Intervalos y regiones de confianza

Estadística (M)

Regiones de confianza

Dado un vector $\mathbf{X}$ con distribución perteneciente a la familia $F(\mathbf{x}, \boldsymbol{\theta})$ con $\boldsymbol{\theta} \in \Theta$, *una región de confianza* $\mathcal{S}(\mathbf{X})$ *para* $\boldsymbol{\theta}$ *con nivel de confianza* $1 - \alpha$ será una función que a cada $\mathbf{X}$ le asigne un subconjunto $\mathcal{S}(\mathbf{X}) \subset \Theta$ de manera que

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in \mathcal{S}(\mathbf{X})) = 1 - \alpha, \qquad \forall \boldsymbol{\theta} \in \Theta$$

Es decir, $\mathcal{S}(\mathbf{X})$ cubre el valor verdadero del parámetro con probabilidad $1 - \alpha$.

Caso Particular: Intervalos de confianza

Caso particular: Si $\boldsymbol{\theta} \in \mathbb{R}$ se dirá que $\mathcal{S}(\mathbf{X})$ es un intervalo de confianza cuando

$$\mathcal{S}(\mathbf{X}) = [a(\mathbf{X}), b(\mathbf{X})]$$

Caso Particular: Intervalos de confianza

Caso particular: Si $\boldsymbol{\theta} \in \mathbb{R}$ se dirá que $\mathcal{S}(\mathbf{X})$ es un intervalo de confianza cuando

$$\mathcal{S}(\mathbf{X}) = [a(\mathbf{X}), b(\mathbf{X})]$$

La longitud de $\mathcal{S}(\mathbf{X})$ es

$$l = b(\mathbf{X}) - a(\mathbf{X})$$

Procedimiento general: Método del Pivote

Sea $\mathbf{X}$ un vector aleatorio cuya distribución pertenece a la familia $F(\mathbf{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$.

Una función $G(\mathbf{X}, \boldsymbol{\theta})$ se llama un pivote si y sólo si la distribución de $G(\mathbf{X}, \boldsymbol{\theta})$ no depende de $\boldsymbol{\theta}$.

Procedimiento general: Método del Pivote

Sea $\mathbf{X}$ un vector aleatorio cuya distribución pertenece a la familia $F(\mathbf{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$.

Una función $G(\mathbf{X}, \boldsymbol{\theta})$ se llama un pivote si y sólo si la distribución de $G(\mathbf{X}, \boldsymbol{\theta})$ no depende de $\boldsymbol{\theta}$.

Teorema *Sea $\mathbf{X}$ un vector aleatorio cuya distribución pertenece a la familia $F(\mathbf{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. Sea*

- $U = G(\mathbf{X}, \boldsymbol{\theta})$ *una variable aleatoria cuya distribución es independiente de $\boldsymbol{\theta}$.*
- A *y* B *tales que* $\mathbb{P}(A \leq U \leq B) = 1 - \alpha$.

Luego, si $\mathcal{S}(\mathbf{X}) = \{\boldsymbol{\theta} : A \leq G(\mathbf{X}, \boldsymbol{\theta}) \leq B\}$, $\mathcal{S}(\mathbf{X})$ es una región de confianza a nivel $(1 - \alpha)$ para $\boldsymbol{\theta}$.

Intervalos para la media μ y la varianza de una normal

Buscando un pivote....

Intervalos para la media μ y la varianza de una normal

Buscando un pivote....

