# M-66 Pr: Castellan Gwanaëlle

## STATISTIQUE MATHÉMATIQUE

### Objectifs

1. avoir un premier aperçu de deux problématiques fondamentales en statistique : l'estimation et la prise de décision
2. pour l'estimation, apprendre à construire rigoureusement un intervalle de confiance
3. pour la prise de décision, découvrir le vocabulaire des tests d'hypothèses, savoir construire et interpréter quelques tests simples.

### Statistiques

1. Simulation de variables aléatoires : simulation de variables aléatoires discrètes, inverse de lafonction de répartition, méthode du rejet
2. Estimation ; introduction du vocabulaire de l'estimation à partir du modèle de Bernoulli ; intervalles de confiance non asymptotiques.
3. Introduction à la notion de tests statistiques : vocabulaire des tests d'hypothèses (hypothèses,risque, région de rejet, puissance) ; test sur une probabilité inconnue dans le cadre du modèle de Bernoulli (test binomial) ; exemples de tests de comparaison d'échantillons appariés (test du signe) et non appariés (test des longueurs, test de la somme des rangs).
4. Application de la loi des grands nombres et du théorème de la limite centrale (TCL) aux intervalles de confiance et aux tests : rappel de la loi des grands nombres ; rappel du théorème de Moivre et énoncé du TCL ; intervalles de confiance asymptotiques ; propriétés des estimateurs, consistance et normalité asymptotique.Les 36h de TD comportent une dizaine d'heures de TD avec Python

## (C1) Démarche statistiq à t^{ts} [M] Bernoulli | II/ Modélisat°

∘ Observat° : n obser. $x_1, \dots, x_n$. On les modélise ĉ réalisat° de (va) : $X_1(w) = x_1, \dots, X_n(w) = x_n$.

∘ Famille de probas : les $(X_i)_{1 \le i \le n}$ st déf s/ espace probabilisable $(\Omega, \mathcal{F})$ (& mesurables). On munit $(\Omega, \mathcal{F})$ d'une famille de probabilités $(\mathbb{P}_\theta)_{\theta \in \Theta}$

∘ Modèle statistiq : $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$

la loi seule des $X_i$ ne suffit pas.

∘ Loi des observat° : pr avoir loi des observat° $(X_1, \dots, X_n)$

ⓔ sondage (N électeurs, n interrogés) pas indep cᵒ rep. tsr. $\mathbb{P}_\theta(S_n = k) = \dfrac{C^k_{\theta N} \, C^{n-k}_{(1-\theta)N}}{C^n_N}$ $\quad S_n = \sum_{i=1}^n X_i$

$S_n = \mathrm{card}\{1 \le i \le n, X_i = 1\}$

$S_n \sim$ Hypergéom$(n, N, \theta)$

⇒ permet d'avoir la loi de $(X_1, \dots, X_n)$.

∘ (CV) loi hypergéom ⟶ loi binom : $\lim_{N \to \infty} \mathbb{P}_\theta(S_n = k) = C^k_n \theta^k (1-\theta)^{n-k}$

2° modél° : "TAR" (indep, remise) $(X_i)_i$ indep, $X_i \sim \mathcal{B}ern(\theta)$, $S_n \sim \mathcal{B}in(n, \theta)$ sous $\mathbb{P}_\theta$. $\left(\mathbb{P}_\theta(S_n = k) = C^k_n \theta^k (1-\theta)^{n-k}\right)$

(Rq) Vers l'∞ (non (d)) qd $n \to \infty$, $\Omega = \{0,1\}^{\mathbb{N}^*}$ ← ens de ttes suites à val^{rs} ds $\{0,1\}$, $\mathcal{F}$ ! $\mathbb{P}_\theta$ sur $(\Omega, \mathcal{F})$ pt ê construit pr coïncider ∀ n. ⇒ $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in [0,1]})$ & un n-échantillon $(X_1, \dots, X_n)$ tq $X_i \sim \mathcal{B}ern(\theta)$ est appelé modèle de Bernoulli

## III/ Estimat° & Interv de confiance

$(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$ & $(X_1, \dots, X_n)$ le n échant. Obtenir $\theta$ ĉ observat° de $(X_1, \dots, X_n)$ ?

### (1) Estimat°

(D) Un estimat^{r} de $\theta$ est f mesurable de $(X_1, \dots, X_n)$ q ne dpd pas de $\theta$.

