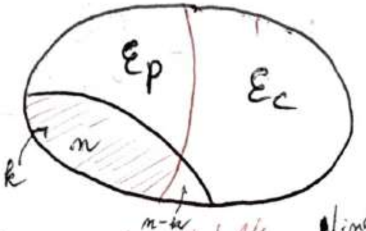
TD - M66 - Statistique

1334.

(pr l'instant plutôt Probas discrètes)

Ex1 **Sondage**



1) Loi de $X_{n,N}$?

- Sondage de n individus parmi N } espace de probabilité
  $\Omega = \{(\omega_1, \dots, \omega_n) : \omega_i \in \{0,1\}, 1 \leq i \leq n\}$
- $X_{n,N}$ va discrète représentant le nbr de "pour" parmi les n.
  - $\mathcal{P}(\Omega)$
  - $\mathbb{P}(\{\omega_1, \dots, \omega_n\})$???

(i)

$\Omega = \{A \subset \{1, \dots, n\},\ \mathrm{card}(A) = n\} \to$ ens des parties de $\{1, \dots, N\}$ à n élts. ens fini.

$\mathrm{card}(\Omega) = C_N^n$ | $\mathcal{F} = \mathcal{P}(\Omega)$ | $\mathbb{P}$ équiproba

$\to$ Tribu : $\mathcal{P}(\Omega)$; muni $(\Omega, \mathcal{P}(\Omega))$ de l'équiprobabilité.

ou bien (ii) $\Omega = \{\underbrace{(\omega_1, \dots, \omega_n)}_{A} \in \{1, \dots, N\}^n,\ \omega_1 < \dots < \omega_n\}$

(l'ordre ds lequel les personnes ont été interrogées ne compte pas). $\mathcal{E} = \{1, \dots, N\} = \mathcal{E}_p \cup \mathcal{E}_c$.

NB Si on voulait prendre en compte l'ordre ds lequel on interroge les individus :

$\Omega = \{(\omega_1, \dots, \omega_n) \in \{1, \dots, N\}^n,\ \omega_i \neq \omega_j \text{ pr } i \neq j\}$

$\mathrm{card}(\Omega) = A_N^n = \dfrac{N!}{(N-n)!} = N(N-1)\dots(N-n-1)$ (1) 12h45

(i) $X_{n,N} : \Omega \longrightarrow \mathbb{R}$

$A \longmapsto \mathrm{card}(A \cap \mathcal{E}_p)$

N inder $X_{n,N}(A) = \mathrm{card}(A \cap \mathcal{E}_p)$

(ii) $\mathcal{E}_p = \{1, \dots, d_N\}$

si $(\omega_1, \dots, \omega_n) \in \Omega$

$X_{n,N}(\omega) = \mathrm{card}\{i \in \{1, \dots, n\},\ \omega_i \leq d_N\}$

(i) $\Rightarrow X_{n,N}$ va discrète à valeurs ds $\{0, \dots, n\}$.

$\mathbb{P}(X_{n,N} = k) = \mathbb{P}(\{A \in \Omega,\ X_{n,N}(A) = k\})$

$$\boxed{\mathbb{P}(X_{n,N} = k) = \frac{C_{d_N}^k \times C_{N-d_N}^{n-k}}{C_N^n}}$$

loi hypergéométriq $\mathcal{H}(n, N, d_N)$

2) On sppse $\frac{d_N}{N} = p + \varepsilon(N)$, $\varepsilon(N) \xrightarrow[N\to\infty]{} 0$, $p \in \mathbb{R}$, $0 < p < 1$

$\lim_{N\to\infty} \mathbb{P}(X_{n,N} = k)$ & k fixé. Supposément ! $0 \leq k \leq n$

$$\mathbb{P}(X_{n,N} = k) = \frac{\frac{d_N!}{k!\,(d_N-k)!} \times \frac{(N-d_N)!}{(n-k)!\,(N-d_N-n+k)!}}{\frac{N!}{n!\,(N-n)!}}$$

$$= \frac{n!}{k!\,(n-k)!} \, \frac{\frac{d_N!}{(d_N-k)!} \times \frac{(N-d_N)!}{(N-d_N-n+k)!}}{\frac{N!}{(N-n)!}}$$

écrire à le produit de k termes

écrire à le produit de n-k termes

écrire à le produit de n termes

$$= C_n^k \cdot \frac{\overbrace{d_N(d_N-1)\cdots(d_N-k+1)}^{k \text{ termes}} \cdot \overbrace{N(N-1)\cdots(N-d_N+1)}^{(n-k) \text{ termes}}}{\underbrace{N\cdot(N-1)\cdots(N-n+1)}_{n \text{ termes}}}$$

$$= C_n^k \cdot \frac{\prod_{i=0}^{k-1}(d_N - i) \cdot \prod_{i=0}^{n-k-1}(N-d_N-i)}{\prod_{i=0}^{n-1}(N-i)}$$

$$= C_n^k \frac{\overbrace{\prod_{i=0}^{k-1}\left(\frac{d_N}{N} - \frac{i}{N}\right)}^{\to p^k} \times \overbrace{\prod_{i=0}^{n-k-1}\left(1 - \frac{d_N}{N} - \frac{i}{N}\right)}^{\xrightarrow[N\to\infty]{} (1-p)^{n-k}}}{\underbrace{\prod_{i=0}^{n-1}\left(1 - \frac{i}{N}\right)}_{\xrightarrow[N\to\infty]{} 1}}$$

le nbr de termes ne bouge pas de N.

$$\Rightarrow \lim_{N\to\infty} \mathbb{P}(X_{n,N} = k) = C_n^k \, p^k \, (1-p)^{n-k}$$
$$= \mathbb{P}(X_n = k) \text{ si } X_n \sim \mathcal{B}in(n,p)$$

$\boxed{\text{Moralité}}$ Un tirage sans remise $(X_{n,N})$ ds une urne "infinie" équivaut à un tirage à remise (X_n).

$0 \leq k \leq n$

Ex2 Inégalité de Tchebitchev

$(X_n)_{n\geq 1}$ suite de va indép. de même loi de (identiqt distribué : i.i.d.) Bern(θ) où $\theta \in]0,1[$. On pose $S_n = \sum_{i=1}^{n} X_i$ & $\tilde{X}_n = \frac{S_n}{n}$ (moy. empiriq)

1) Quelle est la loi de S_n ? Espérance ? Variance ?

$$S_n \sim \mathcal{B}in(n,\theta) \; ; \; \mathbb{P}(S_n = k) = C_n^k \, \theta^k (1-\theta)^{n-k}$$

$$\Omega = \{0,1\}^n$$
$$\mathcal{F} = \mathcal{P}(\Omega)$$

si $(\omega_1, \dots, \omega_n) \in \Omega$:

$$\mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = \theta^{\sum_{i=1}^{n} \omega_i} (1-\theta)^{n - \sum_{i=1}^{n} \omega_i}$$

$$S_n(\omega) = \sum_{i=1}^{n} \omega_i = \text{card}(\{i \in \{1,\dots,n\}, \omega_i = 1\})$$

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n\theta$$

$$Var(S_n) = Var\left(\sum_{i=1}^{n} X_i\right) \underset{\text{par indépendance}}{=} \sum_{i=1}^{n} Var(X_i) = n\theta(1-\theta)$$

car $Var(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2$

$= \mathbb{E}(X_i) - \mathbb{E}(X_i)^2 = \theta - \theta^2 = \theta(1-\theta)$

③

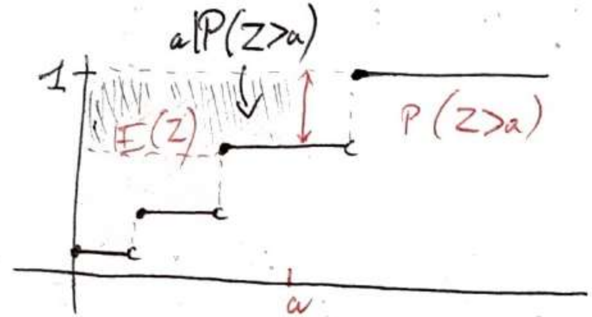
Jan 1

2) Soit $\varepsilon > 0$, majorer proba $\mathbb{P}(|\tilde{X}_n - \theta| \geq \varepsilon)$ en f de n, θ et ε.

indic : inég. Markov

$$\mathbb{P}(|\tilde{X}_n - \theta| \geq \varepsilon) \leq \frac{\mathbb{E}(|\tilde{X}_n - \theta|)}{\varepsilon}$$

pas simple : $\mathbb{E}(|\tilde{X}_n - \theta|) = \sum_{k=0}^{n} \left|\frac{k}{n} - \theta\right| \theta^k (1-\theta)^{n-k}$



Rp Markov $\mathbb{P}(Z \geq a) \leq \frac{\mathbb{E}(Z)}{a}$

- Z va ≥ 0
- $a > 0$

$$Var(\bar{X}_n) = Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2} Var(S_n) = \frac{\theta(1-\theta)}{n}$$

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{S_n}{n}\right) = \theta$$

Rq : $|\tilde{X}_n - \theta| \leq \tilde{X}_n - \theta \leq 1$

$\Rightarrow \mathbb{E}(|\tilde{X}_n - \theta|) \leq \mathbb{E}(\tilde{X}_n) + \theta = 2\theta$

et $\mathbb{E}(|\tilde{X}_n - \theta|) \leq 1$.

$\mathbb{E}(|\tilde{X}_n - \theta|) \leq \min(1, 2\theta)$

$\mathbb{P}(|\bar{X}_n - \theta| \geq \varepsilon) \leq \frac{\mathbb{E}(|\bar{X}_n - \theta|)}{\varepsilon} \leq \frac{\min(1, 2\theta)}{\varepsilon}$

pas intéressant : on perd le n.

Brillon

$\frac{\mathbb{E}(|\bar{X}_n - \theta|)}{\varepsilon} = \frac{1}{\varepsilon} \sum_{k=0}^{n} C_n^k \left|\frac{k}{n} - \theta\right| \theta^k (1-\theta)^{n-k}$

$\leq \frac{1}{\varepsilon} \sum_{k=0}^{n} C_n^k (\theta + 1) \theta^k (1-\theta)^{n-k}$

$\leq \frac{\theta + 1}{\varepsilon} 1^n$. On applique Markov à $\varepsilon = n$

$\leq \frac{\theta + 1}{n}$, dc en passant à la limite :

$\lim_{n \to \infty} \frac{\theta + 1}{n} = 0$

$\underbrace{\mathbb{P}(|\bar{X}_n - \theta|^2 \geq \varepsilon^2)}_{\text{va} \geq 0} \overset{\text{Markov}}{\leq} \frac{\mathbb{E}((\bar{X}_n - \theta)^2)}{\varepsilon^2}$

$= \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\theta(1-\theta)}{n \varepsilon^2}$

Rq : $\mathbb{P}(|\bar{X}_n - \theta| \geq \varepsilon) = \mathbb{P}(|S_n - n\theta| \geq n\varepsilon)$

à θ fixé, $\lim_{n \to \infty} \frac{\theta(1-\theta)}{n\varepsilon^2} = 0$

$\lim_{n \to \infty} \mathbb{P}(|\bar{X}_n - \theta| \geq \varepsilon) = 0$

Rq $\mathrm{Var}(Z) = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = \mathbb{E}[(Z - \mathbb{E}(Z))^2]$ ≥ 0

$\bar{X}_n \xrightarrow[n \to \infty]{\text{Proba}} \theta$

Ex 3 Amélioration de l'inégalité de Tchebitchev :

But $\mathbb{P}\left(\left|\frac{S_n}{n} - \theta\right| \geq \varepsilon\right) \leq \frac{1}{4n^2\varepsilon^4}$

$\frac{S_n}{n} = \bar{X}_n$

Soit $(X_n)_{n\geq 1}$ suite de va idd m̂ loi $\mathrm{Bern}(\theta)$, $\theta \in]0,1[$.

$S_n = \sum_{i=1}^{n} X_i$, pour $Y_i = X_i - \theta$ $\forall i \in \mathbb{N}^*$.

1) en usant identité $\left(\sum_{i=1}^{n} Y_i\right)^4 = \sum_{i,j,k,\ell=1} Y_i Y_j Y_k Y_\ell$

Mq $\mathbb{E}\left(\left(\sum_{i=1}^{n} Y_i\right)^4\right) = n\,\mathbb{E}(Y_i^4) + 3n(n-1)\left(\mathbb{E}(Y_i^2)\right)^2$

$(i,j,k,\ell) \in \{1,\dots,n\}^4$,

$\mathbb{E}(Y_i Y_j Y_k Y_\ell)$? $\quad Y_i = X_i - \theta$, $X_i \sim \mathrm{Bern}(\theta)$

- $\mathbb{E}(Y_i) = 0$ (car $\mathbb{E}(X_i) = \mathbb{E}(\theta)$)
- $Y_1, \dots, Y_n$ m̂ loi + indépendantes. si (i,j,k,ℓ) st tous distincts.

$$\mathbb{E}(Y_i Y_j Y_k Y_\ell) = \mathbb{E}(Y_i)\,\mathbb{E}(Y_j)\,\mathbb{E}(Y_k)\,\mathbb{E}(Y_\ell)$$

(R1) Si X est indépendante de Z, $g(X)$ est indépendante de $h(Z)$. (pour g & h mesurables)

$(i,j,k,\ell) \in \{1,\dots,n\}^4$ plusieurs cas :

- i,j,k,ℓ tous distincts : $\mathbb{E}(Y_i Y_j Y_k Y_\ell) = 0$.
- i,j,k,ℓ tous égaux $i=j=k=\ell$

$\mathbb{E}(Y_i Y_j Y_k Y_\ell) = \mathbb{E}(Y_i^4) = \mathbb{E}(Y_1^4)$ car Y_i a m̂ loi que Y_1

il y a n 4-uplets (i,j,k,ℓ) de ce type

- 3 égaux & 1 $\neq$: ex $i=j=k$ & $\ell \neq i$

$\mathbb{E}(Y_i Y_j Y_k Y_\ell) = \mathbb{E}(Y_i^3 Y_\ell) = \mathbb{E}(Y_i^3)\,\underbrace{\mathbb{E}(Y_\ell)}_{=0} = 0$

- 2 égaux : – 2 restants dist$_1$ – 2 restants égx $\Rightarrow \mathbb{E}(Y_i Y_j Y_k Y_\ell) = 0$.
- 2 égx + 2 égx : ex $i=j$ et $k=\ell$ & $i \neq k$.

$\mathbb{E}(Y_i Y_j Y_k Y_\ell) = \mathbb{E}(Y_i^2 Y_k^2) = \mathbb{E}(Y_i^2)\,\mathbb{E}(Y_k^2)$

$= \left[\mathbb{E}(Y_1^2)\right]^2$ car $\mathbb{E}(Y_i^2) = \mathbb{E}(Y_k^2) = \mathbb{E}(Y_1^2)$

→ dénombrer le nb de (i,j,k,ℓ) ⇔ 2 indices égx & les 2 restants égaux (mais distincts des 2 autres).

ex $(1,1,2,2)$; $(1,2,1,2)$; $(1,2,2,1)$; $(2,2,1,1)$; $(2,1,2,1)$, $(2,1,1,2)$.

↳ 6 cas une fois les 2 indices fixés.

$$\Rightarrow 6 \times C_n^2 = 6 \times \frac{n(n-1)}{2} = 3n(n-1)$$

$x : n$ $\quad (x,x,\circ,\circ)$

$\circ : n-1$ $\quad (x,\circ,x,\circ)$ $\Rightarrow 3n(n-1)$

$(x,\circ,\circ,x)$

$$\Rightarrow \mathbb{E}\left(\left(\sum_{i=1}^{n} Y_i\right)^4\right) = n\,\mathbb{E}(Y_1^4) + 3n(n-1)\left[\mathbb{E}(Y_1^2)\right]^2$$

et $\mathbb{E}\left[(S_n - n\theta)^4\right] = n\theta(1-\theta)\left[3n\theta(1-\theta) + 1 - 6\theta + 6\theta^2\right]$.

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} (X_i - \theta) = S_n - n\theta.$$

$$\mathbb{E}(Y_1^4) = \mathbb{E}\left((X_1-\theta)^4\right)$$ ("moment centré d'ordre 4 de la loi de Bernoulli")

$$\mathbb{E}(Y_1^2) = \mathbb{E}\left((X_1-\theta)^2\right) = Var(X_1) = n\theta(1-\theta)$$

$$\mathbb{E}\left((X_1-\theta)^4\right) = (1-\theta)^4\theta + \theta^4(1-\theta)$$
$$= \theta(1-\theta)\left[(1-\theta)^3 + \theta^3\right]$$
$$= \theta(1-\theta)\left[1 + 3\theta^2 - 3\theta - \theta^3 + \theta^3\right]$$
$$= \theta(1-\theta)(1 - 3\theta + 3\theta^2)$$

NB : $Z_1 = (X_1-\theta)^4 = \begin{cases} (1-\theta)^4 & \text{à proba } \theta \\ \theta^4 & \text{à proba } 1-\theta \end{cases}$

$$\mathbb{E}(g(X)) = \sum_{x_k \in X(\Omega)} g(x_k)\,\mathbb{P}(X = x_k)$$

$$\Rightarrow \mathbb{E}(Y_1^4) = \mathbb{E}\left((X_1-\theta)^4\right) = \theta(1-\theta)(1 - 3\theta + 3\theta^2)$$

$$\mathbb{E}\left[(S_n - n\theta)^4\right] = n\theta(1-\theta)(1 - 3\theta + 3\theta^2) + 3n(n-1)\theta^2(1-\theta)^2.$$
$$= n\theta(1-\theta)\left[3n\theta(1-\theta) + 1 - 6\theta + 6\theta^2\right]$$

2) a) $\forall n \in \mathbb{N}^*, \forall \varepsilon > 0,\ \mathbb{P}\left(\left|\frac{S_n}{n} - \theta\right| \geq \varepsilon\right) \leq \frac{1}{4n^2\varepsilon^4}$

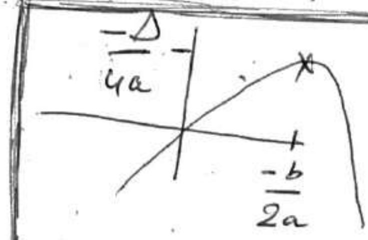
$$\mathbb{P}\left(\left|\frac{S_n}{n} - \theta\right| \geq \varepsilon\right) = \mathbb{P}\left(|S_n - n\theta| \geq n\varepsilon\right) = \mathbb{P}\left((S_n - n\theta)^4 \geq n^4\varepsilon^4\right)$$

par Markov $\leq \dfrac{\mathbb{E}\left((S_n - n\theta)^4\right)}{n^4\varepsilon^4} = \dfrac{n(\theta(1-\theta)\left[3n\theta(1-\theta) + 1 - 6\theta + 6\theta^2\right]}{n^4\varepsilon^4}$

$$\frac{\frac{n}{4}\,\theta^2(6-3n) + \theta(3n-6) + 1}{n^4\varepsilon^4}$$

$$\max\left(\theta^2(6-3n) + \theta(3n-6) + 1\right) = \frac{3}{4}n - \frac{1}{2}$$

$$\leq \frac{3}{16}\,\frac{n}{n^2\varepsilon^4} \leq \frac{1}{4n^2\varepsilon^4}$$



Sur $[0,1]$, $\theta(1-\theta) \leq \frac{1}{4}$

En quoi cela améliore Tchebychev ? Vitesse en $\frac{1}{n^2}$ plutôt que en $\frac{1}{n}$ à Tchebychev.

(R9) $\forall \varepsilon > 0,\ \mathbb{P}\left(\left|\frac{S_n}{n} - \theta\right| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}$

$\forall \delta > 0,\ \exists n_0 \in \mathbb{N}^*,\ \forall n \geq n_0,$

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \theta\right| \geq \varepsilon\right) \leq \delta \iff \mathbb{P}\left(\left|\frac{S_n}{n} - \theta\right| \leq \varepsilon\right) \geq 1-\delta$$

On veut exprimer $\mathbb{P}\left(\forall n \geq n_0,\ \left|\frac{S_n}{n} - \theta\right| < \varepsilon\right) = \mathbb{P}(B_{n_0})$

en f de $A_n(\varepsilon) = \left\{\left|\frac{S_n}{n} - \theta\right| \geq \varepsilon\right\}$

& $B_{n_0}(\varepsilon) = \left\{\omega \in \Omega,\ \forall n \geq n_0,\ \left|\frac{S_n(\omega)}{n} - \theta\right| < \varepsilon\right\}$

et $\mathbb{P}(A_n(\varepsilon)) \leq \frac{1}{4n^2\varepsilon^4}$.

$\omega \in B_{n_0}(\varepsilon)$ ssi $\forall n \geq n_0$ $\omega \in A_n^c(\varepsilon)$.

$$B_{n_0}(\varepsilon) = \bigcap_{n \geq n_0} A_n^c(\varepsilon)$$

Comment gérer intersect° events indep ?

(R9) $\bigcap_{n \in \mathbb{N}^*} A_n$

$$C_N = \bigcap_{n=1}^{N} A_n, \quad \bigcap_{n \in \mathbb{N}^*} A_n = \overbrace{\bigcap_{N \in \mathbb{N}^*} C_N}^{\text{intersect° décroissante}}$$

$$\mathbb{P}\left(\bigcap_{n \in \mathbb{N}^*} A_n\right) = \mathbb{P}\left(\bigcap_{n \in \mathbb{N}^*} C_N\right) = \lim_{N\to\infty} \mathbb{P}(C_N)$$

$$= \lim_{N\to\infty} \prod_{n=1}^{N} \mathbb{P}(A_n)$$

car les $(A_n)_{n \in \mathbb{N}^*}$ st indépendants.

ici les $(A_n^c(\varepsilon))_{n \in \mathbb{N}^*}$ ne st pas indep.

On passe au complémentaire.

$$B_{n_0}^c(\varepsilon) = \bigcup_{n \geq n_0} A_n(\varepsilon)$$

$$\mathbb{P}(B_{n_0}^c(\varepsilon)) \leq \sum_{n \geq n_0} \mathbb{P}(A_n(\varepsilon)) \leq \sum_{n \geq n_0} \frac{1}{4n^2\varepsilon^4}$$

→ Trouver $n_0 \in \mathbb{N}^*$ (en f de ε) tq $\sum_{n \geq n_0} \frac{1}{4n^2\varepsilon^4} \leq 0{,}01$.

Indic : $\sum_{n=N+1}^{\infty} \frac{1}{n^2} \leq \frac{1}{N}$.

$$\sum_{n=n_0-1}^{\infty} \frac{1}{4n^2\varepsilon^4} \leq \frac{1}{4\varepsilon^4} \times \frac{1}{n_0-1}.$$

Il suffit de prendre n_0 tq $\frac{1}{4\varepsilon^4(n_0-1)} \leq 0{,}01$.

$\iff n_0 \geq 1 + \frac{100}{4\varepsilon^4}$.

Mq l'indication : $\sum_{n=N+1}^{\infty} \frac{1}{n^2} \leq \frac{1}{N}$

On a que $\forall n > 1$, $\forall x \in [n-1, n]$

$\frac{1}{n^2} \leq \frac{1}{x^2}$, $\frac{1}{n^2} = \int_{n-1}^{n} \frac{1}{n^2} dx \leq \int_{n-1}^{n} \frac{1}{x^2} dx$

$\Rightarrow \sum_{n=N+1}^{\infty} \frac{1}{n^2} \leq \sum_{n=N+1}^{\infty} \int_{n-1}^{n} \frac{1}{x^2} dx \underset{\text{Chales}}{=} \int_N^{\infty} \frac{1}{x^2} dx = \left[\frac{-1}{x}\right]_N^{\infty} = \frac{1}{N}$

Ex4 Inégalité de Hoeffding

1) ici $X \sim \mathcal{B}er(\theta)$, $\theta \in]0,1[$.

