Estimación puntual

Estimación

Supongamos que tenemos una muestra aleatoria de una distribución F_θ.

$X_1, \ldots, X_n \sim F_\theta \quad i.i.d,$

Interesa conocer *aproximadamente* $\boldsymbol{\theta}$.

Definición. Un *estimador puntual* de θ es una función de la muestra $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$.

Ejemplos

1- Se quiere conocer el peso medio de los paquetes de arroz producido por una fábrica. Para ello , se toman 30 cajas de arroz al azar y se las pesa. Se obtiene:

0.96	0.97	1.12	1.16	1.03	0.95	0.91	0.87	0.96	1.04
0.77	0.99	0.84	1.08	1.12	0.78	0.95	0.93	1.09	0.92
1.00	0.92	1.02	0.90	0.87	0.85	1.03	1.04	0.92	1.07

Suponemos que el peso de un paquete elegido al azar es una variable aleatoria $X \sim \mathcal{N}(\mu, \sigma^2)$.

2- Una fábrica de lámparas sabe que el tiempo de vida, en días, de las lámparas que fabrica, sigue una distribución $Exp(\lambda)$. Obtener una fórmula para estimar λ partir de una muestra aleatoria $X_1 \dots X_n$.

Un estimador es una variable aleatoria

- Antes de probar las lámparas no sabemos cuánto durará cada una. Así, la duración de la primera puede ser considerada una v.a. X_1, la segunda una v.a. X_2, etc.
- Cualquier función de la muestra es una v.a. por ejemplo: $\bar{X}, \tilde{X}, S^2, \max(X_1, \ldots, X_n)$, etc. Una vez observado el resultado del experimento los valores calculados serán denotados $\bar{x}, \tilde{x}, s^2, \max(x_1, \ldots, x_n)$, o $\bar{X}_{obs}, \tilde{X}_{obs}, S^2_{obs}, \max(X_1, \ldots, X_n)_{obs}$,

Estimación

- Un buen estimador deberá tener la propiedad de que cualquiera sea el valor de θ, que es desconocido,la diferencia $\hat{\theta} - \theta$ sea pequeña.

- Veremos dos métodos:
  - Método de los momentos
  - Método de máxima verosimilitud

Método de los momentos

- **Definición**: Sea X una v.a. Se denomina momento (poblacional) de orden k o k-ésimo momento de X a $E(X^k)$, si esa esperanza existe.
- **Definición:** Dada una muestra aleatoria $X_1, X_2, \ldots, X_n$, se denomina momento muestral de orden k o k-ésimo momento muestral de $X_1 \ldots X_n$ a

$$\frac{\sum_{i=1}^{n} X_i^k}{n}$$

Método de los momentos

Definición: Sea $X_1, X_2, \ldots, X_n$ una m.a. de una distribución con función de distribución F_θ (que depende de un parámetro θ)

- El estimador de momentos de θ basado en el primer momento es el valor $\hat{\theta}$ tal que

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

- El estimador de momentos de θ basado en el k-ésimo momento es el valor $\hat{\theta}$ tal que

$$\frac{\sum_{i=1}^{n} X_i^k}{n} = E_{\hat{\theta}}\left(X^k\right)$$

.

donde $E_{\hat{\theta}}(X) = E_{F_{\hat{\theta}}}(X)$.

Motivación: Por **LGN** $\overline{X} \approx E_\theta(X)$ para n grande.

Ejemplo 1: Normal

Se quiero conocer el peso medio de los paquetes de arroz producido por una fábrica. Para ello , se toma una muestra de 30 cajas de arroz y se las pesa. Se obtiene:

0.96	0.97	1.12	1.16	1.03	0.95	0.91	0.87	0.96	1.04
0.77	0.99	0.84	1.08	1.12	0.78	0.95	0.93	1.09	0.92
1.00	0.92	1.02	0.90	0.87	0.85	1.03	1.04	0.92	1.07

Suponemos que el peso de un paquete elegido al azar es una variable aleatoria $X \sim \mathcal{N}(\mu, \sigma^2)$