→ Un estimat^{r} $\hat{\theta}_n$ de $\theta$ est (va) : $\hat{\theta}_n = \varphi(X_1, \dots, X_n)$ où $\varphi$ mesurable.

(D) f mesurable si $f : (E_1, \mathcal{F}_1) \to (E_2, \mathcal{F}_2)$, $\forall B \in \mathcal{F}_2$, $f^{-1}(B) \in \mathcal{F}_1$.

∘ $\hat{\theta}_n(w)$ est le param $\theta$ q maximise $\theta \mapsto \mathbb{P}_\theta((X_1, \dots, X_n) = (x_1, \dots, x_n))$

∘ $\hat{\theta}_n$ est le param le plus vraisemblable.

Généralisat° : $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in [0,1]})$, $X_1, \dots, X_n$ iid $\mathcal{B}ern(\theta)$ sous $\mathbb{P}_\theta$. $\forall \theta \in [0,1]$

$\forall (x_1, \dots, x_n) \in \{0,1\}^n$, $\forall (\theta, x_1, \dots, x_n) = \mathbb{P}_\theta((X_1, \dots, X_n) = (x_1, \dots, x_n)) = \theta^{\sum_n x_i} (1-\theta)^{n - \sum_n x_i}$

$V(\theta, x_1, \dots, x_n)$ f de vraisemblance. $\ln(V(\theta, x_1, \dots, x_n)) = \sum x_i \ln \theta + (n - \sum x_i) \ln(1-\theta)$

$\varphi'(\theta) = \frac{1}{\theta} \sum x_i - \frac{1}{1-\theta} n - \sum x_i$ ; $\varphi'(\theta) \geq 0 \iff \theta \leq \frac{\sum x_i}{n}$

• Estimateur du max de vraisemblance : $\hat{\theta}_n = \underset{\theta \in [0,1]}{\mathrm{argmax}} \; V(\theta, X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$

$\iff V(\theta, X_1, \dots, X_n) \leq V(\hat{\theta}_n, X_1, \dots, X_n)$.

∘ Moyenne Empirique des $(X_i)_i$ : $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

(Pk) estimat^{r} ? → est maximise vraisblce : le param le + vraisblble à la vue des observat° $x_1, \dots, x_n$

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

$\mathbb{E}_\theta(\overline{X}_n) = \theta$. $\overline{X}_n$ est un estimat^{r} sans biais de $\theta$. (+ n grd, + $\overline{X}_n$ se rapproche $\theta$)

Comm^t évaluer (CV)ce de $\overline{X}_n$ vers $\theta$ : évaluer $\mathbb{P}_\theta(|\overline{X}_n - \theta| \geq \varepsilon)$.

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

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

3) Hoeffding (meilleur) $\quad \mathbb{P}_\theta(|\overline{X}_n - \theta| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}$

(Rq) En généralisant :

→ (BT) valable pr (va) de carré intégrable. → (H) valable pr (va) bornée.

pr loi (va) non-bornée

$\forall \varepsilon > 0$, $\lim_{n \to \infty} \mathbb{P}_\theta(|\overline{X}_n - \theta| \geq \varepsilon) = 0$, on dit $\overline{X}_n$ (cv) vers $\theta$ sous $\mathbb{P}_\theta$, $\theta \in [0,1]$, $\quad \overline{X}_n \xrightarrow[n \to \infty]{\text{proba}} \theta$.

• CVce simple ? $\overline{X}_n(w) \xrightarrow[n \to \infty]{} \theta \quad \forall w \in \Omega$ ? $\quad \Omega = \{0,1\}^{\mathbb{N}^*}$ (∞ non (d))

$\mathcal{F}$, $\mathbb{P}$ tq $E_i$ = "pile au i° lancer", $\mathbb{P}_\theta(E_i) = \theta$ et les $(E_i)_{i \in \mathbb{N}^*}$ st indépdts.

$X_i(w) = w_i$ si $w = (w_n)_{n \in \mathbb{N}^*} \in \Omega$ (résultat du i° lancer)

$\mathbb{P}_\theta(X_i = 1) = \mathbb{P}_\theta(E_i) = \theta = 1 - \mathbb{P}_\theta(X_i = 0)$, $X_i \sim \mathcal{B}ern(\theta)$ et $(X_i)_i$ indep.

