Probabilidad

Convergencias
12/05/2023

Lara Sánchez Peña[1]

UTDT 2023 - MEC

[1] Basado en las notas de Andrea Rotnitzky

Objetivos

- Ley débil de grandes números
- Convergencia en probabilidad
- Convergencia en media cuadrática
- Teorema central del límite
- Convergencia en distribución
- Método Delta

En estas slides vamos a ver distintas nociones de convergencias de sucesiones de variables aleatorias y cómo se relacionan éstas diferentes nociones de convergencia entre sí.

Sucesiones de números $(a_n)_{n\in\mathbb{N}}$

- Una sucesión de números $(a_n)_{n\in\mathbb{N}}$ es una lista de números indexados por $n \in \mathbb{N}$.
- Decimos que $(a_n)_{n\in\mathbb{N}}$ es **convergente** si existe $\lim_{n\to\infty} a_n = a$. Es decir, existe un número a tal que para cualquier $\varepsilon > 0$ que se considere, existe $n(\varepsilon)$ de manera que si $n > n(\varepsilon)$ entonces

$$|a_n - a| < \varepsilon$$

- Una sucesión de variables aleatorias $(V_n)_{n\in\mathbb{N}}$ es una lista de variables aleatorias indexadas por $n \in \mathbb{N}$.
- Vamos a considerar sucesiones de variables aleatorias $(V_n)_{n\in\mathbb{N}}$ y vamos a estudiar **de qué tipos de manera** pueden converger cuando $n \to \infty$.

Muestra aleatoria y media muestral $\overline{X}_n$

Antes de empezar, definimos las siguientes variables aleatorias:

- Llamamos una **muestra aleatoria** de v.a. $X_1, \ldots, X_n$ si las variables $\{X_i\}_{i=1}^n$ son variables aleatorias independientes e idénticamente distribuidas.
- Definimos a la variable aleatoria $\overline{X}_n$ la **media muestral** definida por

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\left(X_1 + \cdots + X_n\right)$$

- Definimos a la variable aleatoria S_n^2 la **varianza muestral** definida por

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

Para los resultados teóricos LGN y TCL vamos a estudiar de qué maneras converge la sucesión de variables aleatorias $V_n = \overline{X}_n$ entre otras.

Propiedades de $\overline{X}_n$ y S^2

- $E(\overline{X}_n) = E(X)$

$$E(\overline{X}_n) = \frac{1}{n} E(X_1 + \cdots + X_n) = \frac{1}{n}\left(E(X_1) + \ldots E(X_n)\right) \underbrace{=}_{X_i \text{ id}} E(X)$$

- $Var(\overline{X}_n) = \frac{Var(X)}{n}$

$$Var(\overline{X}_n) = \frac{1}{n^2} Var(X_1 + \cdots + X_n) \underbrace{=}_{X_i \text{ indep}} \frac{1}{n^2}\left(Var(X_1) + \ldots Var(X_n)\right) \underbrace{=}_{X_i \text{ id}} \frac{Var(X)}{n}$$

- $E(S_n^2) = Var(X)$

$$\begin{aligned}
E(S_n^2) =& \frac{1}{n-1} \sum_{i=1}^{n} E(X_i^2) - \frac{n}{n-1} E\left[\left(\overline{X}_n\right)^2\right] \\
=& \frac{1}{n-1} \sum_{i=1}^{n} \left[(E(X))^2 + Var(X)\right] - \frac{n}{n-1}\left[(E(X))^2 + \frac{Var(X)}{n}\right] \\
=& Var(X)
\end{aligned}$$

Ley débil de los grandes números para $X_i \overset{iid}{\sim}$

Sea $X_1, \ldots, X_n$ una muestra aleatoria tal que $E(X) = \mu$. Entonces para todo $\varepsilon > 0$

$$P\left(|\overline{X}_n - \mu| \geq \varepsilon\right) \underset{n\to\infty}{\rightarrow} 0$$

Notacionalmente escribimos $\overline{X}_n \xrightarrow{p} \mu$ o plím $\overline{X}_n = \mu$.

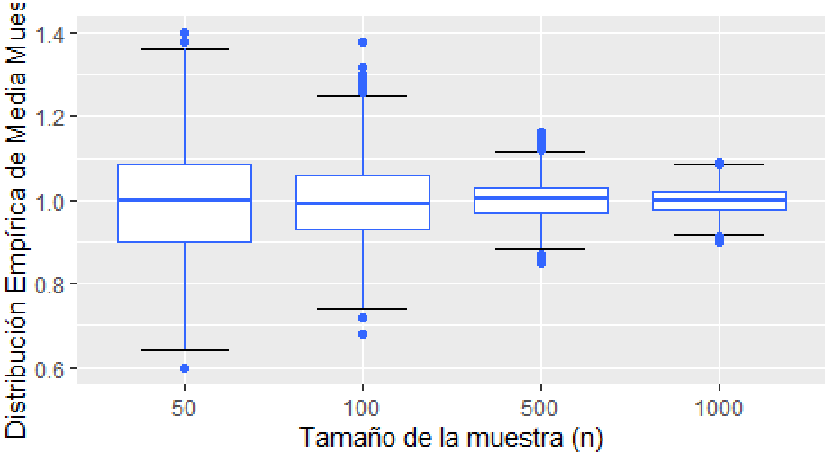
Interpretación frecuentista: Si n es grande con probabilidad alta el estimador $\overline{X}_n$ tomará valores 'cerca'de $E(X)$. Notemos que es una propiedad deseable de $\overline{X}_n$ como estimador de μ.

Notemos que la conclusión es más general y débil que la que obtuvimos antes usando Chebyshev, pero también que las hipótesis son más débiles ya para que valga la LGN que no requerimos que exista $Var(X)$. Con Cheby al pedir que exista $Var(X)$ sabemos que la probabilidad se achica con velocidad $\frac{1}{n}$.

