Distribución Empírica

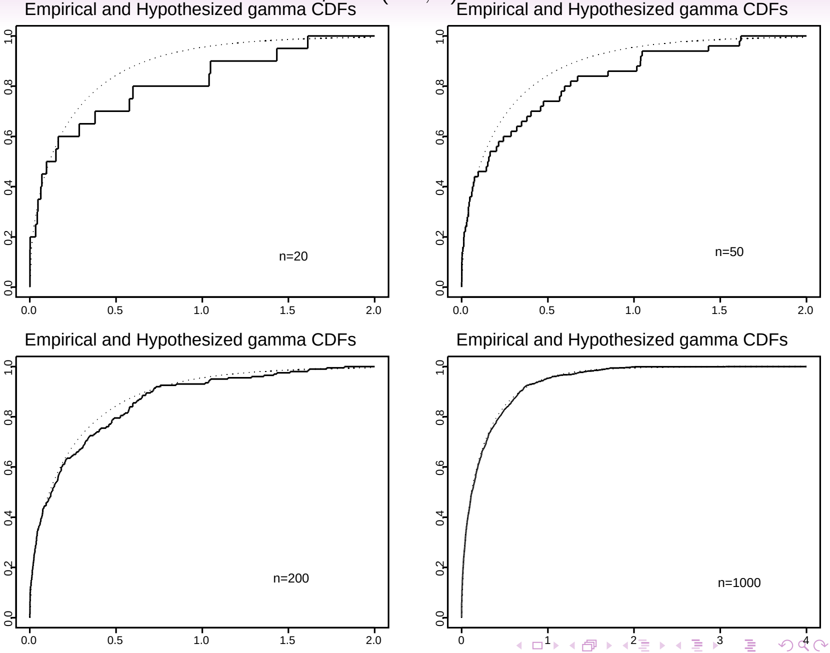
$X_1, \ldots, X_n$ i.i.d. $X_i \sim F$, $X_i \sim P$

$$\mathbb{I}_A(X_j(\omega)) = \begin{cases} 1 & X_j(\omega) \in A \\ 0 & X_j(\omega) \notin A \end{cases}$$

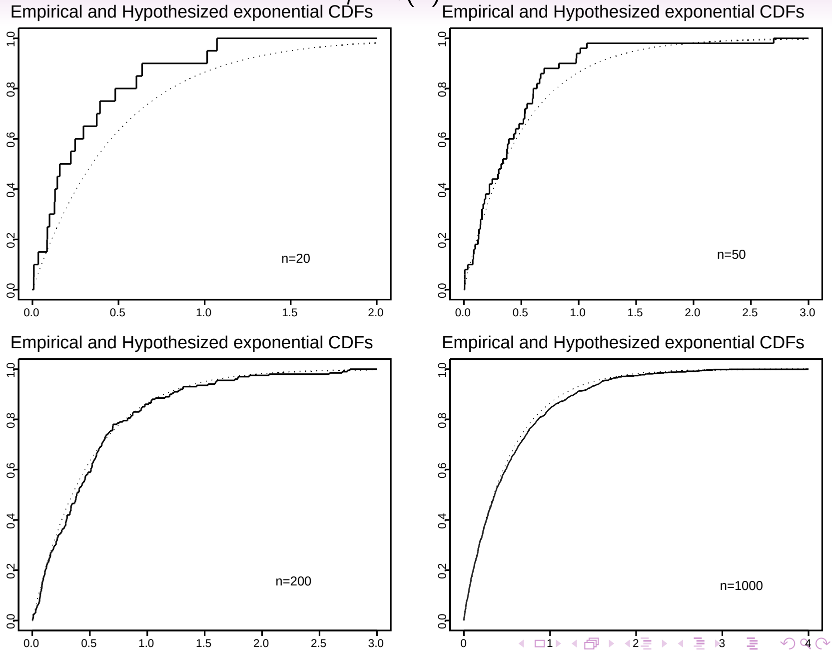
$$\begin{aligned} F_n(x,\omega) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(-\infty,x]}(X_j(\omega)) = \frac{\#\{X_j(\omega) \leq x\}}{n} \\ P_n(A,\omega) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_A(X_j(\omega)) = \frac{\#\{X_j(\omega) \in A\}}{n} \end{aligned}$$

