

Exercices de statistiques mathématiques

Guillaume Lecué

1^{er} septembre 2017

Table des matières

1	Rappels de probabilités	1
2	Vraisemblance, EMV, IC, Information de Fisher	11
3	Tests	23
4	Modèle de régression	28
5	Examen du lundi 26 octobre 2015	36
6	Rattrapage 2015-2016	40
7	Examen du lundi 14 novembre 2016	44
8	Rattrapage 2016-2017	51

1 Rappels de probabilités

Exercice 1.1 (Théorème de la limite centrale)

Soit $(X_n)_n$ une suite de variables aléatoires i.i.d. centrées de variance $\sigma^2 > 1$. Soit

$$Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n X_j.$$

Par le théorème de la limite centrale, cette variable converge en loi vers la loi normale centrée réduite, c'est-à-dire, pour tout $t \in \mathbb{R}$, on a $\lim_{n \rightarrow +\infty} \mathbb{E}[e^{itZ_n}] = e^{-\frac{t^2}{2}}$. L'objet de cet exercice est de montrer que la suite Z_n ne peut pas converger en probabilité.

1. Calculer la fonction caractéristique de $Z_{2n} - Z_n$ et montrer que cette différence converge en loi.
2. En étudiant $\mathbb{P}(|Z_{2n} - Z_n| \geq \epsilon)$, montrer que Z_n ne converge pas en probabilité.

Correction de l'exercice 1.1 L'objectif de cet exercice est de manipuler les différents types de convergence. On commence donc par rappeler les différentes convergences en probabilités. Soit (X_n) une suite de variables aléatoires et X une autre variable aléatoire. On dit que :

- (X_n) converge presque sûrement vers X quand $\{\omega \in \Omega : \lim X_n(\omega) = X(\omega)\}$ est de mesure 1 (on vérifiera que cet ensemble est bien mesurable).
- (X_n) converge en probabilité vers X quand pour tout $\epsilon > 0$, $\mathbb{P}[|X_n - X| \geq \epsilon] \rightarrow 0$ quand n tend vers $+\infty$.
- (X_n) converge en loi vers X quand pour toute fonction continue bornée f on a $\mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X)$.
- si $p \geq 1$, on dit que (X_n) converge dans L_p vers X quand $\mathbb{E}|X_n - X|^p \rightarrow 0$ quand n tend vers $+\infty$.

On a les implications suivantes :

$$\begin{array}{ccccc}
 \text{[cv presque sure]} & \xrightarrow{(1)} & \text{[cv en proba]} & \xrightarrow{(2)} & \text{[cv en loi]} \\
 & & \uparrow (3) & & \\
 & & \text{[cv dans } L_p] & &
 \end{array}$$

Démo et contre-exemple de “(1)” : Soit $\epsilon > 0$. On a $\{X_n \rightarrow X\} \subset \liminf_n \{|X_n - X| \leq \epsilon\}$. En passant, au complémentaire, on a :

$$\begin{aligned}
 0 &\leq \limsup_n \mathbb{P}[|X_n - X| > \epsilon] \leq \mathbb{P}[\limsup_n \{|X_n - X| > \epsilon\}] \\
 &= \mathbb{P}[(\liminf_n \{|X_n - X| \leq \epsilon\})^c] \leq 0.
 \end{aligned}$$

Il n'y a pas équivalence dans “(1) \Rightarrow ”. Voici une exemple d'une suite qui converge en probabilité mais pas presque sûrement : (X_n) des v.a. indépendantes telles que

$$\mathbb{P}[X_n = 1] = \frac{1}{n} \text{ et } \mathbb{P}[X_n = 0] = 1 - \frac{1}{n}.$$

La suite (X_n) converge en probabilité vers 0 car pour tout n , on $\mathbb{P}[|X_n| > \epsilon] = \mathbb{P}[X_n = 1] = 1/n$. Mais elle ne converge pas presque sûrement vers car on a $\sum_n \mathbb{P}(\{X_n = 1\}) = \infty$ donc d'après le “second lemme de Borel-Cantelli” (les événements $(\{X_n = 1\})$ sont indépendants), on a $\mathbb{P}[\limsup_n \{X_n = 1\}] = 1$. Notamment, (X_n) ne converge pas presque sûrement vers 0.

Démo et contre-exemple de “(2)” : Soit f une fonction continue bornée. Soit $\epsilon > 0$ et $N_\epsilon \in \mathbb{N}$ tel que $\mathbb{P}[|f(X_n) - f(X)| \geq \epsilon] \leq \epsilon$ (on rappelle que si f est continue et (X_n) converge en probabilité vers X alors $(f(X_n))$ converge en probabilité vers $f(X)$). On a donc

$$\begin{aligned}
 |\mathbb{E}f(X_n) - \mathbb{E}f(X)| &\leq |\mathbb{E}(f(X_n) - f(X))I(|f(X_n) - f(X)| \geq \epsilon)| \\
 &+ |\mathbb{E}(f(X_n) - f(X))I(|f(X_n) - f(X)| < \epsilon)| \\
 &\leq 2 \|f\|_\infty \mathbb{P}[|f(X_n) - f(X)| \geq \epsilon] + \epsilon \leq (2 \|f\|_\infty + 1)\epsilon.
 \end{aligned}$$

La réciproque est trivialement fautive. Il suffit de prendre la suite stationnaire (X_n) où pour tout n , $X_n = g$ où g est une gaussienne. Comme g est symétrique, $-g$ est aussi distribuée comme g . Donc

(X_n) converge en loi vers g et donc aussi vers $-g$. Par contre $|X_n - (-g)| = 2|g|$ ne converge pas en probabilité vers 0. Donc (X_n) ne converge pas vers $-g$ en probabilité.

Démo et contre-exemple de “(3) \uparrow ” : D’après l’inégalité de Markov, $\mathbb{P}[|X_n - X| \geq \epsilon] \leq \epsilon^{-p} \mathbb{E}|X_n - X|^p$. Pour le contre-exemple, on prend X_n de loi $(n^{-1}\delta_{n^2} + (1 - n^{-1})\delta_0)$. On a $\mathbb{P}[|X_n| \geq \epsilon] \leq n^{-1}$ donc (X_n) converge en probabilité mais $\mathbb{E}|X_n| = n$ donc (X_n) ne converge pas dans L_1 vers 0.

Correction de l’exercice

1. Pour tout $t \in \mathbb{R}$, on a par indépendance

$$\mathbb{E} \exp(it(Z_{2n} - Z_n)) = \mathbb{E} \exp\left(\frac{it}{\sigma\sqrt{n}}\left(\frac{1}{\sqrt{2}} - 1\right) \sum_{j=1}^n Z_j\right) \mathbb{E} \exp\left(\frac{it}{\sigma\sqrt{2n}} \sum_{j=n+1}^{2n} Z_j\right).$$

En appliquant le TCL sur chacun des membres du produit, quand n tend vers l’infini, on obtient que $(Z_{2n} - Z_n)_n$ tend vers une loi dont la fonction caractéristique est $t \mapsto \exp(-t^2(2 - \sqrt{2})/2)$, c’est donc une Gaussienne centrée de variance $\sqrt{2 - \sqrt{2}}$.

2. Supposons que (Z_n) converge en probabilité. Alors il existe une variable aléatoire Z telle que pour tout $\epsilon > 0$, on a $\mathbb{P}[|Z_n - Z| > \epsilon] \rightarrow 0$. Soit $\epsilon > 0$, on a

$$\{|Z_{2n} - Z_n| \geq 2\epsilon\} \subset \{|Z_n - Z| \geq \epsilon\} \cup \{|Z_{2n} - Z| \geq \epsilon\}.$$

Alors, par une borne de l’union :

$$\mathbb{P}[|Z_{2n} - Z_n| \geq 2\epsilon] \leq \mathbb{P}[|Z_n - Z| \geq \epsilon] + \mathbb{P}[|Z_{2n} - Z| \geq \epsilon]$$

et donc en passant à la limite, on obtient $\mathbb{P}[|Z_{2n} - Z_n| \geq 2\epsilon] \rightarrow 0$. Donc $(Z_{2n} - Z_n)_n$ converge en probabilité vers 0. En particulier, cette suite converge en loi vers 0. Ce qui est en contradiction avec 1..

Exercice 1.2 (Lemme de Slutsky)

1. Donner un exemple de suites (X_n) et (Y_n) telles que $X_n \xrightarrow{\text{loi}} X$ et $Y_n \xrightarrow{\text{loi}} Y$, mais $X_n + Y_n$ ne converge pas en loi vers $X + Y$.
2. Soient $(X_n), (Y_n)$ deux suites de variables aléatoires réelles, X et Y des variables aléatoires réelles, telles que
 - (i) $X_n \xrightarrow{\text{loi}} X$ et $Y_n \xrightarrow{\mathbb{P}} Y$,
 - (ii) Y est indépendante de (X_n) et X .

Montrer que le couple (X_n, Y_n) converge en loi vers (X, Y) .

3. En déduire que si (X_n) et (Y_n) sont deux suites de variables aléatoires réelles telles que (X_n) converge en loi vers une limite X et (Y_n) converge en probabilité vers une constante c , alors $(X_n + Y_n)$ converge en loi vers $X + c$ et $(X_n Y_n)$ converge en loi vers cX .

Correction de l'exercice 1.2

1. Soit (δ_n) une suite de v.a. i.i.d. de Bernoulli de moyenne $1/2$ (càd $\mathbb{P}[\delta_n = 0] = \mathbb{P}[\delta_n = 1] = 1/2, \forall n$). D'après le TCL, on sait que

$$X_n := \frac{2}{\sqrt{n}} \sum_{i=1}^n (\delta_i - 1/2) \rightsquigarrow \mathcal{N}(0, 1).$$

On le démontre facilement, en utilisant le Théorème de Levy et en voyant que quand n tend vers l'infini, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E} \exp \left(\frac{2it}{\sqrt{n}} \left(\sum_{i=1}^n (\delta_i - 1/2) \right) \right) &= \left(\left(\frac{1}{2} \right) \left(\exp \left(\frac{-it}{\sqrt{n}} \right) + \exp \left(\frac{it}{\sqrt{n}} \right) \right) \right)^n \\ &= \left(1 - \frac{t^2}{2n} + \mathcal{O} \left(\frac{t^3}{n^{3/2}} \right) \right)^n \rightarrow \exp \left(\frac{-t^2}{2} \right). \end{aligned}$$

Soit g une variable Gaussienne standard. Comme g est symétrique, $-g$ est aussi une Gaussienne Standard. On a donc, (X_n) converge en loi vers g et aussi (X_n) converge en loi vers $-g$. Mais $(X_n + X_n)$ converge en loi vers $2g \neq g + (-g) = 0$. Cet exercice souligne le fait que la convergence en loi est une convergence des lois de distribution et non des variables aléatoires elles mêmes.

2. On note par $\mathcal{C}_b(\mathbb{R})$ l'ensemble des fonctions continues bornées sur \mathbb{R} . Pour montrer que (X_n, Y_n) converge en loi vers (X, Y) , il suffit de prouver que pour tout $f, g \in \mathcal{C}_b(\mathbb{R})$, on a $\mathbb{E}f(X_n)g(Y_n) \rightarrow \mathbb{E}f(X)g(Y)$ quand n tend vers l'infini. Par ailleurs, on sait que si (Y_n) converge en probabilité vers Y et si g est continue alors $(g(Y_n))$ converge en probabilité vers $g(Y)$.

Soit $f, g \in \mathcal{C}_b(\mathbb{R})$ et $\epsilon > 0$. Soit $N_\epsilon \in \mathbb{N}$ tel que pour tout $n \geq N_\epsilon$,

$$\mathbb{P}[|g(Y_n) - g(Y)| \geq \epsilon] \leq \epsilon \text{ and } |\mathbb{E}f(X_n) - \mathbb{E}f(X)| \leq \epsilon.$$

On a pour tout $n \geq N_\epsilon$, par indépendance de $g(Y)$ avec $f(X_n)$ et $f(X)$,

$$\begin{aligned} |\mathbb{E}f(X_n)g(Y_n) - \mathbb{E}f(X)g(Y)| &\leq |\mathbb{E}f(X_n)(g(Y_n) - g(Y))I(|g(Y_n) - g(Y)| \geq \epsilon)| \\ &\quad + |\mathbb{E}f(X_n)(g(Y_n) - g(Y))I(|g(Y_n) - g(Y)| < \epsilon)| + |\mathbb{E}g(Y)(f(X_n) - f(X))| \\ &\leq 2 \|f\|_\infty \|g\|_\infty \mathbb{P}[|g(Y_n) - g(Y)| \geq \epsilon] + \|f\|_\infty \epsilon + |\mathbb{E}g(Y)\mathbb{E}f(X_n) - \mathbb{E}f(X)| \\ &\leq (2 \|f\|_\infty \|g\|_\infty + \|f\|_\infty + \|g\|_\infty) \epsilon. \end{aligned}$$

3. Comme (Y_n) converge en probabilité vers $Y = c$ p.p. qui est indépendante de toutes variables aléatoires, on peut appliquer la question 2. : $((X_n, Y_n))$ converge en probabilité vers (X, c) . Notamment, comme les applications somme et produit sont des fonctions continues de \mathbb{R}^2 dans \mathbb{R} , on voit que $(X_n + Y_n)$ converge en loi vers $X + c$ ainsi que $(X_n Y_n)$ converge en loi vers cX .

Exercice 1.3 (Convergence dans L^p)

Soit (X_n) une suite de variables aléatoires réelles bornées par une même constante. Montrer que si (X_n) converge en probabilité, alors X_n converge dans L^p pour tout $p \geq 1$.

Correction de l'exercice 1.3 Pour cet exercice, on va démontrer un résultat plus fort. On rappelle qu'une suite (X_n) est *équi-intégrable* quand

$$\lim_{a \rightarrow +\infty} \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|I(|X_n| > a)] = 0.$$

Soit $p \geq 1$ et (X_n) une suite d'éléments de L^p . On montre que les deux assertions suivantes sont équivalentes :

1. la suite (X_n) converge dans L^p .
2. la suite (X_n) converge en probabilité et la suite $(|X_n|^p)$ est équi-intégrable.

b) implique a) : On montre d'abord que si (Y_n) est équi-intégrable alors elle est équi-continue : càd pour tout $\epsilon > 0$, il existe $\eta > 0$ tel que si $\mathbb{P}(A) \leq \eta$ alors $\sup_{n \in \mathbb{N}} \mathbb{E}[|Y_n|\mathbb{1}_A] \leq \epsilon$. Soit $\epsilon > 0$ et $a_0 > 0$ tel que pour tout $a \geq a_0$ et tout $n \in \mathbb{N}$, $\mathbb{E}[|X_n|I(|X_n| > a)] \leq \epsilon$. On a pour tout ensemble mesurable A , tout $n \in \mathbb{N}$ et tout $a \geq a_0$,

$$\begin{aligned} \mathbb{E}[|X_n|\mathbb{1}_A] &= \mathbb{E}[|X_n|I(A \cap \{|X_n| \leq a\})] + \mathbb{E}[|X_n|I(A \cap \{|X_n| > a\})] \\ &\leq a\mathbb{P}(A) + \mathbb{E}[|X_n|I(|X_n| > a)] \leq a\mathbb{P}(A) + \epsilon. \end{aligned}$$

On en déduit que (Y_n) est bien équi-continue.

Soit $\epsilon > 0$. Pour tout $q, r \in \mathbb{N}$, on a

$$\begin{aligned} \mathbb{E}|X_r - X_q|^p &\leq \mathbb{E}[|X_r - X_q|^p I(|X_r - X_q|^p \leq \epsilon)] + 2^{p-1} \mathbb{E}[(|X_r|^p + |X_q|^p) I(|X_r - X_q|^p > \epsilon)] \\ &\leq \epsilon + 2^{p-1} \mathbb{E}[(|X_r|^p + |X_q|^p) I(|X_r - X_q|^p > \epsilon)]. \end{aligned}$$

Comme $(|X_n|^p)$ est équi-continue, il existe $\eta > 0$ tel que pour tout A tel que $\mathbb{P}[A] \leq \eta$, on a

$$\sup_{r \in \mathbb{N}} \mathbb{E}[|X_r|^p \mathbb{1}_A] + \sup_{q \in \mathbb{N}} \mathbb{E}[|X_q|^p \mathbb{1}_A] \leq \epsilon/2^{p-1}.$$

Comme (X_n) converge en probabilité, il existe un N_ϵ tel que pour tout $r, q \geq N_\epsilon$, $\mathbb{P}[|X_r - X_q| \geq \epsilon^{1/p}] \leq \eta$. On en déduit, que $\limsup_{r, q} \mathbb{E}|X_r - X_q|^p \leq 2\epsilon$ pour tout $r, q \geq N_\epsilon$. Alors (X_n) est une suite de Cauchy dans L^p , qui est complet, donc elle est convergente dans L^p .

a) implique b) : Par Markov, on a pour tout $\epsilon > 0$,

$$\mathbb{P}[|X_n - X| \geq \epsilon] \leq \epsilon^{-p} \mathbb{E}|X_n - X|^p.$$

Soit $N_\epsilon \in \mathbb{N}$ tel que pour tout $n \geq N_\epsilon$, $\mathbb{E}|X_n - X|^p \leq \epsilon/2^{p-1}$. L'inégalité de Markov donne

$$\mathbb{P}[|X_n|^p > a] \leq a^{-1} \mathbb{E}|X_n|^p \leq B a^{-1} \leq \epsilon.$$

où B majore uniformément la suite $(\mathbb{E}|X_n|^p)$ (qui est bien bornée vu que c'est une suite convergente). Soit $a_0 > 0$ tel que $\sup_{n \in \mathbb{N}} \mathbb{P}[|X_n|^p > a_0] \leq \eta$ où η est tel que $\mathbb{E}[|X|^p \mathbb{1}_A] \leq \epsilon/2^{p-1}$ pour tout A tel que $\mathbb{P}(A) \leq \eta$ (par définition $X \in L^p$). On a donc pour $n \geq N_\epsilon$ et tout $a \geq a_0$,

$$\mathbb{E}[|X_n|^p I(|X_n|^p > a)] \leq 2^{p-1} \mathbb{E}[|X_n - X|^p I(|X_n|^p > a)] + 2^{p-1} \mathbb{E}[|X|^p I(|X_n|^p > a)] \leq \epsilon.$$

De plus, il est facile de voir que toute famille finie de variables aléatoires est équi-intégrable. C'est le cas pour $(X_n : 1 \leq n \leq N_\epsilon)$.

Exercice 1.4 (Lemme de Fatou)

si (f_n) est une suite de fonctions mesurables alors

$$\int \liminf_n f_n \leq \liminf_n \int f_n.$$

En déduire que si (A_n) est une suite d'événements alors

$$\limsup_n \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_n A_n),$$

où on rappelle que $\limsup_n A_n = \bigcap_N \bigcup_{n \geq N} A_n$.

Correction de l'exercice 1.4

1. Pour tout $n \in \mathbb{N}$, on note $g_n = \inf_{p \geq n} f_p$. La suite (g_n) est monotone et converge presque sûrement vers $\liminf_n f_n$. Le théorème de convergence monotone donne :

$$\lim_n \int g_n = \int \lim_n g_n = \int \liminf_n f_n.$$

Par ailleurs, on a pour tout $n \in \mathbb{N}$,

$$\int g_n = \int \inf_{p \geq n} f_p \leq \inf_{p \geq n} \int f_p.$$

Par convergence des deux membres, on peut passer à la limite et obtenir le résultat.

