$\longrightarrow$ Convergences.

La suite de variables aléatoires X_n converge vers X d'accord, mais il est impératif de préciser de quelle manière.

■ La motivation.

D – Tiens, posons S_n la proportion de piles sur n lancers de pièce équilibrées. Alors S_n tend presque sûrement vers $\frac{1}{2}$, et la loi $2\sqrt{n}(X_n - \frac{1}{2})$ tend vers une Gaussienne centrée réduite ...
T - ... par la Loi Forte des Grands Nombres et le Théorème Central Limite ...
D - mais qu'est-ce qu'on peut dire de $Y_n = \frac{1}{X_n}$?
T - $X \mapsto 1/X$ est continue, donc Y_n tend presque sûrement vers 2, et un théorème dit que $\frac{\sqrt{n}}{2}(Y_n - 2)$ tend aussi vers une $\mathcal{N}(0,1)$.

■ Convergences.

Nous allons définir les différents types de convergence, dans ces définitions, $(X_n)_{n\geqslant 0}$ est une suite de variables aléatoires et X est une variable aléatoire.

- On dit que $(X_n)_{n\geqslant 0}$ converge **en loi** vers X, et on note $X_n \xrightarrow[n\to\infty]{\text{Loi}} X$, si en tout point de continuité t de la fonction de répartition de X, F_X, $F_{X_n}(t) \xrightarrow[n\to\infty]{} F_X(t)$ Ou alors si, pour tout t réel, $\varphi_{X_n}(t) \xrightarrow[n\to\infty]{} \varphi_X(t)$ (φ_X fonction caractéristique de X).
- On dit que $(X_n)_{n\geqslant 0}$ converge **en probabilité** vers X, et on note $X_n \xrightarrow[n\to\infty]{\text{proba}} X$, si pour tout ϵ strictement positif $\mathbb{P}(|X_n - X| > \epsilon) \xrightarrow[n\to\infty]{} 0$
- On dit que $(X_n)_{n\geqslant 0}$ converge **presque sûrement** vers X, et on note $X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-p.s.}} X$, s'il existe un ensemble négligeable $N \subseteq \Omega$ tel que pour tout ω dans $\Omega \backslash N$, $X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)$

Si $X_n \xrightarrow[\mathbb{P}-\text{ps}]{n\to\infty} X$, et $g : \mathbb{R} \to \mathbb{R}$ continue, alors $g(X_n) \xrightarrow[\mathbb{P}-\text{ps}]{n\to\infty} g(X)$. Et si $X_n \xrightarrow[\mathbb{P}-\text{ps}]{n\to\infty} X$ et $Y_n \xrightarrow[\mathbb{P}-\text{ps}]{n\to\infty} Y$, alors $(X_n, Y_n) \xrightarrow[\mathbb{P}-\text{ps}]{n\to\infty} (X,Y)$. C'est aussi vrai avec une convergence en probabilité. La première est vraie pour la convergence en loi.

■ Loi Forte des Grands Nombres.

Soit $(X_n)_{n\geqslant 0}$ une suite de v.a. i.i.d. et L^1 alors

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbb{P}\text{ ps}} \mathbb{E}(X_1)$$

■ Théorème Central Limite.

Soit $(X_n)_{n\geqslant 0}$ une suite de v.a. i.i.d. et L^2 alors,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\right) \xrightarrow[n\to\infty]{\text{Loi}} Z \sim \mathcal{N}(0, \text{Var}(X_1))$$

■ Lemme de Borel-Cantelli.

Soit $(A_n)_{n\geqslant 0}$ une suite d'évènements tels que la somme de la probabilité de ces évènements est finie, alors la probabilité qu'une infinité d'entre eux se réalise simultanément est nulle. Explicitement, ça veut dire :

$$\sum_{k=0}^{\infty} \mathbb{P}(A_k) < \infty \Longrightarrow \mathbb{P}(|\{n \in \mathbb{N} : A_n \text{ se réalise}\}| < \infty) = 1$$

■ Lemme de Slutsky.