Teorema *Sean* $X_1, \ldots, X_n$ *variables aleatorias independientes donde* $X_i \sim N(\mu, \sigma^2)$*. Luego*

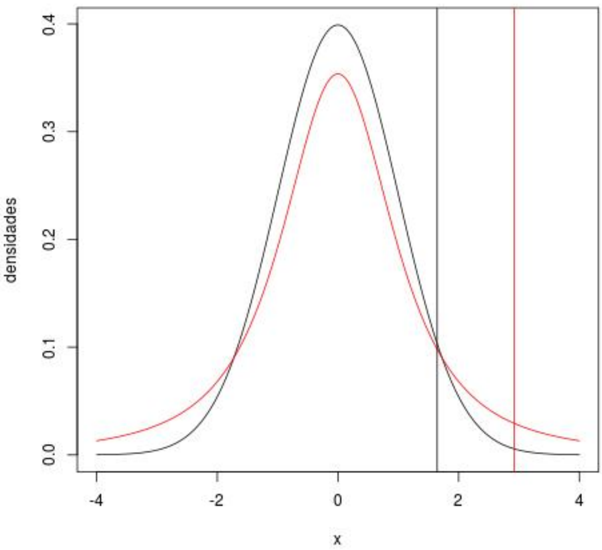
a) $U = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim N(0,1)$, *con* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

b) $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ *es independiente de* $\bar{X}_n$.

c) $V = \frac{(n-1)s_n^2}{\sigma^2} \sim \chi^2_{n-1}$.

d) $W = \sqrt{n}\frac{(\bar{X}_n - \mu)}{s_n} \sim \mathcal{T}_{n-1}$.

Normal vs. t2: mirando percentiles...



Intervalo para la media μ de una normal con varianza desconocida

Sea T una v.a. tal que $T \sim \mathcal{T}_m$. Dado $0 < \eta < 1$ llamamos $t_{m,\eta}$ al punto que satisface:

$$\mathbb{P}(V > t_{m,\eta}) = \eta$$

Intervalo para la media μ de una normal con varianza desconocida

Sea T una v.a. tal que $T \sim \mathcal{T}_m$. Dado $0 < \eta < 1$ llamamos $t_{m,\eta}$ al punto que satisface:

$$\mathbb{P}(V > t_{m,\eta}) = \eta$$

Teorema *Sean $X_1, \ldots, X_n$ variables aleatorias independientes donde $X_i \sim N(\mu, \sigma^2)$ con μ y σ desconocidas. Luego, si*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \text{ y } s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2,$$

tenemos que

$$\left[\bar{X}_n - t_{n-1,\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}}, \bar{X}_n + t_{n-1,\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}}\right]$$

es un intervalo de confianza para μ de nivel $1 - \alpha$.

Intervalo para la media μ de una normal con varianza desconocida

Hagamos unas pruebas en R.

Intervalo para la varianza σ^2 de una normal

Sea V una v.a. tal que $V \sim \chi^2_m$. Dado $0 < \eta < 1$ llamamos $\chi^2_{m,\eta}$ al punto que satisface:

$$\mathbb{P}(V > \chi^2_{m,\eta}) = \eta$$

Intervalo para la varianza σ^2 de una normal

Sea V una v.a. tal que $V \sim \chi^2_m$. Dado $0 < \eta < 1$ llamamos $\chi^2_{m,\eta}$ al punto que satisface:

$$\mathbb{P}(V > \chi^2_{m,\eta}) = \eta$$

Teorema *Sean $X_1, \ldots, X_n$ variables aleatorias independientes donde $X_i \sim N(\mu, \sigma^2)$ con σ desconocida. Luego, sean β y γ son dos números positivos tales que $\beta + \gamma = \alpha$.*

a) *[Ejercicio] Si μ es conocida un intervalo de confianza de nivel $1 - \alpha$ para σ^2 es*

$$\left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n,\beta}}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n,\gamma}} \right).$$

b) *Si μ es desconocida un intervalo de confianza de nivel $1 - \alpha$ para σ^2 es*

$$\left(\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\chi^2_{n-1,\beta}}, \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\chi^2_{n-1,\gamma}} \right).$$

Mundo Normal

- $X_i \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d. Buscamos intervalo de confianza para μ.
- σ conocido: IC nivel $1-\alpha$ para μ

$$\left(\bar{X}_n - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} \quad , \quad \bar{X}_n + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right)$$

- longitud $\rightarrow l = 2\; z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}$