La demostración está fuera del alcanza del curso. Es fácil probar que el resultado vale si asumimos que $Var(X)$ es finita. En este caso, por Chebyshev, para todo $\varepsilon > 0$ vale que

$$P\left(|\overline{X}_n - \mu| \geq \varepsilon\right) \leq \frac{Var(X)}{n}\frac{1}{\varepsilon^2} \underset{n\to\infty}{\longrightarrow} 0\,.$$

Distribución empírica de $\overline{X}_n$ para$X_i \overset{iid}{\sim} Poi(1)$



LGN visto gráficamente $X_i \overset{iid}{\sim} Poi(\lambda)$ si $\lambda = 1$

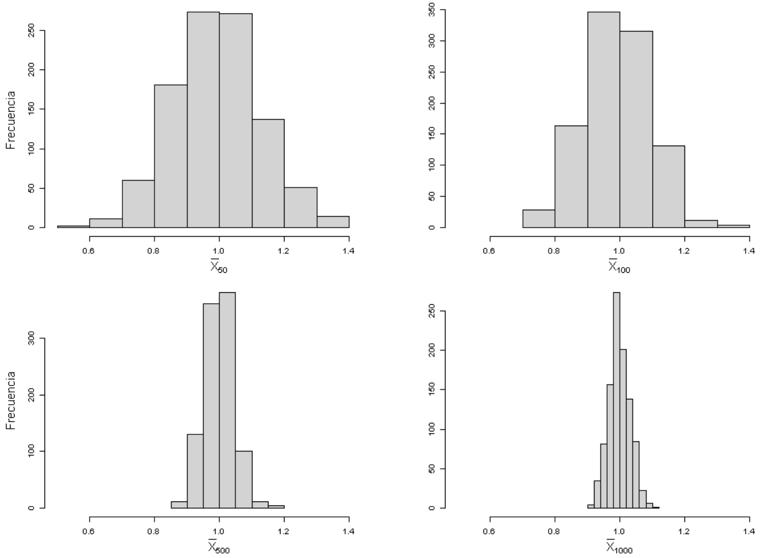


Figura: Distribución de $\overline{X}_n$ para distintos tamaños de muestra

LGN débil - ejercicios

Es importante que noten que si:

- g es una función continua, si $X_i = g(Y_i)$ entonces vale que $\frac{1}{n}\sum_{i=1}^{n} g(Y_i) \xrightarrow{p} E(g(Y))$
- $X_i = Y_i^2$ entonces vale que $\frac{1}{n}\sum_{i=1}^{n} Y_i^2 \xrightarrow{p} E(Y^2)$
- $X_i = \ln(W_i)$ entonces vale que $\frac{1}{n}\sum_{i=1}^{n} \ln(W_i) \xrightarrow{p} E(\ln(W))$

Ejercicios

1. Si $T_i \sim_{iid}$, $E(T) = 5$ y $Var(T) = 144$ calculá θ si $\frac{1}{n}\sum_{i=1}^{n} T_i^2 \xrightarrow{p} \theta$
2. Si $S_i \sim_{iid}$, $E(S) = 1$ y $E(\ln(S)) = -1$ calculá θ si $\ln\left[\left(\prod_{i=1}^{n} S_i\right)^{\frac{1}{n}}\right] \xrightarrow{p} \theta$

Convergencia en probabilidad

Sea $V_1, \dots, V_n, \dots$ una sucesión de variables aleatorias. Decimos que la sucesión $(V_n)_{n\in\mathbb{N}}$ **converge en probabilidad** a $a \in \mathbb{R}$ si

$$\lim_{n\to\infty} P\left(\{\omega \in \Omega : |V_n(\omega) - a| > \varepsilon\}\right) = 0$$

Escribimos $V_n \xrightarrow{p} a$ para denotar convergencia en probabilidad. En la *LGN* débil $V_n = \overline{X}_n$ y $a = E(X)$.

Más en general, decimos que la sucesión $(V_n)_{n\in\mathbb{N}}$ **converge en probabilidad** a V (una v.a.) si

$$\lim_{n\to\infty} P\left(\{\omega \in \Omega : |V_n(\omega) - V(\omega)| > \varepsilon\}\right) = 0$$

En ese caso escribimos $V_n \xrightarrow{p} V$. Note que en cierta bibliografía se escribe también $plim V_n = V$.

Notemos que si $V_n \xrightarrow{p} V$ y $V_n \xrightarrow{p} W$ entonces $P(V = W) = 1$.

Ejemplo 1: convergencia en probabilidad

Sean $X_1, \ldots, X_n$ v.a.i.i.d. de manera que $X_i \sim U(0,1)$, $i = 1, \ldots, n$. Definimos $V_n = \min\{X_1, \ldots, X_n\}$. Muestre que V_n converge en probabilidad a $a = 0$. Sea $\varepsilon > 0$, entonces[2]

$$
\begin{aligned}
P\left(|V_n - 0| \geq \varepsilon\right) &= P\left(\min\{X_1, \ldots, X_n\} \geq \varepsilon\right), \qquad \text{porque } X_i \geq 0 \\
&= P\left(X_1 \geq \varepsilon \text{ y } \cdots \text{ y } X_n \geq \varepsilon\right) \\
&\underbrace{=}_{X_i \text{ indep}} P\left(X_1 \geq \varepsilon\right) \cdots P\left(X_n \geq \varepsilon\right) \underbrace{=}_{X_i \text{ id}} (1 - \varepsilon)^n
\end{aligned}
$$

Tomando límite de $n \to \infty$, concluimos que

$$
\lim_{n \to \infty} P\left(|V_n - 0| \geq \varepsilon\right) = \lim_{n \to \infty} (1 - \varepsilon)^n = 0.
$$

Por lo tanto, $V_n \xrightarrow{p} 0$.