2. On utilise le lemme de Fatou pour $f_n = 1 - \mathbb{1}_{A_n} = \mathbb{1}_{A_n^c}$. On a $\liminf_n f_n = \mathbb{1}_{\liminf_n A_n^c}$ et $(\liminf_n A_n^c)^c = \limsup_n A_n$ donc

$$1 - \mathbb{P}[\limsup_n A_n] = \mathbb{P}[\liminf_n A_n^c] \leq \liminf_n \mathbb{P}[A_n^c].$$

Exercice 1.5 (lemmes de Borel-Cantelli)

1. Le premier lemme de Borel-Cantelli dit que si (A_n) est une suite d'événements telle que $\sum_n \mathbb{P}[A_n] < \infty$ alors $\mathbb{P}[\limsup_n A_n] = 0$.
2. Le deuxième lemme de Borel-Cantelli dit que si (A_n) est une suite d'événements indépendants tels que $\sum_n \mathbb{P}[A_n] = \infty$ alors $\mathbb{P}[\limsup_n A_n] = 1$.

Correction de l'exercice 1.5

1. On note $B_n = \bigcup_{p \geq n} A_p$. On a $\mathbb{P}[B_n] \leq \sum_{p \geq n} \mathbb{P}[A_p]$. Alors par hypothèse, $(\mathbb{P}[B_n])$ tend vers 0 en décroissant. Par convergence monotone, $\lim_n \mathbb{P}[B_n] = \mathbb{P}[\lim_n B_n] = \mathbb{P}[\inf_n B_n] = \mathbb{P}[\liminf_n A_n]$. Donc $\mathbb{P}[\liminf_n A_n] = 0$.

2. Comme $\limsup_n A_n = (\liminf_n A_n^c)^c$, il suffit de montrer que $\mathbb{P}[\liminf_n A_n^c] = 0$. On note $B_n = \bigcap_{p \geq n} A_p$. La suite (B_n) est croissante et converge presque sûrement vers $\liminf_n A_n^c$. Alors, par convergence monotone, $(\mathbb{P}[B_n])$ converge vers $\mathbb{P}[\liminf_n A_n^c]$. Par ailleurs, comme $\log(1-x) \leq -x$ pour $x \in [0, 1)$,

$$\begin{aligned} \mathbb{P}[B_n] &= \mathbb{P}[\bigcap_{p \geq n} A_p^c] = \prod_{p \geq n} \mathbb{P}[A_p^c] = \prod_{p \geq n} (1 - \mathbb{P}[A_p]) \\ &= \exp\left(\sum_{p \geq n} \log(1 - \mathbb{P}[A_p])\right) \leq \exp\left(-\sum_{p \geq n} \mathbb{P}[A_p]\right) = 0. \end{aligned}$$

On en déduit le résultat.

Exercice 1.6 (la loi du 0 – 1 de Kolmogorov)

Soit (σ_n) une suite de tribus indépendantes. La tribu asymptotique est $\sigma_\infty = \bigcap_n \sigma(\bigcup_{p \geq n} \sigma_p)$. La loi du 0 – 1 de Kolmogorov dit que pour tout $A \in \sigma_\infty$, $\mathbb{P}[A] \in \{0, 1\}$.

Correction de l'exercice 1.6 On note $\alpha_n = \sigma(\bigcup_{p \geq n} \sigma_p)$ et $\beta_n = \sigma(\bigcup_{p < n} \sigma_p)$. Les deux tribus α_n et β_n sont indépendantes. Comme $\sigma_\infty \subset \alpha_n$ alors σ_∞ est indépendante de β_n pour tout n . Notamment, σ_∞ est indépendante de $\bigcup_{n \in \mathbb{N}} \beta_n$ et donc de $\sigma(\bigcup_n \beta_n) = \sigma(\bigcup_n \sigma_n) = \alpha_0$. Or $\sigma_\infty \subset \alpha_0$ donc σ_∞ est indépendante d'elle-même. En particulier, si $A \in \sigma_\infty$ alors $\mathbb{P}[A] = \mathbb{P}[A]\mathbb{P}[A]$ donc $\mathbb{P}[A] \in \{0, 1\}$.

Exercice 1.7 (convergence en loi vers une constante)

La convergence en loi vers une constante implique la convergence en proba : On suppose $X_n \rightsquigarrow c$ alors (X_n) converge en probabilité vers c .

Correction de l'exercice 1.7 On peut démontrer que (Y_n) converge en loi vers Y si et seulement si pour tout Borélien A \mathbb{P}^Y -continue (càd $\mathbb{P}[\partial A] = 0$), on a $\mathbb{P}^{Y_n}[A] \rightarrow \mathbb{P}^Y[A]$.

Soit $\epsilon > 0$. On a $\delta_c(B(c, \epsilon)) = 0$. Alors $\mathbb{P}^{X_n}[B(c, \epsilon)] \rightarrow \delta_c(B(c, \epsilon)) = 1$. Donc $\mathbb{P}[|X_n - c| \leq \epsilon] \rightarrow 1$. C'est donc une convergence en probabilité vers c .

Exercice 1.8 (convergence en probabilité et convergence p.s.)

Soit $(X_n)_n$ une suite de variables aléatoires réelles et X une variable aléatoire réelle. L'objectif de cet exercice est de montrer le lien suivant entre convergence en probabilité et convergence presque sûre : il y a équivalence entre :

- a) $(X_n)_n$ converge en probabilité vers X ,
- b) toute sous-suite de $(X_n)_n$ admet une sous-suite qui converge p.s. vers X .

Pour démontrer ce résultat, on va d'abord montrer l'équivalence suivante

- c) $(X_n)_n$ converge en probabilité,
- d) $(X_n)_n$ est une suite de Cauchy en probabilité ; càd $(X_n - X_m)_{n,m}$ converge en probabilité vers 0 quand n et m tendent vers $+\infty$.

Pour démontrer que c) et d) sont équivalents, on procède par étapes :

- 1) Montrer que c) implique d)
- 2) On suppose d).
 - 2.1) En utilisant de lemme de Borel-Cantelli montrer qu'il existe une sous-suite de $(X_n)_n$ qui converge p.s.. On note par X sa limite.
 - 2.2) En déduire que $(X_n)_n$ converge en probabilité vers X .

On montre maintenant l'équivalence entre c) et d).

- 3) On suppose a). En utilisant 2.1) montrer b).
- 4) On suppose b) et on raisonne par contraposé : on suppose que a) n'est pas vrai.
 - 4.1) Ecrire la contraposé.
 - 4.2) Obtenir une contradiction.

Correction de l'exercice 1.8

- 1) On suppose que c) est vrai. Pour tout $\epsilon > 0$ et n, m , on a

$$\mathbb{P}[|X_n - X_m| \geq \epsilon] \leq \mathbb{P}[|X_n - X| \geq \epsilon/2] + \mathbb{P}[|X_m - X| \geq \epsilon/2].$$

Comme le membre de droite tend vers 0 quand n et m tendent vers $+\infty$, on en déduit que le membre de droite tend aussi vers 0 dans ce cas là, càd, d) est vrai.

- 2.1) Comme $(X_n)_n$ est une suite de Cauchy en probability, on peut construire par récurrence en commençant à $n_1 = 1$, une suite strictement croissante d'entiers $(n_j)_j$ telle que

$$\mathbb{P}[|X_{n_j} - X_{n_{j+1}}| > 2^{-j}] < 2^{-j}.$$

Comme $\sum_j \mathbb{P}[|X_{n_j} - X_{n_{j+1}}| > 2^{-j}] < \infty$, le lemmes de Borel-Cantelli dit que $\mathbb{P}[\Omega_0] = 0$ où

$$\Omega_0 = \limsup_{j \rightarrow \infty} \{|X_{n_j} - X_{n_{j+1}}| > 2^{-j}\} = \bigcap_j \bigcup_{k \geq j} \{|X_{n_k} - X_{n_{k+1}}| > 2^{-j}\}.$$

Soit $\omega \in \Omega_0^c$. La suite $(X_{n_j}(\omega))_j$ est une suite (de nombres réels) de Cauchy car pour tout j suffisamment grand et tout $k > j$, on a

$$|X_{n_k}(\omega) - X_{n_j}(\omega)| \leq \sum_{p \geq j} |X_{n_p}(\omega) - X_{n_{p+1}}(\omega)| \leq \sum_{p \geq j} \frac{1}{2^p} = 2^{j-1}.$$

Ainsi, par complétude de \mathbb{R} , on en déduit qu'il existe $X(\omega)$ tel que $(X_{n_j}(\omega))_j$ converge vers $X(\omega)$. Ceci étant vrai pour tout $\omega \in \Omega_0^c$ et $\mathbb{P}[\Omega_0^c] = 1$, on en déduit que $(X_{n_j})_j$ converge presque sûrement.

2.2) On note par X la limite p.s. de $(X_{n_j})_j$. Soit $\epsilon > 0$. On a

$$\mathbb{P}[|X_n - X| \geq \epsilon] \leq \mathbb{P}[|X_n - X_{n_j}| \geq \epsilon/2] + \mathbb{P}[|X_{n_j} - X| \geq \epsilon/2].$$

Comme $(X_{n_j})_j$ converge presque surement vers X , elle converge aussi en probabilité donc $\mathbb{P}[|X_{n_j} - X| \geq \epsilon/2]$ tends vers 0 quand j tends vers $+\infty$. Par ailleurs, $(X_n)_n$ est de Cauchy en probabilité donc quand n et j tendent vers $+\infty$, $\mathbb{P}[|X_n - X_{n_j}| \geq \epsilon/2]$ tends vers 0. En en déduit que $\mathbb{P}[|X_n - X| \geq \epsilon]$ tends vers 0 quand n tends vers $+\infty$, càd $(X_n)_n$ converge vers X en probabilité.

3) On suppose que $(X_n)_n$ converge en probabilité vers X . Soit $(X_{n_j})_j$ une sous suite. Elle converge donc aussi en probabilité vers X . D'après 2.1), c'est aussi une suite de Cauchy en probabilité et donc elle admet une sous-suite qui converge presque surement vers X . Donc *b*) est vrai.

4.1) Dire que *a*) n'est pas vrai, c'est dire qu'il existe un $\epsilon > 0$ et un $\delta > 0$ et une sous-suite $(n_k)_k$ d'entiers tels que pour tout k ,

$$\mathbb{P}[|X_{n_k} - X| \geq \epsilon] \geq \delta. \tag{1}$$

4.2) D'un autre côté, *b*) est vrai donc $(X_{n_k})_k$ admet une sous-suite qui converge presque surement vers X . Ceci contredit (1).

Exercice 1.9 (L'asymptotique normalité implique la convergence en probabilité)

Soit (r_n) une suite de réels positifs tendant vers $+\infty$. Soit (ζ_n) une suite de v.a.r. telle que $r_n(\zeta_n - \mu) \rightsquigarrow \zeta$. Alors (ζ_n) converge en probabilité vers μ .

Correction de l'exercice 1.9 On dit qu'une suite de v.a.r. (ζ_n) est tendue quand pour tout $\epsilon > 0$, il existe $M_\epsilon > 0$ tel que pour tout n , $\mathbb{P}[|\zeta_n| \geq M_\epsilon] \leq \epsilon$. Si une suite converge en probabilité alors elle est tendue. (Car on peut approcher la fonction $I(\cdot \in [-M_\epsilon, M_\epsilon])$ par une suite croissante de fonctions continues bornées). Alors $(r_n(\zeta_n - \mu))$ est tendue. Soit $\epsilon > 0$ et $M_\epsilon > 0$ tels que $\sup_{n \in \mathbb{N}} \mathbb{P}[|\zeta_n - \mu| \geq M_\epsilon/r_n] \leq \epsilon$. Ce qui implique la convergence en probabilité car (r_n) tend vers $+\infty$.

Exercice 1.10 (Loi conditionnelle)

Soit X une variable aléatoire qui suit une loi Gamma $(2, \lambda)$ de densité

$$f(x) = \lambda^2 x e^{-\lambda x} \mathbb{1}_{[0, +\infty)}(x)$$

et soit Y une variable aléatoire dont la loi conditionnelle à $X = x$ est uniforme sur $[0, x]$.

1. Donner la loi jointe de (X, Y) .
2. Donner la loi marginale de Y et montrer que Y est indépendant de $X - Y$.

Correction de l'exercice 1.10

1. Soit f une fonction continue bornée sur \mathbb{R}^2 . On a

$$\begin{aligned} \mathbb{E}f(X, Y) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) d\mathbb{P}^{Y|X=x}(y) \right) d\mathbb{P}^X(x) \\ &= \int_0^\infty \left(\int_0^x f(x, y) \frac{dy}{x} \right) \lambda^2 x e^{-\lambda x} dx = \int_{\mathbb{R}^2} f(x, y) \mathbb{1}_{[0,x]}(y) \mathbb{1}_{\mathbb{R}^+}(x) \lambda^2 e^{-\lambda x} dx dy. \end{aligned}$$

Donc la loi jointe du couple (X, Y) a une densité donnée pour tout $x, y \in \mathbb{R}$ par

$$f^{(X,Y)}(x, y) = \mathbb{1}_{[0,x]}(y) \mathbb{1}_{\mathbb{R}^+}(x) \lambda^2 e^{-\lambda x}$$

2. La loi marginale de Y a pour densité : pour tout $y \in \mathbb{R}$,

$$f^Y(y) = \int_{\mathbb{R}} f^{(X,Y)}(x, y) dx = \mathbb{1}_{y \geq 0} \int_y^\infty \lambda^2 e^{-\lambda x} dx = \lambda e^{-\lambda y} \mathbb{1}_{y \geq 0}.$$

Soit f et g deux fonctions continues bornées. Un changement de variable $x - y \rightarrow t$ donne

$$\begin{aligned} \mathbb{E}f(Y)g(X - Y) &= \int_{\mathbb{R}^2} f(y)g(x - y) \mathbb{1}_{[0,x]}(y) \mathbb{1}_{\mathbb{R}^+}(x) \lambda^2 e^{-\lambda x} dx dy \\ &= \int_{\mathbb{R}} f(y) \mathbb{1}_{y \geq 0} \left(\int_y^\infty g(x - y) \lambda^2 e^{-\lambda x} dx \right) dy = \int_{\mathbb{R}} f(y) \mathbb{1}_{y \geq 0} \left(\int_0^\infty g(t) \lambda^2 e^{-\lambda(t+y)} dt \right) dy \\ &= \left(\int_{\mathbb{R}} f(y) \mathbb{1}_{y \geq 0} \lambda e^{-\lambda y} dy \right) \left(\int_{\mathbb{R}} g(t) \mathbb{1}_{t \geq 0} \lambda e^{-\lambda t} dt \right) = \mathbb{E}f(Y) \mathbb{E}g(X - Y) \end{aligned}$$

(pour avoir la loi de $X - Y$, il suffit de prendre $f \equiv 1$ dans le calcul précédent). Donc Y et $X - Y$ sont bien indépendants.

Exercice 1.11 (quartile)

Soit la loi de probabilité de densité $f(x) = 2x \mathbb{1}\{0 \leq x \leq 1\}$.

1. Trouver les quartiles (y compris la médiane) de cette loi.
2. Considérons un échantillon i.i.d. (X_1, \dots, X_n) de cette loi. Soit \widehat{F}_n la fonction de répartition empirique associée. Donner la loi limite de $\sqrt{n}(\widehat{F}_n(1/2) - 1/4) / \widehat{F}_n(3/4)$ quand $n \rightarrow \infty$, où \widehat{F}_n est la fonction de répartition empirique.

Correction de l'exercice 1.11

1. $q_{1/4} = 1/2$, $q_{1/2} = 1/\sqrt{2}$ et $q_{3/4} = \sqrt{3}/2$
2. Le tCL donne :

$$\sqrt{n}(\widehat{F}_n(1/2) - F(1/2)) \rightsquigarrow \mathcal{N}(0, F(1/2)(1 - F(1/2)))$$

et la LFGN : $\widehat{F}_n(3/4) \xrightarrow{\text{p.s.}} F(3/4)$. Comme $F(1/2) = 1/4$ et $F(3/4) = 9/16$, on obtient

$$\frac{\sqrt{n}(\widehat{F}_n(1/2) - F(1/2))}{\widehat{F}_n(3/4)} \rightsquigarrow \mathcal{N}\left(0, \frac{16}{27}\right)$$

2 Vraisemblance, EMV, IC, Information de Fisher

Exercice 2.1 (Les statistiques d'ordre)

Soient X_1, \dots, X_n des variables aléatoires i.i.d. de fonction de répartition F . On suppose que F admet une densité f par rapport à la mesure de Lebesgue. On note $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ les variables aléatoires X_1, \dots, X_n réordonnées par ordre croissant.

1. Donner l'expression de la loi de la statistique d'ordre $(X_{(1)}, \dots, X_{(n)})$ en fonction de f .
2. Déterminer la fonction de répartition $F_k(x)$ puis la densité $f_k(x)$ de $X_{(k)}$.
3. Sans utiliser les résultats des questions précédentes, calculer les fonctions de répartition de $X_{(1)}, X_{(n)}$, du couple $(X_{(1)}, X_{(n)})$ et la loi de la statistique $W = X_{(n)} - X_{(1)}$ (on appelle W étendue). Les variables $X_{(1)}$ et $X_{(n)}$ sont-elles indépendantes ?

Correction de l'exercice 2.1

1. Comme les X_i ont des densités par rapport à Lebesgues, on a $X_i \neq X_j$ λ -p.p.. Alors p.p.

$$f(X_{(1)}, \dots, X_{(n)}) = \sum_{\sigma \in \mathcal{P}(n)} f(X_{\sigma(1)}, \dots, X_{\sigma(n)}) I(X_{\sigma(1)} < \dots < X_{\sigma(n)}).$$

Soit $\sigma \in \mathcal{P}(n)$. Comme les X_i sont i.i.d., on voit que $(X_{\sigma(1)}, \dots, X_{\sigma(n)})^\top \sim (X_1, \dots, X_n)^\top$. Alors, pour tout $f \in \mathcal{C}_b(\mathbb{R}^n)$,

$$\begin{aligned} \mathbb{E}f(X_{\sigma(1)}, \dots, X_{\sigma(n)}) I(X_{\sigma(1)} < \dots < X_{\sigma(n)}) &= \mathbb{E}f(X_1, \dots, X_n) I(X_1 < \dots < X_n) \\ &= \int_{\mathcal{R}^n} f(x_1, \dots, x_n) \left(\prod_{i=1}^n f(x_i) \right) I(x_1 < \dots < x_n) dx_1 \dots dx_n. \end{aligned}$$

On en déduit que la loi de $(X_{(1)}, \dots, X_{(n)})$ admet une densité par rapport à Lebesgue donnée par

$$f(x_1, \dots, x_n) = n! \left(\prod_{i=1}^n f(x_i) \right) I(x_1 < \dots < x_n).$$

2. On calcul la fonction de répartition de $X_{(k)}$. Soit $t \in \mathcal{R}$,

$$\mathbb{P}[X_{(k)} \leq t] = \mathbb{P}[\exists I \subset \{1, \dots, n\} : |I| \geq k, \forall i \in I, X_i \leq t] = \mathbb{P}[M \geq k]$$

où $M = \sum_{i=1}^n I(X_i \leq t)$ est une multinomiale de paramètre n et $\mathbb{P}[X_1 \leq t] = F(t)$. On a donc

$$\mathbb{P}[X_{(k)} \leq t] = \sum_{j=k}^n \binom{n}{j} F(t)^j (1 - F(t))^{n-j}.$$

Comme F est absolument continue la cdf de $X_{(k)}$ l'est aussi. Donc $X_{(k)}$ admet une densité par rapport à Lebesgues donnée par :

$$\begin{aligned} f(t) &= \sum_{j=k}^n \binom{n}{j} (j f(t) F(t)^{j-1} (1 - F(t))^{n-j} + (n - j) F(t)^j (-f(t)) (1 - F(t))^{n-j-1}) \\ &= \frac{n!}{(k-1)!(n-k)!} F(t)^{k-1} (1 - F(t))^{n-k}. \end{aligned}$$

3. La fonction de répartition de $X_{(1)}$ vérifie :

$$1 - F_{X_{(1)}}(t) = \mathbb{P}[X_{(1)} > t] = \mathbb{P}[X_1 > t, \dots, X_n > t] = \left(\mathbb{P}[X_1 > t]\right)^n = (1 - F(t))^n.$$

La fonction de répartition de $X_{(n)}$ est donnée par :

$$F_{X_{(n)}}(t) = \mathbb{P}[X_{(n)} \leq t] = \mathbb{P}[X_1 \leq t, \dots, X_n \leq t] = \left(\mathbb{P}[X_1 \leq t]\right)^n = (F(t))^n.$$

Pour la fonction de répartition du couple $(X_{(1)}, X_{(n)})$, on calcul la répartition du couple $(X_{(1)}, X_{(n)})$ dans le quadrant inférieur droit. On a pour tout x, y réels :

$$\begin{aligned} \mathbb{P}[X_{(1)} > x, X_{(n)} \leq y] &= \mathbb{P}[x < X_1 \leq y, \dots, x < X_n \leq y] \\ &= \left(\mathbb{P}[x < X_1 \leq y]\right)^n = I(x \leq y)(F(y) - F(x))^n. \end{aligned}$$

On a :

$$\mathbb{P}[X_{(1)} > x, X_{(n)} \leq y] + \mathbb{P}[X_{(1)} \leq x, X_{(n)} \leq y] = \mathbb{P}[X_{(n)} \leq y] = F(y)^n.$$

Alors,

$$F(x, y) = \mathbb{P}[X_{(1)} \leq x, X_{(n)} \leq y] = F(y)^n - I(x \leq y)(F(y) - F(x))^n.$$

La densité de $(X_{(1)}, X_{(n)})$ est donnée par

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}(x, y) = n(n-1)I(x \leq y)f(x)f(y)(F(y) - F(x))^{n-2}.$$

La loi de la statistique $W = X_{(n)} - X_{(1)}$ est donnée par ce qui suit. Soit $f \in \mathcal{C}_b(\mathcal{R})$, on a

$$\begin{aligned} \mathbb{E}f(W) &= \int_{\mathcal{R}^2} f(y-x)d\mathbb{P}^{(X_{(1)}, X_{(n)})}(x, y) \\ &= n(n-1) \int_{\mathcal{R}^2} f(y-x)I(x \leq y)(F(y) - F(x))^{n-2} dx dy \\ &= \int_0^\infty f(u) \left(n(n-1) \int_{\mathcal{R}} (F(u+x) - F(x))^{n-2} dx \right) du. \end{aligned}$$

Alors W a pour densité

$$u \mapsto I(u \geq 0)n(n-1) \int_{\mathcal{R}} (F(u+x) - F(x))^{n-2} dx.$$

Les variables $X_{(1)}$ et $X_{(n)}$ sont indépendantes si et seulement si pour tout x et y , on a

$$\begin{aligned} F(y)^n - I(x \leq y)(F(y) - F(x))^n &= \mathbb{P}[X_{(1)} \leq x, X_{(n)} \leq y] \\ &= \mathbb{P}[X_{(1)} \leq x]\mathbb{P}[X_{(n)} \leq y] = \left(1 - (1 - F(x))^n\right)F(y)^n. \end{aligned}$$

Il faut donc $I(x \leq y)(F(y) - F(x))^n = (F(y) - F(x))^n$ pour tout x, y . Ce qui n'est pas vrai en générale.

