Chap 2 : Théorèmes limites en probabilités utilisés en statistique.

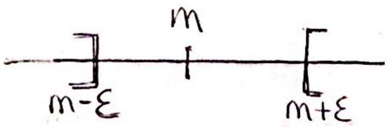
n nombre d'observations sera grand (il va tendre vers l'infini).

- **Loi (faible) des grands nombres (1er thm).**

$(X_n)_{n \geq 1}$ suite de variables aléatoires de même loi, intégrable ($E(|X_1|) < +\infty$) de carré intégrables ($V(X_n) = \sigma^2$, $E(X_1) = m$).

Soit $\varepsilon > 0$, $P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - m\right| \geq \varepsilon\right) \xrightarrow[n \to +\infty]{} 0$

Convergence en probabilité.



Clef de la preuve : Bienaymé-Tchebychev.

Soit Y une v.a d'espérance m et de variance σ^2. Alors pour tout $\varepsilon > 0$

$$P(|Y - m| \geq \varepsilon) \leq \frac{V(Y)}{\varepsilon^2}.$$

BC : Bienaymé-Tchebychev
LGN : Loi des grands nombres

BC $\Rightarrow$ LGN.

$Y = \overline{X_n}$. $E(Y) = E\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{E(X_1) + \dots + E(X_n)}{n} = \frac{n \times m}{n} = m$.

$Var(Y) = Var\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2}\left(Var(X_1) + \dots + Var(X_n)\right)$

car elles sont indépendantes.

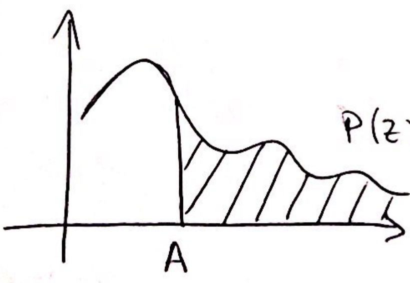
$Var(Y) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$. donc $Var(\overline{X_n}) = \frac{\sigma^2}{n}$.

Soit $\varepsilon > 0$. $P(|\overline{X_n} - m| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n \to +\infty]{} 0$ $\square$

Inégalité de Markov.

Soit Z une v.a positive. Soit r un réel > 0. Soit $A \geq 0$.

$$P(Z \geq A) \leq \frac{E(Z^r)}{A^r}$$



$P(Z \geq A)$.

$E(Z^r) = E(Z^r \text{ sur l'ens où } Z < A) + E(Z^r \text{ sur l'ens où } Z \geq A)$

$= E(Z^r \times \mathbb{1}_{Z<A}) + E(Z^r \times \mathbb{1}_{Z \geq A})$.

$E(Z^r) \geq 0 + E(A^r \times \mathbb{1}_{Z \geq A})$

$E(Z^r) \geq A^r \, E(\mathbb{1}_{Z \geq A})$

$E(Z^r) \geq A^r \, \mathbb{P}(Z \geq A) \Rightarrow \boxed{\frac{E(Z^r)}{A^r} \geq \mathbb{P}(Z \geq A).}$ $\square$

Markov $\Rightarrow$ Bienaymé-Tchebychev.

$Z = |Y-m|$, $A = \varepsilon$ $r=2$.

- Théorème central limite (TLC) (2^e théorème).

Thm de la limite centrée (TLC).

$\overline{X}_n - m$. $E(\overline{X}_n - m) = 0$

$Var(\overline{X}_n - m) = Var(\overline{X}_n) = \frac{\sigma^2}{n}$

On s'intéresse à $\frac{\sqrt{n}}{\sigma}(\overline{X}_n - m)$. On a $\frac{\sqrt{n}(\overline{X}_n - m)}{\sigma} \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1)$

Pour tout $t \in \mathbb{R}$, $P\left(\frac{\sqrt{n}(\overline{X}_n-m)}{\sigma} \leq t\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx = \Phi(t)$.

$$P\left(a \leq \frac{\sqrt{n}(\overline{X}_n-m)}{\sigma} \leq b\right) \xrightarrow[n\to+\infty]{} \Phi(b) - \Phi(a).$$

Applications: $(E(e^{itx}))$

- Bernouilli: $\theta \in]0;1[$. $E_\theta(X_1) = \theta$ $Var_\theta(X_1) = \theta \times (1-\theta)$.

$$\frac{\sqrt{n}(\overline{X}_n - \theta)}{\sqrt{\theta(1-\theta)}} \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0;1).$$

- Poisson $\lambda > 0$: $E_\lambda(X_1) = \lambda$, $Var_\lambda(X_1) = \lambda$

$$\frac{\sqrt{n}(\overline{X}_n - \lambda)}{\sqrt{\lambda}} \xrightarrow[n\to+\infty]{} \mathcal{N}(0,1).$$

- Loi exponentielle $\mathcal{E}(\lambda)$, $\lambda > 0$ $E_\lambda(X_1) = 1/\lambda$, $Var_\lambda(X_1) = 1/\lambda^2$.

$$\frac{\sqrt{n}(\overline{X}_n - \frac{1}{\lambda})}{\frac{1}{\lambda}} \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1). \quad \sqrt{n}(\lambda\overline{X}_n - 1) \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1)$$

TCL amélioré (Slutzky).