(1) MEG stats $\quad \overline{X}_n(w) \xrightarrow[n \to \infty]{} \theta$ n'est pas vraie $\forall w \in \Omega$

@ $\omega=(0,\dots,0,\dots)$ $\overline{X}_n(\omega)=0$ et $\omega=(1,\dots,1,\dots)$ $\overline{X}_n(\omega)=1$

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

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

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

$|\overline{X}_p(\omega)-\theta|\leq\varepsilon$, se traduit en termes d'évents :

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

$\rightarrow \omega\in\Omega'$ 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 $\mathbb{P}_\theta(\Omega'^c)=0$.

$$\Omega'^c=\bigcup_{\varepsilon>0}\bigcap_{n\in\mathbb{N}^*}\bigcup_{p\geq n}A_p^c(\varepsilon)=\bigcup_{m\in\mathbb{N}^*}\bigcap_{n\in\mathbb{N}^*}\bigcup_{p\geq n}A_p^c\left(\frac{1}{m}\right)$$

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

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

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

$$B_{n+1}(\varepsilon)\subset B_n(\varepsilon)=A_n^c(\varepsilon)\cup B_{n+1}(\varepsilon)$$

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

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

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

Reste d'une série CV, $\lim_{n\to\infty}\mathbb{P}_\theta(B_n(\varepsilon))=0$. $\Rightarrow$ on a $(*)$

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

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

Cl $\exists\,\Omega'\in\mathcal{F},\ \mathbb{P}_\theta(\Omega')=1,\ \forall\omega\in\Omega',\ \lim_{n\to\infty}\overline{X}_n(\omega)=\theta$

On dit que $\overline{X}_n$ CV presq sûrement vers $\theta$.

Rq
- $\overline{X}_n$ CV p.s. vers $\theta$ $\Rightarrow$ $\overline{X}_n$ CV en proba vers $\theta$.
- $\overline{X}_n$ CV en proba vers $\theta$ + $\sum_{n\in\mathbb{N}}\mathbb{P}_\theta(|\overline{X}_n-\theta|\geq\varepsilon)<\infty$ $\Rightarrow$ $\overline{X}_n$ CV ps vers $\theta$.

• Risque quadratiq : $\mathbb{E}_\theta\left((\overline{X}_n-\theta)^2\right) = Var(\overline{X}_n)$

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

par indep des $(X_i)$ $= \frac{1}{n^2} n\theta(1-\theta) = \frac{\theta(1-\theta)}{n}$

## 2) Intervalles de confiance

Un IdC est interv aléat. où se trve vrai param inconnu ac grde proba.

(D) $\alpha \in ]0,1[$, un IdC de niv° de confiance $1-\alpha$ est un interv aléa $\hat{I}_n^\alpha$ en f de $X_1, \dots, X_n$ (indépdt de $\theta$) tq $\mathbb{P}_\theta(\hat{I}_n^\alpha \ni \theta) \geq 1-\alpha$. (Outils : $\boxed{BT}$ & $\boxed{H}$)

@ $\boxed{M}$ de Bernoulli $\theta$ inconnu

0 — $\overline{X}_n(\omega)$ — $\hat{I}_n^\alpha$ — 1

@ on 7, $\overline{X}_{10}(\omega) = \frac{3}{10}$,

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

Il suffit de prdre $\varepsilon = \varepsilon(\alpha,n)$ tq $\frac{1}{4n\varepsilon^2} \leq \alpha$, $\varepsilon \leq \sqrt{\frac{1}{4n\alpha}}$

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

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

$\boxed{H}$ $\mathbb{P}_\theta(|\overline{X}_n-\theta| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}$, il suffit de prendre $\varepsilon = \varepsilon(\alpha,n)$ tq

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

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

@ on 7 $n=10$, IdC de niveau 90 %, $\alpha = 10\%$

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

$n=1000$, $\alpha=10\%$ (BT) $\frac{1}{2\sqrt{n\alpha}} = \frac{1}{20} = 0,05$ (H) $\sqrt{\frac{1}{2n}\ln\frac{2}{\alpha}} = 0,0387$.

## IV / Tests Stats

@ médicament : on teste nv° tt$^t$ sr pet$^t$ nbr cobayes ; nv° tt$^t$ est-il meill$^r$ ?

Erreur$_1$ : ccl$^r$ tt$^t$ est mll$^r$ alors q ce n'est pas le cas. $E_2$ : ccl tt$^t$ n'est ps mll$^r$ ⇒ en réalité il l'est.

On ne pourra pas minimiser les 2 err$^{rs}$ : on privilégie $E_1$.

