# Más sobre convergencias y un poco de Estadística

Pablo L. De Nápoli

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Probabilidades y Estadística para Matemática
Segundo cuatrimestre de 2021

# Parte I

# Teorema de Slutksky

# Recordamos algunas definiciones

Sean $(\Omega, \mathcal{E}, P)$ un espacio de probabilidad, y $(X_n) : \Omega \to \overline{\mathbb{R}}$ una sucesión de variables aleatorias finitas en casi todo punto. Recordamos dos nociones de convergencia (cuando $n \to +\infty$)

- Convergencia en probabilidad:

$$X_n \xrightarrow{P} X \Leftrightarrow \forall \delta > 0, \quad P\{|X_n - X| > \delta\} \to 0$$

- Convergencia en distribución:

$$X_n \xrightarrow{D} X \Leftrightarrow F_{X_n}(x) \to F_X(x)$$

  para todo punto de continuidad $x$ de $F_X$.
  Equivalentemente, tenemos la caracterización dada por el teorema de Helly-Bray

$$X_n \xrightarrow{D} X \Leftrightarrow \forall \psi \in C_c^\infty(\mathbb{R}), E[\psi(X_n)] \to E[\psi(X)]$$

  y también en términos de funciones características (teorema de Levy)

$$X_n \xrightarrow{D} X \Leftrightarrow \varphi_{X_n}(t) \to \varphi_X(t)$$

# Teorema de la aplicación continua

Como vimos en la clase 18, el teorema de Helly-Bray inmediatamente implica que:

**Corolario**

*Si* $X_n \xrightarrow{D} X$ *y* $g : \mathbb{R} \to \mathbb{R}$ *es una función continua, entonces* $g(X_n) \xrightarrow{D} g(X)$*.*

En particular:

**Corolario**

*Si* $X_n \xrightarrow{D} X$ *y* $c \in \mathbb{R}$ *es un constante, entonces* $X_n + c \to X + c$ *y* $X_n \cdot c \to X \cdot c$*.*

Esto sale porque $g_1(x) = x + c$ y $g_2(x) = c \cdot x$ son funciones continuas.

Sin embargo, ¡hay que tener cuidado! En general no es cierto que si $X_n \xrightarrow{D} X$, $Y_n \xrightarrow{D} Y$, entonces $X_n + Y_n \xrightarrow{D} X + Y$.

# una versión simple del teorema de Slutsky

**Lema (una versión simple del teorema de Slutsky)**

*Sean $(X_n)$ e $(Y_n)$ dos sucesiones de variables aleatorias finitas con probabilidad 1. Supongamos que $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} 0$. Entonces $X_n + Y_n \xrightarrow{D} X$.*

En el apunte de Victor Yohai pueden ver una prueba usando directamente la definición de convergencia en distribución. Yo les voy a presentar una prueba alternativa usando la caracerización dada por el teorema de Helly-Bray.

# Demostración del lema (1)

Usando la caracerización dada por el teorema de Helly-Bray, queremos probar que para toda $\psi \in C_c^\infty(\mathbb{R})$,

$$E[\psi(X_n + Y_n)] \to E[\psi(X)$$

y sabemos por hipótesis que:

$$E[\psi(X_n)] \to E[\psi(X)]$$

Luego nos bastará probar que para cada $\psi$ fija,

$$E[\psi(X_n + Y_n)] - E[\psi(X_n)] \to 0$$

Notamos que como $\psi_c(\mathbb{R})$, $\psi$ será acotada

$$|\psi(x)| \leq C \quad \text{para todo } x, y \in \mathbb{R}$$

y cumplirá la condición de Lipschitz

$$|\psi(x) - \psi(y)| \leq M|x - y| \quad \text{para todo } x \in \mathbb{R}$$

donde $M$ es cualquier cota de $|\psi'|$ (por el teorema del valor medio).

