Entrega 2 - Parte 1

Ignacio Spiousas

6 de junio de 2023

Ejercicio 4 de la práctica 2

Voy a elegir la distribución exponencial ($\mathcal{E}(\theta)$) para hallar el estimador de momentos, el estimador de máxima verosimilitud, estudiar la consistencia, el sesgo y la distribución asintótica.

Estimador de momentos

Supongamos que las variables aleatorias X_i son independintes y tienen todas distribución exponencial de parámetro θ:

$$X_i \overset{i.i.d}{\sim} \mathcal{E}(\theta)$$

De acuerdo al método de los momentos basta con plantear el primer momento (ya que tenemos un sólo parámetro).

$$\frac{1}{n}\sum_{i=1}^{n} X_i = E_{\hat{\theta}}(X_1)$$

La esperanza de la distribución exponencial es justamente la inversa de su parámetro, por lo tanto obtenemos que el estimador por el método de los momentos ($\hat{\theta}$) se define como:

$$\bar{X}_n = \frac{1}{\hat{\theta}}$$
$$\hat{\theta} = \frac{1}{\bar{X}_n}$$

Es decir, el estimador del parámetro es la inversa del promedio.

Estimador de máxima verosimilitud (MV)

Vamos a calcular el estimador de MV de la misma variable aleatoria que el ítem anterior. Lo primero que vamos a hacer es plantear la función de versosimilitud ($L(\theta; \underset{\sim}{x})$). La misma se construye con la probabilidad conjunta de que las X_i tomen los valores x_i, pero al tratase de variables $i.i.d$ se transforma en la productoria de sus funciones de densidad de probabilidades (por ser una V.A. continua). La podemos expresar de la siguiente forma:

$$L(\theta; \underset{\sim}{x}) = \prod_{i=1}^{n} f(x_i, \theta)$$

Reemplazando $f(x_i, \theta)$ por las funciones de densidad de probabilidad de la exponencial nos queda:

$$L(\theta;\underset{\sim}{x}) = \prod_{i=1}^{n} \theta e^{-x_i\theta} \mathbb{I}_{(0,\infty)}(x_i)$$

Asumiendo que todos los x_i son mayores que cero puedo sacarme de encima la indicadora. Ahora apliquemos log a ambos lados de la igualdad para asi calcular la log-verosimilitud $\ell(\theta;\underset{\sim}{x})$.

$$\ell(\theta;\underset{\sim}{x}) = nlog(\theta) - \theta\sum_{i=1}^{n} x_i$$

Ahora tenemos que encontrar el valor del parámetro que maximiza la log-verosimilitud (que a su vez, por ser el log creciente también maximiza la verosimilitud). Para esto vamos a derivar con respecto al parámetro e igualar a cero. Empecemos obteniendo la derivada:

$$\frac{\partial\ell(\theta;\underset{\sim}{x})}{\partial\theta} = n\frac{\partial log(\theta)}{\partial\theta} - \sum_{i=1}^{n} x_i\frac{\partial\theta}{\partial\theta}$$
$$\frac{\partial\ell(\theta;\underset{\sim}{x})}{\partial\theta} = n\frac{1}{\theta} - \sum_{i=1}^{n} x_i$$

Ahora igualemos a cero y obtengamos el valor del parámetro que maximiza la log-versimilitud, es decir, el estimador de máxima verosimilitud ($\tilde{\theta}$):

$$n\frac{1}{\tilde{\theta}} - \sum_{i=1}^{n} x_i = 0$$
$$n\frac{1}{\tilde{\theta}} = \sum_{i=1}^{n} x_i$$
$$\tilde{\theta} = \frac{n}{\sum_{i=1}^{n} x_i}$$
$$\tilde{\theta} = \frac{1}{\bar{x_n}}$$

Podemos ver que el estimador de momentos y el de máxima verosimilitud coinciden.

