Propiedades de estimadores

Ejercicio 1

Sean $X_1, \dots, X_n$ variables aleatorias i.i.d con densidad dada por

$$f(x, \theta) = \theta^4 x e^{-\theta^2 x} I_{[0,+\infty)}$$

a- Hallar el EMV de θ.

b- Decidir si el estimador hallado es consistente.

c- Hallar el estimador de máxima verosimiltud de θ^2.

e- Estimar el ECM del estimador de θ para distintos valores de n, suponiendo que el verdadero θ es 2.

Resolución

a- El EMV se define como

$$\hat{\theta} = \text{argmax}_\theta L(\theta, x_1, \dots, x_n),$$

donde

$$L(\theta, x_1, \dots, x_n)$$

es la verosimilitud, o sea

$$\begin{aligned} L(\theta, x_1, \dots, x_n) &= \Pi_{i=1}^n f(x_i, \theta) \\ &= \Pi_{i=1}^n \theta^4 x_i e^{-\theta^2 x_i} I_{[0,+\infty)}(x_i) \\ &= \theta^{4n} \Pi_{i=1}^n x_i e^{-\theta^2 \sum_{i=1}^n x_i} I_{[0,+\infty)}(x_i) \\ &= \theta^{4n} \Pi_{i=1}^n x_i e^{-\theta^2 \sum_{i=1}^n x_i} \end{aligned}$$

ya que, como x_i son realizaciones de v.a. X_i con densidad $f(x, \theta)$, podemos suponer que $x_i > 0$ y, por lo tanto, que $I_{[0,+\infty)}(x_i) = 1$ para todo i.

Como log es una función creciente de θ,

$$\text{argmax}_\theta L(\theta, x_1, \dots, x_n) = \text{argmax}_\theta \log L(\theta, x_1, \dots, x_n)$$

Notación:

$$l(\theta, x_1, \dots, x_n) = \log L(\theta, x_1, \dots, x_n)$$

$$l(\theta, x_1, \ldots, x_n) = 4n \log \theta + \sum_{i=1}^{n} \log(x_i) + -\theta^2 \sum_{i=1}^{n} x_i$$

Tenemos que hallar el θ donde $l(\theta, x_1, \ldots, x_n)$ es máximo

$$\frac{\partial l}{\partial \theta}(\theta, x_1, \ldots, x_n) = \frac{4n}{\theta} - 2\theta \sum_{i=1}^{n} x_i$$

$$\frac{\partial l}{\partial \theta}(\theta, x_1, \ldots, x_n) = 0 \Leftrightarrow \frac{4n}{\theta} = 2\theta \sum_{i=1}^{n} x_i \Leftrightarrow \frac{4n}{2\sum_{i=1}^{n} x_i} = \theta^2 \Leftrightarrow \theta = \sqrt{\frac{2}{\overline{x}}}$$

Entonces

$$\hat{\theta} = \sqrt{\frac{2}{\overline{X}}}$$

b) Recordar que el estimador $\hat{\theta} = \hat{\theta}_n$ es consistente si

$$\hat{\theta}_n \longrightarrow_{n \to \infty} \theta$$

Sabemos que, por LGN,

$$\overline{X} \to E(X_1)$$

$$\hat{\theta} = \sqrt{\frac{2}{\overline{X}}}$$

Recordar además que, si $X_n \to^P a$ y g es continua en a entonces $g\left(X_n\right) \xrightarrow{p} g(a)$.

Por lo tanto

$$\sqrt{\frac{2}{\overline{x}}} \xrightarrow{P} \sqrt{\frac{2}{E(X_1)}}$$

Para ver cuánto da ese límite sólo falta calcular $E(X_1)$.

Para ello, observemos que $X_i \sim \Gamma(2, \theta^2)$.