F_n es la distribución empírica
P_n es la medida empírica asociada

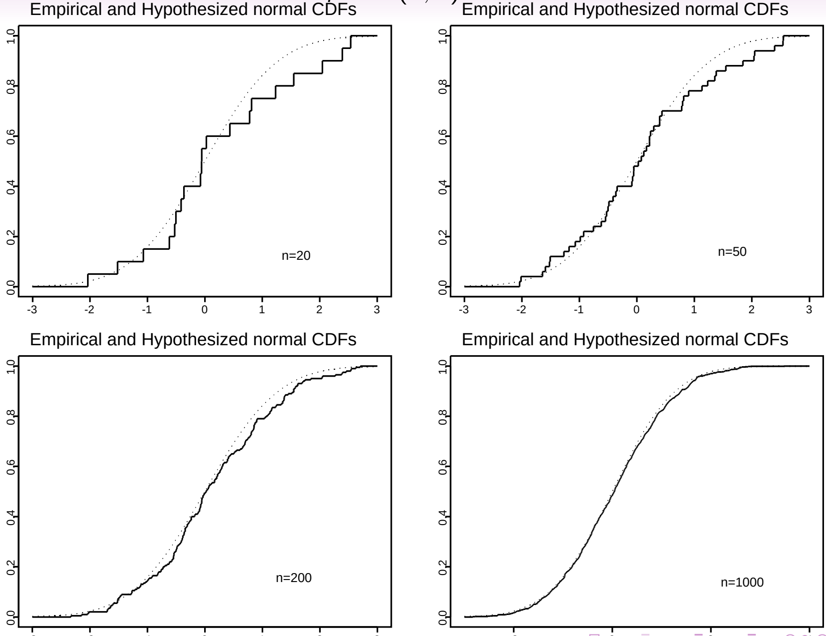
$X_i \sim \Gamma(0.5, 1)$



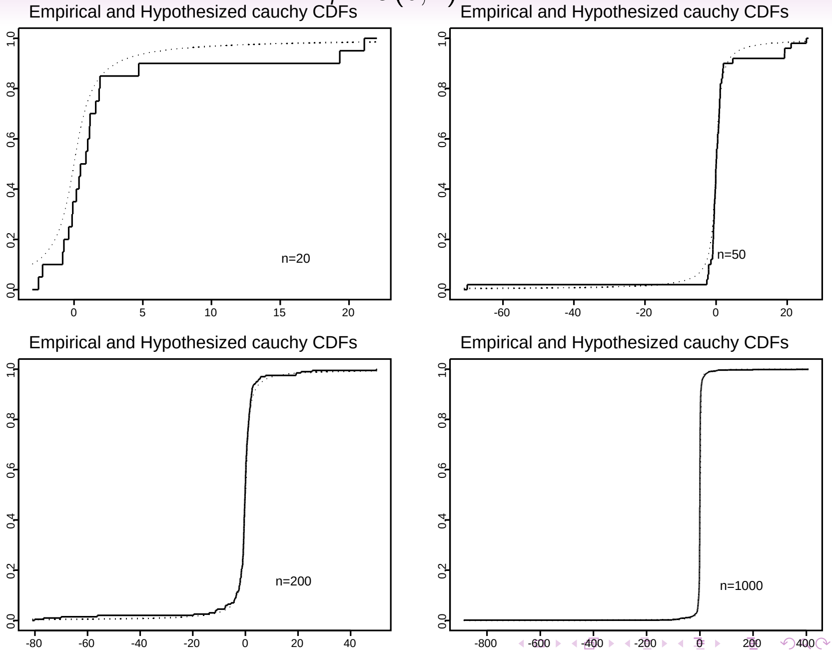
$X_i \sim \epsilon(2)$



$$X_i \sim N(0,1)$$


Empirical and Hypothesized normal CDFs
n=20
Empirical and Hypothesized normal CDFs
n=50
Empirical and Hypothesized normal CDFs
n=200
Empirical and Hypothesized normal CDFs
n=1000

$X_i \sim \mathcal{C}(0,1)$



Teorema de Glivenko–Cantelli

$X_1, \ldots, X_n$ i.i.d. $X_i \sim F$, $X_i \sim P$

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0$$

Teorema de Glivenko–Cantelli

$X_1, \ldots, X_n$ i.i.d. $X_i \sim F$, $X_i \sim P$

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0$$

Observaciones:

- $\sup_{A \in \mathcal{B}} |P_n(A) - P(A)| \mathrel{\not\xrightarrow{a.s.}} 0$
- $\sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = O_{\mathbb{P}}(1)$

Estimación

$\mathbf{X} = (X_1, \ldots, X_n) \sim F,$

$$F \in \mathcal{F} = \{F(x_1, \ldots, x_n, \boldsymbol{\theta}) \text{ donde } \boldsymbol{\theta} = (\theta_1, \ldots, \theta_p) \in \Theta \subset \mathbb{R}^p\}$$

Interesa conocer *aproximadamente* $q(\boldsymbol{\theta})$, donde $q(\cdot) : \Theta \to \mathbb{R}$.

Definición. Un *estimador puntual* de $q(\boldsymbol{\theta})$ será cualquier función $\delta(\mathbf{X})$ de las observaciones que resulte una v.a., o sea, una función $\delta : \mathbb{R}^n \to \mathbb{R}$ que se usa para conocer el valor de $q(\boldsymbol{\theta})$.

Estimación

$\mathbf{X} = (X_1, \ldots, X_n) \sim F,$

$$F \in \mathcal{F} = \{F(x_1, \ldots, x_n, \boldsymbol{\theta}) \text{ donde } \boldsymbol{\theta} = (\theta_1, \ldots, \theta_p) \in \Theta \subset \mathbb{R}^p\}$$

Usualmente X_i son i.i.d. con densidad $f(x, \boldsymbol{\theta})$ y por lo tanto, su densidad conjunta será $\prod_{i=1}^n f(x_i, \boldsymbol{\theta})$.

Un estimador de $\boldsymbol{\theta}$, o más generalmente de $q(\boldsymbol{\theta})$ será una función de $X_1, \ldots, X_n$ y por lo tanto, será una variable aleatoria con una distribución llamada *distribución muestral*.

Estimación

$\mathbf{X} = (X_1, \ldots, X_n) \sim F,$

$$F \in \mathcal{F} = \{F(x_1, \ldots, x_n, \boldsymbol{\theta}) \text{ donde } \boldsymbol{\theta} = (\theta_1, \ldots, \theta_p) \in \Theta \subset \mathbb{R}^p\}$$

Usualmente X_i son i.i.d. con densidad $f(x, \boldsymbol{\theta})$ y por lo tanto, su densidad conjunta será $\prod_{i=1}^{n} f(x_i, \boldsymbol{\theta})$.

Un estimador de $\boldsymbol{\theta}$, o más generalmente de $q(\boldsymbol{\theta})$ será una función de $X_1, \ldots, X_n$ y por lo tanto, será una variable aleatoria con una distribución llamada *distribución muestral*.

Un buen estimador $\delta(\mathbf{X})$ deberá tener la propiedad de que cualquiera sea el valor de $\boldsymbol{\theta}$, que es desconocido,la diferencia $\delta(\mathbf{X}) - q(\boldsymbol{\theta})$ sea pequeña en algún sentido.

