Teorema del límit central

El TLC ens diu que la distribució de la suma de moltes variables aleatòries independents i idènticament distribuïdes, independentment de la seva llei, es comporta aproximadament com una normal.

Primer parlarem segurament de la variable més utilitzada a l'estadística. Apareix en molts fenòmens naturals.

Definició. Una v.a. X segueix una llei normal o Gaussiana amb paràmetres μ i σ^2, i ho escriurem $N(\mu, \sigma^2)$, si té per densitat

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \quad \forall x \in \mathbb{R}.$$

Observacions.

1. $\mathbf{E}(X) = \mu$ i $\mathrm{Var}(X) = \sigma^2$.
2. La seva funció de distribució no es pot calcular.

Definició. Una v.a. X segueix una llei normal o Gaussiana estàndard si $X \sim N0, 1)$, és a dir, si té per densitat

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}, \quad \forall x \in \mathbb{R}.$$

Proposició. Tenim:

- $X \sim N(\mu, \sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0, 1)$.
- $Y \sim N(0, 1) \Rightarrow \sigma Y + \mu \sim N(\mu, \sigma^2)$.

Proposició.

- $X \sim N(\mu, \sigma^2) \Rightarrow aX + b \sim N(a\mu + b, a^2\sigma^2)$.
- $X \sim N(\mu, \sigma^2)$ i $Y \sim N(\nu, \rho^2)$ independents $\Rightarrow X + Y \sim N(\mu + \nu, \sigma^2 + \rho^2)$.
- $X \sim N(\mu, \sigma^2)$ i $Y \sim N(\nu, \rho^2)$ independents, aleshores

$$aX + bY \sim N(a\mu + b\nu, a^2\sigma^2 + b^2\rho^2).$$

La mitjana mostral

Tenim $X_1, \ldots, X_n$ una mostra aleatòria d'una distribució donada independents i idènticament distribuïdes (vaiid).

La **mitjana mostral** $\bar{X}_n$ d'una mostra aleatòria $X_1, \ldots, X_n$ és

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_n.$$

Proposició. Suposem $\mu = \mathbf{E}(X_1)$ i $\sigma^2 = \mathrm{Var}(X_1)$. Aleshores

- $\mathbf{E}(\bar{X}_n) = \mu$.
- $Var(\bar{X}_n) = \frac{\sigma^2}{n}$.

Proposició. Si $X_i \sim N(\mu_i, \sigma_i^2), i = 1, \ldots, n$, independents, aleshores

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma^2\right).$$

Proposició. Siguin $X_1, \ldots, X_n$ vaiid amb distribució $N(\mu, \sigma^2)$, aleshores

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Per tant,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

La desviació mostral

La **desviació mostral** és

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

Proposició. Siguin $X_1, \ldots, X_n$ vaiid amb distribució $N(\mu, \sigma^2)$, aleshores

$$\sqrt{n-1}\frac{\bar{X}_n - \mu}{S_n} \sim t_{n-1},$$

on t_{n-1} és una distribució t de Student amb n-1 graus de llibertat.

La t_n és una variable aleatòria simètrica al voltant del 0, amb un dibuix semblant al de la normal, amb $\mathbf{E}(X) = 0$, $\mathrm{Var}(X) = \frac{n}{n-2}$. Quan n és superior a 30 es comporta com una normal.

Teorema del límit central

Siguin $X_1, \ldots, X_n$ vaiid amb esperança μ i variància σ^2. Aleshores, quan n és superior a 30,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Si desconeixem σ,

$$\sqrt{n-1}\frac{\bar{X}_n - \mu}{S_n} \sim N(0,1).$$