# estimación puntual

Carlos Carleos, Norberto Corral

21 de noviembre de 2022

## Índice



## 1. generalidades

### 1.1. definiciones

**estimador** estadígrafo $T = \hat{\theta}$ que toma valores en el espacio paramétrico $\Theta$

**estimación** valor $t = T(\vec{x}) = \hat{\theta}(\vec{x}) = \hat{\theta}$ que toma el estimador dada una muestra concreta $\vec{x}$

### 1.2. propiedades deseables de un estimador

**insesgadez** $E(T) = E(\hat{\theta}) = \theta$

**eficiencia** mínima varianza (cota de Fréchet, Cramér y Rao)

**ECM** error cuadrático medio, $\mathrm{ECM}(T) = E[(T-\theta)^2] = E[(T-E[T])^2] + [E(T)-\theta]^2 = \mathrm{Var}(T) + [\mathrm{sesgo}(T)]^2$

**consistencia** $T_n \xrightarrow[n\to\infty]{P} \theta$

## 2. método de los momentos

### 2.1. propiedades

- el más antiguo
- sencillo
- versátil
- puede dar valores fuera de $\Theta$

### 2.2. descripción

- la distribucíón de $X$ depende de $\vec{\theta} = (\theta_1, \dots, \theta_k)$
- sean $\alpha_1, \dots, \alpha_r$ momentos finitos de la población $X$

$$\forall\, i = 1, \dots, r, \quad \exists\, g_i, \quad \alpha_i = g_i(\vec{\theta})$$

- sean $a_1, \dots, a_n$ momentos muestrales de $\vec{x}$ extraída de $X$
- hállense las $h$ despejando las ecuaciones

$$\left.\begin{array}{c} a_1 \approx \alpha_1 = g_1(\vec{\theta}) \\ \vdots \\ a_r \approx \alpha_r = g_r(\vec{\theta}) \end{array}\right\} \implies \hat{\theta}_i = h_i(a_1, \dots, a_r)$$

### 2.3. ejemplo

- $X \hookrightarrow N(\mu, \sigma) \implies \mu = \alpha_1,\, \sigma^2 = \alpha_2 - \alpha_1^2 \implies \hat{\mu} = \bar{x}, \hat{\sigma}^2 = s^2$
- $X \hookrightarrow U(0, \theta) \implies \alpha_1 = \frac{\theta}{2} \implies \hat{\theta} = 2\bar{x}$ que no es función del estadígrafo suficiente $X_{(n)}$
- $X \hookrightarrow U(-\theta, \theta) \implies \alpha_1 = 0, \alpha^2 = \mathrm{Var}(X) = \frac{\theta^2}{3} \implies \hat{\theta} = \sqrt{3\overline{x^2}}$ que no es función del estadígrafo suficiente $(X_{(1)}, X_{(n)})$

### 2.4. método delta

#### 2.4.1. nombre

deriva del «método $\delta$» de Dorfman 1938

### 2.4.2. teorema

1. enunciado

 - sea $(T_n)_{n\in\mathbb{N}}$ sucesión de estadígrafos muestreados a partir de $X \hookrightarrow F_\theta$ y tales que

$$\frac{T_n-\theta}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{L} N(0,1) \qquad \text{es decir} \qquad T_n \overset{\sim}{\hookrightarrow} N\left(\theta, \frac{\sigma}{\sqrt{n}}\right)$$

 - sea $g$ derivable con $g'(\theta)\neq 0\;\forall\,\theta\in\Theta$
 - entonces

$$\frac{g(T_n)-g(\theta)}{\frac{\sigma}{\sqrt{n}}|g'(\theta)|} \xrightarrow[n\to\infty]{L} N(0,1) \qquad \text{es decir} \qquad g(T_n) \overset{\sim}{\hookrightarrow} N\left(g(\theta), \frac{\sigma}{\sqrt{n}}|g'(\theta)|\right)$$

2. demostración

 - se sabe que $X_n \xrightarrow{L} X \implies cX_n \xrightarrow{L} cX$
 - por tanto, $\frac{T_n-\theta}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{L} N(0,1) \implies (T_n-\theta)\sqrt{n} \xrightarrow[n\to\infty]{L} N(0,\sigma) \implies (T_n-\theta)\sqrt{n}|g'(\theta)| \xrightarrow[n\to\infty]{L} N\left(0,\sigma|g'(\theta)|\right)$
 - $[g(T_n)-g(\theta)]\sqrt{n} = \frac{g(T_n)-g(\theta)}{T_n-\theta}(T_n-\theta)\sqrt{n}$
 - se sabe que

$$\left.\begin{array}{r} X_n \xrightarrow{L} X \\ Y_n \xrightarrow{P} c \end{array}\right\} \implies X_nY_n \xrightarrow{L} cX$$

