Generalización

Sean $X_1, \ldots, X_n$ v.a. i.i.d. con distribución $F(x, \theta)$ y $g : \mathbb{R} \to \mathbb{R}$ una función con esperanza finita.
El método generalizado de los momentos estima a θ como la solución de

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

Generalización

Sean $X_1, \ldots, X_n$ v.a. i.i.d. con distribución $F(x, \theta)$ y $g : \mathbb{R} \to \mathbb{R}$ una función con esperanza finita.
El método generalizado de los momentos estima a θ como la solución de

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

Resolvamos el siguiente ejemplo:

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con densidad de la forma

$$f(x, \theta) = (1 - \theta)\mathbb{I}_{\left(-\frac{1}{2}, 0\right)}(x) + (1 + \theta)\mathbb{I}_{\left(0, \frac{1}{2}\right)}(x), \qquad -1 < \theta < 1$$

Hallar un estimador de momentos para $q(\theta) = \mathbb{P}_\theta(X > 0)$.

Estimadores de Máxima Verosimilitud

¿Qué es la verosmilitud?

- Antes de realizar el *experimento* el **resultado** es **desconocido**.
- Las probabilidades nos permiten predecir un resultado **desconocido** a partir de parámetros **conocidos**: por ejemplo

$$P(resultado|\theta) \overset{\text{por ej. caso Bernoulli}}{=} P(x|\theta) = \theta^x(1-\theta)^{(1-x)}$$

En Estadística se invierte el paradigma:

- Al realizar el experimento el **resultado** se hace **conocido**: **dato**.
- Nos interesa cuán **verosímil** es que un determinado parámetro haya generado el dato.

Ejemplo

El Método Máxima Verosimilitud es uno de los procedimientos más populares.

Sea X una variable aleatoria que corresponde al valor que toma un tetraedro con caras numeradas del 1 al 4 al ser arrojado de acuerdo con la ley $\mathbb{P}_\theta$. Hay dos posibilidaes: $\theta = \theta_0$ o θ_1 y los valores de $\mathbb{P}_{\theta_j}(\{i\})$ están dados por la siguiente tabla:

	$x = 1$	$x = 2$	$x = 3$	$x = 4$
$\theta = \theta_0$	0.7	0.1	0.1	0.1
$\theta = \theta_1$	0.1	0.3	0.3	0.3

Si $X = 1$ es observado, es más probable que provenga de P_{θ_0}, dado que $P_{\theta_0}(\{1\})$ es mucho mayor que $P_{\theta_1}(\{1\})$.
Luego , estimaríamos a θ por θ_0.

Ejemplo

Por otro lado, si $X = 2$ o 3 o 4, es más probable que provenga de P_{θ_1}, aunque en este caso la diferencia entre las probabilidades no es tan grande como cuando $X = 1$.
Esto sugiere como estimador de θ:

$$T(X) = \begin{cases} \theta_0 & X = 1 \\ \theta_1 & X \neq 1. \end{cases}$$

Ejemplo

Por otro lado, si $X = 2$ o 3 o 4, es más probable que provenga de P_{θ_1}, aunque en este caso la diferencia entre las probabilidades no es tan grande como cuando $X = 1$.
Esto sugiere como estimador de θ:

$$T(X) = \begin{cases} \theta_0 & X = 1 \\ \theta_1 & X \neq 1. \end{cases}$$

Esta idea puede extenderse fácilmente al caso en que $\mathbb{P}_\theta$ es una distribución discreta y $\theta \in \Theta \subset \mathbb{R}^k$.

Si $X = x$ es observada, θ_1 es más probable que θ_0 si y sólo si $\mathbb{P}_{\theta_1}(\{x\}) > \mathbb{P}_{\theta_0}(\{x\})$.
Podemos estimar θ por el valor $\hat{\theta}$ que maximiza $\mathbb{P}_\theta(\{x\})$ sobre $\theta \in \Theta$, si tal $\hat{\theta}$ existe.
Esto se extiende en forma directa al caso de X continua.

Estimador de Máxima Verosimilitud: Caso discreto

- Modelo: $\mathcal{F} = \{p(\cdot,\theta)\ , \theta \in \Theta\}$.