# Demostración del lema (2)

Usando las observaciones anteriores, tenemos que dado $\varepsilon > 0$,

$$|\psi(X_n + Y_n) - \psi(X_n)| \leq M|Y_n| < \frac{\varepsilon}{2}$$

si

$$|Y_n| < \delta = \frac{\varepsilon}{2M}$$

Entonces, introducimos los eventos:

$$A_{n,\delta} = \{\omega \in \Omega : |Y_n(\omega)| < \delta\}$$

y podemos estimar:

$$E[|\psi(X_n + Y_n)] - E[\psi(X_n)| \cdot I_{A_{n,\delta}}] \leq \frac{\varepsilon}{2}$$

# Demostración del lema (3)

Ahora vamos a necesitar mirar que pasa en

$$A^c_{n,\delta} = \{\omega \in \Omega : |Y_n(\omega)| \geq \delta\}$$

Ahí vamos a usar la estimación más bruta

$$|\psi(X + Y_n) - \psi(X_n)| \leq 2C$$

Entonces:

$$E[|\psi(X_n + Y_n) - \psi(X_n)| \cdot I_{A^c_{n,\delta}}] \leq 2C \cdot E[I_{A^c_{n,\delta}}] = 2C \cdot P(A_{n,C}) < \frac{\varepsilon}{2}$$

si $n \geq n_0(\varepsilon, \delta)$ pues $Y_n \xrightarrow{P} 0$. Pero $\delta = \delta(\varepsilon)$, así que en definita $n_0$ depende sólo de $\varepsilon$.

# Demostración del lema (4)

Finalmente acotamos

$$
\begin{aligned}
|E[\psi(X_n + Y_n)] - E[\psi(X_n)]| &\leq E[|\psi(X_n + Y_n)] - E[\psi(X_n)|] \\
&\leq E[|\psi(X_n + Y_n)] - E[\psi(X_n)|I_{A_{n,\delta}}] \\
+ &\leq E[|\psi(X_n + Y_n)] - E[\psi(X_n)|I_{A^c_{n,\delta}}] \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}
$$

si $n \geq n_0$. O sea, que efectivamente hemos probado que:

$$E[\psi(X_n + Y_n)] - E[\psi(X_n)] \to 0$$

y como observamos antes, esto implica la validez del lema.

# Un lema para el teorema de Slutsky

## Lema

*Sea* $(X_n)$ *una sucesión de variables aleatorias finitas con probabilidad 1, tales que* $X_n \xrightarrow{P} c$ *donde* $c \in \mathbb{R}$ *es una constante. Entonces si* $g : \mathbb{R} \to \mathbb{R}$ *es una función boreliana continua en* $c$*, entonces:*

$$Y_n = g(X_n) \xrightarrow{P} g(c)$$

# Demostración del lema

Dado $\varepsilon > 0$ por definición de continuidad, existirá un $\delta > 0$ tal que $|x - c| \leq \delta$ implica $|g(x) - g(c)| \leq \varepsilon$. Luego,

$$\{|g(x) - g(c)| \geq \varepsilon\} \subset \{|x - c| \geq \delta\}$$

En particular,

$$\{|g(X_n) - g(c)| \geq \varepsilon\} \subset \{|X_n - c| \geq \delta\}$$

tomando probabilidades:

$$0 \leq P\{|g(X_n) - g(c)| \geq \varepsilon\} \leq P\{|X_n - c| \geq \delta\}$$

por lo que si el lado derecho tiende a cero cuando $n \to +\infty$, también el término del medio. O sea que si $X_n \xrightarrow{P} c$, se deduce que $g(X_n) \xrightarrow{P} g(c)$.

# Teorema de Slutksky

## Teorema

*Sean* $(X_n)$ *e* $(Y_n)$ *dos sucesiones de variables aleatorias finitas con probabilidad 1. Supongamos que* $X_n \xrightarrow{D} X$ *e* $Y_n \xrightarrow{P} c$ *donde* $X$ *es otra variable aleatoria finita con probabilidad* 1 *y* $c$ *una constante. Entonces,*

- $X_n + Y_n \xrightarrow{D} X + c.$
- $X_n Y_n \xrightarrow{D} cX.$
- *Si* $c \neq 0$,

$$\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{c}$$

# Demostración del teorema de Slutsky (1)

Para probar que $X_n + Y_n \xrightarrow{D} X + c$, escribimos:

$$X_n + Y_n = (X_n + c) + (Y_n - c)$$

Comomo $X_n \xrightarrow{D} X$, tendremos que $X_n + c \xrightarrow{D} X + c$ por los resultados previos.
También

$$Y_n \xrightarrow{P} c \Rightarrow Y_n - c \xrightarrow{P} 0$$

(esto sale directamente la definición).
El resultado se deduce entonces de la versión simple del teorema de Slutsky que probamos antes.