Exercice 2.2 (Estimateur de la variance)

Soient X_1, \dots, X_n des variables aléatoires i.i.d., $X_i \sim f(\cdot - \theta)$, où f est une densité de probabilité sur \mathbb{R} symétrique dont on note $\mu_k = \int_{\mathbb{R}} x^k f(x) dx$ les moments d'ordre $k = 2$ et $k = 4$. On note $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Montrer que l'estimateur $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ de la variance des X_i vérifie un théorème central limite.

Indication : on montrera d'abord que l'on peut se ramener au cas où $\theta = 0$, puis on exprimera l'estimateur comme une transformation de $S_n = \frac{1}{n} \sum_{i=1}^n X_i^2$ et de \bar{X}_n .

Correction de l'exercice 2.2 On commence par quelques remarques préliminaires :

- a) Comme $n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est invariant par translation des X_i et que si $X \sim f(\cdot - \theta)$ et $Y \sim f(\cdot)$ alors $X \sim Y + \theta$, on peut donc supposer que $\theta = 0$. Notamment comme f est symétrique, on a $\mathbb{E}X_i = 0, \forall i$.
- b) On note $\hat{\sigma}_n^2 := n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. On a :

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 = \bar{X}_n^2 - \bar{X}_n.$$

(On écrit $\hat{\sigma}_n^2 = \mathbb{E}_I (X_I - \mathbb{E}_I X_I)^2$.)

- c) On remarque d'abord que $\hat{\sigma}_n^2$ n'est pas un estimateur sans biais de la variance :

$$\begin{aligned} \mathbb{E}\hat{\sigma}_n^2 &= \mathbb{E}X^2 - \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2 = \mathbb{E}X^2 - \frac{1}{n^2} \sum_{i,j} \mathbb{E}X_i X_j \\ &= \left(1 - \frac{1}{n}\right) (\mathbb{E}X^2 - (\mathbb{E}X)^2) = \frac{n-1}{n} \text{var}(X). \end{aligned}$$

Par la LFGN, la suite $(\hat{\sigma}_n^2)$ converge presque sûrement vers σ^2 .

On considère la décomposition suivante :

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) = \sqrt{n}(\bar{X}_n^2 - \mathbb{E}X^2) - \sqrt{n}(\bar{X}_n)^2.$$

Par le TCL, on a :

$$\sqrt{n}(\bar{X}_n^2 - \mathbb{E}X^2) \rightsquigarrow \mathcal{N}(0, \mathbb{E}(X^2 - \mathbb{E}X^2)^2) \text{ et } \mathbb{E}(X^2 - \mathbb{E}X^2)^2 = \mu_4 - \mu_2^2.$$

Par ailleurs, $(\sqrt{n}\bar{X}_n)$ converge en loi vers une Gaussienne et (\bar{X}_n) converge en probabilité vers 0. Alors d'après Slutsky, $(\sqrt{n}(\bar{X}_n)^2)$ converge en loi vers 0, elle converge donc aussi en probabilité vers 0. On applique une seconde fois Slutsky : $(\sqrt{n}(\bar{X}_n^2 - \mathbb{E}X^2))$ converge en loi vers $\mathcal{N}(0, \mathbb{E}(X^2 - \mathbb{E}X^2)^2)$ et $(\sqrt{n}(\bar{X}_n)^2)$ converge en probabilité vers 0. On en déduit que

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \rightsquigarrow \mathcal{N}(0, \mathbb{E}(X^2 - \mathbb{E}X^2)^2).$$

Exercice 2.3 (Stabilisation de la variance)

On dispose d'un échantillon X_1, \dots, X_n i.i.d. de loi de Bernoulli de paramètre $0 < \theta < 1$.

1. On note \bar{X}_n la moyenne empirique des X_i . Appliquer la loi forte des grands nombres et le TCL dans ce modèle.
2. Cherchez une fonction g telle que $\sqrt{n}(g(\bar{X}_n) - g(\theta))$ converge en loi vers Z de loi $\mathcal{N}(0, 1)$.
3. On note z_α le quantile d'ordre $1 - \alpha/2$ de la loi normale standard. En déduire un intervalle de confiance $\hat{I}_{n,\alpha}$ fonction de z_α, n, \bar{X}_n tel que $\lim_{n \rightarrow \infty} \mathbb{P}(\theta \in \hat{I}_{n,\alpha}) = 1 - \alpha$.

Correction de l'exercice 2.3

1. La LFGN dit que (\bar{X}_n) converge presque sûrement vers $\mathbb{E}X_1 = \theta$. Le TCL dit que $((\sqrt{n}/\sigma)(\bar{X}_n - \mathbb{E}X))$ converge en loi vers une Gaussienne centrée réduite où $\sigma = \sqrt{\theta(1 - \theta)}$.
2. D'après le TCL :

$$\sqrt{n}(\bar{X}_n - \theta) \rightsquigarrow \sigma g.$$

On dit que \bar{X}_n est asymptotiquement normale de moyenne θ et de variance asymptotique σ^2 . On peut alors appliquer la Proposition 1.10 (Méthode delta) du cours (en fait, on applique une version plus faible de ce résultat qu'on peut trouver page 26 au théorème 3.1 de [van der Vaart, asymptotic Statistics]) : si (ζ_n) est asymptotiquement normale de moyenne asymptotique θ et de variance asymptotique σ^2 et si $g : \mathcal{D} \subset \mathbb{R} \mapsto \mathbb{R}$ est une fonction différentiable en θ , alors $(g(\zeta_n))$ est aussi asymptotiquement normale et on a :

$$\sqrt{n}(g(\zeta_n) - g(\theta)) \rightsquigarrow \mathcal{N}(0, \sigma^2(g'(\theta))^2). \tag{2}$$

Dans notre cas, on cherche à trouver g tel que $(g(\bar{X}_n))$ est asymptotiquement normal de moyenne asymptotique θ et de variance asymptotique $\theta(1 - \theta)(g'(\theta))^2 = 1$. On est donc amené à résoudre l'équation :

$$\forall \theta \in (0, 1), \quad g'(\theta) = \frac{1}{\sqrt{\theta(1 - \theta)}}.$$

L'ensemble des solutions de cette équation est donnée, à une constante absolue additive près, par $g : \theta \in [0, 1] \mapsto 2\arcsin(\sqrt{\theta})$ (on rappelle que $(\arcsin x)' = (1 - x^2)^{-1/2}, \forall x \in [-1, 1]$). Cette fonction est continûment différentiable en tout $\theta \in (0, 1)$, alors d'après Proposition 1.10 (voir (2)), on a

$$\sqrt{n}(g(\bar{X}_n) - g(\theta)) \rightsquigarrow \mathcal{N}(0, 1).$$

(On rappelle que g a été choisit tel que $\theta(1 - \theta)(g'(\theta))^2 = 1$ pour tout $\theta \in (0, 1)$).

3. Pour tout $\alpha \in]0, 2[$, le quantile d'ordre $1 - \alpha/2$ de la gaussienne est l'unique réel tel que $\mathbb{P}[g \in (-\infty, q_\alpha] = 1 - \alpha/2$. On a

$$\mathbb{P}[\theta \in \hat{I}_{n,\alpha}] = \mathbb{P}\left[\left|\sqrt{n}(g(\bar{X}_n) - g(\theta))\right| \leq z_\alpha\right] \longrightarrow \mathbb{P}[g \in [-z_\alpha, z_\alpha]] = 1 - \alpha$$

pour

$$\hat{I}_{n,\alpha} = \left[\sin^2\left(g(\bar{X}_n) - \frac{z_\alpha}{\sqrt{n}}\right), \sin^2\left(g(\bar{X}_n) + \frac{z_\alpha}{\sqrt{n}}\right) \right]$$

Exercice 2.4 (Modèle probit)

Nous disposons d'une information relative au comportement de remboursement ou de non-remboursement d'emprunteurs :

$$Y_i = \begin{cases} 1 & \text{si l'emprunteur } i \text{ rembourse,} \\ 0 & \text{si l'emprunteur } i \text{ est défaillant.} \end{cases}$$

Afin de modéliser ce phénomène, on suppose l'existence d'une variable aléatoire Y_i^* normale, d'espérance m et de variance σ^2 , que l'on appellera « capacité de remboursement de l'individu i », telle que :

$$Y_i = \begin{cases} 1 & \text{si } Y_i^* > 0, \\ 0 & \text{si } Y_i^* \leq 0. \end{cases}$$

On note Φ la fonction de répartition de la loi normale $\mathcal{N}(0, 1)$.

1. Exprimer la loi de Y_i en fonction de Φ .
2. Les paramètres m et σ^2 sont-ils identifiables ?

Correction de l'exercice 2.4

1. On calcul la loi de Y tel que $Y = 1$ quand $Y^* \geq 0$ et $Y = 0$ quand $Y^* < 0$ où $Y^* \sim \mathcal{N}(m, \sigma^2)$. La loi de Y est donnée par $\mathbb{P}[Y^* \geq 0]\delta_1 + \mathbb{P}[Y^* < 0]\delta_0$. On note par φ la densité d'une gaussienne $\mathcal{N}(0, 1)$, en particulier, on a $\Phi(x) = \int_{-\infty}^x \varphi(t)dt$. Le changement de variable $(x - m)/\sigma \rightarrow t$ donne

$$\mathbb{P}[Y^* < 0] = \int_{-\infty}^0 \varphi\left(\frac{x - m}{\sigma}\right) \frac{dx}{\sigma} = \int_{-\infty}^{-m/\sigma} \varphi(t)dt = \Phi\left(\frac{-m}{\sigma}\right).$$

La loi de Y est donc $(1 - \Phi(-m/\sigma))\delta_1 + \Phi(-m/\sigma)\delta_0$.

2. Les paramètres m et σ^2 ne sont pas identifiable vu que n'importe quels couples (m_1, σ_1^2) et (m_2, σ_2^2) tels que $m_1/\sigma_1^2 = m_2/\sigma_2^2$ donne la même loi pour Y .

Exercice 2.5 (Répartition de génotypes dans une population)

Quand les fréquences de gènes sont en équilibre, les génotypes AA, Aa et aa se manifestent dans une population avec probabilités $(1 - \theta)^2$, $2\theta(1 - \theta)$ et θ^2 respectivement, où θ est un paramètre inconnu. Plato *et al.* (1964) ont publié les données suivantes sur le type de haptoglobine dans un échantillon de 190 personnes :

Type de haptoglobine	Hp-AA	Hp-Aa	Hp-aa
effectifs	10	68	112

1. Comment interpréter le paramètre θ ? Proposez un modèle statistique pour ce problème.
2. Calculez l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ .

3. Donnez la loi asymptotique de $\sqrt{n}(\hat{\theta}_n - \theta)$.

4. Proposez un intervalle de confiance de niveau asymptotique 95% pour θ .

Correction de l'exercice 2.5

1. On propose deux modélisations pour ces données. Seule la deuxième sera utilisée pour le traitement mathématique du problème.

Modèle 1 : On modélise ce problème par une famille de n couples $(\delta_1^{(1)}, \delta_1^{(2)}), \dots, (\delta_n^{(1)}, \delta_n^{(2)})$ où les $\delta_i^{(j)}, i = 1, \dots, n, j = 1, 2$ sont i.i.d. Bernoulli sur $\{A, a\}$ de paramètre θ . On dit que $\delta_i^{(j)} = a$ quand l'allèle a est présent chez l'individu i au gène numéro 2. On a donc bien les probabilités du génotype AA qui est $(1 - \theta)^2$, Aa qui est de probabilité $2\theta(1 - \theta)$ et aa qui est θ^2 . Dans ce modèle θ est la probabilité d'avoir l'allèle a pour chacun des deux gènes.

Modèle 2 : On peut modéliser ce problème par une famille de n variables aléatoires X_1, \dots, X_n i.i.d. à valeurs dans $\{AA, Aa, aa\}$ telles que $\mathbb{P}[X = AA] = (1 - \theta)^2$, $\mathbb{P}[X = Aa] = 2\theta(1 - \theta)$ et $\mathbb{P}[X = aa] = \theta^2$. On choisit ce modèle pour la suite. On peut voir que $X = \{\delta^{(1)}, \delta^{(2)}\}$. Donc θ s'interprète comme étant la probabilité d'avoir l'allèle a pour chacun des deux gènes.

2. Dans le modèle 2, la loi de X est $\mathbb{P}_\theta = (1 - \theta)^2\delta_{AA} + 2\theta(1 - \theta)\delta_{Aa} + \theta^2\delta_{aa}$, elle admet une densité f_θ par rapport à la mesure $\delta_{AA} + \delta_{Aa} + \delta_{aa}$ qui est définie sur $\{AA, Aa, aa\}$ donnée par $f_\theta(AA) = (1 - \theta)^2$, $f_\theta(Aa) = 2\theta(1 - \theta)$ et $f_\theta(aa) = \theta^2$. La Log-vraisemblance est donnée par

$$L : \theta \in (0, 1) \mapsto \sum_{i=1}^n \log f_\theta(X_i) \\ = N_n(AA) \log[(1 - \theta)^2] + N_n(Aa) \log[2\theta(1 - \theta)] + N_n(aa) \log[\theta^2]$$

où $N_n(\square)$ est le nombre de génotypes \square dans l'échantillon $\{X_1, \dots, X_n\}$. On a pour tout $\theta \in (0, 1)$,

$$L'(\theta) = \frac{2n}{\theta} - \frac{1}{\theta(1 - \theta)} [2N_n(AA) + N_n(Aa)].$$

Alors l'estimateur du maximum de vraisemblance est donné par

$$\hat{\theta}_n = 1 - \frac{1}{2n} [2N_n(AA) + N_n(Aa)].$$

Ici, on a $\hat{\theta}_n = 1 - 22/95 \approx 0.77$.

3. On peut appliquer le TCL ou la méthode générale du cours sur la normalité asymptotique des EMV. Pour le TCL, on a directement que

$$\sqrt{n}(\theta - \hat{\theta}_n) = \sqrt{n}\left(\frac{1}{n} \sum_{i=1}^n (I(X_i = AA) + (1/2)I(X_i = Aa)) - (1 - \theta)\right) \rightsquigarrow \mathcal{N}\left(0, \frac{\theta - \theta^2}{2}\right)$$

car

$$\mathbb{E}(I(X = AA) + (1/2)I(Z = Aa)) = (1 - \theta)^2 + \theta(1 - \theta) = 1 - \theta$$

et

$$\mathbb{E}(I(X = AA) + (1/2)I(Z = Aa))^2 = 1 - \frac{3\theta}{2} + \frac{\theta^2}{2}$$

alors

$$\text{var}(I(X = Aa) + (1/2)I(Z = Aa)) = \frac{\theta - \theta^2}{2}.$$

4. On applique la méthode Delta. On cherche une fonction g telle que pour tout $\theta \in (0, 1)$, on a :

$$g'(\theta)^2 \frac{\theta - \theta^2}{2} = 1$$

alors $g(\theta) = 2\sqrt{2}\arcsin(\sqrt{\theta})$. On applique la méthode Delta : $(\sqrt{n}(g(\hat{\theta}_n) - g(\theta)))$ converge en loi vers $\mathcal{N}(0, 1)$. Alors si $\mathbb{P}[|G| \leq z_\alpha] = 1 - \alpha$, où G est Gaussienne Standard, on aura, quand n tend vers ∞ ,

$$\mathbb{P}\left[\hat{\theta}_n \in g^{-1}\left(\left[g(\theta) - z_\alpha/\sqrt{n}, g(\theta) + z_\alpha/\sqrt{n}\right]\right)\right] \rightarrow 1 - \alpha.$$

Exercice 2.6 (Modèle d'autorégression)

On considère les observations X_1, \dots, X_n , où les X_i sont issus du modèle d'autorégression d'ordre 1 :

$$X_i = \theta X_{i-1} + \xi_i, \quad i = 1, \dots, n, \quad X_0 = 0, \tag{3}$$

où ξ_i i.i.d. de loi normale $\mathcal{N}(0, \sigma^2)$ et $\theta \in \mathbb{R}$.

1. Explicitez l'expérience statistique associée à la donnée (X_1, \dots, X_n) .
2. Calculez l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ pour ce modèle.

Correction de l'exercice 2.6

1. Une expérience statistique est un triplet de la forme :

$$\mathcal{E} = (\mathfrak{Z}, \mathcal{Z}, \{\mathbb{P}_\theta : \theta \in \Theta\})$$

où \mathfrak{Z} est l'espace des observations, \mathcal{Z} est la tribu sur l'espace des observations et $\{\mathbb{P}_\theta : \theta \in \Theta\}$ est le modèle : c'est l'ensemble des mesures de probabilités dont on suppose a priori que les données sont issues.