- $\mathbf{x}= x_1, \cdots, x_n$ realización de $X_1, \cdots, X_n$ i.i.d. donde $X_i \sim p(\cdot,\theta)$.

- Función de verosimilitud asociada a $\mathbf{x}= x_1, \cdots, x_n$:

$$L(\cdot,\ \mathbf{x}\ )\ : \Theta \to \mathbb{R}$$

$$\begin{aligned} L(\theta,\ \mathbf{x}\ ) &= \mathbb{P}_\theta(X_1 = x_1, \cdots, X_n = x_n)\ , X_i \sim p(\cdot,\theta) \\ L(\theta,\ \mathbf{x}\ ) &= \prod_{i=1}^{n} p(x_i,\theta)\ , \end{aligned}$$

Estimador de Máxima Verosimilitud: Caso discreto

- Modelo: $\mathcal{F} = \{p(\cdot, \theta)\ , \theta \in \Theta\}$.

- $\mathbf{x}= x_1, \cdots, x_n$ realización de $X_1, \cdots, X_n$ i.i.d. donde $X_i \sim p(\cdot, \theta)$.

- Función de verosimilitud asociada a $\mathbf{x}= x_1, \cdots, x_n$:

$$L(\cdot,\ \mathbf{x}\ )\ : \Theta \to \mathbb{R}$$

$$\begin{aligned} L(\theta,\ \mathbf{x}\ )\ &=\ \mathbb{P}_\theta(X_1 = x_1, \cdots, X_n = x_n)\ , X_i \sim p(\cdot, \theta) \\ L(\theta,\ \mathbf{x}\ )\ &=\ \prod_{i=1}^{n} p(x_i, \theta)\ , \end{aligned}$$

- Propuesta de máxima verosimilitud: $\widehat{\theta}$ es el EMV si

$$L(\widehat{\theta}(\mathbf{X}),\ \mathbf{X}\ ) = \underset{\theta \in \Theta}{\operatorname{argmax}}\ L(\theta,\ \mathbf{X}\ )\ .$$

o sea

$$L(\widehat{\theta}(\mathbf{X}), \mathbf{X}) \geq L(\theta, \mathbf{X})$$

Ejemplo: Bernoulli

- $(X_i)_{i \geq 1}$ i.i.d., $X_i \sim \mathcal{B}(1, \theta)$, $\theta \in [0, 1]$
- $L(\theta; \mathbf{x})$: probabilidad de observar $\mathbf{x} = (\mathbf{x_1}, \ldots, \mathbf{x_n})$ en n lanzamientos cuando la probabilidad de cara es θ. Notemos que $x_i = 1$ o 0.

- $(X_i)_{i \geq 1}$ i.i.d., $X_i \sim \mathcal{B}(1, \theta)$, $p(x, \theta) = \theta^x (1 - \theta)^{1-x}$

$$
\begin{aligned}
L(\theta; \mathbf{x}) &= \prod_{i=1}^{n} p(x_i, \theta) \\
&= \prod_{i=1}^{n} \theta^{x_i} (1 - \theta)^{1 - x_i} = \theta^{\sum_{i=1}^{n} x_i} (1 - \theta)^{n - \sum_{i=1}^{n} x_i} \\
&= \theta^{n\overline{x}} (1 - \theta)^{n - n\overline{x}} = \theta^{n\overline{x}} (1 - \theta)^{n(1 - \overline{x})}
\end{aligned}
$$

Ejemplo: Bernoulli

- $(X_i)_{i\geq 1}$ i.i.d., $X_i \sim \mathcal{B}(1,\theta)$, $\theta \in [0,1]$
- $L(\theta;\mathbf{x})$: probabilidad de observar $\mathbf{x} = (\mathbf{x_1},\ldots,\mathbf{x_n})$ en n lanzamientos cuando la probabilidad de cara es θ. Notemos que $x_i = 1$ o 0.

