,

Clase 4 - De todo un poco

Probabilidades

Carrera de Especialización en Estadística,
Facultad de Ciencias Exactas y Naturales, UBA

2023

¿Qué vamos a ver hoy?

- Otros tipos de convergencia
- Relación entre convergencias
- Método Delta
- Repaso

Tipos de Convergencia

Sean $W_1, W_2, \ldots$ y W v.a. definidas en $(\Omega, \mathcal{A}, \mathbb{P})$. Diremos que W_n converge a W

- en probabilidad, denotado con $W_n \xrightarrow{p} W$, si

$$\lim_{n\to\infty} \mathbb{P}(|W_n - W) > \varepsilon) = 0 \,, \quad \forall \varepsilon > 0$$

.

- en casi todo punto, denotado con $W_n \xrightarrow{ctp} W$, si

$$\mathbb{P}\left(\{s : W_n(s) \to W(s)\}\right) = 1$$

.

- en media cuadrática si

$$\lim_{n\to\infty} \mathbb{E}\{(W_n - W)^2\} = 0$$

Propiedades

- $W_n \xrightarrow{ctp} W$ entonces $W_n \xrightarrow{p} W$.
- $\mathbb{E}\{(W_n - W)^2\} \to 0$, entonces $W_n \xrightarrow{p} W$.

Teorema: Ley de los Grandes Números (LGN)

Sean $(X_i)_{i \geq 1}$ i.i.d., con $\mathbb{E}(X_i) = \mu$ para todo i. Entonces, el promedio converge a μ:

$$\overline{X}_n \xrightarrow{p} \mu \quad \text{(y q.t.p.)}$$

Es decir,

$$\lim_{n \to \infty} \mathbb{P}\left(|\overline{X}_n - \mu| > \varepsilon\right) = 0 \ , \quad \forall \varepsilon > 0.$$

Muestra (aleatoria simple) - De Wasserman

2.41 Definition. *If $X_1, \ldots, X_n$ are independent and each has the same marginal distribution with* CDF *F, we say that $X_1, \ldots, X_n$ are* IID *(independent and identically distributed) and we write*

$$X_1, \ldots X_n \sim F.$$

If F has density f we also write $X_1, \ldots X_n \sim f$. We also call $X_1, \ldots, X_n$ a **random sample of size n from F.**

Utilizamos la Ley de los Grandes Números para Estimación

Estimación de la Esperanza

- $(X_i)_{i\geq 1}$, i.i.d. $X_i \sim F$.
- $\mathbb{E}(X_i) = \mu$, $V(X_i) = \sigma^2$
- Por la ley de los grandes números

$$\overline{X}_n \to \mathbb{E}(X_1) = \mu \quad \text{en probabilidad}$$

- En tal caso, definimos

$$\widehat{\mu} = \widehat{\mu}_n = \overline{X}_n$$

.

- Propiedades de $\widehat{\mu}_n$:
  - $\mathbb{E}(\widehat{\mu}_n) = \mu$
  - $\widehat{\mu}_n \to \mu$ en probabilidad.

Variables Bernoulli

- $X_i \sim \mathcal{B}(1, p)$, i.i.d.
- $\mathbb{E}(X_i) = p$, $V(X_i) = p(1-p)$
- Por la ley de los grandes números

$$\overline{X}_n \to \mathbb{E}(X_1) = p \quad \text{en probabilidad}$$

- En tal caso, definimos

$$\widehat{p} = \widehat{p}_n = \overline{X}_n$$

.

- Propiedades de $\widehat{p}_n$:
  - $\mathbb{E}(\widehat{p}_n) = p$
  - $\widehat{p}_n \to p$ en probabilidad.

Estimación de Probabilidades - Ejemplo

- $(X_i)_{i\geq 1}$, i.i.d. $X_i \sim F$.
- Nos interesa estimar $F(3) = \mathbb{P}(X \leq 3)$.
- Definimos

$$Y_i = I_{X_i\leq 3} = I_{(-\infty,3]}(X_i)$$

- $Y_i \sim \mathcal{B}(1,p)$, con

$$p = \mathbb{E}(Y_1) = \mathbb{P}(Y_1 = 1) = \mathbb{P}(X_1 \leq 3) = F(3)$$

- Por la ley de los grandes números (caso Bernoulli)

$$\overline{Y}_n \to \mathbb{E}(Y_1) = F(3) \quad \text{en probabilidad}$$

- En tal caso, definimos

$$\widehat{F}(3) = \widehat{F}_n(3) = \overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} I_{X_i\leq 3}$$

.

Estimación de Probabilidades.