Mundo Normal

- $X_i \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d. Buscamos intervalo de confianza para μ.
- σ conocido: IC nivel $1-\alpha$ para μ

$$\left(\bar{X}_n - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} \quad , \quad \bar{X}_n + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} \right)$$

- longitud $\rightarrow l = 2\; z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} \leq l_o$

Mundo Normal

- $X_i \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d. Buscamos intervalo de confianza para μ.
- σ conocido: IC nivel $1-\alpha$ para μ

$$\left(\bar{X}_n - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} \quad , \quad \bar{X}_n + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} \right)$$

- longitud $\rightarrow l = 2\, z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}} \leq l_o$
- Si queremos $l \leq l_o \rightarrow \frac{4 z_{\alpha/2}^2 \sigma^2}{l_o^2} \leq n$

Mundo Normal

- $X_i \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d. Buscamos intervalo de confianza para μ.
- σ desconocido .....llegó la t...
- IC nivel $1-\alpha$ para μ

$$\left(\bar{X}_n - t_{n-1,\alpha/2} \, \frac{s_n}{\sqrt{n}} \quad , \quad \bar{X}_n + t_{n-1,\alpha/2} \, \frac{s_n}{\sqrt{n}} \right)$$

- longitud $\rightarrow L = 2 \, t_{n-1,\alpha/2} \, \frac{s_n}{\sqrt{n}}$

Mundo Normal

- $X_i \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d. Buscamos intervalo de confianza para μ.
- σ desconocido .....llegó la t...
- IC nivel $1-\alpha$ para μ

$$\left(\bar{X}_n - t_{n-1,\alpha/2}\,\frac{s_n}{\sqrt{n}} \quad , \quad \bar{X}_n + t_{n-1,\alpha/2}\,\frac{s_n}{\sqrt{n}}\right)$$

- longitud $\rightarrow L = 2\, t_{n-1,\alpha/2}\, \frac{s_n}{\sqrt{n}}$
- L es una v.a. Si queremos que $l \leq l_o$ ¿Cómo hacemos?

Intervalo de longitud prefijada para μ en una $N(\mu, \sigma^2)$

Propuesta

- Tomemos una muestra inicial $X_1, \ldots, X_n$.
- Estimamos σ^2 por

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

con

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Sea m tal que

$$\frac{2\ s_n\ t_{\frac{\alpha}{2}, n-1}}{\sqrt{n+m}} \leq l_o$$

- m es una variable aleatoria

Intervalo de longitud prefijada para μ en una $N(\mu, \sigma^2)$

- Sea $X_{n+1}, \ldots, X_{n+m}$ una muestra complementaria y

$$\bar{X}_{n+m} = \frac{1}{n+m} \sum_{i=1}^{n+m} X_i$$

El intervalo de confianza

$$[\bar{X}_{n+m} - t_{\frac{\alpha}{2},n-1} \frac{s_n}{\sqrt{n+m}}, \bar{X}_{n+m} + t_{\frac{\alpha}{2},n-1} \frac{s_n}{\sqrt{n+m}}]$$

tiene longitud menor o igual a l_o.

Probemos que la propuesta nos lleva a un IC de nivel $1-\alpha$.

Intervalo de longitud prefijada para μ en una $N(\mu, \sigma^2)$

Sean $X_1, ...X_n$ variables aleatorias independientes con distribución N(μ, σ^2) y sea m como antes.

(i) $W = (n-1)s_n^2/\sigma^2 \sim \chi^2_{n-1}$

(ii) $V = \sqrt{m+n}(\bar{X}_{m+n} - \mu)/\sigma \sim N(0,1)$

(iii) V y W son independientes

(iv) $\sqrt{m+n}(\bar{X}_{m+n} - \mu)/s_n \sim \mathcal{T}_{n-1}$

Luego,

$$\left[\bar{X}_{n+m} - t_{\frac{\alpha}{2}, n-1}\frac{s_n}{\sqrt{n+m}}, \bar{X}_{n+m} + t_{\frac{\alpha}{2}, n-1}\frac{s_n}{\sqrt{n+m}}\right]$$

es un intervalo de confianza para μ de nivel $1-\alpha$ con longitud menor o igual a l_o.