- $(X_i)_{i\geq 1}$ i.i.d., $X_i \sim \mathcal{B}(1,\theta)$, $p(x,\theta) = \theta^x(1-\theta)^{1-x}$

$$
\begin{aligned}
L(\theta;\mathbf{x}) &= \prod_{i=1}^{n} p(x_i,\theta) \\
&= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \\
&= \theta^{n\overline{x}}(1-\theta)^{n-n\overline{x}} = \theta^{n\overline{x}}(1-\theta)^{n(1-\overline{x})}
\end{aligned}
$$

- Maximizar $L(\theta;\mathbf{x})$ equivale a maximizar $\ell(\theta;\mathbf{x}) = \log \mathbf{L}(\theta;\mathbf{x})$ ya que $\log$ es estrictamente creciente.

Función de Verosimilitud basada en $\mathbf{x}$

Caso Bernoulli

- Maximizar $L(\theta; \mathbf{x})$ equivale a maximizar $\ell(\theta; \mathbf{x}) = \log \mathbf{L}(\theta; \mathbf{x})$

$$\ell(\theta; \mathbf{x}) \quad = \quad n\overline{x} \ \log(\theta) + n(1-\overline{x}) \ \log(1-\theta)$$

- Derivemos e igualemos a 0:

$$\frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} \quad = \quad \frac{n\overline{x}}{\theta} - \frac{n(1-\overline{x})}{1-\theta} = 0$$

- Cuando $0 < \bar{x} < 1$, hay un único punto crítico $\bar{x}$
- La derivada segunda es:

$$\frac{\partial^2 \ell(\theta; \mathbf{x})}{\partial^2 \theta} \quad = \quad -\frac{n\overline{x}}{\theta^2} - \frac{n(1-\overline{x})}{(1-\theta)^2}$$

que es siempre negativa.

Estimador de Máxima Verosimilitud: Caso discreto

Caso Bernoulli

- $L(\theta, \mathbf{x})$ tiende a 0 cuando θ tiende a 0 o a 1 (la frontera del espacio paramétrico): $\overline{x}$ es el único máximo de θ.
- Cuando $\overline{x} = 0$, $L(\theta) = (1-\theta)^n$ es estrictamente decreciente como función de θ y por lo tanto 0 es el único máximo. Análogamente si $\overline{x} = 1$.
- Combinando todo esto, cada vez que se observan los datos $x_1, \ldots, x_n$ tenemos que $\overline{x}$ maximiza, el estimador de máxima verosimilitud es
$$\widehat{\theta} = \overline{X}$$

Observación: si el espacio paramétrico fuera $\Theta = (0,1)$, cuando $\overline{x} = 0$ o 1, tenemos que el máximo no se alcanza en Θ, sin embargo el supremo en $(0,1)$ es 1, con lo cual el EMV toma un valor fuera de $(0,1)$, que no sería un valor razonable. Sin embargo, a medida qe n crece la probabilidad de que esto ocurra tiende a 0.

Estimador de Máxima Verosimilitud

Algunas observaciones

- Algunos textos trabajan con la clausura del espacio paramétrico $\overline{\Theta}$ en lugar de Θ. Esto es porque el máximo de $L(\theta)$ podría no existir en Θ.
- Cuando el espacio paramétrico consiste en un conjunto finito de puntos, entonces $\overline{\Theta} = \Theta$ y el MLE puede obtenerse haciendo un número finito de comparaciones.
- Como hemos visto, en casos regulares, podemos trabajar con $\ell(\theta)$ en lugar de la verosimilitud y los candidatos a EMV se obtienen resolviendo
$$\frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} = 0$$
Una raíz de esto podría ser un mínimo local o global, idem máximo o un punto estacionario. También los máximos podrían ocurrir en la frontera de Θ o cuando $\|\theta\| \to \infty$.
- En algunos casos $\ell(\theta)$ no es diferenciable.
- Un MLE podría no existir o podría haber más de uno.
- Un MLE puede no tener una forma explícita.

Estimador de Máxima Verosimilitud: Caso continuo

- Modelo: $\mathcal{F} = \{f(\cdot, \theta)\ , \theta \in \Theta\}$, con $f(\cdot, \theta)$ función de densidad.

- $\mathbf{x} = x_1, \cdots, x_n$ realización correspondiente a $X_1, \ldots, X_n$ con $X_i \sim f(\cdot, \theta)$.