Estimación

- $\delta(\mathbf{X})$ es insesgado si $\mathbb{E}_{\boldsymbol{\theta}}(\delta(\mathbf{X})) = q(\boldsymbol{\theta})$, $\forall \boldsymbol{\theta}$

- $\delta_n(X_1, \ldots, X_n)$ es una sucesión débilmente consistente de $q(\boldsymbol{\theta})$ si $\forall \boldsymbol{\theta}$
$$\delta_n(X_1, \ldots, X_n) \xrightarrow{p} q(\boldsymbol{\theta})$$

- $\delta_n(X_1, \ldots, X_n)$ es una sucesión fuertemente consistente de $q(\boldsymbol{\theta})$ si $\forall \boldsymbol{\theta}$
$$\delta_n(X_1, \ldots, X_n) \xrightarrow{a.s.} q(\boldsymbol{\theta})$$

- $\delta_n(X_1, \ldots, X_n)$ es una sucesión de estimadores de $q(\boldsymbol{\theta})$ asintóticamente normal si $\forall \boldsymbol{\theta}$

$$\sqrt{n}\,(\delta_n(X_1, \ldots, X_n) - q(\boldsymbol{\theta})) \xrightarrow{D} N(0, \sigma^2(\boldsymbol{\theta}))$$

Criterios para medir la bondad de un estimador

Para poder elegir el estimador $\delta(\mathbf{X})$ de $q(\boldsymbol{\theta})$ que se utilizará, se deberá dar un criterio para comparar dos estimadores cualesquiera.

Error cuadrático medio

$$\mathrm{ECM}_{\boldsymbol{\theta}}(\delta) = \mathbb{E}_{\boldsymbol{\theta}}(\delta(\mathbf{X}) - q(\theta))^2$$

$$\mathrm{ECM}_{\boldsymbol{\theta}}(\delta) = Var_{\boldsymbol{\theta}}(\delta) + \{\mathbb{E}_{\boldsymbol{\theta}}(\delta(\mathbf{X})) - q(\theta)\}^2$$

Si $\delta(\mathbf{X})$ es insesgado

$$\mathrm{ECM}_{\boldsymbol{\theta}}(\delta) = Var_{\boldsymbol{\theta}}(\delta)$$

Criterios para medir la bondad de un estimador

Un estimador $\delta_1(\mathbf{X})$ de $q(\theta)$ es mejor que $\delta_2(\mathbf{X})$ si

$$\text{ECM}_{\boldsymbol{\theta}}(\delta_1) \leq \text{ECM}_{\boldsymbol{\theta}}(\delta_2) \qquad \forall\, \boldsymbol{\theta} \in \Theta$$

y existe $\boldsymbol{\theta}_0$ tal que

$$\text{ECM}_{\boldsymbol{\theta}_0}(\delta_1) \leq \text{ECM}_{\boldsymbol{\theta}_0}(\delta_2)$$

Diremos que δ^* es un estimador óptimo si para cualquier otro estimador δ se tiene

$$\text{ECM}_{\boldsymbol{\theta}}(\delta^*) \leq \text{ECM}_{\boldsymbol{\theta}}(\delta) \qquad \forall\, \boldsymbol{\theta} \in \Theta$$

Criterios para medir la bondad de un estimador

Un estimador $\delta_1(\mathbf{X})$ de $q(\theta)$ es mejor que $\delta_2(\mathbf{X})$ si

$$\mathrm{ECM}_{\boldsymbol{\theta}}(\delta_1) \leq \mathrm{ECM}_{\boldsymbol{\theta}}(\delta_2) \qquad \forall\, \boldsymbol{\theta} \in \Theta$$

y existe $\boldsymbol{\theta}_0$ tal que

$$\mathrm{ECM}_{\boldsymbol{\theta}_0}(\delta_1) \leq \mathrm{ECM}_{\boldsymbol{\theta}_0}(\delta_2)$$

Diremos que δ^* es un estimador óptimo si para cualquier otro estimador δ se tiene

$$\mathrm{ECM}_{\boldsymbol{\theta}}(\delta^*) \leq \mathrm{ECM}_{\boldsymbol{\theta}}(\delta) \qquad \forall\, \boldsymbol{\theta} \in \Theta$$

Salvo en casos triviales no existirán tales estimadores óptimos.

Criterios para medir la bondad de un estimador

Restringimos la clase de estimadores.

Estimadores IMVU

Un estimador $\delta(\mathbf{X})$ de $q(\boldsymbol{\theta})$ es IMVU para $q(\boldsymbol{\theta})$ si

a) $\delta(\mathbf{X})$ es insesgado

b) Dado otro estimador insesgado δ^* de $q(\theta)$ se tiene que

$$\mathrm{ECM}_{\boldsymbol{\theta}}(\delta) \leq \mathrm{ECM}_{\boldsymbol{\theta}}(\delta^*) \qquad \forall\, \boldsymbol{\theta} \in \Theta$$

Propiedades

- Si
  a) $\mathbb{E}_{\boldsymbol{\theta}}(\delta(\mathbf{X})) \to q(\boldsymbol{\theta})$, $\forall \boldsymbol{\theta}$

  b) $Var_{\boldsymbol{\theta}}(\delta(\mathbf{X})) \to 0$, $\forall \boldsymbol{\theta}$

  entonces $\delta_n(X_1, \ldots, X_n)$ es una sucesión débilmente consistente de $q(\boldsymbol{\theta})$

- Si $\delta(\mathbf{X})$ es un estimador IMVU para $q(\boldsymbol{\theta})$ entonces es un estimador débilmente consistente de $q(\boldsymbol{\theta})$

Pérdidas generales

Llamaremos **Función de Pérdida** a una función $L(\boldsymbol{\theta}, d)$ tal que