Extra

Probemos que el punto crítico $\tilde{\theta} = 1/\bar{x_n}$ es un máximo. Derivemos de nuevo con respecto a θ:

$$\frac{\partial^2\ell(\theta;\underset{\sim}{x})}{\partial\theta^2} = \frac{\partial(n\frac{1}{\theta} - \sum_{i=1}^{n} x_i)}{\partial\theta}$$
$$\frac{\partial^2\ell(\theta;\underset{\sim}{x})}{\partial\theta^2} = -n\frac{1}{\theta^2}$$

Ahora reemplacemos $\theta = \tilde{\theta} = 1/\bar{x_n}$:

$$\frac{\partial^2\ell(\theta;\underset{\sim}{x})}{\partial\theta^2} = -n\frac{1}{\theta^2} = -n(\bar{x_n})^2$$

Que como $\bar{x_n} > 0$ porque todas las x_i son mayores que cero, resulta que $\frac{\partial^2\ell(\theta;\underset{\sim}{x})}{\partial\theta^2} < 0$, es decir, estamos en un máximo de la función log-verosimilitud.

Sesgo

El sesgo del estimador $\hat{\theta}$ se define como:

$$sesgo(\hat{\theta}, \theta) = b(\hat{\theta}, \theta) = E_\theta(\hat{\theta}) - \theta$$

Como ambos estimadores son iguales vamos a estudiar el sesgo de uno de ellos (Momentos) pero el cálculo es el mismo que para el estimador de máxima verosimilitud.

Empecemos calculando la esperanza del estimador. Definamos a Y como la suma de variables aleatorias X, de forma que $Y = \sum_{i=1}^n X_i$. Entonces:

$$E_\theta(\hat{\theta}) = E_\theta(1/\bar{X}_n) = E_\theta(\frac{n}{\sum_{i=1}^n X_i}) = nE_\theta(\frac{1}{Y})$$

Notemos que Y, por ser la suma de n variables aleatorias con distribución exponencial de parámetro θ, tiene una distribución gamma de parámetros n y θ ($Y \sim \Gamma(n, \theta)$). Teniendo en cuenta esto, podemos calcular la esperanza de Y^{-1}:

$$E_\theta(Y^{-1}) = \int_0^\infty y^{-1} \frac{\theta^n}{\Gamma(n)} y^{n-1} e^{-\theta y} dy = \int_0^\infty \frac{\theta^n}{\Gamma(n)} y^{n-2} e^{-\theta y} dy$$

La integral que nos queda tiene casi la forma de la integral de una $\Gamma(n-1, \theta)$ de cero a infinito. Lo primero con lo que tenemos que lidiar es con la función $\Gamma(n)$, transformándola en $\Gamma(n-1)$. Para esto recordemos que $\Gamma(n) = (n-1)!$ y $\Gamma(n-1) = (n-2)!$, por lo que $\Gamma(n-1)(n-1) = (n-1)! = \Gamma(n)$. También podemos reescribir θ^n como $\theta\theta^{n-1}$.Reemplacemos en la integral:

$$E_\theta(Y^{-1}) = \int_0^\infty \frac{\theta\theta^{n-1}}{\Gamma(n-1)(n-1)} y^{n-2} e^{-\theta y} dy = \frac{\theta}{n-1} \int_0^\infty \frac{\theta^{n-1}}{\Gamma(n-1)} y^{n-2} e^{-\theta y} dy = \frac{\theta}{n-1} \times 1$$

Y volviendo a $E_\theta(\hat{\theta}) = nE_\theta(Y^{-1})$ y reemplazando $E_\theta(Y^{-1}) = \frac{\theta}{n-1}$ nos queda que:

$$E_\theta(\hat{\theta}) = nE_\theta(Y^{-1}) = \frac{n}{n-1}\theta$$

Volviendo a la definición de sesgo podemos ver que los estimadores son sesgados.

$$b(\hat{\theta}, \theta) = E_\theta(\hat{\theta}) - \theta = \frac{n}{n-1}\theta - \theta = \frac{1}{n-1}\theta$$