Ici, on a $\mathfrak{Z} = \mathbb{R}^n$ qui est muni de sa tribu des Boréliens \mathcal{Z} . Le modèle est donné par l'équation d'autorégression : $X_i = \theta X_{i-1} + \zeta_i$ où ζ_i sont i.i.d. $\mathcal{N}(0, \sigma^2)$. Pour le modèle, on suppose connu σ^2 . Ainsi le modèle est seulement paramétré par θ (sinon, il serait paramétré par (θ, σ^2)). La loi \mathbb{P}_θ est donc la loi de (X_1, X_2, \dots, X_n) sous l'hypothèse "AR(1)" de l'équation 3. On a $\mathbb{P}_\theta^{X_i|X_{i-1}, \dots, X_1} = \mathbb{P}_\theta^{X_i|X_{i-1}} \sim \mathcal{N}(\theta X_{i-1}, \sigma^2)$. On montre par récurrence que

$$\mathbb{P}_\theta = \mathbb{P}_\theta^{(X_1, \dots, X_n)} = f_\theta \cdot \lambda$$

où λ est la mesure de Lebesgues sur \mathbb{R}^n et f_θ est une fonction de densité définie sur \mathbb{R}^n pour tout $(x_1, \dots, x_n) \in \mathbb{R}^n$ par :

$$f_\theta(x_1, \dots, x_n) = f(x_1)f(x_2 - \theta x_1)f(x_3 - \theta x_2) \cdots f(x_n - \theta x_{n-1})$$

où f est la densité d'une Gaussienne de moyenne nulle et de variance σ^2 . Pour la récurrence, on utilise l'identité $\mathbb{P}^{(X,Y)} = \mathbb{P}^X \otimes \mathbb{P}^{Y|X}$.

2. La fonction de Log-vraisemblance est donnée par :

$$L : \begin{cases} \mathbb{R} & \rightarrow & \mathbb{R} \\ \theta & \mapsto & \log f_{\theta}(X_1, \dots, X_n) = \log f(X_1) + \sum_{i=1}^{n-1} \log f(X_{i+1} - \theta X_i). \end{cases}$$

où $f(x) = (\sigma\sqrt{2\pi})^{-1/2} \exp(-x^2/(2\sigma^2))$. Alors pour tout θ ,

$$L(\theta) = \frac{-n \log(\sigma\sqrt{2\pi})}{2} - \frac{X_1^2}{2\sigma^2} + \sum_{i=1}^{n-1} \frac{(X_{i+1} - \theta X_i)^2}{2\sigma^2}$$

et aussi

$$L'(\theta) = \sum_{i=1}^{n-1} \frac{-X_i(X_{i+1} - \theta X_i)}{\sigma^2} = \sigma^{-2} \left(\theta \sum_{i=1}^{n-1} X_i^2 - \sum_{i=1}^{n-1} X_i X_{i+1} \right).$$

Alors l'EMV est donné par :

$$\hat{\theta}_n = \left(\sum_{i=1}^{n-1} X_i X_{i+1} \right) / \left(\sum_{i=1}^{n-1} X_i^2 \right).$$

Exercice 2.7 (Durées de connection)

On peut modéliser la durée d'une connection sur le site www.Cpascher.com par une loi gamma(2, 1/θ) de densité

$$\theta^{-2} x e^{-x/\theta} 1_{[0, +\infty[}(x).$$

Pour fixer vos tarifs publicitaires, vous voulez estimer le paramètre θ à partir d'un échantillon X_1, \dots, X_n de n durées de connexion. On vous donne $\mathbb{E}_{\theta}(X_i) = 2\theta$ et $\text{var}_{\theta}(X_i) = 2\theta^2$.

1. Calculez l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ .
2. Que vaut $\mathbb{E}(\hat{\theta}_n)$? Quelle est la variance de $\hat{\theta}_n$?

Correction de l'exercice 2.7

1. On note par f_{θ} la densité donnée pour tout $x \in \mathbb{R}$, par $f_{\theta}(x) = \theta^{-2} x e^{-x/\theta} I(x \geq 0)$. La log-vraisemblance du modèle est la fonction $L : \theta \in \mathbb{R}_+^* \mapsto \sum_{i=1}^n \log f_{\theta}(X_i)$. On a pour tout $\theta > 0$,

$$L(\theta) = -2n \log \theta + \sum_{i=1}^n \log X_i - \frac{n}{\theta} \bar{X}_n,$$

où $\bar{X}_n = n^{-1} \sum_i X_i$. Alors $L'(\theta) = -2n\theta^{-1} + n\theta^{-2} \bar{X}_n$ et donc $\hat{\theta}_n \in \text{argmax}_{\theta > 0} L(\theta) = \{(1/2)\bar{X}_n\}$.

2. $\mathbb{E}\hat{\theta}_n = \theta$. Pour la variance, on a

$$\text{var}(\hat{\theta}_n) = \frac{1}{4n} \text{var}(X_1) = \frac{\mathbb{E}X^2 - (\mathbb{E}X)^2}{4n} = \frac{2\theta^2}{4n}.$$

Exercice 2.8 (Durée de vie)

Un système fonctionne en utilisant deux machines de types différents. Les durées de vie X_1 et X_2 des deux machines suivent des lois exponentielles de paramètres λ_1 et λ_2 . Les variables aléatoires X_1 et X_2 sont supposées indépendantes.

1. Montrer que

$$X \stackrel{\text{Loi}}{=} \mathcal{E}(\lambda) \Leftrightarrow \forall x > 0, \mathbb{P}(X > x) = \exp(-\lambda x).$$

2. Calculer la probabilité pour que le système ne tombe pas en panne avant la date t . En déduire la loi de la durée de vie Z du système. Calculer la probabilité pour que la panne du système soit due à une défaillance de la machine 1.

3. Soit $I = 1$ si la panne du système est due à une défaillance de la machine 1, $I = 0$ sinon. Calculer $\mathbb{P}(Z > t; I = \delta)$, pour tout $t \geq 0$ et $\delta \in \{0, 1\}$. En déduire que Z et I sont indépendantes.

4. On dispose de n systèmes identiques et fonctionnant indépendamment les uns des autres dont on observe les durées de vie Z_1, \dots, Z_n .

(a) Écrire le modèle statistique correspondant. A-t-on suffisamment d'information pour estimer λ_1 et λ_2 ?

(b) Si on observe à la fois les durées de vie des systèmes et la cause de la défaillance (machine 1 ou 2), a-t-on alors suffisamment d'information pour estimer λ_1 et λ_2 ?

5. On considère maintenant un seul système utilisant une machine de type 1 et une machine de type 2, mais on suppose que l'on dispose d'un stock de n_1 machines de type 1, de durées de vie $X_1^1, \dots, X_1^{n_1}$ et d'un stock de n_2 machines de type 2, de durées de vie $X_2^1, \dots, X_2^{n_2}$. Quand une machine tombe en panne, on la remplace par une machine du même type, tant que le stock de machines de ce type n'est pas épuisé. Quand cela arrive, on dit que le système lui-même est en panne. On note toujours Z la durée de vie du système. Le cas $n_1 = n_2 = 1$ correspond donc aux trois premières questions.

(a) Montrer que la densité de la somme U de k variables indépendantes qui suivent une loi exponentielle de même paramètre λ s'écrit, pour $x \geq 0$:

$$f_U(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} \exp(-\lambda x).$$

(b) Écrire Z en fonction des X_i^j et en déduire $\mathbb{P}(Z \geq t)$ en fonction $n_1, n_2, \lambda_1, \lambda_2$ et t .

Correction de l'exercice 2.8

1. Par définition, une v.a.r. suit une loi exponentielle $\mathcal{E}(\lambda)$ quand elle admet une densité de la forme $f_\lambda : x \in \mathbb{R} \mapsto \lambda \exp(-\lambda x) I(x > 0)$. Si $X \sim \mathcal{E}(\lambda)$ alors, pour tout $x > 0$,

$$\mathbb{P}[X > x] = \int_x^\infty \lambda \exp(-\lambda x) dx = \exp(-\lambda x).$$

Réciproquement, si X est une v.a.r. telle que pour tout $x > 0$, $1 - F_X(x) = \mathbb{P}[X > x] = \exp(-\lambda x)$. Alors X est portée sur \mathbb{R}^+ et comme F_X est dérivable, X admet une densité donnée par F'_X càd $x \mapsto \lambda \exp(-\lambda x)I(x > 0)$. C'est donc une variable exponentielle.

2. On note par Z la durée de vie du système. On a donc $Z > t$ ssi $X_1 > t$ et $X_2 > t$ et donc par indépendance

$$\mathbb{P}[Z > t] = \mathbb{P}[\{X_1 > t\} \cap \{X_2 > t\}] = \mathbb{P}[X_1 > t]\mathbb{P}[X_2 > t] = \exp(-(\lambda_1 + \lambda_2)t).$$

Donc $Z \sim \mathcal{E}(\lambda_1 + \lambda_2)$. Par ailleurs, la machine sera en panne due à l'élément 1 quand $X_1 < X_2$. On calcul $\mathbb{P}[X_1 < X_2]$:

$$\begin{aligned} \mathbb{P}[X_1 < X_2] &= \mathbb{E}I(X_1 < X_2) = \int_{\mathbb{R}_+^2} I(x_1 < x_2) f_{\lambda_1}(x_1) f_{\lambda_2}(x_2) dx_1 dx_2 \\ &= \int_0^\infty f_{\lambda_1}(x_1) \left(\int_{x_1}^\infty f_{\lambda_2}(x_2) dx_2 \right) dx_1 = \int_0^\infty f_{\lambda_1}(x_1) \exp(-\lambda_2 x_1) dx_1 \\ &= \int_0^\infty \lambda_1 \exp(-(\lambda_1 + \lambda_2)x_1) dx_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2}. \end{aligned}$$

3. [$I = 1$ ssi $X_1 < X_2$] et [$I = 0$ ssi $X_1 > X_2$]. On a

$$\begin{aligned} \mathbb{P}[\{Z > t\} \cap \{I = 1\}] &= \mathbb{P}[\{X_1 \wedge X_2 > t\} \cap \{X_1 < X_2\}] = \mathbb{P}[t < X_1 < X_2] \\ &= \int_t^\infty \lambda_1 \exp(-(\lambda_1 + \lambda_2)x_1) dx_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2} \exp(-(\lambda_1 + \lambda_2)t) = \mathbb{P}[Z > t]\mathbb{P}[I = 1] \end{aligned}$$

Par symétrie,

$$\begin{aligned} \mathbb{P}[Z > t | I = 0] &= \mathbb{P}[X_1 \wedge X_2 > t | X_1 > X_2] = \mathbb{P}[X_1 > X_2 > t] \\ &= \frac{\lambda_2}{\lambda_2 + \lambda_1} \exp(-(\lambda_1 + \lambda_2)t) = \mathbb{P}[I = 0]\mathbb{P}[Z > t]. \end{aligned}$$

On en déduit que Z et I sont indépendantes.

4. a) Le modèle statistique est $\{\mathcal{E}(\lambda_1 + \lambda_2)^{\otimes n} : \lambda_1, \lambda_2 > 0\}$. Ce modèle n'est pas identifiable en le paramètre (λ_1, λ_2) .
4. b) On observe $(X_{1i} \wedge X_{2i}, I_i)$ ou $I_i = 1$ si $X_{1i} < X_{2i}$ et $I_i = 0$ sinon. On peut estimer la moyenne de Z par $n^{-1} \sum_{i=1}^n Z_i$ et on peut estimer la moyenne de I par $n^{-1} \sum_{i=1}^n I_i$. On peut donc estimer $\lambda_1 + \lambda_2$ et $\lambda_1/(\lambda_1 + \lambda_2)$. On peut donc estimer λ_1 et λ_2 .

Exercice 2.9 (Taux de défaillance)

Une chaîne de production doit garantir une qualité minimale de ses produits. En particulier, elle doit garantir que la proportion θ des produits défectueux reste inférieure à un taux fixé par le client. Un échantillon de n produits est prélevé et analysé. On note $\hat{\theta}_n$ la proportion de produits défectueux dans l'échantillon.

1. Proposer un modèle statistique pour ce problème. Quelle est la loi de $n\hat{\theta}_n$?

2. Quelle information donne la loi des grand nombres et le théorème centrale limite sur le comportement asymptotique de $\hat{\theta}_n$?
3. On donne $\mathbb{P}(N > 1.64) = 5\%$ pour $N \sim \mathcal{N}(0, 1)$. En déduire ϵ_n (dépendant de n et θ) tel que $\mathbb{P}(\theta \geq \hat{\theta}_n + \epsilon_n) \xrightarrow{n \rightarrow \infty} 5\%$.
4. La valeur ϵ_n précédente dépend de θ . A l'aide du lemme de Slutsky, donner ϵ'_n ne dépendant que de n et $\hat{\theta}_n$ tel que $\mathbb{P}(\theta \geq \hat{\theta}_n + \epsilon'_n) \xrightarrow{n \rightarrow \infty} 5\%$.

Correction de l'exercice 2.9

1. On modélise ce problème par une famille de n variables de Bernoulli $\delta_1, \dots, \delta_n$ i.i.d. telle que $\mathbb{P}[\delta_i = 1] = \theta = 1 - \mathbb{P}[\delta_i = 0]$. Où $\delta_i = 1$ signifie que le i -ième produit prélevé est défectueux et $\delta_i = 0$ signifie qu'il n'est pas défectueux. On a donc $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n \delta_i$. En particulier, $n\hat{\theta} = \sum_{i=1}^n \delta_i$ donc pour tout $0 \leq k \leq n$,

$$\mathbb{P}[n\hat{\theta}_n = k] = \binom{n}{k} \theta^k (1 - \theta)^{n-k}.$$

On reconnaît la loi d'une multinomiale de paramètre n, θ .

2. La loi des grands nombres assure que $(\hat{\theta}_n)$ converge presque sûrement vers $\mathbb{E}\delta = \theta$. Comme $\text{var}(\delta) = \mathbb{E}\delta^2 - (\mathbb{E}\delta)^2 = \theta - \theta^2$, le TCL dit que $\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, (\theta - \theta^2))$.
3. Le TCL dit que, quand n tend vers l'infini,

$$\mathbb{P}\left[\theta \geq \hat{\theta}_n + \epsilon \sqrt{\frac{\theta - \theta^2}{n}}\right] = \mathbb{P}\left[\sqrt{\frac{n}{\theta - \theta^2}}(\theta - \hat{\theta}_n) \geq \epsilon\right] \longrightarrow \mathbb{P}[g \geq \epsilon].$$

Si on choisit $\epsilon > 0$ tel que $\mathbb{P}[g \geq \epsilon] = 5\%$, on obtient le résultat pour $\epsilon_n = \sqrt{(\theta - \theta^2)/n} \epsilon$.

4. La fonction $x \mapsto \sqrt{1/(x - x^2)}$ est continue sur $(0, 1)$ alors si $\theta \in (0, 1)$, comme $\hat{\theta}_n$ converge presque sûrement vers θ , il existe un $N \in \mathbb{N}$ tel que pour tout $n \geq N$, $\hat{\theta}_n \in (0, 1)$ p.s. et donc $(f(\hat{\theta}_n))_{n \geq N}$ est p.s. définie et elle converge vers $f(\theta)$ presque sûrement. Comme $\sqrt{n}(\theta - \hat{\theta}_n)$ converge en loi vers $\mathcal{N}(0, \theta - \theta^2)$ et $(f(\hat{\theta}_n))_{n \geq N}$ converge presque sûrement vers $f(\theta)$, on en déduit par le lemme de Slutsky que $(f(\hat{\theta}_n)\sqrt{n}(\theta - \hat{\theta}_n))_{n \geq N}$ converge en loi vers $f(\theta)\mathcal{N}(0, \theta - \theta^2) = \mathcal{N}(0, 1)$. On définit la suite de v.a. (ϵ'_n) par

$$\epsilon'_n = \sqrt{\frac{\hat{\theta}_n - \hat{\theta}_n^2}{n}} \epsilon.$$

On a alors :

$$\mathbb{P}\left[\theta \geq \hat{\theta}_n + \epsilon'_n\right] = \mathbb{P}\left[f(\hat{\theta}_n)\sqrt{n}(\theta - \hat{\theta}_n) \geq \epsilon\right] \longrightarrow \mathbb{P}[g \geq \epsilon] = 5\%.$$

Exercice 2.10 (Cas des défaillances rares)

La chaîne produit des composants électroniques utilisés dans le secteur aéronautique. Le taux de défaillance doit donc être très bas. En particulier, comme la taille de l'échantillon ne peut être très grosse (question de coût), il est attendu que θ soit du même ordre

de grandeur que $1/n$. On supposera donc par la suite que la proportion de composants défectueux est $\theta_n = \lambda/n$ pour un certain $\lambda > 0$ et on cherche à estimer λ par $\hat{\lambda}_n = n\hat{\theta}_n$. La valeur λ est supposée indépendante de n (le cas intéressant est quand λ est petit).

1. Quelle est la limite de $\mathbb{P}(\hat{\lambda}_n = k)$ lorsque $n \rightarrow +\infty$? En déduire que $\hat{\lambda}_n$ converge en loi vers une variable de Poisson de paramètre λ .
2. On suppose qu'il y a une proportion $\theta_n = 3/n$ de composants défectueux. Sachant que $\mathbb{P}(Z = 0) \approx 5\%$ pour Z de loi de Poisson de paramètre 3, montrer que $\mathbb{P}(\theta_n > \hat{\theta}_n + 2/n) \approx 5\%$ pour n grand.

Correction de l'exercice 2.10

1. On rappelle qu'une variable de Poisson Z de paramètre λ est portée par \mathbb{N} telle que pour tout $k \in \mathbb{N}$, $\mathbb{P}[Z = k] = (\lambda^k/k!)e^{-\lambda}$. On note par $\delta_1, \dots, \delta_n$ des Bernoulli de paramètre $\theta = \lambda/n$. Pour tout $0 \leq k \leq n$, on a

$$\begin{aligned} \mathbb{P}[\hat{\lambda}_n = k] &= \mathbb{P}\left[\sum_{i=1}^n \delta_i = k\right] = \binom{n}{k} \theta^k (1-\theta)^{n-k} = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n!}{k!(n-k)!} \left(1 - \frac{\lambda}{n}\right)^n \left(\frac{n}{\lambda} - 1\right)^{-k} = \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n!}{(n-k)!} (n-\lambda)^{-k}. \end{aligned}$$

Comme $(1 - \lambda/n)^n$ tend vers $e^{-\lambda}$, il suffit de prouver que $\frac{n!}{(n-k)!} (n-\lambda)^{-k}$ tend vers 1 quand n tend vers $+\infty$. La formule de Stirling est : quand n tend vers $+\infty$, $n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$. Alors, on a

$$\frac{n!}{(n-k)!} (n-\lambda)^{-k} \sim \left(1 + \frac{k}{n-k}\right)^n e^{-k} \left(\frac{n-\lambda}{n-k}\right)^k$$

qui converge bien vers 1. Donc $\hat{\lambda}_n$ converge en loi vers une variable de Poisson de paramètre λ .

2. Comme $\hat{\lambda}_n$ converge en loi vers une Poisson de paramètre 3. On a en particulier, quand n tend vers l'infini,

$$\mathbb{P}[\theta_n > \hat{\theta}_n + 2/n] = \mathbb{P}[1 > \hat{\lambda}_n] \longrightarrow \mathbb{P}[Z = 0] \approx 5\%.$$

Exercice 2.11 (Information de Fisher : entraînement)

Dans les modèles suivants, calculer l'information de Fisher associée aux n observations (si elle est bien définie), l'estimateur du maximum de vraisemblance et sa loi asymptotique :

1. $X_1, \dots, X_n \stackrel{i.i.d}{\sim} \mathcal{B}(\theta)$.
2. $X_1, \dots, X_n \stackrel{i.i.d}{\sim} \mathcal{N}(m, v)$.
3. $X_1, \dots, X_n \stackrel{i.i.d}{\sim} \mathcal{U}[0, \theta]$.

