STT 1000 - STATISTIQVES

ARTHVR CHARPENTIER

Loi de Student

Soient $X_1, \cdots, X_n$ des variables i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$, posons

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ et } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

alors

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \text{ mais } \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim ?$$

Cette loi est importante car la variance est (généralement) inconnue ?

C'est la loi étudiée par William Gosset (Student)

Loi de Student

Loi de Student - $\mathcal{S}td(\nu)$

Y suit une loi de Student de paramètre ν (appelé degré de liberté),

$$f(y) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

Soient

$$\begin{cases} Z \text{ suit un loi normale } \mathcal{N}(0,1) \\ V \text{ suite une loi du chi-deux à } \nu \text{ degrés de liberté} \\ Z \text{ et } V \text{ sont indépendent} \end{cases}$$

$$T = \frac{Z}{\sqrt{V/\nu}} = Z\sqrt{\frac{\nu}{V}} \sim \mathcal{S}td(\nu)$$

Loi de Student

$$\mathbb{E}[Y] = 0,\ \text{Var}[Y] = \frac{\nu}{\nu - 2}\ (\text{si } \nu > 2)$$

Note: si $\nu \to \infty$, $Y \sim \mathcal{N}(0, 1)$

Loi de Student

Soient $X_1, \cdots, X_n$ des variables i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$, posons

- $\overline{X}_n = \dfrac{1}{n}(X_1 + \cdots + X_n)$ suit une loi $\mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$
- $Z = (\overline{X}_n - \mu)\,\dfrac{\sqrt{n}}{\sigma}$
- si $S_n^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$, $V = (n-1)\dfrac{S_n^2}{\sigma^2} \sim \chi^2(n-1)$

(cf théorème de Cochrane)

Loi de Student et théorème central limite

$$T = \frac{Z}{\sqrt{V/\nu}} = (\overline{X}_n - \mu)\,\frac{\sqrt{n}}{S_n} \sim \mathcal{S}td(\nu)$$

Exemple

EXEMPLE

Dans une usine, une machine est réglée de telle sorte que le poids du produit qu'elle verse dans un contenant est distribué selon une loi normale. Un échantillon aléatoire de 10 contenants prélevé dans la production d'une journée donne un poids moyen $\overline{x} = 200{,}1$ g et un écart type corrigé $s = 2{,}5$ g. Estimer, à partir de cet échantillon, le poids moyen par contenant pour l'ensemble de la production de la journée. Le niveau de confiance est fixé à 95 %.

(via Simard (2015))