $\hat{\theta}_{MOM,2}$ y $\hat{\theta}_{MOM,3}$.

Para $\hat{\theta}_{MOM,2}$ tenemos que:

$$2 \cdot \hat{\theta}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$
$$\hat{\theta}^2 = \frac{1}{2n}\sum_{i=1}^{n} X_i^2$$
$$\hat{\theta}_{MOM,2} = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} X_i^2}$$

Y para $\hat{\theta}_{MOM,3}$ tenemos que:

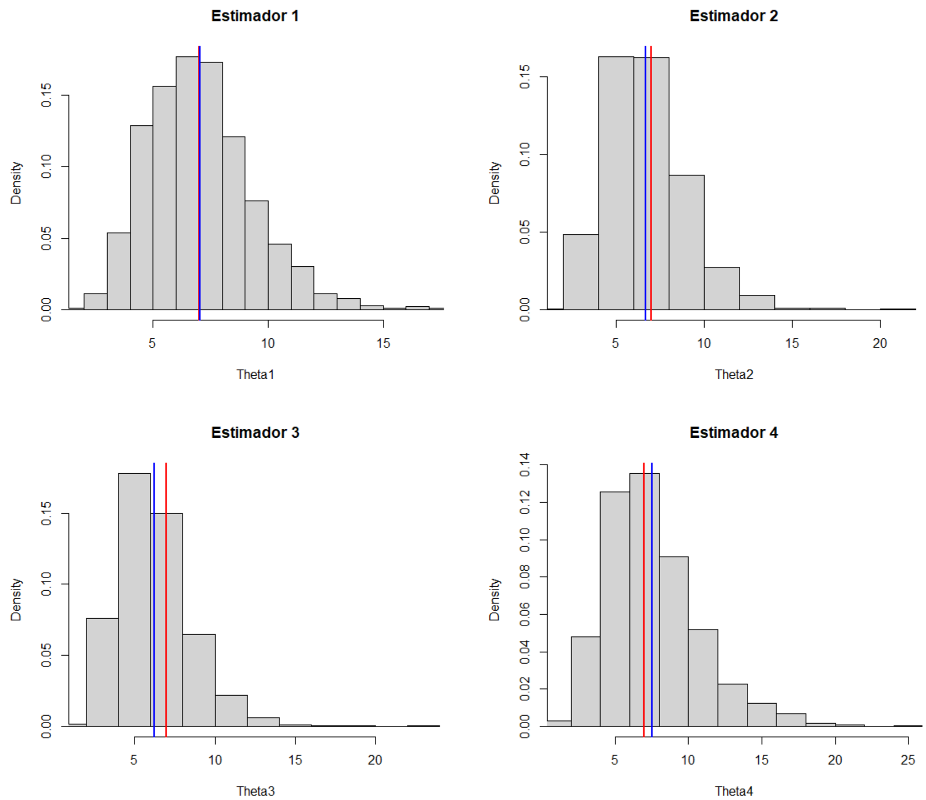
$$6 \cdot \hat{\theta}^3 = \frac{1}{n}\sum_{i=1}^{n} X_i^3$$
$$\hat{\theta}^3 = \frac{1}{6n}\sum_{i=1}^{n} X_i^3$$
$$\hat{\theta}_{MOM,3} = \sqrt[3]{\frac{1}{6n}\sum_{i=1}^{n} X_i^3}$$

Bonus: Propondremos un cuarto estimador de θ, el cuál llamaremos $\hat{\theta}_4$ y estará en términos de la mediana muestral. Y para esto, lo haremos por analogía. Sabemos que la mediana poblacional de una distribución exponencial es $\ln(2) \cdot \theta$, y que por analogía, la mediana muestral se comporta de la misma manera, así que igualaremos esto y despejaremos a θ para hallar a $\hat{\theta}_4$.

$$\ln(2) \cdot \hat{\theta} = M_e$$
$$\hat{\theta}_4 = \frac{M_e}{\ln(2)}$$

c. Genere $m = 1000$ simulaciones de una muestra aleatoria de tamaño $n = 10$ de la distribución $Exp(\theta = 7)$. Para cada una de las m muestras, calcule y almacene las estimaciones de los tres estimadores de momentos construidos. Haga un histograma de las m estimaciones de cada uno. Añada una línea vertical en cada histograma para indicar el verdadero valor del parámetro y otra línea para indiciar el promedio de las m estimaciones en cada caso. ¿Hay indicios de que los tres estimadores sean insesgados? Reporte en una tabla el promedio y la varianza obtenidos de las m estimaciones para este tamaño de muestra. Establezca qué estimador es mejor en términos de los valores de error cuadrático medio (estimado) y añada el MSE estimado a la tabla para este tamaño de muestra.