Mq $\forall s > 0$, $\mathbb{E}\left(e^{s(X-\theta)}\right) \leq e^{s^2/8}$

- calculer $\mathbb{E}\left(e^{s(X-\theta)}\right)$
- poser $g_\theta(s) = \ln\left(\mathbb{E}\left(e^{s(X-\theta)}\right)\right)$ & dériver 2 fois

si $X \sim \mathcal{B}er(\theta)$, $\mathbb{E}(\varphi(X)) = \varphi(1)\theta + \varphi(0)(1-\theta)$

$\mathbb{E}\left(e^{s(X-\theta)}\right) = \theta e^{s(1-\theta)} + (1-\theta)e^{-s\theta}$

$g_\theta(s) = \ln\left[\theta e^{s(1-\theta)} + (1-\theta)e^{-s\theta}\right]$

$\forall \theta \in]0,1[$, $g_\theta'(s) = -\theta + \frac{\theta e^s}{\theta e^s + 1 - \theta}$

$g_\theta''(s) = \frac{\theta(1-\theta)e^s}{(\theta e^s + 1 - \theta)^2} \overset{?}{\leq} \frac{1}{4}$

$g_\theta'(s) \leq \frac{s}{4}$, $g_\theta(s) \leq \frac{s^2}{8}$ $\quad \begin{pmatrix} g_\theta'(0) = 0 \\ g_\theta(0) = 0 \end{pmatrix}$

$g_\theta''(s)$ est de la forme $\frac{ab}{(a+b)^2}$. $a = \theta e^s$, $b = 1-\theta$

$4ab \leq (a+b)^2 \iff \frac{a^2 - 2ab + b^2}{(a+b)^2} \geq 0$

$\Rightarrow g_\theta''(s) \leq \frac{1}{4}$, $g_\theta'(0) = 0$

$\Rightarrow g_\theta'(s) \leq \frac{s}{4}$ & $g_\theta(0) = 0$

$\Rightarrow g_\theta(s) \leq \frac{s^2}{8}$

$g_\theta(s) = \ln\left(\mathbb{E}\left(e^{s(X-\theta)}\right)\right)$

$g_\theta(s) \leq \frac{s^2}{8} \iff \mathbb{E}\left(e^{s(X-\theta)}\right) \leq e^{s^2/8}$

$\forall \varepsilon > 0$

2) et $S_n \sim \mathcal{B}in(n,\theta)$ alors $\mathbb{P}\left(\frac{S_n}{n} - \theta \geq \varepsilon\right) \leq e^{-2n\varepsilon^2}$

et $\mathbb{P}\left(\left|\frac{S_n}{n} - \theta\right| \geq \varepsilon\right) \leq 2e^{-2n\varepsilon^2}$

indic: use $\forall$ va réelle Z & $\forall x \in \mathbb{R}$, $s > 0$:

$$\mathbb{P}(Z \geq x) \leq e^{-sx}\, \mathbb{E}(e^{sZ})$$

Pour $\mathbb{P}\left(\underbrace{\frac{S_n}{n} - \theta}_{\text{va pas positive}} \geq \varepsilon\right)$ mais $\mathbb{P}\left(\underbrace{e^{\frac{S_n}{n}-\theta}}_{\text{va. positive}} \geq e^{s\varepsilon}\right)$

$x \mapsto e^{sx}$ est croissante & bijective.

$\forall \varepsilon > 0$ et $S_n \sim \mathcal{B}in(n,\theta)$; $S_n \overset{\text{loi}}{=} \sum_{i=1}^n X_i$ & $X_i \sim \mathcal{B}ern(\theta)$

et les $(X_i)_i$ indép.

Par Markov :

$$\mathbb{P}\left(\frac{S_n}{n} - \theta \geq \varepsilon\right) \leq e^{-s\varepsilon}\, \mathbb{E}\left(e^{s\left(\frac{S_n}{n} - \theta\right)}\right)$$

or $e^{s\left(\frac{S_n}{n} - \theta\right)} = e^{\frac{s}{n}\sum_{i=1}^n (X_i - \theta)} = e^{\sum_{i=1}^n \frac{s}{n}(X_i - \theta)}$

$$\mathbb{E}\left(e^{s\left(\frac{S_n}{n} - \theta\right)}\right) = \mathbb{E}\left(e^{\sum_{i=1}^n \frac{s}{n}(X_i - \theta)}\right)$$

$$= \mathbb{E}\left(\underbrace{\prod_{i=1}^n e^{\frac{s}{n}(X_i - \theta)}}_{Y_i \text{ indépend.}}\right)$$

$$= \prod_{i=1}^n \mathbb{E}\left(e^{\frac{s}{n}(X_i - \theta)}\right) = \left[g_\theta\left(\frac{s}{n}\right)\right]^n$$

$$\leq \left[e^{\left(\frac{s}{n}\right)^2 \cdot \frac{1}{8}}\right]^n = e^{\frac{s^2}{8n}}$$

$\mathbb{P}\left(\frac{S_n}{n} - \theta \geq \varepsilon\right) \leq e^{-s\varepsilon}\, \mathbb{E}\left(e^{s\left(\frac{S_n}{n} - \theta\right)}\right) \leq e^{-s\varepsilon + \frac{s^2}{8n}}$ $\forall s > 0$ à optimiser en s

$$\Rightarrow \mathbb{P}\left(\frac{S_n}{n} - \theta \geq \varepsilon\right) \leq \inf_{s > 0} e^{-s\varepsilon + \frac{s^2}{8n}}$$

$$= e^{-4n\varepsilon \times \varepsilon + \frac{(4n\varepsilon)^2}{8n}} = e^{-2n\varepsilon^2}$$

$$\Rightarrow \boxed{\mathbb{P}\left(\frac{S_n}{n} - \theta \geq \varepsilon\right) \leq e^{-2n\varepsilon^2}}$$

On a $\mathbb{P}\left(\left|\frac{S_n}{n} - \theta\right| \geq \varepsilon\right) = \mathbb{P}\left(\frac{S_n}{n} - \theta \geq \varepsilon\right) + \mathbb{P}\left(\frac{S_n}{n} - \theta \leq -\varepsilon\right)$

$S_n \sim \mathcal{B}in(n,\theta)$ et $n - S_n \sim \mathcal{B}in(n, 1-\theta)$

et on applique l'inégalité précédente à $n - S_n = S_n' \sim \mathcal{B}in(n, 1-\theta)$

$$\left\{\frac{S_n'}{n} - (1-\theta) \geq \varepsilon\right\} = \left\{\frac{n - S_n}{n} - (1-\theta) \geq \varepsilon\right\} = \left\{-\frac{S_n}{n} + \theta \geq \varepsilon\right\}$$

$$= \left\{\frac{S_n}{n} - \theta \leq -\varepsilon\right\}$$

$$\mathbb{P}\left(\frac{S_n'}{n} - (1-\theta) \geq \varepsilon\right) \leq e^{-2n\varepsilon^2}$$

$$\Leftrightarrow \mathbb{P}\left(\frac{S_n}{n} - \theta \leq -\varepsilon\right) \leq e^{-2n\varepsilon^2}$$

cl : $\mathbb{P}\left(\left|\frac{S_n}{n} - \theta\right| \geq \varepsilon\right) \leq 2e^{-2n\varepsilon^2}$

Applicat° des inégalités de concentrat°

Ex5 Un central téléphoniq dessert 5000 abonnés. À un instant donné, chaq abonné a une proba à 2 % d'user son téléphone & les appels des abonnés st supposés indépendants. Q^e nbr minimal d'appels doit pouvoir traiter simultaném^t le central p^r que sa proba d'ê saturé à un instant donné soit inférieure à 2,5 % ?

$N = 5000$, chq individu a une proba $p = 0{,}02$ d'appeler. Si on voulait spécifier Ω :

$\omega = (\omega_1, \dots, \omega_{5000})$; $\omega_i = \begin{cases} 1 & \text{si la } i^{\circ} \text{ personne appelle} \\ 0 & \text{sinon} \end{cases}$

$\Omega = \{0,1\}^{5000}$.

$X_i = \begin{cases} 1 & \text{si la } i^{\circ} \text{ pers. appelle} \\ 0 & \text{sinon} \end{cases}$

Le nbr d'appels sur le central téléphoniq est la variable aléatoire $S_n = \sum_{i=1}^{n} X_i = \operatorname{card}\{i : X_i = 1\}$

$S_n \sim \mathrm{Bin}(n,p)$ car les $(X_i)_i$ st i.i.d. de loi $\mathrm{Bern}(p)$.

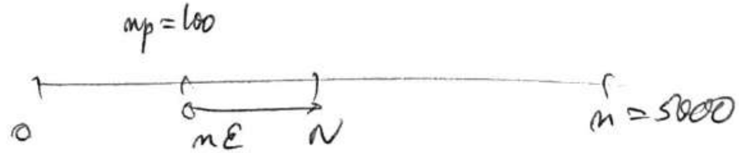
$(X_i(\omega) = \omega_i)$

On cherche N tq $\mathbb{P}(S_n \geq N) \leq 0{,}025$ minimal suffisant

Valeur exacte à py. $\sum_{k=N}^{n} C_n^k p^k (1-p)^{n-k}$

Rq : $\mathbb{P}(S_n - n\theta \geq n\varepsilon) \leq e^{-2n\varepsilon^2}$

$= \mathbb{P}(S_n \geq n\varepsilon + n\theta) = \sum_{k \geq n\varepsilon + n\theta}^{n} C_n^k \theta^k (1-\theta)^{n-k}$



$N = np + n\varepsilon$.

$\mathbb{P}(S_n \geq np + n\varepsilon) = \mathbb{P}\left(\frac{S_n}{n} - p \geq \varepsilon\right) \leq e^{-2n\varepsilon^2}$

Il suffit de prendre ε tq $e^{-2n\varepsilon^2} \leq 0{,}025$

$\varepsilon \geq \sqrt{-\frac{1}{2n}\ln(0{,}025)}$ $\quad n = 5000$.

$\varepsilon \simeq 0{,}019$.

$N = \lceil n(p+\varepsilon) \rceil$ $\quad p = 0{,}02$

$\boxed{N = 196}$ si $k \geq 196$: $\mathbb{P}(S_n \geq k) \leq 0{,}025$

(N n'est pas optimal)

(le vrai $\inf_{k \in \{0,\dots,n\}} \mathbb{P}(S_n \geq k) \leq 0{,}025 \Rightarrow N = 121$)

Ex6 : Surréservation aérienne :

Nbr réservations > nbr passagers effectifs

pr chq vol : $n > 300$ réservations.

$p \leq 300 \leftarrow$ vol

$\downarrow p+k$ } dédommagés

$k \in \mathbb{N}$.

1) Passagers mutuellement indépendants.

- $P(D) = 0,2$ où $D =$ "désistement". ($\theta = 0,8 \leftarrow$ présents / succès)
- n : nbr réservations prises par la compagnie pour un vol.
- S_n : nbr aléatoires de passagers se présentant à l'embarquement.

$\rightarrow S_n \sim \mathcal{B}in(n, p)$

$\rightarrow$ ed sans calcul : $\mathbb{E}(S_n)$ et $Var(S_n)$

$\mathbb{E}(S_n) = n \cdot P(D) = 0,2n$

$Var(S_n) = nP(D)(1-P(D)) = 0,2n(0,8) = 1,6n$.

2) Valeur max n ? $\mathbb{P}(S_n \leq 300) \geq 0,95$.

$$1 - \mathbb{P}(S_n \geq 300) \geq 0,95$$

$$1 - \mathbb{P}\left(\frac{S_n}{n} \geq \frac{300}{n}\right) \geq 0,95$$

$$1 - \mathbb{P}\left(\frac{S_n}{n} - \theta \geq \frac{300}{n} - \theta\right) \geq 0,95$$

$$1 - \mathbb{P}\left(\frac{S_n}{n} - \theta \geq \frac{300}{n} - 0,8\right) \geq 0,95$$

$$-\mathbb{P}\left(\frac{S_n}{n} - \theta \geq \frac{300}{n} - 0,8\right) \geq -0,05$$

$$\mathbb{P}\left(\frac{S_n}{n} - \theta \geq \frac{300}{n} - 0,8\right) \leq 0,05$$

Par l'inégalité de Hoeffding

$$\mathbb{P}\left(\frac{S_n}{n} - \theta \geq \varepsilon\right) \leq e^{-2n\varepsilon^2}$$

$$\mathbb{P}\left(\frac{S_n}{n} - \theta \geq \frac{300 - 0,8n}{n}\right) \leq \exp\left(-2n\left(\frac{300-0,8n}{n}\right)^2\right)$$

Il suffit de prendre ε tq $e^{-2n\varepsilon^2} \leq 0,05$

$\Leftrightarrow \varepsilon \geq \sqrt{-\frac{1}{2n}\ln(0,05)}$; $\varepsilon > 0$. $n = 300$

$\varepsilon \simeq 0,0707$.

$\mathbb{P}\left(\frac{S_n}{n} - \theta \geq \varepsilon\right) \Leftrightarrow \mathbb{P}(S_n \geq n\varepsilon + n\theta)$

$$\exp\left(-2n\left(\frac{300 - 0,8}{n}\right)^2\right) \leq 0,05.$$

$$\cancel{\left(\frac{300-0,8n}{n}\right)^2 \geq \sqrt{-\frac{1}{2n}\ln(0,05)}}$$

$\min(a,b) = \frac{(a+b) - |a-b|}{2}$

$1 - \mathbb{P}(S_n > 300)$

$(1)\ = 1 - \mathbb{P}(S_n \geq 300) + \underbrace{\mathbb{P}(S_n = 300)}_{\geq 0}$

On veut $\mathbb{P}\left(\frac{S_n}{n} - \theta \geq \varepsilon\right) \leq e^{-2n\varepsilon^2}$

$N = n\theta + n\varepsilon$

$N = \lceil n(\theta + \varepsilon) \rceil$

$N = \left\lceil 300\left(0,8 + \frac{300 - 0,8 \times 300}{300}\right) \right\rceil$

$N = 300$

$N = \lceil 300(0,8 + 0,0707) \rceil$

$N = 262$

$\infty \quad (1-\alpha)^k$

Rq en 5 : On pouvait faire de Markov, Tchebitchev ou trouve des valeurs de k trop grande (≈ 4000).

exo seul

Ex 6 : Moyenne empiriq & autre.

- $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in]0,1[})$; $(X_1, \dots, X_n)$ iid Bern(θ)
- 2 estimateurs :
  - $\hat{\theta}_n^1 = \frac{1}{n} \sum_{i=1}^n X_i$
  - $\hat{\theta}_n^2 = \min\{X_1, \dots, X_n\}$

1) Les estimateurs sont-ils des estimateurs sans biais de θ ?
(ie calculer $\mathbb{E}(\hat{\theta}_n^i)$)

- $\mathbb{E}(\hat{\theta}_n^1) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \underset{\text{linéarité de l'espérance}}{=} \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^n X_i\right)$

$\underset{\text{linéarité}}{=} \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta(X_i) = \theta$.

⚠ $\mathbb{E}(\min(A)) \neq \min \mathbb{E}(A)$.

$\hat{\theta}_n^2 = \min\{X_1, \dots, X_n\}$ prend les valeurs $\{0, 1\}$.

$\mathbb{P}_\theta(\hat{\theta}_n^2 = 1) = \mathbb{P}_\theta\left(\bigcap_{i=1}^n \{X_i = 1\}\right) = \prod_{i=1}^n \mathbb{P}_\theta(X_i = 1) = \theta^n$

$\mathbb{P}_\theta(\hat{\theta}_n^2 = 0) = 1 - \theta^n$.

$\mathbb{E}_\theta(\hat{\theta}_n^2) = 1 \times \mathbb{P}_\theta(\hat{\theta}_n^2 = 1) + 0 \times \mathbb{P}_\theta(\hat{\theta}_n^2 = 0) = \theta^n$

$\rightarrow$ $\hat{\theta}_n^2$ n'est pas un estimateur sans biais de θ. (sauf pr $n=1$)

2) Pr $i \in \{1, 2\}$, calculer le risque $\mathbb{E}_\theta[(\hat{\theta}_n^i - \theta)^2]$

$\mathbb{E}_\theta((\hat{\theta}_n^1 - \theta)^2) = Var(\hat{\theta}_n^1)$ car $\mathbb{E}_\theta(\hat{\theta}_n^1) = \theta$

$= Var_\theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2} Var_\theta\left(\sum_{i=1}^n X_i\right)$

$= \frac{1}{n^2}\sum_{i=1}^n Var_\theta(X_i) = \frac{\theta(1-\theta)}{n}$.

Rq $X \sim$ Bern(θ) : $Var(X) = \mathbb{E}(X^2) - \underbrace{\mathbb{E}(X)}_{\theta}{}^2 = \theta - \theta^2 = \theta(1-\theta)$.

$X^2 = X$ $\quad = \mathbb{E}((X - \mathbb{E}(X))^2)$

$\mathbb{E}_\theta((\hat{\theta}_n^2 - \theta)^2) = (1-\theta)^2\, \mathbb{P}_\theta(\hat{\theta}_n^2 = 1) + \theta^2\, \mathbb{P}_\theta(\hat{\theta}_n^2 = 0)$

$= (1-\theta)^2 \theta^n + \theta^2(1 - \theta^n) = (1 - 2\theta + \theta^2)\theta^n + \theta^2 - \theta^{n+2}$

$= -2\theta^{n+1} + \theta^n + \theta^2$

ou bien

$\mathbb{E}_\theta((\hat{\theta}_n^2 - \theta)^2) = \mathbb{E}_\theta((\hat{\theta}_n^2)^2) - 2\theta\, \mathbb{E}_\theta(\hat{\theta}_n^2) + \theta^2$

$= \theta^n - 2\theta^{n+1} + \theta^2$.

3) Lequel de ces 2 estimateurs est le meilleur qd $n = 2$.

$\mathbb{E}_\theta[(\hat{\theta}_n^1 - \theta)^2] = \frac{\theta(1-\theta)}{n}$, $\mathbb{E}_\theta[(\hat{\theta}_n^2 - \theta)^2] = \theta^n - 2\theta^{n+1} + \theta^2$

Le meilleur estimateur : c'est celui q a le plus petit risq quadratiq

$n=2$: $\mathbb{E}_\theta[(\hat{\theta}_2^1 - \theta)^2] = \frac{\theta(1-\theta)}{2}$, $\mathbb{E}_\theta[(\hat{\theta}_2^2 - \theta)^2] = \theta^2 - 2\theta^3 + \theta^2$

$= 2\theta^2(1-\theta)$

$\mathbb{E}_\theta[(\hat\theta_2^2-\theta)^2] \leq \mathbb{E}_\theta[(\hat\theta_2^1-\theta)^2]$

ssi $\theta \leq \frac{1}{4}$.

$\Rightarrow$ d'où $\theta \leq \frac{1}{4}$, $\hat\theta_2^2$ est meilleur que $\hat\theta_2^1$.

Ex 9 : Estimateur de la variance.

$(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in]0,1[})$, $(X_1,\dots,X_n) \sim \mathrm{Bern}(\theta)$.

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, On veut estimer la variance $v(\theta) = \theta(1-\theta)$.

1) Mq égalité valable pr loi de Bernoulli

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

$$\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - 2\bar{X}_n \underbrace{\frac{1}{n}\sum_{i=1}^n X_i}_{\bar{X}_n} + \bar{X}_n$$

$$= \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2$$

$\boxed{\text{FF}: \mathbb{E}((X-\mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2}$

$\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n^2$ tjrs vrai

$= \bar{X}_n - \bar{X}_n^2$ uniqt pr Bernoulli

$= \bar{X}_n(1-\bar{X}_n)$

D'où $\bar{X}_n(1-\bar{X}_n)$ est la variance empirique des $(X_i)_n$.

2) L'estimateur $\hat{v}_n = \bar{X}_n(1-\bar{X}_n)$ est-il un estimateur sans biais de $v(\theta)$?

ie calculer $\mathbb{E}(\hat{v}_n) = \mathbb{E}(\bar{X}_n(1-\bar{X}_n)) = \mathbb{E}(\bar{X}_n - \bar{X}_n^2)$

$= \mathbb{E}(\bar{X}_n) - \mathbb{E}(\bar{X}_n^2) = \theta - \mathbb{E}(\bar{X}_n^2)$

$= \theta - \mathbb{E}_\theta\left[\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2\right] = \frac{1}{n^2}\mathbb{E}_\theta\left[\sum_{1\leq i,j\leq n} X_iX_j\right]$

Autre méthode :

$\mathbb{E}_\theta(\bar{X}_n^2) = \mathrm{Var}_\theta(\bar{X}_n) + \mathbb{E}(\bar{X}_n)^2 = \frac{\theta(1-\theta)}{n} + \theta^2$.

$= \frac{n-1}{n}\theta(1-\theta)$

L'estimateur est asymptotiquement sans biais de $v(\theta)$

3) Lequel de ces 2 estimateurs est le meilleur qd $n=2$?

$\tilde{v}_n = (X_1)(1-X_2)$, $\mathbb{E}_\theta(\tilde{v}_n) = \mathbb{E}_\theta[X_1(1-X_2)]$.

$= \mathbb{E}_\theta(X_1)\,\mathbb{E}_\theta(1-X_2)$ car X_1 & X_2 st indep.

$= \theta(1-\theta) = v(\theta)$

$\Rightarrow \tilde{v}_n$ est un estimateur sans biais de $v(\theta)$

(R) $\hat{v}_n = \overline{X}_n(1-\overline{X}_n)$

$\mathbb{E}_\theta(\hat{v}_n) = \frac{n-1}{n} v(\theta)$ et $\hat{v}_n$ est un estimat^{r} asymptotiquement sans biais de $v(\theta)$.

$\left(\lim_{n\to\infty} \mathbb{E}_\theta(\hat{v}_n) = v(\theta)\right)$.

4) Quel estimat^{r} préférez-vous ?

$\tilde{v}_n$ utilise peu d'informat^{o} : juste X_1 et X_2 alors $\hat{v}_n$ utilise tt l'échantillon $X_1, \dots, X_n$.

$\hat{v}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta \text{ ps}} \theta(1-\theta)$ et $\overline{X}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta \text{ ps}} \theta$

$\iff \exists\, \Omega_1 \in \mathcal{F}$, $\mathbb{P}_\theta(\Omega_1) = 1$,

tq $\forall \omega \in \Omega_1 : \overline{X}_n(\omega) \xrightarrow[n\to\infty]{} \theta$

$\Rightarrow \overline{X}_n(\omega)(1-\overline{X}_n(\omega)) \xrightarrow[n\to\infty]{} \theta(1-\theta)$.

de $\{\omega \in \Omega : \hat{v}_n(\omega) \xrightarrow[n\to\infty]{} v(\theta)\} \supset \Omega_1$.

de $\mathbb{P}_\theta(\{\omega\in\Omega : \hat{v}_n(\omega) \xrightarrow[n\to\infty]{} v(\theta)\}) = 1$.

Intervalles de confiance

Ex 10 Échauffement du dé ou des candidats I (extrait partiel)

Mon fils fabrique dé, on se demande s'il est sûr que son dé est bien équilibré. Il s'arme de patience, lance 3 600 fois son dé et recueille le nb X de 6 obtenus. On note θ la proba d'obtenir 6 à le dé.

1) si le dé est bien équilibré, donner en justifiant, une minorat^{o} de $\mathbb{P}(500 \leq X \leq 700)$. $(\mathbb{E}_\theta(X) = 3600\cdot\theta)$

X nb 6 sur 3600 lancers. ⚠ $X \sim \mathcal{B}in(3600, \theta)$,

si le dé est bien équilibré $\theta = \frac{1}{6}$.

$P(500 \leq X \leq 700) = \sum_{k=500}^{700} \binom{3600}{k} \frac{1}{6}^k (1-\frac{1}{6})^{3600-k}$

$= F_X(700) - F_X(499)$ où $F_X(t) = P(X \leq t)$

(R) Hoeffding $\overline{X}_n = \frac{S_n}{n}$, $S_n = \sum_{i=1}^n X_i$, $\mathbb{P}_\theta(|\overline{X}_n - \theta| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}$

$\mathbb{P}_\theta(|S_n - \underbrace{n\theta}_{\mathbb{E}_\theta(S_n)}| \geq n\varepsilon)$

$\mathbb{P}_{\frac{1}{6}}(500 \leq X \leq 700) = \mathbb{P}_{\frac{1}{6}}(|X-600| \leq 100)$

$= 1 - \mathbb{P}_{1/6}(|X-600| > 100)$

X est très concentrée autour de 600.