- Función de verosimilitud: $L(\cdot\ ;\ \mathbf{x}\ )\ : \Theta \to \mathbb{R}$

$$
\begin{aligned}
L(\theta\ ;\ \mathbf{x}\ ) &= f_{X_1, \cdots, X_n}(x_1, \ldots, x_n)\ , X_i \sim f(\cdot, \theta)\ . \\
L(\theta\ ;\ \mathbf{x}\ ) &= \prod_{i=1}^{n} f(x_i, \theta)\ ,
\end{aligned}
$$

- Propuesta de máxima verosimilitud: $\widehat{\theta}$ es el EMV si

$$
L(\widehat{\theta}(\mathbf{X}),\ \mathbf{X}\ ) = \underset{\theta \in \Theta}{\operatorname{argmax}}\ L(\theta,\ \mathbf{X}\ )\ .
$$

o sea

$$
L(\widehat{\theta}(\mathbf{X}), \mathbf{X}) \geq L(\theta, \mathbf{X})
$$

Ejemplo $\mathcal{E}(\lambda)$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{E}(\lambda), \lambda > 0$. $f(x, \lambda) = \lambda e^{-x\lambda}\, \mathcal{I}_{[0,\infty)}(x)$.

$$L(\lambda; \mathbf{x}) = \prod_{i=1}^{n} f(x_i, \lambda) = \prod_{i=1}^{n} \lambda e^{-x_i\lambda}\, \mathcal{I}_{[0,\infty)}(x_i)$$

- si $x_i \geq 0 \; \forall i$

$$L(\lambda; \mathbf{x}) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

- Si consideramos el $\log L$, resulta

$$\ell(\lambda; \mathbf{x}) = n \log(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

- Derivando e igualando a 0 queda

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \text{punto crítico es } 1/\bar{x}$$

- $\Rightarrow \widehat{\lambda} = 1/\bar{X}_n$

Ejemplo $\mathcal{N}(\mu, \sigma^2)$

$X_1, \cdots, X_n$ v.a. i.i.d. $f\left(x, \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

$$L(\mu, \sigma^2; \mathbf{x}) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x_i-\mu)^2}{2\sigma^2}}$$

Ejemplo $\mathcal{N}(\mu, \sigma^2)$

$X_1, \cdots, X_n$ v.a. i.i.d. $f\left(x, \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\; e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

$$L(\mu, \sigma^2; \mathbf{x}) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x_i-\mu)^2}{2\sigma^2}}$$

$$L(\mu, \sigma^2; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\displaystyle\sum_{i=1}^{n} \dfrac{(x_i-\mu)^2}{2\sigma^2}}$$

Ejemplo $\mathcal{N}(\mu, \sigma^2)$

$X_1, \cdots, X_n$ v.a. i.i.d. $f\left(x, \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

$$L(\mu, \sigma^2; \mathbf{x}) \quad = \quad \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$$

- Tomando logaritmo y resolviendo las ecuaciones

$$\frac{\partial \ell(\mu, \sigma^2; \mathbf{x})}{\partial \mu} = 0 \quad \text{y} \quad \frac{\partial \ell(\mu, \sigma^2; \mathbf{x})}{\partial \sigma^2} = 0$$

se obtiene que los EMV de μ y σ^2 son

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

Consideremos la matriz de derivadas segundas H
Comprobar que

$$\frac{\partial^2}{\partial \mu^2}\ell\left(\mu,\sigma^2\right) = -\frac{n}{\sigma^2}, \quad \frac{\partial^2}{\partial \left(\sigma^2\right)^2}\ell\left(\mu,\sigma^2\right) = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}\left(X_i-\mu\right)^2$$

$$\frac{\partial^2}{\partial \mu \partial \sigma^2}\ell\left(\mu,\sigma^2\right) = \frac{\partial^2}{\partial \sigma^2 \partial \mu}\ell\left(\mu,\sigma^2\right) = -\frac{\sum_{i=1}^{n}\left(X_i-\mu\right)}{\sigma^4} = \frac{n\mu - n\bar{X}}{\sigma^4}.$$

