

Intervalos y regiones de confianza

Estadística (M)

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

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

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

Mundo Normal

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

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

- 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, \dots, 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}, \dots, 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

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

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, \dots, X_n variables aleatorias independientes con distribución $N(\mu, \sigma^2)$ y sea m como antes.

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

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

(iii) V y W son independientes

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

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

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

Procedimiento para obtener RC con nivel asintótico

Teorema Sea X_1, \dots, X_n una muestra aleatoria de una distribución perteneciente a la familia $F(x, \theta)$, $\theta \in \Theta$.

Supongamos que

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

Luego, si

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

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

Ejemplos

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

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

tiene nivel asintótico $1 - \alpha$.

Vayamos al pizarrón.

Ejemplos

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

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

•

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

Vayamos al pizarrón.

Usando la distribución asintótica de los EMV

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

Bajo condiciones de regularidad

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



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

- La región

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

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 $\hat{\theta}_n \xrightarrow{p} \theta$, bajo condiciones de regularidad obtendremos que $I_1(\hat{\theta}_n) \xrightarrow{p} I_1(\theta)$, luego

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

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

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

Otra mirada...

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

Funcionales

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

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

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

Consideramos funcionales de la forma

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

Otra mirada...

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

La Distribución Empírica

Sean X_1, X_2, \dots, 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, \dots, X_n .

La Distribución Empírica

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

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

$$\mathbb{E}_F \{g(X_1, \dots, X_n)\} \iff \mathbb{E} \{g(X_1, \dots, X_n)\} ,$$

cuando X_1, \dots, X_n i.i.d., $X_i \sim F$.

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

$$? \mathbb{P}_{\hat{F}_n}(A) = \dots\dots$$

$$? \mathbb{E}_{\hat{F}_n}(X) = \dots\dots$$

$$? \mathbb{V}_{\hat{F}_n}(X) = \dots\dots$$

La Distribución Empírica

Sean X_1, X_2, \dots, 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, \dots, 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, \dots, 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, \dots, X_n .
- Ley de los Grandes Números:

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

- Glivenko Cantelli:

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

Estimación Plug-in

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

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

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