$100 = 3600\,\varepsilon \Rightarrow \varepsilon = \frac{1}{36}$.

$\mathbb{P}_{1/6}(|X-100| \geq 600) \leq 2e^{-2\times 3600 \times \frac{1}{36^2}} = 2e^{-2\frac{100}{36}}$.

$\Rightarrow \mathbb{P}_{1/6}(500 \leq X \leq 700) \geq 1 - 2e^{-\frac{100}{18}} \approx 0{,}9923$.

13

"Vraie" valeur 0,9999926

2) On suppose dé non équilibré, donner un intervalle de confiance de niveau de confiance 95% pour θ, à inégalité de Hoeffding.

$\theta \in]0,1[$ inconnu. Intervalle de confiance $\hat{I}$.

$P_\theta(\theta \in \hat{I}) \geq 0,95$.

Estimateur de θ : $\frac{X}{3600} = \hat{\theta}$.

$\hat{\theta} - \frac{X}{3600}$ $\qquad \hat{I} = [\hat{\theta} - \varepsilon, \hat{\theta} + \varepsilon]$.

ε à déterminer tq $P_\theta(\theta \in \hat{I}) \geq 0,95$.

$\theta \in \hat{I} \iff |\theta - \hat{\theta}| \leq \varepsilon$.

$P_\theta(\theta \in \hat{I}) \geq 0,95$.

$\iff P_\theta(|\theta - \hat{\theta}| \leq \varepsilon) \geq 0,95$

$\iff P_\theta(|\theta - \hat{\theta}| \geq \varepsilon) \leq 0,05$

$\iff P_\theta\left(\left|\theta - \frac{X}{3600}\right| > \varepsilon\right) \leq 2e^{-2 \times 3600 \times \varepsilon^2}$

Il suffit de prendre $\varepsilon > 0$ tq $2e^{-7200 \times \varepsilon^2} \leq 0,05$.

$\varepsilon \geq \sqrt{\frac{1}{7200} \ln \frac{2}{0,05}} \approx \varepsilon = 0,023$.

$\hat{I} = \left[\frac{X}{3600} - 0,023 ; \frac{X}{3600} + 0,023\right]$

3) Sur les 3600 lancers, il observe 640 fois le chiffre 6. A la vue de cette observation, peut-il conclure que le dé est mal équilibré à moins de 5% de chances de se tromper?

Brouillon

Test statistique? Pour résoudre la question à $\theta = 1/6$.

H_0 : "dé équilibré" $\theta = \frac{1}{6}$ $\quad H_1$: "dé pas équilibré" $\theta \neq \frac{1}{6}$.

$P_{H_0}(\text{conclure } H_1) \leq 0,05$. (ie $p_0 = \frac{1}{6}$, $p_1 \neq \frac{1}{6}$)

Zone de Rejet : $R_{5\%} = \{X \geq k\}$ tq $P_{\frac{1}{6}}(X \geq k) \leq 0,05$

or $P_{\frac{1}{6}}(X \geq k) \approx P(\text{Unif}(\frac{1}{6}) \geq k)$

$R_0 =$

Ex 11 Élections présidentielles

Entre 2 tours, sondage $N = 1000$ p. $\to A$ / $\to B$ gagne ?

(H) $\emptyset$ abstent°.

S_1 : 507 A / 493 B | S_2 : 494 A / 506 B.

50,7 % 49,4 %

1) En tant que statisticien, qu'en pensez-vous ? (p^t pas le même ?)

$\hat{\theta}(\omega_1) = 0{,}507$ $\hat{\theta}(\omega_2) = 0{,}494$.

θ = proba de voter pr A inconnue.

$\mathbb{P}_\theta(\theta \in \hat{I}) \geq 0{,}95$.

$\hat{\theta} = \overline{X}_{1000}$, X_i : réponse du i^e ind.

2) Donner un intervalle de confiance pr θ de niveau de confiance 95 %.

$\mathbb{P}_\theta(\theta \in \hat{I}) \geq 0{,}95$ $(*)$.

$\hat{I} = [\hat{\theta} - \varepsilon, \hat{\theta} + \varepsilon]$.

$(*) \iff \mathbb{P}_\theta(|\theta - \hat{\theta}| > \varepsilon) \leq 0{,}05$.

or $\mathbb{P}_\theta(|\hat{\theta} - \theta| > \varepsilon) \leq 2e^{-2000\varepsilon^2}$

Il suffit de prendre $\varepsilon > 0$ tq

$$2e^{-2000\varepsilon^2} \leq 0{,}05 \iff \varepsilon \geq \sqrt{\frac{1}{2000} \ln \frac{2}{0{,}05}}$$

$$\varepsilon \geq \sqrt{\frac{1}{2000}\ln(40)} \simeq 0{,}043.$$

$\hat{I} = [\hat{\theta} - 0{,}043 ; \hat{\theta} + 0{,}043]$

$\hat{I}(\omega_1) = [0{,}507 - 0{,}043 ; 0{,}507 + 0{,}043]$

$\hat{I}(\omega_2) = [0{,}494 - 0{,}043 ; 0{,}494 + 0{,}043]$.

$\to$ P^t-on conclure que le candidat A va gagner à moins de 5% d'erreur ?

Ex 12 : Consommat° de carburant.

Conso litres au 100 km pr 100 chauffeurs sr m̂ trajet.
Conso carburant est (VA) X de loi inconnue & interpréter tabl° pr 100 observat° $x_1, \dots, x_{100}$ ≈ les observat° $X_1(\omega), \dots, X_n(\omega)$ d'une suite de (VA) indp de m̂ loi que X.
On a $p := \mathbb{P}(X > 7)$

1) Proposez une valeur numériq pr estimer p en justifiant.

Généralités : $x_1 = X_1(\omega), \dots, x_{100} = X_{100}(\omega)$

$\frac{1}{100}\sum_{i=1}^{100} X_i$ est estimat^R de la conso moyenne = (VA).

$\frac{1}{100}\sum_{i=1}^{100} x_i = \frac{1}{100}\sum_{i=1}^{100} X_i(\omega)$ est la val^R observée de l'estimateur.

- $p = \mathbb{P}(X > 7)$, Valeur numériq : 47 %.
C'est la fréquence d'apparit° des $X_i(\omega) > 7$.

$\hat{p} = \frac{1}{100} \cdot \mathrm{card}\{i \in \{1, \dots, 100\} \; X_i > 7\}$

$\hat{p} = \frac{1}{100}\sum_{i=1}^{100} \mathbb{1}\{X_i > 7\}$

$1 \leq i \leq 100$ $\quad Y_i = \mathbb{1}\{X_i > 7\}$

$(Y_i)_{1 \leq i \leq 100}$ st indépendants, $Y_i \sim \mathrm{Bern}(p)$.

- $\mathbb{E}_p(\hat{p}) = \frac{1}{100}\sum_{i=1}^{100} \mathbb{E}_p(Y_i) = p$.

$\Rightarrow$ $\hat{p}$ est un estimateur sans biais de p.

- si on note $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$

$\hat{p}_n \xrightarrow[n \to \infty]{\mathbb{P}_p \text{ ps}} p$. $\quad \hat{p}_n$ est un estimateur consistant.

2) Proposez un intervalle de confiance de niveau 95 % pr p.

On cherche ε tq $\mathbb{P}_p(p \in \hat{I}) \geq 0{,}95$.
$p \in \hat{I} \Leftrightarrow |p - \hat{p}| \leq \varepsilon$.
$\mathbb{P}_p(p \in \hat{I}) \geq 0{,}95 \Leftrightarrow \mathbb{P}_p(|p - \hat{p}| \leq \varepsilon) \geq 0{,}95$
$\Leftrightarrow \mathbb{P}_p(|\hat{p} - p| \geq \varepsilon) \leq 2e^{-200\varepsilon^2}$

Il suffit de choisir ε tq $2e^{-200\varepsilon^2} \leq 0{,}05$

soit $\varepsilon \geq \sqrt{\frac{1}{200}\ln\frac{2}{0{,}05}} \simeq 0{,}136$

$\hat{I} =]\hat{p} - 0{,}136, \hat{p} + 0{,}136[$ (fermé marche aussi)

$\mathbb{P}_p\left(p \in \hat{I}\right) \geqslant 0,95$.

$\hat{I}(\omega) =]0,334 ; 0,606[$

~~Surtout pas $\mathbb{P}_p(p \in]0,334 ; 0,606[)$~~

Tests Statistiques.

Ex 14 Mieux vaut ne pas (trop) se tromper.
Ds situaᵒs, décrivez ce q correspondrait au choix de l'hypo nulle H_0 & de l'hypo alternative H_1.

1) Promenade le long voie ferrée, bruit derrière nous. E q consiste l'err de 1° espèce ? 2° espèce ? Laquelle est la + grave. Que faites-vous ?

Généralités : Choix de H_0 & H_1 est tq on minimise la proba de se tromper qd on conclut H_1.

E_1 : Conclure H_1 alors que H_0 q est vraie.

" Un train passe " , " pas de train en vue "

H_0 ? H_1 ?

• 2 erreurs possibles :

▸ conclure que le train passe alors que ce n'est pas le cas
⇒ on se décalle pour rien.

▸ conclure que le train ne passe pas alors qu'il y en a un q arrive ⇒ pb ! H_1.

erreur à privilégier : E_1 : (erreur de 1° espèce)
Conclure H_1 alors que c'est H_0 q est vraie.

2) Test hypothèse ⟺ jugemt. Erreur de 1° espèce est de condamner le coupable. Vrai / Faux ?

→ Condamner un coupable n'est pas une erreur.

Erreurs : → condamner un innocent
→ innocenter un coupable.

H_0 : innocent / H_1 : coupable.

On conclut H_1 si on a suffisamment de preuves.

Conséqce si on souhaitait rendre l'erreur de 1° espèce nulle. ⇒ tt le monde sont innocent.

3) Voyage ds l'espace & vs voulez tester si la navette spatiale est sans danger.

H_0 : "la navette est dangereuse"

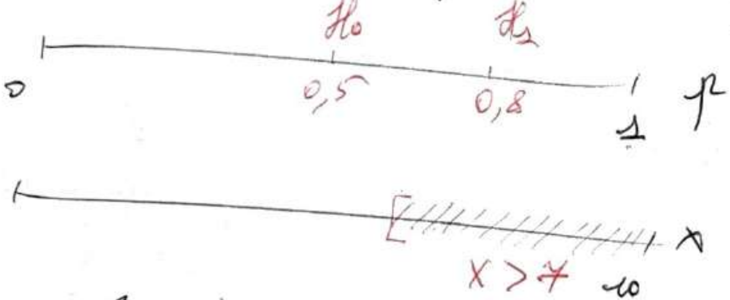
H_1 : "la navette est sûre".

Ex 15 Les bases

On observe une (va) X de loi $\mathcal{B}in(10,p)$ à p ayant 2 valeurs possibles 0,5 ou 0,8.
On décide que $H_0 : p=0,5$ est rejetée & $H_1 : p=0,8$ est acceptée si la vl observée $X \geq 7$.

1) Quelle est la zone de rejet ?

$X \sim \mathcal{B}in(10,p)$; 2 hypo. simples
$H_0 : p=0,5$ contre $H_1 : p=0,8$.



$\mathcal{R} = \{X \geq 7\}$ est l'événement sur lequel on conclut H_1 = zone de rejet
(on rejette H_0).

1. Les hypo se traduisent tjrs en f du param inconnu
- La zone de rejet est un event en f (va) (observations) X ou $X_1, \dots, X_n$.

2) Calculer la taille de ce test

$$\alpha_0 = \mathbb{P}_{0,5}(\mathcal{R}) = \mathbb{P}_{0,5}(X \geq 7) = 1 - \mathbb{P}_{0,5}(X \leq 6)$$
$$= 1 - 0,828 = 0,172.$$

3) Calculer la puissance de ce test.

$$\mathbb{P}_{0,8}(\mathcal{R}) = \mathbb{P}_{0,8}(X \geq 7) = 1 - \mathbb{P}_{0,8}(X \leq 6)$$
$$= 1 - 0,121 = 0,879.$$

Taille H_0, $\theta \in \Theta_0$ ← ici un seul point

$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\mathcal{R})$ ici $\Theta_0 = \{0,5\}$.

4) Qu'elles sont erreurs de 1° & 2° espèce ?

Erreur de 1° espèce : $\alpha_0 = 0,172$.
2° espèce : "conclure H_0 alors que c'est H_1"

$$\beta = \mathbb{P}_{0,8}(\mathcal{R}^c) = 1 - \mathbb{P}_{0,8}(\mathcal{R}) = 1 - \pi = 0,121.$$

5) Construire le test de H_0 contre H_1 à $\alpha = 10\ \%$.

Donner la zone de rejet. $\mathcal{R}_0 = \{X \geq k_\alpha\}$.

$\alpha = 10\%$

$\alpha = 5\%$ où $k_\alpha = \min\{k \in \{0, \dots, 10\} : \mathbb{P}_{0,5}(X \geq k) \leq \alpha\}$

$$\mathbb{P}_{0,5}(X \geq k) = 1 - \mathbb{P}_{0,5}(X \leq k-1)$$

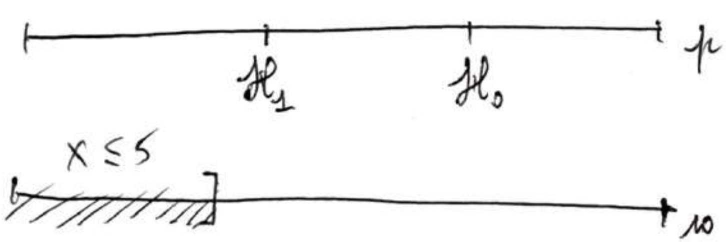
$$\mathbb{P}_{0,5}(X \geq k) \leq \alpha.$$

$$\Leftrightarrow \mathbb{P}_{0,5}(X \leq k-1) \geq 1 - \alpha.$$

- $k_{10\%} - 1 = 7 \Rightarrow k_{10\%} = 8$, $\mathcal{R}_0 = \{X \geq 8\}$
- $k_{5\%} - 1 = 8 \Rightarrow k_{5\%} = 9$, $\mathcal{R} = \{X \geq 9\}$

NB : $\mathbb{P}_{H_0}(\mathcal{R}) \leq \alpha$ proba de conclure H_1 alors que H_0 est vraie

Autre @ test $H_0 : p = 0,8$ contre $H_1 : p = 0,5$



$k_{10\%} = \max\{k \in \{0, \dots, 10\}\ \mathbb{P}_{0,8}(X \leq k) \leq 0,1\}$

Ex 16 : Un enseignant donne un questionn^re (10 propos) vraies ou fausses. Pr tester l'(H) que les étud répondent au hasard, il adopte la règle de décis :

- si + 7 réponses (+) : l'étud conn^t son cours
- si moins de 7 rép (+) : l'étud répd au hasard.

1) Déterminer proba de rejeter l'hypothèse (que l'étud joue aux devinettes) lorsqu'elle est correcte. C'est la proba de conclure à tort que l'étud connaît son cours

- H_0 : "étud répd au hasard" / H_1 : "étud conn^t son cours".
- X (v.a) = nbr réponses correctes.
- $X \sim \mathcal{B}in(10, \theta)$.

$\rightarrow H_0 : \theta = 0,5$ contre $H_1 : \theta > 0,5$

si $X \geq 7$ alors on conclut $\theta > 0,5$ sinon on conclut $\theta = 0,5$.

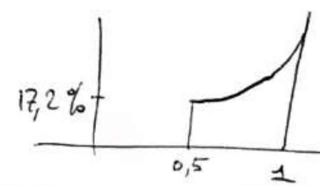
$$\mathcal{R}_0 = \{X \geq 7\}.$$

1) $\mathbb{P}_{0,5}(X \geq 7) = 0,172$.

2) Tracer f puissance de ce test

Puissance : $\pi :]0,5 ; 1] \longrightarrow [0,1]$

$\theta \longmapsto \mathbb{P}_\theta(X \geq 7)$



3) Déterminer le + petit nbr de réponses correctes qu'un étud doit avoir pr que le professeur soit sûr au risq de 5 % ou 1 % que l'étudiant ne joue pas aux devinettes.

$$R_0 = \{X \geq k_\alpha\}$$

$$k_\alpha = \min\{k \in \{0, \dots, 10\} : \mathbb{P}_{0,5}(X \geq k) \leq \alpha\}$$

$k_{6\%} = 8$, $k_{5\%} = 9$, $k_{1\%} = 10$.

Ex 18 ElecO présidentielle II

4 en 11, pt-on conclure que le candidat A va gagner dp moins de 5 % d'erreur.

Sondage sur $N = 1000$.

$X_1, \dots, X_N$ idd $\mathcal{B}ern(\theta)$.

Le candidat A ↓ ?

⇒ test statistiq.

↳ ① choix hypo

② ConstrcO zone de rejet

③ Ccl.

Rq en 5 (Central téléphoniq)

↳ Tchebitchev, on obtient un meilleur résultat qu'avec Hoeffding.

① Choix hypo.

H_0 : "A ne gagne pas", H_1 : "A gagne"

$H_0 : \theta \leq \frac{1}{2}$ $\qquad H_1 : \theta > \frac{1}{2}$.

② ConstrcO de la zone de rejet

PropoO : $S_N = \sum_{i=1}^{N} X_i$, $R_0 = \{S_N > \cancel{500}\ k\}$

$$\sup_{\theta \leq 0,5} \mathbb{P}_\theta(S_N > 500) = \mathbb{P}_{0,5}(S_N > 500) \approx \frac{1}{2}$$

or sous $\mathbb{P}_{0,5}$, $S_N \sim \mathcal{B}in(N, \frac{1}{2})$

$N - S_N \sim \mathcal{B}in(N, \frac{1}{2})$

si on affirme que A gagne si $S_N > 500$ alors on a une chance $\leq \frac{1}{2}$ de se tromper.

$\mathcal{R} = \{S_N \geq k\}$ tq $\underbrace{\sup_{\theta \leq 0,5} \mathbb{P}_\theta(S_N \geq k)}_{\mathbb{P}_{0,5}(S_N \geq k)} \leq 0,05$.

car $\theta \longmapsto \mathbb{P}_\theta(S_N \geq k)$ est croissante.

Hoeffding : $\mathbb{P}_\theta(S_n - n\theta \geq n\varepsilon) \leq e^{-2n\varepsilon^2}$

$= \mathbb{P}_\theta\left(\frac{S_n}{n} - \theta \geq \varepsilon\right) \leq e^{-2n\varepsilon^2}$

$k = N \times \frac{1}{2} + N\varepsilon$, $\varepsilon = \sqrt{\frac{1}{2000}\ln\left(\frac{1}{0,05}\right)}$, $N = 1000$

$e^{-2N\varepsilon^2} \leq 0,05$.

$\mathcal{R}_f = \{S_{1000} \geq 539\}$.

(3) ccl : $S_{1000}(\omega_1) = 507$, $\omega_1 \notin \mathcal{R}$
$(\omega_2) = 494$, $\omega_2 \in \mathcal{R}$.

Ex 19 Un nouveau traitement !
Tumeurs à ¢ géantes. N° huit. Étude sur 35 pat.
↪ rép ⊕ au ttt sur 30 pat.
Lui affirme que ce ttt est efficace à 85 %.

1) Mettant en doute les propos de son médecin, elle effectue un test stat.

Pt de vue de la patiente ?

H_0 : "le ttt n'est pas efficace" $\theta \leq 0,85$

H_1 : "le ttt est efficace". $\theta > 0,85$

2) Pt de vue médecin

H_0 : "le ttt est efficace" $\theta \geq 0,85$

H_1 : "le ttt n'est pas efficace"

3) Construct° de 2 tests. d'avoir une rpat.

- S_{35} : nb pers. ayant rep ⊕ à ttt. θ proba inconnue
- $\mathcal{R} = \{S_{35} \geq k_\alpha : k_\alpha = \min\{k \in \{0, \dots, 35\}, \mathbb{P}_{0,85}(S_{35} \geq k) \leq \alpha\}\}$

On vt pr un niveau α que $\sup_{\theta \leq 0,85} \mathbb{P}_\theta(\mathcal{R}^\alpha) \leq \alpha$

or $\theta \longmapsto \mathbb{P}_\theta(S_{35} \geq k_\alpha) \nearrow$ de $\sup_{\theta \leq 0,85} \mathbb{P}_\theta(S_{35} \geq k_\alpha) = \mathbb{P}_{0,85}(S_{35} \geq k_\alpha)$

$\alpha = 5\%$.

Fiche n°2 : Variables aléat. discrètes

Ex 1 : Contrôle qualité

Mibon Plumé

On souhaite contrôler qualité f^t d'une usine produisant des ampoules. p : proba 💡 défectueuse. On considère que la fabricaⁿ a un régime normal qd $p = 10^{-3}$.

1) On prélève au hasard 500 ampoules indep & on note X le nbr d'ampoules défectueuses de ce lot. Loi de (va) X ? Approx loi qd régime est normal.

$X \sim \mathcal{B}in(500, p)$ (sous $\mathbb{P}_p$).

p la proba inconnue de produire une ampoule défectueuse.

→ si le régime est normal sous $\mathbb{P}_{10^{-3}}$: $X \sim \mathcal{B}in(500, 10^{-3})$ la loi de X pt ê approchée p loi de Poisson $\mathcal{P}(\frac{1}{2})$.

"$\mathcal{B}in(n,p)$ approchée p $\mathcal{P}(np)$ qd n grd".

2) A (va) X, construire un test permettant de déterminer si fabricaⁿ a un régime normal ou si elle est détériorée, sachant que l'on souhaite que la proba de conclure à tort que le régime est détérioré soit inférieure à 5%. Quelle est la taille du test ? Que conclure si on observe $X(\omega) = 3$? Obtient-on m̂ cclⁿ a niv 1% ?

"régime détérioré" : H_1 / "rég. normal" : H_0

$\mathbb{P}_{H_0}$("conclure H_1") $\leq \alpha$.

$H_0 : p = 10^{-3}$ contre $H_1 : p > 10^{-3}$

ou m̂ $\leq$. (*)

⇒ Construire le test à l'aide de la va X :

zone de rejet : $\mathcal{R}_{0\,5\%} = \{X \geq k\}$

tq $\mathbb{P}_{10^{-3}}(X \geq k) \leq 0{,}05$.

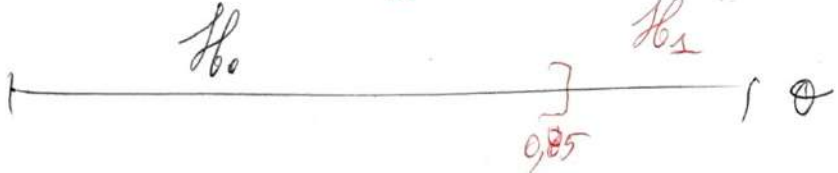
(ou

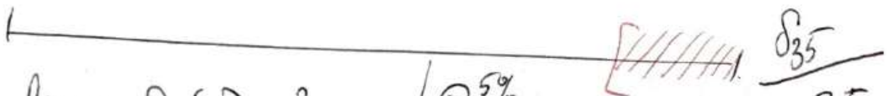
(*) $\sup_{p \leq 10^{-3}} \mathbb{P}_p(X \geq k) \leq 0{,}05$

<u>Retour en 19</u> ; $\alpha = 0{,}05$

$\mathbb{P}_{0,85}(S_{35} \geq k_{5\%}) \leq 0{,}05$

$\Leftrightarrow \mathbb{P}_{0,85}(S_{35} \leq \underbrace{k_{5\%} - 1}_{33\%}) \geq 0{,}95$ $\quad \boxed{k_{5\%} = 34}$





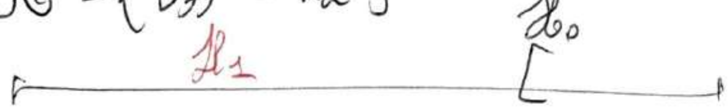
On donne $S_{35}(\omega) = 30$, $\omega \notin \mathcal{R}^{5\%}$ $\quad \left\{ \frac{S_{35}}{35} \geq \frac{34}{35} \right\}$

On ne pt pas conclure que tt est efficace à + 85 %.