 - basta probar $Y_n = \frac{g(T_n)-g(\theta)}{T_n-\theta} \xrightarrow{P} g'(\theta)$
 - por definición de derivada

$$\forall\,\epsilon>0, \quad \exists\,\delta_\epsilon>0, \quad |T_n-\theta|<\delta_\epsilon \implies \left|\frac{g(T_n)-g(\theta)}{T_n-\theta}-g'(\theta)\right|<\epsilon$$

 luego

$$\{\omega\in\Omega : |T_n(\omega)-\theta|<\delta_\epsilon\} \subset \left\{\omega\in\Omega : \left|\frac{g(T_n)-g(\theta)}{T_n-\theta}-g'(\theta)\right|<\epsilon\right\}$$

 luego

$$\begin{aligned}
\lim_{n\to\infty}\Pr\left[\left|\frac{g(T_n)-g(\theta)}{T_n-\theta}-g'(\theta)\right|<\epsilon\right] &\geq \lim\Pr[|T_n-\theta|<\delta_\epsilon] \\
&= \lim\Pr\left[\frac{\sqrt{n}|T_n-\theta|}{\sigma}<\frac{\sqrt{n}\delta_\epsilon}{\sigma}\right] \\
&\geqslant \lim\Pr\left[\frac{\sqrt{n}|T_n-\theta|}{\sigma}<\frac{\sqrt{n_0}\delta_\epsilon}{\sigma}\right] \\
&= \Pr\left[|N(0,1)|<\frac{\sqrt{n_0}\delta_\epsilon}{\sigma}\right] \\
&> 1-\epsilon
\end{aligned}$$

 tomando $n_0$ suficientemente grande

 - por tanto

$$\forall\,\epsilon>0 \qquad \lim_{n\to\infty}\Pr\left[\left|\frac{g(T_n)-g(\theta)}{T_n-\theta}-g'(\theta)\right|<\epsilon\right]>1-\epsilon$$

 luego

$$\lim_{n\to\infty}\Pr\left[\left|\frac{g(T_n)-g(\theta)}{T_n-\theta}-g'(\theta)\right|<\epsilon\right]=1$$

 o sea

$$\frac{g(T_n)-g(\theta)}{T_n-\theta} \xrightarrow{P} g'(\theta)$$

### 2.4.3. ejemplo

- $X \hookrightarrow \mathrm{Exp}(\lambda)$
- $\mathrm{E}(X) = \frac{1}{\lambda} \implies \hat{\lambda}_{\mathrm{MM}} = \frac{1}{\bar{X}}$
- por TCL, $T = \bar{X} \overset{\sim}{\hookrightarrow} N\left(\theta, \frac{\sigma}{\sqrt{n}}\right)$ con $\theta = \frac{1}{\lambda}$ y $\sigma = \frac{1}{\lambda}$
- sea $g(x) = \frac{1}{x}$ con $g'(x) = -\frac{1}{x^2}$
- entonces $\hat{\lambda}_{\mathrm{MM}} = g(T) \overset{\sim}{\hookrightarrow} N\left(g(\theta), \frac{\sigma}{\sqrt{n}}|g'(\theta)|\right) = N\left(g\left(\frac{1}{\lambda}\right), \frac{\sigma}{\sqrt{n}}\left|g'\left(\frac{1}{\lambda}\right)\right|\right) = N\left(\lambda, \frac{1/\lambda}{\sqrt{n}}\left|-\frac{1}{\left(\frac{1}{\lambda}\right)^2}\right|\right) = N\left(\lambda, \frac{1/\lambda}{\sqrt{n}}\lambda^2\right)$
  $= N\left(\lambda, \frac{\lambda}{\sqrt{n}}\right)$

```
## X = Exp(landa)
landa <- 2                              # parAmetro
n <- 100                                # tamaNo muestral
m <- 1e6                                # tamaNo montecarlo
zita <- 1/landa                         # E(X)
sigma <- 1/landa                        # DT(X)
T <- mean                               # estadIgrafo T=N(zita,sigma/n)
g <- function (x) 1/x                   # estimador g(T) = N (landa, landa/raIz(n))
gprima <- function (x) -1/x^2           # g(T) = N (g(zita), sigma/n*abs(gprima(zita)))
gprima <- function (t) eval (D(expression(1/x),"x"), list(x=t)) # otra forma
d <- replicate (m, g(T(rexp(n,landa)))) # distribuciOn de g(T)
c (landa, mean(d))                      # esperanza y media de montacarlo
c (landa/sqrt(n), sigma/n*abs(gprima(zita)), sd(d)) # desvIos tIpicos
plot (ecdf (d))                                     # ojiva empIrica
plot (function (x) pnorm (x, landa, landa/sqrt(n)), # ojiva asintOtica
      min(d), max(d), col=2, add=TRUE)              # por mEtodo delta
dev.new ()                              # nueva ventana grAfica
plot (density (d))                      # densidad estimada mediante "nUcleos"
plot (function (x) dnorm (x, landa, landa/sqrt(n)), # densidad asintOtica
      min(d), max(d), col=2, add=TRUE)              # por mEtodo delta
```

# 3. método de máxima verosimilitud

## 3.1. intro

### 3.1.1. ejemplo

- dos máquinas, A y B, fabrican tornillos
- la máquina A produce un 1 % de tornillos defectuosos
- la máquina B produce un 8 % de tornillos defectuosos
- una caja contiene diez tornillos, dos de ellos defectuosos, fabricados todos por la misma máquinas
- ¿qué máquina los fabricó?
- sea $X|i =$ "número de tornillos defectuosos entre diez tornillos fabricados en la máquina $i$"
- entonces $X|\mathrm{A} \hookrightarrow B(10, \frac{1}{100})$ y $X|\mathrm{B} \hookrightarrow B(10, \frac{8}{100})$
- $\Pr(X = 2 \mid \mathrm{A}) = \texttt{dbinom}(2, 10, ,01) = 0{,}004152351 \ll 0{,}147807 = \texttt{dbinom}(2, 10, ,08) = \Pr(X = 2 \mid \mathrm{B})$

- la muestra de la caja es mucho más verosímil si viene de B que si viene de A
- OJO: no estamos calculando $\Pr(\mathrm{A} \mid X = 2)$ ni $\Pr(\mathrm{B} \mid X = 2)$; para eso necesitaríamos $\Pr \mathrm{A}$ y $\Pr \mathrm{B}$ y aplicar la fórmula de Bayes