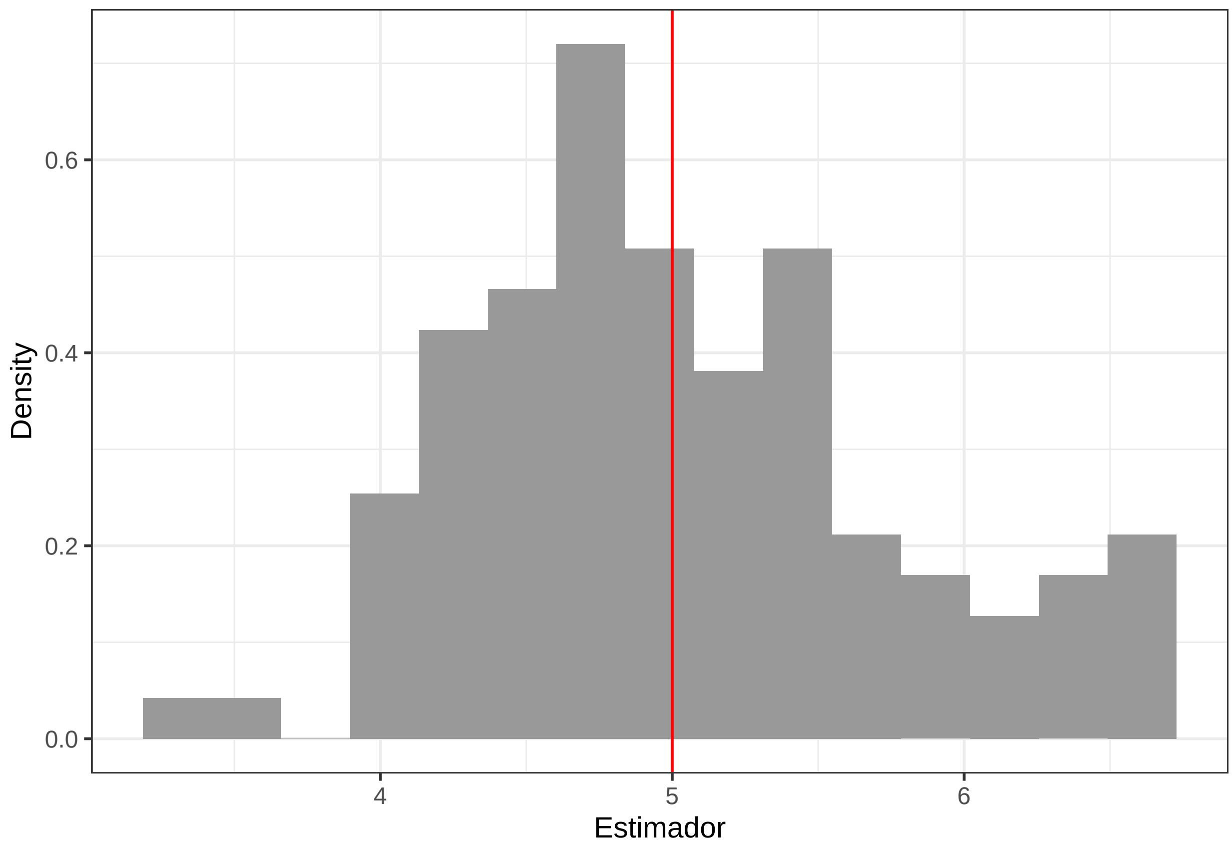
Extra computacional

Vamos a simular unas estimaciones de variables aleatorias i.i.d. con distribución exponencial y ver la distribución de los estimadores. Vamos a correr 100 repeticiones (`Nrep`) del experimento en el que medimos 50 realizaciones (`n`) de variables exponenciales de parámetro 5 (`theta`)

```r
set.seed(1)
Nrep <- 100
n <- 50
theta <- 5
est <- c()
for (i in 1:Nrep) {
  x <- rexp(n = n, rate = theta)
  est[i] <- 1/mean(x)
}

tibble(est) %>%
```

```r
ggplot(aes(x = est, y = after_stat(density))) +
geom_histogram(bins = 15, fill = "gray60") +
geom_vline(xintercept = theta, color = "red") +
labs(x = "Estimador", y = "Density") +
theme_bw()
```



Podemos ver que, efectivamente, los valores del estimador están sesgados hacía valores más altos (en su esperanza el parámetro está multiplicado por un factor mayor a uno $\frac{n}{n-1}$).