→ si on inverse :

$H_0 : \theta \geq 0{,}85$ contre $H_1 : \theta < 0{,}85$.

$\mathcal{R}_0^{\alpha} = \{ S_{35} \leq k_\alpha \}$



$k_\alpha = \max \{ k \in \{0, \dots, 35\} : \sup_{\theta \geq 0{,}85} \mathbb{P}_\theta(S_{35} \leq k) \leq \alpha \}$

or $\theta \mapsto \mathbb{P}_\theta(S_{35} \leq k) = 1 - \mathbb{P}_\theta(S_{35} \geq k+1)$ est ↘ en θ. Donc $\sup_{\theta \geq 0{,}85} \mathbb{P}_\theta(S_{35} \leq k) = \mathbb{P}_{0,85}(S_{35} \leq k)$

$\alpha = 0{,}05$, $k_{5\%} = 25$

$\mathcal{R}_{5\%} = \{ S_{35} \leq 25 \}$.

<u>Ccl</u> : $S_{35}(\omega) = 30$. $\omega \notin \mathcal{R}_{5\%}$.

On ne pt pas conclure que tt n'est pas efficace (à moins de 85 %).

[Dans les 2 cas, on conclut H_0 sans contrôler la proba de se tromper.]

(21)

or $\mathbb{P}_{10^{-3}}(X \geq k) \approx \mathbb{P}(\mathcal{P}ois(\frac{1}{2}) \geq k)$

$\boxed{\mathcal{R} = \{X \geq 3\}}$, $\mathbb{P}(\mathcal{P}(\frac{1}{2}) \geq k) \leq 0,05$.

$\mathbb{P}(\mathcal{P}(\frac{1}{2}) \leq k-1) \geq 0,95$.

$k-1=2$, $k=3$

Taille du test : $\mathbb{P}_{10^{-3}}(X \geq 3) = 1 - 0,986 = 0,014$.

si on observe $X(\omega) = 3$, $\omega \in \mathcal{R}$.

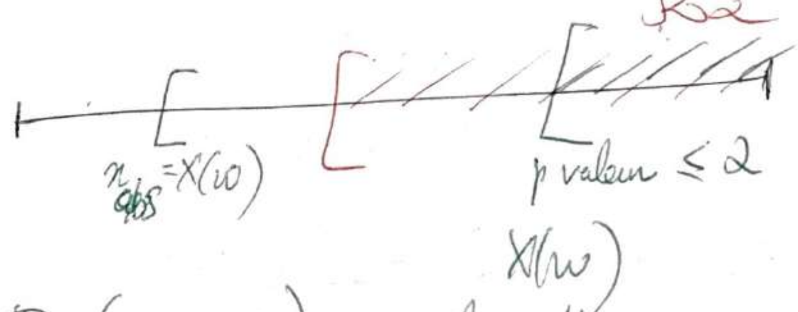
⇒ on rejette H_0 & on conclut que le régime est détériorée.

→ la p-valeur est $\mathbb{P}_{10^{-3}}(X \geq 3) = 1,4\% \leq 5\%$

• Pour un test à 1% :

p-valeur ≥ nivo 1%

$\omega \notin \mathcal{R}$, on ne pt pas conclure que le régime est détérioré : $\mathcal{R}_{0,01} = \{X \geq 4\}$



« $\mathbb{P}_{H_0}(X \geq x_{abs}) = p$ valeur »

3) Spps p soit en réalité 2.10^{-3}.
Q^elle est la proba de déclarer correctement, à l'issue du test de nivo 5%, que le régime de fabricat° est normal ?

$p = 2.10^{-3}$, $\mathbb{P}_{2.10^{-3}}(\mathcal{R}^c) = \mathbb{P}_{2.10^{-3}}(X \leq 2)$

erreur de 2nde espèce $= 92\%$.

Ex2 : Attente aux guichets.

Nbr pers. ds file d'attente donnée est (va) de loi de Poisson d^t param d'pd $\nearrow$ lieu $\searrow$ date. On a dénombré 6 dates $\to$ ville A, tq (va) $X_1, \dots, X_{10}$ $\to$ A s^t indep & m loi Pois(θ), $\theta > 0$ inconnu.

1) Calculer $\boxed{\text{l'estimat}^{\text{r}} \text{ du max de vraisembl}^{\text{ce}}}$ du param θ ? Q$^{\text{lle}}$ est sa loi sous $\mathbb{P}_\theta$?

$\to$ f de vraisemblance : pr $\theta > 0$: $\forall (x_1, \dots, x_n) \in \mathbb{N}^*$

$$V(\theta, x_1, \dots, x_n) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i) = \prod_{i=1}^{n} \left(\frac{\theta^{x_i}}{x_i!} e^{-\theta} \right)$$

$$= \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} e^{-n\theta}$$

$\to$ $\hat{\theta}_n = \underset{\theta>0}{\operatorname{argmax}}\, V(\theta, X_1, \dots, X_n) = \underset{\theta>0}{\operatorname{argmax}} \ln V(\theta, X_1, \dots, X_n)$

$\to$ $\ln V(\theta, X_1, \dots, X_n) = \ln\left(\frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} e^{-n\theta} \right) = \ln\left(\frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \right) - n\theta$

$$= \ln\left(\theta^{\sum_{i=1}^n x_i}\right) - n\theta - \ln\left(\prod_{i=1}^n x_i!\right) = \sum_{i=1}^n x_i (\ln\theta) - n\theta - \ln\left(\prod x_i!\right)$$

Or $\frac{\partial}{\partial\theta} \ln V(\theta, X_1, \dots, X_n) = \frac{\sum_{i=1}^n X_i}{\theta} - n$

θ	0		$\overline{X}_n$		∞
		$+$	ϕ	$-$	
$\theta \mapsto V(\theta, x_1, \dots, x_n)$		$\nearrow$		$\searrow$	

L'estimateur du max de vraisemblance est

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i = \overline{X}_n$$

Résoudre $\frac{\sum X_i}{\theta} - n = 0$

Quelle est sa loi sous $\mathbb{P}_\theta$?

loi de $\hat{\theta}_n$ sous $\mathbb{P}_\theta$, $\mathbb{P}_\theta(\hat{\theta}_n = \theta) = 0$

$\hat{\theta}_n$ pd ses valeurs ds $\{\frac{k}{10}, k \in \mathbb{N}\}$

pr $n = 10$, $\hat{\theta}_{10} = \frac{1}{10} \sum_{i=1}^{10} X_i$

$\hat{\theta}_{10}$ à val$^{\text{rs}}$ ds $\{\frac{k}{10}, k \in \mathbb{N}\}$

$$\mathbb{P}_\theta\left(\hat{\theta}_{10} = \frac{k}{10}\right) = \mathbb{P}_\theta\left(\sum_{i=1}^{10} X_i = k\right) = \frac{(10\theta)^k}{k!} e^{-10\theta}$$

car $\sum_{i=1}^{10} X_i \sim \text{Pois}(10\theta)$.

(somme de (va) indep. de loi de Poisson)

si $X \sim \text{Pois}(\alpha)$, $Y \sim \text{Pois}(\beta)$, X & Y indep

$\Rightarrow X + Y \sim \text{Pois}(\alpha + \beta)$.

$\hat{\theta}_{10}(\Omega) = \left\{ \frac{k}{10}, k \in \mathbb{N} \right\}$, $\forall x \in \hat{\theta}_{10}(\Omega)$.

$$\mathbb{P}_\theta(\hat{\theta}_{10} = x) = \frac{(10\theta)^{10x}}{(10x)!} e^{-10\theta}$$

2) Les val^rs observées st : 3, 0, 1, 4, 2, 3, 1, 4, 0, 2. Pt-on accepter hypo selon lq^elle il y a en moy. moins d'une pers. attendant au guichet au niv^o 5 % ?

Test statistique : Choix des Hypo :

$\mathbb{P}_\theta(X = x) = \frac{\theta^x}{x!} e^{-\theta}$, $\theta = \mathbb{E}_\theta(X_i)$

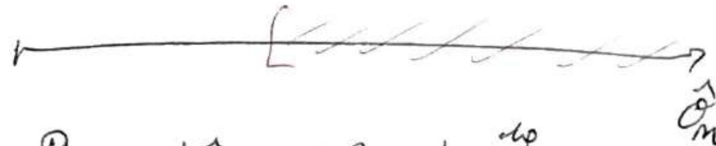
"Il y a en moyenne moins d'une personne q attend"
$\Leftrightarrow$ "$\theta \leq 1$"

Pt de vue des clients : ∘ $H_0 : \theta \geq 1$
∘ $H_1 : \theta < 1$.

Pt de vue de la banque ∘ $H_0 : \theta \leq 1$.
∘ $H_1 : \theta > 1$.

Admis $\theta \mapsto \mathbb{P}_\theta(X \geq k)$,
si $X \sim \mathcal{P}ois(\theta)$ sous $\mathbb{P}_\theta$ est croissante (en θ)

Zone de Rejet



$$R_0 = \{\hat{\theta}_{10} \geq t\} = \left\{ \sum_{i=1}^{10} X_i \geq k \right\}.$$

∘ Seuil est déterminé par : $\sup_{\theta \leq 1} \mathbb{P}\left(\sum_{i=1}^{10} X_i \geq k\right) \leq 0{,}05$

Or $\sup_{\theta \leq 1} \mathbb{P}_\theta\left(\sum_{i=1}^{10} X_i \geq k\right) = \mathbb{P}_1\left(\sum_{i=1}^{10} X_i \geq k\right)$

et sous $\mathbb{P}_1$, $\sum_{i=1}^{10} X_i \sim \mathcal{P}ois(10)$.

$\mathbb{P}_1\left(\sum_{i=1}^{10} X_i \geq k\right) \leq 0{,}05 \Leftrightarrow \mathbb{P}_1\left(\sum_{i=1}^{10} X_i \leq k-1\right) \geq 0{,}95$

$k - 1 = 15$, $k = 16$.

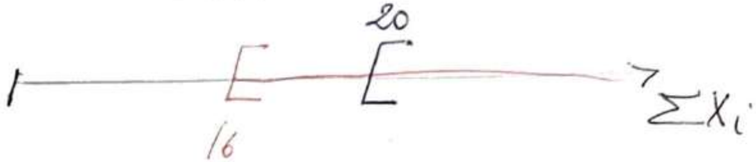
$R_0 = \left\{ \sum_{i=1}^{10} X_i \geq 16 \right\} = \{ \hat{\theta}_n \geq 1{,}6 \}$.

Ccl Test On observe $\sum_{i=1}^{10} X_i(\omega) = 20$, $\hat{\theta}_{10}(\omega) = 2$.

Dc $\omega \in R_0 \Rightarrow$ dc on rejette H_0.

On conclut qu'il y a + d'une personne en moyenne q attend au guichet.

- p-valeur : $\sup_{\theta \leq 1} \mathbb{P}_\theta \left(\sum_{i=1}^{10} X_i \geq 20 \right) = 0{,}003$



On aurait aussi rejeté pr un test à 1%.

$$\mathbb{P}_1 \left(\sum_{i=1}^{10} X_i \geq 20 \right) = 1 - \mathbb{P}_1 \left(\sum_{i=1}^{10} X_i \leq 19 \right)$$

$$= 1 - 0{,}997 = 0{,}003$$

Ex 3 : Estimat° du param d'une loi géom.

- loi géom param $\theta \in]0,1[$ est celle du tps d'attente du 1° succès de suite épreuves répétées ind pr chq épr de proba de succès θ.

Une va X sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ suit cette loi si

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}_\theta(X = k) = (1-\theta)^{k-1} \theta.$$

1) Vérifier $\mathbb{E}_\theta(X) = \frac{1}{\theta}$, $\mathrm{Var}_\theta(X) = \frac{1-\theta}{\theta^2}$

2) $(\Omega, \mathcal{F}, \mathbb{P}_\theta)_{\theta \in]0,1[}$ et un n-échantillon $X_1, \dots, X_n$ associé ie $\forall \theta$, vas X_i st $\mathbb{P}_\theta$-indép & m̂ loi ss $\mathbb{P}_\theta$. où $S_n = \sum_{i=1}^{n} X_i$

a) Donner la vraisemblance associée à l'échantillon $X_1, \dots, X_n$

$$V(\theta, x_1, \dots, x_n) = \prod \mathbb{P}_\theta(X_i = x_i) = \prod \left[\theta (1-\theta)^{x_i - 1} \right]$$

$$= \theta^n (1-\theta)^{\sum_{i=1}^{n} x_i - n}$$

à maximiser en θ.

$$\ln V(\theta, X_1, \dots, X_n) = n \ln(\theta) + \left(\sum_{i=1}^{n} X_i - n \right) \ln(1-\theta)$$

$$\frac{\partial L(\theta, X_1, \dots, X_n)}{\partial \theta} = \frac{n}{\theta} + \frac{\sum X_i - n}{1-\theta} = \frac{n(1-\theta) + \theta\left(\sum X_i - n\right)}{\theta(1-\theta)}$$

s'annule en $\hat{\theta}_n = \frac{n}{S_n}$ où $S_n = \sum_{i=1}^{n} X_i$

b) $\hat{\theta}_n = \frac{n}{S_n}$ est l'estimat^r du max de vraisemblance.

$$\hat{\theta}_n = \frac{n}{S_n}$$

e) Cet estimat^R est-il sans biais ? fortem^t consistant ?

$\hat{\theta}_n$ fort^t consistant de θ. $\mathbb{P}_\theta(\hat{\theta}_n \xrightarrow[n\to\infty]{} \theta) = 1$.

- LFGN ←
- Définiti° (condit° suffisante de CV^{ce} p.s).

si $\sum_{n\geq 1} \mathbb{P}_\theta(|\hat{\theta}_n - \theta| \geq \varepsilon) < \infty \Rightarrow \hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta\, p.s} \theta$

R LFGN : si $(X_i)_{i\in\mathbb{N}}$ st iid $\mathbb{E}(X_1) < \infty$ alors $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{p.s} \mathbb{E}(X_1)$

LFGN $\frac{S_n}{n} \xrightarrow[n\to\infty]{\mathbb{P}_\theta\ ps} \mathbb{E}_\theta(X_1) = \frac{1}{\theta}$

$\Rightarrow \frac{n}{S_n} \xrightarrow[n\to\infty]{\mathbb{P}_\theta\ ps} \theta$

CV^{ce} en proba : $\forall \varepsilon > 0$, $\mathbb{P}(|\overline{X}_n - \theta| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0$

$$\{\overline{X}_n \xrightarrow[n\to\infty]{} \theta\} = \bigcap_{\varepsilon>0} \bigcup_{N\in\mathbb{N}^*} \bigcap_{n\geq N} \{|X_n - \theta| \leq \varepsilon\}$$

R $\hat{\mu}_n$ fortement consistant de μ si $\hat{\mu}_n \xrightarrow[\mu\to\infty]{\mathbb{P}_\theta\ ps} \mu$

$\hat{\mu}_n = \overline{X}_n^2$, $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\overline{X}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta\ p.s} \mathbb{E}_\theta(X_1) = \theta$ LFGN

$\overline{X}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta\ p.s} \theta$

$\mathbb{P}_\theta(\Omega') = 1$: $\forall \omega \in \Omega'$ $\overline{X}_n(\omega) \xrightarrow[n\to\infty]{} \theta$

$\forall \omega \in \Omega'$ $\overline{X}_n^2(\omega) \xrightarrow[n\to\infty]{} \theta^2$

Ex 20 Charlatan

Médicament efficace à 90 % (fabricant)

Si 200 p., 160 ont guéris. Est charlatan ?

Pt de vue du fabricant :

$H_0 : \theta \geq 0,90$ contre $H_1 : \theta < 0,9$

Test à 5 % : $R_0 = \{ S_{200} \leq k \}$.

$X_i = \begin{cases} 1 & \text{si le } i^{\circ} \text{ pers. guérit} \\ 0 & \text{sinon.} \end{cases}$

$S_{200} = \sum_{i=1}^{200} X_i \sim \mathcal{B}in(200, \theta)$ sous $\mathbb{P}_\theta$.

Avec Hoeffding : $\{ S_{200} \leq 162 \}$. → explicaθ

$\mathbb{P}_{0,9}(S_{200} \leq 162) \leq 0,05$

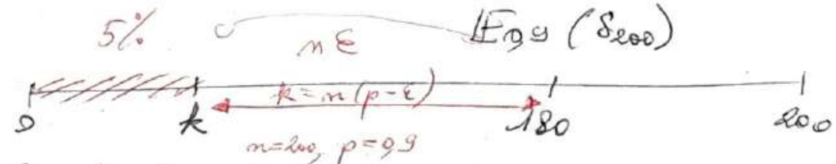
(Rq : $\mathbb{P}_{0,9}(S_{200} \leq 173) \leq 0,05$.

explicaθ : k est déterminé par $\sup_{\theta \geq 0,9} \mathbb{P}_\theta(S_{200} \leq k) \leq 0,05$

"proba de conclure H_1 alors que c'est H_0 q est vraie"

$\theta \longmapsto \mathbb{P}_\theta(S_{200} \leq k) = 1 - \mathbb{P}_\theta(S_{200} \geq k+1)$ est décroissante en θ.

$\Rightarrow \sup_{\theta \geq 0,9} \mathbb{P}_\theta(S_{200} \leq k) = \mathbb{P}_{0,9}(S_{200} \leq k)$



On cherche k : $\mathbb{P}_{0,9}(S_{200} \leq k) \leq 0,05$

$k = 200 \times 0,9 - 200\,\varepsilon$

$\mathbb{P}_{0,9}(S_{200} \leq 200 \times 0,9 - 200\,\varepsilon) = \mathbb{P}_{0,9}\left(\frac{S_{200}}{200} - 0,9 \leq -\varepsilon \right) \leq e^{-2 \times 200 \times \varepsilon^2}$

On choisit $\varepsilon > 0$ tq $e^{-400 \times \varepsilon^2} \leq 0,05$

$\varepsilon \geq \sqrt{-\frac{1}{400} \ln(0,05)}$

$k = \left\lfloor 200 \times 0,9 - 200 \times \sqrt{-\frac{1}{400} \ln(0,05)} \right\rfloor$

Ccl On observe $S_{200}(\omega) = 160$, $\omega \in R_0$ donc on rejette H_0 et on conclut que le tt n'est pas efficace à 90 %. ($\leq$ 5 % de chance de se tromper)

(28) $\boxed{\text{Arrondir à l'entier inférieur}}$

p-valeur : val^{rs} observé

$$\mathbb{P}_{0,9}(S_{200} \leq 160) = \mathbb{P}_{0,9}(S_{200} - 180 \leq -20)$$

vraie valeur $1,7 \cdot 10^{-5}$

$$\leq e^{-400\left(\frac{1}{10}\right)^2} = e^{-4} = 0,018$$

Autre pt de vue :

- $\mathcal{H}_0 : \theta \leq 0,9$ contre $\mathcal{H}_1 : \theta > 0,9$
- $\mathcal{R}_0 = \{S_{200} \geq k'\}$

k' tq $\sup_{\theta \leq 0,9} \mathbb{P}_\theta(S_{200} \geq k') \leq 0,05 \iff \mathbb{P}_{0,9}(S_{200} \geq k') \leq 0,05$

$k = 200 \times 0,9 + 200\varepsilon$

$$\mathbb{P}_{0,9}\left(\frac{S_{200} - 180}{200} \geq \varepsilon\right) \leq e^{-400\varepsilon^2} \leq 0,05.$$

Pour $\varepsilon \geq \sqrt{\frac{-1}{400}\ln(0,05)}$, $k = 180 + 18 = 198$.

$$\{S_{200} \geq 197,31\} = \{S_{200} \geq 198\}$$

$\omega \notin \mathcal{R}_0$, on ne rejette pas $\mathcal{H}_0$.

p valeur $\mathbb{P}_{0,9}(S_{200} \geq 160) \approx 99\%$.

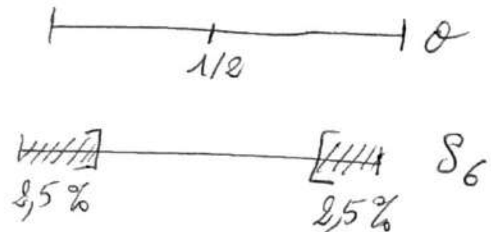
Ex 21 : Une pièce truquée (⚠ test bilatéral)

On lance pièce 6 fois, on obtient 6 "face".
Pt-on conclure au RG de 5% ou 1% que la pièce n'est pas équilibrée.
On constituera un test unilatéral ou bilatéral?

test bilatéral $\mathcal{H}_0 : \theta = \frac{1}{2}$ contre $\mathcal{H}_1 : \theta \neq \frac{1}{2}$.

test unilatéral $\mathcal{H}_0 : \theta = \frac{1}{2}$ contre $\mathcal{H}_1 : \theta < \frac{1}{2}$

Test bilatéral :



$$\mathcal{R}_0 = \{S_6 \leq k_1\} \cup \{S_6 \geq k_2\}.$$

k_1 & k_2 st déterminés par

$$\mathbb{P}_{\frac{1}{2}}(S_6 \leq k_1) + \mathbb{P}_{\frac{1}{2}}(S_6 \geq k_2) \leq 0,05.$$

$k_1 : \mathbb{P}_{1/2}(S_6 \leq k_1) \leq 0,025$

$k_2 : \mathbb{P}_{\frac{1}{2}}(S_6 \geq k_2) \leq 0,025$

$$\mathcal{R}_0 = \{ S_6 \leq 0 \} \cup \{ S_6 \geq 6 \} = \{ S_6 \in \{0, 6\} \}$$

On observe $S_6(w) = 0$: $w \in \mathcal{R}_0$: on conclut que la pièce n'est pas équilibrée.

$$\mathbb{P}_{1/2}(S_6 - 3 \leq -6\varepsilon) \leq e^{-12\varepsilon^2}$$

$\varepsilon \geq \sqrt{-\frac{1}{12}\ln(0{,}05)}$?) ne convient pas.

Taille du test : $\frac{1}{2^5} = 3{,}12\ \%$

$\mathbb{P}_{\frac{1}{2}}(S_6 = 0) = \frac{1}{2^6}$, $\mathbb{P}(S_6 \leq 1) = \frac{1}{2^6}(1+6) \approx 9{,}1$

$= 0{,}0156$ ~~$\mathbb{P}(S_6 \leq 0)$~~ $=$ ~~$\frac{1}{64}$~~

Ex 22 Dé — 3600 lancers, 640 fois → chiffre 6.

Peut-il conclure que dé mal équilibré à moins 5% de chances de se tromper ? On effectuera test stat.

1) $\mathbb{P}(500 \leq X \leq 700) \geq 0{,}9923$ Hoeffding

$\geq 0{,}95$ BT.

2) Interv de conf : $X \sim \mathcal{B}in(3600, \theta)$

$$\hat{I} = \left[\frac{X}{3600} - 0{,}023 ; \frac{X}{3600} + 0{,}023 \right]$$

$$\mathbb{P}_\theta(\theta \in \hat{I}) \geq 0{,}95$$

3) Test stat : $H_0 : \theta = \frac{1}{6}$ contre $H_1 : \theta \neq \frac{1}{6}$.

"peut-on dire dé mal équilibré à $\geq 5\%$ chance de se tromper".

H_1 : mal équilibré.

$$\mathcal{R}_0 = \{ X \leq 500 \} \cup \{ X \geq 700 \}$$

On observe $X(w) = 640$.

On a $w \notin \mathcal{R}_0$; on ne rejette pas H_0, on ne peut pas conclure que le dé est mal équilibré.

Réflexe : $\frac{\mathbb{P}_\theta \text{ p.s}}{X_n}$ } LFGN.

Proba : $g : \mathbb{R} \to \mathbb{R}$ est cont en ℓ & si $X_n \xrightarrow[n\to\infty]{p.s} \ell$
alors $g(X_n) \xrightarrow[n\to\infty]{p.s} g(\ell)$.

Preuve : $\Omega' = \{\omega \in \Omega : X_n(\omega) \xrightarrow[n\to\infty]{} \ell\}$.

$\mathbb{P}(\Omega') = 1$.

$\forall \omega \in \Omega' : g(X_n(\omega)) \xrightarrow{n\to\infty} g(\ell)$ car g cont en ℓ.