[2] Notemos que como $X_i \sim U(0,1)$, entonces $P(X_i < \varepsilon) = \varepsilon$ y $P(X_i \geq \varepsilon) = 1 - \varepsilon$.

Ejemplo 2: convergencia en probabilidad

Sea $X \sim \text{Exponencial}(1)$.
Podemos calcular que $P(X \geq x) = e^{-x}$. Sea $V_n = \frac{X}{n}$. Muestre que V_n converge en probabilidad a $a = 0$. Sea $\varepsilon > 0$, entonces

$$\begin{aligned} P\left(|V_n - 0| \geq \varepsilon\right) &= P\left(V_n \geq \varepsilon\right) \\ &= P(X \geq n\varepsilon) \\ &= e^{-n\varepsilon} \end{aligned}$$

Tomando $n \to \infty$

$$\lim_{n\to\infty} P\left(|V_n - 0| \geq \varepsilon\right) = \lim_{n\to\infty} e^{-n\varepsilon} = 0$$

Por lo tanto, $V_n \xrightarrow{p} 0$.

$V_n \xrightarrow{p} V$ **no implica** que $E(V_n) \overset{n\to\infty}{\longrightarrow} E(V)$: Ejemplo 1

Para el siguiente ejemplo tomamos $V = a = 0$, una variable aleatoria degenerada. Sea $U \sim U(0,1)$ y sean

$$V_n(u) = \begin{cases} 0 & \text{si } u > \frac{1}{n}, \text{ con prob. } 1 - \frac{1}{n} \\ n & \text{si } u \leq \frac{1}{n}, \text{ con prob. } \frac{1}{n} \end{cases}$$

Notamos que:

- $E(V_n) = 0 \cdot \left(1 - \frac{1}{n}\right) + n \cdot \frac{1}{n} = 1$ para cualquier valor de n. O sea,

$$\lim_{n\to\infty} E(V_n) = 1$$

- Sin embargo, $V_n \xrightarrow{P} 0$ porque

$$P(|V_n - 0| \geq \varepsilon) = P(V_n = n) = \frac{1}{n}$$

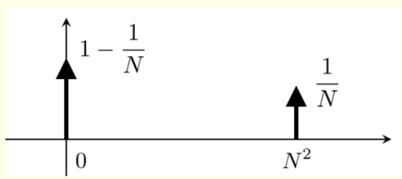
Entonces

$$\lim_{n\to\infty} P(|V_n - 0| \geq \varepsilon) = \lim_{n\to\infty} \frac{1}{n} = 0 \text{ pero } \lim_{n\to\infty} E(V_n) = 1 \not\to E(V) = 0.$$

$V_n \xrightarrow{p} V$ **no implica** que $E(V_n) \overset{n\to\infty}{\to} E(V)$: Ejemplo 2

Para el siguiente ejemplo tomamos $V = a = 0$, una variable aleatoria degenerada. Sea $U \sim U(0,1)$ y sean

$$V_n(u) = \begin{cases} 0 & \text{si } u > \frac{1}{n}, \text{ con prob. } 1 - \frac{1}{n} \\ n^2 & \text{si } u \leq \frac{1}{n}, \text{ con prob. } \frac{1}{n} \end{cases}$$



Notamos que:

- $E\left(V_n\right) = 0 \cdot \left(1 - \frac{1}{n}\right) + n^2 \cdot \frac{1}{n} = n$. O sea,

$$\lim_{n\to\infty} E\left(V_n\right) = +\infty$$

- Sin embargo, $V_n \xrightarrow{P} 0$ porque

$$P\left(\left|V_n - 0\right| \geq \varepsilon\right) = P\left(V_n = n\right) = \frac{1}{n}$$

$V_n \xrightarrow{p} V$ **no implica** que $E(V_n) \overset{n\to\infty}{\to} E(V)$: ¿por qué?

Lo importante para llevarse de estos ejemplos “patológicos” es que las variables que estamos considerando V_n toman con probabilidad cada vez menor $\frac{1}{n}$ un valor cada vez más grande, (n o n^2) de manera que la relación entre esa probabilidad $\frac{1}{n}$ y el valor extremo (n o n^2) hace que la media no sea cero o aumente indefinidamente, pero no modifica el límite en probabilidad que sigue siendo $a = 0$.

Esto es porque la probabilidad de que la variable aleatoria tome ese valor extremo es cada vez menor, de manera que tiende a cero.

Propiedades de la convergencia en probabilidad

Sea a_n es una sucesión de números reales tal que $a_n \underset{n\to\infty}{\longrightarrow} a$. Sean V_n y W_n dos variables aleatorias tales que $V_n \xrightarrow{P} V$ y $W_n \xrightarrow{P} W$. Entonces:

1. $a_n + V_n \xrightarrow{P} a + V$.
2. $a_n V_n \xrightarrow{P} aV$.
3. $V_n + W_n \xrightarrow{P} V + W$.
4. $V_n W_n \xrightarrow{P} VW$.
5. Si $P(W = 0) = 0$, $V_n / W_n \xrightarrow{P} V / W$.
6. Si $f : \mathbb{R} \to \mathbb{R}$ es continua luego $f(V_n) \xrightarrow{P} f(V)$.
7. Si $f : \mathbb{R}^2 \to \mathbb{R}$ es continua luego $g(V_n, W_n) \xrightarrow{P} g(V, W)$.