### 3.1.2. definiciones

**verosimilitud** función de verosimilitud (*likelihood* en inglés) asociada a la muestra $\vec{x}$ de la población $X \hookrightarrow F_\theta$

$$L(\vec{x}, \theta) = L_{\vec{x}}(\theta) = f(\vec{x}, \theta)$$

es función de $\theta$, pues $\vec{x}$ es fija

**EMV** estimador máximo-verosímil de $\theta$

$$\hat{\theta}_{\mathrm{MV}} = \arg\sup_{\theta \in \Theta} L(\vec{x}, \theta)$$

luego

$$L(\vec{x}, \hat{\theta}_{\mathrm{MV}}) \geqslant L(\vec{x}, \theta) \quad \forall\, \theta \in \Theta$$

**logverosimilitud** logaritmo de la verosimilitud, $\log L$ o $\ln L$; a menudo es más sencillo y robusto maximizar la logverosimilitud $\ln L$ que directamente $L$

**informante** si la logverosimilitud es derivable, se define el informante (*score*)

$$u(\theta) = \frac{\partial}{\partial \theta} \ln L(\vec{x}, \theta)$$

**ecuación de verosimilitud** a menudo el EMV está entre las soluciones de

$$u(\theta) = 0$$

puede haber varios EMV

### 3.1.3. ejemplo (exponencial)

- $X =$ tiempo de "vida" de cierto tipo de aparato
- $X \hookrightarrow \mathrm{Exp}(\theta)$

1. información completa
   - calcula el EMV de $\theta$ si se dispone de la duración de $n$ aparatos, es decir, una muestra $\vec{x}$ de $X$

   $f(x,\theta) = \theta e^{-\theta x} \implies L(\vec{x}, \theta) = \theta^n e^{-\theta \sum x_i} \implies \ln L(\vec{x}, \theta) = n \ln \theta - \theta \sum x_i \implies 0 = u(\theta) = \frac{\partial}{\partial \theta} \ln L(\vec{x}, \theta) = \frac{n}{\theta} - \sum x_i \implies \frac{n}{\theta} = \sum x_i \implies \hat{\theta}_{\mathrm{MV}} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$

2. información censurada
   - calcula el EMV de $\theta$ si se dispone de la duración de $n - r$ aparatos; de los otros $r$ se sabe sólo que han durado un tiempo mayor que $t_0$ (por ejemplo, por un límite en la duración del experimento)

   $f(x,\theta) = \theta e^{-\theta x}, \Pr(X > t_0) = 1 - F(t_0) = e^{-\theta t_0} \implies L(\vec{x}, \theta) = e^{-r\theta t_0} \times \theta^{n-r} e^{-\theta \sum_{i=1}^{n-r} x_{(i)}} = \theta^{n-r} e^{-\theta (\sum_{i=1}^{n-r} x_{(i)} + r t_0)} \implies \ln L(\vec{x}, \theta) = (n-r) \ln \theta - \theta (\sum_{i=1}^{n-r} x_{(i)} + r t_0) \implies 0 = \frac{n-r}{\theta} - (\sum_{i=1}^{n-r} x_{(i)} + r t_0) \implies \frac{n-r}{\theta} = \sum_{i=1}^{n-r} x_{(i)} + r t_0 \implies \hat{\theta}_{\mathrm{MV}} = \frac{n-r}{\sum_{i=1}^{n-r} x_{(i)} + r t_0}$

### 3.1.4. ejemplo (bernuli)

$X \hookrightarrow B(1,p) \implies L(\vec{x},p) = p^{\sum x_i}(1-p)^{n-\sum x_i} \implies \ln L(\vec{x},p) = \sum x_i \log p + (n - \sum x_i)\ln(1-p) \implies 0 = \frac{\partial}{\partial p}\ln L(\vec{x},p) = \frac{\sum x_i}{p} + \frac{n-\sum x_i}{1-p} \implies \frac{\sum x_i}{p} = \frac{n-\sum x_i}{1-p} \implies (1-p)\sum x_i = p(n-\sum x_i) \implies \hat{p}_{\mathrm{MV}} = \frac{\sum x_i}{n} = \bar{x}$
es decir, la proporción de unos

### 3.1.5. ejemplo con verosimilitud no derivable

$X \hookrightarrow U(0,\theta) \implies L(\vec{x},\theta) = \frac{1}{\theta^n} I(x_{(n)} < \theta) = \begin{cases} 0 & \theta < x_{(n)} \\ \frac{1}{\theta^n} & \theta > x_{(n)} \end{cases}$ (no derivable) $\implies$ el supremo se alcanza en $\theta = x_{(n)} \implies \hat{\theta}_{\mathrm{MV}} = x_{(n)}$

### 3.1.6. ejemplo EMV no único

$X \hookrightarrow U(\theta,\theta+3) \implies f(\vec{x},\theta) = \frac{1}{3} I(\theta < x < \theta+3) \implies L(\vec{x},\theta) = \frac{1}{3^n} I(x_{(n)} - 3 < \theta < x_{(1)}) \implies$ cualquier valor entre $X_{(n)} - 3$ y $X_{(1)}$ es EMV, por ejemplo $T = \frac{X_{(n)}-3+X_{(1)}}{2}$

### 3.1.7. ejemplo sin solución explícita

$Y \hookrightarrow P(\lambda),\ X = \begin{cases} 1 & Y \leqslant 1 \\ 2 & Y = 2 \\ 3 & Y \geqslant 3 \end{cases},\ n_i = \sum_{j=1}^n I(X_j = i)\,\forall i = 1,2,3$