Oriente le choix (dissymétriq) des hypo :

$H_0$ : hypo nulle (hypo p défaut) | $H_1$ : hypo alternative.

→ le test $H_0$ contre $H_1$ est constr$^t$ dsq la proba de se tromper qd on conclut $H_1$ est faible & contrôlée.

@ $H_0$ : "tt$^t$ n'est ps mll$^r$" | $H_1$ : "tt$^t$ est meill$^r$" $\theta > \theta_0$

$\theta \leq \theta_0$

↳ chx dissymétriq, dépd pt de vue : @ confinement.

$\theta_0$ : proba guérison de ancien tt$^t$ (connu) / $\theta$ : — " — (inconnu)

### Construct° du test :

@ médoc, on vt tester $H_0 : \theta \leq \theta_0$ contre $H_1$ $\theta > \theta_0$, à la vue de l'observat° $X_1, \dots, X_n$ (réact°s de n pat. au nv° tt$^t$)

$X_1, \dots, X_n$ iidd $Bern(\theta)$ : $X_i = \begin{cases} 1 & \text{si } i^\circ \text{ pat guérit} \\ 0 & \text{sinon.} \end{cases}$

$\theta$ inconnu, on a estimat$^r$ $\overline{X}_n$

$\theta_0$ connu ⇒ Comment conclure $H_1$ ?

"Réponse naïve" si $\overline{X}_n > \theta_0$.

(Pb) On veut contrôler proba de se tromper $\mathbb{P}_{H_0}$ ("conclure $H_1$")

Erreur : $\sup_{\theta \leq \theta_0} \mathbb{P}_\theta(\overline{X}_n \geq \theta_0) = \mathbb{P}_{\theta_0}(\overline{X}_n \geq \theta_0)$

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

car la médiane d'une $Bin(n, \theta_0)$ 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. On cherche un seuil $t_\alpha$ tq on conclut $H_1$ si $\overline{X}_n \geq t_\alpha$. ($\overline{X}_n$ suffisamment + grd $\theta_0$) dsq $\sup_{\theta \leq \theta_0} \mathbb{P}_\theta(\overline{X}_n \geq t_\alpha) \leq \alpha$.

→ Zone de rejet : $\mathcal{R}_n^\alpha = \{\overline{X}_n \geq t_\alpha\}$ (event sur lql on rejette $H_0$, on conclut $H_1$)

Règle de décision : Si on observe $\overline{X}_n(\omega) \geq t_\alpha$ alors on rejette $H_0$ (& err$^r$ contrôlée $\leq \alpha$) sinon on ne rejette pas $H_0$. (mais l'err$^r$ n'est pas contrôlée)

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

$X_1, \dots, X_n$ iid $Bern(\theta)$.

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

(D) Construire un test de $H_0$ contre $H_1$ au niv° $\alpha$, c'est construire une zone de rejet ($f$ de $X_1, \dots, X_n$) $\mathcal{R}_n^\alpha$ tq $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\mathcal{R}_n^\alpha) \leq \alpha$.

Règle de décision :

- si $\omega \in \mathcal{R}_n^\alpha \Rightarrow$ on conclut $H_1$.
- si $\omega \notin \mathcal{R}_n^\alpha \Rightarrow$ on ne rejette pas $H_0$.

$k_\alpha$ ? & ordi ou (BT)/(P) ou (TC) & lim.

(4)

Vocab : • taille du test : $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\mathcal{R}_n^\alpha)$

• Erreur de 1$^e$ espèce ($E_1$) : conclure $H_1$ à tort

• " 2$^{nde}$ " ($E_2$) : " $H_0$ " " si $\theta \in \Theta_1$, $\beta(\theta) = \mathbb{P}_\theta(\mathcal{R}_n^{\alpha c})$

Proba de conclure $H_0$ à tort : $\sup_{\theta \in \Theta_1} \mathbb{P}_\theta(\mathcal{R}_n^{\alpha c})$.

• Puissance du test : proba cclre $H_1$ à raison. si $\theta \in \Theta_1$,

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

$f$ puissance $\Theta_1 \longrightarrow \mathbb{R}$

$\theta \longmapsto \mathbb{P}_\theta(\mathcal{R}_n^\alpha)$.