Regiones de confianza con nivel asintótico $(1-\alpha)$

Sea $X_1, X_2, \ldots, X_n$ m.a. $X_i \sim F(x, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. Se dice que $S_n(X_1, \ldots, X_n)$ es una sucesión de regiones de confianza con nivel asintótico $1-\alpha$ si:

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in S_n(X_1, \ldots, X_n)) = 1-\alpha \quad \forall\, \theta \in \Theta\ .$$

Procedimiento para obtener RC con nivel asintótico

Teorema Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución perteneciente a la familia $F(x, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. Supongamos que

- $\forall n$, $\exists$ v.a. $U_n = G_n(X_1, \ldots, X_n, \boldsymbol{\theta})$ tales que $U_n \xrightarrow{D} U$, donde U es una variable aleatoria con distribución independiente de $\boldsymbol{\theta}$
- A y B puntos de continuidad de F_U tales que $\mathbb{P}(A \leq U \leq B) = 1 - \alpha$.

Luego, si

$$\mathcal{S}_n(X_1, \ldots, X_n) = \{\boldsymbol{\theta} : A \leq G_n(X_1, \ldots, X_n, \boldsymbol{\theta}) \leq B\}$$

$\mathcal{S}_n(\mathbf{X})$ es una sucesión de RC con nivel asintótico $(1 - \alpha)$.

Ejemplos

- Consideremos el caso en que $X_1, \ldots, X_n$ son una muestra aleatoria $\mathbb{E}(X_i) = \mu$ y $\mathbb{V}(X_i) = \sigma^2$, ambas desconocidas. Veremos que el intervalo

$$[\bar{X}_n - z_{\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}}, \bar{X}_n + z_{\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}}]$$

  tiene nivel asintótico $1 - \alpha$.

Vayamos al pizarrón.

Ejemplos

- Consideremos el caso en que $X_1, \ldots, X_n$ son una muestra aleatoria donde $x_i \sim Bi(1,p)$
  Vamos a deducir dos intervalos de confianza asintóticos diferentes para p.
- $[\widehat{p}_{1,n}, \widehat{p}_{2,n}]$ raíces del polinomio en p

$$n\bar{X}_n^2 - p(2n\bar{X}_n + z_{\frac{\alpha}{2}}^2) + p^2(z_{\frac{\alpha}{2}}^2 + n)$$

- $$[\bar{X}_n - z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}]$$

Vayamos al pizarrón.

Usando la distribución asintótica de los EMV

$X_1, \ldots, X_n$ i.i.d. donde X_i tienen función de densidad o de probabilidad puntual $f(x, \theta)$.

Bajo condiciones de regularidad

- si $\widehat{\theta}_n = \widehat{\theta}_n^{EMV}$, entonces $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{D} N(0, \frac{1}{I_1(\theta)})$
- $$\sqrt{n}\sqrt{I_1(\theta)}(\widehat{\theta}_n - \theta) \xrightarrow{D} N(0, 1)$$
- La región

$$S(\mathbf{X}) = \{\theta : -z_{\frac{\alpha}{2}} \leq \sqrt{n}\sqrt{I_1(\theta)}(\widehat{\theta}_n - \theta) \leq z_{\frac{\alpha}{2}}\}$$

no tiene porqué ser un intervalo y puede ser difícil de calcular.

Usando la distribución asintótica de los EMV

- Si $I_1(\theta)$ es una función continua de θ, como $\widehat{\theta}_n \xrightarrow{p} \theta$, bajo condiciones de regularidad obtendremos que $I_1(\widehat{\theta}_n) \xrightarrow{p} I_1(\theta)$, luego

$$\sqrt{n}\sqrt{I_1(\widehat{\theta}_n)}(\widehat{\theta}_n - \theta) \xrightarrow{D} N(0,1)$$

- Entonces, un intervalo de nivel asintótico $1-\alpha$ será

$$[\widehat{\theta}_n - z_{\frac{\alpha}{2}}\sqrt{\frac{1}{n\ I_1(\widehat{\theta}_n)}}, \widehat{\theta}_n + z_{\frac{\alpha}{2}}\sqrt{\frac{1}{n\ I_1(\widehat{\theta}_n)}}]$$

Otra mirada...