Consistencia

La consistencia la podemos probar utilizando la ley débil de los grandes números. Empecemos planteando el estimador y su convergencia en probabilidad por ley de grandes números:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \frac{1}{\theta}$$

Es decir, el promedio converge en probabilidad a la esperanza de la exponencial ($\frac{1}{\theta}$). Ahora recordemos la propiedad de la convergencia:

$$X_n \xrightarrow{p} X \quad \Rightarrow \quad g(X_n) \xrightarrow{p} g(X)$$

Entonces, si tenemos en cuenta que $g(.)$ es elevar a la potencia $(.)^{-1}$, podemos decir que:

$$\frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i} \xrightarrow{p} \theta$$
$$\hat{\theta} \xrightarrow{p} \theta$$

Es decir, el estiamdor $\hat{\theta} \xrightarrow{p} \theta$, y es consistente.

Otra alternativa es desarrollar el error cuadrático medio del estimador ($MSE(\hat{\theta})$) y verificar que tiende a cero cuando n tiende a infinito. El $MSE(\hat{\theta})$ se puede calcular como:

$$MSE(\hat{\theta}) = Var_\theta(\hat{\theta}) + b(\hat{\theta})$$

$b(\theta)$ ya lo tenemos calculado, por lo que sólo necesitamos desarrollar $Var(\theta)$. Para esto, vamos a reescribirlo la misma variable Y utilizada anteriormente de la siguiente forma:

$$Var_\theta(\hat{\theta}) = E_\theta(\hat{\theta}^2) - (E_\theta(\hat{\theta}))^2$$
$$Var_\theta(\hat{\theta}) = E_\theta(\frac{n^2}{(\sum_{i=1}^{n} X_i)^2}) - (E_\theta(\frac{n}{\sum_{i=1}^{n} X_i}))^2$$
$$Var_\theta(\hat{\theta}) = n^2 E_\theta(Y^{-2}) - n^2(E_\theta(Y^{-1}))^2$$

Ya sabemos que $E_\theta(Y^{-1}) = \frac{\theta}{n-1}$, así que sólo deberíamos calcular $E_\theta(Y^{-2})$. Y lo vamos a hacer operando de la misma forma que en el sesgo:

$$E_\theta(Y^{-2}) = \int_0^\infty \frac{\theta^2\theta^{n-2}}{\Gamma(n-2)(n-1)(n-2)} y^{n-3} e^{-\theta y} dy = \frac{\theta^2}{(n-1)(n-2)} \int_0^\infty \frac{\theta^{n-2}}{\Gamma(n-2)} y^{n-3} e^{-\theta y} dy = \frac{\theta^2}{(n-1)(n-2)} \times 1$$

Por lo tanto, la varianza se puede escribir como:

$$Var_\theta(\hat{\theta}) = n^2\theta^2(\frac{1}{(n-1)(n-2)} - \frac{1}{(n-1)^2})$$
$$Var_\theta(\hat{\theta}) = \frac{n^2\theta^2}{(n-1)^2}(\frac{n-1}{n-2} - 1)$$
$$Var_\theta(\hat{\theta}) = \frac{n^2\theta^2}{(n-1)^2(n-2)}$$

Y el $MSE(\hat{\theta})$ como:

$$
\begin{aligned}
MSE(\hat{\theta}) &= Var_\theta(\hat{\theta}) + b(\hat{\theta}) \\
MSE(\hat{\theta}) &= \frac{n^2\theta^2}{(n-1)^2(n-2)} + \frac{\theta^2}{(n-1)^2} \\
MSE(\hat{\theta}) &= \frac{\theta^2}{(n-1)^2}(\frac{n^2}{n-2} + 1) \\
MSE(\hat{\theta}) &= \frac{\theta^2}{(n-1)^2}(\frac{n^2+n-2}{n-2}) \\
MSE(\hat{\theta}) &= \frac{\theta^2}{(n-1)^2}(\frac{n^2+n-2}{n-2}) \\
MSE(\hat{\theta}) &= \frac{\theta^2}{(n-1)^2}(\frac{(n-1)(n+2)}{n-2}) \\
MSE(\hat{\theta}) &= \theta^2\frac{n+2}{(n-2)(n-1)} \\
MSE(\hat{\theta}) &= \theta^2\frac{n+2}{n^2-3n+2}
\end{aligned}
$$