Solución. Veamos los histogramas de los estimadores aplicados a las $m = 1000$ muestras con $n = 10$. El verdadero valor del parámetro estará indicado con una línea vertical de color rojo, mientras que el promedio de las estimaciones estará indicado con una línea vertical de color azul.



Como podemos observar, hay indicios de que $\hat{\theta}_{MOM,1}$ resulta insesgado, teóricamente lo es, pero gráficamente, podemos darnos cuenta de que esto parece ser cierto. Por otra parte, $\hat{\theta}_{MOM,2}$, $\hat{\theta}_{MOM,3}$ y $\hat{\theta}_4$ no parecen ser insesgados a simple vista, por lo que en un primer instante, no tenemos indicios para afirmar que son insesgados.

Ahora, veamos el promedio, varianza y el MSE estimado de los cuatro estimadores, con tamaño de muestra $n = 10$ en una tabla.

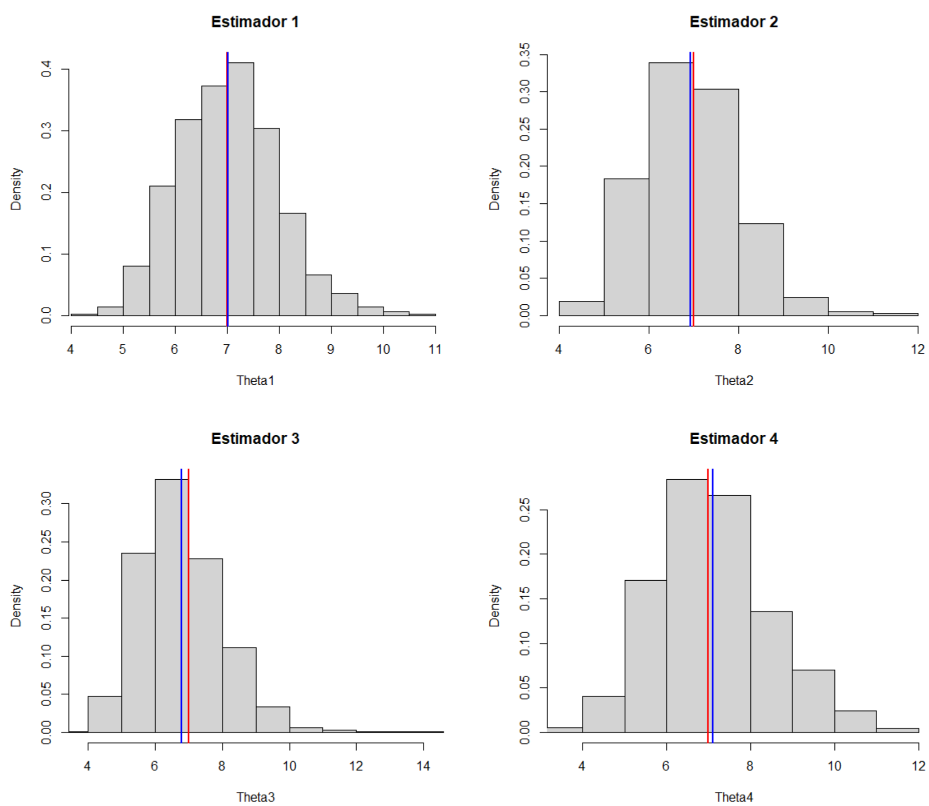
$n = 10$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta}_{MOM,1}$	7.037965	5.11891	5.120352
$\hat{\theta}_{MOM,2}$	6.682708	5.333399	5.434073
$\hat{\theta}_{MOM,3}$	6.242798	5.487592	6.060946
$\hat{\theta}_4$	7.549434	10.29698	10.59886

Para $n = 10$, el mejor estimador en términos de MSE estimado es $\hat{\theta}_{MOM,1}$, seguido por $\hat{\theta}_{MOM,2}$, luego $\hat{\theta}_{MOM,3}$ y por último $\hat{\theta}_4$ como el estimador menos eficiente en MSE estimado.

d. Repita el procedimiento descrito en **c.**, variando los tamaños de muestra por los valores $n = 50, 100, 200, 500, 1000$. ¿Se mantiene la misma conclusión acerca de qué estimador es mejor en términos de error cuadrático medio (estimado)? ¿Qué estimador de momentos resulta ser "mejor.en términos de error cuadrático medio?

Solución.

Tamaño de muestra de 50: Veamos los histogramas de los estimadores, aplicados a las $m = 1000$ muestras, aumentando el tamaño de muestra a $n = 50$, y veamos si podemos concluir algo diferente.