$\frac{\sqrt{n}(\overline{X}_n - m)}{\sigma} \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1)$. Soit S_n un observateur de σ qui converge vers σ en proba. $\frac{\sqrt{n}(\overline{X}_n - m)}{S_n} \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1)$ (Intervalle de confiance).

- <u>Bernoulli</u> : $\overline{X}_n \xrightarrow[n\to+\infty]{} \theta$

$$\sqrt{\overline{X}_n \times (1-\overline{X}_n)} \xrightarrow[n\to+\infty]{} \sqrt{\theta \times (1-\theta)}.$$

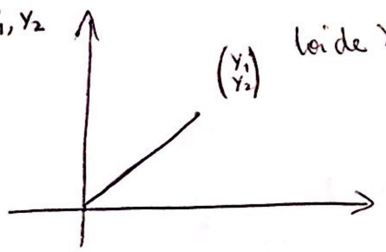
$$\frac{\sqrt{n}(\overline{X}_n - \theta)}{\sqrt{\overline{X}_n \times (1-\overline{X}_n)}} \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1).$$

- <u>Poisson</u> : $\frac{\sqrt{n}(\overline{X}_n - \lambda)}{\sqrt{\overline{X}_n}} \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1)$.

- <u>loi gaussienne</u> : $X_1, \ldots X_n$ n échantillon $\mathcal{N}(0,1)$.

$n=2$

Y_1, Y_2



loi de $Y_1^2 + Y_2^2$?

On s'intéresse à $Y_1^2 + Y_2^2 + \ldots + Y_n^2$

<u>Définition</u>,
La loi de $Z = Y_1^2 + Y_2^2 + \ldots + Y_n^2$ est appelée loi de χ^2 à n degré de liberté.

$X_1, \ldots, X_n$ et $\mathcal{N}(0,1)$ indépendante.

$$Z \equiv X_1^2 + X_2^2 + \ldots + X_d^2 = \|X\|^2 \text{ avec } X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

$\boxed{\text{loi de } Z \sim \chi^2(d)}$

Définition : soit $U = \begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix}$. U est un vecteur gaussien si et seulement si toute combinaison linéaire de ses coordonnées est une gaussienne.

Convention : toute constante se déterministe est une gaussienne (dégénéré). $\mathcal{N}(x,0)$.

Proposition : toute combinaison linéaire de gaussiennes indépendantes est une gaussienne.

Application : X est un vecteur gaussien. Soient $a_1, \dots, a_d \in \mathbb{R}$.

$Y = a_1 X_1 + a_2 X_2 + \dots + a_d X_d$. $Y \sim \mathcal{N}(E(Y), V(Y))$ avec

$$E(Y) = a_1 \times 0 + \dots + a_d \times 0 = 0$$

$$V(Y) = V(a_1 X_1) + V(a_2 X_2) + \dots + V(a_d X_d) =$$

$$= a_1^2 V(X_1) + a_2^2 V(X_2) + \dots + a_d^2 V(X_d) = \|\vec{a}\|^2.$$

$\vec{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_d \end{bmatrix}$ $\quad \vec{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix}$ $\quad \vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ $\rightsquigarrow V = v_1 \times X_1 + v_2 \times X_2 + \dots + v_d X_d$

$\rightsquigarrow \mathcal{N}(0, \|\vec{v}\|^2)$

$U = u_1 X_1 + u_2 X_2 + \dots + u_d X_d$

$$Cov(U, V) = E\big((U - E(U)) \times (V - E(V))\big)$$

$$= E(UV) - E(U) \times E(V) = E(UV).$$

$$E(UV) = E\left[(u_1 X_1 + \dots + u_d X_d) \times (v_1 X_1 + \dots + v_d X_d)\right]$$

$$= E\left[u_1 v_1 X_1^2 + \dots + u_d v_d X_d^2 + \sum u_i v_j X_i X_j\right]$$

$$= u_1 v_1 E(X_1^2) + \dots + u_d v_d E(X_d^2) + \sum u_i v_j E(X_i X_j)$$

$$= u_1 v_1 + u_2 v_2 + \dots + u_d v_d = \langle \vec{u}, \vec{v} \rangle = \vec{u} \cdot \vec{v} = \vec{u} * \vec{v}.$$

$E(X_i^2) = 1 = V(X_1) = E(X_1^2) - (E(X_1))^2$

$i \neq j$ $E(X_i, X_j) = 0$.

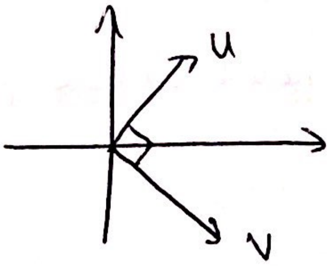
Proposition : U et V sont indépendantes $\Leftrightarrow Cov(U, V) = 0$.