Correction de l'exercice 2.11 On rappelle les formules du cours pour le calcul de l'information de Fisher :

$$I_n(\theta) = nI_1(\theta) = n\mathbb{E}_\theta \nabla_\theta \log f(\theta, X) \nabla_\theta \log f(\theta, X)^\top = -n\mathbb{E}_\theta \nabla_\theta^2 \log f(\theta, X) = -n\nabla_a^2 \mathcal{D}(a, \theta)|_{a=\theta}$$

où $\mathcal{D}(a, \theta) = \mathbb{E}_\theta[\log f(a, X)]$. En utilisant une des trois formules précédentes, on obtient dans les différents modèles :

1. modèle de Bernoulli :

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

L'EMV est ici la moyenne empirique et on vérifie bien qu'il est asymptotiquement normal de variance asymptotique l'inverse de l'information de Fisher (grâce au TCL).

2. modèle Gaussien (moyenne et variance inconnues) :

$$I_n(m, v) = \begin{pmatrix} \frac{n}{v} & 0 \\ 0 & \frac{n}{2v^2} \end{pmatrix}.$$

L'EMV est ici $(\bar{X}_n, \hat{\sigma}_n^2)$ où $\bar{X}_n = \frac{1}{n} \sum_i X_i$ et $\hat{\sigma}_n^2 = \frac{1}{n} \sum_i (X_i - \bar{X}_n)^2$. L'EMV est asymptotiquement normal (soit parce que le modèle est régulier, soit en appliquant le TCL, méthode Delta et Slutsky en dimension 2) de variance asymptotique l'inverse de l'info de Fisher.

3. modèle uniforme : ce modèle n'est pas régulier – en particulier l'info de Fisher n'est pas définie (de manière classique). On peut néanmoins calculer, l'EMV qui est $\hat{\theta}_n^{mv} = \max_i X_i$ et son comportement asymptotique en étudiant sa fonction de répartition :

$$\mathbb{P}_\theta \left[\frac{n(\theta - \hat{\theta}_n^{mv})}{\theta} > x \right] = \mathbb{P}_\theta \left[\forall i = 1, \dots, n : \frac{n(\theta - X_i)}{\theta} > x \right] = \begin{cases} 1 & \text{si } x < 0 \\ \left(1 - \frac{x}{n}\right)^n & \text{quand } 0 < x < n \\ 0 & \text{sinon} \end{cases}$$

car pour tout i , sous \mathbb{P}_θ , $(\theta - X_i)/\theta \sim \mathcal{U}([0, 1])$. Alors quand $n \rightarrow \infty$, on obtient :

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\theta \left[\frac{n(\theta - \hat{\theta}_n^{mv})}{\theta} > x \right] = \begin{cases} 1 & \text{si } x < 0 \\ \exp(-x) & \text{quand } x > 0 \end{cases}$$

donc

$$\frac{n(\theta - \hat{\theta}_n^{mv})}{\theta} \rightsquigarrow \mathcal{E}(1)$$

où $\mathcal{E}(1)$ est une loi exponentielle de paramètre 1.

3 Tests

Exercice 3.1 (Test de Neyman-Pearson)

Chercher la région de rejet du test de Neyman-Pearson dans les cas suivants.

1. Loi exponentielle $\mathcal{E}(\theta)$. Test de $\theta = \theta_0$ contre $\theta = \theta_1$ avec $\theta_1 > \theta_0$.

2. Loi de Bernoulli $\mathcal{B}(\theta)$. Test de $\theta = \theta_0$ contre $\theta = \theta_1$ pour $\theta_1 > \theta_0$. Quel problème rencontre-t-on dans ce cas ?

Correction de l'exercice 3.1

1. La vraisemblance en θ du modèle est

$$L(\theta, (X_i)_i) = \prod_{i=1}^n \theta \exp(-\theta X_i) = \theta^n \exp\left(-\theta \sum_i X_i\right).$$

Le rapport de vraisemblance est

$$\frac{L(\theta_0, (X_i)_i)}{L(\theta_1, (X_i)_i)} = \exp\left(-(\theta_0 - \theta_1) \sum_i X_i\right).$$

Le rapport de vraisemblance est donc une fonction croissante de \bar{X}_n (on a ici $\theta_1 > \theta_0$). Alors, le test de Neyman-Pearson de niveau α est de la forme :

$$\varphi_\alpha((X_i)_i) = \begin{cases} H_0 & \text{quand } \bar{X}_n \geq t_\alpha \\ H_1 & \text{sinon} \end{cases}$$

où t_α est un seuil à choisir tel que

$$\mathbb{P}_{\theta_0}[\bar{X}_n < t_\alpha] = \alpha.$$

On sait qu'une telle solution existe car \bar{X}_n est une v.a.r. admettant une densité par rapport à la mesure de Lebesgue. Néanmoins, cette quantité reste difficile à calculer, on préfère alors fixer le seuil de manière asymptotique vue que \bar{X}_n est asymptotiquement Gaussien (par le TCL).

2. Comme précédemment, il suffit de calculer la vraisemblance et le rapport de vraisemblance dans ce modèle. On a pour la vraisemblance :

$$L(\theta, (X_i)_i) = \prod_{i=1}^n \theta^{X_i} (1 - \theta)^{1 - X_i}.$$

Le rapport de vraisemblance est

$$\frac{L(\theta_0, (X_i)_i)}{L(\theta_1, (X_i)_i)} = \left(\frac{\theta_0}{\theta_1}\right)^{\sum_i X_i} \left(\frac{1 - \theta_0}{1 - \theta_1}\right)^{\sum_i (1 - X_i)}.$$

Le rapport de vraisemblance est donc une fonction décroissante de \bar{X}_n (on a ici $\theta_1 > \theta_0$). Alors, le test de Neyman-Pearson de niveau α est de la forme :

$$\varphi_\alpha((X_i)_i) = \begin{cases} H_0 & \text{quand } \bar{X}_n \leq t_\alpha \\ H_1 & \text{sinon} \end{cases}$$

où t_α est un seuil à choisir tel que

$$\mathbb{P}_{\theta_0}[\bar{X}_n > t_\alpha] = \alpha.$$

Ici, cette équation n'admet pas nécessairement de solution car \bar{X}_n est une variable discrète. Dans ce cas, on peut avoir recours à des tests "randomisés" (hors programme), mais on préférera fixer le seuil t_α de manière asymptotique vue que \bar{X}_n est asymptotiquement Gaussien.

Exercice 3.2 (Test de Wald)

Lors des essais d'un type d'appareils ménagers, une association de consommateurs envisage les 3 issues suivantes : fonctionnement normal, mauvais fonctionnement et défaillance. Les probabilités de fonctionnement normal et de défaillance sont égales à p^2 et à $(1 - p)^2$ respectivement, où $p \in]0, 1[$ est un paramètre inconnu. Pour un échantillon de $n = 200$ appareils, on a observé que 112 appareils fonctionnent normalement, 12 sont défectueux et 76 fonctionnent mal. A partir de ces données, on cherche à inférer le paramètre p .

1. Proposer un modèle statistique pour ce problème.
2. Chercher l'estimateur du maximum de vraisemblance \hat{p}_n de p . Montrer qu'il est consistant et donner la loi limite de $\sqrt{n}(\hat{p}_n - p)$ quand $n \rightarrow \infty$.
3. À l'aide du test de Wald, tester l'hypothèse que $p = 1/2$ contre l'alternative $p \neq 1/2$ (on donnera la forme de la région critique et la p -value du test). On suppose connues les valeurs de la fonction de répartition de la loi normale standard.

Correction de l'exercice 3.2

1. C'est le modèle d'échantillonnage $\{\mathbb{P}_p^{\otimes n} : 0 < p < 1\}$ où

$$\mathbb{P}_p = p^2\delta_N + 2p(1 - p)\delta_{MF} + (1 - p)^2\delta_D$$

où N signifie normal, MF signifie mauvais fonctionnement et D signifie défectueux.

2. On note par #N, #MF, #D le nombre d'appareils dans chacune des trois catégories. On a $\#MF = n - \#N - \#D$.

La vraisemblance en p du modèle est

$$\begin{aligned} L(p, (X_i)_i) &= \prod_{i=1}^n \left[p^2 I(X_i = N) + 2p(1 - p) I(X_i = MF) + (1 - p)^2 I(X_i = D) \right] \\ &= (p^2)^{\#N} [2p(1 - p)]^{\#MF} [(1 - p)^2]^{\#D}. \end{aligned}$$

et la log-vraisemblance est

$$\ell_n(p, (X_i)_i) = \log \left(\frac{p}{1 - p} \right) [\#N - \#D] + (\#D - \#N) \log 2 + n \log [2p(1 - p)].$$

En étudiant la fonction de log-vraisemblance, on voit que la vraisemblance est maximale en

$$\hat{p}_n = \frac{1}{2} \left(\frac{\#N - \#D}{n} + 1 \right),$$

qui est donc l'estimateur du maximum de vraisemblance. Par la loi forte des grands nombres, on a :

$$\frac{\#N}{n} \xrightarrow[n \rightarrow +\infty]{} p^2, \text{ et } \frac{\#D}{n} \xrightarrow[n \rightarrow +\infty]{} (1 - p)^2$$

et donc $\hat{p}_n \xrightarrow[n \rightarrow +\infty]{} p$, càd \hat{p}_n est consistant. L'étude du comportement asymptotique de \hat{p}_n se déduit du TCL :

$$\frac{\#N - \#D}{n} = \frac{1}{n} \sum_{i=1}^n I(X_i = N) - I(X_i = D) := \frac{1}{n} \sum_{i=1}^n Z_i$$

et $\sqrt{n}(\bar{Z}_n - \mathbb{E}Z_1) \rightsquigarrow \mathcal{N}(0, \text{Var } Z_1)$. On obtient alors :

$$\sqrt{n}(\hat{p}_n - p) = \frac{\sqrt{n}}{2}(\bar{Z}_n - \mathbb{E}Z_1) \rightsquigarrow \mathcal{N}\left(0, \frac{p(1-p)}{2}\right)$$

3. On considère le problème de test

$$H_0 : p = \frac{1}{2} \text{ contre } H_1 : p \neq \frac{1}{2}.$$

La forme du test de Wald pour ce problème de test est

$$\varphi_\alpha((X_i)_i) = \begin{cases} H_0 & \text{quand } T_n \leq t_\alpha \\ H_1 & \text{sinon} \end{cases}$$

où t_α est un seuil à choisir tel que le niveau asymptotique du test est α et la statistique du test T_n est donnée ici par :

$$T_n = \sqrt{8n}|\hat{p}_n - 1/2|.$$

Sous H_0 , on a $T_n \rightsquigarrow \mathcal{N}(0, 1)$. On prend alors $t_\alpha = q_{1-\alpha/2}^{\mathcal{N}(0,1)}$.

Sous H_1 , on a $T_n \xrightarrow[n \rightarrow +\infty]{} +\infty$ donc le test est consistant.

(rem. : le test de Wald utilise T_n^2 pour statistique du test (ce qui fait intervenir une $\chi^2(1)$ en loi limite). Mais, en dimension $d = 1$, on peut utiliser directement T_n , les deux tests sont identiques : dans le premier cas la zone de rejet est $T_n^2 > q_{1-\alpha}^{\chi^2(1)}$ et dans le deuxième cas elle vaut $T_n > q_{1-\alpha/2}^{\mathcal{N}(0,1)}$. Ces deux zones sont identiques.

Numériquement, on obtient $\hat{p}_n = 0.5 * ((112 - 12)/200 + 1) = 0.75$ et $T_n = \sqrt{200 * 8}|0.75 - 1/2| = 10$. La p-value est $\mathbb{P}[|g| > 10]$ qui est très petite ; on va donc rejeter avec confiance.

Exercice 3.3 (Test de support)

Soient X_1, \dots, X_n de loi $\mathcal{U}[0, \theta]$ et $M = \max(X_i)$, $1 \leq i \leq n$. On cherche à tester $H_0 : \theta = 1$ contre $H_1 : \theta > 1$.

1. Pourquoi ne peut-on pas utiliser ici le test de Neyman-Pearson ?
2. On propose le test suivant : on rejette H_0 lorsque $M > c$ (c constante donnée). Calculer la fonction de puissance.
3. Quelle valeur prendre pour c pour obtenir un niveau de 5% ?
4. Si $n = 20$ et que la valeur observée de M est 0.96, que vaut la p-value ? quelle conclusion tirer sur H_0 ? Même question pour $M^{obs} = 1.04$.

Correction de l'exercice 3.3

1. Les densités n'ont pas même support. Le rapport de vraisemblance n'est donc pas défini.
2. La puissance d'un test est l'application qui mesure "le rejet à raison" : $\theta \in \Theta_1 \rightarrow \mathbb{P}_\theta[\text{rejet}]$. Etant donné la zone de rejet considérée ici, la fonction puissance est donnée pour tout $\theta > 1$ par

$$\mathbb{P}_\theta[\max X_i > c] = \begin{cases} 0 & \text{si } c \geq \theta \\ 1 & \text{si } c \leq 0 \\ 1 - \left(\frac{c}{\theta}\right)^n & \text{sinon.} \end{cases}$$

3. Pour avoir un niveau $\alpha \in (0, 1)$, il suffit de choisir c tel que $\mathbb{P}_{\theta=1}[\max_i X_i > c] = \alpha$ càd $c = (1 - \alpha)^{1/n}$. Pour $\alpha = 0.05$, on prend $c = (0.95)^{1/n}$.
4. Pour $n = 20$ et $M = 0.96$ la p-value vaut $\mathbb{P}_1[\max_i X_i > 0.96] \approx 0.56$: on va accepter H_0 . Pour $M = 1.04$, la p-value vaut $\mathbb{P}_1[\max_i X_i > 1.04] = 0$ on rejete donc avec un très haut niveau de confiance (c'est normal de rejeter vu qu'au moins un des X_i est plus grand que 1).

Exercice 3.4 (Peut-on retarder sa mort ?)

On prétend couramment que les mourants peuvent retarder leur décès jusqu'à certains événements importants. Pour tester cette théorie, Philips et King (1988, article paru dans *The Lancet*, prestigieux journal médical) ont collecté des données de décès aux environs d'une fête religieuse juive. Sur 1919 décès, 922 (resp. 997) ont eu lieu la semaine précédente (resp. suivante). Comment utiliser de telles données pour tester cette théorie grâce à un test asymptotique ?

Correction de l'exercice 3.4

1. On modélise ce problème par le modèle d'échantillonnage $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{B}(p)$ où

$$X_i = \begin{cases} 1 & \text{si décès avant la fête} \\ 0 & \text{sinon} \end{cases}$$

p est donc la probabilité de décéder avant la fête.

2. Pour la construction du test, le choix des hypothèses est très important. L'idée est de choisir les hypothèses telles que quand on rejette alors on obtient une information qui a de l'intérêt. Ici, on choisit les hypothèses telles que si on rejette alors on pourra dire que "les mourants peuvent retarder leur décès jusqu'à un certain événement important". On choisit alors le problème de test :

$$H_0 : p = \frac{1}{2} \text{ contre } H_1 : p < \frac{1}{2}$$

3. La famille de Bernoulli est une famille à rapport de vraisemblance monotone : le rapport de vraisemblance dépend de manière monotone de la moyenne empirique \bar{X}_n . On va donc utiliser la moyenne empirique pour construire la statistique de test. On considère le test

$$\varphi_\alpha((X_i)_i) = \begin{cases} H_0 & \text{quand } T_n \geq t_\alpha \\ H_1 & \text{sinon} \end{cases}$$

où $T_n = \sqrt{n}(\bar{X}_n - 1/2)$.

- Sous H_1 : pour tout $p < 1/2$, sous \mathbb{P}_p , T_n tend p.s. vers $-\infty$ (c'est pour ça qu'on a choisit cette forme de test).
- Pour le calcul du seuil t_α , on veut :

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_{1/2}[T_n < t_\alpha] = \alpha.$$

Sous $p = 1/2$: $T_n \rightsquigarrow \mathcal{N}(0, 1/4)$, on prend alors $t_\alpha = q_\alpha^{\mathcal{N}(0,1)}/2$.

4. numériquement, on obtient $2T_n = 2\sqrt{1919}(922/1919 - 1/2) \approx -1.712$. La p-value du test est $\mathbb{P}[g < -1.712] = 0.04$ où $g \sim \mathcal{N}(0, 1)$. On rejette donc l'hypothèse avec confiance. On en déduit que les gens "peuvent retarder leur mort".

4 Modèle de régression

Exercice 4.1 (Modèle de régression multiple)

On considère le modèle de regression multiple

$$y = \theta_0 e + X\theta + \xi, \quad \text{où } \mathbb{E}[\xi] = 0, \mathbb{E}[\xi\xi^T] = \sigma^2 I_n, e = (1, 1, \dots, 1)^T$$

avec X une matrice $n \times k$ de rang k et y, ξ des vecteurs de \mathbb{R}^n . Les paramètres $\theta_0 \in \mathbb{R}$ et $\theta \in \mathbb{R}^k$ sont inconnus. On note $\hat{\theta}_0$ et $\hat{\theta}$ les estimateurs des moindres carrés de θ_0 et θ .

1. On note $\hat{y} = \hat{\theta}_0 e + X\hat{\theta}$. Montrer que $\bar{\hat{y}} = \bar{y}$, où \bar{y} (resp. $\bar{\hat{y}}$) est la moyenne des y_i (resp. des \hat{y}_i). En déduire que $\bar{y} = \hat{\theta}_0 + \bar{X}\hat{\theta}$ où $\bar{X} = \frac{1}{n}e^T X = (\bar{X}_1, \dots, \bar{X}_k)$.
2. Montrer l'équation d'analyse de la variance :

$$\|y - \bar{y}e\|^2 = \|y - \hat{y}\|^2 + \|\hat{y} - \bar{y}e\|^2.$$

En déduire que le coefficient de détermination

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

est toujours inférieur à 1.

3. Supposons que $Z = [e, X]$ est de rang $k + 1$. Calculez en fonction de Z la matrice de covariance de $(\hat{\theta}_0, \hat{\theta})$. Comment accède-t-on à $\text{Var}(\hat{\theta}_j)$, pour $j = 0, \dots, p$?

4. On suppose dorénavant que $\theta_0 = 0$ et donc

$$y = X\theta + \xi, \quad \mathbb{E}[\xi] = 0, \quad \mathbb{E}[\xi\xi^T] = \sigma^2 I_n.$$

L'estimateur des moindres carrés $\tilde{\theta}$ dans ce modèle est-il égal à $\hat{\theta}$?

5. A-t-on la relation $\bar{y} = \bar{y}$? Que dire du R^2 dans ce modèle ?

Correction de l'exercice 4.1

1. Par définition, l'estimateur des moindres carrés est donné par :

$$(\hat{\theta}_0, \hat{\theta})^T \in \operatorname{argmin}_{(\theta'_0, \theta')^T \in \mathbb{R} \times \mathbb{R}^k} \|y - \theta'_0 e - X\theta'\|_2.$$

Alors $\hat{y} = \hat{\theta}_0 e + X\hat{\theta}$ est la projection orthogonale de y sur $\operatorname{vect}(e, X^{(1)}, \dots, X^{(k)})$ où $X^{(1)}, \dots, X^{(k)}$ sont les vecteurs colonnes de X . En particulier, pour tout $\theta'_0 \in \mathbb{R}, \theta' \in \mathbb{R}^k$, on a

$$\langle y - \hat{y}, \theta'_0 e + X\theta' \rangle = 0.$$

En particulier, pour $\theta'_0 = 1, \theta' = 0$, on a $\langle y - \hat{y}, e \rangle = 0$ et comme $\bar{y} = n^{-1} \langle y, e \rangle$ (de même $\bar{\hat{y}} = n^{-1} \langle \hat{y}, e \rangle$), on a bien $\bar{y} = \bar{\hat{y}}$. De plus,

$$\bar{\hat{y}} = n^{-1} \langle \hat{y}, e \rangle = n^{-1} \langle \hat{\theta}_0 e + X\hat{\theta}, e \rangle = \hat{\theta}_0 + \bar{X}\hat{\theta}$$

où $\bar{X} = (\bar{X}^{(1)}, \dots, \bar{X}^{(k)})$.