- $L(\boldsymbol{\theta}, d) \geq 0$
- $L(\boldsymbol{\theta}, q(\boldsymbol{\theta})) = 0$

Se define el **Riesgo** del estimador $\delta(\mathbf{X})$ de $q(\boldsymbol{\theta})$ como

$$R(\delta, \boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \delta(\mathbf{X}))$$

Ejemplos de pérdidas son

- $L(\boldsymbol{\theta}, d) = |d - q(\boldsymbol{\theta})|$
- $L(\boldsymbol{\theta}, d) = (d - q(\boldsymbol{\theta}))^2 \quad \rightarrow$ ECM
- $L(\boldsymbol{\theta}, d) = \mathbb{I}_{|d - q(\boldsymbol{\theta})| > c}$

Admisibilidad

- Se dice que un estimador $\delta(\mathbf{X})$ de $q(\boldsymbol{\theta})$ es *inadmisible* respecto de la pérdida $L(\boldsymbol{\theta}, d)$, si existe otro estimador $\delta^*(\mathbf{X})$ mejor que él, es decir, si existe $\delta^*(\mathbf{X})$ tal que

$$\begin{array}{rcl} R(\delta^*, \boldsymbol{\theta}) & \leq & R(\delta, \boldsymbol{\theta}) \quad \forall\, \boldsymbol{\theta} \in \Theta \\ & \text{y} & \\ R(\delta^*, \boldsymbol{\theta}_0) & < & R(\delta, \boldsymbol{\theta}_0) \text{ para algún } \boldsymbol{\theta}_0 \end{array}$$

- El estimador $\delta(\mathbf{X})$ se dirá *admisible* si no es inadmisible.

Admisibilidad

Teorema.

Sea $L(\boldsymbol{\theta}, d)$ es una pérdida estrictamente convexa en d

$$L(\boldsymbol{\theta}, \alpha d_1 + (1-\alpha)d_2) < \alpha L(\boldsymbol{\theta}, d_1) + (1-\alpha)L(\boldsymbol{\theta}, d_2)$$

para todo $0 < \alpha < 1$ y para todo $\boldsymbol{\theta} \in \Theta$.

Si $\delta(\mathbf{X})$ es admisible para $q(\boldsymbol{\theta})$ y $\delta^*(\mathbf{X})$ es otro estimador de $q(\boldsymbol{\theta})$ tal que

$$R(\delta^*, \boldsymbol{\theta}) = R(\delta, \boldsymbol{\theta})$$

entonces

$$P_{\boldsymbol{\theta}}(\delta(\mathbf{X}) = \delta^*(\mathbf{X})) = 1$$

Estimador de los momentos

- $X_1, \ldots, X_n$ i.i.d. $X_i \sim F_\theta$, $\theta \in \mathbb{R}$
- $g : \mathbb{R} \to \mathbb{R}$. Usualmente $g(x) = x^k$ para algún $k \geq 1$.

El método de los momentos estima θ, por el valor $\widehat{\theta} = \delta(\mathbf{X})$ que satisface la ecuación

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) = \mathbb{E}_{\widehat{\theta}}\, g(X_1)$$

donde $\mathbb{E}_\theta(X)$ significa la esperanza de X cuando $X \sim F(x, \theta)$.

Estimador de los momentos

Ejemplo. Sea $X_1, X_2, \ldots, X_n$ i.i.d. $X_i \sim \mathcal{P}(\lambda)$, Estimador de los momentos de orden 1

$$\widehat{\lambda}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i \ .$$

Estimador de los momentos de orden 2 es

$$\widehat{\lambda}_2 = \delta_2(X_1, X_2, \ldots, X_n) = -\frac{1}{2} + \sqrt{\frac{1}{4} + \sum_{i=1}^{n} \frac{X_i^2}{n}}$$

Estimador de los momentos

Ejemplo. Sea $X_1, X_2, \ldots, X_n$ i.i.d. $X_i \sim \mathcal{P}(\lambda)$, Estimador de los momentos de orden 1

$$\widehat{\lambda}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i \ .$$

Estimador de los momentos de orden 2 es

$$\widehat{\lambda}_2 = \delta_2(X_1, X_2, \ldots, X_n) = -\frac{1}{2} + \sqrt{\frac{1}{4} + \sum_{i=1}^{n} \frac{X_i^2}{n}}$$

Estudio hecho en el Instituto Nacional de Ciencia y Tecnología de U.S.A. Fibras de asbesto en filtros. Se tomaron muestras de 3 mm de diámetro del filtro y mediante un microscopio electrónico, se contaron el número de fibras en cada una de las 23 grillas obteniéndose

31	29	19	18	31	28	34	17	27	34	30	16
18	26	27	24	27	18	24	22	28	24	21	

$$\widehat{\lambda}_1 = 24.913 \qquad \widehat{\lambda}_2 = 24.988$$

Distribución de $\widehat{\lambda}_1$

$S = \sum_{i=1}^{n} X_i \sim \mathcal{P}(n\lambda) \qquad \Rightarrow$

$$P_\lambda\left(\widehat{\lambda}_1 = \nu\right) = P_\lambda\left(S = n\,\nu\right) = \frac{(n\lambda)^{n\nu} e^{-n\lambda}}{(n\nu)!} \quad \text{si} \quad n\nu \in N_0 \,.$$

$$\mathbb{E}_\lambda\left(\widehat{\lambda}_1\right) = \lambda \qquad\qquad \mathrm{Var}_\lambda\left(\widehat{\lambda}_1\right) = \frac{\lambda}{n}$$

$\mathbb{E}_\lambda\left(\widehat{\lambda}_1\right) = \lambda$ dice que $\widehat{\lambda}_1$ es un estimador insesgado de λ.