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

• Test sans biais : si $\forall \theta \in \Theta_1$, $\pi(\theta) > \alpha$ err$^r$ de 2$^{nde}$ espèce

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

→ Voir exemples (croisem$^t$ de fleurs) (hypo simple)

(Rq) Une (va) $X$ est dite stochastiq$^t$ + grd qu'une (va) $Y$ si $\forall x \in \mathbb{R}$ : $F_X(x) \leq F_Y(x)$

(Rq) Si on ↘ niv° $\alpha \Rightarrow$ puiss$^{ce}$ ↘ $\Leftrightarrow E_2$ ↗

(hypo multiple : @ médicament)

si $S_n \sim Bin(n, \theta)$ ss $\mathbb{P}_\theta \Rightarrow \theta \mapsto \mathbb{P}_\theta(S_n > k)$ ↗ sur $]0,1[$.

↳ $\boxed{M1}$ mq $f'(\theta) \geq 0$ où $f(\theta) = \mathbb{P}_\theta(S_n > k) = \sum_{i=k}^n C_n^i \theta^i (1-\theta)^{n-i}$

↳ $\boxed{M2}$ simuler loi binom & $U$ (va) uniforme sur $]0,1[$

↳ Il en : $X = \mathbb{1}\{U \leq \theta\}$, $\mathbb{P}(X=1) = \mathbb{P}(U \leq \theta) = \theta$
loi de Bern $\mathbb{P}(X=0) = 1 - \theta$

Hypothèse simple : $\Theta_0$ ou $\Theta_1$ st singletons.

Test unilatéral $H_0 : \theta = \theta_0$ Zone de Rejet : $R_n^\alpha = \{S_n \geq k_\alpha\}$
$H_1 : \theta = \theta_1$ où $k_\alpha$ vérifie $\mathbb{P}_{\theta_0}(S_n \geq k_\alpha) \leq \alpha$.

$$k_\alpha = \underset{0 \leq k \leq n}{\operatorname{argmin}} \{ \mathbb{P}_{\theta_0}(S_n \geq k) \leq \alpha \}.$$

Test bilatéral : $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$ | $R_n^\alpha = \{S_n \leq k_\alpha^1\} \cup \{S_n \geq k_\alpha^2\}$ tq $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(R_n) \leq \alpha$.

$k_\alpha^1, k_\alpha^2$ vérifient $\mathbb{P}_\theta(\{S_n \leq k_\alpha^1\} \cup \{S_n \geq k_\alpha^2\}) \leq \alpha$.

$$k_\alpha^1 = \max\left(k \mid \mathbb{P}_\theta(S_n \leq k) \leq \frac{\alpha}{2}\right)$$

$$k_\alpha^2 = \min\left(k \mid \mathbb{P}_\theta(S_n \geq k) \leq \frac{\alpha}{2}\right)$$

$H_0 : \theta \leq \theta_0$
$H_1 : \theta > \theta_0$ : $k_\alpha = \min\{k \mid \sup_{\theta \leq \theta_0} \mathbb{P}_\theta(S_n \geq k) \leq \alpha\}$.

$\overline{X}_n \xrightarrow[n\to\infty]{\text{proba}} \theta$ $\forall \varepsilon > 0$ si $\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \theta| \geq \varepsilon) = 0$

$$\begin{cases} \overline{X}_n \xrightarrow[n\to\infty]{\text{proba}} \theta \\ \sum_{n \in \mathbb{N}} \mathbb{P}(|\overline{X}_n - \theta| \geq \varepsilon) < \infty \end{cases} \Rightarrow \overline{X}_n \xrightarrow[n\to\infty]{p.s} \theta$$

Estimat$^{r}$ sans biais si $\mathbb{E}(\hat{\theta}) = \theta$ sous $\theta$.

Puissance du test : conclure $H_1$ à raison.

$B_n = \bigcup_{k=n}^{\infty} A_k \searrow$ , $C_n = \bigcap_{k=n}^{\infty} A_k \nearrow$

de limite nulle étant le reste d'une série CV.