POBLACION $\leftrightarrow F$	MUESTRA $X_1, \ldots X_n$ i.i.d. $X_i \sim F$
Parámetro: Valor asociado de F $\theta = \theta(F)$ θ: valor poblacional	

Funcionales

$$\mathbb{E}_F(X) \quad \Longleftrightarrow \quad \mathbb{E}(X) \text{ cuando } X \sim F.$$

Funcionales

$$\mathbb{E}_F(X) \quad \Longleftrightarrow \quad \mathbb{E}(X) \text{ cuando } X \sim F.$$

$$\mathbb{E}_F(X) =$$

Funcionales

$$\mathbb{E}_F(X) \quad \Longleftrightarrow \quad \mathbb{E}(X) \text{ cuando } X \sim F.$$

$$\mathbb{E}_F(X) = \mu(F) = \int x \, dF(x)$$

Funcionales

$$\mathbb{E}_F(X) \quad \Longleftrightarrow \quad \mathbb{E}(X) \text{ cuando } X \sim F.$$

$$\mathbb{E}_F(X) = \mu(F) = \int x \, dF(x)$$

$$\mathbb{V}_F(X) = \sigma(F) = \int (x - \mu)^2 \, dF(x)$$

Funcionales

$$\mathbb{E}_F(X) \quad \Longleftrightarrow \quad \mathbb{E}(X) \text{ cuando } X \sim F.$$

$$\mathbb{E}_F(X) = \mu(F) = \int x \, dF(x)$$

$$\mathbb{V}_F(X) = \sigma(F) = \int (x - \mu)^2 \, dF(x)$$

Consideramos funcionales de la forma

$$T(F) = \int r(x) \, dF(x)$$

Otra mirada...

POBLACION $\leftrightarrow F$	MUESTRA $X_1, \ldots X_n$ i.i.d. $X_i \sim F$
Parámetro: Valor asociado de F	Estimador:estadístico para estimar θ
$\theta = \theta(F)$	$\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$
θ: valor poblacional	$\widehat{\theta}_n$ NUEVA VARIABLE ALEATORIA

La Distribución Empírica

Sean $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim F$. Definimos

$$\widehat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq t\}}$$

- $\widehat{F}_n(t)$ es una función aleatoria.
- $\widehat{F}_n(t)$ representa a una acumulada que da peso $1/n$ a $X_1, X_2, \ldots, X_n$.

La Distribución Empírica

$$\mathbb{P}_F(X \in A) \iff \mathbb{P}(X \in A) \text{ cuando } X \sim F.$$

$$\mathbb{E}_F(X) \iff \mathbb{E}(X) \text{ cuando } X \sim F.$$

$$\mathbb{E}_F\left\{g(X_1, \ldots, X_n)\right\} \iff \mathbb{E}\left\{g(X_1, \ldots, X_n)\right\} \ ,$$
$$\text{cuando } X_1, \ldots, X_n \text{ i.i.d., } X_i \sim F.$$

t	X_1	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	X_n
$p(t)$	$1/n$	$1/n$	$1/n$	$1/n$	$1/n$	$1/n$	$1/n$	$1/n$

$$? \quad \mathbb{P}_{\widehat{F}_n}(A) = \ldots\ldots$$

$$? \quad \mathbb{E}_{\widehat{F}_n}(X) = \ldots\ldots$$

$$? \quad \mathbb{V}_{\widehat{F}_n}(X) = \ldots\ldots$$

La Distribución Empírica

Sean $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim F$. Definimos

$$\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq t\}}$$

- $\widehat{F}_n(t)$ es una función aleatoria.
- $\widehat{F}_n(t)$ representa a una acumulada que da peso $1/n$ a $X_1, X_2, \ldots, X_n$.
- Ley de los Grandes Números:

