M66 CM

Statistiques

C1 La démarche statistique à t → le M de Bernoulli

I / Introd°

→ bcp de données à analyser (med (→ nv° équid, F^{ie} ofie, nv° trait bon...), météo, finance)

→ Diff types de stats : → stats descriptives : visualisat° données (moy, variance, histog)

maths, probas
- → stats inférentielles : tirer de l'info sr données
- → stats décisionnelles : prdr décis° à la vue données (tests)

→ Probas VS Stats : · En proba, on étudie qté de (va) dt m connt loi.
· En stats, on dispose de données (ou observat°s) qu'on modélise ĉ réalisat° de (va) dt on ne connt pas la loi.

→ On cherche des infos sr cette loi inconnue ∈ des observat°s : stats inférentielle

→ "Modéliser, estimer, tester"

II / Modélisat°

∘ Les observat°s :

n observat°s $x_1, \dots, x_n$ (@ n réponses à sondage {oui / non})

On les modélise ĉ réalisat° de (va)

$$X_1(w) = x_1, \dots, X_n(w) = x_n \,.\, \left(@\ x_1 = \begin{cases} 1 & \text{si oui} \\ 0 & \text{si non} \end{cases}\right)$$

X_i est à valeur ds $\{0,1\}$ & suit une loi de Bernoulli

• Une famille de probabilités :

On suppose que les $(X_i)_{1\leq i\leq n}$ st déf sr un espace probabilisable $(\Omega, \mathcal{F})$ (& mesurables) & on munit $(\Omega, \mathcal{F})$ d'une famille de probabilités $(\mathbb{P}_\theta)_{\theta\in\Theta}$.

→ $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ modèle statistique

@ $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in[0,1]})$, θ représente la proba inconnue de voter "oui"

tq $X_i \sim \mathcal{B}ern(\theta)$ sous $\mathbb{P}_\theta$ ie $\begin{cases} \mathbb{P}_\theta(X_i = 1) = \theta \\ \mathbb{P}_\theta(X_i=0) = 1-\theta \end{cases}$ (1) $_{1\leq i\leq n}$

∘ lois des observat°s : Pour avoir la loi des observat°s $(X_1, \dots, X_n)$ la loi seule des X_i ne suffit pas.

⇝ Plusieurs modélisat°s (sondage)

• On n'interroge pas 2x m̂ personne. (tirage sans remise) : pas indépdce entre réponses.

Les $(X_i)_{1\leq i\leq n}$ ne st pas indépendantes, si on pose $S_n = \sum_{i=1}^n X_i$

On sait que θN ... N Electeurs

$(1-\theta)N$ ← n interrogés

$S_n = \text{card}\{1\leq i\leq n,\ x_i = 1\}$

$$\mathbb{P}_\theta(S_n = k) = \frac{C_{\theta N}^k \, C_{(1-\theta)N}^{n-k}}{C_N^n}$$

$S_n \sim \text{Hypergéométriq}(n, N, \theta)$

(Rq) Cela permet d'avoir la loi de $(X_1, \dots, X_n)$ (car ttes infos contenue ds S_n).

⚠ Soit $(x_1, \dots, x_n) \in \{0,1\}^n$, $k = \sum_{i=1}^n x_i$.

Calculer $\mathbb{P}_\theta((X_1, \dots, X_n) = (x_1, \dots, x_n) \mid S_n = k)$ $\forall k \in \{0, \dots, n\}$

∘ (cv) loi hypergéom → loi binomiale : $\lim_{N\to\infty} \mathbb{P}_\theta(S_n = k) = C_n^k \theta^k (1-\theta)^{n-k}$

∘ 2° modélisat° : On fait ĉ si on interrogeait les 1dvds indépendamment des autres (tirage av remise), ce q permet de sppsr les $(X_i)_{1\leq i\leq n}$ indépendtes

$X_i \sim \mathcal{B}ern(\theta)$ & l'indépdce. Rq : $\mathbb{P}_\theta(S_n = k) = C_n^k \theta^k (1-\theta)^{n-k}$

$S_n \sim \mathcal{B}in(n, \theta)$ sous $\mathbb{P}_\theta$.

Ccl : Modèle statistiq : $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ sr lequel on observe un

∘ n-échantillon $(X_1, \dots, X_n)$ → n (va) iid (indep & identiqt distribués)

• Observat°s : Réalisat°s du n-échantillons (données) $(x_1, \dots, x_n)$ tq $x_i = X_i(w)$ $_{1\leq i\leq n}$

(Rq) $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ est rarement explicite.

@ Sondage : $\Omega = \{0,1\}^n$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}_\theta(\{(w_1, \dots, w_n)\}) = \prod_{i=1}^n \theta^{w_i}(1-\theta)^{w_i}$

ens des n-uplets à valr ds $\{0,1\}$: tous les ss-ens de Ω.

$w = (w_1, \dots, w_n) \in \{0,1\}^n$ représente répses n individus interrogés

$= \begin{cases} \theta & \text{si } w_i = 1 \\ 1-\theta & \text{si } w_i = 0 \end{cases}$, $\theta \in [0,1]$

⇒ $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in[0,1]})$ modèle statistiq.

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

$X_i(\omega) = X_i(\omega_1, \dots, \omega_n) = \omega_i$, vérifier que

$$\mathbb{P}_\theta(X_i = 1) = \theta = 1 - \mathbb{P}_\theta(X_i = 0)$$

car
$$\mathbb{P}_\theta(X_i=1) = \mathbb{P}_\theta(\{\omega \in \Omega : X_i(\omega) = 1\})$$
$$= \mathbb{P}_\theta(\{\omega = (\omega_1, \dots, \omega_n) \in \Omega : \omega_i = 1\})$$
$$= \sum_{\substack{(\omega_1, \dots, \omega_n) \in \Omega, \\ \omega_i = 1}} \mathbb{P}_\theta(\{(\omega_1, \dots, \omega_n)\}).$$

cas de toutes les suites à valeurs de $\{0,1\}$ →

Rq Vers l'infini (non ②) qd on fait tendre $n \to \infty$,
$\Omega = \{0,1\}^{\mathbb{N}^*}$, $\mathcal{F} = ?$ $\mathbb{P}_\theta$ sur $(\Omega, \mathcal{F})$ pt ê construit de telle sorte qu'elle coïncide sur la précédente.

$\Rightarrow (\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in [0,1]})$ & un n-échantillon $(X_1, \dots, X_n)$ tq $X_i \sim Bern(\theta)$ $1 \leq i \leq n$ est appelé modèle de Bernoulli.

III / Estimation & intervalles de confiance

OS : Une fois le modèle posé $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$ & $(X_1, \dots, X_n)$ le n-échantillon. Comment obtenir de l'information sur le paramètre θ inconnu à partir de l'observation de $(X_1, \dots, X_n)$.

→ Estimation ponctuelle de θ & fourchette de marge d'erreur.

@ (ex 7) $n = 10$. 11 00 1 00 000.

$\Omega = \{0,1\}^{10}$, $\mathcal{F} = \mathcal{P}(\Omega)$,

$$\mathbb{P}_\theta(\{(\omega_1, \dots, \omega_{10})\}) = \theta^{\sum_{i=1}^{10} \omega_i} (1-\theta)^{10 - \sum_{i=1}^{10} \omega_i}$$

$X_i(\omega_1, \dots, \omega_n) = \omega_i$, $\mathbb{P}_\theta(X_i = 1) = \theta$.

On a envie d'estimer θ par $\frac{3}{10}$.

Marge d'erreur ?

① Estimation :

Ⓓ Un estimateur de θ est f mesurable de $(X_1, \dots, X_n)$ qui ne dépend pas de θ.

ie Un estimateur $\hat{\theta}_n$ de θ est de la forme $\hat{\theta}_n = \varphi_n(X_1, \dots, X_n)$ où φ est mesurable. (Rq) f mesurable $f : (E_1, \mathcal{F}_1) \longrightarrow (E_2, \mathcal{F}_2)$
$\forall B \in \mathcal{F}_2, f^{-1}(B) \in \mathcal{F}_1$.

Idée : Trouver des estimateurs de θ "pas trop loin" de θ.

@ $\hat{\theta}_{10} = \dfrac{\sum_{i=1}^{10} X_i}{10}$, $\hat{\theta}_{10}(\omega) = \dfrac{3}{10}$

(Rq) $\mathbb{P}_\theta((X_1, \dots, X_n) = (1,1,0,0,1,0,0,0,0,0)) = \theta^3 (1-\theta)^7 = f(\theta)$

$\hat{\theta}_{10}(\omega) = \frac{3}{10}$ est le paramètre θ qui maximise $\theta \mapsto \mathbb{P}_\theta((X_1, \dots, X_n) = (1,1,0,0,1,\dots))$ sur $[0,1]$ (calcul de $f'(\theta)$).

→ $\hat{\theta}_{10}$ est le paramètre le plus vraisemblable.

Généralisation

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

$\forall \theta \in [0,1]$, $\forall (x_1, \dots, x_n) \in \{0,1\}^n$

$$V(\theta, x_1, \dots, x_n) = \mathbb{P}_\theta((X_1, \dots, X_n) = (x_1, \dots, x_n)) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}$$

$V(\theta, x_1, \dots, x_n)$ est la f de vraisemblance (vraisemblance du param. θ à la vue de l'observation $(x_1, \dots, x_n)$).

$$\ln(V(\theta, x_1, \dots, x_n)) = \sum_{i=1}^n x_i \ln\theta + \left(n - \sum_{i=1}^n x_i\right) \ln(1-\theta)$$

$$\varphi'(\theta) = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} \Leftrightarrow \varphi'(\theta) \overset{\geq}{=} 0 \Leftrightarrow \frac{\sum x_i}{\theta} \overset{\geq}{=} \frac{n - \sum x_i}{1-\theta}$$

$$\Leftrightarrow (1-\theta)\sum x_i \overset{\geq}{=} \theta(n - \sum x_i)$$

$$\Leftrightarrow \theta \overset{\leq}{=} \frac{\sum x_i}{n}$$

θ	0		$\frac{\sum x_i}{n}$		1
$\varphi'(\theta)$		$+$	0	$-$	
φ		$\nearrow$		$\searrow$	

(3')

→ L'estimateur du maximum de vraisemblance est

$$\boxed{\hat{\theta}_n = \underset{\theta \in [0,1]}{\arg\max}\ V(\theta, X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

$\Leftrightarrow V(\theta, X_1, \dots, X_n) \leq V(\hat{\theta}_n, X_1, \dots, X_n)$ $\forall \theta \in [0,1]$

$\frac{1}{n}\sum_{i=1}^{n} X_i$ est appelé moyenne empiriq des $(X_i)_{1 \leq i \leq n}$ noté $\overline{X}_n$

→ Pq cet estimateur ?

* estimateur q maximise la vraisemblance : le paramètre le plus vraisemblable à la vue des observats $x_1, \dots, x_n$.
* il a de bonnes pptés :

$$\underbrace{\mathbb{E}_\theta(\overline{X}_n)}_{\text{espérance sous } \mathbb{P}_\theta} = \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\,\mathbb{E}_\theta\left(\sum_{i=1}^{n} X_i\right)$$

↑ par linéarité de l'espérance.

Comme $X_i \sim \mathrm{Bern}(\theta)$ sous $\mathbb{P}_\theta$. $\mathbb{E}_\theta(X_i) = \theta$ $\forall i \in \{1, \dots, n\}$

$\Rightarrow \mathbb{E}_\theta(\overline{X}_n) = \theta$.

On dit que $\overline{X}_n$ est un estimateur sans biais de θ.

• On a envie de dire que "plus n est grand plus $\overline{X}_n$ se rapproche de θ"

Comment évaluer la convergence de $\overline{X}_n$ vers θ.

* évaluer $\mathbb{P}_\theta(|\overline{X}_n - \theta| \geq \varepsilon)$

① Bienaymé-Tchebichev

$$\mathbb{P}_\theta(|\overline{X}_n - \theta| \geq \varepsilon) \leq \frac{\theta(1-\theta)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

② Moments d'ordre 4

$$\mathbb{P}_\theta(|\overline{X}_n - \theta| \geq \varepsilon) \leq \frac{1}{4n^2\varepsilon^4}.$$

③ Hoeffding $\mathbb{P}_\theta(|\overline{X}_n - \theta| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}$) pour loi va non bornée.

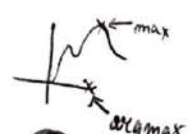
Ici c'est l'inégalité d'Hoeffding la meilleure

Rq Pr une généralisation :

* BT valable pr va de carré intégrable
* H valable pr va bornées

$\forall \varepsilon > 0$, $\lim \mathbb{P}_\theta(|\overline{X}_n - \theta| \geq \varepsilon) = 0$, on dit que $\overline{X}_n$ cv vers θ sous $\mathbb{P}_\theta$, $\theta \in [0,1]$

$$\overline{X}_n \xrightarrow[n \to \infty]{\text{proba}} \theta.$$



* CVce simple ? $\overline{X}_n(\omega) \xrightarrow[n\to\infty]{} \theta$ $\forall \omega \in \Omega$???

$\Omega = \{0,1\}^{\mathbb{N}^*}$ (infini non-dénombrable)

$\mathcal{F}$, $\mathbb{P}$ tq E_i = "pile au i^o lancer"

$\mathbb{P}_\theta(E_i) = \theta$ et les $(E_i)_{i\in\mathbb{N}^*}$ st indépendts.

$X_i(\omega) = \omega_i$ si $\omega = (\omega_n)_{n\in\mathbb{N}^*} \in \Omega$

(résultat du i^o lancer)

$\mathbb{P}_\theta(X_i = 1) = \mathbb{P}_\theta(E_i) = \theta = 1 - \mathbb{P}_\theta(X_i = 0)$

$X_i \sim \mathcal{B}ern(\theta)$ sous $\mathbb{P}_\theta$ et les $(X_i)_{i\in\mathbb{N}^*}$ st indépendantes)

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

@ $\omega = (0, \dots, 0, \dots)$ $\quad \overline{X}_n(\omega) = 0$ $\quad \forall n \in \mathbb{N}$

$\omega = (1, \dots, 1, \dots)$ $\quad \overline{X}_n(\omega) = 1$

$\exists \Omega' \in \mathcal{F}$ tq $\mathbb{P}_\theta(\Omega') = 1$, $\forall \omega \in \Omega'$, $\overline{X}_n(\omega) \xrightarrow[n\to\infty]{} \theta$.

DM

$\Omega' = \{\omega \in \Omega, \lim_{n\to\infty} \overline{X}_n(\omega) = \theta\}$

$\lim_{n\to\infty} \overline{X}_n(\omega) = \theta$, $\quad \forall \varepsilon > 0, \exists n \in \mathbb{N}^*, \forall p \geq n,$

$|\overline{X}_p(\omega) - \theta| \leq \varepsilon$.

Se traduit en termes d'événements :

$A_p(\varepsilon) = \{\omega \in \Omega, |\overline{X}_p(\omega) - \theta| \leq \varepsilon\} = \{|\overline{X}_p - \theta| \leq \varepsilon\}$

$\to \omega \in \Omega'$ ssi $\forall \varepsilon > 0$ $\exists n \in \mathbb{N}^*$ $\forall p \geq n$, $\omega \in A_p(\varepsilon)$

ssi $\forall \varepsilon > 0$, $\omega \in \bigcup_{n\in\mathbb{N}^*} \bigcap_{p\geq n} A_p(\varepsilon)$

ssi $\omega \in \bigcap_{\varepsilon>0} \bigcup_{n\in\mathbb{N}^*} \bigcap_{p\geq n} A_p(\varepsilon)$

Rq : On se ramène à une intersect° dénombrable en mq $\bigcap_{\varepsilon>0} A(\varepsilon) = \bigcap_{m\in\mathbb{N}^*} A\left(\frac{1}{m}\right)$.

En notant $A(\varepsilon) = \bigcup_{n\in\mathbb{N}^*} \bigcap_{p\geq n} A_p(\varepsilon)$.

On va mq que $\mathbb{P}_\theta(\Omega'^c) = 0$.

$\Omega'^c = \bigcup_{\varepsilon>0} \bigcap_{n\in\mathbb{N}^*} \bigcup_{p\geq n} A_p^c(\varepsilon)$

$\Omega'^c = \bigcup_{m\in\mathbb{N}^*} \bigcap_{n\in\mathbb{N}^*} \bigcup_{p\geq n} A_p^c\left(\frac{1}{m}\right)$.

4

On mq $\mathbb{P}_\theta\left(\bigcap_{m\in\mathbb{N}^*}\bigcup_{p\geq m} A_p^c(\varepsilon)\right)=0 \quad \forall \varepsilon>0$!

$\mathbb{P}_\theta(A_p^c(\varepsilon)) = \mathbb{P}_\theta(|\bar{X}_p-\theta|>\varepsilon) \leq 2e^{-2p\varepsilon^2} \xrightarrow[m\to\infty]{} 0$

$B_m(\varepsilon)=\bigcup_{p\geq m} A_p^c(\varepsilon)$, la suite $(B_m(\varepsilon))$ est $\searrow$:

$B_{m+1}(\varepsilon)\subset B_m(\varepsilon)=A_m^c(\varepsilon)\cup B_{m+1}(\varepsilon)$

$\Rightarrow \mathbb{P}_\theta\left(\bigcap_{m\in\mathbb{N}^*} B_m(\varepsilon)\right)=\lim_{m\to\infty}\mathbb{P}_\theta(B_m(\varepsilon))$

$\mathbb{P}_\theta(B_m(\varepsilon))=\mathbb{P}_\theta\left(\bigcup_{p\geq m} A_p^c(\varepsilon)\right)\leq\sum_{p=m}^{\infty}\mathbb{P}_\theta(A_p^c(\varepsilon))$.

Hoeffding: $\mathbb{P}_\theta(B_m(\varepsilon))\leq\sum_{p=m}^{\infty} 2e^{-2p\varepsilon^2}=2\dfrac{e^{-2m\varepsilon^2}}{1-e^{-2\varepsilon^2}}$

Reste d'une série convergente, $\lim_{m\to\infty}\mathbb{P}_\theta(B_m(\varepsilon))=0$.

$\Rightarrow$ On a mqé que $\forall\varepsilon>0$, $\mathbb{P}_\theta\left(\bigcap_{m\in\mathbb{N}^*}\bigcup_{p\geq m} A_p^c(\varepsilon)\right)=0$

$\Omega'^c=\bigcup_{m\in\mathbb{N}^*}\bigcap_{m\in\mathbb{N}^*}\bigcup_{p\geq m} A_p^c(\varepsilon)$

$\mathbb{P}_\theta(\Omega'^c)\leq\sum_{m\in\mathbb{N}^*}\mathbb{P}_\theta\left(\bigcap_{m\in\mathbb{N}^*}\bigcup_{p\geq m} A_p^c(\varepsilon)\right)\Rightarrow\mathbb{P}_\theta(\Omega')=1$.

Ccl $\exists\,\Omega'\in\mathcal{F}$, $\mathbb{P}_\theta(\Omega')=1$,

$\forall\omega\in\Omega'$, $\lim_{m\to\infty}\bar{X}_m(\omega)=\theta$.

On dit que $\bar{X}_m$ cv presque sûrement vers θ.

Rq si $\bar{X}_m$ cv presque sûrement vers θ alors $\bar{X}_m$ cv en proba vers θ. La cve en proba impliq cv ce presque sûre (p.s.) à la condition suffisante que $\boxed{\sum_{m\in\mathbb{N}}\mathbb{P}_\theta(|\bar{X}_m-\theta|\geq\varepsilon)<\infty}$

- Risque quadratiq

$\mathbb{E}_\theta\left[(\bar{X}_m-\theta)^2\right]=Var(\bar{X}_m)=Var_\theta\left(\frac{1}{m}\sum X_i\right)$

$=\frac{1}{m^2}Var_\theta\left(\sum X_i\right)=\frac{1}{m^2}\sum Var_\theta(X_i)$

car les (X_i) indep. $=\frac{1}{m^2}m\theta(1-\theta)=\dfrac{\theta(1-\theta)}{m}$

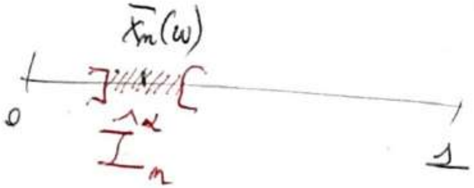
2) Intervalles de confiance

Un <u>IdC</u> est un intervalle aléatoire dans lequel se trouve le vrai paramètre inconnu de grande proba.

(D) soit $\alpha \in]0,1[$, un intervalle de confiance de niveau de confiance $1-\alpha$ est un intervalle aléatoire $\hat{I}_n^\alpha$ en f de $X_1, \dots, X_n$ (indépendant de θ) tq $\mathbb{P}_\theta(\hat{I}_n^\alpha \ni \theta) \geq 1-\alpha$

<u>Outils</u> : Inégalité de Concentration (BT ou Hoeffding)

@ Modèle de Bernoulli



θ inconnu

ex : ex 7 $\bar{X}_{10}(\omega) = \frac{3}{10}$.

$\boxed{BT}$ $\mathbb{P}_\theta(|\bar{X}_n - \theta| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$

$\iff \mathbb{P}_\theta(\theta \notin]\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon[) \leq \frac{1}{4n\varepsilon^2}$

Il suffit de prendre $\varepsilon = \varepsilon(\alpha, n)$ tq

tq $\frac{1}{4n\varepsilon^2} \leq \alpha$, $\varepsilon \leq \sqrt{\frac{1}{4n\alpha}}$.

$$\hat{I}_n^\alpha = \left]\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}}\right[\cap]0,1[$$

$$\mathbb{P}_\theta(\theta \in \hat{I}_n^\alpha) \geq 1-\alpha.$$

$\boxed{\text{Hoeffding}}$ $\mathbb{P}_\theta(|\bar{X}_n - \theta| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}$.

Il suffit de prendre $\varepsilon = \varepsilon(\alpha, n)$ tq

$$2e^{-2n\varepsilon^2} \leq \alpha \iff \varepsilon \geq \sqrt{\frac{1}{2n}\ln\frac{2}{\alpha}}$$

$$\hat{J}_n^\alpha = \left]\bar{X}_n - \sqrt{\frac{1}{2n}\ln\frac{2}{\alpha}}, \bar{X}_n + \sqrt{\frac{1}{2n}\ln\frac{2}{\alpha}}\right[\cap]0,1[$$

@ ex 7 $n = 10$.

Intervalle de confiance de niveau 90 %, $\alpha = 10\%$

$\bar{X}_{10}(\omega) = \frac{3}{10}$, $\hat{I}_{10}^{10\%} =]0; 0,8[$; $\hat{J}_{10}^{10\%} =]0; 0,687[$

$n = 1000$, $\alpha = 10\%$.

(BT) : $\frac{1}{2\sqrt{n\alpha}} = \frac{1}{20} = 0,05$, (H) $\sqrt{\frac{1}{2n}\ln\frac{2}{\alpha}} = 0,0387$

fourchette de 3,9 %

Rq : • Pas optimal si $\hat{I}_n^\alpha$ est un intervalle de confiance de niveau $1-\alpha$ alors tout $\hat{J}_n^\alpha \supset \hat{I}_n^\alpha$ est aussi un intervalle de confiance $1-\alpha$.

IV) Tests statistiques

1) Intro

Aide à la prise de décision :

- Sondage pour le second tour ; on observe 50,3% d'intention de votes pr A (sur 1000 électeurs) Peut-on conclure que A va gagner ?
- médicament : on teste un nv traitement sur un petit nbr de malades, le nv trt est-il meilleur ?

@ ex n=10 → On observe 3 piles sur 10 : pt-on conclure que la pièce est truquée ?

2 choix possibles (hypothèses), potentiellement, on peut se tromper

2) Choix des hypothèses

Chaq décision a des conséquences ≠les ± graves.

@ "Le traitement est meilleur" ou "le trait n'est pas meilleur"

2 erreurs possibles

E_1 = celle que le trait est meilleur alors que ce n'est pas le cas. Conséquence : traiter l'ens des patients avec ce traitement qui n'est pas bon

E_2 = Celle que le trait n'est pas meilleur alors qu'en réalité il l'est.

Csqce : On retravaille le trait.

On ne pourra pas minimiser les 2 erreurs & on en privilégie une : E_1.

Oriente le choix (dissymétriq) des hypothèses

H_0 : hypothèse nulle (hypo par défaut)

H_1 : hypothèse alternative

Le test de H_0 contre H_1 est construit de telle sorte que la proba. de se tromper quand on conclut H_1 est faible et contrôlée.

H_0 : "le traitement n'est pas meilleur"

H_1 : "le traitement est meilleur"

E_1 : Conclure H_1 alors qu'en réalité c'est H_0.

choix dissymétriq, dépd pt de vue. @ confinement "l'épidémie reprend" "l'épidémie ne reprend pas"

<u>Point de vue des médecins</u> :

$\mathcal{H}_0$: "l'épidémie reprend".

$\mathcal{H}_1$: "l'épidémie ne reprend pas".

<u>Pt de vue économiq</u> :

$\mathcal{H}_0$: "l'épidémie ne reprend pas"

$\mathcal{H}_1$: "l'épidémie reprend".

→ tradu° des hypothèses en f du param les hypothèses posent une question sur la loi (ou son paramètre) inconnue.

<u>Exemple médicament</u>

$\mathcal{H}_0$: "le trait n'est pas meilleur" $\quad \theta \leq \theta_0 \quad (\theta = \theta_0)$

$\mathcal{H}_1$: "le nv trait est meill^x" $\quad \theta > \theta_0$

θ_0 proba de guérison de l'ancien trait^t (connue)

θ — — — — (inconnue)

3) Construct° du test

ⓐ médicament, On veut tester $\mathcal{H}_0 : \theta \leq \theta_0$ contre $\mathcal{H}_1 : \theta > \theta_0$, à la vue de l'observa° $x_1, \dots, x_n$. (les réac°s de n patients au nouv° trait^t).

, $X_1, \dots, X_n$ iid $\mathcal{B}er(\theta)$

$$X_i = \begin{cases} 1 & \text{si le } i^{\circ} \text{ patient guérit} \\ 0 & \text{sinon} \end{cases} \quad 1 \leq i \leq n$$

θ inconn., on a estimateur $\overline{X_n}$

θ_0 connu $\Rightarrow$ Comment conclure $\mathcal{H}_1$?

↳ Réponse "naïve" si $\overline{X_n} > \theta_0$

(Pb) On veut contrôler la proba de se tromper.

$$\mathbb{P}_{\mathcal{H}_0}(\text{"conclure } \mathcal{H}_1\text{"})$$

<u>Erreur</u> :

$$\underbrace{\sup_{\theta \leq \theta_0}}_{\text{sous } \mathcal{H}_0} \mathbb{P}_\theta(\overline{X_n} \geq \theta_0) = \mathbb{P}_{\theta_0}(\overline{X_n} \geq \theta_0) \quad \text{(dans de la suite)}$$

$$= \mathbb{P}_{\theta_0}\left(\sum_{i=1}^n X_i \geq n\theta_0\right) \approx \frac{1}{2}.$$

car la médiane d'une $\mathcal{B}in(n, \theta_0)$ est $\lfloor n\theta_0 \rfloor$ ou $\lfloor n\theta_0 + 1 \rfloor$

→ On se donne une marge d'erreurs $\alpha \in]0,1[$ appelé niveau de test. Et on cherche un seuil t_α tq on conclut $\mathcal{H}_1$ si $\overline{X_n} \geq t_\alpha$ ($\overline{X_n}$ suffisamment + grd que θ_0) de telle sorte que $\sup_{\theta \leq \theta_0} \mathbb{P}_\theta(\overline{X_n} \geq t_\alpha) \leq \alpha$

$\mathcal{R}_n^\alpha = \{\overline{X_n} \geq t_\alpha\}$ = zone de rejet (évent sur lequel on rejette $\mathcal{H}_0$ & on conclut $\mathcal{H}_1$).