$\Rightarrow \{\omega \in \Omega : g(X_n(\omega)) \xrightarrow{n\to\infty} g(\ell)\} \supset \Omega'$

$\Rightarrow \mathbb{P}(g(X_n) \xrightarrow[n\to\infty]{} g(\ell)) = 1$.

- $\mathbb{E}_\theta(\hat{\theta}_n) = \mathbb{E}_\theta\left(\frac{n}{S_n}\right) = n\, \mathbb{E}_\theta\left(\frac{1}{S_n}\right)$

@ $n = 1$, $\mathbb{E}(X_1) = \sum_{k \in \mathbb{N}^*} \frac{1}{k} \mathbb{P}(X_1 = k)$

$= \sum_{k\in\mathbb{N}^*} \frac{1}{k}\theta(1-\theta)^{k-1} = \frac{\theta}{1-\theta} \sum_{k=1}^{\infty} \frac{(1-\theta)^{k-1}}{k}$

$f(\theta)$

$$f(\theta) = \sum_{k=1}^{\infty} \frac{(1-\theta)^k}{k}.$$

$$f'(\theta) = \sum_{k=1}^{\infty} -(1-\theta)^{k-1} = -\frac{1}{1-(1-\theta)} = \frac{-1}{\theta}$$

$$f(\theta) = -\ln(\theta)$$

$$\mathbb{E}_\theta\left(\frac{1}{X_1}\right) = \frac{-\theta \ln \theta}{1-\theta} \neq 0.$$

Rq On pourra faire un développement asymptotiq de $\mathbb{E}_\theta\left(\frac{n}{S_n}\right)$.

Rq : $\mathbb{E}_\theta\left(\frac{n}{S_n}\right) = n \sum_{k \leq n} \frac{1}{k} \mathbb{P}_\theta(S_n = k)$ & on connaît μ $k \geq n$.

$$\mathbb{P}_\theta(S_n = k) = \binom{k-1}{n-1} \theta^k (1-\theta)^{n-k}$$

loi binomiale négative.

3) Construire un intervalle de confiance $\hat{I}_n$ au niveau 95 % pour θ à l'aide de I(BT). si on a la certitude que $\theta > \frac{1}{4}$? Pk n'utilise-t-on pas l'inégalité de Hoeffding ?

$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| \geq \varepsilon\right)$

R: (BT) : $(X_i)_{i\in\mathbb{N}^*}$ iid de variance finie.

$$\mathbb{P}\left(|S_n - n\,\mathbb{E}(X_1)| \geq n\varepsilon\right) \leq \frac{Var(X_1)}{n\varepsilon^2}$$

à appliquer aux $(X_i)_{i\in\mathbb{N}}$ iid de loi Géo(θ) sous $\mathbb{P}_\theta$.

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{\theta}\right| \geq \varepsilon\right) \leq \frac{Var_\theta(X_1)}{n\varepsilon^2} = \frac{1-\theta}{\theta^2 n \varepsilon^2}$$

$$\underset{\uparrow}{\leq} \frac{1-1/4}{\left(\frac{1}{4}\right)^2} \times \frac{1}{n\varepsilon^2} = \frac{12}{n\varepsilon^2} \leq 0{,}05.$$

$\theta \geq \frac{1}{4}$ par hyp de l'énoncé

↳ la f $\theta \mapsto \frac{(1-\theta)}{\theta^2}$

Pour $\underbrace{\varepsilon \geq \sqrt{\frac{12}{0{,}05 n}}}_{\varepsilon(n)}$

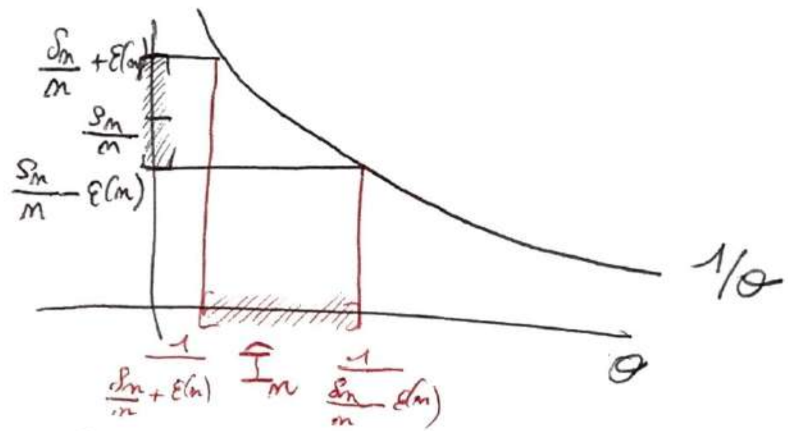
$$\mathbb{P}_\theta\left(\left|\frac{S_n}{n} - \frac{1}{\theta}\right| \geq \varepsilon(n)\right) \leq 0{,}05$$

$\Rightarrow$ intervalle $\hat{I}_n$ tq $\mathbb{P}_\theta(\theta \in \hat{I}_n) \geq 0{,}95$

fournit $\hat{J}_n$ pour $\frac{1}{\theta}$ de niveau 95%

$$\mathbb{P}_\theta\left(\frac{1}{\theta} \in \hat{J}_n\right) \geq 0{,}95.$$

$$\hat{J}_n = \left[\frac{S_n}{n} - \varepsilon(n), \frac{S_n}{n} + \varepsilon(n)\right]$$



$$\hat{I}_n = \left[\frac{1}{\varepsilon(n) + \frac{S_n}{n}} ; \frac{1}{\frac{S_n}{n} - \varepsilon(n)}\right] \cap [0,1] = \left[\frac{\hat{\theta}_n}{1+\varepsilon(n)\hat{\theta}_n}, \frac{\hat{\theta}_n}{1-\varepsilon(n)\hat{\theta}_n}\right] \cap [0,1]$$

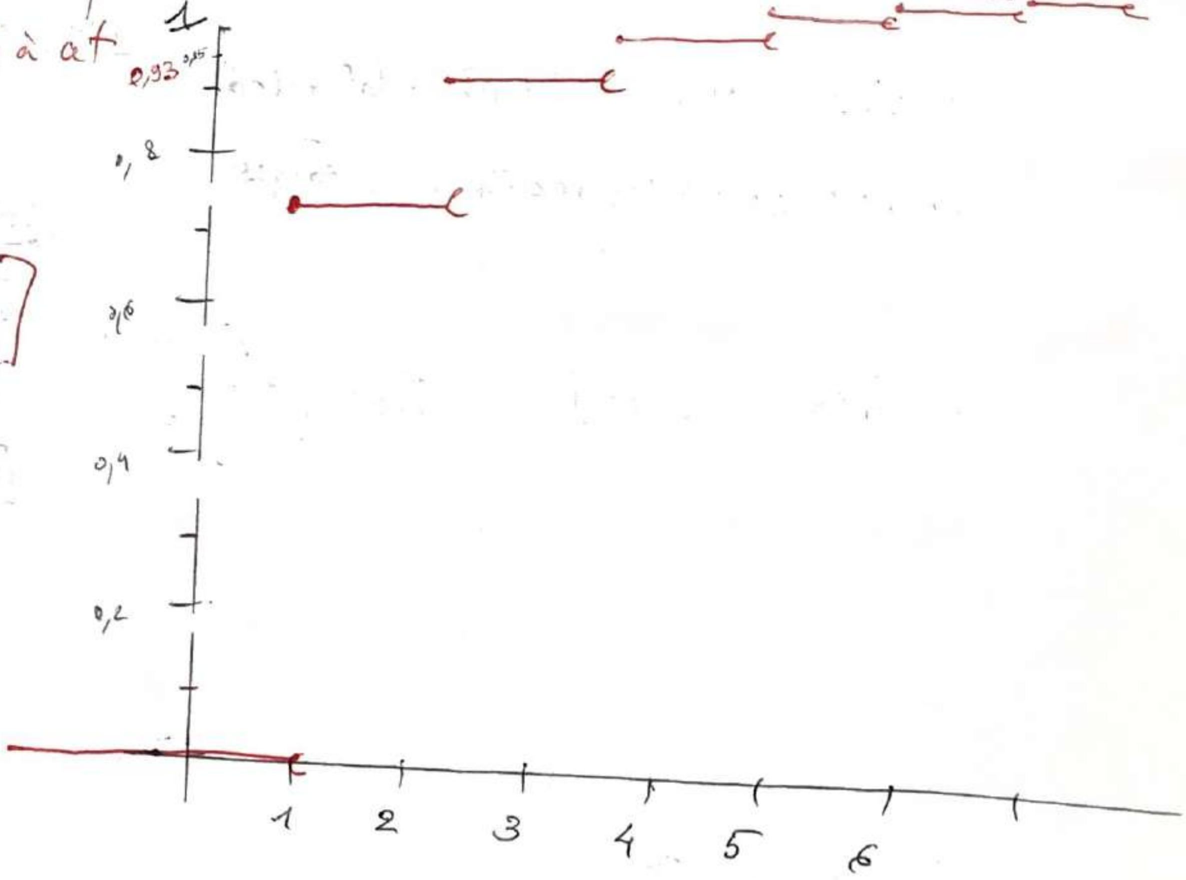
car $\frac{S_n}{n} - \varepsilon(n) \leq \frac{1}{\theta} \leq \frac{S_n}{n} + \varepsilon(n) \iff \frac{1}{\frac{S_n}{n} + \varepsilon(n)} \leq \theta \leq \frac{1}{\frac{S_n}{n} - \varepsilon(n)}$

On ne peut pas appliquer Hoeffding à $S_n = \sum_{i=1}^{n} X_i$ car les $(X_i)_{1\leq i\leq n}$ ne sont pas bornées.

4) a) cf table échantillon observé de taille 200.
Reproduisez & complétez le tablo. User le pr dessiner f de Répartiθ empiriq associée à cet échantillon. Unité verticale 10cm.

	146	40	7	5	1	1
Valeur	1	2	3	4	5	6
Fréquence	0,73	0,2	0,035	0,025	0,005	0,005

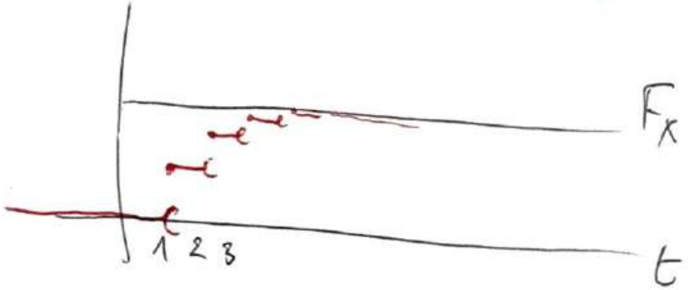
$$146 = \sum_{i=1}^{200} \mathbb{1}\{X_i(\omega) = 1\}, \quad 40 = \sum_{i=1}^{200} \mathbb{1}\{X_i(\omega)=2\}$$

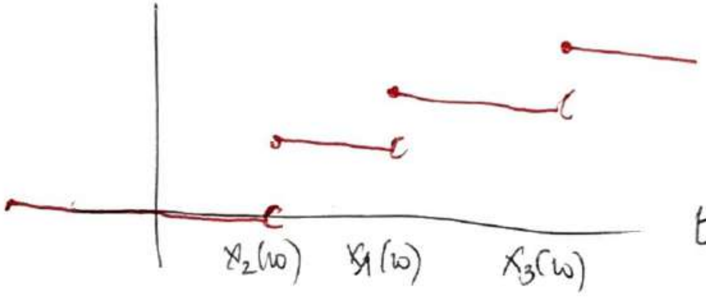


RS Foncθ de Répartiθ empirique

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{E}\left[\mathbb{1}_{\{X\leq t\}}\right]$$

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \leq t\}$$





b) Calculez l'estimat^r $\hat{\theta}_n$ de θ sur cet échantillon. (⚠ c'est une approximat°)

$n = 200$, $X_i(\omega)$: $1 \leq i \leq 200$

$$\hat{\theta}_{200}(\omega) = \frac{200}{S_{200}(\omega)} \approx 0,72.$$

$$S_{200}(\omega) = 146 \times 1 + 40 \times 2 + 7 \times 3 + 5 \times 4 + 1 \times 5 + 1 \times 6$$

$$\frac{1}{\hat{\theta}_{200}(\omega)} = 1 \times 0,73 + 2 \times 0,035 + 3 \times 0,025 + 4 \times \dots + 6 \times 0,005$$

c) Donnez l' IdC I_{200} observé de cet échantillon.

$$\hat{I}_{200}(\omega) = [0,402 ; 3,375] \quad , \quad \varepsilon(200) = \sqrt{\frac{12}{0,05 \times 200}}.$$

5) Autre idée pr estimer θ : ⓡ param va Bernoulli Y_i valant 1 si succès à i^e épreuve, 0 sinon.

Comment tablo $X_k(\omega)$ de 4) permet de reconstituer $Y_i(\omega)$.

cd val^r $\overline{Y_n}(\omega)$ des $Y_i(\omega)$ puis $IdC_{95\%}$ pr θ sur l'observat° de $\overline{Y_n}(\omega)$. Rq ?

Y_i, $P_\theta(Y_i = 1) = \theta$

(y_i), $\overset{x_1}{\overbrace{1}}, \overset{x_2}{\overbrace{0, 1}}, 1, 1 \dots$

X_i 1° succès après X_{i-1}.

$\sum y_i = 200$ nbr de succès.

Nbr d'échecs : $40 + 2 \times 7 + 3 \times 5 + 4 \times 1 + 5 \times 1 = 78$

40 échecs

$x_i = 3$

2 échecs

$\times 7$

$$\sum_{i=1}^{278} y_i = 200 \quad \text{nbr de succès}$$

$$\overline{Y}_{278}(\omega) = \frac{200}{278}$$

Rq. Le "278" est le résultat d'une va.

Variables Aléatoires à densité

1

Ex 4 Estimat du param d'une loi exponentielle

1) Soit X (va) $\sim Exp(\theta)$, de densité $f_\theta : x \mapsto \theta e^{-\theta x} \mathbb{1}_{\mathbb{R}_+}(x)$ où $\theta \in]0,\infty[$, Vérifier $\mathbb{E}(X) = \frac{1}{\theta}$, $Var(X) = \frac{1}{\theta^2}$

Astuce : $\int_{-\infty}^{\infty} (x-\frac{1}{\theta})^2 \theta e^{-\theta x} dx$ p.v.

$$\mathbb{E}_\theta(X) = \frac{1}{\theta} = \int_{\mathbb{R}} x f_\theta(x) dx = \int_0^\infty x \theta e^{-\theta x} dx.$$

$$Var_\theta(X) = \mathbb{E}_\theta(X^2) - \mathbb{E}_\theta(X)^2 = \int_0^\infty x^2 \theta e^{-\theta x} dx - \left[\mathbb{E}_\theta(X)\right]^2$$

⚠ $$\mathbb{E}_\theta(\varphi(X)) = \int_{\mathbb{R}} \varphi(x) f_\theta(x)\, dx.$$

2) On se propose d'estimer $\theta > 0$ d'une loi expo $Exp(\theta)$, au vu d'un échantillon $X_1, \dots, X_n$ de grde taille.
Plus précisément, on considère que X_i est des (va) δ^T positives, def sur $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in]0,\infty[})$ & st $\forall \theta \in]0,\infty[$, $\mathbb{P}_\theta$ - indépendantes & m̂ loi $Exp(\theta)$ sous $\mathbb{P}_\theta$. Pr $n \geq 1$

$S_n = \sum_{i=1}^n X_i$ & $T_n = \frac{n}{S_n}$.

Calculer l'estimateur du max de vraisemblance

La vraisemblance $\theta > 0$ $(x_1, \dots, x_n) \in \mathbb{R}$.

$$V(\theta, x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i)$$

$$= \prod_{i=1}^n \left(\theta e^{-\theta x_i} \mathbb{1}_{[0,\infty[}(x_i)\right)$$

$$= \theta^n e^{-\theta \sum_{i=1}^n x_i} \mathbb{1}_{[0,\infty[}(\min(x_i)) = \theta^n e^{-\theta \sum_{i=1}^n X_i}$$

$$L(\theta, X_1, \dots, X_n) = n \ln \theta - \theta \sum_{i=1}^n X_i$$

$$\frac{\partial}{\partial \theta} L(\theta, X_1, \dots, X_n) = \frac{n}{\theta} - \sum_{i=1}^n X_i$$

$\theta \mapsto V(\theta, X_1, \dots, X_n)$ atteint son max en $T_n = \frac{n}{S_n}$

T_n est l'estimateur du max de vraisemblance de θ. T_n est-il fortement consistant ?

⚠ Reflexe à avoir utiliser <u>Loi Forte des Grands Nombres</u>

<u>LFGN</u> : si $(X_n)_{n\in\mathbb{N}^*}$ suite iid tq $\mathbb{E}(|X_1|)<\infty \Rightarrow \frac{1}{n}\sum X_i \xrightarrow[\text{sous } \mathbb{P}_\theta]{\mathbb{P}_\theta \ p.s} \mathbb{E}(X_1)$

$$T_n = \frac{n}{\sum_{i=1}^{n} X_i} \xrightarrow[n\to\infty]{\mathbb{P}_\theta \ p.s} \theta \quad \text{"}\tfrac{1}{\theta}\text{"}$$

3) <u>Est-ce que</u> $\mathbb{E}_\theta(T_n)=\theta$ (si oui, on dit que T_n est sans biais)

$$\mathbb{E}_\theta(T_n) = \mathbb{E}_\theta\left(\frac{n}{S_n}\right) = n\,\mathbb{E}_\theta\left(\frac{1}{S_n}\right)$$

$n=1$, $S_1 = X_1$,

$$\mathbb{E}_\theta\left(\frac{1}{X_1}\right) = \int_{\mathbb{R}} \frac{1}{x} f_\theta(x)\,dx.$$

$$\mathbb{E}_\theta\left(\frac{1}{X_1}\right) = \int_0^\infty \frac{1}{x}\theta e^{-\theta x}\,dx = \infty$$

T_1 n'est pas sans biais.

4) Justifiez la CV en loi sous $\mathbb{P}_\theta$ ...

$$\frac{\theta S_n - n}{\sqrt{n}} \xrightarrow[n\to\infty]{loi} Z$$

$\boxed{TCL}$ Appliqué aux $(X_i)_{i\in\mathbb{N}^*}$ iid & de carré intégrables sous $\mathbb{P}_\theta$.

$$Z_n = \sqrt{n}\,\frac{\overline{X}_n - \mathbb{E}_\theta(X_1)}{\sqrt{Var_\theta(X_1)}} = \frac{S_n - n\mathbb{E}_\theta(X_1)}{\sqrt{n Var_\theta(X_1)}}$$

$\mathbb{E}_\theta(Z_n)=0$ et $Var_\theta(Z_n)=1$.

(R) $Var_\theta(S_n) = Var_\theta\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var_\theta(X_i)$ car les $(X_i)_{1\le i\le n}$ st indépdts

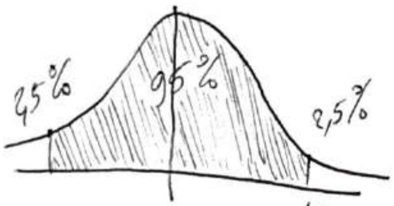
$= n\,Var_\theta(X_1)$

$$\frac{S_n - \underbrace{n\mathbb{E}_\theta(X_1)}_{=n\mathbb{E}_\theta(X_1)}}{\sqrt{n Var_\theta(X_1)}} = \frac{S_n - \frac{n}{\theta}}{\sqrt{\frac{n}{\theta^2}}} = \frac{\theta S_n - n}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathbb{P}_\theta \ loi} Z$$

5) En déduire, IdC pour θ à niveau de confiance 0,95 en négligeant l'erreur due à l'approx. gaussienne. (Appli num. pour $n=400$, $S_{400}(\omega) = 1460$).

$$\mathbb{P}_\theta\left(\left|\frac{\theta S_n - n}{\sqrt{n}}\right| \le \varepsilon\right) \xrightarrow[n\to\infty]{} \mathbb{P}(|Z|\le \varepsilon)$$

ε tq $\mathbb{P}(|Z| \le \varepsilon) = 0,95$, $\varepsilon = \Phi^{-1}(0,975) = 1,96$.



$\Phi^{-1}(0,975) = 1,96$.

$$\left\{\left|\frac{\theta S_n - n}{\sqrt{n}}\right| \le \varepsilon\right\} = \left\{\theta \in \hat{I}_n(\varepsilon)\right\}$$

$$= \Big\{\theta \in \underbrace{\left[\frac{n-\varepsilon\sqrt{n}}{S_n}, \frac{n+\varepsilon\sqrt{n}}{S_n}\right]}_{\hat{I}_n(\varepsilon) = \left[\frac{n}{S_n}\left(1-\frac{\varepsilon}{\sqrt{n}}\right);\ \frac{n}{S_n}\left(1+\frac{\varepsilon}{\sqrt{n}}\right)\right]}\Big\}$$

$\hat{I}_n(1,96)$ est un intervalle de confiance pour θ de niveau asymptotiq 95%..

$$\mathbb{P}_\theta\left(\theta \in \hat{I}_n(1,96)\right) \xrightarrow[n\to\infty]{} 0,95.$$

Ex5 Estimation de la durée d'une panne.

n ordi ; $t=0$: panne ordi, redémarre au bout de τ.
Pour i de 1 à n : quel tps T_i après début panne, 1° requête ordi i est acceptée par serveur.

$X_i = T_i - \tau$ entre redémarrage & 1° requête ordi i

$X_i \sim \mathcal{E}xp(a)$ $1\le i\le n$; a supposé connu, $a>0$, τ inconnu.

X_i indep.

↳ $\frac{1}{a}$ tps moyen entre le redémarrage du serveur & 1° requête

1) Quelle est la loi de (va) $V_n = \inf(X_1, \dots, X_n)$.

⚠ ↳ f de répartition ou f de survie.

$F_X(t) = \mathbb{P}(X\le t)$ $\qquad G_X(t) = \mathbb{P}(X>t)$

$$\forall t\in\mathbb{R},\ \mathbb{P}_\tau(V_n>t) = \mathbb{P}_\tau\left(\bigcap_{i=1}^n \{X_i>t\}\right)$$
$$= \prod_{i=1}^n \mathbb{P}_\tau(X_i>t) = \left[\mathbb{P}_\tau(X_1>t)\right]^n$$ car les $(X_i)_{1\le i\le n}$ ont la même loi.

or $\mathbb{P}_\tau(X_1>t) = \int_t^\infty a e^{-ax}\,dx = e^{-at}$ si $t\ge 0$

$= 1$ si $t<0$

$$\mathbb{P}_\tau(V_n) = \begin{cases} 1 & \text{si } t<0 \\ e^{-nat} & \text{si } t\ge 0 \end{cases}$$

, $V_n \sim \mathcal{E}xp(na)$

$F_{V_n}(t) = (1-e^{-nat})\mathbb{1}_{[0,\infty[}(t)$

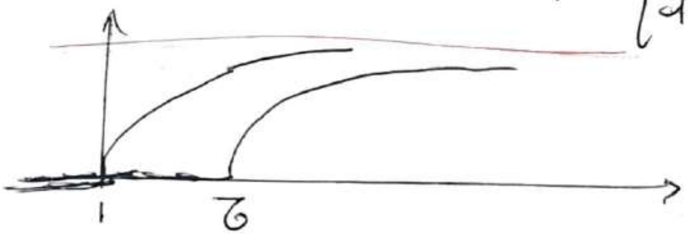
(R°) f de survie d'une loi exp(a)

$$G_a(t) = \begin{cases} 1 & \text{si } t<0 \\ e^{-at} & \text{si } t\geq 0 \end{cases} \qquad F_a(t) = \begin{cases} 0 & \text{si } t<0 \\ 1-e^{-at} & \text{si } t\geq 0 \end{cases}$$

2) Estimer la durée $\tau \in T_1, \dots, T_n$ en utilisant $\boxed{M}$ max de vraisemblance. On note W_n l'estimat^R obtenu.

