Teorema Central del límite

Basado en la clase de Estadistica para químicos por
Mariela Sued, Manuela Cerdeiro y Florencia Statti

Suma de variables Normales

- X_i i.i.d., $X_i \sim \mathcal{N}(\mu, \sigma^2)$
- Suma de normales es normal:

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \sim \mathcal{N}(0,1)$$

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{n} X_i \leq u\right) &= \mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \leq \frac{u - n\mu}{\sqrt{n\sigma^2}}\right) \\
&= \phi\left((u - n\mu)/\sqrt{n\sigma^2}\right) \\
&= \texttt{pnorm}\left((u - n\mu)/\sqrt{n\sigma^2}\right)
\end{aligned}$$

Promedio de variables Normales

- X_i i.i.d., $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mathbb{E}(\overline{X}_n), V(\overline{X}_n)\right) \ , \quad \overline{X}_n \sim \mathcal{N}\left(\mu, \sigma^2/n\right) \ ,$$

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

$$\begin{aligned} \mathbb{P}\left(\overline{X}_n \leq u\right) &= \mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \leq \frac{u-\mu}{\sqrt{\sigma^2/n}}\right) \\ &{\color{red}=} \phi\left(\sqrt{n}(u-\mu)/\sqrt{\sigma^2}\right) = \texttt{pnorm}\left(\sqrt{n}(u-\mu)/\sqrt{\sigma^2}\right) \end{aligned}$$

Teorema Central del Límite (TCL): versión suma

- Sean $(X_i)_{i\geq 1}$ v.a. i.i.d. con $\mathbb{E}(X) = \mu$ y $V(X) = \sigma^2$.
- ¿Qué sabemos de la suma ?

$$\sum_{i=1}^{n} X_i \ , \quad \mathbb{E}(\sum_{i=1}^{n} X_i) = n\mu, \quad V(\sum_{i=1}^{n} X_i) = n\sigma^2$$

TEOREMA: La distribución de la suma $\sum_{i=1}^{n} X_i$ SE PARECE a la de una normal, con su esperanza $(\mathbb{E}(\sum_{i=1}^{n} X_i))$ y su varianza $(V(\sum_{i=1}^{n} X_i))$

TCL: La suma tiene distribución Aproximadamente Normal

- X_i i.i.d., $\mathbb{E}(X_i) = \mu$, $V(X_i) = \sigma^2$

$$\sum_{i=1}^{n} X_i \overset{a}{\approx} \mathcal{N}\left(\mathbb{E}(\sum_{i=1}^{n} X_i), V(\sum_{i=1}^{n} X_i)\right) , \quad \sum_{i=1}^{n} X_i \overset{a}{\approx} \mathcal{N}\left(n\mu, n\sigma^2\right) ,$$

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \overset{a}{\approx} \mathcal{N}(0,1) .$$

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{n} X_i \leq t\right) &= \mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \leq \frac{t - n\mu}{\sqrt{n\sigma^2}}\right) \\ &\approx \texttt{pnorm}\left((t - n\mu)/\sqrt{n\sigma^2}\right) \end{aligned}$$

Teorema Central del Límite (TCL) - Suma:

Sean $(X_i)_{i\geq 1}$ v.a. i.i.d. $X \sim X_i$, con $\mathbb{E}(X) = \mu$ y $V(X) = \sigma^2$, entonces tenemos que

$$\mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \leq x\right) \underset{n\to\infty}{\longrightarrow} \phi(x) = \texttt{pnorm(x)} \ , \quad \forall x \in \mathbb{R} \ ,$$

TCL - Versión Promedio

- X_i i.i.d. $\mathbb{E}(X_i) = \mu$, $V(X_i) = \sigma^2$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \ , \quad \mathbb{E}(\overline{X}_n) = \mu \ , \quad V(\overline{X}_n) = \sigma^2/n$$

TEOREMA: La distribución del promedio $\overline{X}_n$ SE PARECE a la de una normal, con su esperanza $\left(\mathbb{E}\left(\overline{X}_n\right)\right)$ y su varianza $(V(\overline{X}_n))$

