Estadística y Optimización

Teorema Central del Límite

Sucesiones de v.a.i.i.d: una sucesión de variables aleatorias:

$$X_1, X_2, \ldots, X_n, \ldots$$

- son *variables aleatorias independientes e idénticamente distribuidas* (v.a.i.i.d) si:
  - todas son independientes,
  - siguen la misma distribución, es decir, $F_{X_i}(x) = F_{X_j}(x)$ para todo x y todo par X_i y X_j.
- Para todo $n \geq 1$, definimos:

$$S_n := X_1 + \cdots + X_n, \qquad \overline{X}_n := \frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n}$$

  Cuidado: no son números, ¡son variables aleatorias!

Ejemplos:

- Repetir un experimento aleatorio un número arbitrario de veces y calcular el promedio de los resultados.
- Daño estructural acumulado a largo plazo en un trozo de puente por el paso de vehículos.
- Las ganancias a largo plazo por un volumen diario de ganancias.
- Acumulación a largo plazo de fallos en distintas máquinas.
- Aproximación de la distribución teórica de la media de una determinada característica en una muestra grande.

Propiedad: Si tenemos una sucesión de v.a.i.i.d. con media $E[X_i] = \mu$ y varianza $V[X_i] = \sigma^2 > 0$ finitas, entonces:

$$\boxed{\begin{aligned} E[S_n] &\overset{\text{linear.}}{=} n\mu \\ V[S_n] &\overset{\text{indep.}}{=} n\sigma^2 \end{aligned}}$$

$$\boxed{\begin{aligned} E[\overline{X}_n] &= \mu \equiv \text{const.}^{\text{e}} \\ V[\overline{X}_n] &= \tfrac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0 \end{aligned}}$$

Ley débil de los grandes números: Para cualquier número $\varepsilon > 0$, se tiene que

$$P\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) \xrightarrow[n\to\infty]{} 1.$$

Interpretación: El promedio de los resultados asociados a repetir un mismo experimento se va concentrando junto a la media esperada del experimento conforme repetimos más y más.

☞ La idea (central) del Teorema Central del Límite es que, de hecho, este promedio $\overline{X}_n$ se comporta como una normal $\mathcal{N}(\mu, \sigma/\sqrt{n})$ para n suficientemente grande.

☞ **Teorema Central del Límite:** Sea $X_1, X_2, \ldots, X_n, \ldots$ una sucesión de v.a.i.i.d. con media $E[X_i] = \mu$ y varianza $V[X_i] = \sigma^2 > 0$ finitas. Se tiene que:

$$P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq t\right) \xrightarrow[n\to\infty]{} \Phi(t), \qquad \text{para todo } t \in \mathbb{R}.$$

donde $\Phi(t)$ es la función de distribución de la normal tipificada $\mathcal{N}(0,1)$.
Equivalentemente,

$$P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq t\right) \xrightarrow[n\to\infty]{} \Phi(t), \qquad \text{para todo } t \in \mathbb{R}.$$

☞ **Notar:** no hemos supuesto ¡nada! sobre como las X_i están distribuidas excepto que tienen la misma distribución (la que sea), que son variables independientes y que tienen media μ y varianza σ^2 finitas.

☞ **Aplicación en la práctica del TCL:** se considera que, para $n \geq 30$, tenemos para todo $x \in \mathbb{R}$,

- ☞ $P(S_n \leq x) \approx P(Y \leq x)$ con $Y \sim \mathcal{N}\left(E[S_n], \sqrt{V[S_n]}\right) = \mathcal{N}(n\mu, \sqrt{n}\sigma)$.
- ☞ $P\left(\overline{X}_n \leq x\right) \approx P(Y \leq x)$ con $Y \sim \mathcal{N}\left(E[\overline{X}_n], \sqrt{V[\overline{X}_n]}\right) = \mathcal{N}(\mu, \sigma/\sqrt{n})$.

☞ **Ejercicio:** Las ventas diarias de una empresa siguen una distribución uniforme entre 2000 y 4000 euros. Suponiendo independientes las ventas de los distintos años, calcular la probabilidad de que el volumen de ventas anual supere los 920 000 euros, si la empresa trabaja 300 días al año.

Solución: Para cada día, tenemos X_i = "volumen de venta diaria (en miles de €) " $\sim \mathcal{U}([2,4])$ con $E[X_i] = \frac{2+4}{2} = 3$ y $V[X_i] = \frac{(4-2)^2}{12} = \frac{1}{3}$. Por lo que:

$$X = \text{"volumen anual de ventas (en miles €)"} = X_1 + \cdots + X_{300} \quad \text{con} \quad \begin{cases} E[X] = 300 \times 3 = 900 \\ V[X] = \frac{300}{3} = 100. \end{cases}$$

Por el TCL, podemos aproximar X por una normal con $\mu = 900$ y $\sigma = \sqrt{100} = 10$, es decir: $Y \sim \mathcal{N}(900, 10)$. Tenemos, mediante aproximación y tipificado $Z = \frac{Y-900}{10} \sim \mathcal{N}(0,1)$:

$$P(X > 920) \approx P(Y > 920) = P\left(Z > \frac{920 - 900}{10}\right) = P(Z > 2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228.$$

□

☞ **Aproximación de la binomial por una normal:** Una binomial $X \sim \mathcal{B}(n,p)$ es ya una suma de v.a.i.i.d.

$$X = X_1 + \cdots + X_n \quad \text{con} \quad X_i \sim \mathcal{B}er(p) \quad \text{y} \quad \begin{cases} E[X] = np \\ V[X] = npq \end{cases},$$

por lo que podemos aproximarla por una normal $Y \sim \mathcal{N}(np, \sqrt{npq})$ (este resultado fue demostrado históticamente por De Moivre y Laplace hace 200 años).