# Demostración del teorema de Slutsky (2)

Similarmente, para ver que $X_n Y_c \xrightarrow{D} cX$, escibimos

$$X_n Y_n = cX_n + (Y_n - c)X_n = U_n + Z_n$$

donde llammamos $U_n = cX_n$ y $Z_n = (Y_n - c)X_n$.
Como $X_n \xrightarrow{D} X \Rightarrow U_n \xrightarrow{D} cX$ por los resultados previos.
También sabemos que $Y_n - c \xrightarrow{P} 0$. Por otra parte, $(X_n)$ está acotada en probabilidad, ya que converge en distribución. [por el ejercicio 14, item d) de la práctica 8. Este resultado se deduce también de un lema que vimos en la clase 18]
Pero entonces $Z_n \xrightarrow{P} 0$ por un resultado que vimos en la clase 15, ya que es el producto que una sucesión que tiende a cero en probabilidad por una que está acotada en probabilidad, y por lo tanto $Z_n \xrightarrow{D} 0$.
Entonces usando la versión simple del teorema de Slutsky, concluimos que $X_n Y_n \xrightarrow{D} cX$.

# Demostración del teorema de Slutsky (3)

Finalmente, para ver que si $c \neq 0$,

$$\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{c}$$

escribimos

$$\frac{X_n}{Y_n} = X_n \cdot \frac{1}{Y_n}$$

y observamos que

$$Y_n \xrightarrow{P} c \Rightarrow \frac{1}{Y_n} \xrightarrow{P} \frac{1}{c}$$

por el lema previo aplicado a la función $g(y) = \frac{1}{y}$ que es continua en $y = c$ si $c \neq 0$. (Este paso es esencialmente el ejercicio 17 de la práctica 8).
Entonce el resultado se deduce del item anterior.

# Parte II

# Un poco de estadística: Estimación Puntual

# Introducción

Sea $X$ una variable aleatoria cuya función de distribución $F$ desconocemos. Por ejemplo, puede tratarse del peso de una lata de arvejas. que es una variable aleatoria que varía de lata en lata. A veces se hace una hipótesis sobre la posible distribución $F$ (por ejemplo que es normal, aunque esto no es siempre necesario).

La distribución $F$ está caracterizada por ciertos parámetros, como por ejemplo $\mu = E[X]$ o $\sigma^2 = \text{Var}(X)$. (o si suponemos que $F$ pertenece a una familia de distribuciones, los que definen esa familia).

Para hacerlo, se toma una muestra aleatoria de la población. Obtenemos de esta forma variables

$$X_1, X_2, \dots X_n$$

que serán variables aleatorias independientes e identicamente distribuidas con distribución $F$.

# Estimadores

Dada una muestra aleatoria $X_1, X_2, \ldots, X_n$ de una distribución $F$ y $\lambda = \lambda(F)$ un parámetro de $F$, un estimador será una una función

$$\widehat{\lambda}_n(X_1, X_2, \ldots, X_n)$$

(sin parámetros desconocidos) que nos permita estimar $\lambda$ en algún sentido.

Por ejemplo, si $\mu$ es la esperanza de la distribución $F$, entonces:

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

(conocido como media muestral es un estimador razonable de $\mu$ ya que por la ley fuerte de los grandes números

$$\overline{X}_n \xrightarrow{c.s.} \mu \quad \text{cuando } n \to +\infty$$

Se dice que $\overline{X}_n$ es un estimador fuertemente consistente para $\mu$.