Recordemos que la densidad de una $\Gamma(\alpha, \lambda)$ es

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} I_{(0,+\infty)}(x)$$

y que la esperanza de una v.a. $\Gamma(\alpha, \lambda)$ es α/λ.

por lo tanto

$$E(X_1) = \frac{2}{\theta^2} \text{ y esto implica que } \sqrt{\frac{2}{\overline{x}}} \xrightarrow{P} \sqrt{\frac{2}{2/\theta^2}} = \theta.$$

Esto demuestra que el EMV es un estimador consistente de θ.

d) Como $h(x) = x^2$ es una función inyectiva en $(0, +\infty)$, que es el conjunto de valores posibles de θ, se tiene que, por la propiedad de invarianza, el EMV de θ^2 es $\hat{\theta}^2$, o sea

$$\widehat{\theta^2} = \hat{\theta}^2 = \frac{2}{\overline{x}}$$

e)

Lo primero que hay que hacer es generar conjuntos de datos que provengan de la densidad f. Para ello usamos el teorema visto hace un tiempo

Sea U una variable aleatoria con distribución $\mathcal{U}(0,1)$ entonces $F^{-1}(U)$ tiene función de distribución F

La densidad en R

```
tita <- 2
f <- function(x){
  tita^4*x*exp(-tita^2* x) *(x>0)
}
```

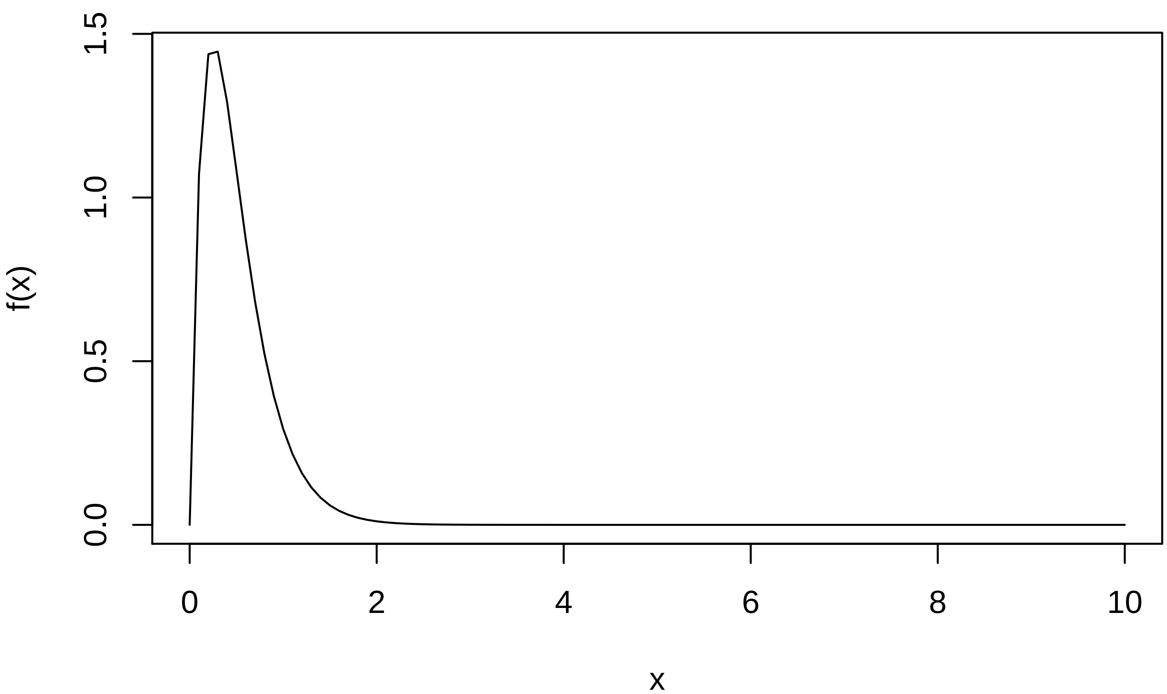
Verifico que integre 1

```
integrate(f,0,10)
```

```
## 1 with absolute error < 6.8e-06
```