Evaluando a H en $\left(\hat{\mu}, \hat{\sigma}^2\right)$ resulta:

$$\frac{\partial^2}{\partial \mu^2} \ell\left(\mu, \sigma^2\right)\Big|_{\left(\hat{\mu}, \hat{\sigma}^2\right)} = -\frac{n}{\hat{\sigma}^2}, \quad \frac{\partial^2}{\partial \left(\sigma^2\right)^2} \ell\left(\mu, \sigma^2\right)\Bigg|_{\left(\hat{\mu}, \hat{\sigma}^2\right)} = \frac{n}{2\hat{\sigma}^4} - \frac{n\hat{\sigma}^2}{\hat{\sigma}^6} = -\frac{n}{2\hat{\sigma}^4}$$

$$\frac{\partial^2}{\partial \mu \partial \sigma^2} \ell\left(\mu, \sigma^2\right)\Bigg|_{\left(\hat{\mu}, \hat{\sigma}^2\right)} = \frac{\partial^2}{\partial \sigma^2 \partial \mu} \ell\left(\mu, \sigma^2\right)\Bigg|_{\left(\hat{\mu}, \hat{\sigma}^2\right)} = \frac{n\hat{\mu} - n\hat{\mu}}{\hat{\sigma}^4} = 0.$$

que resulta una matriz diagonal y es fácil comprobar que es definida negativa pues $\hat{\sigma}^2 > 0$.

Invarianza de los E.M.V.

¿Cuál sería el EMV de σ en el caso anterior?

Invarianza de los E.M.V.

¿Cuál sería el EMV de σ en el caso anterior?

Sea $\lambda = h(\theta) : \Theta \rightarrow \Lambda$ una función *biyectiva*.

Luego, la densidad $f(\mathbf{x}, \theta)$ se puede expresar en función de λ ya que $\theta = h^{-1}(\lambda)$.
Notamos la densidad de $\mathbf{X}$ como función de λ por $f^*(\mathbf{x}, \lambda)$. Se tiene

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

Entonces,

$$L^*(\mathbf{x}, \lambda) = L(\mathbf{x}, h^{-1}(\lambda))$$

Luego, se definen los E.M.V. $\widehat{\theta}$ y $\widehat{\lambda}$ por

$$L(\mathbf{x}, \widehat{\theta}) = \max_{\theta \in \Theta} L(\mathbf{x}, \theta)$$

y

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

Propiedad de Invarianza del EMV

Teorema: *Si $\widehat{\theta}$ es E.M.V. de θ, entonces $\widehat{\lambda} = h(\widehat{\theta})$ es E.M.V. de λ.*

En efecto:

$$L(\mathbf{x}, \widehat{\theta}) = \max_{\theta \in \Theta} L(\mathbf{x}, \theta) \stackrel{(1)}{=} \max_{\lambda \in \Lambda} L(\mathbf{x}, h^{-1}(\lambda)) \stackrel{(2)}{=} \max_{\lambda \in \Lambda} L^*(\mathbf{x}, \lambda) = L^*(\mathbf{x}, \widehat{\lambda})$$

ya que

- (1): coinciden los conjuntos $\{\theta \in \Theta\}$ y $\{h^{-1}(\lambda) : \lambda \in \Lambda\}$
- (2): por definición de L^*.

Finalmente, como h es biyectiva $\widehat{\lambda} = h(\widehat{\theta})$.

Propiedad de Invarianza del EMV

Teorema: *Si $\widehat{\theta}$ es E.M.V. de θ, entonces $\widehat{\lambda} = h(\widehat{\theta})$ es E.M.V. de λ.*

En efecto:

$$L(\mathbf{x}, \widehat{\theta}) = \max_{\theta \in \Theta} L(\mathbf{x}, \theta) \overset{(1)}{=} \max_{\lambda \in \Lambda} L(\mathbf{x}, h^{-1}(\lambda)) \overset{(2)}{=} \max_{\lambda \in \Lambda} L^*(\mathbf{x}, \lambda) = L^*(\mathbf{x}, \widehat{\lambda})$$

ya que

- (1): coinciden los conjuntos $\{\theta \in \Theta\}$ y $\{h^{-1}(\lambda) : \lambda \in \Lambda\}$
- (2): por definición de L^*.

Finalmente, como h es biyectiva $\widehat{\lambda} = h(\widehat{\theta})$.