# Estimación de la varianza

Similarmente, ¿Cómo podríamos estimar $\sigma^2 = \text{Var}(X) = E[(X-\mu)^2]$. Un estimador que podríamos considerar razonable es

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

Este sería el valor de la varianza de la distribución empírica generada a partir de la muestra.
Vamos a ver que nuevamente, este estimador es fuertemente consistente, o sea:

$$\widehat{\sigma}_n \xrightarrow{c.s.} \sigma$$

# Demostración de la consistencia fuerte del estimador de la variancia

Recordando que $\text{Var}(X) = E(X^2) - E(X)^2$ también tenemos:

$$\widehat{\sigma}_n^2 = \left[ \frac{1}{n} \sum_{i=1}^{n} X_i^2 \right] - \overline{X}_n^2$$

Por la ley fuerte de los grandes números

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \xrightarrow{c.s.} E(X^2)$$

Como elevar al cuadrado es una función continua

$$\widehat{\sigma}_n^2 \to E[X^2] - E(X)^2 = \sigma^2$$

y por lo tante

$$\widehat{\sigma}_n \xrightarrow{c.s.} \sigma$$

# Estimadores (2)

Ahora bien: dado un parámetro $\lambda$ pueden pensarse diferentes estimadores para $\lambda$ que pueden parecer igualmente razonables.
Por ejemplo, supongamos que tenemos una población cuya distribución $F$ sabemos que es normal $N(\mu, \sigma^2)$ con ciertos parámetros $\mu$ y $\sigma$ como vimos antes.
Entonces, para estimar $\mu$ podríamos usar la media muestral como vimos antes, porque $\mu$ es la esperanza de $F$.
Pero para la distribución normal $\mu$ también es la mediana. Por lo tanto otra forma de estimar $\mu$ podría ser usar la mediana muestral Me. Para definirla ordenamos las variables (o sea consideramos los estadísticos de orden):

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$$

y definimos

$$\mathrm{Me} = X_{((n+1)/2)} \text{ si } n \text{ es impar}$$

$$\mathrm{Me} = \frac{1}{2}\left(X_{(n/2)} + X_{(n/2+1)}\right) \text{ si } n \text{ es par}$$

Esto lleva a preguntarnos qué propiedades es deseable que tenga un estimador, para tener un criterio para elegir un estimador sobre otro.

# Sesgo de un estimador

Dado un estimador $\widehat{\lambda}_n$ de un parámetro $\lambda = \lambda(F)$, se define el sesgo del estimador como

$$\text{sesgo}(\widehat{\lambda}_n) = E[\widehat{\lambda_n}] - \lambda$$

Un estimador se dice insesgado si

$$\text{sesgo}(\widehat{\lambda}_n) = 0$$

y asintóticamente insesgado si

$$\text{sesgo}(\widehat{\lambda}_n) \to 0$$

# Sesgo de la media muestral

Consideramos el estimador

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

para la esperanza $\mu = E[X]$.
Por la linealidad de la esperanza,

$$E[\overline{X}_n] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

Luego $\overline{X}_n$ es un estimador insesgado de $\mu$.

# Sesgo para el estimador de la varianza

Ahora repitamos la cuenta con el estimador de la varianza que definimos antes. Recordamos que:

$$\widehat{\sigma}_n^2 = \left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] - \overline{X}_n^2$$

Así que empezemos calculando:

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i^2] = c$$

donde

$$c = E(X_i^2) = \text{Var}(X_i) + E(X_i)^2 = \sigma^2 + \mu^2$$

# Sesgo para el estimador de la varianza (2)

Por otra parte, necesitamos calcular $E[\overline{X}_n^2]$. Para ello, la observación clave es que como las variables $X_i$ son independientes

$$\mathrm{Var}(\overline{X}_n^2) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{\sigma^2}{n}$$

Entonces:

$$E(\overline{X}_n^2) = \mathrm{Var}(\overline{X}_n^2) + E[\overline{X}_n]^2 = \frac{\sigma^2}{n} + \mu^2$$

Entonces:

$$E[\widehat{\sigma}_n^2] = \sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n} + \mu^2\right) = \sigma^2 \left(1 - \frac{1}{n}\right) = \sigma^2 \left(\frac{n-1}{n}\right)$$

Luego este estimador no resulta insesgado, pero sí asintóticamente insesgado.