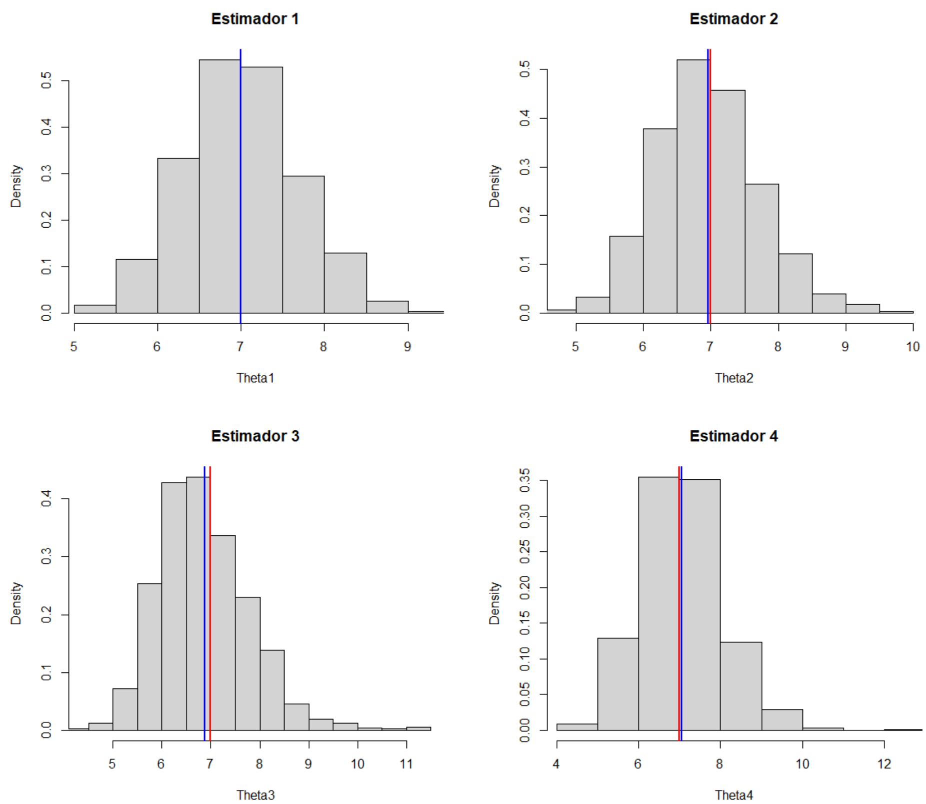
Como podemos ver, siguen habiendo indicios de que $\hat{\theta}_{MOM,1}$ resulta insesgado, pero con un tamaño de muestra de $n = 50$, $\hat{\theta}_{MOM,2}$ y $\hat{\theta}_4$ parecen ser un poco más eficientes que en el caso anterior. Aún no tenemos indicios de que $\hat{\theta}_{MOM,3}$ sea insesgado.

Ahora, veamos el promedio, varianza y el MSE estimado de los cuatro estimadores, con tamaño de muestra $n = 50$ en una tabla.

$n = 50$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta}_{MOM,1}$	7.014526	0.9330667	0.9332777
$\hat{\theta}_{MOM,2}$	6.934594	1.137246	1.141524
$\hat{\theta}_{MOM,3}$	6.784956	1.534249	1.580493
$\hat{\theta}_4$	7.097445	1.894192	1.903687

Para $n = 50$, el mejor estimador en términos de MSE estimado es $\hat{\theta}_{MOM,1}$, seguido por $\hat{\theta}_{MOM,2}$, luego $\hat{\theta}_{MOM,3}$ y por último $\hat{\theta}_4$ como el estimador menos eficiente en MSE estimado. Aunque en este caso, con $n = 50$, los cuatro estimadores son mucho más eficientes y tienen una varianza notoriamente más pequeña.

Tamaño de muestra de 100: Veamos los histogramas de los estimadores aplicados a las $m = 1000$ muestras, aumentando el tamaño de muestra de nuevo, pero esta vez a $n = 100$.