$\implies L(\vec{x},\lambda) = (e^{-\lambda}[1+\lambda])^{n_1} \left(e^{-\lambda}\frac{\lambda^2}{2}\right)^{n_2} \left(1 - e^{-\lambda}\left[1+\lambda+\frac{\lambda^2}{2}\right]\right)^{n_3}$

No se puede encontrar una solución explícita de la ecuación de verosimilitud:

```
(%i17) L(n1,n2,n3,landa) :=
          (exp(-landa)*(1+landa))^n1 *
          (exp(-landa)*landa^2/2)^n2 *
          (1-exp(-landa)*(1+landa+landa^2/2))^n3 $

(%i18) tex (diff (L(n1,n2,n3,landa), landa)) $
```

$$\frac{(\lambda+1)^{n_1}\left(1-\left(\frac{\lambda^2}{2}+\lambda+1\right)e^{-\lambda}\right)^{n_3-1}\left(\left(\frac{\lambda^2}{2}+\lambda+1\right)e^{-\lambda}-(\lambda+1)\,e^{-\lambda}\right)|\lambda|^{2\,n_2}\,e^{-\lambda\,n_2-\lambda\,n_1}\,n_3}{2^{n_2}}$$
$$+$$
$$\frac{(\lambda+1)^{n_1}\left(1-\left(\frac{\lambda^2}{2}+\lambda+1\right)e^{-\lambda}\right)^{n_3}|\lambda|^{2\,n_2}\,(-n_2-n_1)\,e^{-\lambda\,n_2-\lambda\,n_1}}{2^{n_2}}$$
$$+$$
$$\frac{(\lambda+1)^{n_1-1}\left(1-\left(\frac{\lambda^2}{2}+\lambda+1\right)e^{-\lambda}\right)^{n_3}|\lambda|^{2\,n_2}\,n_1\,e^{-\lambda\,n_2-\lambda\,n_1}}{2^{n_2}}$$
$$+$$
$$\frac{(\lambda+1)^{n_1}\left(1-\left(\frac{\lambda^2}{2}+\lambda+1\right)e^{-\lambda}\right)^{n_3}|\lambda|^{2\,n_2}\,n_2\,2^{1-n_2}\,e^{-\lambda\,n_2-\lambda\,n_1}}{\lambda}$$

```
(%i23) logexpand:all $

(%i24) define (lnL(landa), log(L(n1,n2,n3,landa))) $

(%i25) define (u(landa), diff(lnL(landa),landa)) $ /* informante */

(%i26) freeof (landa, rhs (solve (u(landa), landa) [1])) ; /* no puede despejar landa */

(%o26) false

(%i27) tex (u(landa)) $
```

$u(\lambda) = \frac{\left(\left(\frac{\lambda^2}{2}+\lambda+1\right)e^{-\lambda}-(\lambda+1)\,e^{-\lambda}\right)n_3}{1-\left(\frac{\lambda^2}{2}+\lambda+1\right)e^{-\lambda}} + \frac{2\,n_2}{\lambda} - n_2 + \frac{n_1}{\lambda+1} - n_1$

Hay que recurrir a métodos numéricos:

```
> L <- function(n1,n2,n3,landa) (exp(-landa)*(1+landa))^n1 *
                                (exp(-landa)*landa^2/2)^n2 *
                                (1-exp(-landa)*(1+landa+landa^2/2))^n3
> optimize (function (landa) L(3,3,4,landa), c(0,10), maximum=TRUE)
$maximum
[1] 2.342104

$objective
[1] 1.803925e-05
```

En este caso, la maximización directa de la verosimilitud no da problemas. Sin embargo, por robustez numérica (para que el producto de probabilidades no dé un cero numérico y la función quede plana) a menudo es recomendable maximizar la logverosimilitud:

```
> optimize (function (landa) log(L(3,3,4,landa)), c(0,10), maximum=TRUE)
$maximum
[1] 2.342112

$objective
[1] -10.92296
```

### 3.1.8. ejemplo gausiano

$X \hookrightarrow N(\mu,\sigma) \implies L(\vec{x},\mu,\sigma) = \frac{1}{\sigma^n\sqrt{2\pi}^n} e^{\frac{-1}{2\sigma^2}\sum(x_i-\mu)^2} \implies \ln L(\vec{x},\mu,\sigma) = -n\ln\sigma - n\ln\sqrt{2\pi} - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2 \implies$

$$\begin{aligned} 0 = \frac{\ln L(\vec{x},\mu,\sigma)}{\partial\mu} &= -\frac{1}{2\sigma^2}\sum(x_i-\mu)2(-1) \implies \sum x_i = n\mu \\ 0 = \frac{\ln L(\vec{x},\mu,\sigma)}{\partial\sigma} &= \frac{-n}{\sigma} - \frac{1}{2\sigma^{-3}}(-2)\sum(x_i-\mu)^2 \implies \sigma^2 n = \sum(x_i-\mu)^2 \end{aligned}$$

$\implies \hat{\mu}_{\text{MV}} = \frac{\sum x_i}{n} = \bar{x},\ \hat{\sigma}^2_{\text{MV}} = \frac{\sum(x_i-\bar{x})^2}{n} = s^2$

### 3.1.9. ejemplos triangular

1. beta escalada, densa cabe $\theta$

   - $X \hookrightarrow \theta\cdot\beta(2,1) \implies f = \frac{2x}{\theta^2}I(x<\theta) \implies L = \frac{2^n\prod x_i}{\theta^{2n}}I(x_{(n)}<\theta) \implies \forall\theta > x_{(n)}, \ln L = n\ln 2 + \sum\ln x_i - 2n\ln\theta \implies \frac{\partial\ln L}{\partial\theta} = -\frac{2n}{\theta} < 0 \implies L$ es decreciente
   - EMV $\hat{\theta} = X_{(n)}$