Règle de décision

Si on observe $\overline{X_n}(\omega) \geq t_\alpha$ alors on rejette $\mathcal{H}_0$
(Lf une erreur contrôlée $\leq \alpha$).
<u>sinon</u> on ne rejette pas $\mathcal{H}_0$ (mais l'erreur n'est pas contrôlée).

Généralisat

$(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$ $\quad \Theta =]0,1[$ (Bernoulli)

$X_1, \dots, X_n$ <u>iid</u> $\mathcal{B}ern(\theta)$

$\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$, $\Theta_0 \cap \Theta_1 = \emptyset$.

(D) Construire un test de $\mathcal{H}_0$ contre $\mathcal{H}_1$ au niveau α, c'est construire une zone de rejet (f de $X_1, \dots, X_n, \alpha$) $\mathcal{R}_n^\alpha$ tq $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\mathcal{R}_n^\alpha) \leq \alpha$.

Règle de décision qui en découle :
- si $\omega \in \mathcal{R}_n^\alpha$ alors on conclut $\mathcal{H}_1$.
- si $\omega \notin \mathcal{R}_n^\alpha$ alors on ne rejette pas $\mathcal{H}_0$.

Vocabulaire

- taille du test : $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\mathcal{R}_n^\alpha)$.
- Erreur de première espèce (E_1) : conclure $\mathcal{H}_1$ à tort contrôlée
- Erreur de seconde espèce (E_2) : conclure $\mathcal{H}_0$ à tort.

si $\theta \in \Theta_1$, $\beta(\theta) = \mathbb{P}_\theta(\mathcal{R}_n^{\alpha\, c})$
erreur de seconde espèce en $\theta \in \Theta_1$

Proba de conclure $\mathcal{H}_0$ à tort $\sup_{\theta \in \Theta_1} \mathbb{P}_\theta(\mathcal{R}_n^{\alpha\, c})$

- Puissance du test : proba de conclure $\mathcal{H}_1$ à raison.

si $\theta \in \Theta_1$ $\quad \pi(\theta) = \mathbb{P}_\theta(\mathcal{R}_n^\alpha) = 1 - \beta(\theta)$.

Plus la puissance est grde meilleur est le test

$$f \text{ puissance } \quad \Theta_1 \longrightarrow \mathbb{R}$$
$$\theta \longmapsto \mathbb{P}_\theta(\mathcal{R}_n^\alpha)$$

- Test convergent

si $\forall \theta \in \Theta_1$, $\lim_{n \to \infty} \mathbb{P}_\theta(\mathcal{R}_n^\alpha) = 1$.

- Test sans biais

si $\forall \theta \in \Theta_1$ $\quad \pi(\theta) > \alpha$,
ie $\forall \theta \in \Theta_1$, $1 - \beta(\theta) > \alpha \Leftrightarrow \beta(\theta) < 1 - \alpha$
$\Leftrightarrow \beta = \sup_{\theta \in \Theta_1} \beta(\theta) \leq 1 - \alpha$

(9). Erreur de seconde espèce.

Point de vue des médecins :

▸ Hypothèse simple : si Θ_0 ou Θ_1 sont réduits à un singleton :

- $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ (test unilatéral)
- $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ (test bilatéral)

Hypothèse multiple sinon.

4) Exemples

▸ Hypothèses simples fleurs.

Croisement entre "fleurs rouges & fleurs blanches". Gène dominant ?

→ si le rouge est dominant alors la proba d'obtenir une fleur rouge est $\frac{3}{4}$.

→ sinon proba $\frac{1}{4}$.

On observe sur n croisements le nbre S_n de FR.

$S_n \sim \mathcal{B}in(n, \theta)$ où θ est la proba d'obtenir une fleur R à l'issue du croisement (inconnue)

▸ test de H_0 : "rouge non dominant" $\theta = 1/4$

$\boxed{H_1}$ "rouge dominant" $\theta = 3/4$.

H_0, H_1 ; 0, 1 ; θ inconnu

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

$$\mathcal{R}_n^\alpha = \{\hat{\theta}_n \geq t_\alpha\}$$

Pour construire un test de niveau α de H_0 : $\theta = \frac{1}{4}$ contre $H_1 : \theta = \frac{3}{4}$. On construit une zone de rejet $\mathcal{R}_n^\alpha = \{\frac{S_n}{n} \geq t_\alpha\} = \{S_n \geq k_\alpha\}$ où k_α vérifie :

$$\underbrace{\mathbb{P}_{1/4}}_{\text{loi sous } H_0}(S_n \geq k_\alpha) \leq \alpha.$$

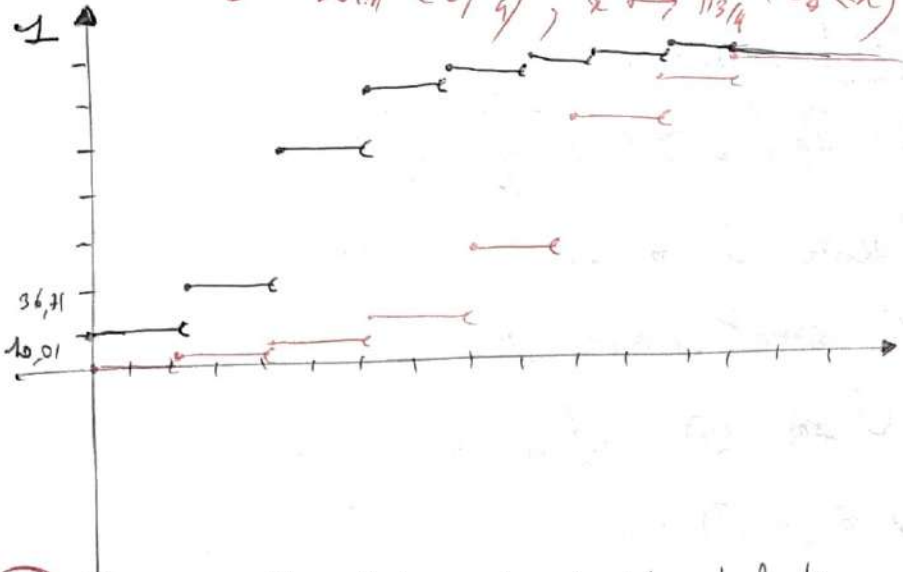
$$k_\alpha = \min\{k \in \{0, \dots, n\} : \mathbb{P}_{1/4}(S_n \geq k) \leq \alpha\}$$

@ $n = 8$ loi de S_8 sous H_0 et sous H_1

k	$\mathbb{P}_{1/4}(S_8 = k)$	$\mathbb{P}_{3/4}(S_8 = k)$	$\mathbb{P}_{1/4}(S_8 \geq k)$
0	10,01	–	100
1	26,70	0,04	89,95
2	31,15		
3	20,77		
4	8,65		11,27
5	2,31		2,72
6	0,37		0,41
7	0,04		0,04
8	–	10,01	0

→ sous $\mathcal{H}_0$: $S_8 \sim \mathcal{B}in(8, \frac{1}{4})$, $x \mapsto \mathbb{P}_{\frac{1}{4}}(S_8 \leq x)$

→ sous $\mathcal{H}_1$: $S_8 \sim \mathcal{B}in(8, \frac{3}{4})$, $x \mapsto \mathbb{P}_{3/4}(S_8 \leq x)$



Rq: Une variable aléatoire X est dite stochastiquement plus grand qu'une (va) Y si $\forall x \in \mathbb{R} : F_X(x) \leq F_Y(x)$ (où F_X et F_Y sont des fn de répartition X & Y).

$k_\alpha = \min\{k \in \{0, .., 8\} \quad \mathbb{P}_{\frac{1}{4}}(S_8 \geq k) \leq \alpha\}$

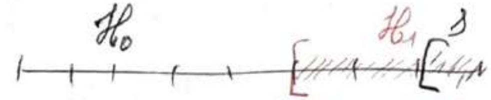
$\alpha = 5\%$ $\quad k_{5\%} = 5$ $\quad$ taille du test 2,72 %

$\alpha = 10\%$ $\quad k_{10\%} = 5$

$\alpha = 15\%$ $\quad k_{15\%} = 4$

Au niveau 5%, si on observe $S_8(\omega) \geq 5$ alors on conclut $\mathcal{H}_1$ $(\theta = \frac{3}{4})$.

$\mathcal{R}_8^{5\%} = \{S_8 \geq 5\}$ sinon on ne rejette pas $\mathcal{H}_0$.



$\{S_8 \geq 5\} = \{\overline{X}_8 \geq \frac{5}{8}\}$

Taille du test : 2,72 %.

Erreur de 2nde espèce : $\mathbb{P}_{3/4}(S_8 \leq 4) = 11,37\%$.

Puissance : $\mathbb{P}_{3/4}(S_8 \geq 5) = 88,63\%$.

Rq : Si on ↘ le niveau : $\alpha = 2\%$, $\mathcal{R}_8^{2\%} = \{S_8 \geq 6\}$ la puissance ↘ ⟺ l'erreur de 2nde espèce ↗.

p-valeur du test : On a observé $S_8(\omega) = s$.

p-valeur $\mathbb{P}_{1/4}(S_8 \geq s) = \alpha(s)$

$\alpha(s) \leq 5\% \iff \omega \in \mathcal{R}_8^{5\%}$.

Ex On observe $S_8(\omega) = 7$, $\alpha(7) = \mathbb{P}_{1/4}(S_8 \geq 7) = 0,04\%$.

$\alpha(7) \leq 5\% \Rightarrow$ on rejette $\mathcal{H}_0$ au niveau 5% (idem $\alpha = 1\%$, $\alpha = 1‰$)

Point de vue des médecins...

- **Hypothèses multiples**

ⓐ médicament, $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in]0,1[})$

$X_1, \dots, X_n$ iid $\mathcal{B}ern(\theta)$, $S_n = \sum_{i=1}^n X_i$

$\mathcal{H}_0$ $\theta \leq \theta_0$ contre $\mathcal{H}_1 : \theta > \theta_0$

$\Theta_0 =]0, \theta_0]$ & $\Theta_1 =]\theta_0, 1[$.

Zone de rejet pr test de niveau α.

k_α tq $\sup_{\theta \leq \theta_0} \mathbb{P}_\theta(S_n \geq k_\alpha) \leq \alpha \}$

$k_\alpha = \min\{k :$

Or $\sup_{\theta \leq \theta_0} \mathbb{P}_\theta(S_n \geq k) = \mathbb{P}_{\theta_0}(S_n \geq k)$

$\Rightarrow k_\alpha = \min\{k \in \{0, \dots, n\} : \mathbb{P}_{\theta_0}(S_n \geq k) \leq \alpha\}$

Preuve : si $S_n \sim \mathcal{B}in(n, \theta)$ sous $\mathbb{P}_\theta$ alors (k fixé) $\theta \longmapsto \mathbb{P}_\theta(S_n \geq k)$ est $\nearrow$ sur $]0,1[$..

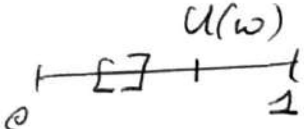
2 [M] possibles :

* $f(\theta) = \mathbb{P}_\theta(S_n \geq k) = \sum_{i=k}^n C_n^i \theta^i (1-\theta)^{n-i}$

On dérive & on mq $f'(\theta) \geq 0$.

* simulation : comment simuler une loi binomiale ? $\mathbb{P}(S_n(\theta) \geq k)$ " $\mathbb{P}_\theta(S_n \geq k)$

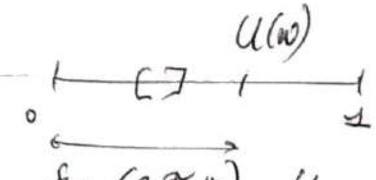
soit U une ⓥⓐ Uniforme sur $]0,1[$



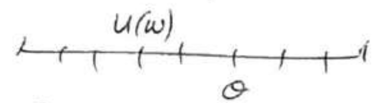
$\mathbb{P}(U \in [a,b]) = b-a$ si $[a,b] \subset]0,1[$

Transforme U en une loi de Bernoulli

$X = \mathbb{1}\{U \leq \theta\}$, $\mathbb{P}(X=1) = \mathbb{P}(U \leq \theta) = \theta$, $\mathbb{P}(X=0) = 1-\theta$



Sur $(\Omega, \mathcal{F}, \mathbb{P})$ $U_1, \dots, U_n$: n ⓥⓐ iid $\mathcal{U}nif(]0,1[)$.



$S_n(\theta) = \sum_{i=1}^n \mathbb{1}\{U_i \leq \theta\}$.

Or si $\theta \leq \theta'$ alors $S_n(\theta) \leq S_n(\theta')$

d'où $\{S_n(\theta) \geq k\} \subset \{S_n(\theta') \geq k\}$

$\Rightarrow \mathbb{P}(S_n(\theta) \geq k) \leq \mathbb{P}(S_n(\theta') \geq k)$

et $\theta \longmapsto \mathbb{P}(S_n(\theta) \geq k)$ est $\nearrow$. $\square$

$k_\alpha^1 = \max\{k \in \{0,\dots,n\} : \mathbb{P}_{\theta_0}(S_n \leq k) \leq \frac{\alpha}{2}\}$

$k_\alpha^2 = \min\{k \in \{0,\dots,n\} : \mathbb{P}_{\theta_0}(S_n \geq k) \leq \frac{\alpha}{2}\}$.

Comment déterminer k_α ?

↳ par calc ou inégalité Hoeffding

↳ à l'aide TCL limite central (si n grand voir MB1)

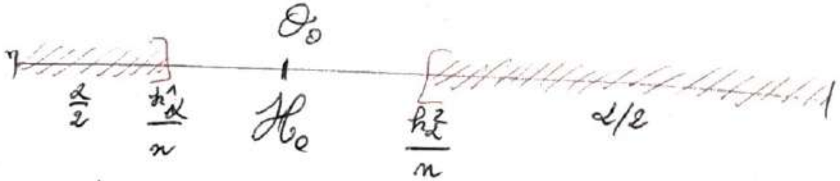
Rq Sur E_1 & E_2 :

$$E_1 = \sup_{\theta \leq \theta_0} \mathbb{P}_\theta(S_n \geq k_\alpha) = \mathbb{P}_{\theta_0}(S_n \geq k_\alpha)$$

$$E_2 = \sup_{\theta > \theta_0} \mathbb{P}_\theta(S_n < k_\alpha) = \mathbb{P}_{\theta_0}(S_n < k_\alpha) \quad \boxed{E_1 + E_2 = 1}$$

• Test bilatéral :

$H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$. → $\Theta_0 = \{\theta_0\}$ → $\Theta_1 =]0,\theta_0[\cup]\theta_0, 1[$.



$$R_n^\alpha = \{S_n \leq k_\alpha^1\} \cup \{S_n \geq k_\alpha^2\}$$

où k_α^1 & k_α^2 $\mathbb{P}_\theta(\{S_n \leq k_\alpha^1\} \cup \{S_n \geq k_\alpha^2\}) \leq \alpha$

Chap II : Généralisat°

I/ Modéliser

- Modèle statistiq : $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ où $(\Omega, \mathcal{F})$ est un espace probabilisable.

$(\mathbb{P}_\theta)_{\theta\in\Theta}$ une famille de probabilités sr cet espace.

→ Θ espace des paramètres, ici on supps $\Theta \subset \mathbb{R}^d$

=> statistiq non paramétriq.

- Observat°s : $(X_1, \dots, X_n)$ un n-échantillon de loi P_θ sous $\mathbb{P}_\theta$ (loi commune des $(X_i)_{1\leq i\leq n}$)

ici on considère des v.a iid sinon $X = (X_1, \dots, X_n)$ une seule observat° à val^rs ds $\mathbb{R}^n$.

=> objectif : ▷ estimer θ sous $g(\theta)$
▷ tester

=> outil : la vraisemblance.

II / La vraisemblance

① Lois discrètes

soit P_θ la loi de X_1 sous $\mathbb{P}_\theta$, que l'on spps discrète. soit $\underset{\sim}{\mathcal{X}}$ $\mathbb{P}_\theta(X \in \mathcal{X}) = 1$

au plus dénom^b.

a) 2 exemples

* capture - recapture

OB : estimer la populat° (les poissons) ds le lac Taille N inconnue.

Démarche : → capture : on en pêche 20 poissons auxquels on fait une marque et on les remet ds le lac.

→ recapture : On en pêche 50 avec remise et on compte le nb de Poissons marqués.

Modélisat° : $(\Omega, \mathcal{F}, (\mathbb{P}_p)_{p\in]0,1[})$ $p = \frac{20}{N}$: proba de pêcher un poisson marqué

$X_1, \dots, X_{50}$ iid $\mathcal{B}ern(p)$

$$X_i = \begin{cases} 1 & \text{si le } i^\circ \text{ poisson pêché est marqué} \\ 0 & \text{sinon} \end{cases}$$

$$S_{50} = \sum_{i=1}^{50} X_i, \qquad S_{50}(\omega) = 4.$$

$\mathbb{P}_p(S_{50}=4) = C_{50}^{4}\, p^4 (1-p)^{46}$.

Idée pr estimer p : choisir $\hat{p}(\omega)$ q maximise p, $p \in]0,1[$; $\mathbb{P}_p(S_{50}=4)$.

En dérivant en $p \in]0,1[$;

$$\hat{p}(\omega) = \underset{p\in]0,1[}{\mathrm{argmax}}\ \mathbb{P}_p(S_{50}=4) = \frac{4}{50} = \frac{S_{50}(\omega)}{50}.$$

Ce q correspond à une estimaθ de N ;

$$\hat{N}(\omega) = \frac{20}{\hat{p}(\omega)} = 250.$$

Plus généralemt, si on observe $S_{50}(\omega) = s$ alors

$$\underset{p\in]0,1[}{\mathrm{argmax}}\ L(p,s) = \frac{s}{50}.$$

où $L(p,s) = \mathbb{P}_p(S_{50}=s) = C_{50}^{s}\, p^s (1-p)^{50-s}$.

$$\hat{p} = \frac{S_{50}}{50} \quad \text{et} \quad \hat{N} = \frac{1000}{S_{50}}.$$

• Loi géométriq : On modélise le nbr de cycles avant de tomber enceinte par une loi géométriq, si θ est proba de tomber enceinte un mois donné. Si X est la va égale au nbr de cycles pr tomber enceinte.

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

On observe les données suivantes sr $n=100$ femmes

$n=100$ femmes

nbr cycles	1	2	3	4	5	6	7	8	9	10	11	12	>12
nbr ♀s	29	16	17	4	3	9	4	5	1	1	1	3	7

On ne dispose pas des données individuelles $(X_i)_{1\le i\le 100}$ mais de $(N_k)_{1\le k\le 13}$

$$N_k = \sum_{i=1}^{100} \mathbb{1}_{\{X_i = k\}} \qquad 1 \le k \le 12$$

$$N_{13} = \sum_{i=1}^{100} \mathbb{1}_{\{X_i > 12\}}.$$

Comment estimer θ ?

$\rightarrow \mathbb{P}_\theta(X_1 = 1) = \theta$

$$\hat{\theta}_1 = \frac{1}{100} \sum_{i=1}^{100} \mathbb{1}_{\{X_i = 1\}}$$

$\hat{\theta}_1$ est un estimatr sans biais de θ car

$$\mathbb{E}_\theta(\hat{\theta}_1) = \frac{1}{100} \sum_{i=1}^{\infty} \mathbb{P}_\theta(X_i = 1) = \theta$$

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i=1\}} \xrightarrow[n\to\infty]{\mathbb{P}_\theta \text{ ps}} \theta$$

$$\hat{\theta}_1(\omega) = \frac{29}{100}$$

Défaut : N'utilise pas ttes les données

→ $\mathbb{E}_\theta(X_1) = \frac{1}{\theta} \iff \theta = \frac{1}{\mathbb{E}_\theta(X_1)}$

⇒ La moyenne empiriq des observaons pourrait ê un bon candidat.

Comment faire cette moyenne empirique ?

→ supprimer la dernière colonne.

$$\hat{\theta}_2 = \frac{93}{\sum_{k=1}^{12} k N_k}, \quad \hat{\theta}_2(\omega) = 0{,}2809.$$

$$\hat{\theta}_2 = \frac{100 - N_{13}}{\sum_{k=1}^{100} k N_k} = \frac{\sum_{i=1}^{100} \mathbb{1}_{\{X_i \leq 12\}}}{\sum_{i=1}^{100} X_i \mathbb{1}_{\{X_i \leq 12\}}}$$

→ On utilise ttes les observaθs.

$$\mathbb{P}_\theta(N_1 = 29, N_2 = 16, \dots, N_{12} = 3, N_{13} = 7) = C\,\theta^{29}((1-\theta)\theta)^{16}(\theta(1-\theta)^{11})^3((1-\theta)^{12})^7 = C\,\theta^{93}(1-\theta)^{322}$$

où $\mathbb{P}_\theta(X > 12) = (1-\theta)^{12}$

$$C = \frac{100!}{29!\,16!\dots 7!}$$

$$L(\theta) = C\theta^{93}(1-\theta)^{322}$$

$$L'(\theta) = C\left[93\theta^{92}(1-\theta)^{322} - 322\,\theta^{93}(1-\theta)^{321}\right]$$

$L'(\theta) = 0$ si $\theta = \frac{93}{415}$,

soit $\hat{\theta}_3(\omega) = 0{,}224$ qui maximise $\mathbb{P}_\theta(N_1 = 29, \dots, N_{12} = 3, N_{13} = 7)$ en θ.

⚠ Vérifier que $\hat{\theta}_3 = \frac{N_1 + \dots + N_{12}}{\sum_{k=1}^{12} k N_k + 12 N_{13}}$

$$\hat{\theta}_3 = \frac{\sum_{i=1}^{100} \mathbb{1}_{\{X_i \leq 12\}}}{\sum_{i=1}^{100} X_i \mathbb{1}_{\{X_i \leq 12\}} + 12 \sum_{i=1}^{\infty} \mathbb{1}_{\{X_i > 12\}}}$$

16 stat royale.

b) Définiθ

(5) La vraisemblance associée à l'observaθ $x_1, \dots, X_n$ est $V(\theta, x_1, \dots, x_n) = \mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n)$

$\theta \in \Theta$, $(x_1, \dots, x_n) \in \mathcal{X}^n$

$(\mathbb{P}_\theta(X_1 \in \mathcal{X})) = 1$.

$V(\theta, x_1, \dots, x_n)$ est la vraisemblance du paramètre θ à la vue de l'observaθ $x_1, \dots, x_n$.

On maximise cette vraisemblance pr avoir le paramètre le plus vraisemblable.

(5) L'estimateur du maximum de vraisemblance est

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \; V(\theta, X_1, \dots, X_n).$$

Rq :

- Modèle de Bernoulli : $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in]0,1[})$

$X_1, \dots, X_n$ iid $Bern(\theta)$, $\mathcal{X} = \{0,1\}$

$\forall (x_1, \dots, x_n) \in \{0,1\}^n$ $\forall \theta \in]0,1[$

$$V(\theta, x_1, \dots, x_n) = \mathbb{P}_\theta(X_1 = x_1, \dots)$$

$$= \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}$$

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

$$\hat{\theta}_n = \underset{\theta \in]0,1[}{\operatorname{argmax}} \; V(\theta, X_1, \dots, X_n)$$

à arg-maximiser en $\theta \in]0,1[$,

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

$$\Leftrightarrow L(\theta, X_1, \dots, X_n) = \ln V(\theta, X_1, \dots, X_n)$$

$$= \left(\sum_{i=1}^n X_i\right)\ln\theta + \left(n - \sum_{i=1}^n X_i\right)\ln(1-\theta).$$

$$L(\theta, X_1, \dots, X_n) = S_n \ln\theta + (n - S_n)\ln(1-\theta)$$

$$\frac{\partial}{\partial\theta} L(\theta, X_1, \dots, X_n) = \frac{S_n}{\theta} - \frac{n - S_n}{1-\theta}$$

pour $\dfrac{S_n}{\theta} = \dfrac{n - S_n}{1-\theta}$, $S_n = \sum_{i=1}^n X_i$

$$\Leftrightarrow (1-\theta)S_n = \theta(n - S_n)$$

$$\Leftrightarrow \theta = \frac{S_n}{n}.$$

$$\hat{\theta}_n = \underset{\theta \in]0,1[}{\operatorname{argmax}} V(\theta, X_1, \dots, X_n) = \underset{\theta \in]0,1[}{\operatorname{argmax}} L(\theta, X_1, \dots, X_n)$$

$$= \frac{S_n}{n}.$$

- Comme $V(\theta, X_1, \dots, X_n)$ est une forme de produit on pt utiliser pr faciliter les calculs la log-vraisemblance

$$L(\theta, X_1, \dots, X_n) = \ln V(\theta, X_1, \dots, X_n)$$

↑ loglood

- Pas tjs $\exists$, pas tjs FF explicite.

e/ Lois à densités

va aléatoires à densité $(\Omega, \mathcal{F}, \mathbb{P})$

soit X une va à valeurs ds $\mathbb{R}$,

$$X : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$
$$\omega \longmapsto X(\omega) \quad \text{mesurable.}$$

La loi est déterminée par la f de répartiθ

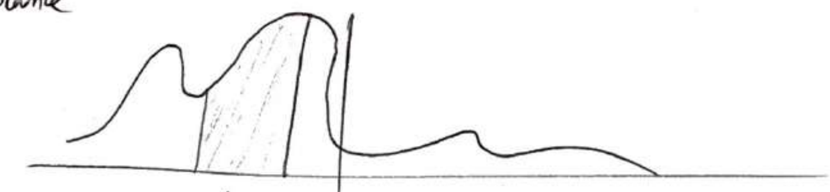
$$F_X : \mathbb{R} \longrightarrow [0,1]$$
$$t \longmapsto F_X(t) = \mathbb{P}(X \leq t)$$

cont à droite
limite à gauche
cadlag

- On dit que X a pr densité f si $\forall A \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}(X \in A) = \int_A f(x)\,dx$.

f est une f positive sr $\mathbb{R}$ & intégrable sr $\mathbb{R}$.