Soient $(X_n)_{n\geqslant 0}$ et $(Y_n)_{n\geqslant 0}$ deux suites de variables aléatoires, X une variable aléatoire et c un réel. Si $X_n \xrightarrow[n\to\infty]{\text{Loi}} X$ et $Y_n \xrightarrow[n\to\infty]{\text{Loi}} c$ (ce qui est équivalent à dire que $Y_n \xrightarrow[n\to\infty]{\mathbb{P}\text{ proba}} c$) alors pour $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ telle que $\mathbb{P}\left((X,c) \in C^0(f)\right) = 1$ alors $f(X_n, Y_n) \xrightarrow[n\to\infty]{\text{Loi}} f(X,c)$

■ Théorème de Cramér.

(Pour cette section, voir d'abord la fiche suivante.)

Supposons qu'on ait $\hat{\theta}_n$ estimateur du paramètre θ, et qu'on ait la convergence presque sûre de $\hat{\theta}_n$ vers θ. On souhaite estimer $h(\theta)$ pour h une fonction sympathique.

Par exemple, la moyenne empirique $\overline{X}_n$ le paramètre p pour une loi binomiale ; et on peut l'utiliser pour estimer la variance de cette loi (pour rappel, $np(1-p)$).

Grâce au Théorème Centrale Limite, on peut dire que $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en loi vers une Gaussienne $\mathcal{N}(0,1)$. Peut-on en dire de même pour notre estimateur de la variance ? C'est là qu'intervient le théorème de Cramér.

Voici le **théorème de Cramér**. Soit $(\alpha_n)_n$ une suite réelle avec $\lim\limits_{n\to+\infty} \alpha_n = +\infty$, $(\hat{\theta}_n)$ une suite de variables aléatoires, θ un réel et $h : \mathbb{R} \to \mathbb{R}$ une fonction $\mathcal{C}^1$.

Si $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}\text{ ps}} \theta$, et $\alpha_n(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{\text{Loi}} Z$ avec Z suivant une loi non-dégénérée, alors $\alpha_n(h(\hat{\theta}_n) - h(\theta)) \xrightarrow[n\to\infty]{\text{Loi}} h'(\theta)Z$.

Typiquement, on aime bien utiliser $\alpha_n = \sqrt{n}$ pour une convergence donnée par le T.C.L.

$\longrightarrow$ Estimateurs.

C'est toujours sympa de modéliser la réalité, encore mieux si on connait les paramètres.

La motivation.

D – J'ai obtenu plein de valeurs d'une variable continue dont ne je ne connais pas la fonction de répartition. Tu penses qu'on peut la retrouver ? Genre, comment tu obtiens une allure approximative de F ?
T - Ici, on peut approximer $F(t)$ par $\hat{F}_n(t)$, qu'on définit comme la proportion des observations inférieures à t.
D - Pourquoi un F "chapeau" ?
T - Ah ! C'est parce que c'est un problème d'estimation, on cherche à approximer la valeur d'un paramètre. Et usuellement, on utilise un "chapeau" pour les estimateurs (qui est une v.a. approximant le paramètre).

Contexte.

On considère n variables aléatoire (v.a.) $(X_1, \dots, X_n)$ indépendantes et identiquement distribuées (i.i.d.) suivant une loi P_X. Soit θ, un paramètre de la loi P_X (comme l'est λ dans la loi de Poisson par exemple) inconnu, on va tenter d'estimer la valeur de θ à l'aide des v.a. $(X_1, \dots, X_n)$. On devra alors créer un estimateur que l'on notera $\hat{\theta}_n$.

Un **estimateur** θ_n de θ est une variable aléatoire de la forme $g(X_1, \dots, X_n)$ où g est une fonction connue.

Propriétés.

Voici plusieurs propriétés que peut avoir un estimateur.