2. beta escalada, densa cabe 0

   - $X \hookrightarrow \theta\cdot\beta(1,2) \implies f = \frac{2(1-\frac{x}{\theta})}{\theta}I(x<\theta) \implies L = \frac{2^n\prod(1-\frac{x_i}{\theta})}{\theta^n}I(x_{(n)}<\theta) \implies \forall\theta > x_{(n)}, \ln L = n\ln 2 + \sum\ln\left(1-\frac{x_i}{\theta}\right) - n\ln\theta \implies \frac{\partial\ln L}{\partial\theta} = \sum\frac{-\frac{x_i}{\theta^2}}{1-\frac{x_i}{\theta}} - \frac{n}{\theta} = \sum\frac{x_i}{\theta x_i-\theta^2} - \frac{n}{\theta} = \implies \left\{ 0 = \frac{\partial\ln L}{\partial\theta} \implies \sum\frac{x_i}{x_i-\theta} = n \right\}$
   - EMV debe obtenerse numéricamente

## 3.2. propiedades

### 3.2.1. EMV función del suficiente

- sea $T$ estadígrafo suficiente
- sea $\hat{\theta}$ el único EMV
- entonces $\hat{\theta}$ es función de $T$
- demostración: $T$ suficiente $\implies L(\vec{x},\theta) = g(T[\vec{x}],\theta)h(\vec{x}) \implies \max_{\theta\in\Theta} L(\vec{x},\theta) = h(\vec{x}) \max_{\theta\in\Theta} g(T[\vec{x}],\theta)$ $\implies \hat{\theta}_{\mathrm{MV}} = \arg\max_{\theta\in\Theta} L(\vec{x},\theta) = \arg\max_{\theta\in\Theta} g(T[\vec{x}],\theta)$
- si el EMV no es único, alguno puede no depender del suficiente:
  - $X \hookrightarrow U(\theta,\theta+3) \implies \forall\lambda \in [0,1],\ \lambda(X_{(n)}-3)+(1-\lambda)X_{(1)}$ es EMV
  - $\lambda$ puede ser aleatoria, por ejemplo $\lambda = \frac{X_1^2}{X_1^2+X_2^2}$, y entonces ese EMV no es función sólo del estadígrafo suficiente $(X_{(1)}, X_{(n)})$
- el EMV puede no ser suficiente; ejemplo:
  - $X \hookrightarrow U(\theta,2\theta) \implies L(\vec{x},\theta) = \frac{1}{\theta^n} I\left(\frac{x_{(n)}}{2} < \theta < x_{(1)}\right) \implies \begin{cases} (X_{(1)}, X_{(n)}) \text{ es suficiente} \\ \hat{\theta}_{\mathrm{MV}} = \frac{X_{(n)}}{2} \end{cases}$
  - $\hat{\theta}_{\mathrm{MV}}$ no es suficiente

### 3.2.2. equivarianza o invariancia funcional

- $\hat{\theta}$ es EMV de $\theta \implies h(\hat{\theta})$ es EMV de $h(\theta)$
  - $h$ biyectiva
    - sea $\eta = h(\theta)$
    - $\hat{\eta}_{\mathrm{MV}} = \arg\sup_\eta L(\vec{x}, h^{-1}[\eta])$
    - $\hat{\theta}_{\mathrm{MV}} = \arg\sup_\theta L(\vec{x},\theta) \implies h^{-1}(\eta) = \hat{\theta}_{\mathrm{MV}}$ maximiza $L(\vec{x}, h^{-1}[\eta]) \implies h^{-1}(\hat{\eta}_{\mathrm{MV}}) = \hat{\theta}_{\mathrm{MV}} \implies$ $\hat{\eta}_{\mathrm{MV}} = h(\hat{\theta}_{\mathrm{MV}})$
  - $h$ cualquiera
    - $\hat{\eta}_{\mathrm{MV}} = h(\hat{\theta})$ maximiza la verosimilitud perfilada

$$\mathcal{L}(\vec{x},\eta) = \sup\left\{L(\vec{x},\theta) \mid \theta \in h^{-1}(\eta)\right\}$$

- por ejemplo, $X \hookrightarrow N(\mu,\sigma) \implies \widehat{\sigma^2}_{\mathrm{MV}} = s^2 \implies \hat{\sigma}_{\mathrm{MV}} = s$

### 3.2.3. comportamiento asintótico

1. condiciones de regularidad
   a) $\theta \neq \theta' \implies F_\theta \neq F_{\theta'}$; si no, no podrían distinguirse y el estimador no podría ser consistente
   b) el soporte $\{x : f(x,\theta) > 0\}$ no cambia (no depende de $\theta$)
   c) $\vec{X}$ es una muestra aleatoria simple de $X \hookrightarrow F_{\theta_0}$
   d) el espacio paramétrico $\Theta$ es un intervalo abierto
   e) $\forall x, f(x,\theta)$ es derivable dos veces respecto a $\theta$
   f) las integrales $\int \left|\frac{\partial^i f(x,\theta)}{\partial\theta^i}\right| dx$ son finitas para $i = 1, 2$
2. teorema de consistencia

- si se cumplen las condiciones de regularidad, entonces existe una sucesión de raíces de la ecuación de verosimilitud, $\hat{\theta}_n$, fuertemente consistente, es decir, tal que $\hat{\theta}_n \xrightarrow{\text{c.s.}} \theta_0$