Convergencia en probabilidad de S_n^2 y $\widehat{\sigma}^2$ a $a = Var(X)$

Sean $X_1, \ldots, X_n, \ldots$ variables aleatorias independientes e idénticamente distribuidas. Muestre que:

1. ambas formas de escribir a $\widehat{\sigma_n^2}$ son equivalentes.
2. $S_n^2 = a_n \widehat{\sigma_n^2}$, donde $a_n = \dfrac{n}{n-1}$.
3. $\widehat{\sigma_n^2}$ y S_n^2 convergen en probabilidad a $\text{Var}(X)$ usando propiedades de convergencia y LGN apropiadamente.

donde:

- $\widehat{\sigma_n^2} = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\overline{X}_n\right)^2 = \dfrac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$
- $S_n^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$

A partir de $\widehat{\sigma_n^2}$ y S_n^2, ¿qué v.a. convergen en probabilidad a $\sqrt{Var(X)}$? Llámelas $\widehat{\sigma}_n$ y S_n respectivamente.

Ejercicios/ Ejemplos

Ejemplo 1

Sean $X_1, \ldots, X_n \overset{iid}{\sim} U(0,1)$, hallar el límite en probabilidad de[a]

$$Y_n = (X_1 X_2 \ldots X_n)^{1/n} = \left(\prod_{i=1}^{n} X_i \right)^{1/n}$$

[a] Tenga en cuenta que si $X \sim U(0,1)$ entonces $E(\ln(X)) = 1$. ¿Por qué vale?

Ejemplo 2

Sean $X_1, \ldots, X_n \overset{iid}{\sim} X$ con $E(X) = Var(X) = 1$, mostrar que

$$\frac{\sum\limits_{i=1}^{n} X_i}{\left(n \sum\limits_{i=1}^{n} X_i^2 \right)^{1/2}} \xrightarrow{P} \frac{1}{\sqrt{2}}$$

Convergencia en media cuadrática

Hasta ahora vimos que se puede demostrar que $V_n \xrightarrow{p} V$ usando la ley débil de los grandes números y por definición. Veamos otra definición que nos permitirá ver convergencia en probabilidad.

- **Definición:** Decimos que las variables V_n convergen en media cuadrática al número a, $V_n \xrightarrow{m.c.} a$, si:
  1. $E(V_n) \xrightarrow{n\to\infty} a$
  2. $Var(V_n) \xrightarrow{n\to\infty} 0$
- **Aplicación:** Si $V_n \xrightarrow{m.c.} a$ entonces vale $V_n \xrightarrow{P} a$.
- **Intuición**: Si el centro del histograma (densidad) de V_n se está acercando a a y la variabilidad de V_n se está acercando a cero cuando n crece, entonces toda la distribución de probabilidad de V_n se está concentrando sobre a cuando n crece (por tanto lo tanto, $V_n \xrightarrow{P} a$).

Demostración de que $V_n \xrightarrow{m.c.} a$ implica que $V_n \xrightarrow{P} a$

Caso 1: Suponemos que $E(V_n) = a$ para todo n. Sea $\varepsilon > 0$. Por Markov

$$P(|V_n - a| \geq \varepsilon) = P(|V_n - a|^2 \geq \varepsilon^2) \leq \frac{E\left((V_n - a)^2\right)}{\varepsilon^2} = \frac{Var(V_n)}{\varepsilon^2} \underset{n\to\infty}{\longrightarrow} 0\,.$$

Caso 2, caso general: Sea $\varepsilon > 0$. Por Markov

$$\begin{aligned}
&P(\,|V_n - a| \geq \varepsilon) = P(|V_n - a|^2 \geq \varepsilon^2) \leq \frac{E\left((V_n - a)^2\right)}{\varepsilon^2} \\
&\leq \frac{1}{\varepsilon^2} E\left((V_n - E(V_n) + E(V_n) - a)^2\right) \\
&= \frac{1}{\varepsilon^2}\left(E\left[V_n - E(V_n)^2\right] - 2E\left[(V_n - E(V_n))(E(V_n) - a)\right] + E\left[(E(V_n) - a)^2\right]\right) \\
&= \frac{1}{\varepsilon^2}\left(Var(V_n) - 2(E(V_n) - a)\cdot \underbrace{E\left[V_n - E(V_n)\right]}_{=0} + (E(V_n) - a)^2\right) \underset{n\to\infty}{\longrightarrow} 0
\end{aligned}$$

Ejemplo para ver que $V_n \overset{m.c.}{\to} \theta$

Si $X_1, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$, definimos

$$V_n = \text{máx}\{X_1, \ldots, X_n\}.$$

Queremos ver que $V_n \overset{m.c.}{\to} \theta$ y por lo tanto $V_n \overset{P}{\to} \theta$. Calculemos $E(V_n)$ y $Var(V_n)$. Para poder hacer esto, necesitamos conocer la distribución de V_n. Calculemos primero la función de distribución acumulada de V_n

$$\begin{aligned} P\left(V_n \le x\right) &= P(X_1 \le x, X_2 \le x, \ldots, X_n \le x) \\ &= P(X_1 \le x) \ldots P(X_n \le x) = P(X \le x)^n. \end{aligned}$$

Como $X \sim U(0,1)$

$$P(X \le x) = \begin{cases} 0 & x \le 0 \\ \frac{x}{\theta} & 0 \le x \le \theta \\ 1 & x \ge \theta \end{cases}$$

Luego, se tiene que

$$P\left(V_n \le x\right) = \begin{cases} 0 & x \le 0 \\ \left(\frac{x}{\theta}\right)^n & 0 \le x \le \theta \\ 1 & x \ge \theta. \end{cases}$$

Ejemplo para ver que $V_n \overset{m.c.}{\to} \theta$

Derivando, obtenemos la densidad de V_n es

$$f_{V_n}(x) = nx^{n-1}\frac{1}{\theta^n} I_{(0,\theta)}(x)\,.$$

Entonces

$$E(V_n) = \int_0^\theta nx^n \frac{1}{\theta^n} dx = \frac{n}{n+1}\theta,$$

$$E(V_n^2) = \int_0^\theta nx^{n+1} \frac{1}{\theta^n} dx = \frac{n}{n+2}\theta^2$$

y

$$Var(V_n) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2\,.$$

Como

$$E(V_n) \underset{n\to\infty}{\longrightarrow} \theta \text{ y } Var(V_n) \underset{n\to\infty}{\longrightarrow} 0,$$

por la proposición que vimos antes

$$V_n \overset{P}{\to} \theta$$

Teorema Central del Límite

Theorem (Teorema central de límite para $X_i \overset{iid}{\sim}$)