Expresado de esta forma podemos evaluar el límite cuando n tiende a infinito:

$$lim_{n\to\infty}(MSE(\hat{\theta})) = lim_{n\to\infty}(\theta^2\frac{n+2}{n^2-3n+2}) = \theta^2 lim_{n\to\infty}(\frac{n+2}{n^2-3n+2})$$

Y si por L'Hopital derivamos:

$$lim_{n\to\infty}(MSE(\hat{\theta})) = \theta^2 lim_{n\to\infty}(\frac{1}{2n-3}) = 0$$

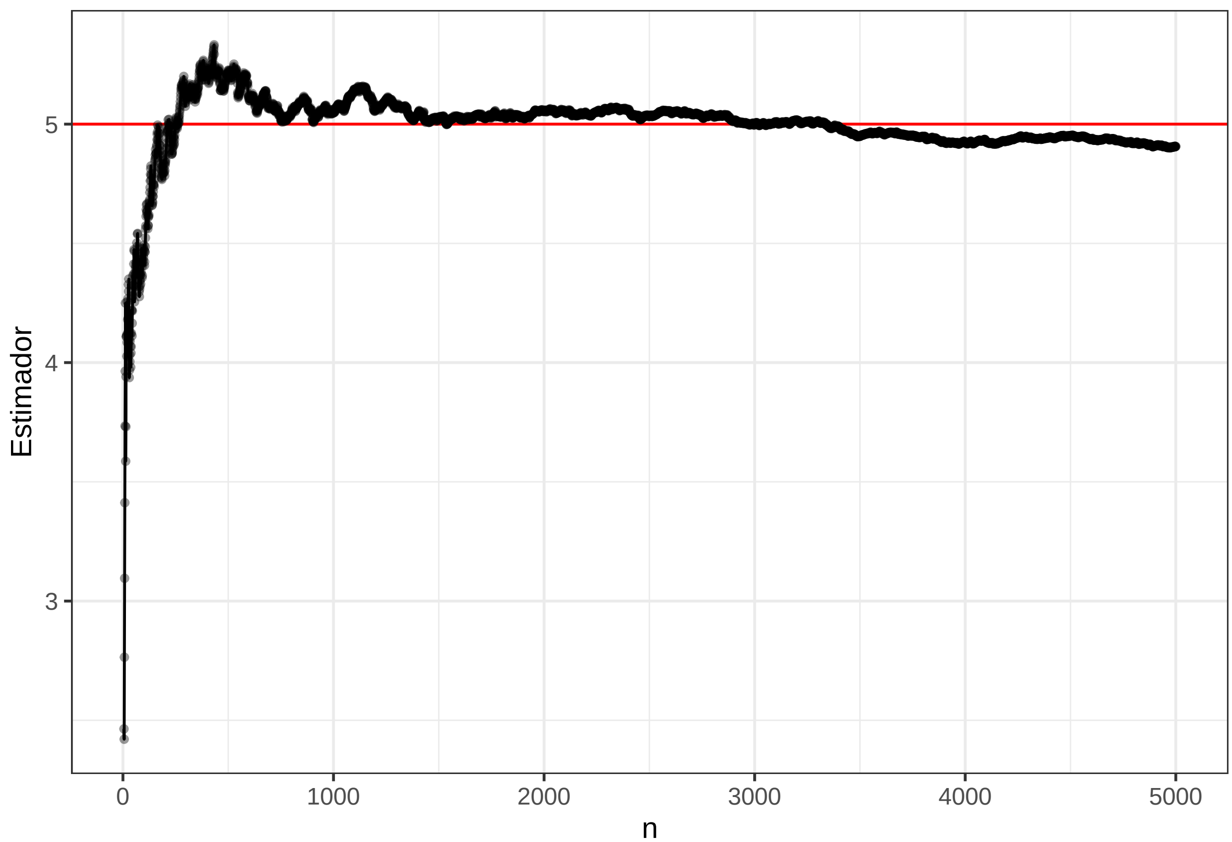
Es decir, el $lim_{n\to\infty}(MSE(\hat{\theta})) = 0$ y el estimador es consistente.