- $\hat{\theta}_n$ est un estimateur **sans biais** de θ si $\mathbb{E}[\hat{\theta}_n] = \theta$, sinon il est dit **biaisé**. Si $\hat{\theta}_n$ est biaisé, son biais est $\mathbb{E}[\hat{\theta}_n] - \theta$.
- $\hat{\theta}_n$ est un estimateur **asymptotiquement sans biais** de θ si $\mathbb{E}[\hat{\theta}_n] \xrightarrow[n\to\infty]{} \theta$ (ou si $\mathbb{E}[\hat{\theta}_n - \theta] \xrightarrow[n\to\infty]{} 0$).
- $\hat{\theta}_n$ est un estimateur **faiblement consistant** de θ si $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}\text{ proba}} \theta$ "en probabilité". (Comme θ est une constante, la convergence en probabilité est équivalente à la convergence en loi)
- $\hat{\theta}_n$ est un estimateur **fortement consistant** de θ si $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}\text{ ps}} \theta$ "presque sûrement".
- $\hat{\theta}_n$ est un estimateur **convergeant au sens** L_p de θ si $\hat{\theta}_n \xrightarrow[n\to\infty]{L^p} \theta$ "dans L^p", c'est à dire si $\mathbb{E}\left[|\hat{\theta}_n - \theta|^p\right] \xrightarrow[n\to\infty]{} 0$

Proposition : Critère de convergence au sens L^2 Soit $\hat{\theta}_n$, un estimateur de θ. Si $\hat{\theta}_n$ est un estimateur asymptotiquement sans biais de θ et que $\mathrm{Var}(\hat{\theta}_n) \xrightarrow[n\to\infty]{} 0$ alors $\hat{\theta}_n$ est un estimateur convergeant au sens L_2 de θ.

Problème-exemple.

On suppose maintenant les X_i i.i.d. réelles ; on s'intéresse à F la fonction de répartition de X_1 ($F(t) = P(X_1 \leq t)$) ; pour rappel, F caractérise la loi de X_1. Le paramètre qu'on souhaite estimer est $\theta_t := F(t)$ (ici F est l'ensemble des lois sur $\mathbb{R}$).

Estimer la fonction de répartition.

Notre estimateur, appelé "processus empirique", est défini comme

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{]-\infty,t]}(x_i) = \frac{|\{i : x_i(\omega) \leq t\}|}{n}$$

Notons que $nF_n(t)$ suit une loi binomiale $\mathcal{B}(n, F(t))$. Ainsi, $\mathbb{E}[F_n(t)] = F(t)$, et $\mathrm{Var}(F_n(t)) = \frac{F(t)(1-F(t))}{n}$. On a donc un estimateur sans biais, convergent au sens L^2. Pour tout ω, $t \mapsto F_n(t,\omega)$ est une fonction de répartition.

Intervalle de confiance.

On a trouvé une méthode pour approximer $F(t)$; on souhaiterait avoir une idée d'à quel point on se trompe. Fixons-nous un $\alpha \in [0, \frac{1}{2}]$. On veut alors trouver un intervalle I, tel que $\mathbb{P}(F(t) \in I) \underset{n}{\to} 1 - \alpha$.

Rappelons rapidement : si $Z_n \underset{loi}{\longrightarrow} \mathcal{N}(0,1)$, alors en aura que $\mathbb{P}(Z_n \in [q_{\alpha/2}, q_{1-\alpha/2}]) \to 1-\alpha$, avec q_a le quantile d'ordre a de la Gaussienne centrée réduite (c'est-à-dire le q tel que $\mathbb{P}(X \leq q_a) = a$ avec $X \sim \mathcal{N}(0,1)$). En l'occurrence, $q_{1-\frac{\alpha}{2}} = -q_{\frac{\alpha}{2}}$. Et si $Z_n = (a - b_n)/c_n$ avec a constant, et b_n, c_n des v.a., alors on aura $\mathbb{P}(a \in [b_n - q_{1-\frac{\alpha}{2}}c_n, b_n + q_{1-\frac{\alpha}{2}}c_n]) \to 1-\alpha$. On va utiliser cette idée plusieurs fois. Rappelons que $\hat{F}_n(t)$ suit une loi binomiale de paramètres $F(t)$ et n. Donc $\sqrt{n}\frac{(\hat{F}_n(t)-F(t))}{\sqrt{F(t)(1-F(t))}}$, et $\sqrt{n}\frac{(\hat{F}_n(t)-F(t))}{\sqrt{\hat{F}_n(t)(1-\hat{F}_n(t))}}$ tendent en loi vers $\mathcal{N}(0,1)$. Et donc, en posant la dispersion

$$D_n = q_{1-\frac{\alpha}{2}}\frac{\sqrt{\hat{F}_n(t)(1-\hat{F}_n(t))}}{\sqrt{n}}$$

(notons que $D_n \sim \frac{c}{\sqrt{n}}$) et l'intervalle de confiance

$$I_n = [\hat{F}_n(t) - D_n, \hat{F}_n(t) + D_n]$$

on a bien que $\mathbb{P}(F(t) \in I_n) \underset{n}{\to} 1 - \alpha$.