Ejemplo:

si $X_1, \ldots, X_n$ son i.i.d. $X_i \sim N(\mu, 1)$, tenemos que

$$\mathbb{P}\left[X_1 \leq x\right] = \mathbb{P}\left[X_1 - \mu \leq x - \mu\right] = \Phi(x - \mu),$$

luego como ϕ es estrictamente creciente y $\overline{X}$ es el EMV de μ, el estimador de máxima versosimilitud de esta probabilidad es

$$\Phi(x - \overline{X})$$

EMV

A veces las cosas se complican un poco....

- ya sea porque $L(\theta; \mathbf{x})$ no es derivable como función de θ:
  Caso $X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{U}[0, \theta]$: familia no regular.
  Notar que la indicadora de la densidad depende del parámetro.

  EMV de θ: $\widehat{\theta} = \max(X_1, \ldots, X_n)$

EMV

A veces las cosas se complican un poco....

- ya sea porque $L(\theta; \mathbf{x})$ no es derivable como función de θ:
  Caso $X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{U}[0, \theta]$: familia no regular.
  Notar que la indicadora de la densidad depende del parámetro.

  EMV de θ: $\widehat{\theta} = \max(X_1, \ldots, X_n)$
- o porque el óptimo no es tan fácil de hallar.

- Tarea: explorar caso $\Gamma(\alpha, \lambda)$

Ejemplo $U[0, \theta]$

$X_1 \dots X_n$, i.i.d., $X_i \sim \mathcal{U}[0, \theta]$. $f(x, \theta) = \frac{1}{\theta} \mathcal{I}_{[0,\theta]}(x)$.

$$f(x_1, \dots x_n, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathcal{I}_{[0,\theta]}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathcal{I}_{[0,\theta]}(x_i)$$

Ejemplo $U[0, \theta]$

$X_1 \dots X_n$, i.i.d., $X_i \sim \mathcal{U}[0, \theta]$. $f(x, \theta) = \frac{1}{\theta} \mathcal{I}_{[0,\theta]}(x)$.

$$f(x_1, \dots x_n, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathcal{I}_{[0,\theta]}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathcal{I}_{[0,\theta]}(x_i)$$

$$= \begin{cases} \dfrac{1}{\theta^n} & \text{si} \quad 0 \le x_i \le \theta \text{ para todo } i \\ 0 & \text{otro caso} \end{cases}$$

Ejemplo $U[0, \theta]$

$X_1 \dots X_n$, i.i.d., $X_i \sim \mathcal{U}[0, \theta]$. $f(x, \theta) = \frac{1}{\theta} \mathcal{I}_{[0,\theta]}(x)$.

$$f(x_1, \dots x_n, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathcal{I}_{[0,\theta]}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathcal{I}_{[0,\theta]}(x_i)$$

$$= \begin{cases} \dfrac{1}{\theta^n} & \text{si} \quad 0 \le x_i \le \theta \text{ para todo } i \\ 0 & \text{otro caso} \end{cases}$$

$$= \begin{cases} \dfrac{1}{\theta^n} & \text{si} \quad \theta \ge \max (x_i) \\ 0 & \text{otro caso} \end{cases}$$

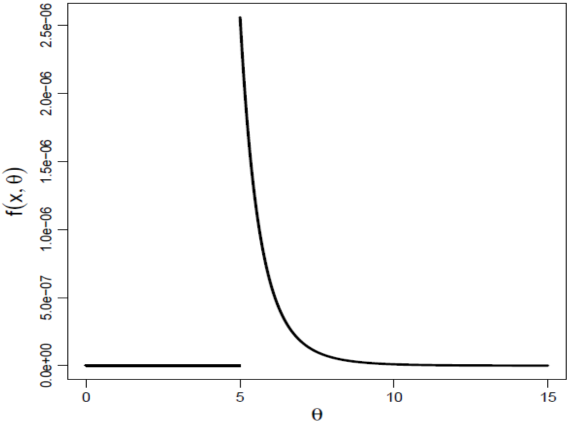
Estimador de Máxima Verosimilitud: Distribución Uniforme

(¡Gracias Marina V.!!)

Gráfico de la función de verosimilitud, $f(\mathbf{x}, \theta)$ correspondiente a $n = 8$ observaciones tales que

$$\max_{1 \leq i \leq n} (x_i) = 5$$

.



Estimador de Máxima Verosimilitud: Distribución Uniforme

El E.M.V. de θ es

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