Chapitre 5 : échantillons gaussiens

• Loi du χ^2 :

$Y_1, \dots, Y_d$ sont i.i.d. de loi $\mathcal{N}(0,1)$

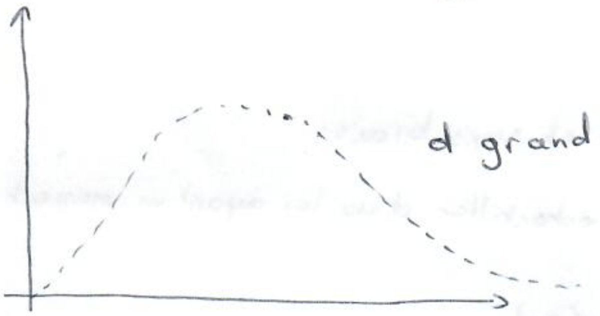
$$Z = Y_1^2 + \dots + Y_2^2 \sim \chi^2(d)$$

degré de liberté.

$E(Z) = d$ $\qquad E(Y_1^2) = E[(Y_1 - 0)^2] = Var(Y_1) = 1$

$Var(Z) = 2d$

Densité de Z : $$f_Z(x) = \frac{1}{2^{d/2}\underbrace{\Gamma\left(\frac{d}{2}\right)}_{\int_0^\infty y^{\frac{d}{2}-1} e^{-y}\,dy}} x^{\frac{d}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{x>0}$$



Application : thermomètre → température X gaussienne, moyenne 1600.

on veut donner un IC à 99% de la qualité d'un nouveau thermomètre.

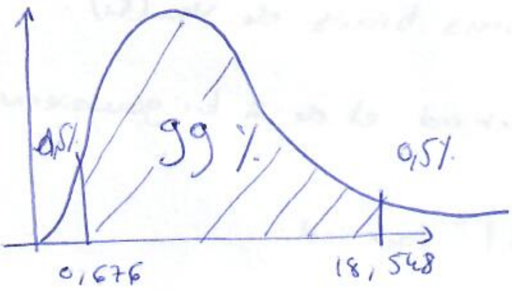
$X_1, \dots, X_6$ i.i.d. $\mathcal{N}(1600, \sigma^2)$

$$\frac{X_1 - 1600}{\sigma}, \dots, \frac{X_6 - 1600}{\sigma} \text{ iid } \mathcal{N}(0,1)$$

donc $$Z = \sum_{i=1}^{6} \left(\frac{X_i - 1600}{\sigma}\right)^2 \sim \chi^2(6)$$

or $P(Z > 18{,}548) = 0{,}005$

$P(Z > 0{,}676) = 0{,}995$



D'où $P(0{,}676 < Z < 18{,}548) = 0{,}99$

$$\Leftrightarrow P\left(0{,}76 < \sum \left(\frac{X_i - 1600}{\sigma}\right)^2 < 18{,}548\right) = 0{,}99$$

$$\Leftrightarrow P\left(\sqrt{\frac{\sum (X_i - 1600)^2}{18{,}548}} < \sigma < \sqrt{\frac{\sum (X_i - 1600)^2}{0{,}76}}\right) = 0{,}99$$

Ainsi IC → intervalle de confiance bilatéral à 99% pour σ :

$$IC = \left[\sqrt{\frac{c}{0{,}76}} \,;\, \sqrt{\frac{c}{18{,}548}}\right]$$

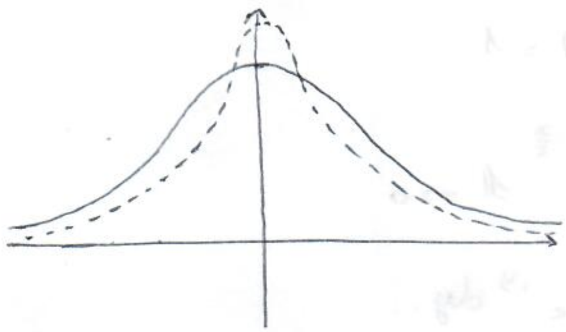
• Lois de Student

Si $Y_0, Y_1, \dots, Y_d$ sont iid de loi $\mathcal{N}(0,1)$

$$T = \frac{Y_0}{\sqrt{\frac{1}{d}(Y_1^2 + \dots + Y_d^2)}} \sim \text{Student}(d)$$

↳ d degrés de liberté.

densité de T : $f_T(x) = \frac{1}{\sqrt{d\pi}} \cdot \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \left(1 + \frac{x^2}{d}\right)^{-\frac{d+1}{2}}$



• $d = 1$ Student(1) = Cau (0, 1)

↳ pas d'espérance

• $d = 2$ $E(T) = 0$ → pas d'espérance

• $d \geq 3$ $E(T) = 0$ $\text{Var}(T) = \frac{d}{d-2}$

(loi tabulée)

• Théorème de Student

Estimateur : • fortement consistant et sans biais :

On prend $X_1, \dots, X_n$ échantillon d'une loi ayant un moment d'ordre 2.