Sea $X_1, \ldots, X_n \overset{iid}{\sim} X$ con $E(X)$ y $Var(X)$, para todo $z \in \mathbb{R}$

$$P\left(\sqrt{n}\frac{(\overline{X}_n - E(X))}{\sqrt{Var(X)}} \leq z\right) \underset{n\to\infty}{\longrightarrow} P(Z \leq z)$$

donde $P(Z \leq z)$ con $Z \sim N(0,1)$. A veces se escribe $\Phi(z) = P(Z \leq z)$.

Escribimos

$$\sqrt{n}\frac{(\overline{X}_n - E(X))}{\sqrt{Var(X)}} \xrightarrow{d} N(0,1)$$

Observaciones sobre el TCL

- El TCL implica que, si $X_i \sim_{iid}$, entonces para cualesquiera $z_1, z_2 \in \mathbb{R}$ que cumplen que $z_1 \leq z_2$, vale

$$P\left(z_1 \leq \frac{\sqrt{n}\,(\overline{X}_n - E(X))}{\sqrt{Var(X)}} \leq z_2\right) \underset{n\to\infty}{\longrightarrow} P(Z \leq z_2) - P(Z \leq z_1)$$

- El TCL nos va a permitir (más adelante en la materia) *cuantificar la incertidumbre* (dar un **margen de error**) que tenemos sobre nuestra estimación de $E(X)$ usando la media muestral $\overline{X}_n$.
- El teorema se refiere a la distribución de $\overline{X}_n$, **no** a la de $X_1, \ldots, X_n$.
- Si $X_i \sim_{iid} N(\mu, \sigma^2)$, **no es necesario usar TCL** porque vale que, para cualquier n que: $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Notación correcta sobre el TCL

Notación correcta:

- **Distribución asintótica** (versión estandarizada)

$$\frac{\sqrt{n}\,(\overline{X}_n - E(X))}{\sqrt{Var(X)}} \xrightarrow{D} N(0,1)$$

- **Distribución aproximada** para **tamaños de muestra grande** ($n \gg 0$), la distribución de $\overline{X}_n$ es **aproximadamente normal**:
  - **Versión estandarizada:**

$$\frac{\sqrt{n}\,(\overline{X}_n - E(X))}{\sqrt{Var(X)}} \sim_a N(0,1) \text{ si } n \gg 0$$

  - **Versión no estandarizada:**

$$\overline{X}_n \sim_a N\left(E(X), \frac{Var(X)}{n}\right) \text{ si } n \gg 0.$$

TCL - ejemplos

Vale TCL para $g(X_i) \overset{iid}{\sim}$ porque $X_i \overset{iid}{\sim}$

- Si $X_i = g(Y_i)$ entonces $\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} g(Y_i) - E(g(Y))\right] \xrightarrow{d} N(0, Var(g(Y))$

<u>Ejercicio</u>

1. Si $T_i \sim_{iid}$, $E(T) = 8$ y $Var(T) = 225$, $g(T) = T^2$ y

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} T_i^2 - \theta\right] \xrightarrow{d} N(0, Var(T^2))$$

   ¿cuánto vale θ?

Teorema Central del Límite

¿Qué tan grande debe ser el tamaño de muestra n?

- Si la distribución de X es simétrica, es decir, si existe algún punto c tal que el histograma o la función de densidad de X es simétrica respecto de c, típicamente con $n > 20$ la aproximación normal del TCL es adecuada.
- Si la distribución de X no es simétrica, no hay una "receta" para decir qué tan grande debe ser n para que la distribución de $\overline{X}_n$ se "parezca" a una distribución normal. Mientras más grande sea el n, mejor.
- Para el caso donde $X_i \sim_{iid} Be(p)$ se suele decir que para que la aproximación de la distribución de $\overline{X}_n$ por una distribución normal sea adecuada, debería ocurrir que $np(1-p) > 5$. Notar que esta regla es una **heurística** y no tiene sustento teórico alguno.

Convergencia en distribución

Sea $V_1, \ldots, V_n, \ldots$ una sucesión de variables aleatorias y sea V una variable aleatoria. **Decimos que** V_n **converge en distribución a** V si, para todo v tal que la función de distribución acumulada de V, F_V es continua en v, vale que

$$F_{V_n}(v) = P(V_n \leq v) \underset{n\to\infty}{\longrightarrow} P(V \leq v) = F_V(v)\,.$$

La convergencia en distribución, más que una convergencia de variables aleatorias, **es una convergencia de funciones de distribución**.

Si $V_n \xrightarrow{d} V$, la función de distribución acumulada de V_n está "cerca" de la distribución de V para valores de n grandes.

Esto no quiere decir que V_n esté cerca de V!

De hecho, $V_1, \ldots, V_n$ y V pueden estar definidas todas sobre espacios muestrales distintos!

Notación: Notamos $V_n \xrightarrow{d} V$. También haremos el siguiente abuso de notación. Por ejemplo, si $V \sim \mathcal{N}(0,1)$, escribiremos $V_n \xrightarrow{d} \mathcal{N}(0,1)$,

Convergencia en dist. vs. convergencia en prob.

Si $V_n \xrightarrow{P} V$, entonces $V_n \xrightarrow{d} V$.

En general no es necesariamente cierto que si $V_n \xrightarrow{d} V$ entonces $V_n \xrightarrow{P} V$.