La grafico

```
curve(f,0,10)
```

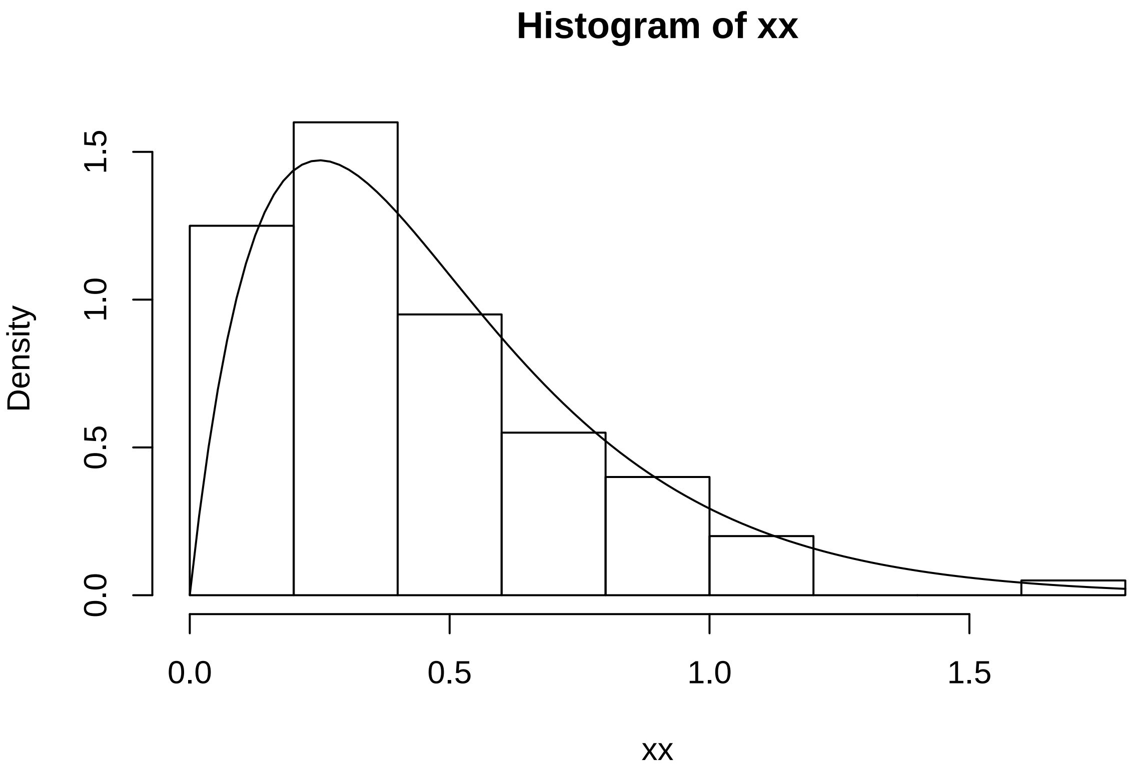


Generamos datos con densidad f

```
generar <- function(n, f, tita){
F <- function(x){
  integrate(f,0,x)$value
}
F <- Vectorize(F)
Finv <- function(x) optimize(function(z)(F(z)-x)^2,c(0,20))$minimum
Finv <- Vectorize(Finv)
uu <- runif(n)
xx <- Finv(uu)
}
```

Histograma de los datos y densidad

```
xx <- generar(100, f, tita = 2)
hist(xx, probability = TRUE)
curve(f, add = TRUE)
```



Calculo EMV con nuestra fórmula

```
est_mv <- sqrt(2/mean(xx))
est_mv
```

```
## [1] 2.140978
```

Calculo EMV con implementación de R

```
f2 <- function(x, tita){tita^4*x*exp(-tita^2* x) *(x>0)}
library(MASS)
est_mv_2 <- fitdistr(xx, densfun=f2, start = list(tita = 1), lower = 0, upper = 10)
est_mv_2
```

```
##       tita
##   2.140978
##  (0.075695)
```

Estimación del ECM

```
tita <- 2
est_mv <- 0
Nrep <- 100
n <- 10
for(i in 1:Nrep){
```

```
xx <- generar(n, f, tita = 2)
est_mv[i] <- sqrt(2/mean(xx))
}
ECM <- mean((est_mv-tita)^2)
ECM
```

Estimación del ECM para varios tamaños de muestra

```
tita <- 2
est_mv <- c()
Nrep <- 100
ECM <- c()
ns <- c(10,50,100,500,1000)
for(j in 1:length(ns)){
  n<-ns[j]
for(i in 1:Nrep){
set.seed(i)
xx <- generar(n, f, tita = 2)
est_mv[i] <- sqrt(2/mean(xx))
}
ECM[j] <- mean((est_mv-tita)^2)
}
plot(ns,ECM)
```

Ejercicio 2

Sean $X_1, \ldots, X_n$ variables aleatorias i.i.d con función de densidad dada por

$$f_X(x, \theta) = \frac{2x}{\theta^2} I_{(0,\theta)}(x) \quad \theta > 0$$

(a) Hallar θ_n el EMV de θ

(b) Decidir si es insesgado.