$\mathrm{Var}_\lambda\left(\widehat{\lambda}_1\right) = \frac{\lambda}{n}$ dice que

- la distribución es más concentrada a medida que n aumenta y
- como es insesgado, $\widehat{\lambda}_1$ es débilmente consistente.

Estimador de los momentos

Si usamos el Teorema Central del Límite obtenemos que

- $\sqrt{n}\left(\widehat{\lambda}_1 - \lambda\right) \xrightarrow{D} N(0, \lambda)$
- $\sqrt{n}\left(\widehat{\lambda}_2 - \lambda\right) \xrightarrow{D} N(0, \sigma_2^2)$ donde

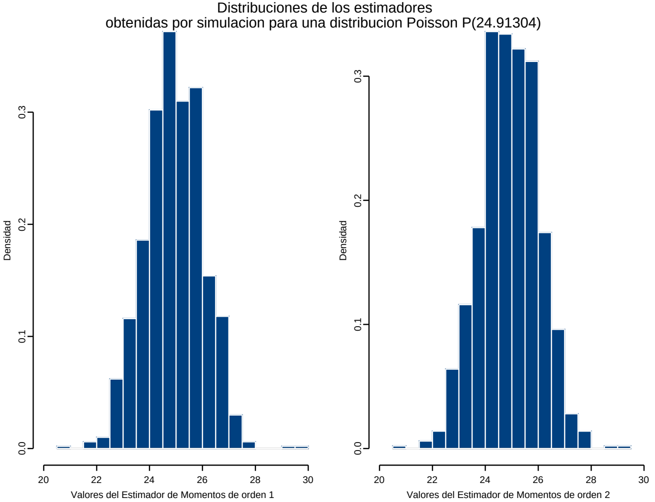
$$\sigma_2^2 = \frac{\lambda + 6\lambda^2 + 4\lambda^3}{1 + 4\lambda + 4\lambda^2}$$

usando que si $X \sim \mathcal{P}(\lambda)$

$$\mathbb{E}_\lambda(X^4) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$$

Luego, la varianza asintótica de $\widehat{\lambda}_2$ es mayor que la de $\widehat{\lambda}_1$

Distribución de $\widehat{\lambda}_1$ y $\widehat{\lambda}_2$ NR=1000



$\mathrm{SD}_{\widehat{\lambda}_1} = 1.086437$

$\mathrm{SD}_{\widehat{\lambda}_2} = 1.092775$

Estimador de los momentos: Generalización cuando hay varios parámetros

$X_1, X_2, \ldots, X_n$ i.i.d.

$X_i \sim F \in \mathcal{F} = \{F(x, \theta_1, \ldots, \theta_p) \text{ con } \boldsymbol{\theta} = (\theta_1, \ldots, \theta_p) \in \Theta \subset \mathbb{R}^p\}$.

Para estimar $\theta_1, \ldots, \theta_p$ por el método de los momentos se procede como sigue:

- Se consideran k funciones $g_1, \ldots, g_p$ de $\mathbb{R}$ en $\mathbb{R}$
- Se resuelve el siguiente sistema

$$\frac{1}{n}\sum_{i=1}^{n} g_j(X_i) = \mathbb{E}_{\widehat{\boldsymbol{\theta}}}\, g_j(X_1) \qquad j = 1, 2, \ldots, p\ .$$

Podemos tomar, por ejemplo, $g_j(x) = x^j$.

Estimador de los momentos

En conclusión, de los ejemplos anteriores vemos que los pasos básicos para estimar $\boldsymbol{\theta} \in \mathbb{R}^p$ son:

a) Calcular p momentos de la distribución de las observaciones

b) Expresar $\boldsymbol{\theta}$ en función de ellos

c) Reemplazar los momentos por los momentos muestrales (o sea, reemplazando F por F_n) en la expresión obtenida en b) para obtener $\widehat{\boldsymbol{\theta}}$.

Caso Gamma

Sea $X_1, X_2, \ldots, X_n$ i.i.d. $X_i \sim \Gamma(\alpha, \lambda)$.

Consideremos $g_1(x) = x$ y $g_2(x) = x^2$. Luego, si

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

los estimadores de los momentos para λ y α resultan ser

$$\begin{aligned} \widehat{\lambda} &= \frac{\overline{X}}{\widehat{\sigma}^2} \\ \widehat{\alpha} &= \frac{\overline{X}^2}{\widehat{\sigma}^2}. \end{aligned}$$

Tiempo de falla de cables Kevlar 49/epoxy sometidos a presión sostenida.

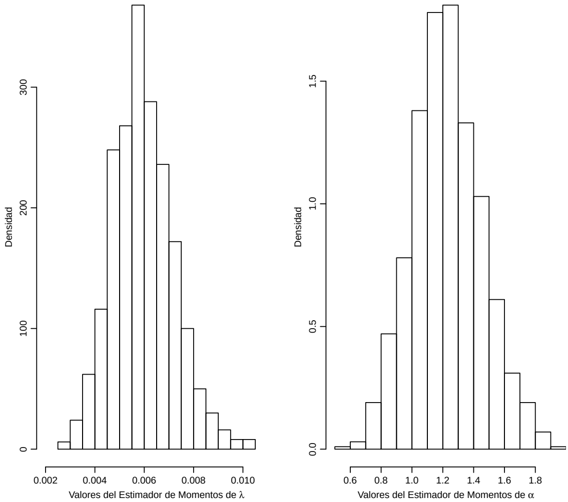
90% de Presión							
0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.04
0.05	0.06	0.07	0.07	0.08	0.09	0.09	0.10
0.10	0.11	0.11	0.12	0.13	0.18	0.19	0.20
0.23	0.80	0.80	0.83	0.85	0.90	0.92	0.95
0.99	1.00	1.01	1.02	1.03	1.05	1.10	1.10
1.11	1.15	1.18	1.20	1.29	1.31	1.33	1.34
1.40	1.43	1.45	1.50	1.51	1.52	1.53	1.54
1.54	1.55	1.58	1.60	1.63	1.64	1.80	1.80
1.81	2.02	2.05	2.14	2.17	2.33	3.03	3.03
3.24	4.20	4.69	7.89				