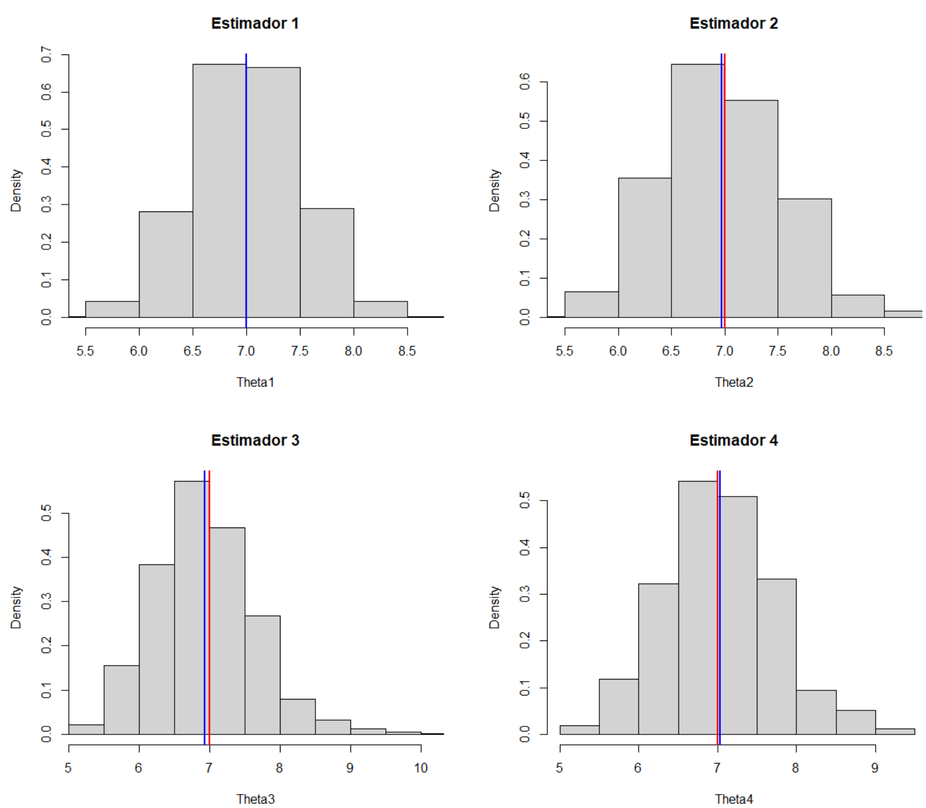
Como podemos ver, $\hat{\theta}_{MOM,1}$ sigue resultando insesgado, y con un tamaño de muestra de $n = 100$, $\hat{\theta}_{MOM,2}$ y $\hat{\theta}_4$ tienen un sesgo muy reducido, tenemos indicios de que son insesgados, pero sigue siendo considerable como para que no lo afirmemos con certeza. Se empiezan a ver indicios de que el sesgo de $\hat{\theta}_{MOM,3}$ se está haciendo cada vez más pequeño, pero sigue siendo lo suficientemente considerable como para afirmar que hay indicios de insesgamiento.

Ahora, veamos el promedio, varianza y el MSE estimado de los cuatro estimadores, con tamaño de muestra $n = 100$ en una tabla.

$n = 100$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta}_{MOM,1}$	7.002645	0.4628251	0.4628321
$\hat{\theta}_{MOM,2}$	6.957651	0.5892087	0.5910021
$\hat{\theta}_{MOM,3}$	6.869533	0.8896814	0.906703
$\hat{\theta}_4$	7.048835	0.9476742	0.9500591

Para $n = 100$, el mejor estimador en términos de MSE estimado sigue siendo $\hat{\theta}_{MOM,1}$, seguido por $\hat{\theta}_{MOM,2}$, luego $\hat{\theta}_{MOM,3}$ y por último $\hat{\theta}_4$ como el estimador menos eficiente en MSE estimado. Aunque en este caso, con $n = 100$, los cuatro estimadores son demasiado eficientes, ya que el sesgo de cada uno es muy reducido y su varianza también es muy pequeña respecto al primer caso con $n = 10$.

Tamaño de muestra de 200: Veamos los histogramas de los estimadores aplicados a las $m = 1000$ muestras, con un tamaño aún mayor, de $n = 200$



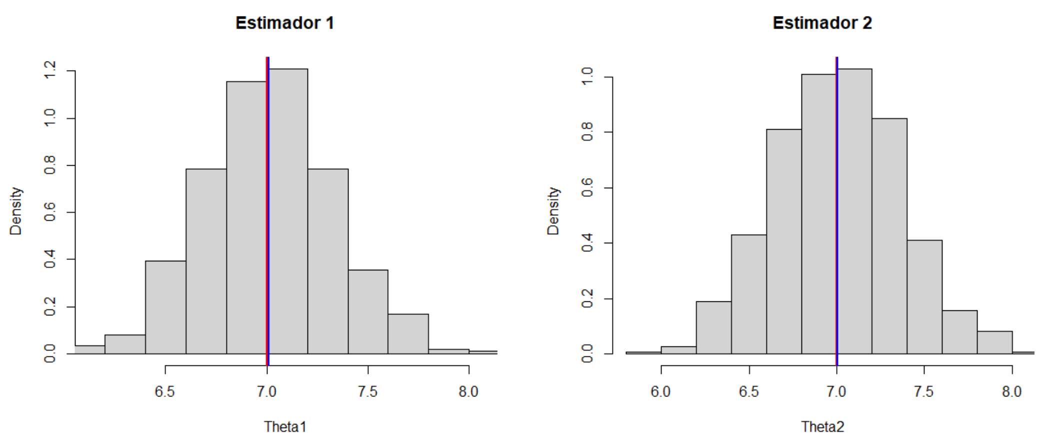
$\hat{\theta}_{MOM,1}$ es insesgado, y con un tamaño de muestra de $n = 200$, $\hat{\theta}_{MOM,2}$ y $\hat{\theta}_4$ tienen un sesgo muy reducido, podríamos afirmar que son insesgados. Como en el caso anterior, hay indicios de que el sesgo de $\hat{\theta}_{MOM,3}$ se está haciendo cada vez más pequeño, pero sigue siendo considerable como para concluir algo de forma contundente. Hay indicios de que los otros tres estimadores son asintóticamente insesgados.