En otras palabras, la convergencia en distribución es más **débil** que la convergencia en probabilidad.

Consideremos una v.a. continua X con una función de probabilidad puntual $f_X(x)$ que cumpla ser una función par, es decir, $f_X(x) = f_X(-x)$. Se sigue que $-X$ tiene la misma PDF. Definimos la sucesión $V_n = X$ si n es impar y $V_n = -X$ si n es par, y sea $V = X$, entonces $F_{V_n}(z) = F_V(z)$ para todo z porque la PDF dw X y $-X$ son iguales. Entonces, $V_n \xrightarrow{d} V$. Sin embargo, $V_n \xrightarrow{p} V$ porque Z_N oscila entre X y $-X$. Son v.a. diferentes, aunque tengan la misma CDF, porque

$$P(X = -X) = P(\{\omega : X(\omega) = -X(\omega)\}) = P(\{\omega : X(\omega) = 0\}) = 0$$

Propiedades de la convergencia en distribución

- **Lema de Slutsky**

  Sea $c \in \mathbb{R}$ una constante. Si $W_n \xrightarrow{P} c$ y $V_n \xrightarrow{d} V$ entonces

  $$W_n V_n \xrightarrow{d} cV \quad \text{y} \quad V_n + W_n \xrightarrow{d} V + c$$

- **Teorema de la función continua**

  Las funciones continuas conservan los límites aún cuando sus argumentos son (sucesiones de) variables aleatorias. En otras palabras, si $g : \mathbb{R} \to \mathbb{R}$ es una función continua y:

  - $V_n \xrightarrow{P} c$ entonces $g(V_n) \xrightarrow{P} g(c)$.
  - $V_n \xrightarrow{d} V$ entonces $g(V_n) \xrightarrow{d} g(V)$.

- **Ojo!** No es cierto que si $V_n \xrightarrow{d} V$ y $W_n \xrightarrow{d} W$ entonces $V_n \cdot W_n \xrightarrow{d} V \cdot W$ o $V_n + W_n \xrightarrow{d} V + W$. Por ejemplo si $V_n = W_n = V \sim N(0,1)$ y $W = -V \sim N(0,1)$ se tiene que $V_n \xrightarrow{d} V$, $W_n \xrightarrow{d} W$, pero $V_n + W_n \sim N(0,4)$ mientras que $V + W = 0$.

Aplicación del lema de Slutsky

Volvamos al ejemplo de antes. Por TCL

$$\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1).$$

Vimos en la slide §36 que $\widehat{\sigma}_n \xrightarrow{P} \sigma$ lo que implica, por las propiedades de convergencia en probabilidad

$$\frac{\sigma}{\widehat{\sigma}_n} \xrightarrow{P} 1.$$

Ahora

$$\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\widehat{\sigma}_n} = \underbrace{\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\sigma}}_{\xrightarrow{d}\mathcal{N}(0,1)} \underbrace{\frac{\sigma}{\widehat{\sigma}_n}}_{\xrightarrow{P}1} \xrightarrow{d} \mathcal{N}(0,1),$$

por el lema de Slutsky. Luego,

$$\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\widehat{\sigma}_n} \xrightarrow{d} \mathcal{N}(0,1)$$

Implicaciones de las convergencias

$$V_n \ \overset{m.c.}{V} \underbrace{\Rightarrow}_{(1)} V_n \ \overset{p}{V} \underbrace{\Rightarrow}_{(2)} V_n \ \overset{D}{V}$$

La implicación (1) vale por la desigualdad de Chebyshev. La demostración de (2) queda fuera del nivel del curso.

Convergencia en distribución: Ejercicio

Ejercicio 1: Sea $X_1, \ldots, X_n$ una muestra aleatoria tal que $E(X_1) = 0$ y $E(X_1^2) = 2$. Calcule el límite en distribución de

$$Y_n = \frac{\sqrt{n}\sum_{i=1}^n X_i}{\sum\limits_{i=1}^n X_i^2} \text{ y de } W_n = e^{Y_n}.$$

Método delta univariado

Theorem (Método delta)

Supongamos que V_n *cumple que*

$$\sqrt{n}\,(V_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

donde μ *y* σ^2 *son constantes. Supongamos que* $g(x)$ *es una función con derivada continua y que* $g'(\mu) \neq 0$*. Entonces*

$$\sqrt{n}\,(g(V_n) - g(\mu)) \xrightarrow{d} N\left(0, \sigma^2 g'(\mu)^2\right)$$

Ejemplos de método delta univariado

- Sean $X_i \sim_{iid} Be(p)$. Muestre que $\sqrt{n}\left(\overline{X}_n - p\right) \xrightarrow{d} N(0, p(1-p))$. Calcule la distribución asintótica de $\sqrt{n}\left(g(\overline{X}_n) - g(p)\right)$, donde $g(x) = \frac{x}{1-x}$.
- Repita el ejercicio anterior con $g(x) = \ln\left(\frac{x}{1-x}\right)$.
- Sean $X_i \sim_{iid} Exp(\lambda)$. Muestre que $\sqrt{n}\left(\overline{X}_n - \frac{1}{\lambda}\right) \xrightarrow{d} N\left(0, \frac{1}{\lambda^2}\right)$. Vea entonces adónde converge en distribución $\sqrt{n}\left(g(\overline{X}_n) - g\left(\frac{1}{\lambda}\right)\right)$, donde $g(x) = \ln(x)$.
- Repita el ejercicio anterior con $g(x) = e^x$.
- Repita el ejercicio anterior con $g(x) = \frac{1}{x}$.