- $(X_i)_{i\geq 1}$, i.i.d. $X_i \sim F$.
- Nos interesa estimar $\mathbb{P}(X \in A)$.
- Definimos
$$Y_i = I_{X_i \in A} = I_A(X_i)$$
- $Y_i \sim \mathcal{B}(1, p)$, con
$$p = \mathbb{E}(Y_1) = \mathbb{P}(Y_1 = 1) = \mathbb{P}(X_1 \in A)$$
- Por la ley de los grandes números (caso Bernoulli)
$$\overline{Y}_n \to \mathbb{E}(Y_1) = \mathbb{P}(X_1 \in A) \quad \text{en probabilidad}$$
- Es decir,
$$\frac{1}{n}\sum_{i=1}^{n} I_{X_i \in A} \longrightarrow \mathbb{P}(X_1 \in A)$$

DEMOSTRAMOS QUE LA FRECUENCIA RELATIVA CONVERGE A LA PROBABILIDAD.

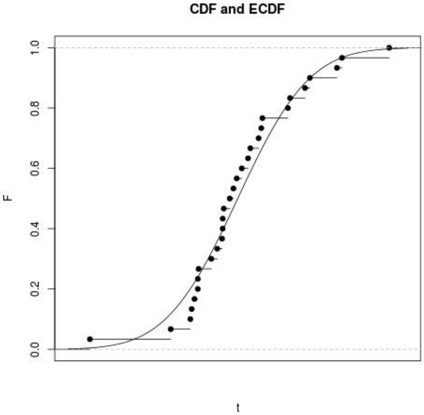
Estimación de $F(t)$

- $(X_i)_{i\geq 1}$, i.i.d. $X_i \sim F$.
- Nos interesa estimar $F(t) = \mathbb{P}(X \leq t)$.
- Definimos
$$Y_i = I_{X_i \leq t} = I_{(-\infty, t]}(X_i)$$
- $Y_i \sim \mathcal{B}(1, p)$, con
$$p = \mathbb{E}(Y_1) = \mathbb{P}(Y_1 = 1) = \mathbb{P}(X_1 \leq t) = F(t)$$
- Por la ley de los grandes números (caso Bernoulli)
$$\overline{Y}_n \to \mathbb{E}(Y_1) = F(t) \quad \text{en probabilidad}$$
- En tal caso, definimos
$$\widehat{F}(t) = \widehat{F}_n(t) = \overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \leq t}$$

.

Empírica: una realización

$$\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le t\}}$$



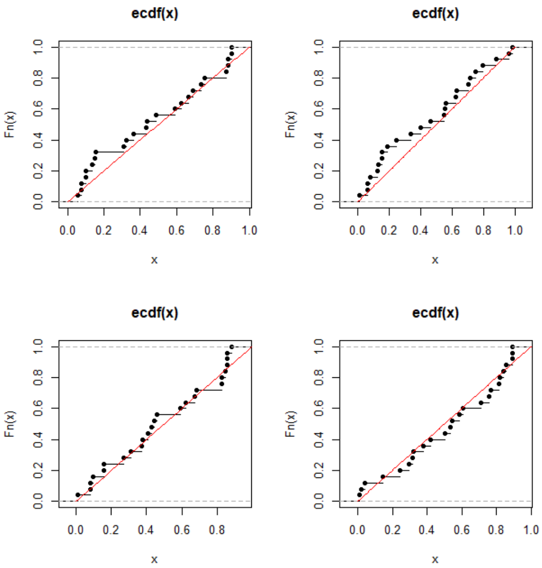
La empírica

Sean $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim F$. Definimos la función de distribución empírica como

$$\widehat{F}_n(t) := \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq t\}}$$

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

Datos simulados: $X_1, \ldots, X_{25}$ i.i.d., $X_i \sim \mathcal{U}(0,1)$



Teorema: Glivenko Cantelli

Sean $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim F$. Consideremos la función de distribución empírica, definida como

$$\widehat{F}_n(t) := \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq t\}}$$

Tenemos entonces que

$$\sup_{t \in \mathbb{R}} |\widehat{F}_n(t) - F(t)| \xrightarrow{p} 0 \quad \text{(y q.t.p.)}$$

LGN - más generalmente

$$\frac{1}{n}\sum_{i=1}^{n}\texttt{mongo}_i \to \mathbb{E}(\texttt{mongo}) \quad \text{en probabilidad}$$

Estimación de varianza.

- $(X_i)_{i \geq 1}$, i.i.d. $X_i \sim F$.
- Nos interesa estimar $\sigma^2 = V(X) = \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2$.

- ¿Qué hacemos?

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \{\overline{X}_n\}^2 \xrightarrow{p} \mathbb{E}(X^2) - \mu^2 = Var(X)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \xrightarrow{p} \sigma^2$$

Propiedades de la convergencia en probabilidad

- Ley de los grandes Números: $\texttt{mongo}_1, \texttt{mongo}_2, \ldots$ iid, $\texttt{mongo}_i \sim \texttt{mongo}$

$$\frac{1}{n}\sum_{i=1}^{n} \texttt{mongo}_i \quad \xrightarrow{p} \quad \mathbb{E}(\texttt{mongo})$$

- Si $W_n \xrightarrow{p} W$ y g continua, entonces $g(W_n) \xrightarrow{p} g(W)$.
- Si $W_n \xrightarrow{p} W$ y $U_n \xrightarrow{p} U$, entonces $W_n \pm U_n \xrightarrow{p} W \pm U$.
- Si $W_n \xrightarrow{p} W$ y $U_n \xrightarrow{p} U$, entonces $W_n U_n \xrightarrow{p} WU$.