• $\mathbb{P}(T_i \leq t) = \mathbb{P}(X_i + \tau \leq t) = \mathbb{P}(X_i \leq t-\tau)$

$$= F_a(t-\tau) = \begin{cases} 0 & \text{si } t-\tau<0 \\ 1-e^{-a(t-\tau)} & \text{si } t-\tau \geq 0. \end{cases}$$



Donc T_i a pr densité, (dérivée de fdR$^{\circ}$)

$$f_\tau(x) = a e^{-a(x-\tau)} \mathbb{1}_{[\tau,\infty[}(x).$$

• Autre $\boxed{M}$:

$$\mathbb{E}_\tau(\varphi(T_1)) \overset{?}{=} \int_{-\infty}^{\infty} \varphi(t) f_\tau(t)\, dt = \int_{-\infty}^{\infty} \varphi(x+\tau) a e^{-ax} \mathbb{1}_{[0,\infty[}(x)$$

$$= \int_0^\infty \varphi(x+\tau) a e^{-ax}\, dx$$

$$= \int_\tau^\infty \varphi(t) a e^{-a(t-\tau)}\, dt.$$

$$= \int_{-\infty}^{\infty} \varphi(t) \underbrace{a e^{-a(t-\tau)} \mathbb{1}_{[\tau,\infty[}}_{f_\tau(t)}\, dt$$

⇒ vraisemblance

$$\boxed{V(\tau, T_1, \dots, T_n) = \prod_{i=1}^{n} f_\tau(T_i)}$$

$$= \prod_{i=1}^{n} a e^{-a(T_i-\tau)} \mathbb{1}_{[\tau,\infty[}(T_i)$$

$$\overset{*}{=} a^n e^{-a\sum_{i=1}^{n} T_i} e^{na\tau} \prod_{i=1}^{n} \mathbb{1}_{[\tau,\infty[}(T_i)$$

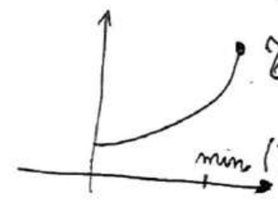
↳ $\prod \mathbb{1}_{[\tau,\infty[}(T_i) = 1$ si $T_i \geq \tau$ $\forall i \in \{1, \dots, n\}$

si $\min T_i \geq \tau$

$$= \mathbb{1}_{]-\infty, \min_{1\leq i\leq n} T_i]}(\tau)$$

$$\overset{*}{=} a^n e^{-a\sum_{i=1}^{n} T_i} e^{na\tau} \mathbb{1}_{]-\infty, \min_{1\leq i\leq n} T_i[}(\tau)$$

• $\tau \mapsto V(\tau, T_1(\omega), \dots, T_n(\omega))$



De l'estimat^R du max de vraisemblance de τ est $W_n = \min\{T_i\}$

3) Calculer le biais de W_n. et un estimateur sans biais Z_n de la durée de panne. Quelle est la variance de ce nouvel estimat^R ?

$$W_n = \min_{1\le i\le n}\{T_i\} \quad \& \quad T_i = \tau + X_i.$$

$$\min_{1\le i\le n}(T_i) = \tau + \min_{1\le i\le n}(X_i) \iff W_n = \tau + V_n$$

$$\mathbb{E}_\tau(W_n) = \tau + \mathbb{E}_\tau(V_n) = \tau + \frac{1}{na}$$

W_n n'est pas un estimateur sans biais de τ.
(il est asymptotiquement sans biais)

$\rightarrow Z_n$ sans biais : $Z_n = W_n - \frac{1}{na}$ car a est connu.

$$\mathbb{E}_\tau(Z_n) = \tau.$$

$$Var_\tau(Z_n) = Var_\tau(W_n) = Var_\tau(V_n) = \frac{1}{(na)^2}$$

4) Fabriquer un autre estimat^R sans biais de τ, en utilisant cette fois la moyenne empiriq $\overline{T_n}$ des T_i. Quelle est sa variance ?

$$\overline{T_n} = \frac{1}{n}\sum_{i=1}^{n} T_i, \quad \mathbb{E}_\tau(\overline{T_n}) = \mathbb{E}_\tau(T_1) = \tau + \frac{1}{a}$$

$R_n = \overline{T_n} - \frac{1}{a}$ est un estimat^R sans biais & fortement consistant de τ.
℘ la loi forte des grds nbres

$$Var_\tau(R_n) = Var_\tau(\overline{T_n}) = \frac{1}{n^2}\sum_{i=1}^{n} Var_\tau(T_i)$$
$$= \frac{1}{n} Var_\tau(T_1) = \frac{1}{n} Var_\tau(X_1) = \frac{1}{na^2}.$$

Rq $Var_\tau(Z_n) << Var_\tau(R_n)$.
$\Rightarrow Z_n$ est meilleur que R_n.

On pt aussi mq Z_n est un estimat^R fortement consistant de τ.

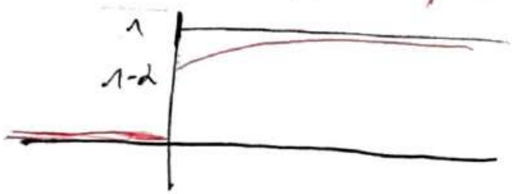
(Va) ni discrète ni à densité

Suppos qte X pluie (mm d'eau tombée en 24h) est modélisée p loi dont f de répartition est

$$F_{\alpha,\beta}(x) = (1-\alpha e^{-\beta x})\, \mathbb{1}_{[0,\infty[}(x).$$

où $\alpha \in]0,1]$ & $\beta > 0$ st param inconnus.

1) Tracer $F_{\alpha,\beta}$ (p @ $\alpha = \frac{1}{4}$, $\beta = 2$)



2) Proba pas précipita° ... 24h : $\mathbb{P}_{\alpha,\beta}(X=0)$

$$\mathbb{P}_{\alpha,\beta}(X=0) = F_{\alpha,\beta}(0) - F_{\alpha,\beta}(0_-) = F_{\alpha,\beta}(0) = 1-\alpha.$$

3) Loi (va) X discrète ? à densité ?

* X à densité ?

• si $\alpha = 1$, $F_{\alpha,\beta}$ est cont sur $\mathbb{R}$ dérivable p morcx dc X est à densité.

• si $\alpha \neq 1$, $F_{\alpha,\beta}$ n'est pas cont dc X n'est pas à densité.

* X est-elle discrète ?

non car la somme des sauts ne vaut pas 1.

4). Mq $\mathbb{E}_{\alpha,\beta}(X) = \frac{\alpha}{\beta}$.

X est une v.a positive.

$$\boxed{\mathbb{E}_{\alpha,\beta}(X) = \int_0^\infty \mathbb{P}_{\alpha,\beta}(X>t)\,dt = \int_0^\infty 1 - F_{\alpha,\beta}(t)\,dt}$$

$$\mathbb{E}_{\alpha,\beta}(X) = \int_0^\infty 1 - (1-\alpha e^{-\beta t})\,\mathbb{1}_{[0,\infty[}(t)\,dt$$

$$= \int_0^\infty \alpha e^{-\beta t}\,dt = \alpha\left[-\frac{e^{-\beta t}}{\beta}\right]_0^\infty = \frac{\alpha}{\beta}.$$

5) $X_1, \dots, X_n$ iid m loi que X.

$X_1(\omega), \dots, X_{30}(\omega)$ observées au mois de mai 2016 en mm st données p : cf la bib

Construire un estimat^R sans biais & consistant $\hat{\alpha}_n$ du param α. On pourra exprimer ce param en f nb jours sans précipitat^ons N_0.

$\hat{\alpha}_n$ estimateur de α.

$$N_n^0 = \sum_{i=1}^{n} \mathbb{1}_{\{X_i=0\}} \quad \text{nbr jours sans précipitat}^{\text{ons}}.$$

$$\hat{\alpha}_n = \frac{n - N_n^0}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i>0\}}$$

- Estimat^R sans biais de α :

car $\mathbb{E}_{\alpha,\beta}(\hat{\alpha}_n) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathbb{E}_{\alpha,\beta}(\mathbb{1}_{\{X_i>0\}})}_{\mathbb{P}_{\alpha,\beta}(X_i>0)=\alpha} = \alpha$

- Estimat^R fortement consistant de α p LFGN (ps) :

appliquée à $Y_i = \mathbb{1}_{\{X_i>0\}}$

iid intégrables.

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{\mathbb{P}_{\alpha,\beta}-p.s} \mathbb{E}_{\alpha,\beta}(Y_i) = \alpha$$

ie $\hat{\alpha}_n \xrightarrow[n\to\infty]{\mathbb{P}_{\alpha,\beta}\ ps} \alpha$

6) Donner cpt asymptotiq de la loi de $\hat{\alpha}_n$ & ed IdC p $\alpha = 95\%$.

à l'aide TLC appliqué aux $(Y_i)_{1\leq i\leq n}$.

iid de carrés intégrables :

$$\sqrt{n}\,\frac{\bar{Y}_n - \mathbb{E}_{\alpha,\beta}(Y_i)}{\sqrt{Var_{\alpha,\beta}(Y_1)}} \xrightarrow[n\to\infty]{\mathbb{P}_{\alpha,\beta}\text{ en loi}} Z \sim \mathcal{N}(0,1).$$

$\bar{Y}_n = \hat{\alpha}_n$, $\mathbb{E}_{\alpha,\beta}(Y_1) = \alpha$

$Var_{\alpha,\beta}(Y_1) = \alpha(1-\alpha)$, $Y_1 \sim Bern(\alpha)$

$$\sqrt{n}\,\frac{\hat{\alpha}_n - \alpha}{\sqrt{\alpha(1-\alpha)}} \xrightarrow[n\to\infty]{\mathbb{P}_{\alpha,\beta}\text{ loi}} Z$$

NB : $Z_n = \sqrt{n}\,\frac{\hat{\alpha}_n - \alpha}{\sqrt{\alpha(1-\alpha)}}$ « $\hat{\alpha}_n \simeq \alpha + \sqrt{\frac{\alpha(1-\alpha)}{n}}\,Z$

La loi de $\hat{\alpha}_n$ est approchée par $\mathcal{N}\left(\alpha, \sqrt{\frac{\alpha(1-\alpha)}{n}}\right)$

$$\mathbb{P}_{\alpha,\beta}\left(\sqrt{n}\,\frac{|\hat{\alpha}_n-\alpha|}{\sqrt{\alpha(1-\alpha)}}\leq\varepsilon\right)\xrightarrow[n\to\infty]{}\mathbb{P}(|Z|\leq\varepsilon)=0{,}95$$

$$\mathbb{P}_{\alpha,\beta}(|Z_n|\leq\varepsilon)$$

pr $\varepsilon = 1{,}96 = \Phi^{-1}(0{,}975)$.

2 Méthodes
- a) → $\hookrightarrow$ majoraⁿ de variance $\alpha(1-\alpha)$
- → $\hookrightarrow$ estimaⁿ de la variance

(a)
$$\mathbb{P}_{\alpha,\beta}\left(\alpha\in\left[\hat{\alpha}_n-1{,}96\sqrt{\frac{\alpha(1-\alpha)}{n}}\,;\hat{\alpha}_n+1{,}96\sqrt{\frac{\alpha(1-\alpha)}{n}}\right]\right)\xrightarrow[n\to\infty]{}0{,}95$$

$$\leq\mathbb{P}_{\alpha,\beta}\left(\alpha\in\left[\hat{\alpha}_n-\frac{1{,}96}{2\sqrt{n}}\,;\hat{\alpha}_n+\frac{1{,}96}{2\sqrt{n}}\right]\right)$$

$$\Rightarrow \liminf_{n\to\infty}\mathbb{P}_{\alpha,\beta}\left(\alpha\in\left[\hat{\alpha}_n-\frac{1{,}96}{2\sqrt{n}}\,;\hat{\alpha}_n+\frac{1{,}96}{2\sqrt{n}}\right]\right)\geq 0{,}95.$$

(b) Estimons la variance : TLC $\hookrightarrow$ autonormalisaⁿ :

$$\sqrt{n}\,\frac{\hat{\alpha}_n-\alpha}{\sqrt{\hat{\alpha}_n(1-\hat{\alpha}_n)}}\xrightarrow[n\to\infty]{\mathbb{P}_{\alpha,\beta}\text{ loi}}Z\sim\mathcal{N}(0,1).$$

$$\Rightarrow\mathbb{P}_{\alpha,\beta}\left(\alpha\in\left[\hat{\alpha}_n-1{,}96\sqrt{\frac{\hat{\alpha}_n(1-\hat{\alpha}_n)}{n}}\,;\hat{\alpha}_n+1{,}96\sqrt{\;''\;}\right]\right)\xrightarrow{n\to\infty}0{,}95$$

$$\hat{I}_n^M=\left[\hat{\alpha}_n-\frac{1{,}96}{2\sqrt{n}}\,,\hat{\alpha}_n+\frac{1{,}96}{2\sqrt{n}}\right].$$

$$\hat{I}_n^E=\left[\hat{\alpha}_n-1{,}96\sqrt{\frac{\hat{\alpha}_n(1-\hat{\alpha}_n)}{n}}\,;\hat{\alpha}_n+1{,}96\sqrt{\;''\;}\right]$$

$n=30$, $\hat{\alpha}_{30}(\omega)=\frac{14}{30}$

$\hat{I}_{30}^M(\omega)=$

$\hat{I}_{30}^E(\omega)=$

7) P^t-on conclure qu'il ne pleut pas plus d'un jour sur 2 $\hookrightarrow$ risq d'erreur d'au plus 5%. On effectuera un test à niv 5%
$H_0 : \alpha\geq\frac{1}{2}$ contre $H_1 : \alpha<\frac{1}{2}$.
en justifiant ce choix d'hypothèses.

$H_0 : \alpha\geq\frac{1}{2}$ contre $H_1 : \alpha<\frac{1}{2}$

$1-\alpha\leq\frac{1}{2}$ $\qquad$ $1-\alpha>\frac{1}{2}$.

"il ne fait pas beau dans le Nord" $\qquad$ "il fait beau dans le Nord"

$\mathcal{R}_{5\%} = \{ \hat{\alpha}_n \leq t_{5\%} \}$

$\sup_{\alpha \geq \frac{1}{2}} \mathbb{P}_{\alpha,\beta}(\hat{\alpha}_n \leq t_{5\%}) \leq 0{,}05$ $\quad k_{5\%} = n\, t_{5\%} \in \mathbb{N}$

$\sup_{\alpha \geq 1/2} \mathbb{P}_{\alpha,\beta}\left(\sum_{i=1}^{n} \mathbb{1}\{X_i > 0\} \leq k_{5\%} \right)$

$= \mathbb{P}_{\frac{1}{2},\beta}\left(\sum_{i=1}^{n} \mathbb{1}\{X_i > 0\} \leq k_{5\%} \right)$ & TLC

Trouver $k_{5\%} = \max\{ k \in \{0, \ldots, n\} : \mathbb{P}_{\frac{1}{2},\beta}\left(\sum_{i=1}^{n} \mathbb{1}\{X_i>0\} \leq k\right) \leq 0{,}05 \}$

$\mathbb{P}_{\frac{1}{2},\beta}\left(\sum_{i=1}^{n} \mathbb{1}\{X_i > 0\} \leq k_{5\%} \right)$

$$= \mathbb{P}_{\frac{1}{2},\beta}\left(\frac{\sum_{i=1}^{n} \mathbb{1}\{X_i>0\} - \frac{n}{2}}{\sqrt{n \frac{1}{2}(1-\frac{1}{2})}} \leq \frac{k_{5\%} - \frac{n}{2}}{\frac{\sqrt{n}}{2}} \right)$$

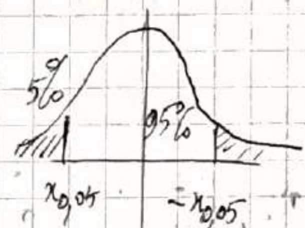
en négligeant l'erreur d'approx gaussienne

$$\simeq \mathbb{P}\left(Z \leq \frac{k_{5\%} - \frac{n}{2}}{\frac{\sqrt{n}}{2}} \right) = 0{,}05$$

$$\Rightarrow \frac{k_{5\%} - \frac{n}{2}}{\frac{\sqrt{n}}{2}} = \Phi^{-1}(0{,}05)$$

NB : $\dfrac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \dfrac{S_n - n\mathbb{E}(X_1)}{\sqrt{n \mathrm{Var}(X_1)}}$

$= \sqrt{n} \dfrac{\overline{X}_n - \mathbb{E}(X_1)}{\sqrt{\mathrm{Var}(X_1)}}$



$\Phi^{-1}(0{,}05) = x_{0,05} = -1{,}645$

$\mathbb{P}(Z \leq x_{0,05}) = 0{,}05$

$k_{5\%} = \left\lfloor \frac{n}{2} - 1{,}645 \frac{\sqrt{n}}{2} \right\rfloor =$

$n = 30, \quad \left\lfloor 15 - 1{,}645 \frac{\sqrt{30}}{2} \right\rfloor = 10$

$\mathcal{R}_{5\%} = \{ \sum_{i=1}^{30} \mathbb{1}\{X_i > 0\} \leq 10 \}$

$10 \notin \mathcal{R}_{5\%}$: on ne pt pas conclure H_1

is
b])
l
r+3
en vl^t abs
9.
$l_i)_{i \in \mathbb{N}^*}$
intégrables
inées.
,J

$\mathbb{P}_{\alpha,\beta}$

On définit $\hat\beta_n = \dfrac{\sum_{i=1}^{n} \mathbb{1}\{X_i \neq 0\}}{\sum_{i=1}^{n} X_i}$

8) Mq, presque sûrement, apcr $\sum_{i=1}^{n} X_i > 0$

Me indic : Calculer $\mathbb{P}_{\alpha,\beta}\left(\sum_{i=1}^{n} X_i = 0\right)$

ⓐ $\mathbb{P}_{\alpha,\beta}\left(\sum_{i=1}^{n} X_i = 0\right) = \mathbb{P}_{\alpha,\beta}\left(\bigcap_{i=1}^{n} \{X_i = 0\}\right)$

$= \prod_{i=1}^{n} \mathbb{P}_{\alpha,\beta}(X_i = 0)$ car les $(X_i)_{1\leq i\leq n}$ st indep.

$= \mathbb{P}_{\alpha,\beta}(X_1 = 0)^n$ car les (X_i) ont la m̂ loi.

$\Rightarrow l_n = (1-\alpha)^n$

$\mathbb{P}_{\alpha,\beta} \leq 1$

apcr $A_n = \left\{\sum_{i=1}^{n} X_i = 0\right\}$

ⓑ Ei $\boxed{BC}$ si $\sum_{n=1}^{\infty} \mathbb{P}_{\alpha,\beta}(A_n) < \infty \iff \mathbb{P}_{\alpha,\beta}\left(\bigcap_{n\in\mathbb{N}^*} \bigcup_{m>0} A_m\right) = 0$

Donc $\mathbb{P}_{\alpha,\beta}\left(\bigcup_{n\in\mathbb{N}^*} \bigcap_{m\geq n} A_m^c\right) = 1$

$\Rightarrow$ Donc presq sûrem^t apdcrg les A_n ne se réalisent plus $\sum_{i=1}^{n} X_i > 0$.

$\exists\, \Omega' \in \mathcal{F}$ tq $\mathbb{P}_{\alpha,\beta}(\Omega') = 1$ & $\forall \omega \in \Omega'$

$\exists\, n(\omega) \in \mathbb{N}^*$ tq $\forall n \geq n(\omega) : \sum_{i=1}^{n} X_i(\omega) > 0$

$\Rightarrow \hat\beta_n(\omega)$ est bien défini.

44

Rq p.s. $\hat\beta_n$ est bien défini pr n assez grand

plus proprement

$$\hat\beta_n(\omega) = \begin{cases} 0 & \text{si } \sum_{i=1}^{n} X_i(\omega) = 0 \\ \dfrac{\sum_{i=1}^{n} \{X_i \neq 0\}}{\sum_{i=1}^{n} X_i(\omega)} & \text{sinon.} \end{cases}$$

9) Mq $\hat\beta_n$ est un estimat^R fortement consistant de β

Rq $\hat I = \{i \in \{1, \dots, n\} : X_i \neq 0\}$.

$\hat\beta_n = \dfrac{\operatorname{card}(\hat I)}{\sum_{i\in \hat I} X_i} = \dfrac{1}{\overline{X}_{\hat I}}$.

mais les indices étant aléatoires, on ne pt pas utiliser directement la loi forte des grds nbres.

LFGN

On sait que $\mathbb{P}_{\alpha,\beta}$ ⓐ p.s.

$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \neq 0\} \xrightarrow[n\to\infty]{\mathbb{P}_{\alpha,\beta}\text{ p.s}} \mathbb{P}_{\alpha,\beta}(X_1 \neq 0) = \alpha$

$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbb{P}_{\alpha,\beta}\text{ p.s}} \mathbb{E}_{\alpha,\beta}(X_1) = \dfrac{\alpha}{\beta}$.

$\exists\, \Omega_1 \in \mathcal{F} : \mathbb{P}_{\alpha,\beta}(\Omega_1) = 1$ et $\forall \omega \in \Omega_1$

$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \neq 0\} \xrightarrow[n\to\infty]{} \alpha$

$\exists\, \Omega_2 \in \mathcal{F} : \mathbb{P}_{\alpha,\beta}(\Omega_2) = 1$ et $\forall w \in \Omega_2$:

$$\frac{1}{n} \sum X_i(w) \xrightarrow[n\to\infty]{} \frac{\alpha}{\beta}.$$

et $\mathbb{P}_{\alpha,\beta}(\Omega_1 \cap \Omega_2) = 1$ car $\mathbb{P}_{\alpha,\beta}(\Omega_1^c \cup \Omega_2^c)$

$$\mathbb{P}_{\alpha,\beta}(\Omega_1^c \cup \Omega_2^c) \leq \mathbb{P}_{\alpha,\beta}(\Omega_1^c) + \mathbb{P}_{\alpha,\beta}(\Omega_2^c) = 0$$

(R1) si $(\Omega_i)_{1\leq i\leq n}$ est une suite d'évent de proba 1 alors $\mathbb{P}\left(\bigcap_{i\in\mathbb{N}^*} \Omega_i\right) = 1$.

(R2) $\bar{X}_{\hat{I}} = \frac{1}{\mathrm{Card}(\hat{I})} \sum_{i\in\hat{I}} X_i$.

$$\sum_{i\in\hat{I}} X_i = \sum_{i=1}^{n} X_i \mathbb{1}_{\{i\in\hat{I}\}}$$

$$\mathrm{card}(\hat{I}) = \sum_{i=1}^{n} \mathbb{1}_{\{i\in\hat{I}\}}$$

Q2 LFGN $Y_i = \mathbb{1}_{\{X_i \neq 0\}} \sim \mathrm{Bern}(\mathbb{P}_{\alpha,\beta}(X_i \neq 0))$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{\mathbb{P}_\beta\ p.s} \mathbb{E}_{\alpha,\beta}(Y_1).$$

$$\mathbb{E}_{\alpha,\beta}(Y_1) = \mathbb{E}_{\alpha,\beta}(\mathbb{1}_{\{X_1\neq 0\}}) = \mathbb{P}_{\alpha,\beta}(X_1 \neq 0).$$

TD 3

Applicat° du TCL

Ex 1 : Cumul d'erreurs

Prgm calcul use J chiff. significatifs ; arrondi à $\frac{1}{2}\cdot 10^{-J}$ près. Sppo effectue 10^6 op succss. Errs indep. Err $\sim \mathrm{Unif}([a,b])$ 4 $a = -\frac{1}{2} 10^{-J}$, $b = \frac{1}{2}10^{-J}$ & err ss résult final est la somme des errs commises à chq opé.

1) Evaluer proba pr err^e finale soit $\leq \frac{1}{2} \cdot 10^{-J+3}$ en v^{l} abs

$U_i \sim \mathrm{Unif}\left(\left[-\frac{1}{2}10^{-J}, \frac{1}{2}10^{-J}\right]\right)$ l'err^e de la i° opérat°.

L'erreur finale est $U = \sum_{i=1}^{n} U_i$.

$$\mathbb{P}\left(|U| \leq \frac{1}{2} 10^{-J+3}\right)$$

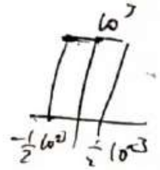
TLC $\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} \xrightarrow[n\to\infty]{loi} Z \sim \mathcal{N}(0,1)$ car les $(U_i)_{i\in\mathbb{N}^*}$ iid et de carrés intégrables car st bornées.