3. teorema de eficiencia

   *a*) enunciado: Si
      - se cumplen las condiciones de regularidad
      - existe $\frac{\partial^3 \ln f(x,\theta)}{\partial \theta^3}$ con $\left|\frac{\partial^3 \ln f(x,\theta)}{\partial \theta^3}\right| < K(x)$ siendo $E_\theta[K(X)] < \infty$

      entonces
      - existe una sucesión de raíces de la ecuación de verosimilitud, $\hat{\theta}_n$, consistente y asintóticamente gausiana, es decir, $(\hat{\theta}_n - \theta_0)\sqrt{nI(\theta_0)} \xrightarrow{\text{L}} N(0,1)$
      - equivalentemente, con $n$ grande $\hat{\theta}_n \overset{\sim}{\hookrightarrow} N\left(\theta_0, \frac{1}{\sqrt{nI(\theta_0)}}\right)$

      donde $I(\theta) = \text{Var}\left(\frac{\partial \ln f(X,\theta)}{\partial \theta}\right) = E\left[\left(\frac{\partial \ln f(X,\theta)}{\partial \theta}\right)^2\right] = -E\left(\frac{\partial^2 \ln f(X,\theta)}{\partial \theta^2}\right)$ es la cantidad información de Físher que $X$ contiene sobre $\theta$.

   *b*) demostración:
      - $u(\theta) = \frac{\partial \ln L(\vec{x},\theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \ln f(x_i,\theta)}{\partial \theta}$
      - $E[u(\theta)] = \sum_{i=1}^n E\left(\frac{\partial \ln f(X_i,\theta)}{\partial \theta}\right)$
      - $E\left(\frac{\partial \ln f(X,\theta)}{\partial \theta}\right) = \int \frac{\partial \ln f(x,\theta)}{\partial \theta} f(x,\theta)dx = \int \frac{1}{f(x,\theta)}\frac{\partial f(x,\theta)}{\partial \theta} f(x,\theta)dx = \int \frac{\partial f(x,\theta)}{\partial \theta}dx = \frac{\partial \int f(x,\theta)dx}{\partial \theta} = \frac{\partial 1}{\partial \theta} = 0$
      - $\text{Var}[u(\theta)] = \text{Var}\left(\sum_{i=1}^n \frac{\partial \ln f(X_i,\theta)}{\partial \theta}\right) = \sum_{i=1}^n \text{Var}\left(\frac{\partial \ln f(X_i,\theta)}{\partial \theta}\right) = \sum_{i=1}^n E\left[\left(\frac{\partial \ln f(X_i,\theta)}{\partial \theta}\right)^2\right]$
      - $E\left[\left(\frac{\partial \ln f(X,\theta)}{\partial \theta}\right)^2\right] = \int \left(\frac{\partial \ln f(x,\theta)}{\partial \theta}\right)^2 f(x,\theta)dx = \int \left(\frac{1}{f(x,\theta)}\frac{\partial f(x,\theta)}{\partial \theta}\right)^2 f(x,\theta)dx =$
        $\int \frac{1}{f(x,\theta)^2}\left(\frac{\partial f(x,\theta)}{\partial \theta}\right)^2 f(x,\theta)dx = \int \frac{1}{f(x,\theta)}\left(\frac{\partial f(x,\theta)}{\partial \theta}\right)^2 dx$
      - $E\left(\frac{\partial^2 \ln f(X,\theta)}{\partial \theta^2}\right) = \int \frac{\partial^2 \ln f(x,\theta)}{\partial \theta^2} f(x,\theta)dx = \int \frac{\partial}{\partial \theta}\left(\frac{\partial \ln f(x,\theta)}{\partial \theta}\right) f(x,\theta)dx =$
        $\int \frac{\partial}{\partial \theta}\left(\frac{1}{f(x,\theta)}\frac{\partial f(x,\theta)}{\partial \theta}\right) f(x,\theta)dx = \int \frac{\frac{\partial^2 f(x,\theta)}{\partial \theta^2}f(x,\theta) - \left(\frac{\partial f(x,\theta)}{\partial \theta}\right)^2}{f(x,\theta)^2} f(x,\theta)dx = \int \frac{\frac{\partial^2 f(x,\theta)}{\partial \theta^2}f(x,\theta) - \left(\frac{\partial f(x,\theta)}{\partial \theta}\right)^2}{f(x,\theta)}dx$
        $= \int \frac{\partial^2 f(x,\theta)}{\partial \theta^2}dx - \int \frac{\left(\frac{\partial f(x,\theta)}{\partial \theta}\right)^2}{f(x,\theta)}dx = \frac{\partial^2 \int f(x,\theta)dx}{\partial \theta^2} - \int \frac{\left(\frac{\partial f(x,\theta)}{\partial \theta}\right)^2}{f(x,\theta)}dx = \frac{\partial^2 1}{\partial \theta^2} - \int \frac{\left(\frac{\partial f(x,\theta)}{\partial \theta}\right)^2}{f(x,\theta)}dx =$
        $0 - \int \frac{\left(\frac{\partial f(x,\theta)}{\partial \theta}\right)^2}{f(x,\theta)}dx = -E\left[\left(\frac{\partial \ln f(X,\theta)}{\partial \theta}\right)^2\right]$
      - luego $u(\theta)$ tiene esperanza y varianza finitas
      - sea $\hat{\theta}_n$ una raíz de la ecuación de verosimilitud, $0 = u(\hat{\theta}_n)$
      - desarrollo de Taylor de $u(\hat{\theta}_n)$ en torno a $\theta_0$: $0 = u(\hat{\theta}_n) = u(\theta_0) + (\hat{\theta}_n - \theta_0)u'(\theta_0) + (\hat{\theta}_n - \theta_0)^2\frac{u''(\tilde{\theta})}{2}$ con $\tilde{\theta}$ tal que $|\tilde{\theta} - \theta_0| < |\hat{\theta}_n - \theta_0|$
      - $0 = u(\theta_0) + (\hat{\theta}_n - \theta_0)\left[u'(\theta_0) + (\hat{\theta}_n - \theta_0)\frac{u''(\tilde{\theta})}{2}\right] \implies \hat{\theta}_n - \theta_0 = \frac{-u(\theta_0)}{u'(\theta_0) + (\hat{\theta}_n - \theta_0)\frac{u''(\tilde{\theta})}{2}} \implies \sqrt{n}(\hat{\theta}_n - \theta_0)$
        $= \frac{-\sqrt{n}u(\theta_0)}{u'(\theta_0) + (\hat{\theta}_n - \theta_0)\frac{u''(\tilde{\theta})}{2}} = \frac{\sqrt{n}\frac{-u(\theta_0)}{n}}{\frac{u'(\theta_0)}{n} + (\hat{\theta}_n - \theta_0)\frac{u''(\tilde{\theta})}{2n}}$
      - numerador $\sqrt{n}\frac{-u(\theta_0)}{n}$: por el TCL, como $\frac{\partial \ln f(X_i,\theta)}{\partial \theta}$ son i.i.d. con esperanza 0 y varianza $I(\theta)$, se tiene $-\sqrt{n}\frac{u(\theta_0)}{n} = -\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \left.\frac{\partial \ln f(X_i,\theta)}{\partial \theta}\right|_{\theta=\theta_0}\right) \xrightarrow{L} N\left(0, \sqrt{I(\theta_0)}\right)$
      - denominador converge en probabilidad a $I(\theta_0)$:
        - por la LFGN $\frac{u'(\theta_0)}{n} = \frac{1}{n}\sum_{i=1}^n \left.\frac{\partial^2 \ln f(X_i,\theta)}{\partial \theta^2}\right|_{\theta=\theta_0} \xrightarrow{\text{c.s.}} -I(\theta_0)$