- Si $W_n \xrightarrow{p} W$ y $U_n \xrightarrow{p} U$ y $\mathbb{P}(U = 0) = 0$, entonces $W_n/U_n \xrightarrow{p} W/U$.

- Si $W_n \xrightarrow{p} W$ y $U_n \xrightarrow{p} U$ y g continua, entonces $g(W_n, U_n) \xrightarrow{p} g(W, U)$.

Convergencia en distribución

Definition

Sea $W_1, W_2, \ldots$ una sucesión de variables aleatorias y sea W otra variable aleatoria. Denotemos con F_n y F a la función de distribución acumulada de W_n y W, respectivamente: $W_n \sim F_n$, $W \sim F$. Diremos que W_n converge en distribución a W si

$$\lim_{n \to \infty} F_n(t) = F(t), \quad \text{para todo } t \text{ punto de continuidad de } F$$

Notación: $W_n \xrightarrow{D} W$.

Teorema Central del límite

Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias iid definidas en $(\Omega, \mathcal{A}, P)$, con $\mathbb{E}(X_i) = \mu$ y $V(X_i) = \sigma^2$, Entonces,

$$\frac{\overline{X}_n - \mu}{\sqrt{\mathbb{V}(\overline{X}_n)}} \xrightarrow{D} Z, \quad Z \sim \mathcal{N}(0, 1).$$

Versiones equivalentes

$$\frac{\overline{X}_n - \mu}{\sqrt{\mathbb{V}(\overline{X}_n)}} \xrightarrow{D} Z, \quad Z \sim \mathcal{N}(0,1).$$

$$\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \xrightarrow{D} Z, \quad Z \sim \mathcal{N}(0,1).$$

$$\frac{\sqrt{n}\{\overline{X}_n - \mu\}}{\sqrt{\sigma^2}} \xrightarrow{D} Z, \quad Z \sim \mathcal{N}(0,1).$$

$$\sum_{i=1}^{n} X_i \overset{a}{\approx} \mathcal{N}(n\mu, n\sigma^2) \quad \overline{X}_n \overset{a}{\approx} \mathcal{N}(\mu, \sigma^2/n).$$

Teorema de Slutzky

1. Si $W_n \xrightarrow{D} W$ y $U_n \xrightarrow{p} c$, con $c \in \mathbb{R}$, entonces

$$W_n + U_n \xrightarrow{D} c + U$$

2. Si $W_n \xrightarrow{D} W$ y $U_n \xrightarrow{p} c$, con $c \in \mathbb{R}$, entonces

$$W_n U_n \xrightarrow{D} cU$$

Teorema de Slutzky

1. Si $W_n \xrightarrow{D} W$ y $U_n \xrightarrow{p} c$, con $c \in \mathbb{R}$, entonces

$$W_n + U_n \xrightarrow{D} c + U$$

2. Si $W_n \xrightarrow{D} W$ y $U_n \xrightarrow{p} c$, con $c \in \mathbb{R}$, entonces

$$W_n U_n \xrightarrow{D} cU$$

¿Cómo se usan estas cosas?

Método Delta

Supongamos que

$$\sqrt{n}\frac{(W_n - \mu)}{\sigma} \xrightarrow{D} \mathcal{N}(0,1)$$

y que g es una función diferenciable tal que $g'(\mu) \neq 0$. Entonces

$$\sqrt{n}\frac{(g(W_n) - g(\mu))}{|g'(\mu)|\sigma} \xrightarrow{D} \mathcal{N}(0,1)$$

Es decir,

$$\text{si} \quad W_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{entonces} \quad g(W_n) \approx \mathcal{N}\left(g(\mu), \frac{(g'(\mu))^2\sigma^2}{n}\right)$$

Bonus Truck: *Tightness*

Una sucesión $(W_n)_{n\geq 1}$ se dice acotada en probabilidad (o tight) si

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}\left(|W_n| > K\right) = 0$$

Notación: $\quad W_n = O_P(1)$

Bonus Truck: *Tightness*

Una sucesión $(W_n)_{n\geq 1}$ se dice acotada en probabilidad (o tight) si

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}\left(|W_n| > K\right) = 0$$

Notación: $\quad W_n = O_P(1)$

Lema: Si $(W_n)_{n\geq 1}$ converge en distribución, entonces $(W_n)_{n\geq 1}$ esta acotada en probabilidad.

Propiedades

- $W_n \xrightarrow{ctp} W$ entonces $W_n \xrightarrow{p} W$.

- $\mathbb{E}\{(W_n - W)^2\} \to 0$, entonces $W_n \xrightarrow{p} W$.

- $W_n \xrightarrow{p} W$, entonces $W_n \xrightarrow{D} W$.
- $W_n \xrightarrow{p} W$, entonces $W_n = O_P(1)$.

- $W_n \xrightarrow{D} c$, con c constante, entonces $W_n \xrightarrow{p} c$.

La esperanza solo preserva convergencia en media cuadrática.