Extra computacional

Ahora simulemos la evolución del estimador (cualquiera de ellos) con el n. Vamos a variar `n` entre 5 y 5000 agregando realizaciones de variables exponenciales de parámetro 5 (`theta`). Después vamos a graficar el estimador en función de n y a marcar en rojo el valor real del parámetro θ.

```
set.seed(3)
x <- rexp(n = 5, rate = 2)
theta <- 5
n_max <- 5e3
est <- c()
for (n in 5:n_max) {
  x <- c(x, rexp(n = 1, rate = theta))
  est[n-4] <- 1/mean(x)
}

tibble(n = 5:n_max, est) %>%
  ggplot(aes(x = n, y = est)) +
  geom_hline(yintercept = theta, color = "red") +
  geom_line() +
  geom_point(alpha = 0.4, size = 1) +
  labs(x = "n", y = "Estimador") +
  theme_bw()
```



Vemos que el estimador tiende al parámetro real θ al aumentar n.s

Distribución asintótica

Como $\{X_i ... X_n\}$ son variables aleatorias i.i.d con esperanza $E[X_i] = \mu = 1/\theta$ y $Var[X_i] = \sigma^2 = 1/\theta^2$, cuando $n \to \infty$ ocurre que:

$$\sqrt{n}(\bar{X_n} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$
$$\sqrt{n}(\bar{X_n} - \frac{1}{\theta}) \xrightarrow{d} \mathcal{N}(0, \frac{1}{\theta^2})$$

También recordemos la definición de método Delta que dice que si:

$$\sqrt{n}[\bar{X_n} - \mu] \xrightarrow{D} \mathcal{N}(0, \sigma^2)$$

Entonces:

$$\sqrt{n}[g(\bar{X_n}) - g(\mu)] \xrightarrow{D} \mathcal{N}(0, \sigma^2 \cdot [g'(\mu)]^2)$$

Siempre que la función $g(.)$ sea continua y no nula. En nuestro caso, volvemos a considerar que $g(.)$ es elevar a la potencia $(.)^{-1}$ y nos queda que:

$$\sqrt{n}(\bar{X_n}^{-1} - (\frac{1}{\theta})^{-1}) \xrightarrow{d} \mathcal{N}(0, \frac{1}{\theta^2}[g'(\frac{1}{\theta})]^2)$$
$$\sqrt{n}(\bar{X_n}^{-1} - \theta) \xrightarrow{d} \mathcal{N}(0, \frac{1}{\theta^2}[g'(\frac{1}{\theta})]^2)$$

Y, como la derivada de $g(.)$ es $\frac{-1}{(.)^2}$, al evaluarla en $\frac{1}{\theta}$ queda θ^2. Reeplazando en la ecuación anterior obtenemos:

$$\sqrt{n}(\bar{X_n}^{-1} - \theta) \xrightarrow{d} \mathcal{N}(0, \frac{1}{\theta^2}(\theta^2)^2)$$
$$\sqrt{n}(\bar{X_n}^{-1} - \theta) \xrightarrow{d} \mathcal{N}(0, \theta^2)$$

O sea que la varianza asintótica sería θ^2.

Extra computacional

Podemos trabajar un poco con la distribución aproximada para un n grande (pero no infinito).

$$\begin{aligned} \sqrt{n}(\bar{X_n}^{-1} - \theta) &\approx \mathcal{N}(0, \theta^2) \\ (\bar{X_n}^{-1} - \theta) &\approx \mathcal{N}(0, \frac{\theta^2}{n}) \\ \bar{X_n}^{-1} &\approx \mathcal{N}(\theta, \frac{\theta^2}{n}) \\ \hat{\theta} &\approx \mathcal{N}(\theta, \frac{\theta^2}{n}) \end{aligned}$$

Es decir, si el n es "grande", la distribución del estimador es una normal con media θ y varianza $\frac{\theta^2}{n}$.