- se sabe que existe sucesión de soluciones de la ecuación de verosimilitud tal que $\hat{\theta}_n - \theta_0 \xrightarrow{P} 0$
- $\frac{u''(\tilde{\theta})}{n} = \frac{1}{n}\sum_{i=1}^{n} \left.\frac{\partial^3 \ln f(X_i,\theta)}{\partial\theta^3}\right|_{\theta=\tilde{\theta}} \implies \left|\frac{u''(\tilde{\theta})}{n}\right| \leqslant \frac{1}{n}\sum_{i=1}^{n} \left|\left.\frac{\partial^3 \ln f(X_i,\theta)}{\partial\theta^3}\right|_{\theta=\tilde{\theta}}\right| \leqslant \frac{1}{n}\sum_{i=1}^{n} K(X_i)$ $\xrightarrow{\text{c.s.}} E[K(X)] = k < \infty \implies$ con $n$ grande, $\left|\frac{1}{n}\sum_{i=1}^{n} K(X_i) - k\right| < \varepsilon \implies$ con $n$ grande, $\frac{1}{n}\sum_{i=1}^{n} K(X_i) < k+\varepsilon \implies$ con $n$ grande, $\left|\frac{u''(\tilde{\theta})}{n}\right| < k+\varepsilon$
- $(\hat{\theta}_n - \theta_0)\frac{u''(\tilde{\theta})}{2n} \xrightarrow{P} 0$
- $\frac{u'(\theta_0)}{n} + (\hat{\theta}_n - \theta_0)\frac{u''(\tilde{\theta})}{2n} \xrightarrow{P} -I(\theta_0)$

- $I(\theta_0) > 0 \implies \frac{1}{\frac{u'(\theta_0)}{n} + (\hat{\theta}_n - \theta_0)\frac{u''(\tilde{\theta})}{2n}} \xrightarrow{P} \frac{-1}{I(\theta_0)}$
- $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L} N\left(0, \frac{1}{\sqrt{I(\theta_0)}}\right)$

*c*) corolario: En las condiciones del teorema anterior, si la ecuación de verosimilitud tiene una única raíz, ésta es el EMV y es consistente, asintóticamente eficiente y asintónticamente gausiano.

*d*) ejemplo

- $X \hookrightarrow \text{Exp}(\theta)$
- el EMV es $\hat{\theta}_{\text{MV}} = \frac{1}{\bar{X}}$
- la información de Física es $\frac{-\partial^2 \ln f(x,\theta)}{\partial\theta^2} = \frac{-\partial^2 \ln(\theta e^{-\theta x})}{\partial\theta^2} = \frac{1}{\theta^2}$
- se verifican las condiciones de regularidad de F-C-R
- además, $\frac{\partial^3 f(x,\theta)}{\partial\theta^3} = \frac{\partial^3 \ln(\theta e^{-\theta x})}{\partial\theta^3} = \frac{2}{\theta^3}$
- considerando que el espacio paramétrico es $\Theta = (\epsilon, \infty)$, entonces $\theta > \epsilon$ y $\frac{\partial^3 f(x,\theta)}{\partial\theta^3} = \frac{2}{\theta^3} < \frac{2}{\epsilon^3}$
- por el teorema, $\sqrt{n}\left(\frac{1}{\bar{X}} - \theta\right) \xrightarrow{L} N(0,\theta)$
- aunque la acotación de la tercera derivada no se verifica para cualquier $\theta > 0$, basta suponer en la práctica el $\epsilon$ muy pequeño, o emplear directamente el método delta con $g(x) = \frac{1}{x}$:
  - por el TCL, $\sqrt{n}(\bar{X} - \mu) \xrightarrow{L} N(0,\mu)$ con $\mu = \frac{1}{\theta}$
  - luego $\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right) = \sqrt{n}\left(\frac{1}{\bar{X}} - \theta\right) \xrightarrow{L} N(0,\theta)$