(c) Hallar el error cuadrático medio de θ_n

(d) Decidir si es consistente.

(e) Repetir los ítems anteriores para el estimador de momentos. ¿Cuál de los dos estimadores prefiere?

Resolución

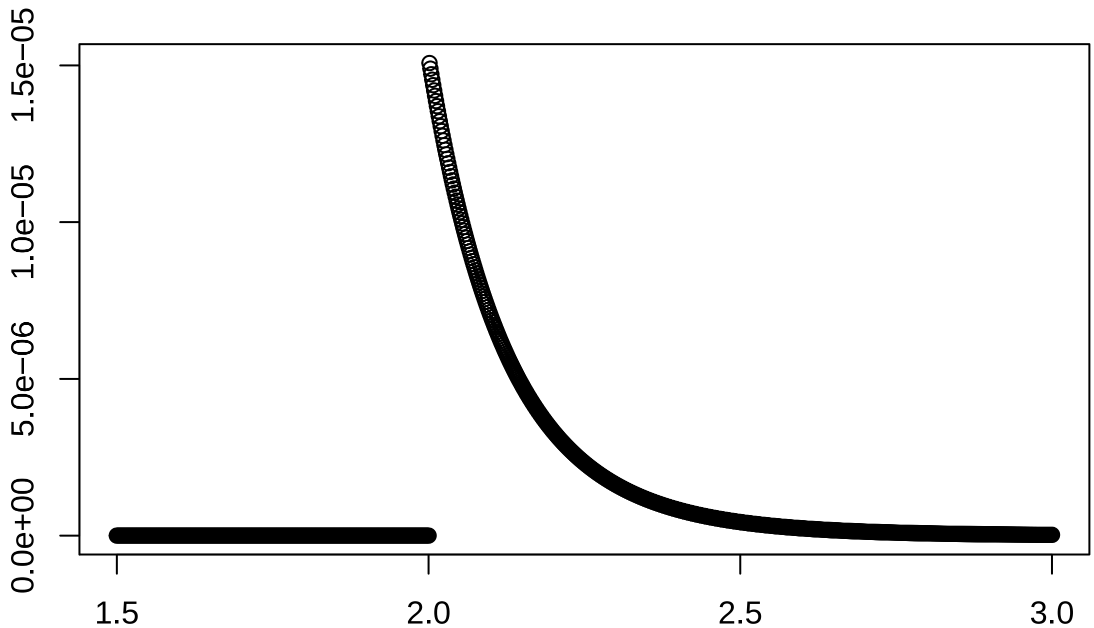
(a) Verosimilitud:

$$L(\theta, x_1, \ldots, x_n) = \Pi_{i=1}^n f(x_i, \theta) = \Pi_{i=1}^n \frac{2x_i}{\theta^2} I_{(0,\theta)}(x_i) = 2^n \Pi_{i=1}^n x_i \theta^{-2n} \Pi_{i=1}^n I_{(0,\theta)}(x_i)$$

$$= \begin{cases} 2^n \Pi_{i=1}^n x_i \theta^{-2n} & \text{si} \quad 0 < x_i < \theta \text{ para todo } i \\ 0 & \text{otro caso} \end{cases}$$

$$= \begin{cases} \text{cte } \theta^{-2n} & \text{si} \quad \theta > \text{ máx } (x_i) \\ 0 & \text{otro caso} \end{cases}$$

```r
n<-8
titas <- c(seq(1.5,3,length.out = 1000))
plot(titas,titas^(-2*n)*(titas>2), xlab="", ylab="")
```



Por lo tanto

$$\hat{\theta} = \max (x_i)$$