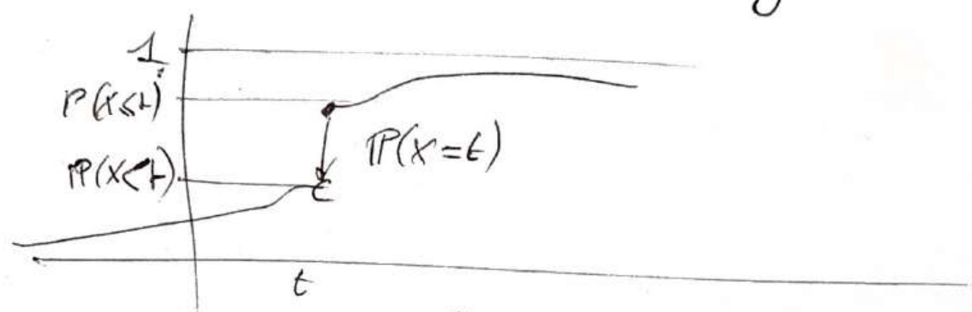
$$\int_{\mathbb{R}} f(x)\,dx = 1$$



- si X a pr densité f alors F_X est cont sr $\mathbb{R}$.

$$F_X(t) - F_X(t^-) = \mathbb{P}(X \leq t) - \mathbb{P}(X < t) = \mathbb{P}(X = t) = 0$$

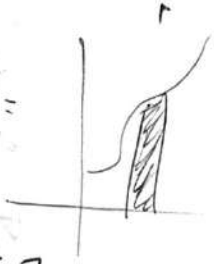
cadlag : cont à droite limitée à gauche



$$\lim_{t_n \to t} \mathbb{P}(X \leq t_n) = \mathbb{P}\Big(\bigcup_{n \in \mathbb{N}^*} \{X \leq t_n\}\Big) = \mathbb{P}(X < t)$$

$\forall t \in \mathbb{R}$, $\mathbb{P}(X = t) = 0$, si X est à densité

mais $\mathbb{P}(X \in]t-\varepsilon, t+\varepsilon[) = \displaystyle\int_{t-\varepsilon}^{t+\varepsilon} f(x)\,dx$



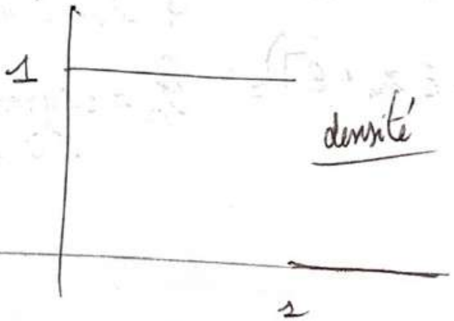
Exemple : ▸ Loi uniforme sur $[0,1]$

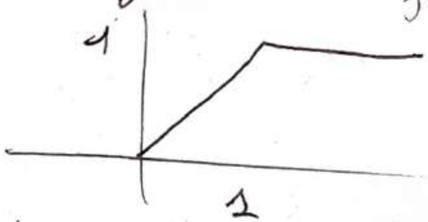
$X \sim \mathrm{Unif}([0,1])$ si X a pr densité $f = \mathbb{1}_{[0,1]}$

$\mathbb{P}(X \in A) = \displaystyle\int_A \mathbb{1}_{[0,1]}(x)\,dx$

$= \lambda(A \cap [0,1])$

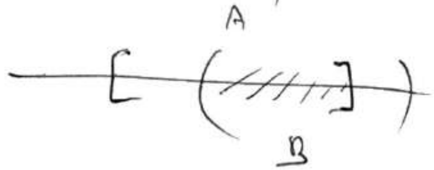
où λ est la mesure de Lebesgue sur $\mathbb{R}$.





si $B \in \mathcal{B}(\mathbb{R})$, on dit que $X \sim \mathrm{Unif}(B)$,

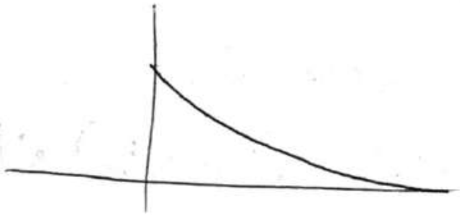
si $\forall A \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}(X \in A) = \dfrac{\lambda(A \cap B)}{\lambda(B)}$

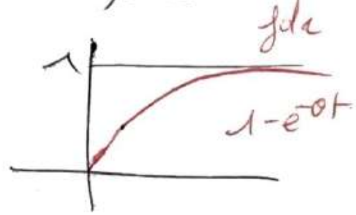


(se généralise à $\mathbb{R}^2, \dots, \mathbb{R}^m$)

▸ Loi exponentielle, $X \sim \mathcal{E}xp(\theta)$,

si X a pr densité $f_\theta(x) = \theta e^{-\theta x} \mathbb{1}_{[0,\infty[}(x)$.

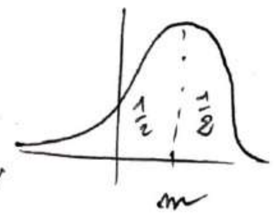


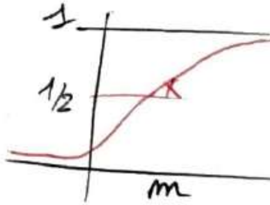


$\mathbb{P}(X \geq 0) = \displaystyle\int_0^\infty \theta e^{-\theta x}\,dx = 1$.

▸ Loi normale

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





si X a pr densité

$f_{m,\sigma}(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-m)^2}$

$f_{0,1}(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$\displaystyle\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx = 1$.

Statistiq $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$

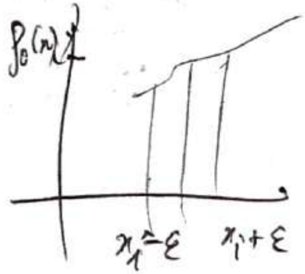
$X_1, \dots, X_n$ iid tq X_1 est à densité f_θ.

$\mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n) = 0$.

Heuristiq : si $\varepsilon > 0$, $(x_1, \dots, x_n) \in \mathbb{R}^n$, $\theta \in \Theta$

$$\mathbb{P}_\theta(X_1 \in [x_1 - \varepsilon, x_1 + \varepsilon], \dots, X_n \in [x_n - \varepsilon, x_n + \varepsilon]) = \prod_{i=1}^n \mathbb{P}_\theta(X_i \in [x_i - \varepsilon, x_i + \varepsilon])$$

$$= \prod_{i=1}^n \int_{x_i - \varepsilon}^{x_i + \varepsilon} f_\theta(x)\,dx$$



$$\simeq \prod_{i=1}^n (2\varepsilon) f_\theta(x_i)$$

$$\simeq (2\varepsilon)^n \prod_{i=1}^n f_\theta(x_i)$$

① La vraisemblance associée à un échantillon $X_1, \dots, X_n$ de loi à densité f_θ est donnée par $\forall \theta \in \Theta$, $\forall (x_1, \dots, x_n) \in \mathbb{R}^n$,

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

l'estimateur du max de vrais. est

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}}\, V(\theta, X_1, \dots, X_n) = \underset{\theta \in \Theta}{\operatorname{argmax}} \prod_{i=1}^n f_\theta(X_i)$$

Exemples Estimateur du max de vraisemblce pr des lois à densité.

1) Loi uniforme sur $[0, \theta]$

Sur $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in]0, \infty[})$, on observe un échantillon $X_1, \dots, X_n$ où $X_i \sim Unif([0, \theta])$.

R $X \sim Unif([0, \theta])$

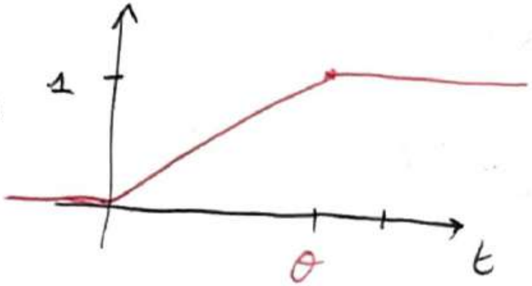
si $\mathbb{P}(X \in B) = \dfrac{\lambda(B \cap [0, \theta])}{\theta}$ $\forall B \in \mathcal{B}(\mathbb{R})$

où λ : mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

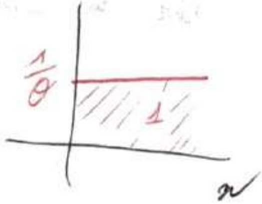
FdR : $B =]-\infty, t]$ $\forall t \in \mathbb{R}$.

$$F_\theta(t) = \mathbb{P}(X \leq t) = \frac{\lambda(]-\infty, t] \cap [0,\theta])}{\theta}$$

$$= \begin{cases} 0 & \text{si } t < 0 \\ t/\theta & \text{si } 0 \leq t < \theta \\ 1 & \text{si } t \geq \theta. \end{cases}$$



Densité $\quad f_\theta(x) = \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x)$



Rq : $F_\theta(t) = \int_{-\infty}^{t} f_\theta(x)\, dx$

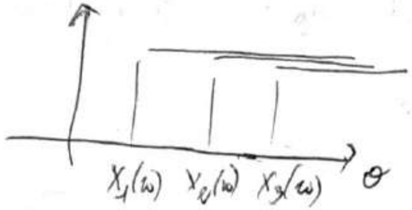
La vraisemblance ~~A~~ à à l'observaθ $X_1, \dots, X_n$ est

$$V(\theta, X_1, \dots, X_n) = \prod_{i=1}^{n} f_\theta(X_i) = \prod_{i=1}^{n} \left(\frac{1}{\theta} \mathbb{1}_{[0,\theta]}(X_i)\right),$$

$$V(\theta, X_1, \dots, X_n) = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbb{1}_{[0,\theta]}(X_i)$$

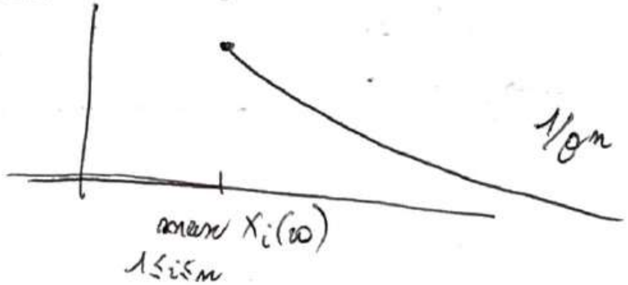
$$\mathbb{1}_{[0,\theta]}(X_i) = \mathbb{1}_{[X_i, \infty[}(\theta).$$

$$\prod_{i=1}^{n} \mathbb{1}_{[0,\theta]}(X_i) = \prod_{i=1}^{n} \mathbb{1}_{[X_i,\infty[}(\theta) = \mathbb{1}_{[\max_{1\leq i\leq n} X_i, \infty[}(\theta)$$

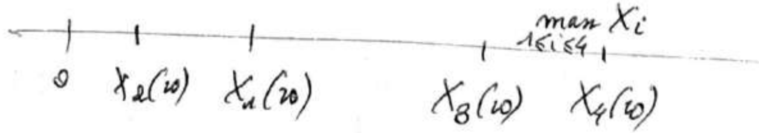


$= 1$ si $\theta \geq X_i$ $\forall i \in \{1, \dots, n\}$ si $\theta \geq \max X_i$.

$$V(\theta, X_1, \dots, X_n) = \frac{1}{\theta^n} \mathbb{1}_{[\max_{1\leq i\leq n} X_i, \infty[}(\theta)$$



$$\hat{\theta}_n = \underset{\theta > 0}{\operatorname{argmax}}\, V(\theta, X_1, \dots, X_n) = \max_{1\leq i\leq n} X_i$$

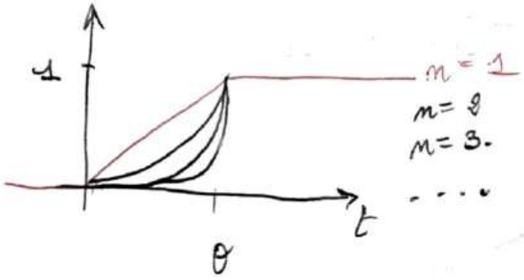


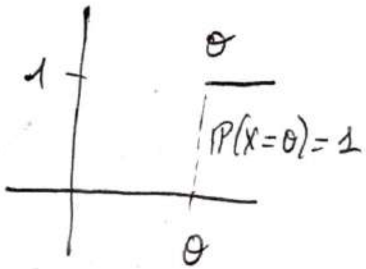
Chercher la loi de $\hat{\theta}_n$ via la Fdr.

$$\mathbb{P}_\theta(\hat{\theta}_n \leq t) = \mathbb{P}_\theta\left(\max_{1\leq i\leq n} X_i \leq t\right)$$

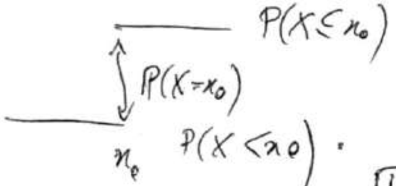
$$= \mathbb{P}_\theta\left(\bigcap_{i=1}^{n} \{X_i \leq t\}\right) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i \leq t) = (F_\theta(t))^n$$

$$\Rightarrow \mathbb{P}_\theta(\hat{\theta}_n \leq t) = \begin{cases} 0 & \text{si } t < 0 \\ \left(\frac{t}{\theta}\right)^n & \text{si } 0 \leq t < \theta \\ 1 & \text{si } t \geq \theta \end{cases}$$





NB



On montrera que $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta \text{ ps}} \theta$.

Densité de $\hat{\theta}_n$: $\quad n \dfrac{t^{n-1}}{\theta^n} \mathbb{1}_{[0,\theta]}(t)$

Autre estimateur. si $X \sim \mathcal{U}nif([0,\theta])$.

$0 \qquad \theta/2 = \mathbb{E}(X) \qquad \theta$

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_\theta(x)\, dx = \int_{\mathbb{R}} x \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x)\, dx = \int_0^\theta \frac{x}{\theta} dx = \frac{\theta}{2}$$

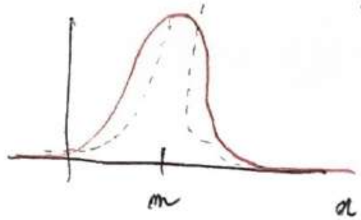
$\Rightarrow \theta = 2\mathbb{E}(X)$.

$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \Rightarrow \tilde{\theta}_n = 2\overline{X}_n$ est un estimateur de θ.

On les comparera par la suite.

2) Loi normale

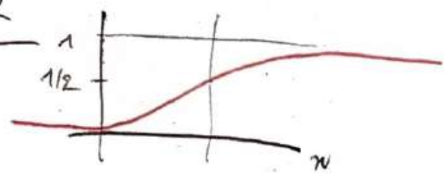
(R1) $X \sim \mathcal{N}(m, \sigma^2)$ si X a pr densité $f_{m,\sigma^2}(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-m)^2}$



si σ petit

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$$

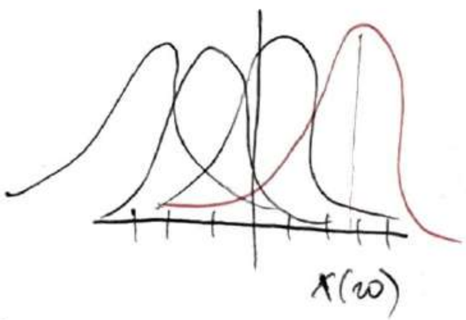
fdR



$\mathbb{E}(X) = m$

$Var(X) = \sigma^2$.

• $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\mathbb{R}})$ sur lequel on observe $X_1, \dots, X_n$ iid : $X_i \sim \mathcal{N}(\theta, 1)$.



si $n = 1$: l'estimateur de θ le + vraisemblable est X.

$$V(\theta, X_1, \dots, X_n) = \prod_{i=1}^{n} f_{\theta,1}(X_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i-\theta)^2}{2}}$$

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

on cherche $\hat{\theta}_n$ q maximise $V(\theta, X_1, \dots, X_n)$ en θ.

$\Leftrightarrow$ minimiser $\theta \longmapsto \sum_{i=1}^{n}(X_i-\theta)^2$.

Or $\sum_{i=1}^{n}(X_i-\theta)^2 = \sum_{i=1}^{n} X_i^2 - 2\theta \sum_{i=1}^{n} X_i + n\theta^2$

admet un minimum en $\hat{\theta}_n = \overline{X}_n$

• $\theta = (m, \sigma^2) \in \mathbb{R} \times]0, \infty[$

$(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \mathbb{R}\times]0,\infty[})$ $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(m, \sigma^2)$

Vraisemblance : $V(m, \sigma^2, X_1, \dots, X_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(X_i-m)^2}$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-m)^2}$$

$$\ln\left(V(m, \sigma^2, X_1, \dots, X_n)\right) = \frac{-n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-m)^2$$

$$\frac{\partial}{\partial m}\ln V(m, \sigma^2, X_1, \dots, X_n) = \frac{1}{2\sigma^2} 2 \sum (X_i - m)$$

$$\frac{\partial}{\partial \sigma^2}\ln V(m, \sigma^2, X_1, \dots, X_n) = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum (X_i-m)^2$$

$\frac{\partial}{\partial m}\ln V(m, \sigma^2, X_1, \dots, X_n) = 0$ ssi $m = \overline{X}_n = \hat{m}_n$.

$\frac{\partial}{\partial \sigma^2}\ln V(m, \sigma^2, X_1, \dots, X_n) = 0$ ssi $\sigma^2 = \sigma^2(m) = \frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2$

$\Rightarrow$ l'estimateur du max de vraisemblance de (m, σ^2) est $(\hat{m}_n, \hat{\sigma}^2_n)$; $\hat{m}_n = \overline{X}_n$, $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

(variance empiriq)

Rq $X \sim Unif(\{x_1,\dots,x_n\})$;

$$Var(X) = \mathbb{E}\left[(X-\mathbb{E}(X))^2\right]$$

$$\mathbb{E}(X) = \sum_{i=1}^{n} \frac{x_i}{n} = \bar{x}_n$$

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

III / Pptés des estimateurs

$(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ un modèle statistique muni d'observ. $X_1,\dots,X_n$ iid soit $\hat{\theta}_n$ un estimateur de θ.

1) Biais & risque quadratiq

Le biais de l'estimateur est $b_n(\theta) = \theta - \mathbb{E}_\theta(\hat{\theta}_n)$ et l'estimateur $\hat{\theta}_n$ est sans biais si $\mathbb{E}_\theta(\hat{\theta}_n) = \theta$.

- le risque quadratiq de l'estimateur $\hat{\theta}_n$ est

$$R_n(\hat{\theta}_n) = \mathbb{E}_\theta\left[(\hat{\theta}_n - \theta)^2\right]$$

Rq Si $\hat{\theta}_n$ est sans biais alors $R_n(\hat{\theta}_n) = Var_\theta(\hat{\theta}_n)$

Décompo. " biais - variance "

$$R_n(\hat{\theta}_n) = \mathbb{E}_\theta\left[\left(\hat{\theta}_n - \mathbb{E}_\theta(\hat{\theta}_n) + \mathbb{E}_\theta(\hat{\theta}_n) - \theta\right)^2\right]$$

$$= \mathbb{E}_\theta\left[(\hat{\theta}_n - \mathbb{E}_\theta(\hat{\theta}_n))^2\right] + (\mathbb{E}_\theta(\hat{\theta}_n) - \theta)^2$$

$$+ 2\mathbb{E}_\theta\left[(\hat{\theta}_n - \mathbb{E}_\theta(\hat{\theta}_n))(\mathbb{E}_\theta(\hat{\theta}_n) - \theta)\right]$$

$$= b_n^2(\theta) + Var(\hat{\theta}_n) + 2(\mathbb{E}_\theta(\hat{\theta}_n) - \theta)\underbrace{\mathbb{E}_\theta\left[(\hat{\theta}_n - \mathbb{E}_\theta(\hat{\theta}_n))\right]}_{=0}$$

$$= b_n^2(\theta) + Var_\theta(\hat{\theta}_n)$$

$$\boxed{R_n(\hat{\theta}_n) = b_n^2(\theta) + Var_\theta(\hat{\theta}_n)}$$

→ Le risq quadratiq est un critère pr comparer des estimateurs entre eux.

Exemples : • modèle de Bernoulli

$\bar{X}_n$ est un estimat^{r} sans biais de θ.

$$R_n(\bar{X}_n) = Var_\theta(\bar{X}_n) = \frac{\theta(1-\theta)}{n} \xrightarrow[n\to\infty]{} 0$$

• loi Uniforme

$X_1,\dots,X_n$ iid $Unif([0,\theta])$

$$\hat{\theta}_n = \max_{1\le i\le n} X_i, \quad \tilde{\theta}_n = 2\bar{X}$$

$$\mathbb{E}_\theta(\overline{X}_n) = \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\theta(X_i) = \frac{\theta}{2}$$

$\Rightarrow \tilde{\theta}_n$ est un estimat^r sans biais de θ.

- $\hat{\theta}_n$ a pr densité $n\dfrac{x^{n-1}}{\theta^n}\mathbb{1}_{[0,\theta]}(x)$

$$\mathbb{E}_\theta(\hat{\theta}_n) = \int_{\mathbb{R}} x \cdot n \cdot \frac{x^{n-1}}{\theta^n}\mathbb{1}_{[0,\theta]}(x)\,dx.$$

$$\mathbb{E}_\theta(\hat{\theta}_n) = \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \frac{n}{n+1}\theta$$

$\hat{\theta}_n = \max\limits_{1\le i\le n} X_i \le \theta$ $\mathbb{P}_\theta$ p.s.

→ $\hat{\theta}_n$ n'est pas un estimat^r sans biais de θ mais il est asymptotiquem^t sans biais.

Calculer les RQs quadratiqs de $\tilde{\theta}_n$ et $\hat{\theta}_n$

$$\tilde{\theta}_n \downarrow \text{ en } \frac{1}{n} \qquad \hat{\theta}_n \downarrow \ \frac{1}{n^2}$$

$\hat{\theta}_n$ a un RQ quad bien + petit que $\tilde{\theta}_n$.

② Consistance

- On dit que $\hat{\theta}_n$ est un estimateur faiblem^t consistant de θ si $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta\text{-proba}} \theta$.

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}_\theta(|\hat{\theta}_n - \theta| \geq \varepsilon) = 0.$$

- On dit que $\hat{\theta}_n$ est un estimat^r fortement consistant de θ si $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta \text{ ps}} \theta$.

$$\Leftrightarrow \mathbb{P}_\theta\left(\{\omega \in \Omega : \hat{\theta}_n(\omega) \xrightarrow[n\to\infty]{} \theta\}\right) = 1 \quad \forall \theta \in \Theta$$

Outils probabilistes.

$\hat{\theta}_n$ fortement consistant de $\theta \Rightarrow$ faiblement consistant de θ.

Réciproqm^t : Conditio suffisante si $\sum_{n=1}^{\infty}\mathbb{P}_\theta(|\hat{\theta}_n - \theta| \geq \varepsilon) < \infty$ alors $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta \text{ ps}} \theta$

Exemple : loi uniforme sur $[0,\theta]$, $\hat{\theta}_n = \max\limits_{1\le i\le n} X_i$.

∘ consistance faible.

$\forall \varepsilon > 0$ $\mathbb{P}_\theta(|\hat{\theta}_n - \theta| \geq \varepsilon) = \mathbb{P}(\theta - \hat{\theta}_n \geq \varepsilon)$ car $\hat{\theta}_n \le \theta$ $\mathbb{P}_\theta$ ps

$$= \mathbb{P}_\theta(\hat{\theta}_n \le \theta - \varepsilon) = (F_\theta(\theta-\varepsilon))^n$$

$$= \left(\frac{\theta-\varepsilon}{\theta}\right)^n \text{ si } \varepsilon > 0 \text{ ou } 0 \text{ si } \varepsilon \geq \theta.$$

Dc $\lim_{n \to +\infty} \left(\frac{\theta-\varepsilon}{\theta}\right)^n = 0$ 1

$\Rightarrow \forall \varepsilon > 0, \ \mathbb{P}_\theta(|\hat{\theta}_n - \theta| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0$

∘ De plus, $\sum_{n=1}^{\infty} \left(\frac{\theta-\varepsilon}{\theta}\right)^n < \infty$ $\varepsilon < \theta$.

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

et $\hat{\theta}_n$ est un estimateur fortement consistant de θ.

Loi forte des grds nbres

si $(X_n)_{n\in\mathbb{N}^*}$ est une suite d va iid tq
$\mathbb{E}(|X_i|) < \infty$ alors $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{ps} \mathbb{E}(X)$

ex Loi uniforme sur $[0,\theta]$.

$\tilde{\theta}_n = 2\overline{X}_n$, $\overline{X}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta \ ps} \frac{\theta}{2} \Rightarrow \tilde{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta \ ps} \theta$

$\tilde{\theta}_n$ est un estimat^r fortement consistant de θ.

③ Loi asymptotiq

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

La loi exacte de la loi somme peut ê difficile à calculer & la loi asymptotiq est plus accessible & universelle

∘ Convergence en loi :

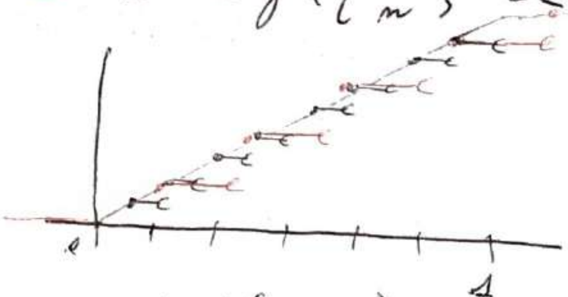
Ⓓ Une suite de var $(X_n)_{n\in\mathbb{N}^*}$ CV en loi vers une var X qd $n\to\infty$ si

$\boxed{\lim_{n\to\infty} F_{X_n}(x) = F_X(x)}$ en tt point x de continuité de F_X.

où F_{X_n} est FdR de X_n.

F_X —— X. sur $(\Omega, \mathcal{F}, \mathbb{P})$

ex $X_n \sim \mathrm{Unif}\left(\left\{\frac{k}{n}, \ 1 \leq k \leq n\right\}\right)$



$X_n \xrightarrow[n\to\infty]{loi} X$

$X \sim \mathrm{Unif}([0,1])$.

⚠ Mq $F_{X_n}(x) \xrightarrow[n\to\infty]{} F_X(x)$ $\forall x \in \mathbb{R}$

STOP

proba ps CV ⇒ en loi

@ $X_i \sim \mathcal{U}nif([0,\theta])$, $T_n = \max_{1 \leq i \leq n} X_i$ ($\leq \theta$ ps)

On pose $Z_n = n(\theta - T_n)$.

On pt mq $Z_n \xrightarrow[n\to\infty]{loi} Z$ où $Z \sim \mathcal{E}xp\left(\frac{1}{\theta}\right)$

Conséquence de la CV en loi : si $X_n \xrightarrow[n\to\infty]{loi} X$.

$\forall (a,b) \in \mathbb{R}^2$, $a < b$,

$$\mathbb{P}(X_n \in]a,b]) \xrightarrow[n\to\infty]{loi} \mathbb{P}(X \in]a,b])$$

car $\mathbb{P}(X_n \in]a,b]) = F_{X_n}(b) - F_{X_n}(a) \xrightarrow[n\to\infty]{} F_X(b) - F_X(a) = \mathbb{P}(X \in]a,b])$

Th de limite centrale (TLC)

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de va réelles iid & de carré intégrables $\mathbb{E}(X_1^2) < +\infty$, (définies sur m̂ espace de proba) $(\Omega, \mathcal{F}, \mathbb{P})$.

On pose $S_n = \sum_{i=1}^{n} X_i$

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \xrightarrow[n\to\infty]{loi} Z \text{ où } Z \sim \mathcal{N}(0,1).$$

Rq • $\mathbb{E}(X_1^2) < \infty \Rightarrow \mathbb{E}(|X_1|) < \infty$ et $Var(X_1) < \infty$.

• $Var(S_n) = \sum_{i=1}^{n} Var(X_i)$ car les $(X_i)_{1 \leq i \leq n}$ sont indep

$= n\, Var(X_1)$ car de m̂ loi.

• $\mathbb{E}(S_n) = n\,\mathbb{E}(X_1)$.

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n\,\mathbb{E}(X_1)}{\sqrt{n\, Var(X_1)}} = \sqrt{n}\,\frac{\overline{X}_n - \mathbb{E}(X_1)}{\sqrt{Var(X_1)}}$$

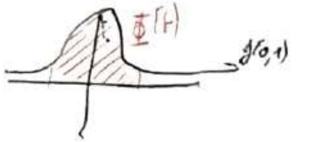
TLC fournit une vitesse de CVce de $\overline{X}_n$ vers $\mathbb{E}(X_1)$ (ps par la LFGN) en $\frac{1}{\sqrt{n}}$.

$$\overline{X}_n = \mathbb{E}(X_1) + \sqrt{\frac{Var(X_1)}{n}}\, Z_n$$

• si $a < b$:

$$\mathbb{P}\left(a \leq \sqrt{n}\,\frac{\overline{X}_n - \mathbb{E}(X_1)}{\sqrt{Var(X_1)}} \leq b\right) \xrightarrow[n\to\infty]{} \mathbb{P}(a \leq Z \leq b)$$

R : $Z \sim \mathcal{N}(0,1)$ a pr densité :



$$f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ sur } \mathbb{R},$$

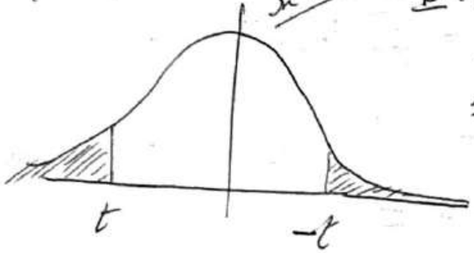
sa FdR notée Φ est déf : $\Phi(t) = \mathbb{P}(Z \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$

2P centrée : d'espérance nulle
réduite : var 1.