Application de Cramér.

On a $\sqrt{n}\cdot(\hat{F}_n(t) - F(t)) \xrightarrow[n\to\infty]{\text{Loi}} \mathcal{N}(0, F(t)\cdot(1-F(t)))$.
Si $F(t) \in]0,1[$, on aura aussi, par théorème de Cramér

$$\sqrt{n}\cdot\left(\hat{F}_n(t)^2 - F(t)^2\right) \xrightarrow[n\to\infty]{\text{Loi}} \mathcal{N}\left(0, 4\cdot F(t)^3(1-F(t))\right)$$

$\longrightarrow$ Construction d'estimateurs.

Bien sûr que X_1 est un estimateur de X_1 ... mais on peut mieux faire ! Comment ?

■ La motivation.

D – Comment peut-on construire un estimateur ?
T - On peut facilement estimer l'espérance par une moyenne empirique, la variance et les moments suivants. Donc si on peut retrouver le paramètre avec, on a un estimateur.
D - On ne pourrait pas juste essayer de prendre le paramètre qui donne aux résultats le plus de chances de se produire ?
T - Si, c'est un estimateur du maximum de vraisemblance.

■ Contexte.

On cherche maintenant des méthodes qui permettent de construire explicitement des estimateurs pour des situations connues. On présente ici deux méthodes particulièrement courantes : l'estimateur des moments, et l'estimateur du maximum de vraisemblance.

Ici, le contexte choisi est celui de variables X_i i.i.d., de même loi P_θ, où $\theta \in \Theta$ et $P_\theta \in \mathcal{F}$, Θ l'ensemble des valeurs possibles du paramètre et $\mathcal{F} = \{P_\theta, \theta \in \Theta\}$ l'ensemble des lois possibles. Le vecteur $\overline{X} = (X_1, \cdots, X_n)$ est appelé échantillon. Et on souhaite estimer θ (ou une fonction de θ) avec une fonction de $\overline{X}$.

■ Moments empiriques.

On rappelle que le moment d'une variable aléatoire X est $\mu_k = \mathbb{E}[X^p]$ quand cette espérance existe. On va d'ailleurs supposer que X est une v.a. $\mathcal{L}^p$ (c'est-à-dire $\mathbb{E}[|X|^p] < +\infty$).

On définit alors les **moments empiriques** comme : $\hat{\mu}_n(k) = \sum_{i=1}^n X_i^k$.

On a bien sûr que $\mathbb{E}[\hat{\mu}_n(k)] = \mu_k$. C'est même un estimateur fortement consistant des μ_k. On a des théorèmes plus précis sur la convergence. Si maintenant $2k \leq p$, alors on peut appliquer le Théorème Central Limite pour avoir $\sqrt{n}(\hat{\mu}_n(k) - \mu_k) \to_{loi} \mathcal{N}\left(0, \sigma_k^2\right)$ avec $\sigma_k^2 = \mu_{2k} - \mu_k^2$.

■ Intervalle de confiance.

Et donc on peut encore donner un intervalle de confiance, à savoir qu'on a $\mathbb{P}(\mu_k \in I_n) \to 1 - \alpha$. On pose $\hat{\sigma}_n^2(k) := \hat{\mu}_n(2k) - \hat{\mu}_n(k)^2$, on a alors

$$I_n = \left[\hat{\mu}_n(k) - q_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}_n(k)}{\sqrt{n}}, \hat{\mu}_n(k) + q_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}_n(k)}{\sqrt{n}}\right]$$

Grâce à la section suivante, on a de quoi donner un intervalle de confiance pour l'espérance et la variance empirique ; utile par exemple pour les paramètres d'une loi normale.

■ Plus de détails.