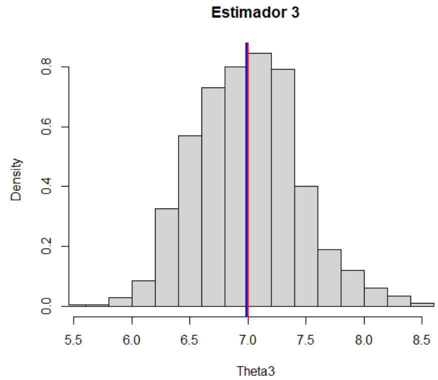
Ahora, veamos el promedio, varianza y el MSE estimado de los cuatro estimadores, con tamaño de muestra $n = 200$ en una tabla.

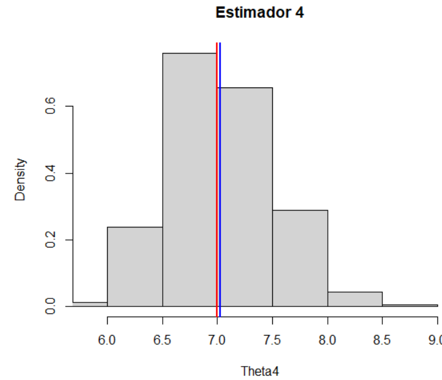
$n = 200$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta}_{MOM,1}$	6.998949	0.2546504	0.2546515
$\hat{\theta}_{MOM,2}$	6.973724	0.3323565	0.333047
$\hat{\theta}_{MOM,3}$	6.927221	0.5248056	0.5301023
$\hat{\theta}_4$	7.029524	0.4883891	0.4892608

Para $n = 200$, el mejor estimador en términos de MSE estimado sigue siendo $\hat{\theta}_{MOM,1}$, seguido por $\hat{\theta}_{MOM,2}$, pero en este caso, ocurre algo interesante, y es que $\hat{\theta}_4$ resulta más eficiente en términos de MSE estimado que $\hat{\theta}_{MOM,3}$. Aunque puede deberse al azar, con $n = 200$, el peor estimador de los cuatro resultó ser ahora $\hat{\theta}_{MOM,3}$. Aún así, los cuatro estimadores son muy eficientes, y la diferencia entre los cuatro estimadores, en términos de MSE estimado, es de apenas décimas.

Tamaño de muestra de 500: Veamos los histogramas de los estimadores aplicados a las $m = 1000$ muestras, con un tamaño de muestra ya muy grande, de $n = 500$.