$$\widehat{F}_n(t) \xrightarrow{p} F(t)$$

La Distribución Empírica

Sean $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim F$. Definimos

$$\widehat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq t\}}$$

- $\widehat{F}_n(t)$ es una función aleatoria.
- $\widehat{F}_n(t)$ representa a una acumulada que da peso $1/n$ a $X_1, X_2, \ldots, X_n$.
- Ley de los Grandes Números:

$$\widehat{F}_n(t) \xrightarrow{p} F(t)$$

- $\mathbb{E}(\widehat{F}_n(t)) = F(t)$
  $\mathbb{V}(\widehat{F}_n(t)) = \frac{F(t)(1-F(t)}{n}$

La Distribución Empírica

Sean $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim F$. Definimos

$$\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq t\}}$$

- $\widehat{F}_n(t)$ es una función aleatoria.
- $\widehat{F}_n(t)$ representa a una acumulada que da peso $1/n$ a $X_1, X_2, \ldots, X_n$.
- Ley de los Grandes Números:

$$\widehat{F}_n(t) \xrightarrow{p} F(t)$$

La Distribución Empírica

Sean $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim F$. Definimos

$$\widehat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq t\}}$$

- $\widehat{F}_n(t)$ es una función aleatoria.
- $\widehat{F}_n(t)$ representa a una acumulada que da peso $1/n$ a $X_1, X_2, \ldots, X_n$.
- Ley de los Grandes Números:

$$\widehat{F}_n(t) \xrightarrow{p} F(t)$$

- Glivenko Cantelli:

$$\lim_{n \to \infty} \sup_{t \in \mathbb{R}} |\widehat{F}_n(t) - F(t)| = 0 \ , \quad \text{c.s.}$$

Estimación Plug-in

Nos interesa estimar un funcional $T(F)$.

El procedimiento **plug-in** propone estimar $\theta = T(F)$ con

$$\widehat{\theta}_n = T(\widehat{F}_n)$$

Aproximación de la distribución de un estimador

- Laboratorio de Física I con 30 mesas.
- Cada grupo realiza el mismo experimento, obtiene n datos y calcula la estimación utilizando sus datos (todos mismo n).
- Cada grupo obtiene una realización del estimador.

$$\hat{\theta}_n^1, \ldots, \hat{\theta}_n^{30}$$

- Cambio de notación. Desaparece n.

$$\hat{\theta}_1, \ldots, \hat{\theta}_{30}$$

- Aproximamos la distribución del estimador.

Apoximación de la distribución de un estimador (si pudiéramos...)

$$\hat{\theta}_n^1, \ldots, \hat{\theta}_n^{N_{simu}}, \quad X_i \sim F$$

Apoximación de la distribución de un estimador (si pudiéramos...)

$$\hat{\theta}_n^1, \ldots, \hat{\theta}_n^{N_{simu}}\,, \quad X_i \sim F$$

$$\hat{\theta}_1, \ldots, \hat{\theta}_{N_{simu}}\,, \quad X_i \sim F$$

Aproximación Bootstrap de la distribución de un estimador

$$X_1 \ldots, X_n \quad X_i \sim F \longrightarrow \hat{\theta}_1, \ldots, \hat{\theta}_{N_{simu}}$$

Aproximación Bootstrap de la distribución de un estimador

$$X_1 \ldots, X_n \quad X_i \sim F \longrightarrow \hat{\theta}_1, \ldots, \hat{\theta}_{N_{simu}}$$

$x_1 \ldots, x_n$: datos originales, realizaciones de $X_i \sim F \longrightarrow \widehat{F}_n$:

distribución construida con los datos originales que aproxima a F

$$X_1^* \ldots, X_n^* \quad X_i^* \sim \widehat{F}_n \longrightarrow \hat{\theta}^*$$

Aproximación Bootstrap de la distribución de un estimador

$$X_1 \ldots, X_n \quad X_i \sim F \longrightarrow \hat{\theta}_1, \ldots, \hat{\theta}_{N_{simu}}$$