Soit $(X_1, \cdots, X_n)$ des v.a. i.i.d. dans $\mathcal{L}^k$, posons $\overline{X}_n = \mu_n(1)$. Alors, on a :

- $k \geq 1$: $\overline{X}_n \underset{p.s.}{\longrightarrow} \mathbb{E}[X_1]$
- $k \geq 2$: $\hat{\mu}_2(n) - \hat{\mu}_1(n)^2 \to \mathrm{Var}(X_1)$ donc $\sqrt{n}(\overline{X}_n - \mu) \to \mathcal{N}(0, \mathrm{Var}(X_1))$.
- $k \geq 4$: avec $\hat{\sigma}_n^2 = \hat{\mu}_n(2) - \hat{\mu}_n(1)^2$, on a même $\sqrt{n}\left(\hat{\sigma}_n^2 - \sigma^2\right) \to \mathcal{N}\left(0, \tau^2\right)$ avec $\tau^2 = \mu_4 - 4\mu_1\mu_3 + 4\mu_1^2\mu_2 - \sigma^4$.

■ Estimateur des moments.

Un **estimateur des moments** est un estimateur de la forme $f(\hat{\mu}_n(1), \cdots, \hat{\mu}_n(k))$ avec $f : \mathbb{R}^k \to \Theta$ sympathique. On suppose bien sûr que X_1 est L^k et que $\theta = f(\mu_1, \cdots, \mu_k)$. Dans l'idée, on cherche une formule qui donne le paramètre en fonction des moments, et ça nous donne un estimateur, en remplaçant les moments par les estimateurs correspondants.

■ Maximum de vraisemblance.

Une autre méthode, souvent meilleure, est de chercher à maximiser la vraisemblance. Celle-ci donne la probabilité d'obtenir l'échantillon, en fonction du paramètre $\theta \in \Theta$. On a deux cas : on étudie des lois discrètes. On utilise alors la loi, $f_\theta(x) = \mathbb{P}_\theta(X = x)$; et $\mathbb{P}(X_1 \in A) = \sum_{x \in A \cap E} f_\theta(x)$. on étudie des lois continues. Dans ce cas, on utilise la fonction de masse comme f_θ, et Dans les deux cas, on posera la fonction de vraisemblance comme $V(\theta) := \prod_{i=1}^n f_\theta(x_i)$. Et on va donc construire l'estimateur de vraisemblance $\hat{\theta}_N^{MV}$ comme $g(X_1, \cdots, X_n)$ où g est la fonction qui à $(x_1, \cdots, x_n)$, associe le θ qui maximise $V(\theta)$. Autrement dit, quand il existe, $\forall \theta \in \Theta, V(\hat{\theta}_N) \geq V(\theta)$ et donc $\hat{\theta}_N^{MV} = \mathrm{argmax}_{\theta \in \Theta} V(\theta)$.

Il est souvent pratique, quand V ne s'annule pas, d'étudier $L(\theta) := \log(V(\theta))$ à la place (souvent plus facile à dériver).

On admet le pseudo-théorème suivant : pour une large classe de modèles, si l'estimateur du maximum de vraisemblance est sujet aux conditions $\hat{\theta}_n^{MV} \xrightarrow{p.s.} \theta$ et $\sqrt{n}\left(\hat{\theta}_n^{MV} - \theta\right) \to \mathcal{N}\left(0, \sigma^2(\theta)\right)$, alors $\sigma^2(\theta)$ est la plus petite variance minimale qu'on peut avoir pour un estimateur ; pour tout estimateur $\hat{\theta}_n$ avec $\sqrt{n}(\hat{\theta}_n - \theta) \to \mathcal{N}(0, \tau(\theta))$, on aura $\tau(\theta) \geq \sigma^2(\theta)$.

$\longrightarrow$ Tests d'hypothèse.

Le formalisme scientifique aime bien formuler des hypothèses, et créer des tests afin de pouvoir les accepter, ou rejeter.

■ La motivation.

D - J'ai lancé une pièce 12 fois, et j'ai beaucoup de piles. Est-ce que ça veut dire que la pièce est biaisée ?
T - Pour une pièce équilibrée, la probabilité qu'on ait entre 3 et 9 piles (inclus) est de 96%. Donc, cela donne un test de niveau 96% : on accepte l'hypothèse "la pièce est biaisée" si on a moins de 3 ou plus de 9 piles.
D - J'en ai eu 8, donc ce n'est pas significatif ?
T - Le test accepte l'hypothèse "la pièce n'est pas biaisée". Mais pour être encore plus sûr, il faudrait faire plus de lancers.