• $\mathbb{E}(U_1) = 0 = \int_{-1/2\cdot 10^{-J}}^{1/2\cdot 10^{-J}} x\, dx$

• $\mathrm{Var}(U_1) = \mathbb{E}(U_1^2) - \mathbb{E}^2(U_1) = \mathbb{E}(U_1^2)$

$= \int x^2\, 10^{J}\, \mathbb{1}_{[-\frac{1}{2}10^{-J}, \frac{1}{2}10^{-J}]}(x)\, dx$

$f(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}$: densité $= 10^{J}$



R1 $\quad \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n\,\mathbb{E}(X_1)}{\sqrt{n\,Var(X_1)}}$

$$= \sqrt{n}\,\frac{\frac{S_n}{n} - \mathbb{E}(X_1)}{\sqrt{Var(X_1)}} = \sqrt{n}\,\frac{\overline{X}_n - \mathbb{E}(X_1)}{\sqrt{Var(X_1)}}$$

car $Var(S_n) = Var\left(\sum_{i=1}^{n} X_i\right) \overset{\text{indep}}{=} \sum_{i=1}^{n} Var(X_i) = n\,Var(X_1)$

$$Var(U_1) = \mathbb{E}(U_1^2) = \int x^2\, 10^J\, \mathbb{1}_{[-\frac{1}{2}10^{-J}, \frac{1}{2}10^{-J}]}(x)\,dx$$

$$= 10^J \int_{-\frac{1}{2}10^{-J}}^{\frac{1}{2}10^{-J}} x^2\,dx = 10^J \left[\frac{x^3}{3}\right]_{-\frac{1}{2}10^{-J}}^{\frac{1}{2}10^{-J}}$$

$$= 10^J \left[\frac{1}{2^3\times 3}10^{-3J} + \frac{1}{2^3\times 3}10^{-3J}\right] = \frac{1}{12}10^{-2J}$$

p le TCL

$$\frac{S_n}{\sqrt{\frac{n}{12}}\,10^{-J}} \xrightarrow[n\to\infty]{\text{loi}} Z \sim \mathcal{N}(0,1)$$

$$\mathbb{P}\left(|S_n| \le \frac{1}{2}10^{-J+3}\right) = \mathbb{P}\left(\frac{|S_n|}{\sqrt{\frac{n}{12}}\,10^{-J}} \le \underbrace{\frac{\frac{1}{2}10^{-J+3}}{\sqrt{\frac{n}{12}}\,10^{-J}}}_{t}\right)$$

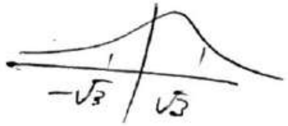
$\simeq \mathbb{P}(|Z| \le t)$ pr $n = 10^6$.

$$\mathbb{P}\left(|S_n| \le \frac{1}{2}10^{-J+3}\right) \simeq \mathbb{P}(|Z| \le \sqrt{3})$$

$$= \Phi(\sqrt{3}) - \Phi(-\sqrt{3})$$

$$= 2\Phi(\sqrt{3}) - 1$$

$$= 2\times 0{,}9582 - 1 \approx 92\ \%$$



(comparer Hoeffding)

2) On parle d'arrondi avec TCL de Berry-Esseen pr majorer l'erreur commise ds approx faite. (B-E)

$$\left|\mathbb{P}\left(|S_{10^6}| \le \frac{1}{2}10^{-J+3}\right) - (2\Phi(\sqrt{3}) - 1)\right| \le$$

$$\left|\mathbb{P}(|S^*_{10^6}| \le \sqrt{3}) - \mathbb{P}(|Z| \le \sqrt{3})\right|$$

$$\le \Big|\mathbb{P}(S^*_{10^6} \le \sqrt{3}) - \mathbb{P}(Z \le \sqrt{3}) - \left(\mathbb{P}(S^*_{10^6} \le -\sqrt{3}) - \mathbb{P}(Z \le -\sqrt{3})\right)\Big|$$

$\le |\cdot| + |\cdot|$ où $S^*_{10^6} = \frac{S_{10^6}\sqrt{12}}{10^{3-J}}$

$$\text{(BE)} \leq 2 \sup_{x\in\mathbb{R}} |\mathbb{P}(S^*_{10^6} \leq x) - \mathbb{P}(Z \leq x)|$$

$$\leq 2\times \frac{1}{2} \frac{\rho^3}{c^3\sqrt{10^6}} \leq 10^{-3} \frac{\rho^3}{\sigma^3}$$

$$\rho^3 = \mathbb{E}\left[|U_1 - \mathbb{E}(U_1)|^3\right] = \mathbb{E}\left[|U_1|^3\right]$$

$$\sigma^3 = (\sigma^2)^{\frac{3}{2}} = \left(\frac{10^{-2J}}{12}\right)^{\frac{3}{2}}, \text{ pose } a = 10^{-2J}.$$

$$= \left(\frac{1}{3^{3/2}} a^3\right)$$

$$\rho^3 = \frac{a^3}{4} = \frac{10^{-3J}}{2^5}, \quad \frac{\rho^3}{\sigma^3} = \frac{3^{3/2}}{4}.$$

$$\rho^3 = \mathbb{E}(|U_1|^3) = \int_{-a}^{a} |x|^3 \frac{1}{2a}\,dx$$

$$= 2\int_0^a |x|^3 \frac{1}{2a}\,dx = 2\int_0^a x^3 \frac{1}{2a}\,dx$$

$$= \frac{1}{a}\left[\frac{x^4}{4}\right]_0^a = \frac{a^3}{4}$$

$$\leq \frac{3^{3/2}}{4} 10^{-3} = 0{,}026 = 2{,}6\,\%$$

Ex2 Contrôle qualité.

Une machine : découpe pièces → lg^r l_0 : 20mm. En tps normal, pièces ont lg^r moyenne 20 mm, de écart-type σ_0 de 1mm. Si lot 120 pièces, lg^r moy 20,9 mm & somme des carrés des lng^rs est 52 692,9.

1) Si on admet σ_0 n'a pas varié, pensez-vs que la machine est déréglée (construire test niv 5%) ?
Q^lle est la proba de conclure que la machine est déréglée si la moy in le lot des 120 pièces est en réalité 21 mm?

Modélisaº : X tq $\mathbb{E}(X) = l_0$, $\text{Var}(X) = 1 = \sigma_0^2$.

$n = 120$ pièces. $X_1, \dots, X_n$ sous $\mathbb{P}_l$, $\mathbb{E}(X_i) = l$, $l \in \mathbb{R}^+$ inconnu.

Test : • Choix des hypo :

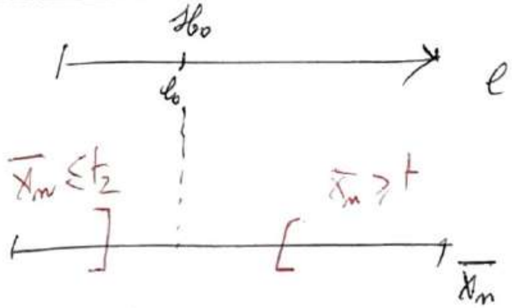
H_0 : "la machine est bien réglée contre H_1 : "machine est déréglée" :
$\underline{l = l_0}$ $\qquad l \neq l_0$

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$: estimat^R sans biais de l.

↳ estimat^R fort^mt consistant de l : $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{\mathbb{P}\,p.s} \mathbb{E}_l(X_i) = l$

p. CFGN

$\sigma_0 = 1$



$R_{5\%} = \{\overline{X}_n \le t_2\} \cup \{\overline{X}_n \ge t_2\}$ t_1 & t_2 tq

$\mathbb{P}_{\ell_0}(R_{5\%}) \le 0{,}05 \Rightarrow$ on choisit t_1 & t_2,

$\mathbb{P}_{\ell_0}(\overline{X}_n \le t_2) \le 0{,}025$ & $\mathbb{P}_{\ell_0}(\overline{X}_n \ge t_1) \le 0{,}025$

M1 TCL $\sqrt{n}\,\dfrac{\overline{X}_n - \ell_0}{\sigma_0} \xrightarrow[n\to\infty]{\mathbb{P}_{\ell_0}\text{ loi}} Z \sim \mathcal{N}(0,1)$

La loi de $\sqrt{120}\,\dfrac{\overline{X}_{120} - \ell_0}{\sigma_0}$ sous $\mathbb{P}_{\ell_0}$ est approchée

p. la loi $\mathcal{N}(0,1)$

op $\mathbb{P}_{\ell_0}\left(\sqrt{120}\,\dfrac{|\overline{X}_{120} - \ell_0|}{\sigma_0} \ge \varepsilon\right) \simeq \mathbb{P}(|Z| \ge \varepsilon) = 0{,}05$

pour $\varepsilon = 1{,}96 = \Phi^{-1}(0{,}975)$

$\mathbb{P}(Z \ge \varepsilon) = 0{,}025 \iff \mathbb{P}(Z \le \varepsilon) = 0{,}975$

$= 2\mathbb{P}(Z \ge \varepsilon)$

⚠ $\mathbb{P}(|Z| \ge \varepsilon) = \mathbb{P}(Z \ge \varepsilon) + \mathbb{P}(Z \le -\varepsilon)$

$$R_{5\%} = \left\{\sqrt{120}\,\frac{|\overline{X}_{120} - \ell_0|}{\sigma_0} \ge 1{,}96\right\}$$

$$= \left\{\overline{X}_{120} \le \ell_0 - 1{,}96\sqrt{\frac{\sigma_0^2}{120}}\right\} \cup \left\{\overline{X}_{120} \ge \ell_0 + 1{,}96\sqrt{\frac{\sigma_0^2}{120}}\right\}.$$

$R_{5\%} = \{\overline{X}_{120} \le 19{,}82\} \cup \{\overline{X}_{120} \ge 20{,}18\}$ $\mathbb{P}^2$ approchée

Cl On observe $\overline{x}_{120}(\omega) = 20{,}9$ mm.

$\omega \in R_{5\%}$: on conclut que machine est déréglée.

M2 Tchebitcher

$\mathbb{P}_{\ell_0}\left(\sqrt{120}\,\dfrac{|\overline{X}_{120} - \ell_0|}{\sigma_0} \ge \varepsilon\right)$ pr $\varepsilon \ge \dfrac{1}{\sqrt{0{,}05}} = 4{,}47$

$\le \dfrac{1}{\varepsilon^2} \le 0{,}05.$

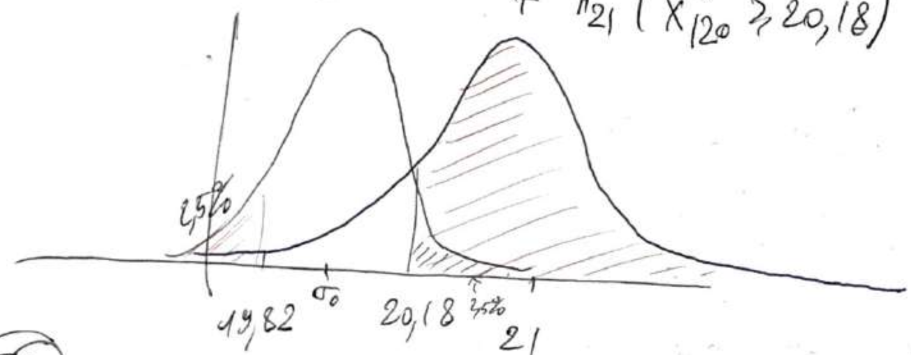
$\mathbb{P}_{\ell_0}(|\overline{X}_{120} - \ell_0| \ge \varepsilon) \le \dfrac{\mathrm{Var}_{\ell_0}(X_1)}{120\,\varepsilon^2} = \dfrac{\sigma_0^2}{120\,\varepsilon^2}$

$R_{5\%}^{\text{Tchebi}} = \{\overline{X}_{120} \le 19{,}59\} \cup \{\overline{X}_{120} \ge 20{,}41\}$ $\mathbb{P}^K$ + petite

$\Rightarrow \hat{I}_{100} = [\overline{X}_{...} - 1,66 \sqrt{\ldots}$

Q^{le} est la proba de conclure que la machine est déréglée si la moyenne sur le lot des 120 pièces est en réalité 21 mm ?

$\ell = 21$, $\mathbb{P}_{21}(\mathcal{R}_{5\%}) = \mathbb{P}_{21}(\overline{X}_{120} \leq 19,82) + \mathbb{P}_{21}(\overline{X}_{120} \geq 20,18)$



(TCL) $\sqrt{n}\,\dfrac{\overline{X}_n - 21}{\sigma_0} \xrightarrow[n \to \infty]{\mathbb{P}_{21}\ \text{loi}} Z \sim \mathcal{N}(0,1)$

La loi de $\sqrt{120}\,\dfrac{\overline{X}_{120} - 21}{\sigma_0}$ est approchée p la loi $\mathcal{N}(0,1)$.

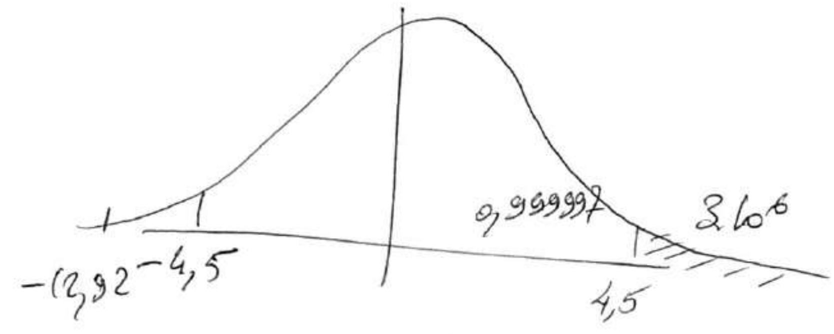
$\mathbb{P}_{21}(\overline{X}_{120} \leq 19,82) + \mathbb{P}_{21}(\overline{X}_{120} \geq 20,18) =$

$= \mathbb{P}_{21}\left(\sqrt{120}\,\dfrac{\overline{X}_{120} - 21}{\sigma_0} \leq \sqrt{120}\,\dfrac{19,82 - 21}{\sigma_0}\right) + \mathbb{P}_{21}\left(\sqrt{120}\,\dfrac{\overline{X}_{120} - 21}{\sigma} \geq \sqrt{120}\,\dfrac{20,18 - 21}{\sigma_0}\right)$

$= \mathbb{P}(Z \leq -12,92) + \mathbb{P}(Z \geq -8,98) \simeq 1$

$\mathbb{P}(Z \leq -12,92) \simeq 1,73 \cdot 10^{-38}$

$\mathbb{P}(Z \geq -8,98) \simeq 1$



2) Sans mettre d'hypothèses sur l'écart-type, est-ce que cela modifie les choses ?

Sans hyp : $\sigma_0 = 1 \Rightarrow \sigma$ inconnu.

TLC + autonormalisat°: $V_{120} = \dfrac{1}{120} \sum_{i=1}^{120} (X_i - \overline{X}_{120})^2$

$\sqrt{120}\,\dfrac{\overline{X}_{120} - \ell}{\sqrt{V_{120}}} \xrightarrow[n \to \infty]{\mathbb{P}_\ell\ \text{loi}} Z$

$\mathcal{R}_{5\%} = \{\overline{X}_{120} \leq \ell_0 - 1,96\sqrt{\frac{V_{120}}{120}}\} \cup \{\overline{X}_{120} \leq \ell_0 + 1,96\sqrt{\frac{V_{120}}{120}}\}$

Appli Num :

$V_{120}(\omega) = 2,2975$, $\omega \in \mathcal{R}_{5\%}$

On rejette encore.

49

$\sigma_0 = 1$

Echantillon gaussien

Ex 3 : Consommat de carburant (suite)

Données ex 12 de fiche 1 : $\sum_{i=1}^{100} x_i = 700,51$ et $\sum_{i=1}^{100} x_i^2 = 5043,99$

On spps $X \sim \text{Normale}(m, \sigma^2)$ inconnues.

1) Donner un intervalle de confiance pr m de niveau de confiance 90 %.

$x_1, \dots, x_{100}$ (ex 12, fiche 1).

Réalisats de $X_1, \dots, X_{100}$ iid de loi $\mathcal{N}(m, \sigma^2)$.

$$\sqrt{n}\,\frac{\bar{X}_n - m}{\sigma} \sim \mathcal{N}(0,1)$$

$$\bar{X}_n = \frac{1}{n}\sum X_i \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \quad ; \quad \frac{\bar{X}_n - m}{\sqrt{\sigma^2/n}}$$

$$\mathbb{E}(\bar{X}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n} \times n \times m = m.$$

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

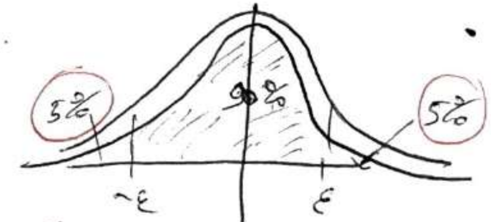
où $V_n^* = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

$$\sqrt{n}\,\frac{\bar{X}_n - m}{\sqrt{V_n^*}} \xrightarrow[n \to \infty]{\text{loi}} Z \sim \mathcal{N}(0,1).$$

(On a des quantiles de la loi de Student(n) vers les quantiles de la loi normale $n \to \infty$).

d = 100 ds la table.

n − 1 = 99.



$$\mathbb{P}(\text{Student}(99) \le t) = \alpha$$

$$T_n = \sqrt{n}\,\frac{\bar{X}_n - m}{\sqrt{V_n^*}}$$

$$\mathbb{P}(|T_n| \le \varepsilon) \geq 0,9$$

n = 100, $T_n \sim \text{Student}(99)$

4 la taille, on se contente des quantiles pr n = 100,

$\varepsilon = 1,66$.

$$\to \mathbb{P}_{m,\sigma^2}\left(\sqrt{100}\,\frac{|\bar{X}_{100} - m|}{\sqrt{V_{100}^*}} \le 1,66\right) = 0,90$$

$$\mathbb{P}_{m,\sigma^2}\left(m \in \underbrace{\left[\bar{X}_{100} - 1,66\sqrt{\frac{V_{100}^*}{100}} ; \bar{X}_{100} + 1,66\sqrt{\frac{V_{100}^*}{100}}\right]}_{\hat{I}_{100}}\right) = 0,9$$

$$\Rightarrow \hat{I}_{100} = \left[\overline{X}_{100} - 1{,}66\sqrt{\frac{V_{100}^{*}}{100}} \,;\, \overline{X}_{100} + 1{,}66\sqrt{\frac{V_{100}^{*}}{100}} \right]$$

$$V_n = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = \frac{1}{n}\sum X_i^2 - \overline{X}_n^2$$

$$V_n^{*} = \frac{n}{n-1}V_n = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}_n^2$$

$$V_{100}^{*}(\omega) = \frac{1}{99}\sum_{i=1}^{100} x_i^2 - \frac{100}{99}\left(\frac{1}{100}\sum_{i=1}^{n} x_i\right)^2$$

$$= \frac{1}{99}\times 5043{,}99 - \frac{100}{99}\left(7{,}0051\right)^2 = 1{,}38.$$

$\overline{X}_{100}(\omega) = 7{,}0051.$

$\hat{I}_{100}(\omega) = \left[6{,}81 \,;\, 7{,}20\right]$

2) Donner un IdConfiance pr σ^2 de niveau de confiance 90%

- 1 intervalle unilatéral : $[0, -] = \hat{I}$.

  $\mathbb{P}_{m,\sigma^2}\left(\sigma^2 \in \hat{I}\right) \geq 0{,}9$

- 1 intervalle bilatéral : $\hat{J}$.

  $\mathbb{P}_{m,\sigma^2}\left(\sigma^2 \in \hat{J}\right) \geq 0{,}9$

, si $U_d \sim \chi^2(d)$, loi asymptotiq de U_d

$$\boxed{U_d \overset{loi}{=} \sum_{i=1}^{d} Z_i^2} \quad \text{où } Z_i \sim \mathcal{N}(0,1)$$

LFGN

$$\frac{1}{d}\sum_{i=1}^{d} Z_i^2 \xrightarrow[d\to\infty]{p.s} \mathbb{E}(Z_1^2) = 1.$$

$$\mathrm{Var}(Z_1) = \mathbb{E}(Z_1^2) - \mathbb{E}(Z_1)^2 = \mathbb{E}(Z_1^2).$$

TCL

$$\frac{\sum_{i=1}^{d} Z_i^2 - d}{\sqrt{2d}} \xrightarrow[d\to\infty]{loi} Z \sim \mathcal{N}(0,1).$$

$$\mathrm{Var}(Z_1^2) = \mathbb{E}(Z_1^4) - \mathbb{E}(Z_1^2) = 2.$$

$$\mathrm{Var}(U_d) = \sum_{i=1}^{d} \mathrm{Var}(Z_1^2) = 2d.$$

$\dfrac{U_d - d}{\sqrt{2d}} \xrightarrow[d\to\infty]{loi} Z$, la loi de U_d pt ê approchée si d est grand par $\mathcal{N}(d, 2d)$.

$\sqrt{2U_d} - \sqrt{2d-1} \xrightarrow[n\to\infty]{loi} Z$ se dmq 4 Slutsky.

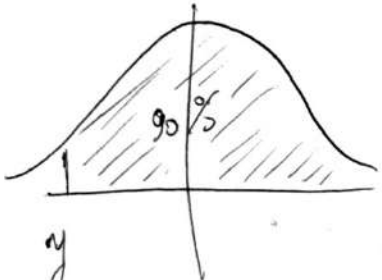
(Permet d'obtenir les quantiles de la loi du $\chi^2(d)$ pr des grdes valeurs de d). @ trouver x tq $\mathbb{P}(U_d \geq x) = 0{,}9$

• $\mathbb{P}(U_d \geq x) = \mathbb{P}(\sqrt{2U_d} - \sqrt{d-1} \geq \sqrt{\ }$

si d gd $\simeq \mathbb{P}(Z \geq \sqrt{2x} - \sqrt{2d-1}) = 0{,}9$

pour $\sqrt{2x} - \sqrt{2d-1} = -1{,}28$.

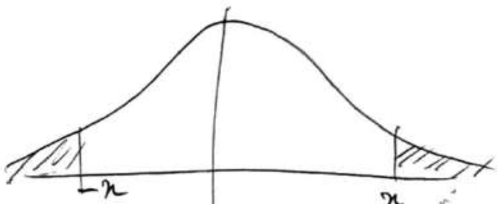
$$x = \frac{(\sqrt{2d-1} - 1{,}28)^2}{2}$$



$\mathbb{P}(Z \geq y) = 0{,}9$

$\mathbb{P}(-Z \geq y) = 0{,}9$

$\mathbb{P}(Z \leq -y) = 0{,}9$.



$\Phi(-x)$
$\parallel$
$1 - \Phi(x)$

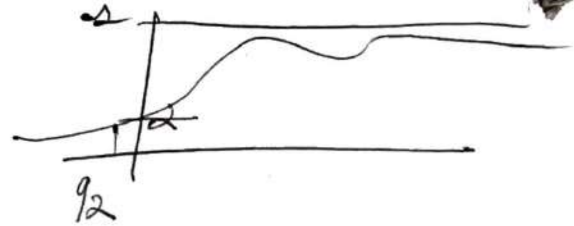
Trouver le quantile d'une loi du $\chi^2(d)$ qd $d > 30$.

(5) Le quantile d'ordre α de la loi de X est le réel x_α tq $\mathbb{P}(X \leq x_\alpha) = \alpha$.

Pas de pb de définit° si X est à densité sinon $x_\alpha = F^{-1}(\alpha) = \inf\{x \in \mathbb{R},\ F(x) \geq \alpha\}$ où F est FdR de X.