Table des val^{RS} de $\Phi(t)$, $t \geq 0$. $\mathbb{P}(-Z \geq -t)$ car Z & $-Z$ ont m̂ loi.
(val^{RS} $\geq 1/2$)

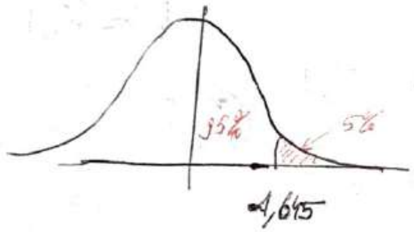
si $t<0$

si $t>0$: $\Phi(t) = \mathbb{P}(Z \leq t) = \mathbb{P}(Z \geq -t)$
$= 1 - \mathbb{P}(Z \leq -t)$
$= 1 - \Phi(-t)$



si $t>0$: $\Phi(t) = 1 - \Phi(-t)$

Valeurs utiles: t tq $\Phi(t) = 0{,}95$.
$t = 1{,}645$.



tq $\Phi(t) = 0{,}975$
$t = 1{,}96$.

Applicat° : $\boxed{M}$ de Bernoulli.

$\boxed{M}$ statistiq $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in]0,1[})$.
$X_1, \dots, X_n$ iid Bern (θ) sous $\mathbb{P}_\theta$.

TLC (les $(X_i)_{1 \leq i \leq n}$ st de carrés intégrables).

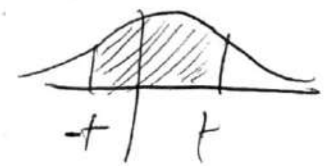
$$\Rightarrow \sqrt{n}\,\frac{\overline{X}_n - \theta}{\sqrt{\theta(1-\theta)}} \xrightarrow[n\to\infty]{\mathbb{P}_\theta\text{-loi}} Z \text{ où } Z \sim \mathcal{N}(0,1).$$

Conséqce : Intervalle de fluctuat°.

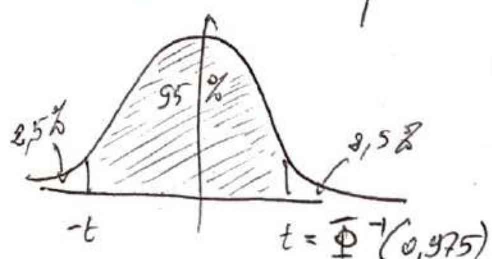
$$\mathbb{P}_\theta\left(-t \leq \sqrt{n}\,\frac{\overline{X}_n - \theta}{\sqrt{\theta(1-\theta)}} \leq t\right) \xrightarrow[n\to\infty]{} \mathbb{P}_\theta(-t \leq Z \leq t)$$

$$\Leftrightarrow \mathbb{P}_\theta\left(\sqrt{n}\,\frac{|\overline{X}_n - \theta|}{\sqrt{\theta(1-\theta)}} \leq t\right) \xrightarrow[n\to\infty]{} \mathbb{P}_\theta(|Z| \leq t)$$

or $\mathbb{P}(|Z| \leq t) = \mathbb{P}(Z \leq t) - \mathbb{P}(Z \leq -t))$
$= \Phi(t) - \Phi(-t)$
$= \Phi(t) - (1 - \Phi(t))$
$= 2\Phi(t) - 1$



On cherche par exemple t tq $\mathbb{P}_\theta(|Z| \leq t) = 0{,}95$
$\Leftrightarrow 2\Phi(t) - 1 = 0{,}95$.
$\Leftrightarrow \Phi(t) = 0{,}975$.
$\Leftrightarrow t = \Phi^{-1}(0{,}975) = 1{,}96$.



$$\mathbb{P}_\theta\left(\theta - 1{,}96\sqrt{\frac{\theta(1-\theta)}{n}} \leq \overline{X}_n \leq \theta + 1{,}96\sqrt{\frac{\theta(1-\theta)}{n}}\right) \xrightarrow[n\to\infty]{} 0{,}95$$

 inég. large = in. stricte car loi à densité.

Car $\theta(1-\theta) \leq \frac{1}{4}$,

$$\lim_{n\to\infty} \mathbb{P}_\theta\left(\theta - \frac{1{,}96}{2\sqrt{n}} \leq \bar{X}_n \leq \theta + \frac{1{,}96}{2\sqrt{n}}\right) \geq 0{,}95$$

$$\mathbb{P}_\theta\left(\bar{X}_n \in \left[\theta - \frac{1{,}96}{2\sqrt{n}} ; \theta + \frac{1{,}96}{2\sqrt{n}}\right]\right) \gtrsim 0{,}95$$

si n est assez grd.

Contrôle de l'approximat gaussienne.

Th (Berry - Esseen)

Soit $(X_n)_{n\in\mathbb{N}^*}$ suite de var iid tq $\mathbb{E}(|X_1|^3) < \infty$.

On note $\sigma^2 = \mathrm{Var}(X_1)$, $\rho^3 = \mathbb{E}\left[|X_1 - \mathbb{E}(X_1)|^3\right]$

$\sigma^2 > 0$ & $\rho^3 > 0$.

alors si $Z_n = \dfrac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}}$ où $S_n = \sum_{i=1}^n X_i$.

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(Z_n \leq x) - \mathbb{P}(Z \leq x)| \leq c \frac{\rho^3}{\sigma^3} \frac{1}{\sqrt{n}}.$$

$c \to \frac{1}{2}$

Esseen 1942 : $c = 7{,}59$

Applicat : $X_i \sim \mathrm{Bern}(\theta)$.

$\sigma^2 = \theta(1-\theta)$, $\rho^3 = \theta(1-\theta)(\theta^2 + (1-\theta)^2)$

$$\frac{\rho^3}{\sigma^3} = \frac{\theta^2 + (1-\theta)^2}{\sqrt{\theta(1-\theta)}} \leq \frac{1}{\sqrt{\theta(1-\theta)}}$$

ex 3600 fois un dé & S nb de 6 obtenus.

$S \sim \mathrm{Bin}\left(3600, \frac{1}{6}\right)$.

$$\mathbb{P}(540 \leq S \leq 660) = \mathbb{P}\left(\left|\frac{S-600}{\sqrt{500}}\right| \leq \frac{6}{\sqrt{5}}\right)$$

$$\simeq \Phi\left(\frac{6}{\sqrt{5}}\right) - \Phi\left(-\frac{6}{\sqrt{5}}\right) = 2\Phi\left(\frac{6}{\sqrt{5}}\right) - 1 \simeq 0{,}993$$

erreur d'approximat.

$$\left|\mathbb{P}\left(\frac{S-600}{\sqrt{500}} \leq \frac{6}{\sqrt{5}}\right) - \Phi\left(\frac{6}{\sqrt{5}}\right)\right| \leq c \frac{6}{\sqrt{5}} \frac{1}{\sqrt{3600}} = c/10\sqrt{5}.$$

$$\left|\mathbb{P}\left(\frac{S-600}{\sqrt{500}} \leq \frac{-6}{\sqrt{5}}\right) - \Phi\left(\frac{-6}{\sqrt{5}}\right)\right| \leq \frac{c}{10\sqrt{5}}$$

$$\Rightarrow \left|\mathbb{P}\left(\left|\frac{S-600}{\sqrt{500}}\right| \leq \frac{6}{\sqrt{5}}\right) - \left(\Phi\left(\frac{6}{\sqrt{5}}\right) - \Phi\left(\frac{-6}{\sqrt{5}}\right)\right)\right| \leq \frac{2c}{10\sqrt{5}} \simeq 0{,}045.$$

$\mathbb{P}(540 \leq S \leq 660) \geq 0{,}993 - 0{,}045.$

Rq Par Bienaymé-Tchebitchev :

$$\mathbb{P}(540 \leq S \leq 660) = \mathbb{P}(|S-600| \leq 60)$$
$$= 1 - \mathbb{P}(|S-600| > 60)$$
$$\geq 1 - \frac{500}{60^2} = 1 - \frac{5}{36} = 0{,}86.$$

Vocab statistiq

Ⓓ Un estimat^R est asymptotiquement normal. s'il Ⓒⓥ en loi (à un chgt d'échelle au préalable) vers une loi $\mathcal{N}(0,1)$. On parle de normalité asymptotiq de l'estimat^R.

IV Intervalles de confiance $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$

$\Theta \subset \mathbb{R}$. $X_1, \dots, X_n$ iid ss Θ.

Ⓓ Un intervalle de confiance pour θ de niveau de confiance $1-\alpha$ ($\alpha \in]0,1[$) est un intervalle aléatoire $\hat{I}_n$ f de $X_1, \dots, X_n$ (indépdt de Θ) tq $\mathbb{P}(\theta \in \hat{I}_n) \geq 1-\alpha$.

Rq $\mathbb{P}_\theta(\theta \in \hat{I}_n) = \mathbb{P}_\theta(\hat{I}_n \ni \theta)$ aléatoire

Généralisaº : $\Theta \in \mathbb{R}^d$.

Région de confiance $\hat{D}_n \subset \mathbb{R}^d$.

$$\mathbb{P}_\theta(\theta \in \hat{D}_n) \geq 1-\alpha$$

② $\boxed{M}$ de Construcº

$\hat{I}_n$ (ou $\hat{D}_n$) est svt basé sr estimat^R de θ

$\hat{I}_n = [\hat{\theta}_n - \varepsilon, \hat{\theta}_n + \varepsilon]$ à $\varepsilon > 0$ (dpdt de n & α, mais pas de θ).

ou $\hat{I}_n = [\frac{1}{c}\hat{\theta}_n, c\,\hat{\theta}_n]$ $c > 1$.

a) Calcul exact

si on connaît la loi d'un estimat^R de θ.

@ (Estimaº nbr taxes all des 2GM ŷ matricules chars détruits).

n tirages à remise ds urne q contient N balles numérotées de 1 à N, N est inconnu.

X_i résultat du i^e tirage $1 \leq i \leq n$.

$(\Omega, \mathcal{F}, (\mathbb{P}_N)_{N \in \mathbb{N}^*})$, $X_i \sim \mathcal{U}nif(\{1, \dots, N\})$ iid

$$\mathbb{P}_N(X_i = k) = \frac{1}{N} \qquad 1 \leq k \leq N.$$

Estimat^R du max de vraisembl^{ce}.

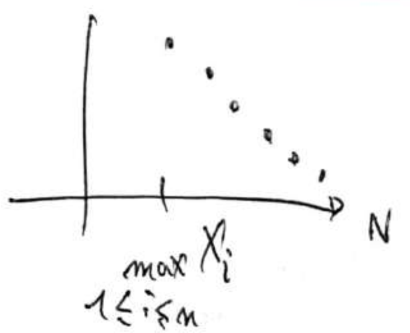
$\forall\, (x_1, \dots, x_n) \in (\mathbb{N}^*)^n$

$$V(N, x_1, \dots, x_n) = \prod_{i=1}^{n} \mathbb{P}_N(X_i = x_i)$$

$$= \prod_{i=1}^{n} \frac{1}{N} \mathbb{1}\{x_i \leq N\}$$

$$= \frac{1}{N^n} \mathbb{1}\{x_i \leq N : \forall i \in \{1, \dots, n\}\}$$

$$\boxed{V(N, x_1, \dots, x_n) = \frac{1}{N^n} \mathbb{1}\{N \geq \max_{1 \leq i \leq n} x_i\}}$$



$\hat{N}_n = \max_{1 \leq i \leq n} X_i$ est l'estimat^r du max de vraisemblance de N. Sa loi est fac. calculable car $\mathbb{P}_N(\hat{N}_n \leq k) = \mathbb{P}_N(\max_{1 \leq i \leq n} X_i \leq k)$

$$= \mathbb{P}_N\left(\bigcap_{i=1}^{n} \{X_i \leq k\}\right) = \prod_{i=1}^{n} \mathbb{P}_N(X_i \leq k) = \left(\frac{k}{N}\right)^n$$

<u>Intervalle de confiance pr N de niv^o $1-\alpha$</u>

$\hat{N}_n \leq N$ et $\mathbb{P}_N(\varepsilon N \leq \hat{N}_n \leq N) =$

$$= 1 - \mathbb{P}_N(\hat{N}_n < \varepsilon N) = 1 - \mathbb{P}_N(\hat{N}_n \leq \lceil \varepsilon N \rceil - 1)$$

$$= 1 - \left(\frac{\lceil \varepsilon N \rceil - 1}{N}\right)^n \geq 1 - \varepsilon^n$$

Car $\varepsilon - \frac{1}{N} \leq \frac{\lceil \varepsilon N \rceil - 1}{N} < \varepsilon$, $\alpha = \varepsilon^n$, $\varepsilon = \sqrt[n]{\alpha}$.

$$\boxed{\Rightarrow \mathbb{P}_N\left(N \in \left[\hat{N}_n, \frac{\hat{N}_n}{\sqrt[n]{\alpha}}\right]\right) \geq 1 - \alpha}$$

ⓑ Δ les inégalités de concentrat° t.q. à n fixé $X_1, \dots, X_n$ iid ss $\mathbb{P}_\theta$.

- (BT) si les $(X_i)_{1 \leq i \leq n}$ est de carrés intégrables

alors $\boxed{\mathbb{P}_\theta\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}_\theta(X_1)\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}_\theta(X_1)}{n\varepsilon^2}}$

- Hoeffding si les $(X_i)_{1 \leq i \leq n}$ st bornées

$a \leq X_i \leq b$ $\qquad \mathbb{P}_\theta(a \leq X_1 \leq b) = 1$

alors $\boxed{\mathbb{P}_\theta\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}_\theta(X_1)\right| > \varepsilon\right) \leq 2e^{-2\frac{n\varepsilon^2}{(b-a)^2}}}$

Plus restrictif ds les hypo H.S mlk ds inégalités.

c) Asymptotiq

Un intervalle de confiance $\hat{I}_n$ pour θ de niv° de confiance asymptotiq $1-\alpha$ si $\lim_{n\to\infty} \mathbb{P}_\theta(\theta \in \hat{I}_n) \geq 1-\alpha$.

Outils : TLC :

si $(X_n)_{n\in\mathbb{N}^*}$ est une suite de v.a. iid de carré intégrable

pas besoin

alors $\sqrt{n}\,\frac{\overline{X}_n - \mathbb{E}_\theta(X_1)}{\sqrt{\mathrm{Var}_\theta(X_1)}} \xrightarrow[n\to\infty]{\text{en loi sous } \mathbb{P}_\theta} Z \sim \mathcal{N}(0,1)$

Rq (BT) variance finie (H) v.a. bornée.

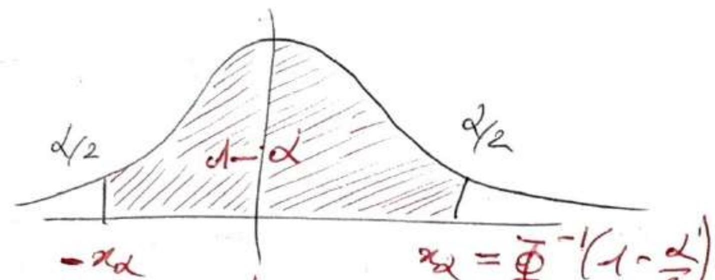
Applicat° au modèle de Bernoulli

X_i iid $\mathrm{Bern}(\theta)$, $\theta \in \,]0,1[$.

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

$\lim_{n\to\infty} \mathbb{P}_\theta\left(\sqrt{n}\,\frac{|\overline{X}_n - \theta|}{\sqrt{\theta(1-\theta)}} \leq \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)$

$= \mathbb{P}\left(|Z| \leq \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right) = 1-\alpha$



Rq $\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx = \mathbb{P}(Z \leq t)$

On cherche $\boxed{x_\alpha > 0}$ tq $\qquad Z \sim \mathcal{N}(0,1)$

$\boxed{\mathbb{P}(|Z| \leq x_\alpha) = 1-\alpha}$

or $\mathbb{P}(|Z| \leq x_\alpha) = \mathbb{P}(Z \leq x_\alpha) - \mathbb{P}(Z \leq -x_\alpha)$

$= \Phi(x_\alpha) - (1-\Phi(x_\alpha))$

$= 2\Phi(x_\alpha) - 1$

car $\mathbb{P}(Z \leq -x_\alpha) = \mathbb{P}(-Z \geq x_\alpha) = \mathbb{P}(Z \geq x_\alpha)$
$= 1 - \mathbb{P}(Z < x_\alpha) = 1 - \Phi(x_\alpha)$

Puis on résoud $2\Phi(x_\alpha) - 1 = 1 - \alpha$

$\iff \Phi(x_\alpha) = 1 - \frac{\alpha}{2}$

$\iff x_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$.

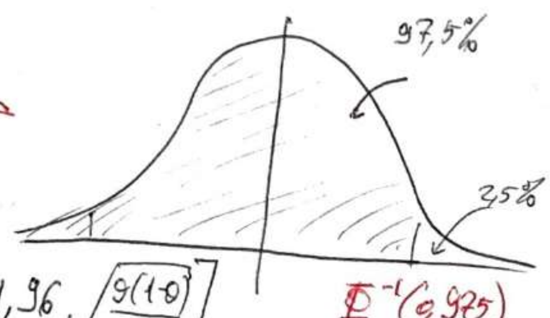
$\boxed{RQ}$ En exercice analytiq si on cherche a_α & b_α tq $\mathbb{P}(Z \in [a_\alpha, b_\alpha]) = 1 - \alpha$ (*)
le plus petit intervalle q vérifie (*), est centré en 0.

⚠ $\alpha = 0,05$, $\Phi^{-1}(1 - \frac{\alpha}{2}) = \Phi^{-1}(0,975) = 1,96$

Pour un intervalle de niveau asymptotiq 95 %, on a :

$$\mathbb{P}_\theta\left(\sqrt{n}\,\frac{|\bar{X}_n - \theta|}{\sqrt{\theta(1-\theta)}} \leq 1,96\right) \xrightarrow[n\to\infty]{} 0,95$$

⚠ la Variance dépend de θ



$$\hat{I}_n(\theta) = \left[\bar{X}_n - 1,96\sqrt{\frac{\theta(1-\theta)}{n}}, \bar{X}_n + 1,96\sqrt{\frac{\theta(1-\theta)}{n}}\right]$$

n'est pas IdC car dépd de θ !

→ Comment se débarrasser de θ (ds la variance) ?

$\boxed{M1}$ Résoudre l'inéquat° (en θ).

$$\sqrt{n}\,\frac{|\bar{X}_n - \theta|}{\sqrt{\theta(1-\theta)}} \leq 1,96 \iff \sqrt{n}\,|\bar{X}_n - \theta| \leq 1,96\sqrt{\theta(1-\theta)}$$

$$\iff n(\bar{X}_n - \theta)^2 \leq 1,96^2\,\theta(1-\theta)$$

$$\iff \theta^2(n + 1,96^2) + \theta(-2n\bar{X}_n - 1,96^2) + n\bar{X}_n^2 \leq 0$$

$\iff \theta \in [\hat{\theta}_n^1, \hat{\theta}_n^2]$, $\Delta_n = (2n\bar{X}_n + 1,96^2)^2 - 4n\bar{X}_n^2(n + 1,96^2)$
$= 4 \times 1,96^2 n \bar{X}_n + 1,96^4 - 4 \times 1,96^2 n \bar{X}_n^2$
$= 4 \times 1,96^2 n \bar{X}_n(1 - \bar{X}_n) + 1,96^4$

$$\hat{\theta}_n^1 = \frac{2n\bar{X}_n + 1,96^2 - \sqrt{\Delta_n}}{2(n + 1,96^2)}$$

$$\hat{\theta}_n^2 = \frac{2n\bar{X}_n + 1,96^2 + \sqrt{\Delta_n}}{2(n + 1,96^2)}$$

$\boxed{M2}$ On majore la variance :

Comme $\theta(1-\theta) \leq \frac{1}{4}$. $\sqrt{\theta(1-\theta)} \leq \frac{1}{2}$

$\hat{I}_n(\theta) \subset \hat{I}_n(\frac{1}{2})$ $\forall \theta \in]0,1[$

$$\hat{I}_n\left(\tfrac{1}{2}\right) = \left[\bar{X}_n - \frac{1,96}{2\sqrt{n}} ; \bar{X}_n + \frac{1,96}{2\sqrt{n}}\right]$$

$$\lim_{n\to\infty} \mathbb{P}_\theta(\theta \in \hat{I}_n(\theta)) = 0,95$$

$$\Longrightarrow \lim_{n\to\infty} \mathbb{P}_\theta\left(\theta \in \hat{I}_n\left(\tfrac{1}{2}\right)\right) \geq 0,95$$

$\boxed{M_3}$ On estime la variance :

OB : Remplacer $\theta(1-\theta)$ par $\overline{X}_n(1-\overline{X}_n)$
(c'est la variance empirique, estimat^R fortement consistant de la variance $Var_\theta(X_1) = \theta(1-\theta)$.

Outil probabiliste

Ⓛ de Slutsky

soit $(Y_n)_{n\in\mathbb{N}^*}$ & $(Z_n)_{n\in\mathbb{N}^*}$ 2 suites de va.
définies sur le même espace $(\Omega, \mathcal{F}, \mathbb{P})$ tq
$Y_n \xrightarrow[n\to\infty]{loi} Y$ & $Z_n \xrightarrow[n\to\infty]{proba} c$

$$\Longrightarrow \quad (Y_n, Z_n) \xrightarrow[n\to\infty]{loi} (Y, c).$$

Conséqce :

$X_n \xrightarrow[n\to\infty]{loi} X$ ssi $\forall f$ h cont bornée

$$\mathbb{E}(h(X_n)) \xrightarrow[n\to\infty]{} \mathbb{E}(h(X))$$

→ permet de définir la CVce en loi pr des vecteurs aléatoires (X_n & X à valeurs ds $\mathbb{R}^d$).

→ si g est cont à valeurs ds $\mathbb{R}$ &
$(Y_n, Z_n) \xrightarrow[n\to\infty]{loi} (Y, c)$.
alors $g(Y_n, Z_n) \xrightarrow[n\to\infty]{loi} g(Y, c)$.

ex $Y_n + Z_n \xrightarrow[n\to\infty]{loi} Y + c$

$Y_n \times Z_n \xrightarrow[n\to\infty]{loi} cY$

Applicaθ du Ⓛ au TLC :

Th TLC + autonormalisaθ :

si $(X_n)_{n\in\mathbb{N}^*}$ est une suite de va iid de variance finie & si V_n est un estimat^R faiblem^t consistant de la variance alors

$$\sqrt{n}\,\frac{\overline{X}_n - \mathbb{E}_\theta(X_1)}{\sqrt{V_n}} \xrightarrow[n\to\infty]{loi} Z \sim \mathcal{N}(0,1)$$

CV loi = CVce en pt de continuité de la f

fblt consist^t
CV en proba.

Preuve : $\sqrt{n}\,\frac{\overline{X}_n - \mathbb{E}_\theta(X_1)}{\sqrt{n}} = \underbrace{\sqrt{n}\,\frac{\overline{X}_n - \mathbb{E}_\theta(X_1)}{\sqrt{\mathrm{Var}_\theta(X_1)}}}_{Y_n} \times \sqrt{\frac{\mathrm{Var}_\theta(X_1)}{V_n}}$

$Y_n \xrightarrow[n\to\infty]{loi} Y$ où $Y \sim \mathcal{N}(0,1)$

$Z_n = V_n$, $Z_n \xrightarrow[n\to\infty]{proba} \mathrm{Var}_\theta(X_1)$

alors $(Y_n, Z_n) \xrightarrow[n\to\infty]{loi} (Y, \mathrm{Var}_\theta(X_1))$

$g(y,z) = y \times \sqrt{\frac{\mathrm{Var}_\theta(X_1)}{z}}$

$g(Y_n, Z_n) \longrightarrow g(Y, \mathrm{Var}_\theta(X_1)) = Y.$

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

Preuve du (L) de Slutsky :

$\forall$ h cont bornée, $\mathbb{E}[h(Y_n, Z_n)] \xrightarrow[n\to\infty]{} \mathbb{E}(h(Y,c))$

admis il suffit de le faire pr $h(y,z) = f(y)g(z)$

f & g cont bornées.

$|\mathbb{E}[f(Y_n)g(Z_n)] - \mathbb{E}[f(Y)g(c)]|$

$= |\mathbb{E}[f(Y_n)g(Z_n) - f(Y)g(c)]|$

$= |\mathbb{E}[f(Y_n)[g(Z_n) - g(c)] + g(c)[f(Y_n) - f(Y)]]|$

$\leq \underbrace{|\mathbb{E}(f(Y_n)(g(Z_n) - g(c)))|}_{A} + \underbrace{|\mathbb{E}[g(c)(f(Y_n) - f(Y))]|}_{B}$

$B \leq |g(c)|\,|\mathbb{E}(f(Y_n)) - \mathbb{E}(f(Y))| \xrightarrow[n\to\infty]{} 0$

$A \leq |\mathbb{E}[|f(Y_n)|\,|g(Z_n) - g(c)|]| \leq \|f\|_\infty \mathbb{E}[|g(Z_n) - g(c)|]$

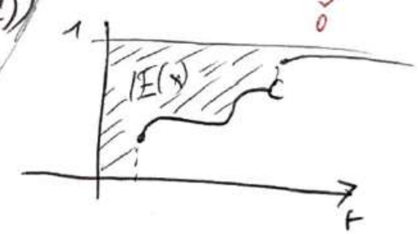
(L) (CVce dominée en probabilité)

soit $(W_n)_{n\in\mathbb{N}^*}$ une suite de va tq

- $\exists M \in \mathbb{R}_+$, $|W_n(\omega)| < M$ $\forall \omega \in \Omega$
- $W_n \xrightarrow[n\to\infty]{proba} 0$ $\Rightarrow \mathbb{E}(|W_n|) \xrightarrow[n\to\infty]{} 0$.

(R) si X va positive

$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq t)\, dt.$



Preuve : $\mathbb{E}(|W_n|) = \int_0^\infty \mathbb{P}(|W_n| \geq t)\,dt$

$\forall \varepsilon > 0$

$= \int_0^M \mathbb{P}(|W_n| \geq t)\,dt$

$= \int_0^\varepsilon \mathbb{P}(|W_n| \geq t)\,dt + \int_\varepsilon^M \mathbb{P}(|W_n| \geq t)\,dt$

$\leq \varepsilon + \mathbb{P}(|W_n| \geq \varepsilon)(M - \varepsilon)$

car $\mathbb{P}(|W_n| \geq t) \leq \mathbb{P}(|W_n| \geq \varepsilon)\ \forall t \geq \varepsilon$.

$\Rightarrow \lim_{n\to\infty} \mathbb{E}(|W_n|) \leq \varepsilon \quad \forall \varepsilon > 0$

$\rightarrow \lim_{n\to\infty} \mathbb{E}(|W_n|) = 0$

On l'applique à $W_n = g(Z_n) - g(c) \xrightarrow[n\to\infty]{\text{Proba}} 0$

Applicat° au $\boxed{M}$ de Bernoulli :

$\sqrt{n}\ \dfrac{\overline{X}_n - \theta}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} \xrightarrow[n\to\infty]{\mathbb{P}_\theta\ \text{loi}} Z$ où $Z \sim \mathcal{N}(0,1)$.