2. $\bar{y}e$ est un élément de $\operatorname{vect}(e, X^{(1)}, \dots, X^{(k)})$. Comme \hat{y} est le projeté orthogonal de y sur cet espace, on voit que $y - \hat{y}$ est orthogonal à $\bar{y}e - \hat{y}$. par Pythagore, on a

$$\|y - \bar{y}e\|_2^2 = \|y - \hat{y}\|_2^2 + \|\hat{y} - \bar{y}e\|_2^2.$$

On a donc

$$R^2 = \frac{\|\hat{y} - \bar{y}e\|_2^2}{\|y - \bar{y}e\|_2^2} \leq 1.$$

1. $R^2 = 1$ signifie que y est dans $\operatorname{vect}(e, X^{(1)}, \dots, X^{(k)})$ (modèle sans bruit).
2. $R^2 = 0$ signifie que $\hat{y} = \bar{y}e$. Donc y est orthogonal à $\operatorname{vect}(X^{(1)}, \dots, X^{(k)})$. Alors $X^{(1)}, \dots, X^{(k)}$ sont des mauvaises variables pour expliquer ou prédire y .
3. Soit Proj l'opérateur de projection sur $\operatorname{vect}(e, X^{(1)}, \dots, X^{(k)})$. On a $Z(\hat{\theta}_0, \hat{\theta})^T = \operatorname{Proj}(y)$. On a pour tout $\theta'_0 \in \mathbb{R}, \theta' \in \mathbb{R}^k, \langle y - Z(\hat{\theta}_0, \hat{\theta})^T, Z(\theta'_0, \theta')^T \rangle = 0$. Par ailleurs,

$$\langle y - Z(\hat{\theta}_0, \hat{\theta})^T, Z(\theta'_0, \theta')^T \rangle = \langle Z^T y - Z^T Z(\hat{\theta}_0, \hat{\theta})^T, (\theta'_0, \theta')^T \rangle.$$

Donc $Z^T y = Z^T Z(\hat{\theta}_0, \hat{\theta})^T$. Comme la matrice carrée $Z^T Z$ de taille $k + 1$ est de rang $k + 1$, elle est de rang plein donc inversible. Alors $(Z^T Z)^{-1} Z^T y = (\hat{\theta}_0, \hat{\theta})^T$.

On peut aussi voir que

$$(\hat{\theta}_0, \hat{\theta})^T \in \operatorname{argmin}_{\theta'_0 \in \mathbb{R}, \theta' \in \mathbb{R}^k} \|y - \theta'_0 e - X\theta'\|_2.$$

Alors, $(\hat{\theta}_0, \hat{\theta})^\top$ minimise la fonction convexe $F(u) = \|y - Zu\|_2^2$ sur \mathbb{R}^{k+1} . Alors $(\hat{\theta}_0, \hat{\theta})^\top$ est solution de $F'(u) = 0$ càd $Z^\top(y - Zu) = 0$. Donc $(Z^\top Z)^{-1}Z^\top y = (\hat{\theta}_0, \hat{\theta})^\top$.

La matrice de covariance de $\hat{\Theta} := (\hat{\theta}_0, \hat{\theta})^\top$ est donnée par

$$\Sigma = \mathbb{E}[(\hat{\Theta} - \mathbb{E}\hat{\Theta})(\hat{\Theta} - \mathbb{E}\hat{\Theta})^\top].$$

L'espérance de $\hat{\Theta}$ est donnée par

$$\mathbb{E}\hat{\Theta} = \mathbb{E}(Z^\top Z)^{-1}Z^\top y = (Z^\top Z)^{-1}Z^\top Z(\theta_0, \theta)^\top = (\theta_0, \theta)^\top.$$

On en déduit que (étant donné que $\mathbb{E}\zeta\zeta^\top = \sigma^2 I_n$)

$$\Sigma = \mathbb{E}(Z^\top Z)^{-1}Z\zeta\zeta^\top Z(Z^\top Z)^{-1} = \sigma^2(Z^\top Z)^{-1}.$$

Pour tout $j = 0, \dots, k$,

$$\text{var}(\hat{\theta}_j) = \text{var}(\langle e_j, (\hat{\theta}_0, \hat{\theta})^\top \rangle) = \sigma^2 e_j^\top (Z^\top Z)^{-1} e_j = \sigma^2 (Z^\top Z)^{-1}_{jj}.$$

4. On a $\tilde{\theta} = (X^\top X)^{-1}X^\top y$ càd, $\tilde{\theta}$ est le projeté de y sur $\text{vect}(X^{(1)}, \dots, X^{(k)})$. En général $\tilde{\theta} \neq \hat{\theta}$ sauf quand e est orthogonal à $\text{vect}(X^{(1)}, \dots, X^{(k)})$.
5. Si $e \notin \text{vect}(X^{(1)}, \dots, X^{(k)})$ alors on n'a pas $\langle e, y - \hat{y} \rangle = 0$ donc $\bar{y} \neq \hat{y}$. Dans ce modèle R^2 n'a pas de sens.

Exercice 4.2 (Régression Ridge)

On considère le modèle de regression

$$Y \underset{(n,1)}{=} X \underset{\substack{(n,k) \\ (k,1)}}{\theta} + \underset{(n,1)}{\xi}.$$

On suppose que X est une matrice déterministe, $\mathbb{E}[\xi] = 0$, $\mathbb{E}[\xi\xi^\top] = \sigma^2 I_n$,

1. On suppose que $k > n$. Que dire de l'estimation par moindres carrés ?
2. On appelle estimateur Ridge regression de paramètre de régularisation $\lambda > 0$ l'estimateur

$$\hat{\theta}_\lambda = \arg \min_{\theta \in \mathbb{R}^k} \{ \|Y - X\theta\|^2 + \lambda \|\theta\|^2 \}.$$

Exprimez $\hat{\theta}_\lambda$ en fonction de X , Y et λ . Cet estimateur est-il défini pour $k > n$?

3. Calculez la moyenne et la matrice de covariance de l'estimateur Ridge. Est-il sans biais ?
4. On suppose maintenant que $k = 1$, ce qui correspond au modèle de régression simple. Montrer qu'il existe une valeur de λ telle que le risque de l'estimateur Ridge de paramètre λ est inférieur au risque de l'estimateur des MC.

Correction de l'exercice 4.2 On peut voir la régression Ridge, comme une relaxation de la méthode MC dans le cas où les variables explicatives sont colinéaires (càd quand il y a de la redondance d'information dans les variables explicatives). Pour définir l'EMC de manière unique, on a besoin que $X^\top X$ soit inversible. Dans ce cas $\theta^{MC} = (X^\top X)^{-1} X^\top Y$. Comme $\ker(X^\top X) = \ker X$, on a vu que $X^\top X$ est inversible si et seulement si les colonnes de X ne sont pas colinéaires. D'un point de vue statistiques, des colonnes de X linéairement dépendantes signifie qu'il y a de la redondance d'information parmi les variables explicatives. Par ailleurs, quand $X^\top X$ est inversible mais que son conditionnement (ratio plus grande valeur singulière sur plus petite valeur singulière) est grand alors un calcul effectif de l'EMC est difficile. On va donc considérer, un estimateur qui "régularise" l'EMC ou "conditionne" la matrice de Gram $X^\top X$. Pour cela, on va inverser $X^\top X + \lambda I_k$ et ainsi considérer l'*estimateur Ridge*

$$\hat{\theta}_\lambda = (X^\top X + \lambda I_k)^{-1} X^\top Y.$$

Cet estimateur n'est plus sans biais mais il peut améliorer le risque quadratique de l'EMC. On peut voir ça comme un compromis biais variance : on perd un peu sur l'espérance mais on gagne sur la variance dans l'égalité

$$\mathbb{E}(\hat{\theta}_\lambda)^2 = (\mathbb{E}\hat{\theta}_\lambda - \mathbb{E}\theta)^2 + \text{var}(\hat{\theta}_\lambda).$$

On doit aussi faire en sorte de bien choisir $\lambda > 0$. Ceci introduit le problème de la sélection de paramètre en statistique (et notamment la méthode de validation croisée).

1. Quand $k > n$, la matrice $X : \mathbb{R}^k \mapsto \mathbb{R}^n$ a un noyau et comme $\ker(X^\top X) = \ker X$, la matrice $X^\top X$ n'est plus inversible. On sait que l'EMC est défini comme solution de l'équation $X^\top X \hat{\theta} = X^\top Y$ qui admet une infinité de solution (un espace affine dirigé par $\ker(X^\top X)$). L'EMC n'est donc pas uniquement défini. On peut alors choisir parmi cet ensemble infini de solutions, une ayant certaines propriétés supplémentaires. On va chercher celle ayant une petite norme 2.
2. On introduit la fonction

$$F(\theta) = \|Y - X\theta\|_2^2 + \lambda \|\theta\|_2^2, \quad \forall \theta \in \mathbb{R}^k.$$

Cette fonction est strictement convexe et tend vers l'infini quand $\|\theta\|_2$ tend vers l'infini donc elle admet un unique minimum $\hat{\theta}_\lambda$ qui est solution de l'équation $\Delta F(\hat{\theta}_\lambda) = 0$ càd $-2X^\top(Y - X\hat{\theta}_\lambda) + 2\lambda\hat{\theta}_\lambda = 0$. On a donc

$$\hat{\theta}_\lambda = (X^\top X + \lambda I_k)^{-1} X^\top Y.$$

3. Le biais de l'ER est donné par :

$$\mathbb{E}\hat{\theta}_\lambda = (X^\top X + \lambda I_k)^{-1} X^\top \theta$$

qui est différent de θ en général. Alors l'ER est en général un estimateur biaisé. La matrice de covariance est donnée par :

$$\begin{aligned} \text{var}(\hat{\theta}_\lambda) &= (X^\top X + \lambda_k)^{-1} X^\top \mathbb{E}\zeta\zeta^\top X (X^\top X + \lambda_k)^{-1} \\ &= \sigma^2 (X^\top X + \lambda_k)^{-1} X^\top X (X^\top X + \lambda_k)^{-1}. \end{aligned}$$

4. Pour $k = 1$, on écrit $Y = X\theta + \zeta$ où X est un vecteur de \mathbb{R}^n . Dans ce cas $X^\top X = \|X\|_2^2$ alors l'EMC et l'ER sont donnés par :

$$\hat{\theta} = \hat{\theta}^{MC} = \frac{\langle X, Y \rangle}{\|X\|_2^2} \text{ et } \hat{\theta}_\lambda = \hat{\theta}^{ER} = \frac{\langle X, Y \rangle}{\|X\|_2^2 + \lambda}.$$

Le risque quadratique de l'EMC est

$$\begin{aligned} \mathbb{E}(\hat{\theta} - \theta)^2 &= \text{var}(\hat{\theta}) = \mathbb{E}\hat{\theta}^2 - (\mathbb{E}\hat{\theta})^2 = \frac{\mathbb{E}\langle X, Y \rangle^2}{\|X\|_2^4} - \theta^2 \\ &= \frac{\mathbb{E}\langle X, X\theta + \zeta \rangle}{\|X\|_2^2} - \theta^2 = \frac{\sigma^2}{\|X\|_2^2}. \end{aligned}$$

La décomposition biais-variance du risque quadratique de l'ER donne :

$$\mathbb{E}(\hat{\theta}_\lambda - \theta)^2 = (\mathbb{E}\hat{\theta}_\lambda - \mathbb{E}\theta)^2 + \text{var}(\hat{\theta}_\lambda) = \left(\frac{\|X\|_2^2 \theta}{\|X\|_2^2 + \lambda} - \theta \right)^2 + \frac{\sigma^2 \|X\|_2^2}{(\|X\|_2^2 + \lambda)^2}.$$

En posant $\mu = \lambda / \|X\|_2^2$, on est amené à chercher $\mu > 0$ tel que

$$\left(\frac{1}{1 + \mu} - 1 \right)^2 \theta^2 + \frac{(\sigma^2 / \|X\|_2^2)}{(1 + \mu)^2} < (\sigma^2 / \|X\|_2^2) \tag{4}$$

càd $\mu(\theta^2 - (\sigma^2 / \|X\|_2^2)) < 2(\sigma^2 / \|X\|_2^2)$. Si $\theta^2 \|X\|_2^2 > \sigma^2$ alors pour tout λ tel que

$$\lambda < \frac{2\sigma^2 \|X\|_2^2}{\theta^2 \|X\|_2^2 - \sigma^2},$$

le risque quadratique de l'ER est moindre que celui de l'EMC. Quand $\theta^2 \|X\|_2^2 < \sigma^2$ alors pour tout $\lambda > 0$, le risque quadratique de l'ER est moindre que celui de l'EMC.

Le ratio θ^2 / σ^2 (et en général pour tout k , $\|\theta\|_2^2 / \sigma^2$) est appelé le "signal sur bruit". Quand il est grand ($\theta^2 / \sigma^2 > \|X\|_2^{-2}$), il faut choisir λ assez petit et quand il est petit, l'ER est toujours meilleur (en terme de risque quadratique) que l'EMC pour n'importe quel λ .

Exercice 4.3 (Théorème de Gauss-Markov)

On considère le modèle de regression

$$Y = X \theta + \xi .$$

$(n,1) \quad (n,k)(k,1) \quad (n,1)$

On suppose que X est une matrice déterministe, $\mathbb{E}[\xi] = 0$, $\mathbb{E}[\xi\xi^T] = \sigma^2 I_n$, $\text{Rang}(X) = k$. On note $\hat{\theta}$ l'estimateur des MC de θ .

1. Montrer que $\hat{\theta}$ est sans biais et expliciter sa matrice de covariance.
2. Soit $\tilde{\theta}$ un estimateur de θ linéaire en Y , i.e., $\tilde{\theta} = LY$ pour une matrice $L \in \mathbb{R}^{k \times n}$ déterministe. Donner une condition nécessaire et suffisante sur L pour que $\tilde{\theta}$ soit sans biais. On supposera maintenant cette hypothèse vérifiée.

3. Calculer la matrice de covariance de $\tilde{\theta}$. En posant $\Delta = L - (X^T X)^{-1} X^T$ montrer que $\Delta X = 0$ et $\text{cov}(\tilde{\theta}) = \text{cov}(\hat{\theta}) + \sigma^2 \Delta \Delta^T$. En déduire que

$$\mathbb{E}[(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T] \geq \mathbb{E}[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \quad (\text{inégalité au sens matriciel}).$$

4. En passant au risques quadratiques $\mathbb{E}[\|\tilde{\theta} - \theta\|_2^2]$ et $\mathbb{E}[\|\hat{\theta} - \theta\|_2^2]$, en déduire que l'estimateur des MC est optimal dans la classe de tous les estimateurs linéaires sans biais.

Correction de l'exercice 4.3

1. Par définition, $\hat{\theta}$ minimise $F(u) = \|y - Xu\|_2^2$ donc $\hat{\theta} = (X^T X)^{-1} X^T y$. On remarque que $\text{rang}(X) = k$ donc $n \geq k$ et X est injective (donc $X^T X$ est inversible : en effet, $X^T X$ est symétrique donc diagonalisable et si λ est une valeur propre de vecteur propre u alors $\|Xu\|_2^2 = \lambda \|u\|_2^2$, donc $\lambda \neq 0$ donc $X^T X$ est inversible).

On a donc $\mathbb{E}\hat{\theta} = (X^T X)^{-1} X^T \mathbb{E}y = (X^T X)^{-1} X^T X\theta = \theta$. Donc $\hat{\theta}$ est bien un estimateur sans biais. La matrice de covariance de $\hat{\theta}$ est donnée par

$$\Sigma := \mathbb{E}(\hat{\theta} - \mathbb{E}\hat{\theta})(\hat{\theta} - \mathbb{E}\hat{\theta})^T = (X^T X)^{-1} X^T \mathbb{E}\zeta \zeta^T X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}$$

2. On a $\mathbb{E}LY = LX\theta$. Pour que $\tilde{\theta} = LY$ soit sans biais, il faut et il suffit que $LX\theta = \theta$. Ceci étant vrai pour tout θ , on doit avoir $LX = I_k$.
3. $\Sigma = \mathbb{E}((\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T) = L \text{var}(Y) L^T = \sigma^2 LL^T$. Comme $LX = I_k$, on a :

$$\Delta X = LX - (X^T X)^{-1} X^T X = I_k - I_k = 0$$

et la covariance de $\tilde{\theta}$ est donnée par :

$$\begin{aligned} \text{var}(\tilde{\theta}) &= \text{var}(\Delta Y + \hat{\theta}) = \text{var}(\Delta Y) + \text{var}(\hat{\theta}) + \text{cov}(\hat{\theta}, \Delta Y) + \text{cov}(\Delta Y, \hat{\theta}) \\ &= \sigma^2 \Delta \Delta^T + \text{var}(\hat{\theta}) + \text{cov}(\hat{\theta}, \Delta Y) + \text{cov}(\Delta Y, \hat{\theta}). \end{aligned}$$

Par ailleurs, comme $\Delta X = 0$, on a $\mathbb{E}\Delta Y = 0$ et

$$\text{cov}(\Delta Y, \hat{\theta}) = \mathbb{E}[\Delta Y \hat{\theta}^T] = \Delta \mathbb{E}[(X\theta + \zeta) \zeta^T X (X^T X)^{-1}] = 0$$

car $\mathbb{E}\zeta \zeta^T = \sigma^2 I_n$. De même $\text{cov}(\hat{\theta}, \Delta Y) = 0$. On en déduit que

$$\text{var}(\tilde{\theta}) = \text{var}(\hat{\theta}) + \sigma^2 \Delta \Delta^T \succeq \text{var}(\hat{\theta}).$$

4. On a

$$\|\tilde{\theta} - \theta\|_2^2 = \sum_{j=1}^k (\tilde{\theta}_j - \theta_j)^2 = \sum_{j=1}^k e_j^T (\tilde{\theta} - \theta) (\tilde{\theta} - \theta)^T e_j$$

alors

$$\mathbb{E} \|\tilde{\theta} - \theta\|_2^2 = \sum_{j=1}^k e_j \text{var}(\tilde{\theta}) e_j$$

de même $\mathbb{E} \left\| \hat{\theta} - \theta \right\|_2^2 = \sum_{j=1}^k e_j \text{var}(\hat{\theta}) e_j$. Mais d'après 3., on a $\text{var}(\tilde{\theta}) \succeq \text{var}(\hat{\theta})$. Notamment, pour tout j , $e_j^\top \text{var}(\tilde{\theta}) e_j \succeq e_j^\top \text{var}(\hat{\theta}) e_j$. On a donc

$$\mathbb{E} \left\| \tilde{\theta} - \theta \right\|_2^2 \geq \mathbb{E} \left\| \hat{\theta} - \theta \right\|_2^2.$$

Exercice 4.4 (La formule de Woodbury)

Etant donné une matrice carré $A \in \mathbb{R}^{n \times n}$ inversible, $U \in \mathbb{R}^{n \times k}$, $V \in \mathbb{R}^{k \times n}$ deux matrices rectangulaires et $C \in \mathbb{R}^{k \times k}$ une matrice carré inversible. On suppose que $A + UCV$ et $C^{-1} + VA^{-1}U$ sont inversibles. La formule de Woodbury dit que

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}. \tag{5}$$

En d'autres termes, l'inverse d'une perturbation de rang k de A peut être obtenue en faisant une perturbation de rang k de A^{-1} . Elle permet de calculer l'inverse de $A + UCV$ seulement à partir des inverses de A et de C .

Quand $k = 1$, c'est-à-dire quand C est un scalaire alors la formule de Woodbury est connue sous le nom de Sherman-Morrison-Woodbury :

$$(A + uv^\top)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}. \tag{6}$$

La formule de Woodbury trouve des applications en statistiques. Par exemple, l'estimateur Ridge pour une matrice de design $A \in \mathbb{R}^{n \times d}$, un vecteur de sorties $y \in \mathbb{R}^n$ et un paramètre de régularisation $\lambda > 0$ est défini par

$$\hat{t} \in \underset{t \in \mathbb{R}^d}{\text{argmin}} (1/2) \|At - y\|_2^2 + \lambda \|t\|_2^2. \tag{7}$$

On montre que

$$\hat{t} = \left(A^\top A + \lambda I_d \right)^{-1} A^\top y \tag{8}$$

et grâce à la formule de Woodbury

$$\hat{t} = \left(A^\top A + \lambda I_d \right)^{-1} A^\top y = A^\top \left(AA^\top + \lambda I_n \right)^{-1} y. \tag{9}$$

L'intérêt de la dernière formule est que si $n < d$ alors il est plus facile d'inverser la matrice $AA^\top + \lambda I_n$ de taille $n \times n$ que la matrice $A^\top A + \lambda I_d$ de taille $d \times d$.