- **La corrección de Yates:** Al aproximar probabilidades de una v.a. discreta X por una contínua Y, se obtienen mejores aproximaciones al "agrandar" en 0.5 los extremos cerrados de los intervalos de valores donde cae nuestra v.a. contínua con la que queremos aproximar (ver Figura 1):

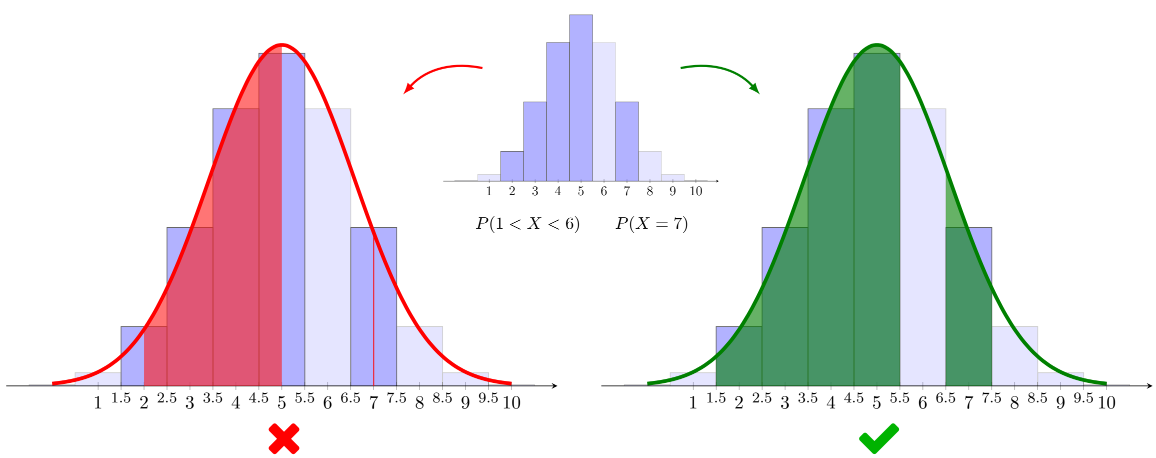
Corrección de Yates:

- $P(X \in [a, b]) \approx P(Y \in [a - 0.5, b + 0.5])$.
- $P(X \in [a, +\infty)) \approx P(Y \in [a - 0.5, +\infty))$.
- $P(X \in (-\infty, b]) \approx P(Y \in (-\infty, b])$.
- $P(X = k) \approx P(Y \in [k - 0.5, k + 0.5])$.

Ejemplos:

- $P(1 < X < 6) = P(2 \leq X \leq 5) \approx P(1.5 \leq Y \leq 5.5)$.
- $P(X > 3) = P(X \geq 4) \approx P(Y \geq 3.5)$.
- $P(2 < X) = P(3 \leq X) \approx P(2.5 \leq Y)$.
- $P(X = 7) \approx P(6.5 \leq Y \leq 7.5)$.

- ¡Esto nos permite calcular probabilidades $P(X = k)$ que de otro modo serían nulas para la contínua!
- **Cuidado:** Cuando tenemos un intervalo abierto en la v.a. discreta X, se puede expresar primero como uno con extremo(s) cerrado(s) antes de aplicar la corrección, como se muestra en los ejemplos.



$$P(1 < X < 6) \approx P(2 \leq Y \leq 5)$$
$$P(X = 7) \approx P(Y = 7) = 0$$

$$P(1 < X < 6) \approx P(1.5 \leq Y \leq 5.5)$$
$$P(X = 7) \approx P(6.5 \leq Y \leq 7.5)$$

Figura 1: La corrección de Yates: conviene agrandar los intervalos cerrados en 0.5 por cada extremo para obtener una mejor aproximación de la suma discreta por el área integral..

- **Ejemplo:** Si $X \sim \mathcal{B}(120, 1/6)$ y queremos calcular $P(15 < X < 20)$, podemos aproximar por una distribución normal

$$Y \sim \mathcal{N}(\mu, \sigma) \text{ con media } \mu = 120 \times (1/6) = 20 \text{ y } \sigma = \sqrt{120 \times (1/6) \times (5/6)} = \sqrt{50/3}.$$

Usando la corrección de Yates y tipificando $Z = \frac{Y-120}{\sqrt{50/3}} \sim \mathcal{N}(0, 1)$:

$$\begin{aligned} P(15 < X < 20) &\approx P(15.5 \leq Y \leq 19.5) \\ &= P\left(\frac{15.5 - 20}{\sqrt{50/3}} \leq \frac{Y - 20}{\sqrt{50/3}} \leq \frac{19.5 - 20}{\sqrt{50/3}}\right) \\ &= P(-1.10 \leq Z \leq -0.12) \overset{\text{sime.}}{=} P(0.12 \leq Z \leq 1.10) \\ &= \Phi(1.10) - \Phi(0.12) = 0.8643 - 0.5478 \\ &= 0.3165. \end{aligned}$$