80% de Presión									
0.18	3.1	4.2	6	7.5	8.2	8.5	10.3	10.6	24.2
29.6	31.7	41.9	44.1	49.5	50.1	59.7	61.7	64.4	69.7
70	77.8	80.5	82.3	83.5	84.2	87.1	87.3	93.2	103.4
104.6	105.5	108.8	112.6	116.8	118	122.3	123.5	124.4	125.4
129.5	130.4	131.6	132.8	133.8	137	140.2	140.9	148.5	149.2
152.2	152.8	157.7	160	163.6	166.9	170.5	174.9	177.7	179.2
183.6	183.8	194.3	195.1	195.3	202.6	220	221.3	227.2	251
266.5	267.9	269.2	270.4	272.5	285.9	292.6	295.1	301.1	304.3
316.8	329.8	334.1	346.2	351.2	353.3	369.3	372.3	381.3	393.5
451.3	461.5	574.2	656.3	663	669.8	739.7	759.6	894.7	974.9

Para el Tiempo de Falla de cables Kevlar 49/epoxy sometidos a presión sostenida del 80%, obtenemos $\overline{X} = 209.1828$ y $\widehat{\sigma}^2 = 37640$, con lo cual $\widehat{\alpha} = 1.1625$ y $\widehat{\lambda} = 0.00556$.

Distribución de $\widehat{\alpha}$ y $\widehat{\lambda}$

La podemos obtener por simulación generando muchas muestras de tamaño 100 de una distribución $\Gamma(1.1625, 0.00556)$.



Distribución de $\widehat{\alpha}$ y $\widehat{\lambda}$

Los gráficos muestran que los estimadores parecen tener una distribución aproximadamente normal. La variabilidad mostrada en los histogramas puede resumirse calculando los desvíos de las 1000 estimaciones, lo que provee el error estándar estimado de $\widehat{\alpha}$ y de $\widehat{\lambda}$ y que en este caso son 0.2251 y 0.00124 respectivamente.

Este uso de la simulación es un ejemplo de lo que en estadística se denomina *bootstrap.*

Estimador de máxima verosimilitud

$\mathbf{X} = (X_1, X_2, \ldots, X_n)$ discreto o continuo cuya función de densidad discreta o continua pertenece a una familia $f(\mathbf{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$ y que se quiera estimar $\boldsymbol{\theta}$.

$\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\mathbf{X})$ es un estimador de máxima verosimilitud (E.M.V.) de $\boldsymbol{\theta}$, si se cumple

$$f(\mathbf{X}, \widehat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \Theta} f(\mathbf{X}, \boldsymbol{\theta})$$

Cómputo del E.M.V..

- $\Theta \subset \mathbb{R}^p$ abierto,
- el soporte de $f(\mathbf{x}, \boldsymbol{\theta})$ no depende de $\boldsymbol{\theta}$
- $f(\mathbf{x}, \boldsymbol{\theta})$ tiene derivadas parciales respecto a θ_i.

Estimador de máxima verosimilitud

Luego, el E.M.V. $\widehat{\boldsymbol{\theta}}$ debe verificar

$$\frac{\partial \ln f(\mathbf{X}, \boldsymbol{\theta})}{\partial \theta_i} = 0 \qquad i = 1, 2, \ldots, p\,. \qquad (1)$$

En particular, si $X_1, \ldots, X_n$ son i.i.d. con función de probabilidad puntual o densidad $f(x, \boldsymbol{\theta})$, se verifica

$$f(\mathbf{x}, \boldsymbol{\theta}) = f(x_1, \ldots, x_n, \boldsymbol{\theta}) = \prod_{j=1}^{n} f(x_j, \boldsymbol{\theta})$$

y bajo las condiciones dadas anteriormente, el sistema de ecuaciones (1) se transforma en

$$\sum_{i=1}^{n} \frac{\partial \ln f(X_i, \widehat{\boldsymbol{\theta}})}{\partial \theta_j} = 0 \qquad j = 1, 2, \ldots, p\,. \qquad (2)$$

Estimador de máxima verosimilitud

Sea

$$\psi_j(x, \boldsymbol{\theta}) = -\frac{\partial \ln f(x, \widehat{\boldsymbol{\theta}})}{\partial \theta_j}$$

entonces (2) puede escribirse como

$$\sum_{i=1}^{n} \psi_j(X_i, \boldsymbol{\theta}) = 0 \qquad\qquad j = 1, 2, \ldots, p\ .$$

Esta ecuación corresponde a la forma general de los denominados $M-$estimadores.

Ejemplo: Caso Normal

Sea $X_1, \ldots, X_n$, $X_i \sim N(\mu, \sigma^2)$. Los E.M.V. de μ y σ^2 son

$$\begin{aligned} \widehat{\mu} &= \sum_{i=1}^{n} X_i/n = \overline{X} \\ \widehat{\sigma}^2 &= \sum_{i=1}^{n} (X_i - \overline{X})^2/n \end{aligned}$$

que son los mismos estimadores que encontramos por el método de los momentos.