■ Intervalle de confiance.

Soit $\alpha \in]0, \frac{1}{2}[$. Deux v.a. $L(X_1, \cdots, X_n)$ et $U(X_1, \cdots, X_n)$ avec $L \leq U$ définissent un **intervalle de confiance** de niveau $1 - \alpha$ si pour tout $\theta \in \Theta$, et X_i i.i.d. suivant une loi P_θ, on a $\mathbb{P}(\theta \in [L(X_1, \cdots, X_n), U(X_1, \cdots, X_n)]) = 1 - \alpha$.

Il est dit asymptotique si on a plutôt que $\mathbb{P}(\theta \in [L(X_1, \cdots, X_n), U(X_1, \cdots, X_n)]) \xrightarrow[n\to\infty]{} 1 - \alpha$.

■ Fonction pivotale.

Soit une fonction $h : \Theta \times E^n \to \mathbb{R}$, et la variable aléatoire $h(\theta, X_1, \cdots, X_n)$ avec X_i i.i.d. suivant P_θ. La fonction h est alors dite **pivotale** si la loi de $h(\theta, X_1, \cdots, X_n)$ ne dépend plus de θ, et **asymptotiquement pivotale** si $h(\theta, X_1, \cdots, X_n) \xrightarrow[loi]{} Z$ avec Z ne dépend pas de θ.

■ Construction.

Supposons maintenant que la fonction [asymptotiquement] pivotale donne [tende vers] une loi Z continue. On peut alors définir les quantiles de Z comme $\zeta(\alpha) := \inf \{x | F_z(x) \geq \alpha\}$ (on souhaite avoir $F_Z(\zeta(\alpha)) = \alpha$. Alors, posons $g_{\underline{X}}(\theta) = h(\theta, X_1, \cdots, X_n)$. En posant $R(\underline{X}, \alpha) = g_{\underline{X}}^{-1}\left(\left[\zeta\left(\frac{\alpha}{2}\right), \zeta\left(1 - \frac{\alpha}{2}\right)\right]\right)$, on a $\mathbb{P}(\theta \in R(\underline{X}, \alpha)) = 1 - \alpha$.

Le $R(\underline{X}, \alpha)$ est un intervalle, on parle d'un intervalle de confiance ; sinon, c'est juste une région de confiance.

Plutôt que $\left[\frac{\alpha}{2}, 1 - \frac{\alpha}{2}\right]$, on peut remplacer par n'importe quel intervalle $[\alpha_1, 1 - \alpha + \alpha_1] \subseteq [0, 1]$, et même choisir α_1 de manière à minimiser la longueur de $g_{\underline{X}}^{-1}([\zeta(\alpha_1), \zeta(1 - \alpha + \alpha_1)])$.

■ Hypothèses.

Nous avons construit, pour $X_1, \cdots, X_n$ suite de v.a. i.i.d. selon une loi P_θ, une estimation ponctuelle $\hat{\theta}_n$ et un intervalle de confiance $I(\underline{X})$.

Soit un $\theta_0 \in \Theta$ fixé ; on veut déterminer un critère pour déterminer si $X_1, \cdots, X_n$ i.i.d. sont distribuées selon F_{θ_0}. Pour ça, on pose l'hypothèse H_0 : "$\theta = \theta_0$, qu'on appelle **hypothèse nulle** et on pose l'hypothèse contraire, dite **hypothèse alternative**, H_1 : "$\theta \neq \theta_0$". Un **test** est une méthode permettant d'accepter ou rejeter l'hypothèse nulle selon les valeurs obtenues.

■ Test d'hypothèse.

Un test est une méthode qui, en fonction des résultats, accepte une des hypothèses, et rejette l'autre. Bien sûr, un test n'est pas infaillible, il peut échouer ; auquel cas, on a une erreur de type I ou de type II.