$\mathbb{P}_\theta\left(\theta \in \left[\overline{X}_n \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right]\right) \xrightarrow[n\to\infty]{} 1-\alpha$

$\Rightarrow$ Pour Bernoulli : 5 $\boxed{M}$ IdC

- BT $\hat{I}_n^{BT} = \left[\overline{X}_n - \frac{1}{2\sqrt{n\alpha}}, \overline{X}_n + \frac{1}{2\sqrt{n\alpha}}\right] \cap]0,1[$
- Hoeffding $\hat{I}_n^{H} = \left[\overline{X}_n - \sqrt{\frac{-1}{2n}\ln\frac{\alpha}{2}}, \overline{X}_n + \sqrt{\frac{-1}{2n}\ln\frac{\alpha}{2}}\right] \cap]0,1[$
- TLC
  - 2 méthodes d'inégalité
  - majoration de la variance $\not\!\!D$ résultat inégalité

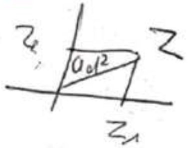
  $\hat{I}_n^{M} = \left[\overline{X}_n - \frac{\Phi^{-1}(1-\frac{\alpha}{2})}{2\sqrt{n}}, \overline{X}_n + \frac{\Phi^{-1}(1-\frac{\alpha}{2})}{2\sqrt{n}}\right)$
  - estimation de la variance

  $\hat{I}_n^{E} = \left[\overline{X}_n \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right]$

③ Cas particulier des échantillons gaussiens

loi du Chi-deux

si $(Z_i)_{1\leq i\leq d}$ est une suite de v.a. iid de loi $\mathcal{N}(0,1)$ alors la loi de $U_d = \sum_{i=1}^{d} Z_i^2 = \|Z\|^2$ où $Z=(Z_1,\dots,Z_d)\in\mathbb{R}^d$ est appelée loi du χ^2 à d degrés de liberté.



U_d est positive.

$\mathbb{E}(U_d) = \sum_{i=1}^{d} \mathbb{E}(Z_i^2) = d$

$Z_i \sim \mathcal{N}(0,1)$, $\mathbb{E}(Z_i)=0$

$Var(Z_i) = 1 = \mathbb{E}(Z_i^2)$

$Var(U_d) = \sum_{i=1}^{d} Var(Z_i^2)$ car les $(Z_i)^2_{1\leq i\leq d}$ sont indep.

$Var(Z_i^2) = \mathbb{E}(Z_i^4) - \mathbb{E}(Z_i^2)^2 = 3-1=2.$

Calculer $\forall n\in\mathbb{N}^*$, $\mathbb{E}(Z^n)$, si $Z\sim\mathcal{N}(0,1)$

$Var(U_d) = 2d$

densité de U_d.

densité $\mathbb{E}(\varphi(Z_1^2)) = \int_{\mathbb{R}} \varphi(z^2)\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz$

car disjoints

$= 2\int_0^{\infty} \varphi(z^2)\frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$

densité $d=1$

$= \int_0^{\infty} \varphi(u)\frac{1}{\sqrt{2\pi u}} e^{-\frac{u}{2}}\,du$ $\quad \left(u=z^2,\ z=\sqrt{u},\ dz=\frac{1}{2\sqrt{u}}\,du\right).$

Rq: φ mesurable bornée : $\mathbb{E}[\varphi(z)] = \int_{\mathbb{R}} \varphi(z) f_Z(z)\,dz.$

$f_1(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} \mathbb{1}_{]0,\infty[}(x)$

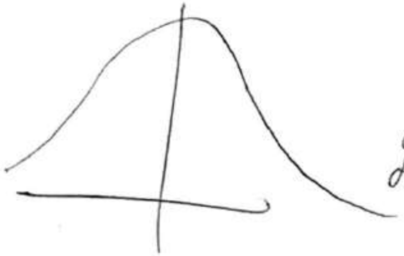
densité de U_1 (de Z_i^2).

$f_d(x) = \frac{1}{2^{d/2}\Gamma\left(\frac{d}{2}\right)} x^{\frac{d}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{]0,\infty[}(x)$

$\Gamma\left(\frac{d}{2}\right) = \int_0^{\infty} t^{\frac{d}{2}-1} e^{-t}\,dt$) on peut par la notation de $(n-1)!$

Loi de Student

si $Z \sim \mathcal{N}(0,1)$ & $U_d \sim \chi^2(d)$
& si Z & U_d st indépendantes alors la loi de $\sqrt{d}\,\dfrac{Z}{\sqrt{U_d}}$ est appelé loi de Student à d degré de liberté



$$g_d(x) = \frac{1}{\sqrt{d\pi}} \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \left(1+\frac{t^2}{d}\right)^{-\frac{d+1}{2}}$$

$\mathbb{E}(T_d) = 0$, $Var(T_d) = \dfrac{d}{d-2}$, $d > 2$.

Uti de Student

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

alors • $\overline{X}_n \sim \mathcal{N}(m, \frac{\sigma^2}{n})$

cad $\sqrt{n}\,\dfrac{|\overline{X}_n| - m}{\sigma} \sim \mathcal{N}(0,1)$

• $\dfrac{n V_n}{\sigma^2} = \displaystyle\sum_{i=1}^{n} \frac{(X_i - \overline{X}_n)^2}{\sigma^2} \sim \chi^2(n-1)$

V_n & $\overline{X}_n$ st indep.

• $\sqrt{n}\,\dfrac{\overline{X}_n - m}{\sqrt{V_n^*}}$, $V_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

$V_n^* = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

(estimat° sans biais de la variance).

Rq/s • si $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$,
X_1, X_2 indépendants
alors $X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Conséquence : $X_1 + \dots + X_n \sim \mathcal{N}(nm, n\sigma^2)$

$\frac{1}{n}(X_1 + \dots + X_n) \sim \mathcal{N}(m, \frac{\sigma^2}{n})$.

• $\displaystyle\sum_{i=1}^{n}\left(\frac{X_i - m}{\sigma}\right)^2 \sim \chi^2(n)$

$Z_i = \dfrac{(X_i - m)}{\sigma} \sim \mathcal{N}(0,1)$

On perd un deg de liberté en remplaçant m par $\overline{X}_n$

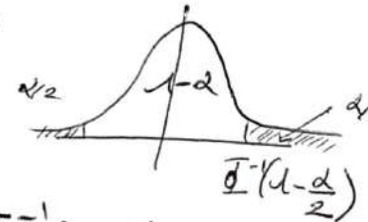
Intervalles de confiance de niveau $1-\alpha$

- $\boxed{\text{Pour la moyenne } m}$

→ $\boxed{\text{à variance } \sigma^2 \text{ connue}}$ $\sqrt{n}\,\dfrac{\bar{X}_n - m}{\sigma} \sim \mathcal{N}(0,1)$

Donc $\mathbb{P}\left(\sqrt{n}\,\dfrac{|\bar{X}_n - m|}{\sigma} \leq \Phi^{-1}\left(1-\dfrac{\alpha}{2}\right)\right) = 1-\alpha$

Φ est FdR de $\mathcal{N}(0,1)$.



d'où $\mathbb{P}_m\left(m \in \left[\bar{X}_n \pm \Phi^{-1}\left(1-\dfrac{\alpha}{2}\right)\dfrac{\sigma}{\sqrt{n}}\right]\right) = 1-\alpha$

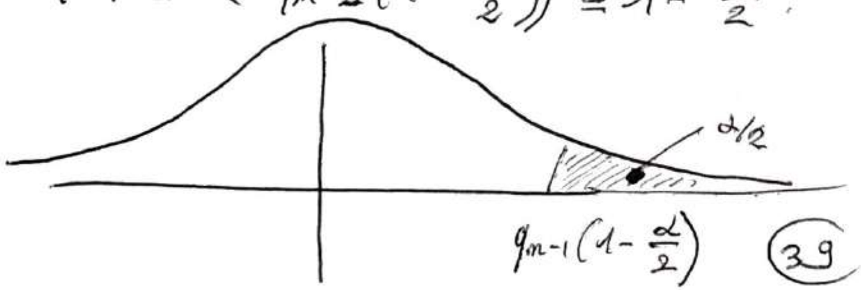
$\hat{I}_n = \left[\bar{X}_n \pm \Phi^{-1}\left(1-\dfrac{\alpha}{2}\right)\dfrac{\sigma}{\sqrt{n}}\right]$ est int d conf pr m de niveau $1-\alpha$.

→ $\boxed{\text{à variance } \sigma^2 \text{ inconnue}}$

$\sqrt{n}\,\dfrac{\bar{X}_n - m}{\sqrt{V_n^*}} = \sqrt{n-1}\,\dfrac{\bar{X}_n - m}{\sqrt{V_n}} \sim \text{Student}(n-1)$

$q_{n-1}\left(1-\dfrac{\alpha}{2}\right)$ le quantile d'ordre $1-\dfrac{\alpha}{2}$ de la loi de Student $(n-1)$ si $T_{n-1} \sim \text{Student}(n-1)$,

$\mathbb{P}\left(T_{n-1} \leq q_{n-1}\left(1-\dfrac{\alpha}{2}\right)\right) = 1-\dfrac{\alpha}{2}$.



, $\forall (m,\sigma^2) \in \mathbb{R}\times\mathbb{R}^+$,

$\mathbb{P}_{m,\sigma^2}\left(\sqrt{n}\,\dfrac{|\bar{X}_n - m|}{\sqrt{V_n^*}} \leq q_{n-1}\left(1-\dfrac{\alpha}{2}\right)\right) = 1-\alpha$.

$\mathbb{P}_{m,\sigma^2}\left(m \in \left[\bar{X}_n \pm q_{n-1}\left(1-\dfrac{\alpha}{2}\right)\sqrt{\dfrac{V_n^*}{n}}\right]\right) = 1-\alpha$.

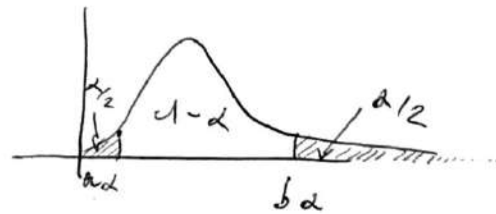
Donc $\hat{J}_n = \left[\bar{X}_n \pm q_{n-1}\left(1-\dfrac{\alpha}{2}\right)\sqrt{\dfrac{V_n^*}{n}}\right]$

- $\boxed{\text{Pour la variance } \sigma^2}$

→ $\boxed{\text{à moyenne } m \text{ connue}}$

Intervalle bilatéral

$\displaystyle\sum_{i=1}^{n}\left(\frac{X_i - m}{\sigma}\right)^2 \sim \chi^2(n)$



$\mathbb{P}_m\left(a_\alpha \leq \displaystyle\sum_{i=1}^{n}\left(\frac{X_i - m}{\sigma}\right)^2 \leq b_\alpha\right) = 1-\alpha$

- a_α quantile d'ordre $\dfrac{\alpha}{2}$ de la loi $\chi^2(n)$

$c_n\left(\dfrac{\alpha}{2}\right)$, $\mathbb{P}\left(U_n \leq c_n\left(\dfrac{\alpha}{2}\right)\right) = \dfrac{\alpha}{2}$

où $U_n \sim \chi^2(n)$

- b_α : quantile d'ordre $1-\dfrac{\alpha}{2}$ de la loi $\chi^2(n)$

$c_n\left(1-\dfrac{\alpha}{2}\right)$

et $c_n\left(\dfrac{\alpha}{2}\right) \leq \displaystyle\sum_{i=1}^{n}\frac{(X_i - m)^2}{\sigma^2} \leq c_n\left(1-\dfrac{\alpha}{2}\right)$

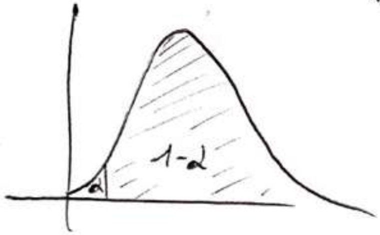
$$\Leftrightarrow \frac{\sum_{i=1}^{n}(X_i-m)^2}{c_n(1-\frac{\alpha}{2})} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n}(X_i-m)^2}{c_n(\frac{\alpha}{2})}$$

D'où l'intervalle de confiance σ^2 de niveau $1-\alpha$

$$\hat{I}_n = \left[\sum_{i=1}^{n} \frac{(X_i-m)^2}{c_n(1-\frac{\alpha}{2})} ; \frac{\sum_{i=1}^{n}(X_i-m)^2}{c_n(\frac{\alpha}{2})} \right]$$

Intervalle unilatéral

$$\mathbb{P}_m\left(c_n(\alpha) \leq \sum_{i=1}^{n} \frac{(X_i-m)^2}{\sigma^2}\right) = 1-\alpha$$



$$\Leftrightarrow \mathbb{P}_m\left(\sigma^2 \leq \frac{\sum (X_i-m)^2}{c_n(\alpha)}\right) = 1-\alpha$$

$$\hat{I}_n^{uni} = \left[0, \frac{\sum (X_i-m)^2}{c_n(\alpha)}\right].$$

→ ~~à moyenne inconnue~~

$$\frac{nV_n}{\sigma^2} = \sum \frac{(X_i-\bar{X}_n)^2}{\sigma^2} \sim \chi^2(n-1) \text{ sous } \mathbb{P}_{m,\sigma^2}$$

$$\mathbb{P}_{m,\sigma^2}\left(\sigma^2 \in \left[\frac{\sum (X_i-\bar{X}_n)^2}{c_{n-1}(1-\frac{\alpha}{2})} ; \frac{\sum (X_i-\bar{X}_n)^2}{c_n(\alpha/2)}\right]\right) = 1-\alpha$$

$$\mathbb{P}_{m,\sigma^2}\left(\sigma^2 \in \left[0, \frac{\sum (X_i-\bar{X}_n)^2}{c_{n-1}(\alpha)}\right]\right) = 1-\alpha$$

40 Int. unilat

@ Pour χ^2, $n=10$, $\alpha = 5\%$.

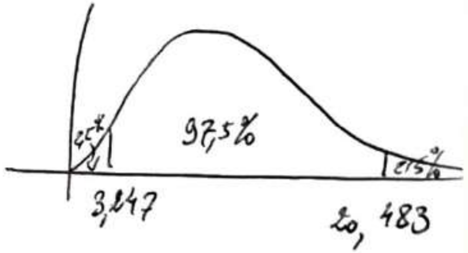
$c_n(\frac{\alpha}{2}) = 3{,}247$

$\mathbb{P}(U_{10} \geq 3{,}247) = 0{,}975$

$\Leftrightarrow \mathbb{P}(U_{10} \leq 3{,}247) = 0{,}025$

$c_n(1-\frac{\alpha}{2}) = 20{,}483$

$\mathbb{P}(U_{10} \geq 20{,}483) = 0{,}025$



où $U_{10} \sim \chi^2(10)$

C3 Simulations

OB : Simuler va ayant une certaine loi

Outil de base : ordinateur "simule des réalisat°s de loi uniforme sur $[0,1]$"

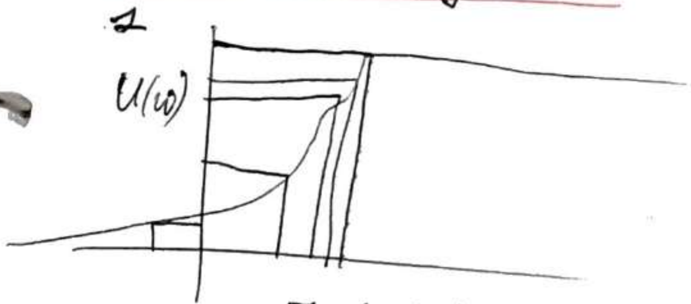
Générat^R de nbr aléatoires ou pseudo aléatoires (voir article Jean-Paul Delayae : aleas du hasard informatiq pr la science).

I / [M] Théoriq pr simuler var

basée sur l'invers° de la f dR°.

soit X une var de FdR F,

1) si F est bijective



$U \sim Unif([0,1])$

si $Y = F^{-1}(U) \Rightarrow X$ et Y ont la m̂ loi.

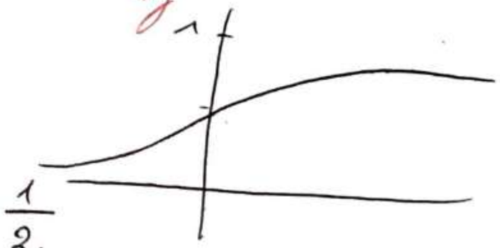
En effet $\forall t \in \mathbb{R}$,

$\mathbb{P}(Y \leq t) = \mathbb{P}(F^{-1}(U) \leq t) = \underbrace{\mathbb{P}(U \leq F(t))}_{=F(t)}$ or $F(t) \in [0,1]$

Donc Y & X ont la m̂ FdR.

Exemple : loi de Cauchy

$$f(x) = \frac{1}{\pi(1+x^2)}$$



$$F(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}$$

$$F(x) = y \iff \frac{1}{\pi}\arctan(x) + \frac{1}{2} = y \qquad y \in]0,1[$$

$$\iff \arctan(x) = \pi\left(y - \frac{1}{2}\right)$$

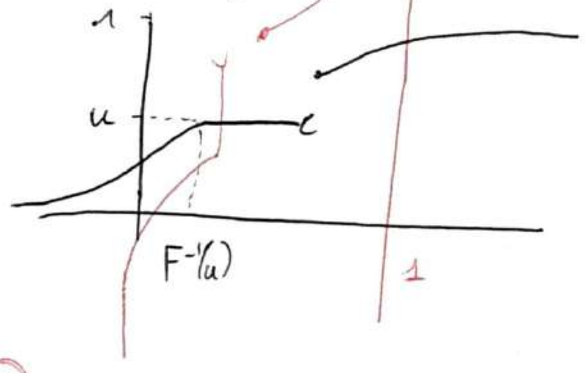
$$\iff x = \tan\left(\pi\left(y - \frac{1}{2}\right)\right)$$

$Y = \tan\left(\pi\left(U - \frac{1}{2}\right)\right)$ suit une loi de Cauchy.

② Inverse généralisé

On va construire l'inverse généralisé F^{-1} de tq $\mathbb{P}(F^{-1}(U) \leq t) = \mathbb{P}(U \leq F^{-1}(t))$

(D) L'inverse généralisé d'un FdR F est déf $\forall u \in]0,1[$, $F^{-1}(u) = \inf\{t \in \mathbb{R} \mid F(t) \geq u\}$



(Rq) • si F est cont & S^T ↗, on retrouve la f réciproq

• On pt mq F^{-1} ↗ & F^{-1} est cont à gauche.

(Th) si $U \sim \text{Unif}([0,1])$ & si F^{-1} est l'inverse généralisé d'un FdR F alors $F^{-1}(U)$ a pr FdR F.

Preuve : $\mathbb{P}(F^{-1}(U) \leq x) \overset{?}{=} \mathbb{P}(U \leq F(x)) = F(x)$

ON Mq $\forall u \in]0,1[$, $u \leq F(x) \Leftrightarrow F^{-1}(u) \leq x$.

$I_u = \{x \in \mathbb{R}, F(x) \geq u\}$, $F^{-1}(u) = \inf I_u$.

• $I_u \neq \emptyset$ car $\lim_{x \to \infty} F(x) = 1$.

• I_u est un intervalle de la forme $[b, \infty[$ ou $]b, \infty[$.

En effet, si $x \in I_u$, $F(x) \geq u$ & si $x' \geq x$ alors $F(x') \geq F(x) \geq u \Rightarrow x' \in I_u$.

F est ↗ $\Rightarrow [x, \infty[\subset I_u$.

• $I_u \neq \mathbb{R}$ car $\lim_{x \to -\infty} F(x) = 0$

• I_u est fermé : $\inf I_u \in I_u$.

soit $b = \inf I_u$, $\forall x > b$: $x \in I_u$ & $F(x) \geq u$. soit $(x_n)_{n \in \mathbb{N}}$ une suite de I_u qui cv vers b ($x_n > b$).

$F(x_n) \geq u \Longrightarrow F(b) \geq u \Longrightarrow b \in I_u$

car F est cont à droite.

(Ccl) $I_u = [b, \infty[$, $b = F^{-1}(u)$

$c_{n-1}(\alpha)$ —/

$b \in I_u$, $F(b) \geq u$

$\iff F^{-1}(F(u)) \geq u$ ...

& on a mqé que $F(x) \geq u \iff x \in I_u$.

$\iff x \geq b \iff x \geq F^{-1}(u)$

Conséquence : $\forall x \in \mathbb{R}$

$$\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$$

③ Exemples :

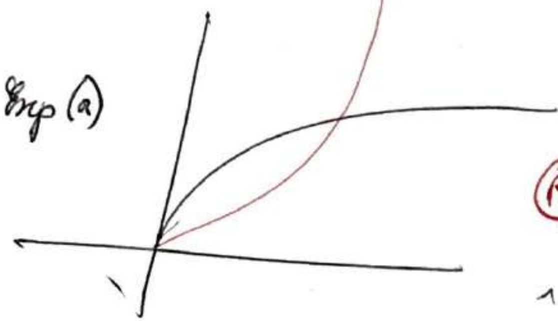
• loi exponentielle : $\mathcal{E}xp(a)$

$$F(t) = (1 - e^{-at}) \mathbb{1}_{[0,\infty[}(t)$$

$$F^{-1}(u) = -\frac{\ln(1-u)}{a}, \quad u \in]0,1[$$

$$F^{-1}(U) = \frac{-\ln(1-U)}{a} \sim \mathcal{E}xp(a)$$

ou $\frac{-\ln(U)}{a} \sim \mathcal{E}xp(a)$



• Loi de Weibull :

Fonction de survie :

$$G_{a,b,c}(x) = \exp\left(-\left(\frac{x-b}{c}\right)^a\right)$$

si $x \geq b$, $c > 0$, $b \geq 0$, $a > 0$

$X \sim Weib(a,b,c)$

$\iff \frac{X-b}{c} \sim \underbrace{Weib(a,0,1)}_{Y_a}$

$Z_a = b + cY_a \sim Weib(a,b,c)$

$$Y_a = (-\ln U)^{1/a} \sim Weib(a,0,1).$$

II / M spécifiqs pr lois usuelles

① Lois discrètes à support fini

• Loi de Bernoulli

si $X \sim Bern(p)$ et $U \sim Unif(]0,1[)$ alors $\mathbb{1}_{]0,p[}(U) \overset{loi}{=} X$ car si $Y = \mathbb{1}_{]0,p[}(U)$

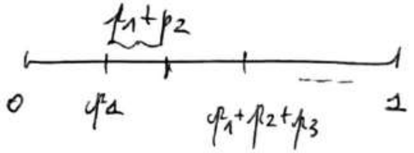
$$\mathbb{P}(Y=1) = \mathbb{P}(U \in]0,p[) = p.$$

Rq Inverse généralisée de la f de répartiθ.

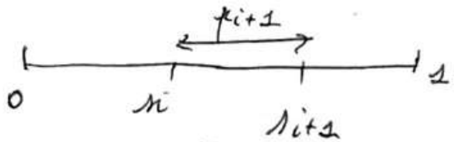
$F^{-1}(u) = \mathbb{1}_{]1-p,1[}(u) \sim Bern(p)$

Généralisat° : $X(\Omega) = \{x_n, n \in \mathbb{N}\}$

X prd nbr ⓓ de valeurs, $p_n = \mathbb{P}(X = x_n)$ $\forall n \in \mathbb{N}^*$



On pose $s_i = \sum_{k=1}^{n} p_k$



$i \in \mathbb{N}^*$, on pose $s_0 = 0$.

$]0,1[= \bigcup_{i=0}^{\infty}]s_i, s_{i+1}[$ car $\lim_{i \to \infty} s_i = 1 \Leftrightarrow \sum_{k=1}^{\infty} p_k = 1$.

$$Y = \sum_{i=0}^{\infty} x_{i+1} \mathbb{1}_{]s_i, s_{i+1}[}(U)$$

$\mathbb{P}(Y = x_{i+1}) = \mathbb{P}(U \in]s_i, s_{i+1}[) = s_{i+1} - s_i = p_{i+1}$

$\Rightarrow$ $Y \sim X$.

② Loi Binomiale (n,p) données.

Si $U_1, U_2, \dots, U_n$ iid de loi $\mathcal{U}nif(]0,1[)$.

alors on pose $Y_i = \mathbb{1}_{]0,p[}(U_i)$ $1 \leq i \leq n$

$Y_i \sim \mathcal{B}ern(p)$ & $(Y_i)_{1 \leq i \leq n}$ iid

$\Rightarrow Y = \sum_{i=1}^{n} Y_i$

$Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \mathbb{1}_{]0,p[}(U_i) = \text{card}(\{i \in \{1,\dots,n\} \mid U_i \in]0,p[\})$

$Y \sim \mathcal{B}in(n,p)$ ↗

Rq Conséq^ce intéressante : $p \mapsto \mathbb{P}_p(S_n \geq k)$ est ↗ $k \in \mathbb{N}$

si $S_n \sim \mathcal{B}in(n,p)$ sous $\mathbb{P}_p$.

Car sur $(\Omega, \tilde{\mathcal{F}}, \mathbb{P})$ sur lequel on définit $U_1, \dots, U_n$;

$\mathbb{P}_p(S_n \geq k) = \mathbb{P}_p\left(\sum_{i=1}^{n} \mathbb{1}_{]0,p[}(U_i) \geq k\right)$, or

si $p \leq p'$ alors $\sum_{i=1}^{n} \mathbb{1}_{]0,p[}(U_i) \leq \sum_{i=1}^{n} \mathbb{1}_{]0,p'[}(U_i)$

donc $\left\{\sum_{i=1}^{n} \mathbb{1}_{]0,p[}(U_i) \geq k\right\} \subset \left\{\sum_{i=1}^{n} \mathbb{1}_{]0,p'[}(U_i) \geq k\right\}$

& de $\mathbb{P}_p\left(\sum \mathbb{1}_{]0,p[}(U_i) \geq k\right) \leq \mathbb{P}_{p'}\left(\mathbb{1}_{]0,p'[}(U_i) \geq k\right)$

$\Rightarrow$ $\mathbb{P}_p$ croissante en p.

Loi multinomiales : $\mathcal{M}(n, p_1, \dots, p_d)$

Répartir en n boules ds d cases.

$N = (N_1, \dots, N_d) \sim \mathcal{M}ult(n, p_1, \dots, p_d)$ où

p_i : prob d'aller ds i° case, N_i : nbr boules ds i° case $(\forall i \in [\![1,d]\!])$.

$\sum N_i = n$, $X_k \in \{0,1\}^d$ représente $(1 \leq k \leq n)$ le n° de la case ds lq^{lle} se trouve la k-i° boule.

$X_k(\omega) = (0,0,\dots,0,\overset{j}{1},0,\dots,0)$ d-uplet $\leftarrow X_k(\omega)$ vaut ça de proba p_j.

alors $N = \sum_{k=1}^{n} X_k \in \mathbb{N}^d$, $e_i = (0,0,\dots,1,0,\dots,0)$

$\lambda_k = \sum_{i=0}^{d-1} e_{i+1} \mathbb{1}_{[s_i, s_{i+1}[}(U_k)$ où $s_i = \sum_{j=1}^{i} p_j$, $s_0 = 0$.

On simule de $U_1, \dots, U_n \sim \mathcal{U}nif(]0,1[)$

$$N = \sum_{k=1}^{n} \sum_{i=0}^{d-1} e_{i+1} \mathbb{1}_{[s_i, s_{i+1}[}(U_k)$$

Alors $N \sim \mathcal{M}ult(n, p_1, \dots, p_d)$.

3 Loi de Poisson ($\mathcal{P}ois(\theta > 0)$).

R$^\circ$ $X(\Omega) = \mathbb{N}$, $\mathbb{P}(X = k) = \frac{\theta^k}{k!} e^{-\theta}$.

L si $(E_i)_{i \in \mathbb{N}^*}$ est une suite de v.a de loi Exp de param θ, si l'on pose $S_1 = E_1$, $S_n = E_1 + \dots + E_n$

alors $\mathbb{P}(S_n \leq 1 < S_{n+1}) = \frac{\theta^n}{n!} e^{-\theta}$ $\forall n \in \mathbb{N}^*$.