U et V $\Leftrightarrow \langle \vec{u}, \vec{v} \rangle = 0$, $\vec{u} \perp \vec{v}$.

<u>Ex :</u> $d = 2$ $\quad X_1, X_2 \sim \mathcal{N}(0, 1)$ indépendantes.

$U = X_1 + X_2$

$V = X_1 - X_2$



U et V sont indépendantes.

$\bar{X} = \dfrac{X_1 + X_2}{2} \sim \mathcal{N}(0, 1/2)$

$\hat{\sigma}^2$ et $\bar{X}$ sont indépendantes.

$X_1 - \bar{X} = X_1 - \left(\dfrac{X_1 + X_2}{2}\right) = \dfrac{X_1 - X_2}{2}$

$X_2 - \bar{X} = \dfrac{X_2 - X_1}{2}$

$$\hat{\sigma}^2 = \frac{2(X_1 - X_2)^2}{2\times 4} = \left(\frac{X_1 - X_2}{2}\right)^2 . \quad X_1 - X_2 \underset{\text{indépendance}}{\perp\!\!\!\perp} X_1 + X_2 .$$

Autre démonstration :

$$\bar{X} \perp\!\!\!\perp X_1 - \bar{X} . \quad Cov(\bar{X}, X_1 - \bar{X}) = Cov(\bar{X}, X_1) - Cov(\bar{X}, \bar{X})$$

$$= Cov\left(\frac{X_1 + X_2}{2}, X_1\right) - V(\bar{X}).$$

$$= \frac{V(X_1)}{2} - \frac{V(X_1)}{2} = 0.$$

loi de $\hat{\sigma}^2$

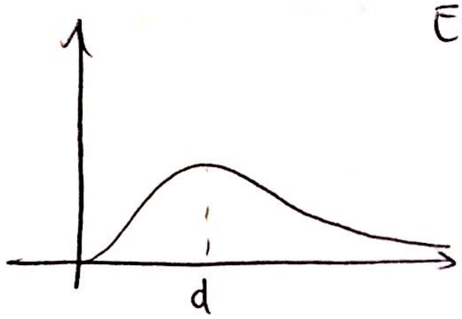
loi de $(X_1 - \bar{X}) + (X_2 - \bar{X})^2$

$$= \frac{(X_1 - X_2)^2}{2} = (\mathcal{N}(0,1))^2$$

$X_1 - X_2 \sim \mathcal{N}(0,2)$

$\frac{X_1 - X_2}{\sqrt{2}} \sim \mathcal{N}(0,1).$

Soit $Z \sim \chi^2(d)$ $d \geq 1$



$$E(Z) = d \times 1 = d$$

$$V(Z) = V(X_1^2 + X_2^2 + \dots + X_d^2)$$

$$= V(X_1^2) + V(X_2^2) + \dots + V(X_d^2) \quad \text{(les } X_i \text{ sont indép)}$$

$$= d \times V(X_1^2)$$

$$= 2d$$

$$V(X_i^2) = E(X_1^4) - (E(X_1^2))^2 = 3 - 1 = 2.$$

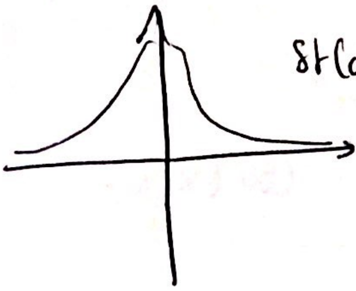
$$E(X_1^4) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^3 x e^{-\frac{x^2}{2}} dx \qquad u' = x e^{-\frac{x^2}{2}} \quad v = x^3 \qquad u = -e^{-x^2/2} \quad v' = 3x^2.$$

$$E(X_1^4) = \frac{1}{\sqrt{2\pi}} \left(\left[-x^3 e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} 3x^2 e^{-\frac{x^2}{2}} dx \right) = 0 \times \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2} dx = 3\,E(X_1^2) = 3$$

Définition : Soit $Y \sim \mathcal{N}(0,1)$, soit $Z \sim \chi^2(d)$. On suppose que Y et Z sont indépendantes. Alors $T = \frac{Y}{\sqrt{\frac{Z}{d}}}$ à une loi appelée la loi de student à (W. Gausset). d degré de liberté.

Exemple : X_1, X_2, $d=1$, $\mathcal{N}(0,1)$ indép.

$$\frac{\sqrt{2}\sqrt{\bar{X}}}{\sqrt{\frac{(X_1-\bar{X})^2+(X_2-\bar{X})^2}{1}}} \sim St(1).$$



$$St(d) \xrightarrow[d\to+\infty]{loi} \mathcal{N}(0,1)$$

$$\bar{X} \sim \mathcal{N}\left(0, \frac{1}{2}\right)$$

$$\frac{Z}{d} = \frac{X_1^2+X_2^2+\dots+X_d^2}{d}$$

$$\frac{X_1^2+X_2^2+\dots+X_d^2}{d} \xrightarrow[d\to+\infty]{} E(X^2)=1.$$