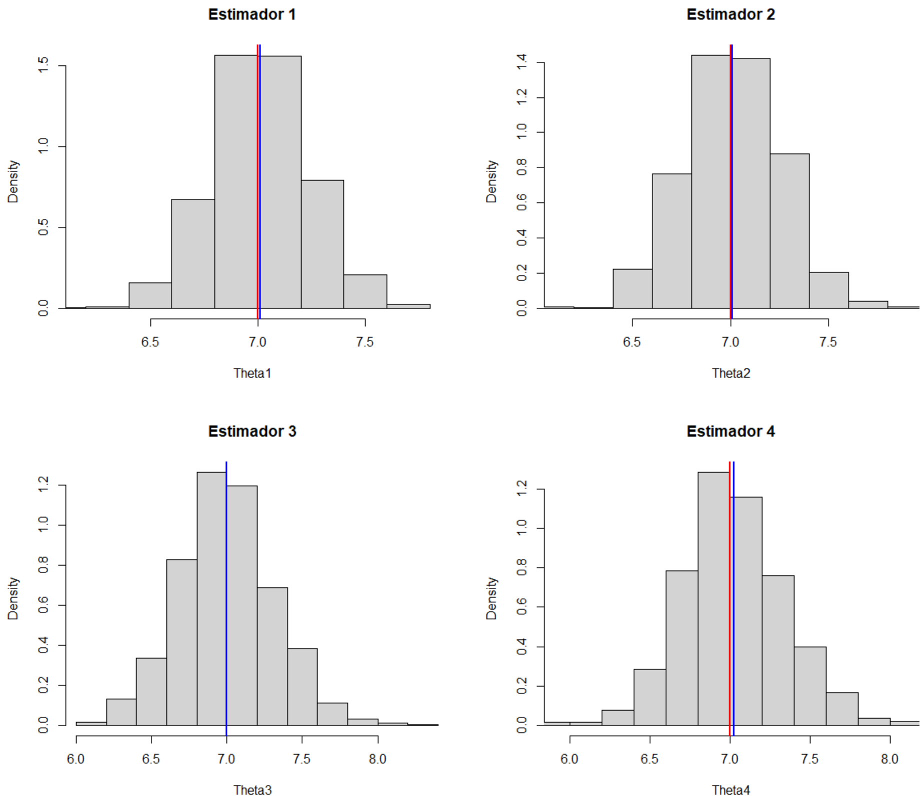
Ya sabemos que $\hat{\theta}_{MOM,1}$ es insesgado, y con un tamaño de muestra de $n = 500$, $\hat{\theta}_{MOM,2}$, $\hat{\theta}_{MOM,3}$ y $\hat{\theta}_4$, también son insesgados, y en este caso, el sesgo de $\hat{\theta}_4$ es mayor que el de $\hat{\theta}_{MOM,3}$, suceso que no había pasado antes. Este es el primer caso en el que el insesgamiento de los cuatro estimadores se puede afirmar, pues su sesgo está muy cercano a cero.

Ahora, veamos el promedio, varianza y el MSE estimado de los cuatro estimadores, con tamaño de muestra $n = 500$ en una tabla.

$n = 500$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta}_{MOM,1}$	7.009512	0.1009294	0.1010199
$\hat{\theta}_{MOM,2}$	7.002794	0.1242783	0.1242861
$\hat{\theta}_{MOM,3}$	6.98788	0.204097	0.2042439
$\hat{\theta}_4$	7.024957	0.2161472	0.2167701

Para $n = 500$, el mejor estimador en términos de MSE estimado sigue siendo $\hat{\theta}_{MOM,1}$, seguido por $\hat{\theta}_{MOM,2}$, luego retomamos la tendencia de que $\hat{\theta}_{MOM,3}$ queda tercero en cuanto a eficiencia de MSE, y por último $\hat{\theta}_4$. En este caso, volvemos a que $\hat{\theta}_4$ es el peor estimador en términos de MSE, pero la diferencia entre $\hat{\theta}_4$ y $\hat{\theta}_{MOM,3}$ es de centécimas, por lo que puede que se deba al azar. Lo que sí podemos concluir es que con muestras tan grandes, los cuatro estimadores resultan ser muy eficientes.

Tamaño de muestra de 1000: Veamos los histogramas de los estimadores aplicados a las $m = 1000$ muestras, con el mayor tamaño de muestra que veremos, $n = 1000$.



Con un tamaño de muestra de $n = 1000$, no resulta extraño que $\hat{\theta}_{MOM,1}$ sea insesgado, y que $\hat{\theta}_{MOM,2}$, $\hat{\theta}_{MOM,3}$ y $\hat{\theta}_4$, también sean insesgados. Lo que si es una novedad, es que en este caso $\hat{\theta}_{MOM,3}$ tiene el sesgo más reducido de todos, de forma muy notoria, algo que no había pasado antes.

Ahora, veamos el promedio, varianza y el MSE estimado de los cuatro estimadores, con tamaño de muestra $n = 1000$ en una tabla.

$n = 1000$	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{MSE}$
$\hat{\theta}_{MOM,1}$	7.012345	0.05168741	0.05183981
$\hat{\theta}_{MOM,2}$	7.007387	0.06130024	0.0613548
$\hat{\theta}_{MOM,3}$	6.999483	0.09902806	0.09902833
$\hat{\theta}_4$	7.024999	0.1048951	0.10552

Para $n = 1000$, el mejor estimador en términos de MSE estimado sigue siendo $\hat{\theta}_{MOM,1}$, seguido por $\hat{\theta}_{MOM,2}$, luego $\hat{\theta}_{MOM,3}$ y por último $\hat{\theta}_4$. En este caso, pese a que el sesgo de $\hat{\theta}_{MOM,3}$ era el menor de todos, la varianza del estimador hizo que su MSE estimado siguiera posicionandose tercero en nuestra lista de cuatro estimadores. Por otra parte, el estimador $\hat{\theta}_4$ sigue siendo el peor de todos en términos de MSE estimado, aún con un tamaño de muestra de $n = 1000$.