Preuve $\mathbb{P}(S_n \leq 1 < S_{n+1}) = \mathbb{E}(\mathbb{1}_{\{E_1 + \dots + E_n \leq 1 \leq E_1 + \dots + E_{n+1}\}})$
$= \mathbb{E}(\varphi(E_1, \dots, E_{n+1}))$

φ une f compliquée.

Si X a pr densité f & Y a pr densité g et indep $\Rightarrow (X,Y)$ a pr densité sur $\mathbb{R}^2$.

h tq $h(x,y) = f(x) g(y)$; le vect^r $(E_1, \dots, E_{n+1})$ a pr densité $f_{n+1}(x_1, \dots, x_{n+1}) = \prod_{i=1}^{n+1} f_\theta(x_i)$ sur $\mathbb{R}^{n+1}$.

où $f_\theta(x) = \theta e^{-\theta x} \mathbb{1}_{]0,\infty[}(x)$

$\Rightarrow f_{n+1}(x_1, \dots, x_{n+1}) = \theta^n e^{-\theta \sum_{i=1}^{n+1} x_i} \prod_{i=1}^{n} \mathbb{1}_{]0,\infty[}(x_i)$ $(\uparrow)$

où $A_{n+1} = \{(x_1, \dots, x_{n+1}) \in \mathbb{R} \}$

d'où $\mathbb{P}(S_n \leq 1 < S_{n+1})$

$$= \int_{\mathbb{R}^{n+1}} \mathbb{1}_{A_{n+1}}(x_1,\dots,x_{n+1})\, f_{n+1}(x_1,\dots,x_{n+1})\, dx_1 \dots dx_{n+1}$$

$$= \int_{A_{n+1}} f_{n+1}(x_1,\dots,x_{n+1})\, dx_1 \dots dx_{n+1} = \int_{A_{n+1}} \theta^n e^{-\theta \sum_{i=1}^{n+1} x_i}\, \mathbb{1}_{(\mathbb{R}^+)^{n+1}}(x_1,\dots,x_{n+1})\, dx_1 \dots dx_{n+1}$$

chgt de var bijectif $y_1 = x_1,\ y_2 = x_1 + x_2, \dots, y_n = x_1 + \dots + x_n,\ y_{n+1} = x_1 + \dots + x_n + x_{n+1}$

$$= \int_{\varphi(A_{n+1})} \theta^n e^{-y_{n+1}}\, \mathbb{1}_{(\mathbb{R}^+)^{n+1}}(y_1, y_2, \dots, y_{n+1} - y_n)\, dy_1 \dots dy_{n+1}$$

et $\varphi(A_{n+1}) = \{(y_1,\dots,y_{n+1}) \in \mathbb{R}^{n+1} \mid y_n \leq 1 < y_{n+1}\}$

$$= \theta^n \int_{\mathbb{R}^{n+1}} \mathbb{1}_{\{(y_1,\dots,y_{n+1}) \in \mathbb{R}^{n+1} \mid 0 \leq y_1 \leq \dots \leq y_n \leq 1 < y_{n+1}\}} \times e^{-\theta y_{n+1}}\, dy_1 \dots dy_n\, dy_{n+1}$$

$$= \theta^n \int_{\mathbb{R}^n} \mathbb{1}_{B_n}(y_1,\dots,y_n)\, \mathbb{1}_{]1,\infty[}(y_{n+1})\, e^{-\theta y_{n+1}}\, dy_1 \dots dy_n\, dy_{n+1}$$

où $B_n = \{(y_1,\dots,y_n) \in \mathbb{R}^n \mid 0 \leq y_1 \leq \dots \leq y_n < 1\}$ simplex

$$= \theta^n \lambda_n(B_n) \int_1^\infty e^{-\theta y_{n+1}}\, dy_{n+1} \quad \& \quad \lambda_n(B_n) = \frac{1}{n!}$$

$n!$ simplexes en permuttant (cf wiki)

$$= \frac{\theta^n}{n!} e^{-\theta}$$ ← valable pr $n=0$ aussi

ccl $\mathbb{P}(X = n) = \mathbb{P}(S_n \leq 1 < S_{n+1})$.

Dc on pt définir $Y = \sum_{n=0}^{\infty} n \cdot \mathbb{1}\{S_n \leq 1 < S_{n+1}\}$

où on le rappelle $S_n = E_1 + \dots + E_n$ ↗

Algo de simulat°

Simulat° des $(E_i)_{i \in \mathbb{N}}$, $E_i = \frac{-1}{\theta} \ln(U_i) \sim \mathcal{E}xp(\theta)$.

$$S_n \leq 1 < S_{n+1} \iff -\sum_{i=1}^{n} \frac{\ln(U_i)}{\theta} \leq 1 < -\sum_{i=1}^{n+1} \frac{\ln(U_i)}{\theta}$$

$$\iff \sum_{i=1}^{n+1} \ln(U_i) < -\theta \leq \sum_{i=1}^{n} \ln(U_i)$$

$$\iff \prod_{i=1}^{n+1} U_i < e^{-\theta} \leq \prod_{i=1}^{n} U_i$$ ↗

dc $Y = \inf\{k \in \mathbb{N} \mid \prod_{i=1}^{k+1} U_i < e^{-\theta}\}$

Conséquence $f : \theta \longrightarrow \mathbb{P}_\theta(X \geq k)$ si $X \sim \mathcal{P}oi(\theta)$ est ↗ en θ.

En effet si $Y(\theta) = \inf\{k \in \mathbb{N} \mid \prod U_i < e^{-\theta}\}$

$\mathbb{P}_\theta(X \geq k) = \mathbb{P}(Y(\theta) \geq k)$.

si $\theta < \theta'$, $e^{-\theta} > e^{-\theta'}$ $\underbrace{\{k \in \mathbb{N},\ \prod_{i=1}^{k+1} U_i < e^{-\theta}\}}_{(1)} \subset \underbrace{\{k \in \mathbb{N} \mid \prod_{i=1}^{k+1} U_i < e^{-\theta}\}}_{(2)}$

$\Rightarrow$ l'inf (2) $\geq$ inf (1) d'où la croissance voulue

si $(U_i)_{i \in \mathbb{N}^*}$ suite de va iid de loi $\mathcal{U}nif(]0,1[)$

46 Mq $M_n = \prod U_i$ ps vers 0.

④ Loi géométriq

$$Y = \inf \{ k \in \mathbb{N}^* \mid U_k \leq p \} \text{ où } Y \sim \mathcal{G}eom(p)$$

⚠ Etudier monotonie de $\theta \longmapsto \mathbb{P}_\theta (X \geq k)$ si $X \sim \mathcal{G}eom(\theta)$ sous $\mathbb{P}_\theta$.

Autre $\boxed{M}$ de simulat°

Si $E \sim \mathcal{E}xp(\theta) \Rightarrow \lfloor E \rfloor \sim \mathcal{G}eom(1-e^{-\theta})$ changer de var pr param custom

$$Y = \left\lceil \frac{\ln U}{\ln(1-p)} \right\rceil \sim \mathcal{G}eom(p)$$

⑤ Loi Gaussienne

$\boxed{M}$ Box-Müller

Ⓛ si $U_1, \dots, U_2 \underset{\text{va indep}}{\sim} \mathcal{U}nif(]0,1[)$ alors si

$$\left. \begin{array}{l} X = \sqrt{-2\ln U_1} \cos(2\pi U_2) \\ Y = \sqrt{-2\ln U_2} \sin(2\pi U_2) \end{array} \right. \Rightarrow (X,Y) \sim \mathcal{N}(0, I_2)$$

$$\Rightarrow \begin{array}{l} X \sim \mathcal{N}(0,1) \\ Y \sim \mathcal{N}(0,1) \end{array} \text{ et } X, Y \text{ indep}$$

NB $(x,y) = g(u_1,u_2)$
$(u_1,u_2) = g^{-1}(x,y)$
$du_1 du_2 = J_{g^{-1}}(x,y) dx dy$

Preuve : $\mathbb{E}(\varphi(X,Y)) = \int_{\mathbb{R}^2} \varphi(x,y) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dx dy$
borel, bornée — densité

or $\mathbb{E}(\varphi(X,Y)) = \mathbb{E}(\varphi \circ g(U_1,U_2))$

$$= \int_{]0,1[^2} \varphi(\sqrt{-2\ln U_1} \cos(2\pi U_2) \sqrt{-2\ln U_2} \sin(2\pi U_2)) dU_1 dU_2$$

@ ça va bien

Reprenons la preuve :

soit $\varphi : \mathbb{R}^2 \to \mathbb{R}$ mesurable bornée.

OB : Mq $\mathbb{E}(\varphi(X,Y)) = \int_{\mathbb{R}^2} \varphi(x,y) f(x,y) dx dy$

$$\mathbb{E}[\varphi(X,Y)] = \mathbb{E}[\varphi(\sqrt{-2\ln U_1} \cos(2\pi U_2) ; \sqrt{-2\ln U_1} \sin(2\pi U_2))]$$

$$= \int_{]0,1[^2} \varphi(\sqrt{-2\ln u_1} \cos(2\pi u_2) ; \sqrt{-2\ln u_1} \sin(2\pi u_2)) du_1 du_2$$

$$= \int_{]0,1[^2} \varphi(g(u_1,u_2)) du_1 du_2 = \int_{\mathbb{R}^2 \setminus \{(0,0)\}} \varphi \circ g \circ g^{-1}(x,y) . J_{g^{-1}}(x,y) dx dy$$

où $g :]0,1[^2 \longrightarrow \mathbb{R}^2 \setminus \{(0,0)\}$
$(u_1,u_2) \longmapsto (\sqrt{-2\ln u_1} \cos(2\pi u_2), \sqrt{-2\ln u_1} \sin(2\pi u_2))$

g est un $\mathcal{C}^1$ difféomorphisme. $J_{g^{-1}}(x,y) = \frac{1}{J_g(g^{-1}(x,y))}$

$$= \int_{\mathbb{R}^2} \varphi(x,y) \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

III / Algorithme du rejet

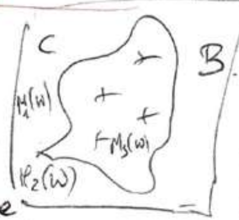
Idée : On sppse qu'on sait simuler un cert1s vctR aléatoire M. On en simule plsrs $M_1, \dots, M_n$ jsqu'à ce qu'une cert1ne conditθ soit réalisée

$\Rightarrow M_T$ ↳ T aléatoire et $M_T \sim$ loi qu'on cherche à simuler.

1) Simulatθ d'un vectR aléatoire de loi uniforme

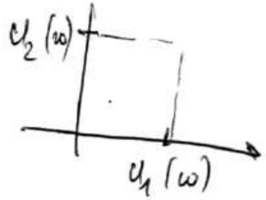
$B \in \mathcal{B}(\mathbb{R}^d)$

OB : Simuler un vectR de loi unif sur B.



On sait simuler des vctRs de loi uniforme sur $[0,1]^d$.

$(U_1, \dots, U_d)$ où $U_i \sim Unif(\mathbb{R}^d)$
$(U_i)_i$ iid.



Sur un pavé qq :

$$\prod_{i=1}^{d} [a_i, b_i]$$

$$V_i = a_i + (b_i - a_i) U_i \sim Unif([a_i, b_i])$$

Principe :

soit C tq $B \subset C$ (si B est borné on pt prendre un pavé pour C) tq on suppose qu'on sait simuler une uniforme sur C

$\Rightarrow$ on simule $M_1, \dots, M_n$ des vectRs de loi Uniforme sur C jusqu'à avoir un vectR $M_i \in B$.

$\Rightarrow M_T \sim Unif(B)$.

Prop : soit $(M_n)_{n \in \mathbb{N}^*}$ une suite de vctRs aléa. à valRs ds $\mathbb{R}^d$ indépendants & de m loi μ.

soit $B \in \mathcal{B}(\mathbb{R}^d)$, $\mu(B) > 0$

$\forall \omega \in \Omega$, $T(\omega) = \inf\{ i \in \mathbb{N}^*, M_i(\omega) \in B \}$.

On définit $M_T(\omega) = \begin{cases} M_T(\omega) & \text{si } T(\omega) < \infty \\ 0 & \text{si } T(\omega) = \infty \end{cases}$ le vctR aléatR alors a) $\mathbb{P}(T<\infty)=1$ et si $T' = T\mathbb{1}_{\{T<\infty\}}$ alors $T' \sim Geom(\mu(B))$.

b) M_T est un vect^R aléat^R de $\mathbb{R}^d$ dont la loi ϑ est $\mu(\cdot \mid B)$.

ie $\forall A \in \mathcal{B}(\mathbb{R}^d)$, $\vartheta(A) = \mathbb{P}(M_T \in A) = \mu(A \mid B) = \frac{\mu(A\cap B)}{\mu(B)}$

Conséq^{ce} :

si μ est la Unif sur C alors

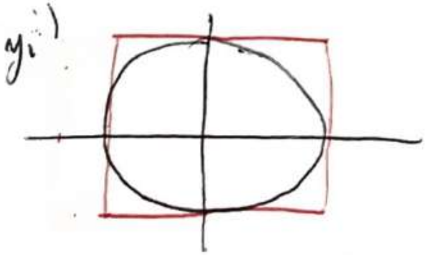
$$\vartheta(A) = \frac{\mu(A\cap B)}{\mu(B)} = \frac{\lambda_d(A\cap B\cap C)/\lambda_d(C)}{\lambda_d(B\cap C)/\lambda_d(C)}$$

$$= \frac{\lambda_d(A\cap B)}{\lambda_d(B)}$$ car $B \subset C \Rightarrow \vartheta$ est la loi unif sur B.

$\Rightarrow M_T \sim Unif(B)$ si $M_1, \dots, M_n \dots$ st iid $Unif(C)$ et $T = \inf\{i \in \mathbb{N}^*, M_i \in B\}$.

@ $B = \{(x,y) \in \mathbb{R}^2, x^2+y^2 \leq 1\}$.

$M_i = (x_i, y_i)$



$M_1, \dots, M_n$ $Unif([-1,1]^2)$

$T = \inf\{i \in \mathbb{N}^* \; M_i \in B$
$= \inf\{i \in \mathbb{N}^*, x_i^2 + y_i^2 \leq 1\}$ $\quad T \sim Geo(p)$.

$\boxed{Rq}$ $U_1 \sim Unif([0,1])$, idem U_2.

$X = U_2 \cos(2\pi U_1)$ $\quad Y = U_2 \sin(2\pi U_1)$

Calculer la loi de (X,Y)

Preuve Prop :

- Mesurabilité de T & T'

$\{T=k\} = \{\inf\{i \in \mathbb{N}^* : M_i \in B\} = k\}$

$= \bigcap_{i=1}^{k-1} \{M_i \notin B\} \cap \{M_k \in B\}$.

$= \bigcap_{i=1}^{k-1} M_i^{-1}(B)^c \cap M_k^{-1}(B) \in \mathcal{F}$ car les (M_i) st vect^{rs} aléa

$\mathbb{P}(T=k) = \mathbb{P}\left(\bigcap_{i=1}^{k-1} \{M_i \notin B\} \cap \{M_k \in B\}\right)$

$= \prod_{i=1}^{k-1} \mathbb{P}(M_i \notin B)\, \mathbb{P}(M_k \in B)$ $\mathcal{P}$ indépdce des $(M_i)_{i\in\mathbb{N}^*}$

$\mathbb{P}(M_k \in B) = \mu(B)$ car μ est la loi de M_k

$\Rightarrow \mathbb{P}(T=k) = \mu(B^c)^{k-1}\mu(B) = \mu(B)(1-\mu(B))^{k-1}$ $\forall k \in \mathbb{N}^*$

$= p(1-p)^{k-1}$ $\quad p = \mu(B)$.

$\{T'=0\} = \{T=\infty\} = \bigcap_{i=1}^{\infty} \{M_i \notin B\}$

$\boxed{1^e M}$ • $\bigcap_{i=1}^{\infty} \{M_i \notin B\} = \bigcap_{n=1}^{\infty} A_n$ où $A_n = \bigcap_{i=1}^{n} \{M_i \notin B\}$

les A_n st $\downarrow$: $\mathbb{P}(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} (\mu(A))^n = 0$, $\mu(B)>0$

$2^e \boxed{M}$

$\bigcap_{i=1}^{\infty} \{M_i \notin B\} \subset \bigcap_{i=1}^{n} \{M_i \notin B\} \quad \forall n \in \mathbb{N}^*$

$\mathbb{P}(\bigcap_{i=1}^{\infty} \{M_i \notin B\}) \leq \mathbb{P}(\bigcap_{i=1}^{n} \{M_i \notin B\})$

$= (\mu(B^c))^n \quad \forall n \in \mathbb{N}^*$

$\Rightarrow \mathbb{P}(\bigcap_{i=1}^{\infty} \{M_i \notin B\}) = 0$

ccl

$\mathbb{P}(T=T') = 1$. & T' v.a à val^{rs} ds $\mathbb{N}$ (finie)

& T et T' suivent $Geo(p)$ $p = \mu(B)$

b) mesurabilité & loi de M_T, soit $A \in \mathcal{B}(\mathbb{R}^d)$

$M_T^{-1}(A) = \{M_T \in A\} = \{w \in \Omega : M_{T(w)}(w) \in A\}$

$= \{w \in \Omega : \exists k \in \mathbb{N}^*$ tq $T(w) = k$ & $M_k(w) \in A$.

$= \{w \in \Omega : \exists k \in \mathbb{N}^*$ tq $T(w) = k$ & $M_k(w) \in A\}$

$= \bigcup_{k\in\mathbb{N}^*} \underbrace{\{T=k, M_k \in A\}}_{\{w\in\Omega : T(w)=k, M_k(w)\in A\}}$

$\{w \in \Omega, T(w)=k, M_k(w) \in A\}$:

$= \{w \in \Omega, M_1(w) \notin B, M_{k-1}(w) \notin B, M_k(w) \in B$

et $M_k(w) \in A\}$.

$\bigcap_{i=1}^{k-1} \{M_i \notin B\} \cap \{M_k \in A \cap B\}$

• si $k=\infty$, $M_\infty = 0$

$\{w \in \Omega : T(w) = \infty, M_\infty^{(w)} \in A$.

$= \begin{cases} \bigcap_{i=1}^{\infty} \{M_i \notin B\} & \text{si } 0 \in A \\ \emptyset & 0 \notin A \end{cases}$

dc $\mathbb{P}(M_T \in A) = \sum_{k\in\mathbb{N}^*} \mathbb{P}(T=k, M_k \in A)$

$= \sum_{k\in\mathbb{N}^*} \mathbb{P}(\bigcap_{i=1}^{k-1} \{M_i \notin B\} \cap \{M_k \in A\cap B\})$

$= \sum_{k=1}^{\infty} (\mu(B^c))^{k-1} \mu(A\cap B) = \frac{\mu(A\cap B)}{\mu(B)} = \mu(A|B)$.

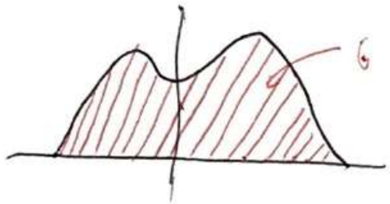
NB : $\mu(\cdot) = \mathbb{P}(M \in \cdot)$

2) Simulation lois à densité

Idée : Se servir de l'hypographe de la densité & appliquer la $\boxed{M}$ précédente.

Densité f.

$G = \{(x,y) \in \mathbb{R}^2,\ 0 < y \leq f(x)\}$ ← va jouer le rôle de B.



Prop 1 : si $(Z,Y) \sim \mathrm{Unif}(G)$ alors Z a pr densité f.

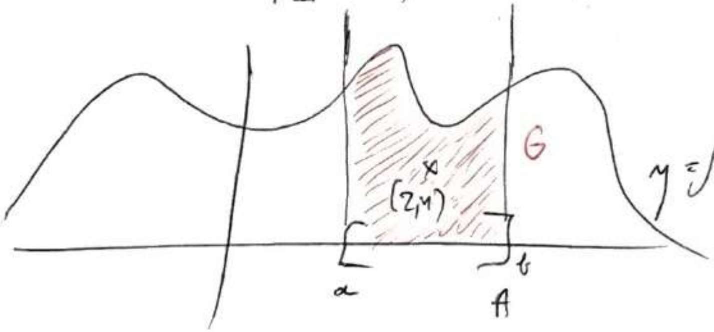
Rq $\lambda_2(G) = 1$.

Preuve : soit $A \in \mathcal{B}(\mathbb{R})$,

$\mathbb{P}(Z \in A) = \mathbb{P}(Z \in A, Y \in \mathbb{R}) = \mathbb{P}((Z,Y) \in A \times \mathbb{R})$

$$= \frac{\lambda_2(A \times \mathbb{R} \cap G)}{\lambda_2(G)} = \int_a^b f(t)\,dt$$

$A = [a,b]$



Prop 2 : soit X une va de densité g.

Posons $M = (X, c \cdot g(X) \cdot U)$ où $U \sim \mathrm{Unif}(]0,1[)$ indépendante de X alors M suit une loi Unif sur $H = \{(x,y) \in \mathbb{R}^2 : 0 \leq y \leq c\,g(x)\}$

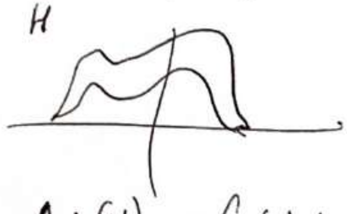
Preuve : soit $\varphi : \mathbb{R}^2 \to \mathbb{R}$ mesurable bornée

$$\mathbb{E}[\varphi(M)] = \int_{\mathbb{R} \times]0,1[} \varphi(x, c.g(x)u)\, g(x)\, dx\,du$$

$$= \int_{\mathbb{R}} \left(\int_0^1 \varphi(x, cg(x)u)\, cg(x)\, du \right) \frac{1}{c}\, dx$$

$$= \int_{\mathbb{R}} \int_0^{cg(x)} \varphi(x,y)\, dy\, \frac{1}{c}\, dx \qquad y = cg(x)u$$

$$= \iint_H \varphi(x,y) \frac{1}{c}\, dx\,dy \quad \text{où } c = \lambda d(H)$$



$$= \iint_H \varphi(x,y) \underbrace{\frac{1}{c} \mathbb{1}_H(x,y)}_{\text{densité de } (X,Y) = M} dx\,dy$$

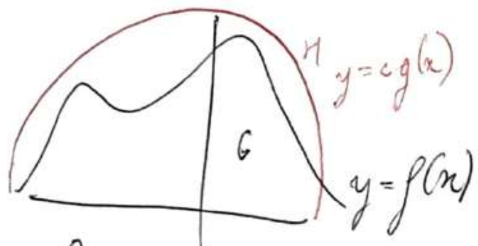
$\lambda d(H) = c \int_{\mathbb{R}} g(x)\,dx.$

$M \sim \mathrm{Unif}(H)$

Applicat° : $(X_i)_{i\in\mathbb{N}^*}$ de densité g

$(U_i)_{i\in\mathbb{N}^*}$ Unif$(]0,1[)$

$M_i = (X_i, cg(X_i)U_i) \sim \text{Unif}(H)$ (p prop 1)



Il suffit de prendre $c>1$ tq $f(x) \le c\,g(x)$.

On applique l'algorithme du rejet

$T = \inf\{i\in\mathbb{N}^* \ M_i \in G\}$

$M_i = (X_i, cg(X_i)U_i) \in G \iff cg(X_i)U_i \le f(X_i)$

→ si $T = \inf\{i\in\mathbb{N}^*, cg(X_i) \le f(X_i)\}$

alors $M_T \sim \text{Unif}(G) \Rightarrow X_T$ a pr densité f. (prop 1).

@ Loi gamma

Loi de densité

$$f_{a,b}(t) = \frac{b^a}{\Gamma(a)} t^{a-1} e^{-bt} \mathbb{1}_{]0,\infty[}(t)$$

Idée : g densité d'une loi de Weibull.

$$g_a(t) = a t^{a-1} e^{-t^a} \mathbb{1}_{[0,\infty[}(t)$$

Exemple d'applicat° :

- loi gamma : $f_a(t) = \frac{1}{\Gamma(a)} t^{a-1} e^{-t} \mathbb{1}_{]0,\infty[}(t)$

$0<a<1$ f_a non bornée, $f_a(t) \underset{0}{\sim} \frac{1}{\Gamma(a)} t^{a-1}$.

OB Appliquer $\boxed{M}$ de Rejet, majorer f_a par $c.g_a$ où g_a est facile à simuler.

On use la loi de Weibull dont la densité est

$$g_a(t) = a.t^{a-1} e^{-t^a} \mathbb{1}_{]0,\infty[}(t)$$

Sa f de survie est $G_a(t) = e^{-t^a} \mathbb{1}_{[0,\infty[}(t) + \mathbb{1}_{]-\infty,0[}(t)$

$F_a(t) = 1 - e^{-t^a}$, $t \ge 0$.

$F_a(u)^{-1} = [-\ln(1-u)]^{1/a}$, si $U \sim \text{Unif}(]0,1[)$

alors $F_a^{-1}(U) = [-\ln(1-U)]^{\frac{1}{a}}$ a pr densité g_a

$X = [-\ln U]^{\frac{1}{a}}$ a pr densité g_a.

Trouver c tq $f_a(t) \le c.g_a(t)$ $\forall t \in \mathbb{R}$.

$t>0$, $\dfrac{f_a(t)}{g_a(t)} = \dfrac{1}{a\Gamma(a)} e^{-t+t^a} = \dfrac{1}{a\Gamma(a)} e^{-t(1-t^{a-1})}$

Le sup en $t \in]0,\infty[$ est atteint

$$t_a = \left(\frac{1}{a}\right)^{\frac{1}{a-1}} \Rightarrow c_a = \frac{1}{e\Gamma(a)} e^{-t_a(1-t_a^{a-1})}$$

$$c_a = \frac{e^{a^{\frac{a}{a-1}}(1-a)}}{\Gamma(a+1)} \qquad (a\Gamma(a) = \Gamma(a+1))$$

Algorithme du Rejet :

On simule $X_i = (-\ln U_i)^{\frac{1}{a}}$

indep. $V_i \sim Unif(]0,1[)$

On s'arrête à i_0 tq $c_a g_a(X_{i_0}) V_{i_0} \leq f_a(X_{i_0})$

aléatoire

alors X_{i_0} a pr densité f_a.

• Autre exemple :

→ loi normale : $f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$

Majorer (à constante près) f par la densité g d'une loi de Cauchy (qu'on sait simuler en inversant la fdr).

3) Simulation d'une loi discrète :

On cherche à simuler Z de loi discrète sur $\mathbb{N}$ et on note

$$f : \mathbb{N} \to \mathbb{R}$$
$$k \mapsto f(k) = \mathbb{P}(Z=k)$$

On suppose que l'on sait simuler X de loi discrète sur $\mathbb{N}$ tq $g(k) = \mathbb{P}(X=k)$; $0 \leq f(k) \leq c.g(k)$, $c>1$, $\forall k \in \mathbb{N}$

Prop : soit $(X_i)_{1\leq i\leq n}$ une va iid de densité g p/ à la mesure de comptage ($P_X = \sum_{k\in\mathbb{N}} g(k)\delta_k$) et $(U_i)_{i\in\mathbb{N}^*}$ une loi uniforme sur $[0,1]$ indépendante des $(X_i)_{i\in\mathbb{N}^*}$.

alors si $T = \inf\{i \in \mathbb{N}^* \text{ tq } c.g(X_i)U_i \leq f(X_i)\}$.

X_T a pr densité f p/ à la mesure de comptage sur $\mathbb{N}$.