$x_1 \ldots, x_n$: datos originales, realizaciones de $X_i \sim F \longrightarrow \widehat{F}_n$:

distribución construida con los datos originales que aproxima a F

$$X_1^* \ldots, X_n^* \quad X_i^* \sim \widehat{F}_n \longrightarrow \hat{\theta}^*$$

$$\hat{\theta}_1^*, \ldots, \hat{\theta}_{N_{simu}}^* \,, \quad X_i^* \sim \widehat{F}_n$$

$$F \longrightarrow \mathbf{X} \longrightarrow \widehat{\theta}$$

$$F \longrightarrow \mathbf{X} \longrightarrow \widehat{\theta}$$

$$\widehat{F}_n \longrightarrow \mathbf{X}^\star \longrightarrow \widehat{\theta}^\star$$

Intervalo de confianza para la media

- $\mu := T_1(F) = \mathbb{E}_F(X)$. Estimador plug-in:

$$\widehat{\mu}_n = T_1(\widehat{F}_n) = \mathbb{E}_{\widehat{F}_n}(X) = \bar{X}_n$$

- Distribución de $\widehat{\mu}_n$: asintóticamente normal

$$\frac{\widehat{\mu}_n - \mu}{\mathsf{se}(\widehat{\mu}_n)} \approx \mathcal{N}(0,1) \quad n \text{ grande}$$

- Desvío del Estimador:

$$\mathsf{se}(\widehat{\mu}) = \sqrt{\mathbb{V}_F(\widehat{\mu})} = \sqrt{\frac{\sigma^2}{n}} = \mathsf{se}$$

se estima con $\widehat{\mathsf{se}} = \sqrt{\frac{\widehat{\sigma}^2}{n}}$ o con $\widehat{\mathsf{se}} = \sqrt{\frac{S^2}{n}}$

$$\text{Intervalo de confianza} \quad \widehat{\mu} \pm z_{\alpha/2}\widehat{\mathsf{se}}$$

¿Intervalo de confianza para la mediana?

- Distribución de la mediana muestral: asintóticamente normal

$$\frac{\mathsf{med}(X_1, \dots, X_n) - \mathsf{med}(X)}{\mathsf{se}} \approx \mathcal{N}(0,1) \quad n \text{ grande}$$

- Desvío del Estimador:

$$\mathsf{se} = \mathsf{se}(\mathsf{med}(X_1, \dots, X_n)) = \sqrt{\mathbb{V}_F\{\mathsf{med}(X_1, \dots, X_n)\}} = ???$$

- $\widehat{\mathsf{se}} = ??$
- Bootstrap! $\widehat{\mathsf{se}}_{boot}$

Intervalo de confianza $\quad \mathsf{med}(X_1, \dots, X_n) \pm z_{\alpha/2}\widehat{\mathsf{se}}_{boot}$

Intevalos Bootstrap Normal

- $\widehat{\theta}_n$ asintóticamente normal si

$$\frac{\widehat{\theta}_n - \theta}{\mathsf{se}} \approx \mathcal{N}(0,1)$$

  con $\mathsf{se} = \mathsf{se}(\widehat{\theta}_n)$

- Sea $\widehat{\mathsf{se}}_{\text{boot}}$ el estimador bootstrap de $\mathsf{se}(\widehat{\theta}_n)$

$$\text{intervalo boot normal nivel } 1-\alpha\text{:} \quad \widehat{\theta}_n \pm z_{\alpha/2}\, \widehat{\mathsf{se}}_{\text{boot}}$$

Intevalos Bootstrap Percentil.

- $N_{boot} = N_{simu}$
- Sean $\widehat{\theta}^*_1, \ldots, \widehat{\theta}^*_{Nboot}$ estadísticos bootstrap de su estimador.

intervalo boot percentil $1-\alpha$: $\quad \left(\widehat{\theta}^*_{\alpha/2}\ ,\ \widehat{\theta}^*_{1-\alpha/2}\right)$