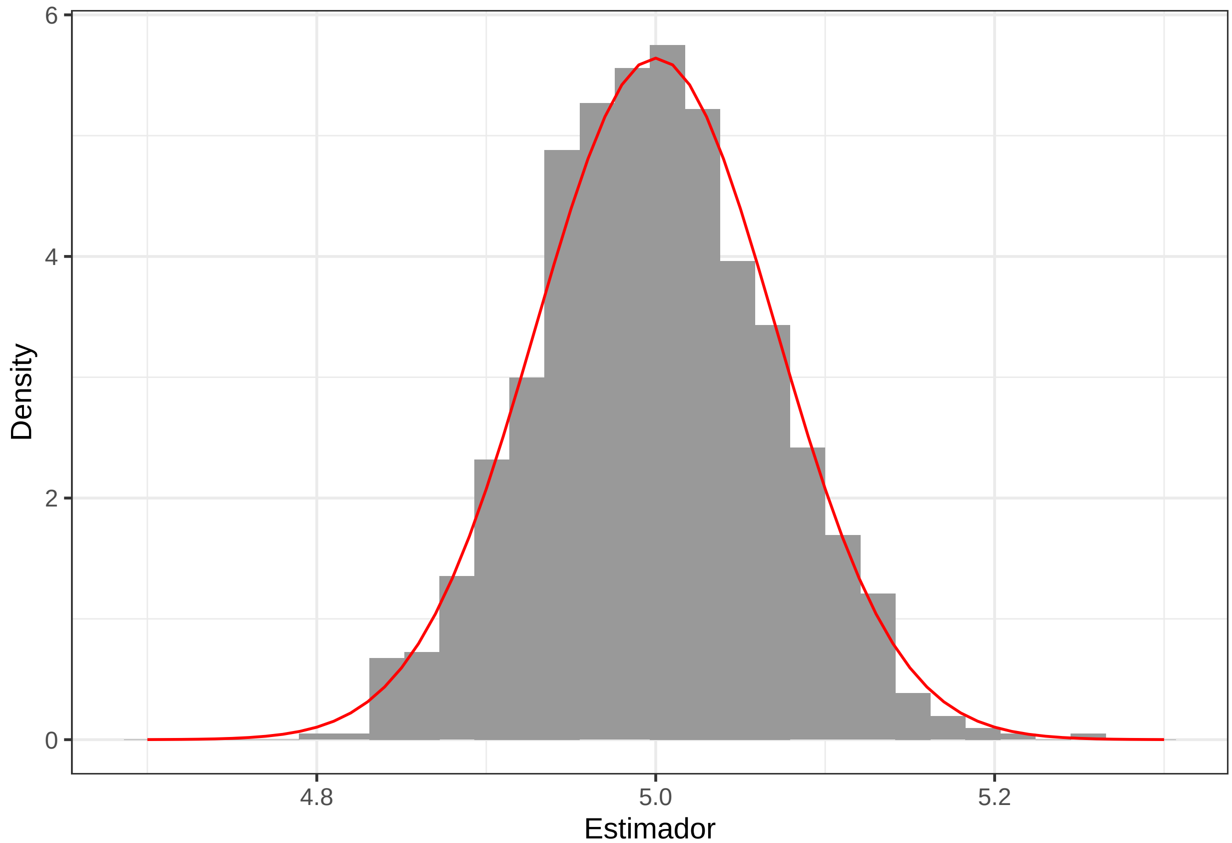
Vamos a simular unas estimaciones de variables aleatorias i.i.d. con distribución exponencial y ver la distribución de los estimadores. Vamos a correr 1000 repeticiones (`Nrep`) del experimento en el que medimos 50 realizaciones (`n`) de variables exponenciales de parámetro 5 (`theta`). Al histograma de las observaciones le vamos a superponer en rojo la distribución asintótica.

```
set.seed(1)
Nrep <- 1e3
n <- 5e3
theta <- 5
est <- c()
for (i in 1:Nrep) {
  x <- rexp(n = n, rate = theta)
  est[i] <- 1/mean(x)
}

mu_asyn <- theta
sigma_asyn <- theta/sqrt(n)

norm_asyn <- tibble(x = seq(4.7, 5.3, .01),
                    y = 1/(sigma_asyn*sqrt(2*pi)) * exp(-1/2 * ((x-mu_asyn)/sigma_asyn)^2))

tibble(est) %>%
  ggplot(aes(x =est, y = after_stat(density))) +
  geom_histogram(bins = 30, fill = "gray60") +
  geom_line(data = norm_asyn, aes(x = x, y = y), color = "red") +
  labs(x = "Estimador", y = "Density") +
  theme_bw()
```



Vemos que, en efecto, la densidad asintótica se eproxima mucho a la observada luego de repetir el experimento 1000 veces.