	On accepte H_0, et rejette H_1	On accepte H_1, et rejette H_0
H_0 vraie	—	erreur de type I
H_1 vraie	erreur de type II	—

On va construire nos tests sous forme d'une région $R \subset E^n$, et du critère $(X_1, \cdots, X_n) \in R$ pour accepter H_1 (et rejeter H_0). Pour une erreur du type I, on a $\theta = \theta_0$ mais $(X_1, \cdots, X_n) \in R$. On peut caractériser une telle erreur par un réel $\alpha = \mathbb{P}((X_1, \cdots, X_n) \in R)$ quand les X_i sont i.i.d. selon une loi F_{θ_0}. Pour une erreur de type II, on a $\theta \neq \theta_0$ mais $(X_1, \cdots, X_n) \notin R$. Il faut donc utiliser une fonction $\beta_n : \Theta \backslash \{\theta_0\} \to [0, 1]$ qui à θ, donne $\mathbb{P}((X_1, \cdots, X_n) \notin R)$ avec des X_i i.i.d. selon F_θ. Quand on construit un test, on dit qu'il est de niveau α, et il est dit consistant si $\beta(\theta) \xrightarrow{n\to\infty} 0$. Si on a trouvé un intervalle de confiance (asymptotique) $[L(X_1, \cdots, X_n), U(X_1, \cdots, X_n)]$ de niveau $1 - \alpha$, on peut prendre la région $R = \{(X_1, \cdots, X_n) | \theta_0 \in [L(X_1, \cdots, X_n), U(X_1, \cdots, X_n)]\}$.

$\longrightarrow$ Exemples.

Pour le dessert des lecteurs·trices, voici un plateau d'exemples.

■ Estimateur des moments.

D'abord X suivant une loi de Poisson dont on veut estimer le paramètre p, on a plusieurs choix, puisque $\theta = \mu_1 = \mu_2 - \mu_1^2 = \frac{\sqrt{1+4\mu_2}-1}{2}$. Donc on pourrait prendre $\hat{\theta}_n = \hat{\mu}_n(1)$ ou $\hat{\theta}_n = \hat{\mu}_n(2) - \hat{\mu}_n(1)^2 = \hat{\sigma}_n^2$. Il s'agit dorénavant de savoir lequel converge le plus vite (ou le mieux). Il se trouve que le premier donne une variance de σ^2, et le second de τ^2 ; il se trouve qu'ainsi, le premier converge plus vite.

■ Maximum de vraisemblance.

Si les X_i suivent une loi de Poisson $\mathcal{P}(\theta)$ de paramètre θ à déterminer, la fonction de vraisemblance est

$$V(\theta) = \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{x_i}}{x_i!} = e^{-n\theta} \frac{\theta^{x_1+\cdots+x_n}}{x_1! \cdots x_n!}$$

Ainsi a-t-on comme logarithme de vraisemblance :

$$L(\theta) = -n\theta + (x_1 + \cdots + x_n) \log(\theta)$$

En dérivant L, on obtient $L'(\theta) = \frac{x_1+\cdots+x_n}{\theta} - n$, donc s'annule et change de signe en $\theta = \frac{x_1+\cdots+x_n}{n}$. Donc on y atteint un maximum, qui est aussi un maximum de la vraisemblance, et on prendra donc $\overline{X}_n$ comme estimateur du maximum de vraisemblance. Fréquemment, on a que μ_1 est le paramètre étudié, et que $\overline{X}_n$ fonctionne à la fois comme estimateur des moments et du maximum de vraisemblance.

Si cette fois on étudie des lois exponentielles $\mathcal{E}(\theta)$, on a alors

$$V(\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} \mathbb{I}_{x_i>0}$$

On supposera donc tous les x_i strictement positifs, alors $L(\theta) = n \log\theta - \theta(x_1 + \cdots + x_n)$ et donc $L'(\theta) = \frac{n}{\theta} - (x_1 + \cdots + x_n)$, ce qui donne encore l'estimateur $\theta = \frac{x_1+\cdots+x_n}{n}$.

■ Fonction pivotale.