$E(\bar{X}_n) = E(X_1)$

$\text{var}(\bar{X}_n) = \frac{1}{n}\text{Var}(X_1)$

On sait que $\boxed{\bar{X}_n \xrightarrow[n \to \infty]{p.s.} E(X_1)}$ LFGN alors $\bar{X}_n$ est un estimateur fortement consistant et sans biais de $E(X_1)$.

• fort. consistant et biaisé

On sait que $\boxed{V_n \xrightarrow[n \to \infty]{p.s.} \text{Var}(X_1)}$ LFGN

V_n est un estimateur fortement consistant et biaisé de $\text{Var}(X_1)$.

car $E(V_n) = \left(\frac{n-1}{n}\right)\text{Var}(X_1)$.

d'où $V_n^* = \frac{n}{n-1} V_n$ est un est. fort. consi. et sans biais de $\text{Var}(X_1)$.

Théorème de Student → Lorsque $X_1, \dots, X_n$ sont ind. et de m̂ loi gaussienne.

$\mathcal{N}(m, \sigma^2)$ avec $\sigma > 0$, on a :

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ et $V_n = \frac{1}{n}\sum(X_i - \bar{X}_n)^2$ sont ind.

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

$$\boxed{\frac{n}{\sigma^2} V_n = \frac{n-1}{\sigma^2} V_n^* \sim \chi^2(n-1)}$$

et $$\boxed{\sqrt{n-1}\,\frac{\bar{X}_n - m}{\sqrt{V_n}} \sim \text{Stu}(n-1)}$$

Exemple : on étudie la taille des arbres dans une forêt.

La hauteur de i ème arbre est notée $X_i \sim \mathcal{N}(m, \sigma^2)$

les X_i sont ind. On veut un intervalle à 95% pour m et σ^2.

- $X_1, \dots, X_{10}$ iid $\mathcal{N}(m, \sigma^2)$, $\sigma > 0$

$$\sqrt{9}\, \frac{\overline{X_{10}} - m}{\sqrt{V_{10}}} \sim \text{Student}(9)$$

on cherche a et b tq $P(T_9 > b) \simeq 0{,}025 \Rightarrow a = -b$

$P(T_9 > a) \simeq 0{,}975 \Rightarrow b = 2{,}62$

$$P\left(-2{,}262 \leq \sqrt{9}\, \frac{\overline{X_{10}} - m}{\sqrt{V_{10}}} \leq 2{,}262\right) \simeq 95\%$$

$$\Leftrightarrow P\left(\overline{X_{10}} - 2{,}262 \times \frac{\sqrt{V_{10}}}{\sqrt{9}} \leq m \leq \overline{X_{10}} + 2{,}262 \times \frac{\sqrt{V_{10}}}{\sqrt{9}}\right) \simeq 0{,}95\%$$

D'où l'interval de confiance pour m à 95% :

$$I = \left[\overline{X_{10}} \pm 2{,}262 \times \frac{\sqrt{V_{10}}}{\sqrt{9}}\right]$$

- $X_1, \dots, X_{10}$ iid $\mathcal{N}(m, \sigma^2)$ $\sigma^2 > 0$.

D'après student : $\frac{10}{\sigma^2} V_{10} \sim \chi^2(9)$

$P\left(\frac{10}{\sigma^2} V_{10} > a\right) \simeq 97{,}5\% \Rightarrow a = 2{,}700$

$P\left(\frac{10}{\sigma^2} V_{10} > b\right) \simeq 0{,}025 \Rightarrow b = 19{,}023.$

$$P\left(2{,}700 < \frac{10}{\sigma^2} V_{10} < 19{,}023\right) \simeq 95\%$$

$$\Leftrightarrow P\left(2{,}700 \times \frac{1}{10V_{10}} < \frac{1}{\sigma^2} < 19{,}023 \times \frac{1}{10V_{10}}\right) \simeq 95\%$$

donc $IC = \left[\frac{10 V_{10}}{19{,}023} ; \frac{10 V_{10}}{2{,}700}\right]$.

Théorème de Student :

si $X_1, \dots, X_n$ iid de loi $\mathcal{N}(m, \sigma^2)$ avec $\sigma^2 > 0$

1) $\overline{X_n}$ et V_n st ind.

2) $\overline{X} \sim \mathcal{N}(m, \sigma^2)$

3) $\frac{nV_n}{\sigma^2} \sim \chi^2(n-1)$

4) $\sqrt{n-1}\, \frac{\overline{X_n} - m}{\sqrt{V_n}} \sim \text{Student}(n-1)$.