Para terminar, haremos una tabla para clasificar a los cuatro estimadores, en la que veremos con cada tamaño de muestra, qué tan buenos son en términos de MSE estimado. Si con un tamaño de muestra n tienen el menor MSE estimado, obtienen 3 puntos, si son los segundos mejores, o sea, tienen el segundo menor MSE estimado, obtienen 2, si fueron el tercer estimador obtienen un punto y si fueron el peor estimador de los cuatro, no obtienen puntos.

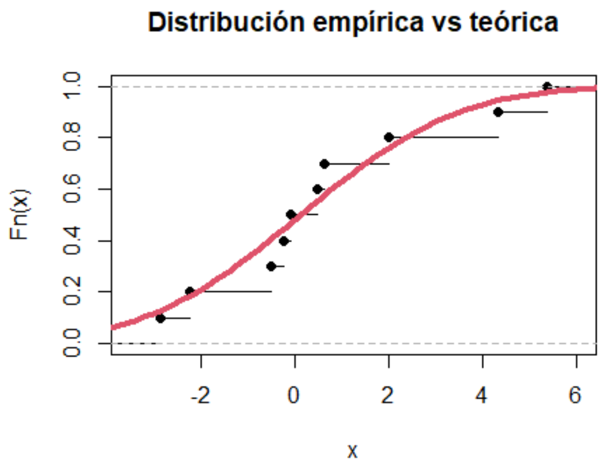
n	$\hat{\theta}_{MOM,1}$	$\hat{\theta}_{MOM,2}$	$\hat{\theta}_{MOM,3}$	$\hat{\theta}_4$
10	3	2	1	0
50	3	2	1	0
100	3	2	1	0
200	3	2	0	1
500	3	2	1	0
1000	3	2	1	0
Total	18	12	5	1

Como vemos, el mejor estimador en términos de MSE estimado indiscutiblemente es $\hat{\theta}_{MOM,1}$, lo cual se ha mantenido desde que la muestra era pequeña con $n = 10$, hasta cuando aumentamos la muestra a un tamaño de $n = 1000$. El segundo mejor estimador es $\hat{\theta}_{MOM,2}$. Aunque en un caso el MSE estimado de $\hat{\theta}_4$ fue mejor que el de $\hat{\theta}_{MOM,3}$, en los demás $\hat{\theta}_{MOM,3}$ fue mejor estimador, dándole la tercera posición en nuestra búsqueda del mejor de los cuatro estimadores, y por tanto, el peor de todos es $\hat{\theta}_4$.

Problema 4:

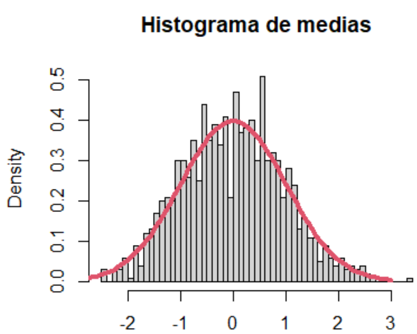
Solución:

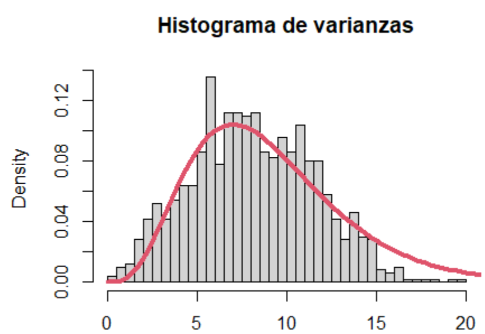
a. Como $X \sim N(\mu = \frac{1}{7}, \sigma^2 = 7)$ tenemos que $\bar{x_n} = 0{,}143$ y $\sigma = 2{,}65$. Por otra parte, del dataset generado obtenemos: $\hat{x_n} = 0{,}492$ y $\hat{\sigma} = 5{,}328$.Son estimaciones relativamente cercanas.
b.



Las dos funciones se comportan de una manera muy diferente, sin embargo podemos empezar a observar como es el comportamiento de la sucesión $\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F_X(x,\theta)|$

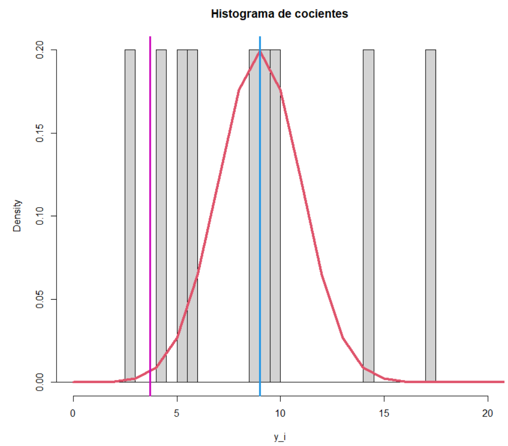
d.