$F(q_\alpha) = \alpha$

$\mathbb{P}(X \leq q_\alpha) = \alpha$



ici $X_d \sim \chi^2(d)$

$$\sqrt{2X_d} - \sqrt{2d-1} \xrightarrow[d \to \infty]{\text{loi}} Z$$

$\mathbb{P}(Z \leq 1{,}28) \simeq 0{,}9$

$\simeq \mathbb{P}(\sqrt{2X_d} - \sqrt{2d-1} \leq 1{,}28) \simeq 0{,}90$

$$\boxed{\mathbb{P}\Big(X_d \leq \underbrace{\left(\frac{(1{,}28 + \sqrt{2d-1})^2}{2}\right)}_{\text{quantile approché de la loi } \chi^2(d) \text{ d'ordre } 0{,}9}\Big)} \simeq 0{,}9$$

Quantile d'ordre 0,1 de $\chi^2(d)$

$\mathbb{P}(Z \leq -1{,}28) = 0{,}1$.

(52) $\Rightarrow \mathbb{P}($

$$\mathbb{P}(Z \leq -1{,}28) = 0{,}1$$

$$\Rightarrow \mathbb{P}\left(X_d \leq \left(\frac{\sqrt{2d-1} - 1{,}28}{2}\right)^2\right) \simeq 0{,}1$$

$$\Leftrightarrow \mathbb{P}\left(X_d \geq \left(\frac{\sqrt{2d-1} - 1{,}28}{2}\right)^2\right) \simeq 0{,}9$$

$$X_d \to \frac{nV_n}{\sigma^2},$$

$$\frac{nV_n}{\sigma^2} \geq \underbrace{\left(\frac{\sqrt{2(n-1)^2-1} - 1{,}28}{2}\right)^2}_{q_{n-1}(0{,}1)}$$

$$\Leftrightarrow \sigma^2 \leq \frac{nV_n}{q_{n-1}(0{,}1)}$$

$$\mathbb{P}_{\sigma^2}\left(\sigma^2 \leq \frac{nV_n}{q_{n-1}(0{,}1)}\right) \geq 0{,}9$$

FF $$\boxed{q_d(\alpha) \simeq \left(\frac{\sqrt{2d-1} + \Phi^{-1}(\alpha)}{2}\right)^2}$$

$$a \leq \frac{nV_n}{\sigma^2} \leq b$$ bilateral

$$\frac{nV_n}{b} \leq \sigma^2 \leq \frac{nV_n}{a}$$

$$\sigma^2 \in \left[0, \frac{nV_n}{q_{n-1}(0{,}1)}\right]$$ Unilatéral (precisión en DS).

TD 4 Ex 1 : M Monte-Carlo pr le calcul d'intégrales

$f : [0,1] \to \mathbb{R}$ cont ; vl^{r} approchée de m

$$m := \int_0^1 f(x)\,dx \quad , (U_i)_i \text{ va ind}$$

$U_i \sim Unif([0,1])$.

$$M_{2n} := \frac{1}{2n}\sum_{i=1}^{2n} f(U_i) \quad \& \quad \tilde{M}_{2n} := \frac{1}{2n}\sum_{i=1}^{n}\left(f(U_i) + f(1-U_i)\right)$$

1) Expliquer pk $X_1 := f(U_1)$ et $Y_1 = f(U_1) + f(1-U_1)$ st intégrables & exprimer l^{r} espérance à l'aide de l'intég m.

$X_1 := f(U_1)$, X_1 est intégrable car f est bornée sur $[0,1]$. (car cont sur un segment).

Donc $\exists M \in \mathbb{R}_+$ tq $|X_1| = |f(U_1)| \leq M \Rightarrow \mathbb{E}(|X_1|) < \infty$.

De m̂, $|Y_1| = |f(U_1) + f(1-U_1)| \leq |f(U_1)| + |f(1-U_1)| \leq 2M$.

Dc Y_1 est une r.a. bornée dc intég.

- $\mathbb{E}(X_1) = \mathbb{E}(f(U_1))$.

Rq si Z a pr densité h et si $\varphi(Z)$ est intégrable

$$\Longrightarrow \mathbb{E}(\varphi(Z)) = \int_{\mathbb{R}} \varphi(z)\, h(z)\, dz$$

$$\mathbb{E}(f(U_1)) = \int_{\mathbb{R}} f(u)\, \underbrace{f_{U_1}(u)}_{\text{densité de } U_1}\, du \quad , \boxed{f_{U_1}(u) = \mathbb{1}_{[0,1]}(u)}$$

$$\Longrightarrow \mathbb{E}[f(U_1)] = \int_{\mathbb{R}} f(u)\, \mathbb{1}_{[0,1]}(u)\, du = \int_0^1 f(u)\,du = m$$

- $\mathbb{E}(Y_1) = \mathbb{E}[f(U_1) + f(1-U_1)]$

$$= \int_{\mathbb{R}} (f(u) + f(1-u))\, \mathbb{1}_{[0,1]}(u)\, du$$

$$= \int_0^1 f(u)\,du + \int_0^1 f(1-u)\,du = 2m.$$

Rq $\mathbb{E}(f(U_1)) = \mathbb{E}(f(1-U_1))$ car $1-U_1 \sim Unif([0,1])$.

2) En vs appuyant sur Th cours, mq
$(M_{2n})_{n\in\mathbb{N}^*}$ & $(\tilde{M}_{2n})_{n\in\mathbb{N}^*}$ CV ps
vers m qd $n\to\infty$.

$$M_{2n}=\frac{1}{2n}\sum_{i=1}^{2n}X_i \quad \text{par } \underline{\text{LFGN}}, \ (X_i)_{i\in\mathbb{N}} \text{ iid de carré intég.}$$

$$M_{2n}\xrightarrow[n\to\infty]{p.s}\mathbb{E}(X_1)=m$$

$$\tilde{M}_{2n}=\frac{1}{2n}\sum_{i=1}^{n}Y_i \quad \text{p LFGN, } (Y_i) \text{ iid carré intég.}$$

$$\frac{1}{n}\sum_{i=1}^{n}Y_i\xrightarrow{p.s}\mathbb{E}(Y_1)=2m$$

$$\Rightarrow \tilde{M}_{2n}\xrightarrow[n\to\infty]{p.s}m$$

3) Expliquer pq X_1 & Y_1 st de carré intégrables & exprimer leur variance à l'aide f $\int$.

$$|X_1|^2\leq M^2 \quad , \quad M=\sup_{x\in[0,1]}|f(x)|$$

$\Rightarrow X_1$ est de carré intégrable idem $|Y_1|^2\leq 4M^2$.
$\Rightarrow Y_1$ est de carré intégrable

$$Var(X_1)=\mathbb{E}(f^2(U_1))-m^2=\int_0^1 f^2(x)\,dx-m^2$$

$$Var(Y_1)=\mathbb{E}(Y_1^2)-\mathbb{E}^2(Y_1)=\mathbb{E}\left[(f(U_1)+f(1-U_1))^2\right]$$

$$=\mathbb{E}\left[f^2(U_1)+f^2(1-U_1)+2f(U_1)f(1-U_1)\right]-\mathbb{E}^2(Y_1)$$

$$=2\int_0^1 f^2(x)\,dx+2\int_0^1 f(x)f(1-x)\,dx-2m^2$$

$$+\cancel{\int_0^1 f(x).f(1-x)\,dx}\quad \cancel{2m^2}$$

4) On suppose $f\nearrow$ sur $[0,1]$.

a) Mq $\forall (x,y)\in[0,1]^2$,

$$(f(x)-f(y))(f(1-x)-f(1-y))\leq 0.$$

$f(x)-f(y)$ est du signe de $x-y$.

$x\mapsto f(1-x)$ est décroissante.

Dc $f(1-x)-f(1-y)$ est du signe opposé de $x-y$.

$\Rightarrow f(x)-f(y)$ et $f(1-x)-f(1-y)$ de signes opposés.

b) est $\mathbb{E}(f(U_1)f(1-U_1)) \leq \mathbb{E}^2(U_1)$

D'après a) $(f(x)-f(y))(f(1-x)-f(1-y)) \leq 0 \quad \forall (x,y) \in [0,1]^2$,

$f(x)f(1-x) + f(y)f(1-y) \leq f(y)f(1-x) + f(x)f(1-y)$

$y = 1-x$. $\boxed{M1}$

$2f(x)f(1-x) \leq f^2(1-x) + f^2(x)$

$\Rightarrow 2f(U_1)f(1-U_1) \leq f^2(1-U_1) + f^2(U_1)$

& $\mathbb{E}(f(U_1)f(1-U_1)) \leq \mathbb{E}[f^2(U_1)] + \mathbb{E}[f^2(1-U_1)]$

$= 2\mathbb{E}[f^2(U_1)]$.

car U_1 et $1-U_1$ ont la m loi.

$\boxed{M2}$ $x = U_1$, $y = U_2$. (et on utilisera l'indépendance de U_1 & U_2)

$\mathbb{E}(-) \leq \mathbb{E}^2(f(U_1))$ plus fort.

c) Comparer les variances de M_{2n} et $\tilde{M}_{2n}$.

$Var(M_{2n}) = Var\left(\frac{1}{2n}\sum_{i=1}^{2n} X_i\right) = \frac{1}{4n^2} Var\left(\sum_{i=1}^{2n} X_i\right)$

$= \frac{1}{4n^2}\sum_{i=1}^{2n} Var(X_i)$ car les (X_i) st indep.

$= \frac{1}{2n} Var(X_1) = \frac{1}{2n}\left[\int_0^1 f^2(x)\,dx - m^2\right]$

$Var(\tilde{M}_{2n}) = Var\left(\frac{1}{2n}\sum_{i=1}^{n} f(U_i) + f(1-U_i)\right)$

$= \frac{1}{4n^2} Var\left(\sum_{i=1}^{n} Y_i\right) = \frac{1}{4n^2}\sum_{i=1}^{n} Var(Y_i)$

$= \frac{1}{4n} Var(Y_1) = \frac{1}{4n} Var(f(U_1) + f(1-U_1))$

$= \frac{1}{4n}\left(Var(f(U_1)) + Var(f(1-U_1)) + 2cov(f(U_1); f(1-U_1))\right)$

$= \frac{1}{4n}\left[2 Var f(U_1) + 2\left(\mathbb{E}(f(U_1)f(1-U_1)) - \mathbb{E}(f(U_1))\right)^2\right]$

$\leq \frac{1}{2n} Var(f(U_1)) = \frac{1}{n} Var(X_1)$.

$cov(f(U_1), f(1-U_1)) \leq 0$.

$= \mathbb{E}[f(U_1)f(1-U_1)] - \mathbb{E}[f(U_1)]\mathbb{E}[f(1-U_1)]$

$= \mathbb{E}[f(U_1)f(1-U_1)] - m^2$

b) $\mathbb{E}[f(U_1)f(1-U_1)] \leq m^2 = [\mathbb{E}(f(U_1))]^2$

$\Leftrightarrow \int_0^1 f(x)f(1-x)\,dx \leq m^2$.

$cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

Comparaison des risques quadratiqs :

$$Var(\tilde{M}_{2n}) \leq Var(M_{2n}).$$

$$R(\tilde{M}_{2n}) = \mathbb{E}\left[(\tilde{M}_{2n} - m)^2\right] = Var(\tilde{M}_{2n})$$

car $m = \mathbb{E}[\tilde{M}_{2n}]$

ccl $\tilde{M}_{2n}$ est un estimateur sans biais de m meilleur que M_{2n} au sens du Rq quadratiq. (lorsq f est $\nearrow$)

5) En utilisant TLC, propose un IdC pr m de niveau 95 % (en négligeant l'erreur due à l'approx gaussienne) en supposant que l'on connaît un majorant M de $\sup_{x \in [0,1]} |f(x)|$

$$\tilde{M}_{2n} = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{2}$$

$$\sqrt{n}\, \frac{\tilde{M}_{2n} - m}{\sqrt{Var\left(\frac{Y_i}{2}\right)}} \xrightarrow[n \to \infty]{loi} Z \quad \text{où } Z \sim \mathcal{N}(0,1)$$

$$= 2\sqrt{n}\, \frac{\tilde{M}_{2n} - m}{\sqrt{Var(Y_1)}} \xrightarrow[n \to \infty]{Loi} Z$$

$\mathbb{P}(\underset{-\varepsilon}{a} \leq Z_n \leq \underset{\varepsilon}{b}) \xrightarrow[n \to \infty]{} \mathbb{P}(a \leq Z \leq b)$

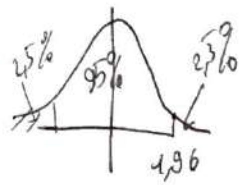
$\Leftrightarrow \mathbb{P}(-\varepsilon \leq Z_n \leq \varepsilon) \xrightarrow[n \to \infty]{} \mathbb{P}(-\varepsilon \leq Z \leq \varepsilon)$

$\Leftrightarrow \mathbb{P}(|Z_n| \leq \varepsilon) \xrightarrow[n \to \infty]{} \mathbb{P}(|Z| \leq \varepsilon)$

$$Var(Y_1) \leq 2\, Var(f(U_1))$$

$$= 2\left(\mathbb{E}[f^2(U_1)] - [\mathbb{E}(f(U_1))]^2\right)$$

$$\leq 2\, \mathbb{E}[f^2(U_1)] \leq 2M^2.$$



$$\varepsilon = \Phi^{-1}(0{,}975) = 1{,}96.$$

$$\mathbb{P}(Z \leq 1{,}96) = 97{,}5\ \%.$$

$$V(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) \leq \mathbb{E}(X^2)$$

$$\{|Z_n| \leq 1{,}96\} = \left\{2\sqrt{n}\,\frac{|\tilde{M}_{2n} - m|}{\sqrt{\mathrm{Var}(Y_1)}} \leq 1{,}96\right\}$$

$$= \left\{m \in \left[\tilde{M}_{2n} - 1{,}96\,\frac{\sqrt{\mathrm{Var}(Y_1)}}{2\sqrt{n}}\,;\,\tilde{M}_{2n} + 1{,}96\,\frac{\sqrt{\mathrm{Var}(Y_1)}}{2\sqrt{n}}\right[\right\}$$

$$\subset \left\{m \in \left[\tilde{M}_{2n} - 1{,}96\,\frac{\sqrt{2}\,M}{2\sqrt{n}}\,;\,\tilde{M}_{2n} + 1{,}96\,\frac{\sqrt{2}\,M}{2\sqrt{n}}\right]\right\}$$

$$\liminf_{n\to\infty} \mathbb{P}\left(m \in \left[\tilde{M}_{2n} - 1{,}96\,\frac{M}{\sqrt{2n}}\,;\,\tilde{M}_{2n} + 1{,}96\,\frac{M}{\sqrt{2n}}\right]\right) \overset{\geq 0{,}95}{\gtrsim} 0{,}95$$

En négligeant l'erreur d'approximat° gaussienne.

C4 Le Forró

I Intro

D Le forró : musiq du Nord & Brésil à base d'accordéon zabumba & triangle. À Lille, le Forró existe depuis 2008 par Cacau Moutinho. @ Aqui de Norte.

II/ Ppté

TH 1) Si on écoute du forró => on ne pt s'empêcher de danser.
2) Le forró rend heureux,

Preuve D Devinette instrument de la prof.

Csqu 30 avril / 8 mai / 28 mai
12 bars - 12 (9)
Aqui de Norte au Relax.

Ex 2 Simulation de la loi de Poisson

$(U_i)_{1\leq i\leq n}$ iid $\mathrm{Unif}([0,1])$.

$$M_n = \prod_{i=1}^{n} U_i$$

1) Mq $(M_n)_{n\in\mathbb{N}^*}$ cv ps vers 0.

$\boxed{M1}$ Passer au log.

$Z_n = \log M_n = \sum_{i=1}^{n} \log U_i$,

si $\log U_1$ est intégrable $\Rightarrow \frac{1}{n}\sum_{i=1}^{n}\log(U_i) \xrightarrow[n\to\infty]{ps} \mathbb{E}(\log(U_1))$

est iid car $(\log U_i)_{i\in\mathbb{N}^*}$.

$$\mathbb{E}(\log(U_1)) = \int_{\mathbb{R}} \log x \, f(x)\, dx$$

$$= \int_{\mathbb{R}} \log x \, \mathbb{1}_{[0,1]}(x)\, dx = \int_0^1 \log x\, dx$$

$$= [x\log x - x]_0^1 = -1.$$

$$\frac{1}{n} Z_n \xrightarrow[n\to\infty]{p.s} -1$$

$$M_n = e^{Z_n} = e^{n(\frac{1}{n}Z_n)}$$

$$M_n \xrightarrow[n\to\infty]{ps} 0$$

$\boxed{M2}$ $\mathbb{P}(M_n \geq \varepsilon) \overset{\text{Markov}}{\leq} \frac{\mathbb{E}(M_n)}{\varepsilon}$

or $\mathbb{E}(M_n) = \mathbb{E}\left(\prod_{i=1}^{n} U_i\right) = \prod_{i=1}^{n}\mathbb{E}(U_i) = \frac{1}{2^n}$

$\Rightarrow \mathbb{P}(M_n \geq \varepsilon) \leq \frac{1}{2^n \varepsilon}$.

- $\Rightarrow \lim_{n\to\infty} \mathbb{P}(M_n \geq \varepsilon) = 0$.

M_n cv en proba vers 0.

- $\Rightarrow \sum \mathbb{P}(M_n \geq \varepsilon) < \infty$

Dc M_n cv ps vers 0.

$\boxed{M3}$ $M_n = \prod_{i=1}^{n} U_i$

- $0 \leq M_n \leq 1$ et $M_{n+1} = U_{n+1} M_n \leq M_n$ ps.
- $(M_n)_{n\in\mathbb{N}^*} \searrow$
- Pour presq tout ω : $(M_n(\omega))_{n\in\mathbb{N}^*} \searrow$ minoré

de cv vers $M(\omega) \geq 0$

Par le Th de CV dominée,

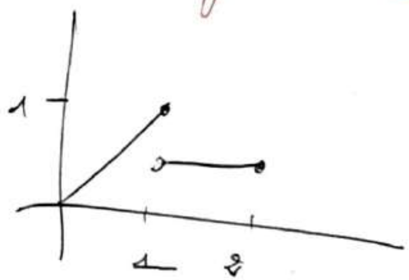
$$\lim_{n\to\infty} \mathbb{E}(M_n) = \mathbb{E}(M) = 0 \quad \text{car } \mathbb{E}(M_n) = \frac{1}{2n}.$$

$\Rightarrow \mathbb{E}(M) = 0 \Rightarrow M = 0$ p.s car M est une v.a positive.

Ex4 Plusieurs méthodes de simulat°

On vt simuler va X de densité f sur $\mathbb{R}$:

$$f(x) = x\,\mathbb{1}_{[0,1]}(x) + \frac{1}{2}\,\mathbb{1}_{]1,2]}(x)$$



$\int_{\mathbb{R}} f(x)\,dx = 1$ ☺

et $f \geq 0$ ☺.

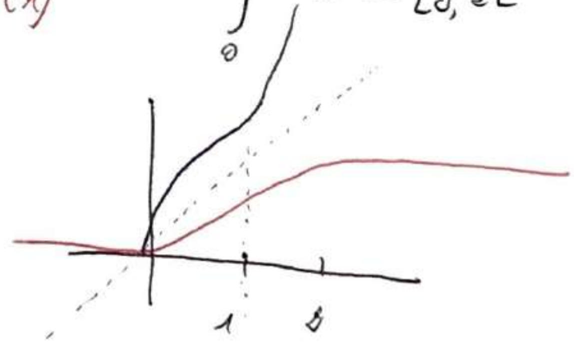
1) Expliquer comment simuler X en utilisant la $\boxed{M}$ par inversion de la f de répartit°.

$\forall t \in \mathbb{R}$, $F(t) = \int_{-\infty}^{t} f(x)\,dx$

$$= \begin{cases} 0 & \text{si } t < 0 \\ t^2/2 & \text{si } 0 \leq t < 1 \\ t/2 & \text{si } 1 \leq t < 2 \\ 1 & \text{si } t \geq 2 \end{cases}$$

→ si $t \in [1,2[$,

$$F(t) = \int_0^1 x\,\mathbb{1}_{[0,1[}(x)\,dx + \int_1^t \frac{1}{2}\,dx = \frac{1}{2} + \frac{t-1}{2} = \frac{t}{2}$$



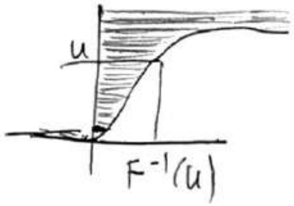
- Calcul de F^{-1}.

- si $u \in]0, \frac{1}{2}[$:
  $F(t) = u$
  $\Leftrightarrow \frac{t^2}{2} = u$
  $\Leftrightarrow t = \sqrt{2u}$

- si $u \in [\frac{1}{2}, 1[$.
  $\Leftrightarrow F(t) = u$
  $\Leftrightarrow \frac{t}{2} = u \Leftrightarrow t = 2u$

CCl $\forall u \in]0,1[\Rightarrow F^{-1}(u) = \sqrt{2u}\,\mathbb{1}_{]0,\frac{1}{2}[}(u) + 2u\,\mathbb{1}_{[\frac{1}{2},1[}(u)$

Pour simuler X de densité f, il suffit : simuler $U \sim Unif(]0,1[)$ et $X = F^{-1}(u)$

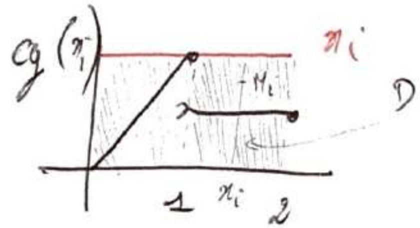


$$= \sqrt{2u}\, \mathbb{1}_{]0,\frac{1}{2}[}(u) + 2u\, \mathbb{1}_{[\frac{1}{2},1[}(u).$$

2) Expliquer comment simuler X en utilisant la $\boxed{M}$ de rejet à partir de la loi uniforme sur $[0,2]$.

▶ Majorer f par cg où g est une densité qu'on sait simuler.

$$g(x) = \frac{1}{2} \mathbb{1}_{[0,2]}(x)$$
$$cg(x) = \mathbb{1}_{[0,2]}(x)$$



$X_i \sim Unif([0,2])$ $\quad U_i \sim Unif([0,1])$

$M_i = (X_i, cg(x_i) U_i) \sim Unif(D)$.

$D = \{(x,y) \in \mathbb{R}^2 : 0 \leq y \leq cg(x)\}$.

$\sim P(|T| < 6) \geq 0.95$

Rq $cg(x_i) U_i = \mathbb{1}_{[0,2]}(x_i) U_i = U_i$.

On simule les points $(M_i)_{i \in \mathbb{N}^*}$ jusqu'à la conditiθ : $cg(x_i) U_i \leq f(x_i) \Leftrightarrow U_i \leq f(x_i)$

$T = \inf\{i \in \mathbb{N}^*, U_i \leq f(x_i)\}$ alors x_T a pr densité f.

3) Soit Y & Z va indep de loi uniforme respectivement sur $[0,1]$ & sur $[0,2]$. Quelle est la densité de $S = \max(Y,Z)$ et une autre $\boxed{M}$ de simulatiθ de X.

Réflexe : calculer f de Répartiθ

$$F_S(t) = P(S \leq t) = P(\max(Y,Z) \leq t) \quad \forall t \in \mathbb{R}$$
$$= P(\{Y \leq t\} \cap \{Z \leq t\})$$
$$= P(Y \leq t) \times P(Z \leq t) \quad \text{car } Y \perp\!\!\!\perp Z.$$

$$P(Y \leq t) = \begin{cases} 0 & \text{si } t < 0 \\ t & \text{si } t \in [0,1[\\ 1 & \text{si } t \geq 1 \end{cases} \qquad P(Z \leq t) = \begin{cases} 0 & \text{si } t < 0 \\ t/2 & \text{si } t \in [0,2[\\ 1 & \text{si } t \geq 2 \end{cases}$$

61 $$F_S(t) = \begin{cases} 0 & \text{si } t < 0 \\ t^2/2 & \text{si } t \in [0,1[\\ t/2 & \text{si } t \in [1,2[\\ 1 & \text{si } t \geq 2 \end{cases}$$

$(fdR)' =$ densité

S et X ont la m̂ loi pr simuler X
on simule X, on simule $U_1 \sim Unif([0,1])$
$U_2 \sim Unif([0,1])$

$X = \max(U_1, 2U_2)$.