Preuve : m̂ chose que ds le cas à densité sur $\mathbb{R}$ (à adapter)

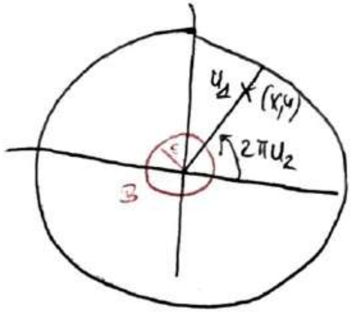
Rq U_1 & U_2 2 va ind. de loi $Unif([0,1])$

$$\begin{cases} X = U_1 \cos(2\pi U_2) \\ Y = U_1 \sin(2\pi U_2) \end{cases}$$

$D = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \leq 1\}$

↓

M de Box-Muller



1) $(X,Y) \in D$ mais ne suit pas une loi uniforme sur D.

@ $\mathbb{P}((X,Y) \in B(0,\varepsilon)) = \mathbb{P}(U_1\cos(2\pi U_2), U_1 \sin(2\pi U_2)) \in B(0,\varepsilon))$ sur $[0,1]$.

$B(0,\varepsilon) = \{(x^\varepsilon, y^\varepsilon) \in \mathbb{R}^2, x^2+y^2 \leq \varepsilon^2\}$.

$= \mathbb{P}(U_1 \in [0,\varepsilon]) = \varepsilon$

Or si (X,Y) suivait une loi uniforme sur D, on aurait $\mathbb{P}((X,Y) \in B(0,\varepsilon)) = \dfrac{\lambda_2(B(0,\varepsilon) \cap D)}{\lambda_2(D)}$

$= \dfrac{\lambda_2(B(0,\varepsilon))}{\lambda_2(D)} = \dfrac{\pi\varepsilon^2}{\pi}$

2) Mq la densité de (X,Y) est $f(x,y) = \dfrac{1}{\pi\sqrt{x^2+y^2}} \mathbb{1}_D(x,y)$

3) Trouver $\boxed{M}$ de simulation de loi uniforme sur D :

$$\begin{cases} X = g(U_1)\cos(2\pi U_2) \\ Y = g(U_1)\sin(2\pi U_2) \end{cases}$$ trouver g.

Correct° du Partiel DS$_1$

Q° de cours /2

Ex 2 /5

$n = 700$ personnels ds 2 restos Univ.

$S_n^{(1)}$: nbr de personnels RU Barrois

$S_n^{(2)}$: Sully.

~~On cherche N tq $\mathbb{P}(S_n^{(1)} \geq N) \leq 0{,}1$~~

Si on met N places ds chq RU, l'évent

A = "chq personnel trouve une place".

$$A = \{S_n^{(1)} \leq N\} \cap \{S_n^{(2)} \leq N\}$$

$$A^c = \{S_n^{(1)} > N\} \cup \{S_n^{(2)} > N\}.$$

$$\mathbb{P}(A) \geq 0{,}9 \iff \mathbb{P}(A^c) \leq 0{,}1.$$

$$\iff \mathbb{P}(\{S_n^{(1)} > N\} \cup \{S_n^{(2)} > N\}) \leq 0{,}1,$$

si on choisit N tq $\mathbb{P}(S_n^{(1)} > N) \leq 0{,}05$

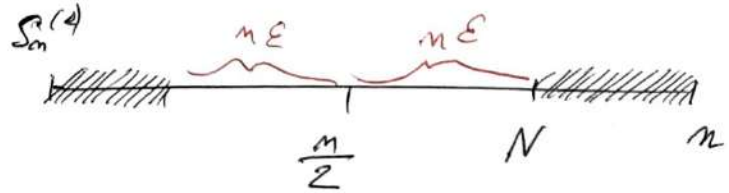
et $\mathbb{P}(S_n^{(2)} > N) \leq 0{,}05 \Rightarrow \mathbb{P}(A^c) \leq 0{,}1.$

$$S_n^{(1)} + S_n^{(2)} = n.$$

$$\mathbb{P}(S_n^{(2)} > N) = \mathbb{P}(n - S_n^{(1)} > N) = \mathbb{P}(S_n^{(1)} < n-N)$$

$\mathbb{P}\left(S_n^{(1)} > N\right) \le 0{,}05$. $\quad \mathbb{P}\left(S_n^{(1)} < n-N\right) \le 0{,}05$

$S_n^{(1)} \sim \mathrm{Bin}\left(n, \frac{1}{2}\right)$.



$N = \frac{n}{2} + n\varepsilon \qquad (= 350 + 700\varepsilon)$

$$(*) \Leftrightarrow \begin{cases} \mathbb{P}\left(S_n^{(1)} > \frac{n}{2} + n\varepsilon\right) \le 0{,}05 \\ \mathbb{P}\left(S_n^{(1)} < \frac{n}{2} - n\varepsilon\right) \le 0{,}05 \end{cases}$$

$$\Leftrightarrow \mathbb{P}\left(\left|S_n^{(1)} - \frac{n}{2}\right| > n\varepsilon\right) \le 0{,}1.$$

Par Hoeffding,

$$\mathbb{P}\left(\left|S_n^{(1)} - \frac{n}{2}\right| > n\varepsilon\right) \le 2e^{-2n\varepsilon^2}$$

On cherche $\varepsilon > 0$ tq $2e^{-2n\varepsilon^2} \le 0{,}1$.

$$\varepsilon \geq \sqrt{-\frac{1}{2n}\ln(0{,}1/2)}$$

$n = 700$: $\varepsilon = 0{,}04626$.

$$\underline{N = 383 = 350 + 700\varepsilon.}$$

2) Test : si θ est la proba de choisir Barrois

H_0 : " — Sully $\qquad \theta \le \frac{1}{2}$

H_1 : "le RU Barrois est le préféré des personnels $\qquad \theta > \frac{1}{2}$

$S_n^{(1)}$ est le nbr de personnels q vont à Barrois.

$$\mathcal{R}_0 = \{S_n^{(1)} \geq k\}.$$

où k est tq $\sup_{\theta \le \frac{1}{2}} \mathbb{P}_\theta(\mathcal{R}_0) \le 0{,}05$

or $\sup_{\theta \le \frac{1}{2}} \mathbb{P}_\theta(S_n^{(1)} \geq k) = \mathbb{P}_{\frac{1}{2}}(S_n^{(1)} \geq k)$.

$$k = \min\left\{ \ell \in \{0, \dots, n\} : \mathbb{P}_{\frac{1}{2}}(S_n^{(1)} \geq \ell) \le 0{,}05 \right\}$$

Or $\mathbb{P}_{\frac{1}{2}}(S_n^{(1)} \geq k) \le 0{,}05$ conduit p. Hoeffding à $k = 383$.

$\mathcal{R}_0 = \{S_{700}^{(1)} \geq 383\}$ or $S_{700}^{(1)}(\omega) = 399$

$\Rightarrow \omega \in \mathcal{R}_0$ dc on conclut que le Barrois est le RU préféré des personnels.

6. __Ex3__ "Les stats expliquées à mon chat".

p : proba d'opinion "oui"

q : proba de répondre "oui" à l'issue du lancer de dé

A = "le lancer tombe sur 6".

1) B = "répondre oui"

$$\mathbb{P}(B) = q = \mathbb{P}(B|A).\,\mathbb{P}(A) + \mathbb{P}(B|A^c)\,\mathbb{P}(A^c)$$
$$= p\,\frac{1}{6} + (1-p)\frac{5}{6} = \frac{5}{6} - \frac{2}{3}p.$$

2) $(X_i)_{1\leq i\leq n}$, $X_i \sim \mathcal{B}ern(q)$

3) $\hat{q}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

Sans biais $\mathbb{E}_p(\hat{q}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_p(X_i) = q$

fort^{mt} consist^t LFGN $\rightarrow$ $\hat{q}_n \xrightarrow[n\to\infty]{\mathbb{P}_p\ p.s} q$.

__Rq quadratiq__

$$\mathbb{E}_p((\hat{q}_n - q)^2) = Var_p(\hat{q}_n) = \frac{1}{n^2}Var_p\Big(\sum_{i=1}^{n} X_i\Big)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n} Var_p(X_i) = \frac{q(1-q)}{n}.$$

4) $q = \frac{5}{6} - \frac{2}{3}p \iff p = \frac{5}{4} - \frac{3}{2}q$

Empirique $\hat{p}_n = \frac{5}{4} - \frac{3}{2}\hat{q}_n$

Rq $\hat{p}_n$ n'est pas une moyenne empiriq de va de loi de Bernoulli.

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{5}{4} - \frac{3}{2}X_i\Big) = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

$$Y_i = \frac{5}{4} - \frac{3}{2}X_i$$

$$\mathbb{E}_p(\hat{p}_n) = \frac{5}{4} - \frac{3}{2}\mathbb{E}_p(\hat{q}_n) = \frac{5}{4} - \frac{3}{2}\Big(\frac{5}{6} - \frac{2}{3}p\Big) = p$$

donc __sans biais__.

- __fortement consistant__: $\hat{q}_n \xrightarrow[n\to\infty]{\mathbb{P}_p\ p.s} q$

$$\Rightarrow \frac{5}{4} - \frac{3}{2}\hat{q}_n \xrightarrow[n\to\infty]{\mathbb{P}_p\ p.s} \frac{5}{4} - \frac{3}{2}q$$

d'où $\hat{p}_n \xrightarrow[n\to\infty]{\mathbb{P}_p\ p.s} p$

$$\mathbb{E}_p((\hat{p}_n - p)^2) = \mathbb{E}_p\Big[\Big(\frac{5}{4} - \frac{3}{2}\hat{q}_n - \Big(\frac{5}{4} - \frac{3}{2}q\Big)\Big)^2\Big]$$
$$= \mathbb{E}_p\Big[\frac{9}{4}(\hat{q}_n - q)^2\Big]$$
$$= \frac{9}{4}\,\frac{q(1-q)}{n}.$$

5) Intervalle de confiance pour q à 90%, $n = 1000$

$\hat{q}_{1000}^{(\omega)} = \frac{642}{1000} = 0,642$.

$\mathbb{P}_p(|\hat{q}_n - q| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2} \leq 0,1$.

On cherche $\varepsilon > 0$ tq $\varepsilon \geq \sqrt{\frac{-1}{2n} \ln(\frac{0,1}{2})}$

$n = 1000$, $\varepsilon \geq 0,039$.

Intervalle de confiance pour q à 90%.

$\mathbb{P}_p(q \in [\hat{q}_{1000} - 0,039, \hat{q}_{1000} + 0,039]) \geq 0,9$

$\Leftrightarrow \mathbb{P}_p(|\frac{5}{6} - \frac{2}{3}\hat{p}_n - (\frac{5}{6} - \frac{2}{3}p)| > \varepsilon) \leq 0,1$.

$\Leftrightarrow \mathbb{P}_p(\frac{2}{3}|\hat{p}_n - p| > \varepsilon) \leq 0,1$

$\Leftrightarrow \mathbb{P}_p(|\hat{p}_n - p| > \frac{3}{2}\varepsilon) \leq 0,1$.

Intervalle de confiance pour p à 90%

$\mathbb{P}_p(p \in \underbrace{[\hat{p}_n - \frac{3}{2}\varepsilon, \hat{p}_n + \frac{3}{2}\varepsilon]}_{\hat{I}_n}) \geq 0,9$

Appli Numériq : $\hat{q}_{1000}(\omega) = 0,642$

$\hat{p}_{1000}(\omega) = \frac{5}{4} - \frac{3}{2}\hat{q}_{1000}(\omega) = 0,287$.

$\hat{I}_{1000}(\omega) = [0,228 ; 0,346]$.

Ex 4

$$S_{n,a} = \sum_{i=1}^{n} \mathbb{1}\{X_i = a\}$$

1) $S_{n,a} \sim \mathcal{B}in(n, \frac{1-\theta}{2})$

$S_{n,b} \sim \mathcal{B}in(n, \frac{1}{2})$, $S_{n,c} \sim \mathcal{B}in(n, \frac{\theta}{2})$

$\mathbb{P}(S_{n,a}) + \mathbb{P}(S_{n,b}) + \mathbb{P}(S_{n,c}) = 1$.

2) $\frac{2}{n} S_{n,c}$ d'espérance θ CV $\mathbb{P}_\theta$ ps vers θ.

cf TD.

3) $S_{n,a} + S_{n,b} + S_{n,c} = n$

car $\mathbb{1}\{X_i = a\} + \mathbb{1}\{X_i = b\} + \mathbb{1}\{X_i = c\} = 1$

car $X_i \in \{a, b, c\}$

4) $\sum_{n\in\mathbb{N}^*} \mathbb{P}_\theta(S_{n,b}=n) < \infty$ car

$\mathbb{P}(S_{n,b}=n) = \mathbb{P}_\theta\left(\bigcap_{i=1}^{n}\{X_i=b\}\right) = \prod \mathbb{P}_\theta(X_i=b)$
$= \frac{1}{2^n}$.

5) $\Omega' = \{\omega\in\Omega,\ \exists N(\omega)\in\mathbb{N}^*,\ \forall n\geq N(\omega)$
a) $S_{n,a}(\omega)+S_{n,c}(\omega)>0\}$.

$\Omega'^c = \{\omega\in\Omega,\ \forall N\in\mathbb{N}^*,\ \exists n\geq N$
$S_{n,a}(\omega)+S_{n,c}(\omega)=0\}$.
$= \{\omega\in\Omega,\ \forall N\in\mathbb{N}^*,\ \exists n>N$
$\omega\in\{S_{n,a}+S_{n,c}=0\}\}$.

$= \bigcap_{N\in\mathbb{N}^*}\bigcup_{n\geq N}\{S_{n,a}+S_{n,c}=0\}$ réflexe Borel-Cantelli

b) Or $\{S_{n,a}+S_{n,c}=0\} = \{S_{n,b}=n\}$.

Par (BC) $\sum \mathbb{P}_\theta(\{S_{n,a}+S_{n,c}=0\}) < \infty$
$\Rightarrow \mathbb{P}_\theta(\Omega'^c)=0 \Rightarrow \mathbb{P}_\theta(\Omega')=1$.

6) $T_n(\omega) = \begin{cases} \dfrac{S_{n,c}(\omega)}{S_{n,a}(\omega)+S_{n,c}(\omega)} & \text{si } S_{n,a}(\omega)+S_{n,c}(\omega)>0 \\ 0 & \text{sinon.} \end{cases}$

si $\omega\in\Omega'$, $\exists N(\omega)\in\mathbb{N}^*;\ \forall n\geq N(\omega)$,
$T_n(\omega) = \dfrac{S_{n,c}(\omega)}{S_{n,a}(\omega)+S_{n,c}(\omega)}$

$\dfrac{S_{n,c}}{n} \xrightarrow[n\to\infty]{\mathbb{P}_\theta\ p.s} \dfrac{\theta}{2}$ et $\dfrac{S_{n,a}+S_{n,c}}{n} \xrightarrow[n\to\infty]{\mathbb{P}_\theta\ p.s} \dfrac{1}{2}$

$\exists\ \Omega_c\in\mathcal{F}^P$ $\mathbb{P}_\theta(\Omega_c)=1$ tq $\forall\omega\in\Omega_c$,
$\dfrac{S_{n,c}(\omega)}{n} \xrightarrow[n\to\infty]{} \dfrac{\theta}{2}$.

$\exists\ \Omega_{a,c}\in\mathcal{F}^P$, $\mathbb{P}_\theta(\Omega_{a,c})=1$ tq
$\forall\omega\in\Omega_{a,c}$ $\dfrac{S_{n,a}(\omega)+S_{n,c}(\omega)}{n} \xrightarrow[n\to\infty]{} \dfrac{1}{2}$

$\omega\in\underbrace{\Omega'\cap\Omega_c\cap\Omega_{a,c}}_{\text{de proba } 1}$

$T_n(\omega) = \dfrac{S_{n,c}/n}{\dfrac{(S_{n,a}(\omega)+S_{n,c}(\omega))}{n}} \xrightarrow[n\to\infty]{} \dfrac{\theta/2}{1/2} = \theta$

7)$_1$ $V(\theta, x_1, \dots, x_n) = \prod_{i=1}^{n} \mathbb{P}_\theta (X_i = x_i)$

$= \prod_{i=1}^{n} \left[\left(\frac{1-\theta}{2}\right)^{\mathbb{1}\{X_i = a\}} \left(\frac{1}{2}\right)^{\mathbb{1}\{X_i = b\}} \left(\frac{\theta}{2}\right)^{\mathbb{1}\{X_i = c\}} \right]$

$= \left(\frac{1-\theta}{2}\right)^{\sum_{i=1}^{n} \mathbb{1}\{X_i = a\}} \left(\frac{\theta}{2}\right)^{\sum_{i=1}^{n} \mathbb{1}\{X_i = c\}} \left(\frac{1}{2}\right)^{\sum \mathbb{1}\{X_i = b\}}$

$= \left(\frac{1-\theta}{2}\right)^{n_a} \left(\frac{1}{2}\right)^{n_b} \left(\frac{\theta}{2}\right)^{n_c}$

$F'_{a,b,c} = f^*_{a,b,c} = \frac{d}{dx}\left(1-e^{-z^a}\right) = \frac{d}{dx}\left(1-e^{-\left(\frac{x-b}{c}\right)^a}\right) = +\frac{a}{c^a}(x-b)^{a-1} e^{-\left(\frac{x-b}{c}\right)^a}$

$(U^a)' = a\,U'\,U^{a-1}$

$-\left(a\left(\frac{x-b}{c}\right)'\left(\frac{x-b}{c}\right)^{a-1}\right) = -\left(\frac{a}{c}\left(\frac{x-b}{c}\right)^{a-1}\right)$

$\frac{d}{dx}\left(-\left(\frac{x-b}{c}\right)^a\right) = -\frac{1}{c^a}\frac{d}{dx}(x-b)^a = -\frac{a}{c^a}(x-b)^{a-1}$

$\cancel{\frac{1}{c^a}(x-b)^a \to \frac{a}{c^a}(x-b)^{a-1}}$

Statistiques Mathématiques M66
TP n° 2

Nous utilisons dans ce TP les bibliothèques `numpy` et `matplotlib`. On fait donc figurer en en-tête de notre code

```
import numpy as np
import matplotlib.pyplot as plt
```

Exercice 1 - Simulation par inversion de la fonction de répartition Le but de cet exercice est d'illustrer la partie du cours sur la simulation de variables aléatoires. On suppose que l'on sait simuler des variables aléatoires de loi uniforme sur l'intervalle $[0,1]$. En particulier, avec Python, le code suivant permet de simuler un vecteur de `n = 1000` variables aléatoires indépendantes et de loi uniforme sur $[0,1]$.

```
n = 1000
U = np.random.uniform(size=n)
```

Soit X une variable aléatoire réelle et $F(x) = P(X \leq x)$ sa fonction de répartition. On définit l'inverse généralisée de F par $F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}$ pour $u \in]0,1[$. Si U est une variable de loi uniforme sur $[0,1]$ alors

$$F^{-1}(U) \overset{(\text{loi})}{=} X.$$

Nous avons donc une méthode pour simuler une variable aléatoire, si l'on sait inverser sa fonction de répartition. Donnons un exemple : la simulation d'une variable de loi exponentielle de paramètre égal à $a > 0$. Dans ce cas la variable aléatoire X a une densité

$$f(x) = ae^{-ax}1_{[0,+\infty[}(x),$$

la fonction de répartition s'écrit

$$F(x) = (1 - e^{-ax})1_{[0,+\infty[}(x),$$

et pour $u \in [0,1[$

$$F^{-1}(u) = -\frac{1}{a}\log(1-u).$$

Le code suivant permet ainsi de simuler une variable aléatoire de loi exponentielle de paramètre égal à 1.

```
U = np.random.uniform()
X = - np.log(1 - U)
```

Si l'on veut simuler efficacement plusieurs variables aléatoires de loi exponentielle, on doit penser à utiliser des *vecteurs*. Par exemple le code suivant est celui d'une fonction qui prend en entrée `a` et `n`, et qui retourne un vecteur de `n` variables aléatoires indépendantes de loi exponentielle de paramètre égal à `a`.

F

Fl

```
def expo(a,n):
        U = np.random.uniform(size=n)
        X = - np.log(1-U)/a
        return(X)
```

Pour vérifier que nous simulons la loi souhaitée, on peut tracer l'histogramme d'un échantillon et lui superposer la densité de la loi exponentielle correspondante.

```
X = expo(1,1000)

plt.hist(X, bins= 20, density=True)

xx = np.linspace(0, np.max(X), 2**10)
plt.plot(xx, np.exp(-xx), 'r-')

plt.show()
```

1. Reprendre le code de simulation de variables exponentielles ci-dessus et tester différentes valeurs du paramètre a.
2. La loi de Gumbel de paramètres $b \in \mathbb{R}$ et $c > 0$ est la loi de sur $\mathbb{R}$ de fonction de répartition
$$F_{b,c}(x) = \exp\left(-\exp\left(-\frac{x-b}{c}\right)\right).$$
Cette loi apparaît lorsque l'on étudie des maxima de variables aléatoires. Écrire une fonction `gumbel` qui les deux paramètres `b` et `c`, un entier `n` et qui retourne un vecteur de `n` variables aléatoires indépendantes qui suivent la loi de Gumbel de paramètres `b` et `c`. Vérifier votre fonction à l'aide d'histogrammes.
3. La loi de Weibull de paramètres $a > 0$, $b \geq 0$ et $c > 0$ est la loi sur $\mathbb{R}$ de fonction de répartition
$$F_{a,b,c}(x) = \begin{cases} 1 - \exp\left(-\left(\frac{x-b}{c}\right)^a\right) & \text{si } x \geq b, \\ 0 & \text{sinon.} \end{cases}$$
Cette loi apparaît également lorsque l'on étudie des maximums de variables aléatoires. Écrire une fonction `weibull` qui prend en entrée les deux paramètres `a`, `b`, et `c`, un entier `n` et qui retourne un vecteur de `n` variables aléatoires indépendantes qui suivent la loi Weibull de paramètres `a`, `b`, et `c`. Vérifier votre fonction à l'aide d'histogrammes.
4. On souhaite simuler des variables aléatoires de densité
$$f(x) = \begin{cases} x & \text{si } x \in [0,1], \\ 1/2 & \text{si } x \in]1,2] \\ 0 & \text{sinon.} \end{cases}$$

$F(t) = \int_{-\infty}^{t} \frac{1}{2} x \, dx = \cancel{2[x^2]_{-\infty}^{t}} \cancel{\frac{t^2}{2}} = \frac{1}{2}\int_1^2 dx = \frac{1}{2}[x]_1^2 = \frac{1}{2}$ $]1,2]$

$F(t) = \int_{-\infty}^{t} x \, 1\!\!1_{[0,1]}(x) \, dx = \int_0^1 x \, dx = \frac{1}{2}[x^2]_0^1 = \frac{1}{2}.$

$F(x) = y$, $x = F^{-1}(y)$

$\sqrt{2y} = x \; 1\!\!1_{[0,1]}$

$y = \frac{1}{2}x^2$

$y = F(x) = \begin{cases} \frac{1}{2}x^2 & \text{si } x \in [0,1] \\ \frac{1}{2}x & \text{si } x \in]1,2] \\ 0 & \text{sinon} \end{cases}$

$1\!\!1_{]1,2]} \; x = 2y$

$f(x) = \begin{cases} x & \text{si } x \in [0,1] \\ 1/2 & \text{si } x \in]1,2]. \\ 0 & \text{sinon} \end{cases}$

Écrire une fonction `mixte` qui prend en entrée un entier `n` et qui retourne un vecteur de `n` variables aléatoires indépendantes de densité f. Vérifier votre fonction à l'aide d'histogrammes.

5. On définit la loi « double exponentielle » de paramètre $a > 0$ comme la loi sur $\mathbb{R}$ de densité

$$f_a(x) = \frac{a}{2} e^{-a|x|}. \quad F_a(t) = \int_{-\infty}^{t} f_a(x)dx = \begin{cases} \int_{-\infty}^{t} \frac{a}{2} e^{ax} dx & \text{si } t < 0 \\ \frac{1}{2} + \int_0^t \frac{a}{2} e^{-ax} dx & \text{si } t \geq 0 \end{cases}$$

$\mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = \frac{1}{2}$

Écrire une fonction `douple_exp` qui prend en entrée le paramètre `a`, un entier `n` et qui retourne un vecteur de `n` variables aléatoires indépendantes de loi double exponentielle de paramètre `a`. Vérifier votre fonction à l'aide d'histogrammes.

6. Dans un exercice de TD, on s'intéressait à une variable aléatoire X de fonction de répartition

$$F_{a,b}(x) = \begin{cases} 1 - a\exp(-bx) & \text{si } x \geq 0, \\ 0 & \text{sinon,} \end{cases}$$

$v_1 = \frac{e^{at}}{2}$

$v_2 = \frac{1}{2} + \frac{1}{2}(1 - e^{-at}) = 1 - \frac{1}{2}e^{-at}.$

avec $a \in]0,1[$ et $b \geq 0$. On avait alors vu qu'avec probabilité $1-a$, $X = 0$ et avec probabilité a, X suit une loi exponentielle de paramètre égal à b. Cela implique que si U et V sont deux variables aléatoires de loi uniforme sur $[0,1]$, alors

$$-\frac{1}{b}\log(1-U) \, 1_{[0,a]}(V) \overset{(\text{loi})}{=} X.$$

Utiliser cette remarque pour écrire une fonction `pluvio` qui prend en entrée les deux paramètres `a` et `b`, un entier `n` et qui retourne un vecteur de `n` variables aléatoires indépendantes qui suivent la même loi que X. Vérifier votre fonction à l'aide d'histogrammes.

$F(x) = y$, $y \in [0,1[$

$x = F^{-1}(y)$

Exercice 2 - Méthode de Box Muller La fonction de répartition de la loi normale (centrée, réduite) n'admet pas d'inverse généralisée explicite. Différentes méthodes permettent de palier ce problème, celle de Box-Muller s'appuie sur la propriété suivante. Si X et Y sont deux variables indépendantes de même loi $\mathcal{N}(0,1)$ et (R,θ) les coordonnées polaires de (X,Y), alors R et θ sont deux variables aléatoires indépendantes et de lois respectives $\text{Exp}(1/2)$ et $\text{Unif}([0,2\pi])$. On peut ainsi simuler R et θ pour construire X et Y.

En utilisant la méthode de Box-Muller, écrire une fonction `normal_BM` qui prend en entrée deux paramètres `mu` et `sigma`, un entier `n`, et qui retourne un vecteur de `n` variables aléatoires indépendantes de loi normale d'espérance `mu` et d'écart-type `sigma`. Vérifier votre fonction à l'aide d'histogrammes.

$Y = F^{-1}(U)$

$F(Y) = U$

$X = F(X)$

Exercice 3 - Méthode de Monte Carlo et réduction de variance La méthode de Monte Carlo permet de calculer des intégrales en utilisant l'aléatoire. Illustrons sur un exemple simple où l'on souhaite évaluer

$$I = \int_0^1 e^u du.$$

$F(t) = \int_{-\infty}^{t} \frac{1}{2} 1\!\!1_{[1,2]}(x) \, dx = \frac{1}{2}\left(\int_0^1 dx + \int_1^t dx\right) = \cancel{\frac{1}{2}(t-1)(2-t) = \frac{-t^2}{2} - 1 + t}$ $\quad \frac{1}{2}(1 + (t-1))$ $\quad \frac{1}{2} + \frac{1}{2}(t-1)$

$F(t) = \int_{-\infty}^{t} x \, 1\!\!1_{[0,1]}(x) \, dx = \int_0^t x \, dx + \cancel{\int_1^t x \, dx} = \left[\frac{1}{2}x^2\right]_0^t + \cancel{\left[\frac{1}{2}x^2\right]_t^1} = \frac{t^2}{2}$

Si U suit une loi uniforme sur $[0,1]$, alors $\mathbb{E}(e^U) = I$, en déduit que si $(U_k)_{1\leq k\leq n}$ sont des variables aléatoires indépendantes de loi uniforme sur $[0,1]$ alors par la loi forte des grands nombres

$$I_n = \frac{1}{n}\sum_{k=1}^{n} e^{U_k} \xrightarrow[n\to+\infty]{\text{p.s.}} I.$$

On peut alors donner un intervalle de confiance de niveau 95% pour I

$$\left[I_n - 1.96\frac{\sigma_n}{\sqrt{n}}\ ;\ I_n + 1.96\frac{\sigma_n}{\sqrt{n}}\right]$$

où σ_n est l'écart-type empirique de l'échantillon $(e^{U_k})_{1\leq k\leq n}$.