Al elegir n paquetes tenemos Sea $X_1, X_2, \ldots, X_n$ i.i.d. $X_i \sim \mathcal{N}(\mu, \sigma^2)$

Estimador de los momentos de μ de orden 1:

$$\frac{1}{n}\sum_{i=1}^{n} X_i = E_{\hat{\mu}}(X) = \hat{\mu}.$$

Con estos datos la estimación que se obtiene es $\widehat{\mu}_{obs} = 0.97$

Ejemplo 2: Exponencial

Una fábrica de lámparas sabe que el tiempo de vida, en días, de las lámparas que fabrica, sigue una distribución $Exp(\theta)$. Obtener una fórmula para estimar θ partir de una muestra aleatoria $X_1 \dots X_n$.

$$X_1, X_2, \dots, X_n \sim Exp(\theta)$$

Ejemplo 2: Exponencial

Una fábrica de lámparas sabe que el tiempo de vida, en días, de las lámparas que fabrica, sigue una distribución $Exp(\theta)$. Obtener una fórmula para estimar θ partir de una muestra aleatoria $X_1 \dots X_n$.

$$X_1, X_2, \dots, X_n \sim Exp(\theta)$$

Para hallar el estimador de momentos de θ, hay que despejar $\hat{\theta}$ de

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

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{\widehat{\theta}} \Rightarrow \widehat{\theta} = \frac{1}{\hat{X}}$$

Ejemplo 3: Distribución Uniforme

$$X_1, X_2, \ldots, X_n \sim \mathcal{U}(0, \theta).$$

Ejemplo 3: Distribución Uniforme

$$X_1, X_2, \ldots, X_n \sim \mathcal{U}(0, \theta).$$

Hay que despejar θ de

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

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\widehat{\theta}}{2} \Rightarrow \widehat{\theta} = \frac{1}{2n}\sum_{i=1}^{n} X_i.$$

Estimador de los momentos: Generalización cuando hay varios parámetros

Definición: Sea $X_1, X_2, \ldots, X_n$ una m.a. de una distribución que depende de m parámetros $\theta_1, \theta_2, \ldots., \theta_m$. Los estimadores de momentos de $\theta_1, \theta_2, \ldots., \theta_m$ son los valores $\hat{\theta}_1, \hat{\theta}_2, \ldots., \hat{\theta}_m$ que se obtienen igualando m momentos poblacionales con los correspondientes momentos muestrales. En general, se obtienen resolviendo el siguiente sistema de ecuaciones

$$\frac{\sum_{i=1}^{n} X_i^k}{n} = E\left(X^k\right) \quad k = 1, 2, \ldots, m$$

Ejemplo 4: Estimación de ambos parámetros de la normal

Supongamos que tenemos una muestra aleatoria $X_1, \dots X_n \sim \mathcal{N}(\mu, \sigma^2)$.

Se tiene $\mathbb{E}_{\mu,\sigma^2}(X) = \mu$ y $\mathbb{E}_{\mu,\sigma^2}(X^2) = \mu^2 + \sigma^2$.

Para encontrar el estimador de momentos de μ y σ^2 hay que resolver el sistema

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}_{\hat{\mu},\hat{\sigma}^2}(X_1) = \widehat{\mu}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mathbb{E}_{\hat{\mu},\hat{\sigma}^2}(X_1^2) = \widehat{\mu}^2 + \widehat{\sigma}^2$$

Ejemplo 3: Estimación de ambos parámetros de la normal

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} X_i = \widehat{\mu} \\ \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \widehat{\mu}^2 + \widehat{\sigma}^2 \end{cases}$$

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2$$

Máxima verosimilitud

Caso discreto

- Supongamos que tenemos una muestra aleatoria de una distribución discreta F_θ: $X_1, \ldots, X_n \sim F_\theta$.
- $\mathbf{x}= x_1, \cdots, x_n$ realización de $X_1, \cdots, X_n$ i.i.d.
- Función de verosimilitud asociada a $\mathbf{x}= x_1, \cdots, x_n$:

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

$$L(\theta\,;\,\mathbf{x}\,)\ =\ \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n)\ , X_i \sim p(\cdot, \theta)\ .$$

$$L(\theta\,;\,\mathbf{x}\,)\ =\ \prod_{i=1}^{n} p(x_i, \theta)\ ,$$

Nos interesa ver cuán **verosímil** es que un determinado parámetro haya generado los datos.