Ejemplo: Caso Gamma

Sean $X_i \sim \Gamma(\alpha, \lambda)$, $1 \leq i \leq n$. Sea $\Gamma'(\alpha)$ la derivada de la función $\Gamma(\alpha)$. Los E.M.V. de α y λ verifican

$$\begin{aligned} n \ln \widehat{\lambda} + \sum_{i=1}^{n} \ln(X_i) - n \frac{\Gamma'(\widehat{\alpha})}{\Gamma(\widehat{\alpha})} &= 0 \\ \frac{n \widehat{\alpha}}{\widehat{\lambda}} - n \overline{X} &= 0 \,, \end{aligned}$$

Este sistema no tiene una solución explícita aunque

$$\widehat{\lambda} = \frac{\widehat{\alpha}}{\overline{X}} \,.$$

Ejemplo: Caso Gamma

Reemplazando obtenemos

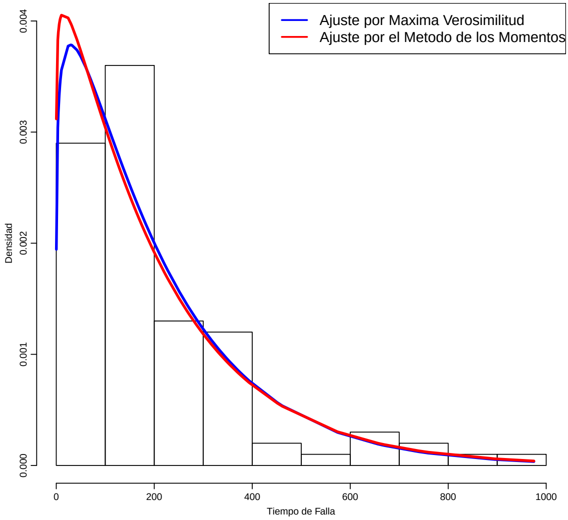
$$n\left(\ln\widehat{\alpha} - \ln(\overline{X})\right) + \sum_{i=1}^{n} \ln(X_i) - n\,\frac{\Gamma'(\widehat{\alpha})}{\Gamma(\widehat{\alpha})} = 0\,,$$

que puede resolverse numéricamente. Para iniciar el proceso, se puede tomar como estimador inicial el estimador de los momentos, por ejemplo.

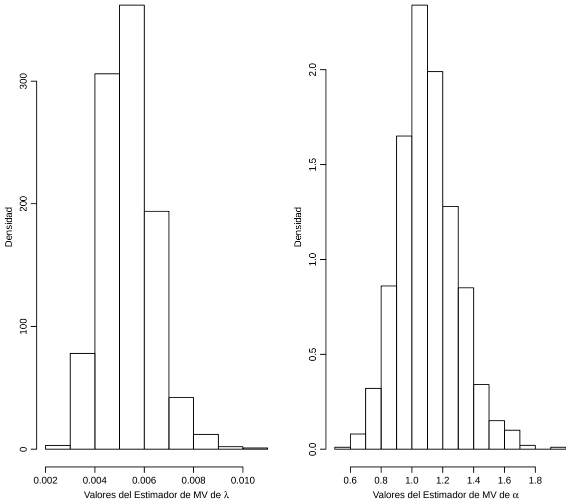
En este caso, el estimador de máxima verosimilitud no coincide con el estimador de los momentos.

Tiempo de Falla de cables Kevlar: presión 80%

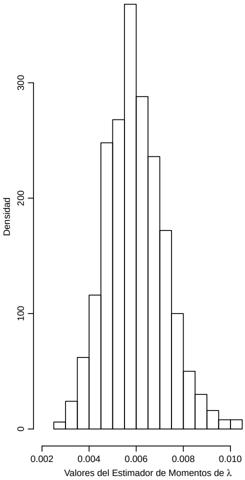
Los E.M.V. resultan ser $\widehat{\alpha} = 1.0774$ y $\widehat{\lambda} = 0.00515$.

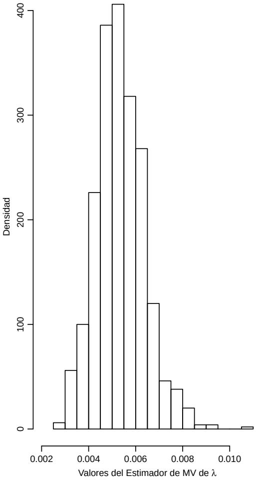


Los E.M.V. no tienen una forma explícita en el caso $\Gamma(\alpha, \lambda)$.
Bootstrap: Generamos $nr = 1000$ muestras cada una de tamaño $n = 100$ de una variable $\Gamma(1.0774, 0.00515)$.



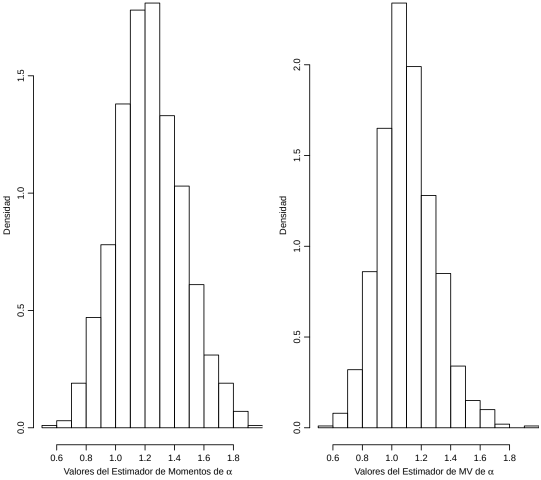
Histogramas de los estimadores de λ utilizando bootstrap ($nr = 1000$)





$s_{\widehat{\lambda}_{\text{Momentos}}} = 0.00124$ $\qquad$ $s_{\widehat{\lambda}_{\text{MV}}} = 0.00102$

Histogramas de los estimadores de α utilizando $nr = 1000$



$s_{\widehat{\alpha}_{\text{Momentos}}} = 0.2251$ $\qquad$ $s_{\widehat{\alpha}_{\text{MV}}} = 0.1865$