TCL: El promedio tiene distribución Aproximadamente Normal

- X_i i.i.d., $\mathbb{E}(X_i) = \mu$, $V(X_i) = \sigma^2$

$$\overline{X}_n \overset{a}{\approx} \mathcal{N}\left(\mathbb{E}(\overline{X}_n), V(\overline{X}_n)\right)\ , \quad \overline{X}_n \overset{a}{\approx} \mathcal{N}\left(\mu, \sigma^2/n\right)\ ,$$

$$\frac{\overline{X}_n - \mu}{\sqrt{V(\overline{X}_n)}} \overset{a}{\approx} \mathcal{N}(0,1)\ ,\ \frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \overset{a}{\approx} \mathcal{N}(0,1)\ ,$$

$$\mathbb{P}\left(\overline{X}_n \leq x\right) \approx \phi\left((x-\mu)/\sqrt{\sigma^2/n}\right)$$

$$\mathbb{P}\left(\overline{X}_n \leq x\right) \approx \texttt{pnorm}\left((x-\mu)/\sqrt{\sigma^2/n}\right)$$

Teorema Central del Límite (TCL) - Promedio:

Sean $(X_i)_{i\geq 1}$ v.a. i.i.d. $X \sim X_i$, con $\mathbb{E}(X) = \mu$ y $V(X) = \sigma^2$, entonces tenemos que

$$\mathbb{P}\left(\frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} \leq x\right) \underset{n\to\infty}{\longrightarrow} \phi(x) = \texttt{pnorm(x)}\ , \quad \forall x \in \mathbb{R}\ ,$$

o equivalentemente

$$\mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \leq x\right) \underset{n\to\infty}{\longrightarrow} \phi(x) = \texttt{pnorm(x)}\ , \quad \forall x \in \mathbb{R}\ .$$

Ejemplo (del apunte de Ana Bianco y Elena Martinez)

Supongamos que, al sumar números, reondeamos cada número a un decimal. Los errores de aproximación se suponen independientes y con distribución U(-0.05,0.05).

a) Si se suman 1500 números, ¿cuál es la probabilidad de que el valor absoluto del error total exceda 1.5?

b) ¿Cuántos números pueden sumarse a fin de que el valor absoluto del error total sea menor o igual que 1 con probabilidad mayor o igual que 0.90?

Si llamamos X_i al error correspondiente al i-ésimo sumando, el error total es

$$T_{1500} = \sum_{i=1}^{1500} X_i.$$

Entonces,

$$E(T_{1500}) = 1500E(X_1) = 0$$

$$V(T_{1500}) = 1500V(X_1) = 1500/1200 = 1.25$$

$$P\left(|T_{1500}| > 1.5\right) = 1 - P\left(|T_{1.500}| \leq 1.5\right)$$

$$= 1 - P\left(-1.5 \leq T_{1500} \leq 1.5\right)$$

$$= 1 - P\left(\frac{-1.5}{\sqrt{1500/1200}} \leq \frac{T_{1500}}{\sqrt{1500/1200}} \leq \frac{1.5}{\sqrt{1500/1200}}\right) \cong_{\text{(por TCL)}}$$

$$1 - \Phi\left(\frac{1.5}{\sqrt{1500/1200}}\right) + \Phi\left(\frac{-1.5}{\sqrt{1500/1200}}\right) =$$

$$= 1 - \Phi(1.34) + \Phi(-1.34) = 0.18$$

b) Buscamos el valor de n tal que

$$P(|T_n| \leq 1) \geq 0.90$$
$$P(T_n| \leq 1) \geq 0.90 \Leftrightarrow P(-1 \leq T_n \leq 1) \geq 0.90$$
$$\Leftrightarrow P\left(\frac{-1}{\sqrt{n/1200}} \leq \frac{T_n}{\sqrt{n/1200}} \leq \frac{1}{\sqrt{n/1200}}\right) \geq 0.90$$

Aplicando la aproximación normal, debemos hallar n tal que

$$\Phi\left(\frac{1}{\sqrt{n/1200}}\right) - \Phi\left(\frac{-1}{\sqrt{n/1200}}\right) \geq 0.90$$

$$\Leftrightarrow 2\Phi\left(\frac{1}{\sqrt{n/1200}}\right) - 1 \geq 0.90 \Leftrightarrow \Phi\left(\frac{1}{\sqrt{n/1200}}\right) \geq 0.95$$
$$\Leftrightarrow \frac{1}{\sqrt{n/1200}} \geq 1.64 \Leftrightarrow \sqrt{n} \leq 21.12 \Leftrightarrow n \leq 446$$

es decir, que se pueden sumar a lo sumo 446 números para que el valor absoluto del error total sea menor o igual que 1 con probabilidad mayor o igual que 0.90 .