Si X_1 suit une loi uniforme $\mathcal{U}[0,\theta]$, alors $\frac{X_1}{\theta}$ suit une loi $\mathcal{U}[0,1]$, qui ne dépend donc plus de θ. Et de même, si on a X_i des v.a. i.i.d. selon une loi uniforme $\mathcal{U}[0,\theta]$, et $h : [0,1]^n \to \mathbb{R}$, alors $h\left(\frac{X_1}{\theta}, \cdots, \frac{X_n}{\theta}\right)$ suit une loi qui ne dépend pas de θ. Donc par exemple, $h(\theta, x_1, \cdots x_n) = \sum \frac{x_i}{\theta}$ ou $h(\theta, x_1, \cdots x_n) = \max\left(\frac{x_i}{\theta}\right)$ sont pivotales. Pour des X_i i.i.d. selon une loi de Poisson $\mathcal{P}(\theta)$, il est difficile de trouver des fonctions pivotales. Mais le T.C.L. nous donne la fonction asymptotiquement pivotale $h(\theta, x_1, \cdots, x_n) = \frac{\sqrt{n}}{\sqrt{\theta}}\left(\frac{1}{n}\sum X_i - \theta\right)$ puisqu'alors, $h(\theta, X_1, \cdots, X_n) \xrightarrow{loi} \mathcal{N}(0,1)$, et la loi de $\mathcal{N}(0,1)$ ne dépend pas de θ.

■ Construction d'un test.

On étudie des variables D_i i.i.d. suivant une loi $\mathcal{N}(\theta, 1)$ de paramètre θ inconnu, on souhaite mettre en place un test pour vérifier si $\theta > 0$. On pose les hypothèses H_0 (nulle) : $\theta = 0$ et H_1 (alternative) : $\theta \neq 0$.

Estimons θ par la moyenne empirique $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} D_i$; alors $\hat{\theta}_n$ suit une loi $\mathcal{N}(\theta, 1/n)$.

Puisque $D_i - \theta$ suit une loi $\mathcal{N}(0,1)$, alors pour n'importe quelle fonction $g : \mathbb{R}^n \to \mathbb{R}$ on a $g(D_1 - \theta, \cdots, D_n - \theta)$ dont la loi ne dépend pas de θ. Donc $g(\theta, x_1, \cdots, x_n) = h(x_1 - \theta, \cdots, x_n - \theta)$ est une fonction pivotale. On va prendre ici $h(x_1, \cdots, x_n) = x_1 + \cdots + x_n$. Alors

$$\frac{g(D_1, \cdots, D_n)}{\sqrt{n}} = \frac{\sum_{i=1}^{n} x_i - n\theta}{\sqrt{n}} \overset{loi}{\sim} \mathcal{N}(0,1)$$

Notons q_α le quantile d'ordre α de la loi normale centrée réduite. Alors

$$\mathbb{P}\left(\frac{g(D_1, \cdots, D_n)}{\sqrt{n}} \in [q_{\alpha/2}, q_{1-\alpha/2}]\right) = 1 - \alpha$$

et donc, après réarrangement :

$$\mathbb{P}\left(\theta \in \left[\hat{\theta}_n - \frac{q_{1-\alpha/2}}{\sqrt{n}}, \hat{\theta}_n + \frac{q_{1-\alpha/2}}{\sqrt{n}}\right]\right) = 1 - \alpha$$

Grâce à ça, on construit le test suivant : on accepte H_1 (et on rejette H_0) si

$$0 \notin \left[\hat{\theta}_n - \frac{q_{1-\alpha/2}}{\sqrt{n}}, \hat{\theta}_n + \frac{q_{1-\alpha/2}}{\sqrt{n}}\right]$$

Sinon on accepte H_0 et on rejette H_1.

Maintenant, on peut vérifier si le test est consistant. Supposons $\theta \neq 0$. On le rappelle, $\beta(\theta) = \mathbb{P}(\theta \in [\hat{\theta}_n - \frac{q_{1-\alpha/2}}{\sqrt{n}}, \hat{\theta}_n + \frac{q_{1-\alpha/2}}{\sqrt{n}}]$. Donc, c'est $\mathbb{P}(\hat{\theta}_n \in [\theta - \frac{q_{1-\alpha/2}}{\sqrt{n}}, \theta + \frac{q_{1-\alpha/2}}{\sqrt{n}}] = \Phi(\theta\sqrt{n} + q_{1-\alpha/2}) - \Phi(\theta\sqrt{n} - q_{1-\alpha/2})$, avec $\Phi(t) = \mathbb{P}(\mathcal{N}(0,1) \geq t)$. Quand $n \to \infty$, $\beta(\theta) \to 0$. Donc le test est bien consistant.