Ejemplo: Distribución Uniforme

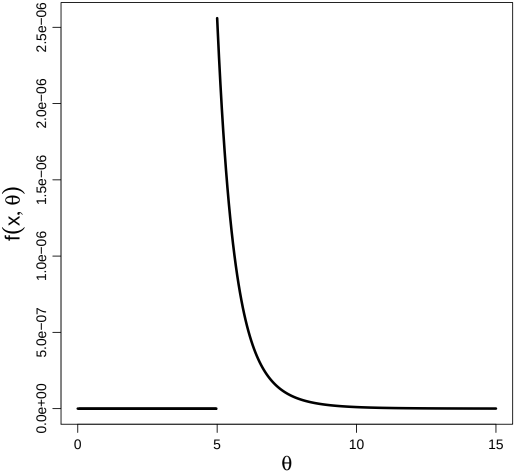
$X_i \sim \mathcal{U}(0, \theta)$.

$$f(\mathbf{x}, \theta) = \begin{cases} \dfrac{1}{\theta^n} & \text{si} \quad 0 < \text{mín}(x_i) \leq \text{máx}(x_i) < \theta \\ 0 & \text{otro caso} \end{cases}$$

El E.M.V. de θ es

$$\widehat{\theta}_n = \max_{1 \leq i \leq n} (X_i)$$

Gráfico de la función de verosimilitud, $f(\mathbf{x}, \theta)$ correspondiente a $n = 8$ observaciones tales que $\max_{1 \leq i \leq n}(X_i) = 5$.



La distribución de $\widehat{\theta}_n = \max_{1\leq i\leq n}(X_i)$ está dada por

$$F_{\widehat{\theta}_n}(t,\theta) = \begin{cases} 0 & t<0 \\ \dfrac{t^n}{\theta^n} & \text{si} \quad 0<t<\theta \\ 1 & \theta<t \end{cases}$$

con lo cual

$$\mathbb{E}_{\boldsymbol{\theta}}\left(\max_{1\leq i\leq n}(X_i)\right) = \frac{n}{n+1}\theta \qquad \mathrm{Var}_{\boldsymbol{\theta}}\left(\max_{1\leq i\leq n}(X_i)\right) = \frac{n}{(n+2)(n+1)^2}\theta^2 \, .$$

Por lo tanto, $\widehat{\theta}_n$ es débilmente consistente para θ.

Invarianza del Estimador de máxima verosimilitud

Sea $\boldsymbol{\lambda} = g(\boldsymbol{\theta})$ donde $g : \Theta \to \Lambda$, $\Theta, \Lambda \subset \mathbb{R}^p$.

Si g es biunivoca, $f(\mathbf{x}, \boldsymbol{\theta})$ se puede expresar en función de $\boldsymbol{\lambda}$.

$$f^*(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}, g^{-1}(\boldsymbol{\lambda}))$$

El estimador de máxima verosimilitud de $\boldsymbol{\lambda}$ es

$$f^*(\mathbf{x}, \widehat{\boldsymbol{\lambda}}) = \max_{\boldsymbol{\lambda} \in \Lambda} f^*(\mathbf{x}, \boldsymbol{\lambda})$$

Teorema. Si $\widehat{\boldsymbol{\theta}}$ es el estimador de máxima verosimilitud de $\boldsymbol{\theta}$, entonces $\widehat{\boldsymbol{\lambda}} = g(\widehat{\boldsymbol{\theta}})$ es el estimador de máxima verosimilitud de $\boldsymbol{\lambda}$.

Estimador de mínimos cuadrados

Supongamos que $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias de la forma

$$Y_i = G_i(\theta_1, \ldots, \theta_p) + \varepsilon_i \qquad\qquad 1 \leq i \leq n$$

donde

- $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p) \in \Theta \subset \mathbb{R}^p$
- ε_i son variables aleatorias con
  (i) $\mathbb{E}(\varepsilon_i) = 0$
  (ii) $\mathrm{Var}(\varepsilon_i) = \sigma^2$
  (iii) $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ son variables aleatorias independientes.

Estimador de mínimos cuadrados

Llamaremos estimador de mínimos cuadrados al valor $\widehat{\boldsymbol{\theta}}(Y_1, Y_2, \ldots, Y_n)$ que minimiza

$$\sum_{i=1}^{n}(Y_i - G_i(\theta_1, \theta_2, \ldots, \theta_p))^2$$

es decir,

$$\sum_{i=1}^{n}(Y_i - G_i(\widehat{\boldsymbol{\theta}}))^2 = \min_{\boldsymbol{\theta}\in\Theta}\sum_{i=1}^{n}(Y_i - G_i(\boldsymbol{\theta}))^2 \; .$$

Si $\varepsilon_i \sim N(0, \sigma^2)$ el estimador de mínimos cuadrados coincide con el de máxima verosimilitud.

Estimador de mínimos cuadrados

Si Θ es abierto y si $G_i(\boldsymbol{\theta})$ son derivables respecto a cada θ_i, $\widehat{\boldsymbol{\theta}}$ satisface

$$\sum_{i=1}^{n}\left(Y_i - G_i(\widehat{\boldsymbol{\theta}})\right)\frac{\partial G_i(\widehat{\boldsymbol{\theta}})}{\partial \theta_j} = 0 \qquad\qquad j = 1, 2, \ldots, p\ .$$

Ejemplo Si $G_i(\theta_1, \theta_2) = \theta_1 X_i + \theta_2$, $1 \leq i \leq n$ entonces

$$\begin{aligned} \widehat{\theta}_1 &= \frac{\sum_{i=1}^{n}(Y_i - \overline{X})(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\,, \\ \widehat{\theta}_2 &= \overline{Y} - \widehat{\theta}_1\overline{X}\,, \end{aligned}$$