Máxima verosimilitud

Caso discreto

- Supongamos que tenemos una muestra aleatoria de una distribución discreta F_θ: $X_1, \ldots, X_n \sim F_\theta$.

- $\mathbf{x}= x_1, \cdots, x_n$ realización de $X_1, \cdots, X_n$ i.i.d.

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

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

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

Nos interesa ver cuán **verosímil** es que un determinado parámetro haya generado los datos.

- Propuesta de máxima verosimilitud:

$$\hat{\theta}(\mathbf{x}) = \underset{\theta \in \Theta}{\operatorname{argmax}}\ L(\theta\ ;\ \mathbf{x}\ )\ .$$

o sea

Máxima Verosimilitud

Caso continuo

- Supongamos que tenemos una muestra aleatoria de una distribución continua F_θ: $X_1, \ldots, X_n \sim F_\theta$.

- $\mathbf{x} = x_1, \cdots, x_n$ realización correspondiente a $X_1, \ldots, X_n$

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

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

- Propuesta de Máxima Verosimilitud:

$$
\hat{\theta}(\mathbf{x}) = \underset{\theta\in\Theta}{\operatorname{argmax}}\; L(\theta\,;\,\mathbf{x}\,)\;.
$$

o sea $\widehat{\theta}_n = h_n(X_1, \cdots, X_n)$.

Estimador de Máxima verosimilitud

Caso discreto y continuo

$X_1, \ldots, X_n \sim F_\theta,$

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

Estimador de máxima verosimilitud de θ:

$$\hat{\theta} = \hat{\theta}(\mathbf{X}) = \underset{\theta \in \Theta}{\text{argmax}} \; L(\theta \, ; \, \mathbf{X}) \, .$$

Recordar:

- $\hat{\theta}(\mathbf{X})$ es una variable aleatoria
- $\hat{\theta}(\mathbf{X})_{obs} = \hat{\theta}(\mathbf{x})$ es una realización, i.e. un número.

Ejemplo $\mathcal{E}(\lambda)$: $f(x,\lambda) = \lambda e^{-x\lambda}\,\mathcal{I}_{(0,\infty)}(x)$.

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{E}(\lambda), \lambda > 0$

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

- si $x_i > 0\ \forall i$

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

- 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

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \text{punto crítico es } 1/\bar{x}_n \ , \quad \text{ver que maximiza!}$$

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

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{N}(\mu, 9)$, $f\,(x, \mu, 9) = \frac{1}{\sqrt{2\pi}}\ \frac{1}{3}\ e^{-\frac{1}{2}\frac{(x-\mu)^2}{9}}$

$$L(\mu, 9; \mathbf{x}) \quad = \quad \prod_{i=1}^{n} f(x_i, \mu, 9)$$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{N}(\mu, 9)$, $f(x, \mu, 9) = \frac{1}{\sqrt{2\pi}} \; \frac{1}{3} \; e^{-\frac{1}{2}\frac{(x-\mu)^2}{9}}$

$$L(\mu, 9; \mathbf{x}) \quad = \quad \prod_{i=1}^{n} f(x_i, \mu, 9) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \; \frac{1}{3} \; e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{9}}$$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{N}(\mu, 9)$, $f(x, \mu, 9) = \frac{1}{\sqrt{2\pi}} \frac{1}{3} e^{-\frac{1}{2}\frac{(x-\mu)^2}{9}}$

$$
\begin{aligned}
L(\mu, 9; \mathbf{x}) &= \prod_{i=1}^{n} f(x_i, \mu, 9) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \frac{1}{3} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{9}} \\
L(\mu, 9; \mathbf{x}) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{3}\right)^n e^{-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{9}}
\end{aligned}
$$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{N}(\mu, 9)$, $f(x, \mu, 9) = \frac{1}{\sqrt{2\pi}} \frac{1}{3} e^{-\frac{1}{2}\frac{(x-\mu)^2}{9}}$

$$L(\mu, 9; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{3}\right)^n e^{-\frac{1}{2}\frac{\sum_{i=1}^n (x_i-\mu)^2}{9}}$$