1. Montrer que pour toute matrice P telle que $I + P$ est inversible, on a $(I + P)^{-1} = I - (I + P)^{-1}P$.
2. Montrer que pour toutes matrices P et Q telles que $I + PQ$ et $I + QP$ sont inversibles, on a $(I + PQ)^{-1}P = P(I + QP)^{-1}$.
3. En utilisant la première question, montrer que

$$(A + UCV)^{-1} = (I + A^{-1}UCV)^{-1}A^{-1} = A^{-1} - (I + A^{-1}UCV)^{-1}A^{-1}UCVA^{-1}.$$

4. Ensuite, en utilisant la deuxième question, montrer la formule de Woodbury.
5. Dédurre la formule de Sherman-Morrison-Woodbury à partir de la formule de Woodbury.
6. Prouver (8).
7. Prouver (9).

Correction de l'exercice 4.4

1. On a

$$(I + P)^{-1} = (I + P)^{-1}(I + P - P) = I - (I + P)^{-1}P.$$

2. On a $P(I + QP) = P + PQP = (I + PQ)P$ et en multipliant à gauche par $(I + PQ)^{-1}$ et à droite par $(I + QP)^{-1}$, on obtient le résultat.

3. On applique la première question à $P = A^{-1}UCV$. Comme A et $A + UCV$ sont inversible on voit que $I + A^{-1}UCV$ est inversible en tant que produit de deux matrices inversibles : $I + A^{-1}UCV = A^{-1}(A + UCV)$. On a alors d'après 1)

$$\begin{aligned} (A + UCV)^{-1} &= [A(I + A^{-1}UCV)]^{-1} = (I + A^{-1}UCV)^{-1}A^{-1} \\ &= [I - (I + A^{-1}UCV)^{-1}A^{-1}UCV]A^{-1} = A^{-1} - (I + A^{-1}UCV)^{-1}A^{-1}UCVA^{-1}. \end{aligned}$$

4. On sait déjà que $I + A^{-1}UCV$ est inversible. On a $I + CVA^{-1}U = C(C^{-1} + VA^{-1}U)$ et donc $I + CVA^{-1}U$ est le produit de deux matrices inversibles, elle est donc elle-même inversible. On applique la relation de la deuxième question à $P = A^{-1}U$ et $Q = CV$:

$$(I + A^{-1}UCV)^{-1}A^{-1}U = A^{-1}U(I + CVA^{-1}U)^{-1}.$$

On utilise la dernière relation dans la question 3) :

$$\begin{aligned} (A + UCV)^{-1} &= A^{-1} - (I + A^{-1}UCV)^{-1}A^{-1}UCVA^{-1} = A^{-1} - A^{-1}U(I + CVA^{-1}U)^{-1}CVA^{-1} \\ &= A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}. \end{aligned}$$

5. On applique la formule de Woodbury quand $k = 1$ et $C = 1$:

$$(A + uv^T)^{-1} = A^{-1} - A^{-1}u(1 + v^T A^{-1}u)v^T A^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1}u}.$$

6. La fonction $F : t \in \mathbb{R}^d \rightarrow \|At - y\|_2^2 + \lambda \|t\|_2^2$ est convexe différentiable. On a donc \hat{t} minimise F si et seulement $\nabla F(\hat{t}) = 0$ qui est équivalent à $A^T(At - y) + \lambda t = 0$. Par ailleurs, $A^T A + \lambda I$ est inversible (car c'est une matrice symétrique dont les valeurs singulières sont plus grandes que $\lambda > 0$), on obtient donc bien que le seul minimiseur de F est $\hat{t} = (A^T A + \lambda I_d)^{-1}A^T y$.

7. La formule de Woodbury appliquée à $A = \lambda I_d$, $U = A^T$, $C = I_n$ et $V = A$ donne :

$$(\lambda I_d + A^T A)^{-1} = \lambda^{-1}I_d - \lambda^{-1}A^T(I_n + \lambda^{-1}AA^T)^{-1}A\lambda^{-1} = \lambda^{-1} \left(I_d - A^T(\lambda I_n + AA^T)^{-1}A \right).$$

On multiplie ensuite à droite par A^T pour obtenir

$$(\lambda I_d + A^T A)^{-1}A^T = \lambda^{-1} \left(A^T - A^T(\lambda I_n + AA^T)^{-1}(AA^T + \lambda I_n - \lambda I_n) \right) = A^T(AA^T + \lambda I_n)^{-1}.$$

5 Examen du lundi 26 octobre 2015

Exercice 5.1 (Estimation de la variance et borne de Cramer-Rao)

On considère le modèle d'échantillonnage $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \theta)$ où $\theta > 0$ (la variance) est le paramètre inconnu à estimer.

1. Calculer l'information de Fisher en $\theta > 0$ contenue dans ce n -échantillon.
2. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{\text{mv}}$ de θ .
3. Calculer le biais $b(\theta) = \mathbb{E}_\theta \hat{\theta}_n^{\text{mv}} - \theta$ et le risque quadratique $R_\theta(\hat{\theta}_n^{\text{mv}}) = \mathbb{E}_\theta (\hat{\theta}_n^{\text{mv}} - \theta)^2$ de $\hat{\theta}_n^{\text{mv}}$.
4. Rappeler la borne de Cramer-Rao pour ce problème. En déduire, que $\hat{\theta}_n^{\text{mv}}$ atteint la borne de Cramer-Rao parmi tous les estimateurs sans biais.

Rappel : si $g \sim \mathcal{N}(0, 1)$ alors $\mathbb{E}g^4 = 3$.

Correction de l'exercice 5.1

1. Soit $\theta > 0$. L'information de Fisher contenue dans un n -échantillon vaut n fois celle contenue dans une seule donnée : $I_n(\theta) = nI_1(\theta)$. L'information de Fisher dans une donnée est :

$$I_1(\theta) = \mathbb{E}_\theta [(\partial_\theta \log f(\theta, X))^2] = \mathbb{E}_\theta \left[\left(\frac{-1}{2\theta} + \frac{X^2}{2\theta^2} \right)^2 \right] = \frac{\text{Var } X^2}{4\theta^4} = \frac{1}{2\theta^2}.$$

2. La fonction de vraisemblance en $\theta > 0$ est

$$L(\theta, (X_i)_i) = \left(\frac{1}{2\pi\theta} \right)^{n/2} \exp \left(-\frac{1}{2\theta} \sum_{i=1}^n X_i^2 \right)$$

et donc la log-vraisemblance est

$$\ell_n(\theta, (X_i)_i) = -\frac{n}{2} \log(2\pi\theta) - \frac{1}{2\theta} \sum_{i=1}^n X_i^2.$$

En étudiant la fonction ℓ_n , on voit que l'EMV est $\hat{\theta}_n^{\text{mv}} = \frac{1}{n} \sum_{i=1}^n X_i^2$.

3. la biais de $\hat{\theta}_n^{\text{mv}}$ est $b(\theta) = \mathbb{E}_\theta \hat{\theta}_n^{\text{mv}} - \theta = \mathbb{E}_\theta X_1^2 - \theta = 0$ (car $\mathbb{E}X_1 = 0$ donc $\text{Var } X_1 = \mathbb{E}X_1^2$). Son risque quadratique est

$$R_\theta(\hat{\theta}_n^{\text{mv}}) = \mathbb{E}_\theta (\hat{\theta}_n^{\text{mv}} - \theta)^2 = \mathbb{E}_\theta \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \theta \right)^2 = \frac{\text{Var } X_1^2}{n} = \frac{2\theta^2}{n}.$$

4. La borne de Cramer-Rao dit que si $\hat{\theta}$ est un estimateur de θ et si $b(\theta) = \mathbb{E}_\theta \hat{\theta} - \theta$ est le biais de cet estimateur alors :

$$\mathbb{E}_\theta (\hat{\theta} - \theta)^2 \geq \frac{(1 + b'(\theta))^2}{I_n(\theta)} + b(\theta)^2.$$

En particulier, si $\hat{\theta}$ est sans biais alors $b(\theta) = 0$ et $R_\theta(\hat{\theta}) \geq I_n(\theta)^{-1} = 2\theta^2/n$. Or le risque quadratique de $\hat{\theta}_n^{\text{mv}}$ est égal à $2\theta^2/n$ donc $\hat{\theta}_n^{\text{mv}}$ atteint la borne de Cramer-Rao parmi tous les estimateurs sans biais.

Exercice 5.2 (Estimateur on-line de la moyenne)

Dans le modèle d'échantillonnage X_1, \dots, X_n où $\mathbb{E}|X_1| < \infty$, on note $\mathbb{E}X_1 = \theta$; construire :

1. un estimateur *batch* de la moyenne θ
2. un estimateur *on-line* de la moyenne θ

Correction de l'exercice 5.2

1. Un estimateur *batch* est donné par la moyenne empirique \bar{X}_n
2. Un estimateur *on-line* est donné par l'algorithme de Robbins-Monro pour les fonctions

$$f(x, X) = x - X \text{ et } F(x) = \mathbb{E}f(x, X) = x - \mathbb{E}X.$$

Comme $x = \mathbb{E}X$ est l'unique zéro de F , on est naturellement amené à considérer une méthode de Newton stochastique :

$$x_{k+1} = x_k - \eta_n(x_k - X_{k+1}).$$

x_n est donc un estimateur *on-line* de la moyenne.

Exercice 5.3 (Deux échantillons gaussiens)

On observe $X_1, \dots, X_m \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu_1, v)$ et $Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu_2, v)$ deux échantillons Gaussiens ayant même variance v mais des moyennes différentes. On suppose que les deux échantillons sont indépendants entre eux.

1. Calculer la vraisemblance en (μ_1, μ_2, v) de l'observation $(X_1, \dots, X_m, Y_1, \dots, Y_n)$.
2. En déduire l'estimateur du maximum de vraisemblance de (μ_1, μ_2, v) .
3. On suppose dorénavant dans toutes les questions qui suivent que $m = n$. Calculer l'information de Fisher en (μ_1, μ_2, v) contenue dans le n -échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$.
4. On suppose que le modèle est régulier; donner le comportement asymptotique de l'estimateur du maximum de vraisemblance.
5. Donner un test de niveau α consistant pour le problème de test

$$H_0 : \mu_1 = 0 \text{ contre } H_1 : \mu_1 \neq 0$$

Correction de l'exercice 5.3

1. La vraisemblance est

$$L((\mu_1, \mu_2, v), (X_i)_i, (Y_j)_j) = (2\pi v)^{-m/2} \exp\left(-\frac{1}{2v} \sum_{i=1}^m (X_i - \mu_1)^2\right) \times (2\pi v)^{-n/2} \exp\left(-\frac{1}{2v} \sum_{j=1}^n (Y_j - \mu_2)^2\right)$$

2. On voit que le gradient de la log-vraisemblance admet un seul et unique zéro donné par \bar{X}_m

$$\begin{pmatrix} \bar{X}_m \\ \bar{Y}_n \\ \hat{v} \end{pmatrix} \text{ où } \bar{X}_m = \frac{1}{m} \sum_{i=1}^m X_i, \bar{Y}_n = \frac{1}{n} \sum_{j=1}^n Y_j \text{ et } \hat{v} = \frac{1}{n+m} \left[\sum_{i=1}^m (X_i - \bar{X}_m)^2 + \sum_{j=1}^n (Y_j - \bar{Y}_n)^2 \right].$$

De plus la Hessienne de la log-vraisemblance en ce point est telle que

$$\nabla^2 \ell_n(\bar{X}_m, \bar{Y}_n, \hat{v}) = \begin{pmatrix} \frac{-m}{v} & 0 & 0 \\ 0 & \frac{-n}{v} & 0 \\ 0 & 0 & \frac{-(m+n)}{v^2} \end{pmatrix} \prec 0$$

Donc le point $(\bar{X}_m, \bar{Y}_n, \hat{v})$ est un maximum local. Par ailleurs, il n'y a qu'un seul maximum local, c'est donc un maximum global. C'est donc l'EMV.

3. La densité (par rapport à la mesure de Lebesgue sur \mathbb{R}^2) du couple (X, Y) est

$$\begin{aligned} f((\mu_1, \mu_2, v), (x, y)) &= \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x - \mu_1)^2}{2v}\right) \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(y - \mu_2)^2}{2v}\right) \\ &= \frac{1}{2\pi v} \exp\left(-\frac{1}{2v}((x - \mu_1)^2 + (y - \mu_2)^2)\right). \end{aligned}$$

La matrice d'information de Fisher en (μ_1, μ_2, v) pour une observation (X_1, Y_1) est donnée par

$$\begin{aligned} I_1(\mu_1, \mu_2, v) &= \mathbb{E}_{(\mu_1, \mu_2, v)} [\nabla \log f((\mu_1, \mu_2, v), (X, Y)) \nabla \log f((\mu_1, \mu_2, v), (X, Y))^T] \\ &= -\mathbb{E}_{(\mu_1, \mu_2, v)} [\nabla^2 \log f((\mu_1, \mu_2, v), (X, Y))] \\ &= \begin{pmatrix} \frac{1}{v} & 0 & 0 \\ 0 & \frac{1}{v} & 0 \\ 0 & 0 & \frac{1}{v^2} \end{pmatrix} \end{aligned}$$

4. Le modèle étant régulier, l'EMV est asymptotiquement normal de matrice de covariance asymptotique égale à l'inverse de la matrice d'information de Fisher :

$$\sqrt{n} \left(\begin{pmatrix} \bar{X}_m \\ \bar{Y}_n \\ \hat{v} \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \\ v \end{pmatrix} \right) \rightsquigarrow \mathcal{N}_3(0, I_1(\mu_1, \mu_2, v)^{-1}) = \mathcal{N}_3\left(0, \begin{pmatrix} v & 0 & 0 \\ 0 & v & 0 \\ 0 & 0 & v^2 \end{pmatrix}\right)$$

5. On a $\sqrt{n}(\bar{X}_m - \mu_1) \rightsquigarrow \mathcal{N}(0, v)$ et $\hat{v} \xrightarrow{p.s.} v$ alors par le lemme de Slutsky,

$$\frac{\sqrt{n}(\bar{X}_m - \mu_1)}{\sqrt{\hat{v}}} \rightsquigarrow \mathcal{N}(0, 1).$$

On considère le test

$$\varphi_\alpha = \begin{cases} H_0 & \text{si } T_n \leq t_\alpha \\ H_1 & \text{sinon} \end{cases}$$

où $t_\alpha = q_{1-\alpha/2}^{\mathcal{N}(0,1)}$ et

$$T_n = \frac{\sqrt{n}|\bar{X}_m|}{\sqrt{\hat{v}}}.$$

Le test est de niveau asymptotique α car sous H_0 , T_n converge en loi vers $|g|$ où $g \sim \mathcal{N}(0, 1)$ et, il est consistant, car sous H_1 , T_n tends vers $+\infty$ p.s..

Exercice 5.4 (Ceinture de sécurité)

Une enquête sur l'influence de la ceinture de sécurité a donné les résultats suivants : sur 10.779 conducteurs ayant subi un accident l'enquête rapporte les effectifs dans le tableau qui suit selon la gravité et le port au non de la ceinture de sécurité :

nature des blessures	port de la ceinture	pas de ceinture
graves ou fatales	5	141
blessures sérieuses	25	330
peu ou pas de blessures	1229	9049

On souhaite répondre à la question : *la ceinture de sécurité a-t-elle une influence sur la gravité des blessures lors d'un accident ?*

1. Modéliser ces données.
2. Définir un problème de test permettant de répondre à la question.
3. Construire un test de niveau asymptotique $\alpha = 0.05$, consistant pour ce problème.
4. Comparer la p-value de ce test à 0,001. Répondre à la question d'origine et donner un niveau de confiance sur votre décision.

On rappelle les quantiles d'ordre $1 - \alpha$ d'une $\chi^2(2)$:

α	0,999	0,995	0,99	0,98	0,95	0,9	0,8	0,2	0,1
$q_{1-\alpha}^{\chi^2(2)}$	0,0020	0,0100	0,0201	0,0404	0,1026	0,2107	0,4463	3,2189	4,6052
α	0,05	0,02	0,01	0,005	0,001				
$q_{1-\alpha}^{\chi^2(2)}$	5,9915	7,8240	9,2103	10,5966	13,8155				

Correction de l'exercice 5.4

1. On modélise ces données par le modèle d'échantillonnage de n couples $(X_1, Y_1), \dots, (X_n, Y_n)$ où pour tout $1 \leq i \leq n$, X_i correspond à la gravité du i ème accident et Y_i au port ou non de la ceinture :
 - a) $X_i \in \{ \text{graves ou fatales, blessures sérieuses, peu ou pas de blessures} \}$
 - b) $Y_i \in \{ \text{port de la ceinture, pas de ceinture} \}$
2. On veut tester si le port de la ceinture est indépendant de la gravité des blessures de l'accident. On va donc faire un test d'indépendance entre X et Y . On considère le problème de test suivant :

$$H_0 : \text{'X et Y sont indépendantes'} \text{ contre } H_1 : \text{'X et Y ne sont pas indépendantes'}$$

3. On considère le test d'indépendance du χ^2 de niveau asymptotique α :

$$\varphi_\alpha((X_i, Y_i)_i) = \begin{cases} H_0 & \text{si } T_n \leq t_\alpha \\ H_1 & \text{sinon} \end{cases}$$

où $t_\alpha = q_{1-\alpha}^{\chi^2(2)}$ (la degrés vient de $(2 - 1)(3 - 1) = 2$) et $T_n = n\chi^2((\hat{p}_{\ell,\ell'}^{(n)})_{\ell,\ell'}, (\hat{p}_{\bullet,\bullet}^{(n)} \times \hat{p}_{\bullet,\ell'}^{(n)})_{\ell,\ell'})$ où

$$\hat{p}_{\ell,\ell'}^{(n)} = \frac{1}{n} \sum_{i=1}^n I((X_i, Y_i) = (\ell, \ell')), \quad \hat{p}_{\ell,\bullet}^{(n)} = \frac{1}{n} \sum_{i=1}^n I(X_i = \ell) \text{ et } \hat{p}_{\bullet,\ell'}^{(n)} = \frac{1}{n} \sum_{i=1}^n I(Y_i = \ell')$$

pour tout $\ell \in \{ \text{graves ou fatales, blessures sérieuses, peu ou pas de blessures} \}$ et $\ell' \in \{ \text{port de la ceinture, pas de ceinture} \}$. On sait que ce test est consistant (d'après le cours). Par ailleurs, pour $\alpha = 0,05$, on a d'après la table $q_{1-\alpha}^{\chi^2(2)} = 5,99$.

4. On a

$$T_n = \frac{(5 - 17,05)^2}{17,05} + \dots + \frac{(9049 - 9077,52)^2}{9077,52} = 17,81.$$

Alors la p-value du test est plus petite que 0,001, on a va donc rejeter et on a un très haut niveau de confiance en cette décision. On peut alors affirmer que le port de la ceinture de sécurité et la nature des blessures sont dépendants.