# Estimador insesgado de la varianza

Si queremos tener un estimador insegado de la varianza, debemos reemplazarlo por:

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

que aparece en el ejercicio 24 de la práctica 8, ya que como

$$S_n^2 = \frac{n}{n-1}\cdot \widehat{\sigma}_n^2$$

ahora tendremos que:

$$E[S_n^2] = \sigma^2$$

(si se fijan, hemos resuelto todo el ejercicio).

# Parte III

# Intervalos de confianza

# Planteo del problema

Hasta ahora vimos como estimar los parámetros de una distribución. Por ejemplo si sabemos (o conjeturamos) que tenemos una muetra de la distribución normal $N(\mu, \sigma^2)$ podemos estimar los parámetros.

¿Pero cómo podemos estimar el error cometido en la estimación? Primero consideraremos el caso más sencillo aunque poco realista en que $\sigma$ es conocido y queremos estimar $\mu$. Sabemos que podemos estimar $\mu$ usando la medida muestral $\widehat{\mu}_n = \overline{X}_n$.

Nos gustaría encontrar un intervalo de confianza para $\mu$, es decir un intervalo alrededor de $\widehat{\mu}_n$ tal que

$$P\{\mu \in I_\alpha\} = 1 - \alpha$$

donde $0 < \alpha < 1$ es un nivel de confianza elegido (típicamente $\alpha = 0,05$ ).

Ya nos encontramos con este concepto en un ejemplo que vimos en la clase 7 sobre las aproximaciones de la normal (encuesta electoral).

# Solución cuando la varianza es conocida

Cuando la distribución es normal y $\sigma$ es conocida podemos razonar así: $\overline{X}_n$ tendrá distribución $N\left(\mu, \frac{\sigma^2}{n}\right)$ Entonces:

$$Z_n = \sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\sigma} \sim N(0,1)$$

Ahora elegimos $z_{\alpha/2}$ de modo que $P(Z_n > z_{\alpha/2}) = \alpha/2$, y por la simetría de la curva normal tenemos que

$$P\{-z_{\alpha/2} \leq -Z_n \leq z_{\alpha/2}\} = 1 - \alpha$$

Depejando obtenemos el intervalo de confianza

$$I_\alpha = \left[\overline{X}_n - \frac{z_{\alpha/2}\sigma}{\sqrt{n}}, \overline{X}_n + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right]$$

para el que podemos garantizar que

$$P\{\mu \in I_\alpha\} = 1 - \alpha$$

# Intervalos de confianza asintóticos

En la realidad, no es realista suponer que la distribución es conocida, o que la varianza lo es. De todos modos, podemos definir un intervalo de confianza asintótico para $\mu = E[X]$, reemplazando a $\sigma$ por un estimador fuertemente consistente $\widehat{\sigma}_n$ de los que vimos antes (da igual cuál consideremos)

$$I_\alpha = \left[ \overline{X}_n - \frac{z_{\alpha/2}\widehat{\sigma}_n}{\sqrt{n}}, \overline{X}_n + \frac{z_{\alpha/2}\widehat{\sigma}_n}{\sqrt{n}} \right]$$

Con sólo suponer que la variancia de la distribución tendremos que:

$$Z_n = \sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{D} N(0,1)$$

por el teorema del límite central, siempre que $\sigma^2 = \mathrm{Var}(X_i) < \infty$.

Como

$$\frac{\sigma}{\widehat{\sigma}_n} \xrightarrow{c.s.} 1$$

tendremos que la convergencia en distribución no se ve alterada:

$$\widehat{Z}_n = \frac{\sigma}{\widehat{\sigma}_n} Z_n = \sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\widehat{\sigma}_n} \xrightarrow{D} N(0,1)$$

por el teorema de Slutsky.
Por lo que nuestro intervalo

$$I_\alpha = \left[ \overline{X}_n - \frac{z_{\alpha/2}\widehat{\sigma}_n}{\sqrt{n}}, \overline{X}_n + \frac{z_{\alpha/2}\widehat{\sigma}_n}{\sqrt{n}} \right]$$

verfica que

$$\lim_{n \to +\infty} P\{\mu \in I_\alpha\} = 1 - \alpha$$