- Tomemos logaritmo

$$\ell(\mu, 9; \mathbf{x}) = cte - \frac{1}{2}\frac{\sum_{i=1}^n (x_i-\mu)^2}{9}$$

Maximizar a $\ell(\mu, 9; \mathbf{x})$ como función de μ equivale a minimizar

$$h(\mu) = \sum_{i=1}^n (x_i - \mu)^2$$

Un par de clases atrás, vimos que $h(\mu)$ se minimiza en $\bar{x}_n$

EMV de μ : $\widehat{\mu} = \bar{X}_n$

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

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

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

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

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

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

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

- 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

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

Ejemplo $U(0,\theta)$: $f(x,\theta) = \frac{1}{\theta}\mathcal{I}_{(0,\theta)}(x)$.

$X_1 \dots X_n \sim \mathcal{U}(0,\theta)$.

$$f(x_1, \dots x_n, \theta) = \Pi_{i=1}^{n} \frac{1}{\theta} \mathbb{I}_{(0,\theta)}(x_i) = \frac{1}{\theta^n} \Pi_{i=1}^{n} \mathbb{I}_{(0,\theta)}(x_i)$$

Ejemplo $U(0,\theta)$: $f\,(x,\theta) = \frac{1}{\theta}\mathcal{I}_{(0,\theta)}(x)$.

$X_1 \ldots X_n \sim \mathcal{U}(0,\theta)$.

$$f(x_1,\ldots x_n,\theta) = \Pi_{i=1}^{n}\frac{1}{\theta}\mathbb{I}_{(0,\theta)}(x_i) = \frac{1}{\theta^n}\Pi_{i=1}^{n}\mathbb{I}_{(0,\theta)}(x_i)$$

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

Ejemplo $U(0,\theta)$: $f\,(x,\theta) = \frac{1}{\theta}\mathcal{I}_{(0,\theta)}(x)$.

$X_1 \dots X_n \sim \mathcal{U}(0,\theta)$.

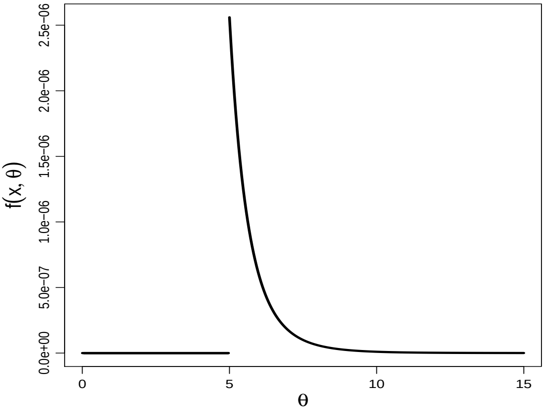
$$f(x_1,\dots x_n,\theta) = \Pi_{i=1}^{n}\frac{1}{\theta}\mathbb{I}_{(0,\theta)}(x_i) = \frac{1}{\theta^n}\Pi_{i=1}^{n}\mathbb{I}_{(0,\theta)}(x_i)$$

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

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

Estimador de máxima verosimilitud: Distribución Uniforme

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



El E.M.V. de θ es

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

Propiedad de Invarianza de los EMV:

Sea $\hat{\theta}$ el EMV de θ y sea h una función inyectiva con dominio en el rango de valores posibles de θ, entonces el EMV de $h(\theta)$ es $h(\hat{\theta})$. Por ejemplo, en el caso de una m.a. de una distribución $\mathrm{N}\left(\mu, \sigma^2\right)$ hemos visto que el EMV de σ^2 es

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n}$$

entonces el EMV de σ es

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n}}$$

pues la función $h(x) = \sqrt{x}$, $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$ es inyectiva.