$\mathbb{P}\left(\bigsqcup_{n=N}^{\infty} A_n\right) = \sum_{n=N}^{\infty} \mathbb{P}(A_n)$ ; $\mathbb{P}\left(\bigcup_{n=N}^{\infty} A_n\right) \leq \sum_{n \geq N}^{\infty} \mathbb{P}(A_n) \leq \sum_{n=0}^{\infty} \mathbb{P}(A_n) = 0$

$$\sum_{n=u_0}^{N} q = \frac{1-q^{N+1}}{1-q} u_0 , \quad \sum_{n=u_0}^{\infty} q = \frac{1}{1-q} \times u_0$$



$\forall q \in ]0,1[$

# Ⓒ Généralisat°

## I/ Modéliser

• Modèle statistiq : $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ où
$(\Omega,\mathcal{F})$ : espace probabilisable, $(\mathbb{P}_\theta)_{\theta\in\Theta}$ famille de probas sr cet espace.
→ $\Theta$ espace de param, spps $\Theta \subset \mathbb{R}^d$ ⇒ stat nn param

• Observats : $(X_1,\dots,X_n)$ un n échantillon de la loi $P_\theta$ sous $\mathbb{P}_\theta$ (loi commune des $(X_i)_{1\le i\le n}$) ; va iid

⇒ OB estimer θ sous g(θ) & tester, ⇒ outils : vraisemblce.

## II/ Vraisemblce

① Lois discrètes : $P_\theta$ la loi de $X_1$ sous $\mathbb{P}_\theta$, spps discrète.
soit $\underbrace{\mathcal{X}}_{\text{au plus dénombrable}}$ $\mathbb{P}_\theta(X\in\mathcal{X})=1$

(taille N inconnue)

ⓐ 2 exemples ▸ Capture / Recapture : OB estimer nb poissons ds lac.
capture → Démarche : on pêche 20 poissons → marqués & on les remet ds lac.
recapture → pêche 50 sans remise & on compte nbr de poissons marqués.

Modélisat° : $(\Omega,\mathcal{F},(\mathbb{P}_p)_{p\in]0,1[})$, $p=\frac{20}{N}$ : proba pêcher poisson marqué.
$X_1,\dots,X_{50}$ iid Bern(p)

$X_i = \begin{cases}1 & \text{si i° poisson pêché est marqué}\\ 0 & \text{sinon.}\end{cases}$ $S_{50}=\sum_{i=1}^{50}X_i$, $S_{50}(\omega)=4$.

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

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

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

Ce qui correspond à une estimat° de N, $\hat N(\omega)=\frac{20}{\hat p(\omega)}=250$

+ général^t, si on observe $S_{50}(\omega)=s$ ⇒ $\underset{p\in]0,1[}{\arg\max}\ L(p,s)=\frac{s}{50}$.

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

$\hat p=\frac{S_{50}}{50}$ & $\hat N=\frac{1000}{S_{50}}$.

▸ Loi géom : Modélise nbr cycles avt tomber enceinte. θ proba TE un mois donné.
X va = nbr cycles pr TE. $\mathbb{P}_\theta(X=k)=\theta(1-\theta)^{k-1}$, $X\in\mathbb{N}^*$,
cf tableau nbr cycles/nbr fs. On ne dispose pas données indi $(X_i)_{1\le i\le 100}$ mais de $(N_k)_{1\le k\le 13}$ : $N_k=\sum_{i=1}^{100}\mathbb{1}\{X_i=k\}=\sum_{i=1}^{100}\mathbb{1}\{X_i>12\}$

Comment estimer θ ?

→ $\mathbb{P}_\theta(X_1=1)=\theta$, $\hat\theta_1=\frac{1}{100}\sum_{i=1}^{100}\mathbb{1}\{X_i=1\}$

$\mathbb{E}_\theta(\hat\theta_1)\overset{(*)}{=}\frac{1}{100}\sum_{i=1}^{\infty}\mathbb{P}_\theta(X_i=1)=\theta$ (*) car $\hat\theta_1$ est estimat^r sans biais de θ.

$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i=1\}\xrightarrow[n\to\infty]{\mathbb{P}_\theta\ ps}\theta$, $\hat\theta_1(\omega)=\frac{29}{100}$.

Défaut → n'utilise pas ttes données, $\mathbb{E}_\theta(X_1)=\frac{1}{\theta}$ ⇔
⇒ La moyenne empiriq des observaᵗˢ $\theta = 1/\mathbb{E}_\theta(X_1)$.
puissance −1 ê un bon candidat.

→ Comment f^r cette moyenne empiriq ? (suppr 3° donnée)

$$\hat\theta_2=\frac{93}{\sum_{k=1}^{12}kN_k}=\frac{100-N_{13}}{\sum_{k=1}^{100}kN_k}=\frac{\sum_{i=1}^{100}\mathbb{1}\{X_i\le 12\}}{\sum_{i=1}^{100}X_i\,\mathbb{1}\{X_i\le 12\}}$$