Método delta multivariado

Sea $(\underline{Z}_n)_{\{n\in\mathbb{N}\}}$ una sucesión de vectores columna con p-coordenadas de manera que para $\underline{a} \in \mathbb{R}^p$ y $\Sigma \in \mathbb{R}^{p\times p}$ vale

$$\sqrt{n}(\underline{Z} - \underline{a}) \xrightarrow{d} N_p(\underline{0}, \Sigma)$$

donde $\underline{0} = (0, \ldots, 0)^{\mathsf{T}} \in \mathbb{R}^p$.
Sea $g : \mathbb{R}^p \to \mathbb{R}^k$ una función continuamente diferenciable de $\underline{z} = (z_1, \ldots, z_p)$ Sea

$$\nabla(\underline{a}) = \begin{pmatrix} \frac{\partial}{\partial z_1} g_1(\underline{z})\Big|_{\underline{z}=\underline{a}} & \frac{\partial}{\partial z_1} g_2(\underline{z})\Big|_{\underline{z}=\underline{a}} & \cdots & \frac{\partial}{\partial z_1} g_k(\underline{z})\Big|_{\underline{z}=\underline{a}} \\ \frac{\partial}{\partial z_2} g_1(\underline{z})\Big|_{\underline{z}=\underline{a}} & \frac{\partial}{\partial z_2} g_2(\underline{z})\Big|_{\underline{z}=\underline{a}} & \cdots & \frac{\partial}{\partial z_2} g_k(\underline{z})\Big|_{\underline{z}=\underline{a}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial}{\partial z_p} g_1(\underline{z})\Big|_{\underline{z}=\underline{a}} & \frac{\partial}{\partial z_p} g_2(\underline{z})\Big|_{\underline{z}=\underline{a}} & \cdots & \frac{\partial}{\partial z_p} g_k(\underline{z})\Big|_{\underline{z}=\underline{a}} \end{pmatrix}$$

Método delta multivariado

Entonces si $\nabla(\underline{a})^{\mathsf{T}}\Sigma\nabla(\underline{a})$ es definido positivo, vale que

$$\sqrt{n}(g(\underline{Z}_n) - g(\underline{a}))^{\mathsf{T}} \xrightarrow{d} N_k(\underline{0}, \nabla(\underline{a})^{\mathsf{T}}\Sigma\nabla(\underline{a}))$$

Ejercicio: Consideremos $X_i \sim_{iid} Be(p)$ e $Y_i \sim_{iid} Be(q)$. Suponga adicionalmente que las variables X_i e Y_j son independientes para cualquier $i, j \in \{1, \ldots, n\}$.
Consideremos el vector $\underline{Z}_n = (\overline{X}_n, \overline{Y}_n)$ Veamos cuál es el límite en distribución de:

1. $g(a, b) = b - a$
2. $g(a, b) = (a - b, a + b)$

Notemos que en ambos casos $\underline{a} = (p, q)^{\mathsf{T}}$ y $\Sigma = \begin{pmatrix} p(1-p) & 0 \\ 0 & q(1-q) \end{pmatrix}$ y que,

por TCL

$$\sqrt{n}(\underline{Z}_n - \underline{a})^{\mathsf{T}} \xrightarrow{d} N\left((0, 0)^{\mathsf{T}}, \Sigma\right)$$

Método delta multivariado: ejemplos

Ejemplo 1: Para $g(a,b) = b - a$, evaluado en $\underline{Z}_n$ queda $g(\underline{Z}_n) = \overline{Y}_n - \overline{X}_n$.
Calculamos $\nabla(\underline{a}) = (-1 \quad 1)^{\mathsf{T}}$
Ejemplo 2: Para $g(a,b) = (a-b, a+b)$, evaluado en $\underline{Z}_n$ queda $g(\underline{Z}_n) = (\overline{Y}_n - \overline{X}_n, \overline{Y}_n + \overline{X}_n)$
Calculamos $\nabla(\underline{a}) = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$
Por lo tanto, por método delta multivariado vale que

$$\sqrt{n}(g(\underline{Z}_n) - g(\underline{a}))^{\mathsf{T}} \xrightarrow{d} N_k(\underline{0}, \nabla(\underline{a})^{\mathsf{T}}\Sigma\nabla(\underline{a}))$$

donde, para el **ejemplo 1**, $\nabla(\underline{a})^{\mathsf{T}}\Sigma\nabla(\underline{a}) = p(1-p) + q(1-q)$ donde, para el **ejemplo 2**, $\nabla(\underline{a})^{\mathsf{T}}\Sigma\nabla(\underline{a}) = \begin{pmatrix} r+s & r-s \\ r-s & r+s \end{pmatrix}$ donde $r = p(1-p)$ y $s = q(1-q)$

Apéndice

A partir de aquí las slides son optativas

Demostración de las propiedades 2 y 3 en §16

2. Si $a = 0$ es trivial. Si $a \neq 0$ $P(|aV_n - aV| > \varepsilon) = P\left(|V_n - V| > \frac{\varepsilon}{|a|}\right) \xrightarrow{n\to\infty} 0$
3. Sea $\varepsilon > 0$. Tenemos que ver que

$$P\left(|V_n + W_n - (V + W)| \geq \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0\,.$$

Notemos que

$$\{|V_n + W_n - (V + W)| \geq \varepsilon\} \subset \{|V_n - V| \geq \varepsilon/2\} \cup \{|W_n - W| \geq \varepsilon/2\}\ .$$

Entonces

$$\begin{aligned} P\left(|V_n + W_n - (V + W)| \geq \varepsilon\right) \leq P\left(|V_n - V| \geq \varepsilon/2\right) + \\ P\left(|W_n - W| \geq \varepsilon/2\right). \end{aligned}$$

Como $V_n \xrightarrow{P} V$ y $W_n \xrightarrow{P} W$

$$P\left(|V_n - V| \geq \varepsilon/2\right) + P\left(|W_n - W| \geq \varepsilon/2\right) \underset{n\to\infty}{\longrightarrow} 0\,.$$