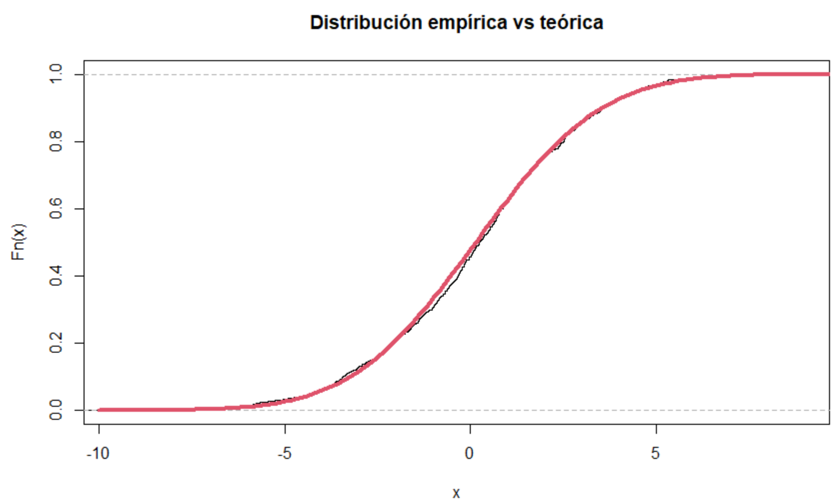


e. De nuevo podríamos basarnos en el resultado del teorema de Glivenko-Cantelli para afirmar que dado la que convergencia casi segura implica convergencia en probabilidad entonces las distribuciones asíntoticas derivadas de $\hat{F}_n(x)$ se van a comportar casi seguramente como las de $F_X(x,\theta)$.Retomando a la idea original del curso, estos conjuntos de datos resultan ser realizaciones de variables aleatorias y estamos tomando 1000 de ellas, sin embargo al tener una muestra inicial de un tamaño pequeño, los datos generados no siguen el modelo de las distribuciones que estamos estudiando por ello se presenta un mayor desajuste que si tomaramos una muestra de un tamaño mayor.

f.En este punto consideramos que proponer un modelo conocido que se ajuste bien a estos datos es bastante difícil. Sin embargo, al observar la gráfica podemos pensar que estos datos parececen tener una distribución normal. Por esto usaremos el modelo $X \sim N(4,4)$ en esta simulación, para otra simulación se debe cambiar la media.

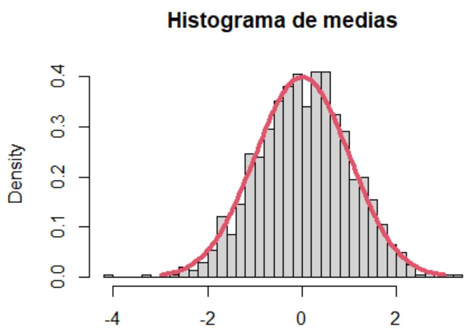


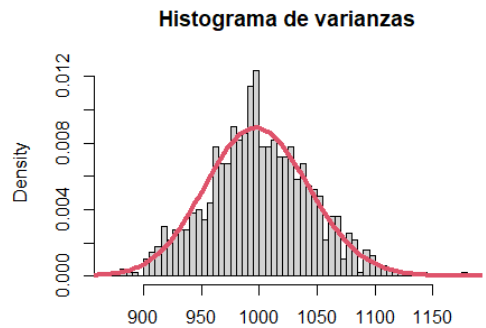
El sesgo calculado en este modelo es: $B[\hat{c}] = 6{,}82$. En efecto, parece ser un estimador sesgado.
g. Repitiendo este experimento para un tamaño de muestra $n = 1000$ obtenemos los siguientes resultados:



Aquí podemos visualizar el efecto del teorema de Glivenko- Cantelli en la simulación pues vemos que la distancia entre la densidad real y la densidad empírica se hace más pequeña a medida que $n \to \infty$

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_X(x, \theta)| \xrightarrow{\text{c.s}} 0 \ \text{ cuando } \ n \to \infty$$





Como mencionamos en e) el hecho de que la función de distribución empirica sea cercana al distribución real permite influye en el ajuste de las curvas sea más preciso.
Respecto al punto del cociente de variación, obtenemos una mejor visualización de sus datos, podemos observar que χ^2_{27} parece ser una buena aproximación, sin embargo el sesgo calculado resulta ser mayor $B[\hat{c}] = 13{,}24$ al punto anterior, lo que nos lleva que el modelo para una muestra de tamaño $n = 10$ es mucho peor.