*e*) ejemplo

- $X \hookrightarrow U(0,\theta)$
- EMV $\hat{\theta}_n = X_{(n)}$ cuya distribución es

$$F_{X_{(n)}}(x) = \Pr(X_{(n)} \leqslant x) = \begin{cases} 0 & x \leqslant 0 \\ \left(\frac{x}{\theta}\right)^n & 0 < x < \theta \\ 1 & x \geqslant \theta \end{cases}$$

- $E(X_{(n)}) = \frac{n}{n+1}\theta \to \theta$
- $\text{Var}(X_{(n)}) = \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2 \to 0$
- luego EMV consistente pero $\sqrt{n}(X_{(n)} - \theta)$ no converge a una ley gausiana pues
  - su varianza tiende a cero: $\text{Var}[\sqrt{n}(X_{(n)} - \theta)] = \left(\frac{n^2}{n+2} - \frac{n^3}{(n+1)^2}\right)\theta^2 = \frac{n^2\,\theta^2}{n^3+4\,n^2+5\,n+2} \to 0$
  - luego $\sqrt{n}(X_{(n)} - \theta)$ tiende a una degenerada en cero, $\sqrt{n}(X_{(n)} - \theta) \xrightarrow{\Pr} 0$:
    - directamente: $\forall \quad \epsilon \quad > \quad 0$, $\Pr\left[\left|\sqrt{n}(X_{(n)} - \theta)\right| < \epsilon\right] = \Pr\left[\sqrt{n}(\theta - X_{(n)}) < \epsilon\right] =$ $\Pr\left[X_{(n)} > \theta - \frac{\epsilon}{\sqrt{n}}\right] = 1 - F_{X_{(n)}}\left(\theta - \frac{\epsilon}{\sqrt{n}}\right) = 1 - \left[\frac{\theta - \frac{\epsilon}{\sqrt{n}}}{\theta}\right]^n = 1 - \left[1 - \frac{\epsilon}{\theta\sqrt{n}}\right]^n =$

$$1-\Bigg(\underbrace{\left[1-\frac{\epsilon}{\theta\sqrt{n}}\right]^{\sqrt{n}}}_{\xrightarrow{n} e^{-\epsilon/\theta}<1}\Bigg)^{\sqrt{n}} \overset{\substack{\exists\ \delta, n_0\\ \forall\ n>n_0}}{\geqslant} 1-\underbrace{\Big(\underbrace{e^{-\epsilon/\theta}+\delta}_{<1}\Big)^{\sqrt{n}}}_{\xrightarrow{n} 0} \xrightarrow{n} 1$$

- ∘ otramente, mediante la desigualdad de Márkov: sea $T = \sqrt{n}(X_{(n)}-\theta)$; $\Pr\left[|T|\leqslant\epsilon\right] = \Pr\left[T^2\leqslant\epsilon^2\right] \geqslant 1-\frac{E\left[T^2\right]}{\epsilon^2} = 1-\frac{\mathrm{Var}[T]+E^2[T]}{\epsilon^2} = 1-\frac{\mathrm{Var}[T]+\left(\sqrt{n}\frac{-1}{n+1}\theta\right)^2}{\epsilon^2} \xrightarrow{n} 1-0=1$

- multiplicando por $n$ en vez de por $\sqrt{n}$ se consigue que la varianza no se anule: $\mathrm{Var}[n(X_{(n)}-\theta)] = \left(\frac{n^3}{n+2}-\frac{n^4}{(n+1)^2}\right)\theta^2 = \frac{n^3\,\theta^2}{n^3+4\,n^2+5\,n+2} \to \theta^2$
- considérese la convergencia en ley de $\frac{X_{(n)}-\theta}{\theta/n}$
  - $\mathrm{lím}_{n\to\infty} F_{\frac{X_{(n)}-\theta}{\theta/n}}(x) = \mathrm{lím}_n \Pr\left[\frac{X_{(n)}-\theta}{\theta/n}\leqslant x\right]$
  - si $x\geqslant 0$ entonces $\Pr\left[\frac{X_{(n)}-\theta}{\theta/n}\leqslant x\right]=1$
  - si $x < 0$ entonces $\mathrm{lím}_n \Pr\left[\frac{X_{(n)}-\theta}{\theta/n}\leqslant x\right] = \mathrm{lím}_n \Pr\left[X_{(n)}\leqslant\left(\frac{x}{n}+1\right)\theta\right] = \mathrm{lím}_n \Pr\left[\frac{X_{(n)}}{\theta}\leqslant\frac{x}{n}+1\right] = \langle Y\hookrightarrow U(0,1)\rangle = \mathrm{lím}_n \Pr\left[Y_{(n)}\leqslant\frac{x}{n}+1\right] = \mathrm{lím}_n\left(\Pr\left[Y\leqslant\frac{x}{n}+1\right]\right)^n = \mathrm{lím}_n\left(\frac{x}{n}+1\right)^n = \mathrm{lím}_n\left(1+\frac{x}{n}\right)^n = e^x$
  - $\frac{X_{(n)}-\theta}{\theta/n}\hookrightarrow -\mathrm{Exp}(1)$ distribución asintótica no gausiana