Ejemplo de variables que no son asintóticamente normales

Consideremos que $X_i \sim_{iid} U(0,\theta]$ y las variables $Y_n = \text{máx}\{X_1,\dots,X_n\}$. Vimos anteriormente que $F_{Y_n}(x) = \frac{x^n}{\theta^n}$ si $0 \le x \le \theta$.
Afirmamos que

$$n(\theta - Y_n) \xrightarrow{d} Exp\left(\frac{1}{\theta}\right)$$

Veamos que $P(n(\theta - Y_n) \le x) \underset{n\to\infty}{\longrightarrow} 1 - e^{x/\theta}$

$$\begin{aligned}
P(n(\theta - Y_n) \le x) &= P\left(\theta - Y_n \le \frac{x}{n}\right) = P\left(\theta - \frac{x}{n} \le Y_n\right) \\
&= P\left(Y_n \ge \theta - \frac{x}{n}\right) = 1 - P\left(Y_n < \theta - \frac{x}{n}\right) \\
&= 1 - P\left(X_1 < \theta - \frac{x}{n}, X_2 < \theta - \frac{x}{n}, \dots, X_n < \theta - \frac{x}{n}\right) \\
&\underbrace{=}_{X_i \text{ indep}} 1 - P\left(X_1 < \theta - \frac{x}{n}\right) \cdots P\left(X_n < \theta - \frac{x}{n}\right) \\
&\underbrace{=}_{X_i \text{ i.d.}} 1 - \left(P\left(X_1 < \theta - \frac{x}{n}\right)\right)^n = 1 - \left(1 - \frac{x}{n\theta}\right)^n \underset{n\to\infty}{\longrightarrow} 1 - e^{-x/\theta}.
\end{aligned}$$

Demostración de TCL

Para hacer la demostración de TCL estamos asumiendo que las variables X_i, $i = 1, \cdots, n$ son i.i.d., que la FGM de X_i está definida en un entorno de $t = 0$.
Sin pérdida de generalidad hacemos la demostración para $\mu = 0$.
La función generatriz de momentos de X es $M_X(t) = E\left(e^{Xt}\right)$. Como las variables X_i son independientes si $i + 1, \ldots, n$, entonces vale que:

$$M_{\sum_{i=1}^{n} X_i}(t) \underbrace{=}_{X_i \text{ indep}} \prod_{i=1}^{n} M_{X_i}(t) = [M_X(t)]^n$$

Por lo tanto, considerando la variable aleatoria

$$Y_n = \frac{\overline{X}_n}{\sigma/\sqrt{n}} = \frac{\sum\limits_{i=1}^{n} X_i/n}{\sigma/\sqrt{n}} = \frac{\sum\limits_{i=1}^{n} X_i}{\sqrt{n} \cdot \sigma}$$

vale que, como $Y_n = \sum\limits_{i=1}^{n} a_i X_i$ con $a_i = \dfrac{1}{\sqrt{n} \cdot \sigma}$:

$$M_{Y_n}(t) = \left[M_X\left(\frac{t}{\sqrt{n} \cdot \sigma}\right)\right]^n$$

Demostración de TCL (cont)

Consideremos ahora el desarrollo de Taylor de la función generatriz de momentos $M_X(t)$ en un entorno de $s = 0$:

$$M_X(s) = \sum_{k=0}^{\infty} \frac{s^k}{k!} M_X^{(k)}(0)$$

donde $M_X^{(k)}$ es la k-ésima derivada de M. Remplazando[3] $s = \frac{t}{\sqrt{n}} \cdot \sigma$

$$M_X\left(\frac{t}{\sqrt{n}\cdot\sigma}\right) = M(0) + \frac{t}{\sqrt{n}\cdot\sigma} M'(0) + \left(\frac{t}{\sqrt{n}\cdot\sigma}\right)^2 \frac{1}{2} M''(0) + \varepsilon$$

donde los términos representados por ε son:[4]

$$\varepsilon = \textstyle\sum_{k=3}^{\infty} \left(\frac{t}{\sqrt{n}\cdot\sigma}\right)^k \frac{1}{k!} M^{(k)}(0) = \mathcal{O}\left(\frac{1}{n}\right)$$

[3] Notemos que si $s \approx 0$ entonces $t \approx 0$.

[4] $\mathcal{O}\left(\frac{1}{n}\right)$ es una función que cumple que $\lim\limits_{n\to+\infty} \frac{\mathcal{O}\left(\frac{1}{n}\right)}{\frac{1}{n}} = 0$.

Demostración de TCL (cont)

Como $M(0) = 1$, $M'(0) = E(X) = 0$ y $M''(0) = E\left(X^2\right) = \sigma^2$:

$$M_X\left(\frac{t}{\sqrt{n}\cdot\sigma}\right) = 1 + \frac{t}{\sqrt{n}\cdot\sigma}\cdot 0 + \left(\frac{t}{\sqrt{n}\cdot\sigma}\right)^2 \frac{1}{2}\sigma^2 + \varepsilon = 1 + \frac{t^2/2}{n} + \varepsilon$$

En base a lo anterior, resulta que:

$$M_{Y_n}(t) = \left[M_X\left(\frac{t}{\sqrt{n}\cdot\sigma}\right)\right]^n = \left(1 + \frac{t^2/2}{n} + \varepsilon\right)^n$$

Tomando límite $n \to \infty$, resulta:

$$\lim_{n\to\infty} M_{Y_n}(t) = \lim_{n\to\infty}\left(1 + \frac{t^2/2}{n} + \varepsilon\right)^n = e^{t^2/2} = M_Z(t)$$

donde $e^{t^2/2}$ es la función generatriz de momentos de la distribución normal estándar. Entonces

$$F_{Y_n}(x) \overset{n\to\infty}{\to} F_Z(x) \text{ y por lo tanto } Y_n \overset{d}{\to} Z, \text{ donde } Z \sim N(0,1).$$