1. Estimer I à l'aide de l'estimateur I_n et donner l'intervalle de confiance de niveau 95% correspondant.

La précision du résultat obtenu par une méthode de Monte Carlo dépend de la largeur de l'intervalle de confiance, donc du rapport $\sigma_n/\sqrt{n}$. Pour le réduire, on peut augmenter le nombre de simulations n ou trouver une méthode pour avoir un écart-type σ_n plus petit, on parle alors de *réduction de variance*. La fin de l'exercice consiste en une présentation succincte de certaines de ces méthodes.

2 **Échantillonnage antithétique** Généralement, c'est l'opération de simulation qui est coûteuse en temps de calcul. Il peut alors parfois être intéressant de « réutiliser » la simulation. Ici remarque que $1 - U_k$ donne une autre variable aléatoire de loi uniforme sur $[0,1]$. Le nouvelle estimateur Monte Carlo est alors donné par

$$I_n^{\text{anti}} = \frac{1}{n}\sum_{i=1}^{n} \frac{e^{U_k} + e^{1-U_k}}{2}.$$

Estimer I à l'aide de l'estimateur I_n^{anti} et donner l'intervalle de confiance de niveau 95% correspondant.

3 **Variable de contrôle** Il s'agit d'une variable centrée Z_k que l'on sait facilement simuler avec U_k et telle que

$$\mathrm{Var}(e^{U_k} - Z_k) \leq \mathrm{Var}(e^{U_k}).$$

Dans notre cas, on peut choisir $Z_k = U_k - 1/2$, et le nouvel estimateur Monte Carlo est donné par

$$I_n^{contr} = \frac{1}{n}\sum_{k=1}^{n}(e^{U_k} - U_k + 1/2).$$

Estimer I à l'aide de l'estimateur I_n^{contr} et donner l'intervalle de confiance de niveau 95% correspondant.

4 Parmi les trois estimateurs précédents, lequel semble être le meilleur ?

DS M66

Ex 1 : 1) $(A_n)_{n\in\mathbb{N}^*}$ suite d'events $(\Omega,\mathcal{F},\mathbb{P})$

$$\sum_{n\in\mathbb{N}^*} \mathbb{P}(A_n) < \infty$$

⚠ $\mathbb{P}\left(\bigsqcup_{n=N}^{\infty} A_n\right) = \sum_{n=N}^{\infty} \mathbb{P}(A_n)$

$$\mathbb{P}\left(\bigcup_{n=N}^{\infty} A_n\right) \le \sum_{n=N}^{\infty} \mathbb{P}(A_n) \le \sum_{n=0}^{\infty} \mathbb{P}(A_n) = 0$$

de ~~$\lim_{N\to\infty}$~~ de limite nulle étant le reste d'une série (a).

$\rightarrow B_n = \bigcup_{k=n}^{\infty} A_k \searrow$

$\rightarrow C_n = \bigcap_{k=n}^{\infty} A_k \nearrow$.

d'où $0 \le \mathbb{P}\left(\bigcup_{n=N} A_n\right) \le \sum_{n=N} \mathbb{P}(A_n) \to 0$

Passage à la limite. $\longrightarrow 0$

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

N°

b) (d) $\mathbb{P}\left(\bigcap_{N\in\mathbb{N}^*} \bigcup_{n=N}^{\infty} A_n\right) = 0$

$B_n = \bigcup_{k=n}^{\infty} A_k \searrow$, $n=1$, $\mathbb{P}(B_1) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$

Il s'ensuit que $\mathbb{P}\left(\bigcap_{n\in\mathbb{N}^*} B_N\right) = \lim \mathbb{P}(B_N) = \lim \mathbb{P}\left(\bigcup_{n=N}^{\infty} A_n\right) = 0$

2) $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in]0,1[})$

$X_i \sim \mathrm{Bern}(\theta)$ sous $\mathbb{P}_\theta$. $\varepsilon > 0$.

$$A_n(\varepsilon) = \{(x_1, \dots, x_n) \mid |\bar{x}_n - \theta| \ge \varepsilon\}$$

a) Signification $A(\varepsilon) = \bigcap_{N\in\mathbb{N}^*} \bigcup_{n=N}^{\infty} A_n(\varepsilon)$

La moyenne empiriq ~~s'éloignera~~ de θ de ε se réalise
$\bar{X}_n$ se rapproche de θ.

b) Mq $\mathbb{P}_\theta(A(\varepsilon)) = 0$. $\forall \varepsilon > 0$.

$$\mathbb{P}_\theta\left(\bigcup_{n=N}^{\infty} A_n(\varepsilon)\right) \le \sum_{n=N}^{\infty} \mathbb{P}_\theta(A_n(\varepsilon)) = \sum_{n=N}^{\infty} \mathbb{P}_\theta(\{|\bar{X}_n - \theta| \ge \varepsilon\})$$

$$\le \sum_{n=N}^{\infty} 2e^{-2n\varepsilon^2} = 2e^{-2N^2} \frac{1}{1-e^{-2\varepsilon^2}} \xrightarrow[N\to\infty]{} 0$$

- $\overline{X}_n \xrightarrow[n\to\infty]{\text{proba}} \theta$: $\forall \varepsilon>0$, $\lim_{n\to\infty} \mathbb{P}_\theta(|\overline{X}_n-\theta|\geq\varepsilon)=0$.

- $\begin{cases} \overline{X}_n \xrightarrow[n\to\infty]{\text{proba}} \theta \\ \sum_{n\in\mathbb{N}} \mathbb{P}_\theta(|\overline{X}_n-\theta|\geq\varepsilon)<\infty \end{cases} \Rightarrow \overline{X}_n \xrightarrow[n\to\infty]{p.s} \theta$.

- $\overline{X}_n \xrightarrow[n\to\infty]{p.s} \theta \Rightarrow \overline{X}_n \xrightarrow[n\to\infty]{\text{proba}} \theta$

---

c) $\underline{\text{cd}}$ $\overline{X}_n \xrightarrow[n\to\infty]{p.s} \theta$

$\mathbb{P}_\theta(A(\varepsilon))=0 \Rightarrow \mathbb{P}_\theta\left(\bigcap_{N\in\mathbb{N}^*}\bigcup_{n=N}^{\infty} A_n(\varepsilon)\right)=0$.

$\Rightarrow \lim_{N\to\infty} \mathbb{P}_\theta\left(\bigcup_{n=N}^{\infty} A_n(\varepsilon)\right)=0$.

Or $A_N(\varepsilon) \subseteq \bigcup_{n=N}^{\infty} A_n(\varepsilon) \Rightarrow \mathbb{P}_\theta(A_N(\varepsilon)) \leq \mathbb{P}_\theta\left(\bigcup_{n=N}^{\infty} A_n(\varepsilon)\right)$

$\to 0$ $\to 0$

$\Rightarrow$ cvce en proba car $\lim_{N\to\infty} \mathbb{P}_\theta(A_N(\varepsilon))=0$.

$\sum \mathbb{P}(A_n(\varepsilon)) <\infty$

iid : independant identiqt distribué

Puis on doit prouver $\sum_{n\in\mathbb{N}} \mathbb{P}_\theta(|\overline{X}_n-\theta|\geq\varepsilon)<\infty$.

Or on a vu à peu près ds b), cette qtté était majorée par $\frac{2}{1-e^{-2\varepsilon^2}} \cdot e^{-2\varepsilon^2}$ dc est finie.

Dc on a cv p.s.

$S_n \sim \text{Bin}\left(3600, \frac{1}{6}\right)$

$S_n=\sum_{n=1}^{3600} X_i$

Ex2 : 3600. $\theta \leftarrow \frac{1}{6}$. $X_i \sim \text{Bern}(\theta)$ indep.

1) Si dé bien équilibré :

$\mathbb{P}_\theta(500 \leq S \leq 700) = \cancel{\mathbb{P}_\theta(|\overline{X}_n-600|\geq 500)} = \mathbb{P}_\theta(|$

$\mathbb{P}_\theta(|S-600| \leq 100) = \mathbb{P}_\theta(|S-600|<101) = \mathbb{P}_\theta\left(\left|\frac{S}{3600}-\frac{1}{6}\right| < \frac{101}{3600}\right)$

600

500 x 700

$= 1-\mathbb{P}_\theta\left(\left|\frac{S}{3600}-\frac{1}{6}\right| \geq \frac{101}{3600}\right) \geq 1-2e^{-2\times3600\left(\frac{101}{3600}\right)^2}$

$\approx 0,993$.

$|X-200| \geq 600$

$-a \leq X \leq a$ $X-200$

$|X| \geq a$ $-X+200 \geq 600$

$X-200 \leq 500$

$|X-200$ $X \leq 700$

2 $|X| \geq$

$\overline{X}_n = \frac{S_n}{n}$, $S_n=\sum X_i$

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

$\mathbb{P}(|X_n-n\theta| \geq n\varepsilon) \leq 2e^{-2n\varepsilon^2}$

2) IdC $_{95\%}$? Hoeffding

$\overline{X}_n = \frac{1}{n}\sum X_i$

$1-\alpha = 0{,}05$.

$\mathbb{P}_\theta(\hat{I}_n \ni \theta) \leq 1-\alpha$?

$\mathbb{P}_\theta(|\overline{X}_n - \theta| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2} \leq 0{,}05$ $\quad \varepsilon > 0$

$e^{-2n\varepsilon^2} \leq \frac{0{,}05}{2}$: $-2n\varepsilon^2 \leq \ln\left(\frac{0{,}05}{2}\right)$

$2n\varepsilon^2 \geq \ln\left(\frac{2}{0{,}05}\right)$

$\varepsilon \geq \sqrt{\frac{1}{2n}\ln\left(\frac{2}{0{,}05}\right)}$.

$\hat{I}_n = \left]\overline{X}_n - \sqrt{\frac{1}{2n}\ln\left(\frac{2}{0{,}05}\right)}, \overline{X}_n + \sqrt{\frac{1}{2n}\ln\left(\frac{2}{0{,}05}\right)}\right[\cap \left]0,1\right[$

$\hat{I}_n = \left]\overline{X}_n - 0{,}0226, \overline{X}_n + 0{,}0226\right[$

$\overline{X}_n = \frac{x}{3600}$

Soit $\theta \in]0,1[$, on cherche ε tq

~~$\mathbb{P}_\theta(|S - 3600\theta| \leq k) \leq 0{,}05$~~

$\Leftrightarrow \mathbb{P}_\theta(|\overline{X}_n - \theta| \geq \varepsilon) \leq 0{,}05$

$\Leftrightarrow 2e^{-7200\varepsilon^2} \leq 0{,}05$.

3) Si 3600 lancers, observe 640 → 6.

H_0 : "dé équil" $\frac{1}{6}$ $\qquad$ H_1 " dé mal équi. $\theta \neq \frac{1}{6}$.

$\mathbb{P}_{H_0}(\text{conclure } H_1) \leq 0{,}05$.

~~$R_{5\%} = \{S_n \geq k\}$ tq $\mathbb{P}_{\frac{1}{6}}(S_n \geq k_\alpha) \leq 0{,}05$~~

~~k_α vérifie $\mathbb{P}_{\theta_0}(S_n \geq k_\alpha) \leq \alpha$.~~

~~$k_\alpha = \underset{0 \leq k \leq n}{\operatorname{argmin}} \{\mathbb{P}_{\theta_0}(S_n \geq k) \leq \alpha\}$~~

$\left|\frac{S_n}{n} - \theta\right| \geq \varepsilon$

$R_{5\%} = \{$

$\frac{640}{3600}$

3) $\mathcal{H}_0 : \theta = \frac{1}{6}$ contre $\mathcal{H}_1 : \theta \neq \frac{1}{6}$.

• Pour k_2^{α} :

$$R_{3600}^{0,05} = \{ S_{3600} \leq k_1^{0,05} \} \cup \{ S_{3600} \leq k_2^{0,05} \}.$$

où $k_1^{0,05}, k_2^{0,05}$ vérifiant $\mathbb{P}_{\frac{1}{6}}(R_{3600}^{0,05}) \leq \alpha$

$$\Rightarrow k_1^{0,05} = \max(k \mid \mathbb{P}_{1/6}(S_{3600} \leq k) \leq 0,025)$$

$$k_2^{0,05} = \min(k \mid \mathbb{P}_{1/6}(S_{3600} \geq k) \leq 0,025)$$

Pour le premier : $\mathbb{P}_{\frac{1}{6}}(S_{3600} \leq k) \leq 0,025$

multiplie par -1

$\mathbb{P}_{1/6}(-S_{3600} \geq -k) \geq 0,025$

$$\Leftrightarrow \cancel{\mathbb{P}_{\frac{1}{6}}\left(\frac{S_{3600}}{3600} - \frac{1}{6} \leq \frac{k+1}{3600} - \frac{1}{6}\right) \leq 0,025}$$

$$\Leftrightarrow \mathbb{P}_{\frac{1}{6}}\left(\frac{S_{3600}}{3600} - \frac{1}{6} \geq \frac{-k}{3600} + \frac{1}{6}\right) \geq 0,975.$$

$$\Leftrightarrow e^{-7200\left(\frac{1}{6} - \frac{k}{3600}\right)^2} \leq 0,975$$

$$\Leftrightarrow 600 - 360\sqrt{\frac{\ln(200)}{7200}} \geq k$$

le max qui vérifie est $k = 302$.

$$k \geq 599 + 3600\sqrt{\ln(0,975)(7200)}$$

4

3 $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in]0,1[})$

$V(\theta, x_1, \dots, x_n) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i)$

2) $\prod_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$

$= \theta^{\sum x_i}(1-\theta)^{\sum(1-x_i)}$

$= \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$.

3) poser $\mu = \theta^2$.

$\tilde{V}(\mu, x_1, \dots, x_n) = \sqrt{\mu}^{\sum x_i}(1-\sqrt{\mu})^{n-\sum x_i} = V(\theta, x_1, \dots, x_n)$

4) L ? Mq estimat^R du maximum de vraisemblance de μ est $\hat{\mu}_n = \text{argmax}_{\mu \in]0,1[} \tilde{V}(\mu, X_1, \dots, X_n)$

$= \text{argmax}_{\mu \in]0,1[} \tilde{L}(\mu, X_1, \dots, X_n)$

$= \overline{X}_n^2$ où $\overline{X}_n = \frac{1}{n}\sum X_i$.

$\mathbb{P}(X = x) = p^x(1-p)^{1-x}$, $x \in \{0,1\}$

, Estimat^R du max de vraisemble de θ est $\hat{\theta}_n =$

$= \text{argmax}_\theta V(\theta, x_1, \dots, x_n)$.

On a $\text{argmax}_{x \in D}(f(x)) = \text{argmax}_{x \in D}(\ln(f(x)))$

$\tilde{L}(\theta, x_1, \dots, x_n)\ \ln\left(\sqrt{\mu}^{\sum x_i}(1-\sqrt{\mu})^{n-\sum x_i}\right)$

$= \frac{1}{2}\left(\sum x_i\right)\ln(\mu) + \left(n - \sum x_i\right)\ln(1-\sqrt{\mu})$. $\quad \frac{-1}{2\sqrt{\mu}}$

On dérive en μ : $\frac{\partial}{\partial \mu}\tilde{L}(\mu, x_1, \dots, x_n) = \frac{\sum x_i}{2\mu} + \frac{(n - \sum x_i)}{1-\sqrt{\mu}}$

Cette dérivée est nulle en on cherche qd la dérivée s'annule μ trouver le maximum.

$\frac{1}{2\mu}\left(\sum x_i\right) + \left(n - \sum x_i\right)\frac{-\frac{1}{2\sqrt{\mu}}}{1-\sqrt{\mu}} = 0$

$\Leftrightarrow \sum x_i - \left(n - \sum x_i\right)\frac{\mu}{\sqrt{\mu} - \mu} = 0$

$\Leftrightarrow \sum x_i - \left(n - \sum x_i\right)\frac{1}{\frac{1}{\sqrt{\mu}} - 1} = 0$

$\Leftrightarrow \frac{\sum x_i}{n - \sum x_i} = \frac{1}{\frac{1}{\sqrt{\mu}} - 1} \Leftrightarrow \frac{1}{\frac{n - \sum x_i}{\sum x_i} + 1} = \sqrt{\mu}$.

$$\Leftrightarrow \frac{1}{\frac{n-\sum x_i}{\sum x_i}+1} = \sqrt{\mu} \Leftrightarrow \frac{\sum x_i}{n-\sum x_i+\sum x_i} = \sqrt{\mu}$$

$$\Leftrightarrow \frac{\sum x_i}{n} = \sqrt{\mu} \Leftrightarrow \mu = (\bar{X}_n)^2$$

6) Calculer $\mathbb{E}_\theta(\bar{X}_n^2) = \mathbb{E}_\theta\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right)$

$$= \frac{1}{n^2}\mathbb{E}_\theta\left(\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} X_i\right)\right) = \frac{1}{n^2}\mathbb{E}_\theta\left(\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n}\left(\prod_{\substack{k=1\\k\neq i}}^{n} X_i\right)\right)$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbb{E}_\theta(X_i)^2 + \sum_{i=1}^{n}\mathbb{E}_\theta\left(\prod_{\substack{k=1\\k\neq i}}^{n} X_i\right)\right)$$

$$= \frac{1}{n^2}\left[\sum\theta^2 + \sum\prod\theta\right] = \frac{1}{n^2}\left[n\theta^2 + n\theta^{n-1}\right]$$

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

L'estimateur est sans biais si

$$\mathbb{E}(\text{estimat}^R) = \theta.$$

ici n'est pas sans biais.

convergent

$$\lim_{n\to\infty} \longrightarrow \theta \qquad \mu = \theta^2$$

7) Proposer un autre estimat^R trivial de μ q est sans biais.

$$\hat{\mu}_n = \frac{1}{n}\sum X_i$$

$$\Rightarrow \mathbb{E}_\theta(\hat{\mu}_n) = \sqrt{\mu}$$

$$\mathrm{Var}_\theta(\bar{X}_n) = \mathbb{E}_\theta(\hat{\mu}_n) - \mathbb{E}_\theta(\bar{X}_n)^2$$

$$\mathbb{E}_\theta(\hat{\mu}_n) = \mathrm{Var}_\theta(\bar{X}_n) + \mathbb{E}_\theta(\bar{X}_n)^2$$

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

$$\left(\sum x_i\right)^2 = \left(\sum x_i\right)\left(\sum x_j\right) = \sum\sum x_i x_j$$

$$= \sum x_i^2 + 2\sum_{1\leq i\leq j\leq n} x_i x_j$$

Sans biais si

$$\mathbb{E}(\text{estimat}^R) = \theta.$$

1 Faire l'amour vs Ménage

1) Modéliser pb. Ds $(\Omega, \mathcal{F}, \mathbb{P}_\theta)_{\theta \in]0,1[}$

on def une suite la $(X_i)_{1\le i \le n}$ de Bernoulli inconnue indiquant (la proba) θ par celle qu'un h marié effectuant au moins 30% des TM d'avoir au moins 4 RS/mois. ($X_i = 1$ si i° h interrogé

la loi $S_n = \cancel{\frac{1}{n}} \sum_{i=1}^{n} X_i$ binomiale $\sim \mathcal{B}in(n, \theta)$ sous $\mathbb{P}_\theta$.

La Q° est de savoir si $\theta \le 0,6$.

2) Déterminer si $\theta \le 0,6$. Faire test stat.

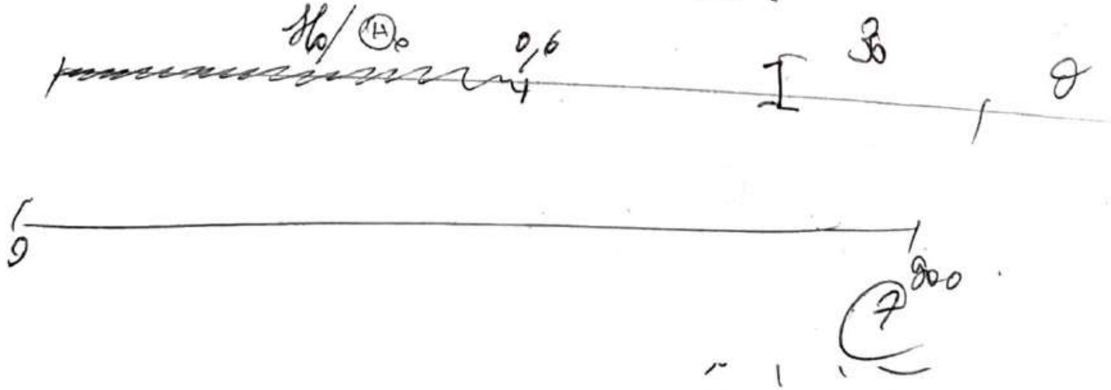
Il faut sélectionner les hypos H_0 & H_1 tq

$H_0 : \theta \le 0,6$

$H_1 : \theta > 0,6$. } à justifier.

minimiser $\mathbb{P}_{H_0}(\cancel{\text{conclure}}\ H_1)$ ← conclure à tort H_1 alors que H_0 vraie.

3) Construire test de niveau 5%.



$1-\alpha$: fiabilité (E_1).

α : incertitude 0,5%.

Zone de Rejet $R_n^\alpha = \{S_n \ge k\} \le \alpha$.

de t^a sorte que $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(R_n) \le \alpha$,

$k = \min\{k \mid \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(R_n) \le \alpha\}$

$= \min\{k \mid \mathbb{P}_{0,6}(\{S_n \ge k\}) \le \alpha\}$

$\mathbb{P}_{0,6}(\{S_n \ge k\}) \le 0,05 \iff \mathbb{P}_{0,6}(\{S_n > k+1\}) \le 0,05$

$\iff 1 - \mathbb{P}_{0,6}(\{S_n < k+1\}) \le 0,05$.

$k = np - n\varepsilon$. $\iff 0,95 \le \mathbb{P}_{0,6}(\{S_n < k+1\})$

→ D'après la table, le + petit entier $k+1$ vérifiant cette inégalité est 564, soit $k_\alpha = 563$

On ne rejette pas H_0 car $513 \le 563$.

Autre rech : $\mathcal{H}_0 : \theta > 0{,}6$

$\mathcal{H}_1 : \theta \leq 0{,}6$.

On aurait déterminé k_α a étant

$k_\alpha = \max\{ k \mid \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(R_n) \leq \alpha \}$

$= \max\{ k \mid \sup \mathbb{P}_\theta(\{S_n \leq k_\alpha\}) \leq 0{,}05 \}$

$= 513$

$513 \leq 515$ et on rejette $\mathcal{H}_0$.

atteindre le sup

$\Rightarrow$ Sans table : user Hoeffding.

Astuce de Lamsi

$\mathbb{P}_\theta(S_n \leq k_\alpha) \leq 0{,}05$

$\mathbb{P}_\theta(-S_n \geq -k_\alpha) \leq 0{,}05$

$\mathbb{P}_\theta\left(0{,}6 - \frac{S_n}{900} \geq 0{,}6 - \frac{k_\alpha}{900}\right) \leq 0{,}05$

$\Leftrightarrow e^{-1800\underbrace{\left(0{,}6 - \frac{k_\alpha}{900}\right)^2}_{\varepsilon}} \leq 0{,}05$.

$k \leq 503$

$e^{-1800\varepsilon^2} \leq 0{,}05$

$\mu\ \varepsilon > 0$ on cherche

~~$0{,}6 - \frac{k}{}$~~ le max

$k = 503$.

$S_n \leq k_\alpha \qquad \frac{S_n}{n} - \theta \geq \varepsilon$

$\boxed{12}$ Consommation carburant. $\mathbb{P}(X \geq 7) = p$

1) Valeur numérique pour estimer p (en justifiant).

Moyenne empirique $\hat{p} = \frac{|\{i \in \{1,\dots,100\} \mid X_i(\omega) > 7\}|}{100} = 4{,}7$.

En effet $\hat{p} = \frac{1}{100} \operatorname{card}\{i \in \{1,\dots,100\} \mid X_i \geq 7\}$.

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

$1 \leq i \leq 100,\ Y_i = \mathbb{1}_{\{X_i \geq 7\}}$

$(Y_i)_{1 \leq i \leq 100}$ sont indep, $Y_i \sim \mathrm{Bern}(p)$

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

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

Biais $\theta - \mathbb{E}(\cdot) = 0$.

$\mathbb{E}(\text{estimateur}) = \theta \Rightarrow$ estimateur est sans biais

(a)

2) On cherche un IdC au niveau 95 %

μ p. On cherche ε tq $\mathbb{P}_p(p \in \hat{I}) \geq 0{,}95$

$p \in \hat{I} \iff |p - \hat{p}| \leq \varepsilon$. $\leftarrow \mathbb{P}_p\left(\left|\frac{Y_n}{n} - p\right| \geq \varepsilon\right) \leq 0{,}05$

$\mathbb{P}_p(p \in \hat{I}) \geq 0{,}95 \iff \mathbb{P}_p(|p - \hat{p}| \geq \varepsilon) \leq 0{,}05$.

$\iff \mathbb{P}_p(|\hat{p} - p| \geq \varepsilon) \leq 2 e^{-2n\varepsilon^2}$

Il suffit $\varepsilon > 0$ tq $2 e^{-2n\varepsilon^2} \leq 0{,}05$.

$e^{-2n\varepsilon^2} \leq \frac{1}{2} 0{,}05$

$-2n\varepsilon^2 \leq \ln\left(\frac{1}{2} 0{,}05\right)$

$2n\varepsilon^2 \geq \ln\left(\frac{2}{0{,}05}\right)$

$\varepsilon \geq \sqrt{\frac{1}{n 2} \ln\left(\frac{2}{0{,}05}\right)}$

$\varepsilon \geq \sqrt{\frac{1}{200} \ln(40)}$

cet intervalle de confiance à 95 % est

$\Rightarrow \left]\frac{Y_n}{n} \pm \sqrt{\frac{\ln 40}{200}}\right[\cap \left]0, 1\right[$

$S_n = \sum Y_i \sim \mathcal{B}in(n, p)$

$\overline{X}_n = \frac{1}{n} \sum Y_i$: estimateur moy empirique

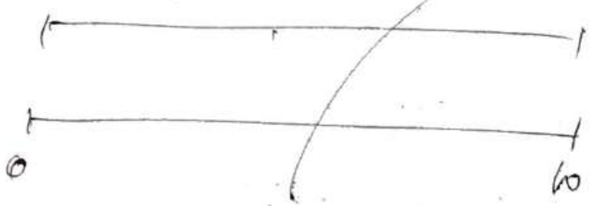
(d) X

$X \sim \mathcal{B}in(10, p)$

$H_0 : p = 0{,}5$ $H_1 : p = 0{,}8$

Unilatéral

$\{7, 8, 9, 10\}$



① Zone de Rejet : H_0 rejetée / H_1 acceptée.

2) Taille du test : $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(R_0)$ ici $\Theta = \{0{,}5\}$

$= \mathbb{P}_{0,5}(\{R_0\}) = \mathbb{P}_{0,5}(\{X > 6\}) = 1 - \mathbb{P}(\{X \leq 6\})$

$= 1 - 0{,}828 = 0{,}172$

Puissance du test : conclure H_1 à raison.

$\mathbb{P}_{0,8}(R)$

9