6 Rattrapage 2015-2016

Exercice 6.1 (Modèle d'uniforme perturbées)

Soit le modèle d'échantillonnage $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta$ pour $\theta \in]-1, 1[$ où \mathbb{P}_θ est une loi admettant une densité par rapport à la mesure de Lebesgue donnée par

$$f(\theta, x) = \frac{d\mathbb{P}_\theta}{d\lambda}(x) = (1 - \theta)I(-1/2 < x < 0) + (1 + \theta)I(0 < x < 1/2).$$

On pose

$$Y_n = \text{card}\{i : X_i > 0\} = \sum_{i=1}^n I(X_i > 0).$$

a) Préliminaires

1. Donner l'expérience statistique associée à ces données.
2. Calculer $\mathbb{P}_\theta([0, 1/2])$, la moyenne $\mathbb{E}_\theta X_1$ et la variance $\text{Var}(X_1)$.
3. Donner la loi de Y_n , sa moyenne et sa variance.
4. Vérifier que

$$f(\theta, x) = (1 - \theta)^{1 - I(0 < x < 1/2)} (1 + \theta)^{I(0 < x < 1/2)}.$$

En déduire l'expression de la vraisemblance de l'échantillon en θ en fonction de Y_n .

5. Calculer l'information de Fisher sur θ contenue dans un n -échantillon de ce modèle.

b) Estimation de θ

1. Proposer un estimateur des moments de θ en fonction de Y_n .
2. Montrer que l'estimateur du maximum de vraisemblance vaut $\hat{\theta}_n^{\text{mv}} = \frac{2}{n} Y_n - 1$.

3. Etudier les propriétés de $\hat{\theta}_n^{\text{mv}}$: biais, variance, consistance.
4. Comparer le risque quadratique de $\hat{\theta}_n^{\text{mv}}$ et la borne de Cramer-Rao. En déduire que $\hat{\theta}_n^{\text{mv}}$ atteint la borne de Cramer-Rao parmi tous les estimateurs sans biais.
5. Montrer que sous \mathbb{P}_θ , $\sqrt{n}(\hat{\theta}_n^{\text{mv}} - \theta)$ converge en loi vers $\mathcal{N}(0, 1 - \theta^2)$.
6. Etudier le comportement asymptotique de

$$\frac{\sqrt{n}(\hat{\theta}_n^{\text{mv}} - \theta)}{\sqrt{1 - \hat{\theta}_n^{\text{mv}2}}}$$

7. Construire un intervalle de confiance pour θ de niveau asymptotique $\alpha = 0.95$ centré en $\hat{\theta}_n^{\text{mv}}$ et de longueur proportionnelle à $n^{-1/2}$.

c) Tests

1. On considère le problème de test :

$$H_0 : \theta = 0 \text{ contre } H_1 : \theta = 1/2$$

Sous quelles condition existe-t'il un test de Neyman-Pearson de niveau α (on ne considère ici que les tests non randomisés). Dans ce cas, existe-t'il un test de même niveau plus puissant ?

2. Pour le même problème de test, construire un test de niveau asymptotique α . Etudier sa puissance.
3. On considère le problème de test :

$$H_0 : \theta = 0 \text{ contre } H_1 : \theta \neq 0$$

Construire un test de niveau asymptotique α . Etudier sa consistance.

d) Application

On considère un n -échantillon $U_1, \dots, U_n \stackrel{i.i.d.}{\sim} \mathcal{U}([-1/2, 1/2])$. Un phénomène aléatoire perturbe les observations des U_i : pour chaque $i = 1, \dots, n$, la quantité $|U_i|$ est observée avec probabilité $\theta \in [0, 1)$ sinon c'est U_i qui est observée. Ces perturbations sont indépendantes entres elles et indépendantes des U_i . On note X_1, \dots, X_n l'échantillon finalement observé après perturbation.

1. Déterminer la loi de X_1 .
2. Proposer une méthode d'estimation de θ .
3. Construire un test de niveau asymptotique α consistant permettant de décider si un tel phénomène de perturbation s'est produit.
4. La loi des U_i n'étant plus uniforme, que suffit-il de connaître sur elle pour que ce test reste valable ?

- a) 1. L'expérience statistique associée aux données est celle d'un n -échantillon dans un modèle dominé par la mesure de Lebesgue :

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{\mathbb{P}_\theta : \theta \in \Theta\})^{\otimes n}.$$

2.

$$\mathbb{P}_\theta([0, 1/2]) = \frac{1+\theta}{2}, \mathbb{E}_\theta X_1 = (1-\theta) \int_{-1/2}^0 x dx + (1+\theta) \int_0^{1/2} x dx = \frac{\theta}{4} \text{ et } \text{Var}_\theta(X_1) = \frac{1}{12} - \frac{\theta^2}{16}$$

3. Y_n est le nombre de succès dans une expérience de n réalisation d'une binomiale de moyenne $\mathbb{P}[X_1 > 0] = (1+\theta)/2$. C'est donc une multinomiale de paramètre $\mathcal{M}(n, (1+\theta)/2)$ càd pour tout $0 \leq k \leq n$, on a

$$\mathbb{P}[Y_n = k] = \binom{n}{k} \left(\frac{1+\theta}{2}\right)^k \left(\frac{1-\theta}{2}\right)^{n-k}$$

On a aussi $\mathbb{E}_\theta Y_n = n(1+\theta)/2$ et $\text{Var}_\theta Y_n = n(1-\theta^2)/4$.

4. On vérifie directement l'égalité en regardant les cas $-1/2 < x < 0$ et $0 < x < 1/2$. La vraisemblance s'obtient alors comme suit :

$$L(\theta, (X_i)_i) = \prod_{i=1}^n (1-\theta)^{1-I(0 < X_i < 1/2)} (1+\theta)^{I(0 < X_i < 1/2)} = (1-\theta)^n \left(\frac{1+\theta}{1-\theta}\right)^{Y_n}.$$

5. L'information de Fisher d'un n -échantillon est $I_n(\theta) = nI_1(\theta)$ et celle contenue dans une seule donnée est

$$I_1(\theta) = \mathbb{E}_\theta(\partial_\theta \log f(\theta, X))^2$$

où $\log f(\theta, x) = \log(1-\theta)I(-1/2 < x < 0) + \log(1+\theta)I(0 < x < 1/2)$. Alors

$$\partial_\theta \log f(\theta, x) = \frac{-1}{1-\theta}I(-1/2 < x < 0) + \frac{1}{1+\theta}I(0 < x < 1/2).$$

Donc

$$\begin{aligned} I_1(\theta) &= \mathbb{E}_\theta \left[\frac{1}{(1-\theta)^2} I(-1/2 < X < 0) + \frac{1}{(1+\theta)^2} I(0 < X < 1/2) \right] \\ &= \frac{1}{(1-\theta)^2} \mathbb{P}(-1/2 < X < 0) + \frac{1}{(1+\theta)^2} \mathbb{P}(0 < X < 1/2) \\ &= \frac{1}{(1-\theta)^2} \frac{1-\theta}{2} + \frac{1}{(1+\theta)^2} \frac{1+\theta}{2} = \frac{1}{2(1-\theta)} + \frac{1}{2(1+\theta)} = \frac{1}{1-\theta^2}. \end{aligned}$$

On a donc $I_n(\theta) = n/(1-\theta^2)$.

- b) 1. On a $\mathbb{P}_\theta[X_1 > 0] = (1+\theta)/2$. L'estimateur des moments d'ordre 1 est $\hat{\theta}$ tel que $Y_n/n = (1+\hat{\theta})/2$ càd

$$\hat{\theta} = 2 \frac{Y_n}{n} - 1.$$

2. D'après la question 4 de la partie précédente, la log-vraisemblance est

$$\theta \in]-1, 1[\mapsto n \log(1-\theta) + Y_n \log\left(\frac{1+\theta}{1-\theta}\right),$$

elle est maximale en $\hat{\theta}_n^{\text{mv}} = 2(Y_n/n) - 1$.

3. $\mathbb{E}_\theta \hat{\theta}_n^{\text{mv}} = 2\mathbb{P}[X > 0] - 1 = \theta$; donc $\hat{\theta}_n^{\text{mv}}$ est sans biais. La variance de l'estimateur est ici égale à son risque quadratique et on a :

$$\text{Var}_\theta(\hat{\theta}_n^{\text{mv}}) = (4/n) \text{Var}_\theta(I(X > 0)) = (4/n)\mathbb{P}_\theta[X > 0]\mathbb{P}_\theta[X < 0] = (1 - \theta^2)/n.$$

De plus, la loi forte des grands nombres dit que $Y_n \xrightarrow{\text{p.s.}} \mathbb{P}_\theta[X > 0]$ donc $\hat{\theta}_n^{\text{mv}} \xrightarrow{\text{p.s.}} \theta$. Donc $\hat{\theta}_n^{\text{mv}}$ est un estimateur fortement consistant.

4. Si $\hat{\theta}$ est un estimateur sans biais de θ alors la borne de Cramer-Rao dit que son risque quadratique vérifie $R_\theta(\hat{\theta}) \geq I_n(\theta)^{-1} = (1 - \theta^2)/n$. Par ailleurs, le risque quadratique de $\hat{\theta}_n^{\text{mv}}$ vaut aussi $(1 - \theta^2)/n$ (et $\hat{\theta}_n^{\text{mv}}$ est sans biais), donc $\hat{\theta}_n^{\text{mv}}$ atteint bien la borne de Cramer-Rao parmi tous les estimateurs sans biais.
5. On écrit $\hat{\theta}_n^{\text{mv}}$ comme une moyenne empirique : $\hat{\theta}_n^{\text{mv}} = \frac{1}{n} \sum_{i=1}^n (2I(X_i > 0) - 1)$. On applique le TCL :

$$\sqrt{n}(\hat{\theta}_n^{\text{mv}} - \theta) \rightsquigarrow \mathcal{N}(0, \text{Var}_\theta(2I(X > 0) - 1)) = \mathcal{N}(0, 1 - \theta^2)$$

car $\text{Var}_\theta(2I(X > 0) - 1) = 4 \text{Var}_\theta(I(X > 0)) = 4\mathbb{P}_\theta[X > 0]\mathbb{P}_\theta[X < 0] = 1 - \theta^2$.

6. En utilisant le résultat de convergence de la question précédente, la consistance de $\hat{\theta}_n^{\text{mv}}$ et le lemme de Slutsky, on obtient que

$$\frac{\sqrt{n}(\hat{\theta}_n^{\text{mv}} - \theta)}{\sqrt{1 - \hat{\theta}_n^{\text{mv}2}}} \rightsquigarrow \mathcal{N}(0, 1).$$

7. On déduit de la question précédente un intervalle de confiance de niveau asymptotique α : $\mathbb{P}_\theta[\theta \in \mathcal{I}_{n,\alpha}] \rightarrow 1 - \alpha$ où

$$\mathcal{I}_{n,\alpha} = \left[\hat{\theta}_n^{\text{mv}} \pm q_{1-\alpha/2}^{\mathcal{N}(0,1)} \sqrt{\frac{1 - \hat{\theta}_n^{\text{mv}2}}{n}} \right]$$

- c) 1. Le rapport de vraisemblance pour $0 = \theta_0 < \theta_1 = 1/2$ est :

$$\frac{L(\theta_0, (X_i)_i)}{L(\theta_1, (X_i)_i)} = \left(\frac{1 - \theta_0}{1 - \theta_1} \right)^n \left(\frac{(1 + \theta_0)(1 - \theta_1)}{(1 - \theta_0)(1 + \theta_1)} \right)^{Y_n}.$$

Comme $\theta_0 < \theta_1$, le rapport de vraisemblance est une fonction croissante de Y_n , le test de Neyman-Pearson de niveau α est donc de la forme

$$\varphi_\alpha = \begin{cases} H_0 & \text{si } Y_n/n \leq t_\alpha \\ H_1 & \text{sinon} \end{cases}$$

Ce test sera exactement de niveau α s'il existe t_α tel que

$$\mathbb{P}_{\theta_0}[Y_n/n > t_\alpha] = \alpha.$$

Cette condition n'est pas toujours satisfaite car Y_n est une variable aléatoire discrète.

2. D'après le TCL, sous H_0 , $\sqrt{n}(Y_n/n - 1/2) \rightsquigarrow \mathcal{N}(0, 1/4)$. Il suffit alors de fixer le seuil dans le test précédent tel que $2\sqrt{n}(t_\alpha - 1/2) = q_{1-\alpha}^{\mathcal{N}(0,1)}$ pour avoir un test de niveau asymptotique α . Sous H_1 , $Y_n/n \xrightarrow{\text{p.s.}} 3/4$, on en déduit que la puissance du test tend vers 1. Donc le test est consistant.

3. On considère

$$\varphi_\alpha = \begin{cases} H_0 & \text{si } \sqrt{n}|\hat{\theta}_n^{mv}| \leq t_\alpha \\ H_1 & \text{sinon} \end{cases}$$

où $t_\alpha = q_{1-\alpha/2}^{\mathcal{N}(0,1)}$. Sous H_0 , $\sqrt{n}|\hat{\theta}_n^{mv}| \rightsquigarrow |g|$ où $g \sim \mathcal{N}(0,1)$ et sous H_1 , $\sqrt{n}|\hat{\theta}_n^{mv}| \xrightarrow{p.s.} +\infty$.
Donc le test est de niveau asymptotique α et il est consistant.

d) 1. On note δ_i la variable aléatoire indiquant la présence de perturbation dans l'observation i , c'ad $\delta_i = 1$ avec probabilité θ et 0 sinon. Par hypothèse les δ_i sont i.i.d. $\mathcal{B}(\theta)$ et indépendantes de U_i . La loi de X_1 est donnée par : si f est une fonction continue à support compact sur \mathbb{R} alors

$$\begin{aligned} \mathbb{E}f(X_1) &= \mathbb{E}f(|U_i|)\delta_i + f(U_i)(1 - \delta_i) = \mathbb{E}f(|U_i|)\theta + f(U_i)(1 - \theta) = \int_{-1/2}^{1/2} [f(|u|)\theta + f(u)(1 - \theta)] du \\ &= 2\theta \int_0^{1/2} f(u)du + \int_{-1/2}^{1/2} f(u)(1 - \theta)du = (1 - \theta) \int_{-1/2}^0 f(u)du + (1 + \theta) \int_0^{1/2} f(u)du. \end{aligned}$$

Donc X_1 est une variable admettant $f(\theta, \cdot)$ pour densité.

2. On va alors estimer θ par $\hat{\theta}_n^{mv} = 2Y_n/n - 1$

3. On va considérer le problème de test de la question 3 de la partie c). On décide donc qu'il y aura perturbation si $\sqrt{n}|\hat{\theta}_n^{mv}| > q_{1-\alpha/2}^{\mathcal{N}(0,1)}$.

4. Le test ne dépend que de $\mathbb{P}[X > 0]$. Donc pour toute loi telle que $\mathbb{P}_\theta[X > 0] = (1 + \theta)/2$, on aura les mêmes résultats.

7 Examen du lundi 14 novembre 2016

Exercice 7.1 (Loi géométrique)

Soit X_1, \dots, X_n un n échantillon de variables aléatoires indépendantes de loi \mathbb{P}_θ pour $\theta \in (0, 1)$, telle que pour tout entier k , on a $\mathbb{P}_\theta[\{k\}] = \theta(1 - \theta)^k$.

1. Calculer $\mathbb{E}X_1$ et déterminer un estimateur de θ par la méthode des moments.
2. Déterminer l'estimateur du maximum de vraisemblance.
3. Calculer l'information de Fisher en θ contenue dans un n -échantillon de ce modèle.
4. On admettra que le modèle est régulier ; établir la normalité asymptotique de l'estimateur du maximum de vraisemblance.

Correction de l'exercice 7.1

1. On a

$$\mathbb{E}X = \sum_{k \geq 0} k\theta(1 - \theta)^k = \theta(1 - \theta) \sum_{k \geq 1} k(1 - \theta)^{k-1} = \frac{1 - \theta}{\theta}.$$

Un estimateur des moments est donc $\tilde{\theta}$ tel que $\overline{X_n} = (1 - \tilde{\theta})/\tilde{\theta}$ c'ad

$$\tilde{\theta} = \frac{1}{1 + \overline{X_n}}.$$

2. La log-vraisemblance du modèle est la fonction

$$\theta \in (0, 1) \longrightarrow \ell_n(\theta, X_1, \dots, X_n) = n \log \theta + \log(1 - \theta) \sum_{i=1}^n X_i.$$

Elle admet pour dérivée

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

qui s'annule uniquement en

$$\hat{\theta} = \frac{1}{1 + \bar{X}_n}$$

et pour lequel on vérifie que $\partial_\theta^2 \ell_n(\hat{\theta}, X_1, \dots, X_n) < 0$. Donc $\hat{\theta}$ est l'EMV.

3. On a pour tout $\theta \in (0, 1)$, l'information de Fisher en θ d'un n -échantillon est

$$I(\theta) = -\mathbb{E}_\theta \partial_\theta^2 \ell_n(\theta, X_1, \dots, X_n) = \frac{n}{\theta^2} + \frac{n \mathbb{E}_\theta X_1}{(1 - \theta)^2} = \frac{n}{\theta^2} + \frac{n}{\theta(1 - \theta)}$$

4. Comme le modèle est régulier, on a

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \theta^2(1 - \theta)).$$

Exercice 7.2 (Test de comparaison de moyenne)

On observe n_1 variables aléatoires i.i.d. X_1, \dots, X_{n_1} de loi $\mathcal{N}(\mu, \sigma^2)$ et n_2 variables aléatoires i.i.d. Y_1, \dots, Y_{n_2} de loi $\mathcal{N}(\mu + \Delta, \sigma^2)$ indépendantes des X_i . On souhaite tester $\Delta = 0$ contre $\Delta \neq 0$; les deux paramètres μ et σ^2 étant inconnus.

1. Quelle est la loi de $\bar{X}_{n_1} - \bar{Y}_{n_2}$ où

$$\bar{X}_{n_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i \text{ et } \bar{Y}_{n_2} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i.$$

2. On note

$$\hat{\sigma}_{X, n_1}^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X}_{n_1})^2 \text{ et } \hat{\sigma}_{Y, n_2}^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y}_{n_2})^2$$

les variances empiriques (débiaisées) associées respectivement à chacun des deux échantillons. Donner la loi de

$$U^2 = (n_1 - 1)\hat{\sigma}_{X, n_1}^2 + (n_2 - 1)\hat{\sigma}_{Y, n_2}^2.$$

3. On introduit

$$T = \frac{\sqrt{n_1 + n_2 - 2}Z}{U/\sigma} \text{ où } Z = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} + \Delta}{\sigma \sqrt{n_1^{-1} + n_2^{-1}}}.$$

Montrer que la loi de T est une loi de Student à $n_1 + n_2 - 2$ degrés de liberté.

4. En déduire la construction d'un test de niveau α pour le problème de test

$$H_0 : \Delta = 0 \text{ contre } H_1 : \Delta \neq 0.$$

5. On observe pour $n_1 = 13$ et $n_2 = 14$, les valeurs

$$\overline{X}_{n_1} = 25.97, \quad \hat{\sigma}_{X,n_1}^2 = 1.36, \quad \overline{Y}_{n_2} = 25.38, \quad \text{and } \hat{\sigma}_{Y,n_2}^2 = 1.77.$$

Calculer la p -value du test. Faut-il rejeter aux niveaux usuels 1%, 5% et 10%.

On rappelle quelques valeurs de la fonction de répartition d'une loi de Student à 25 degrés de liberté. On note $F(t) = \mathbb{P}[X \leq t]$ quand X suit une loi de Student à 25 degrés de liberté.

t	0,9	1.0125	1.125	1.2375	1.35	1.4625	1.575	1.6875	1.8
$F(t)$	0.811	0.839	0.864	0.886	0.905	0.921	0.936	0.948	0.958

Correction de l'exercice 7.2

1. La variable aléatoire $\overline{X}_{n_1} - \overline{Y}_{n_2}$ est une combinaison linéaire de variables Gaussiennes, c'est donc aussi une variable Gaussienne. Pour la caractériser, il suffit de donner sa moyenne et sa variance.

On a :

$$\mathbb{E}[\overline{X}_{n_1} - \overline{Y}_{n_2}] = -\Delta \text{ et } \text{var}(\overline{X}_{n_1} - \overline{Y}_{n_2}) = \text{var}(\overline{X}_{n_1}) + \text{var}(\overline{Y}_{n_2}) = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right).$$

On en déduit que $\overline{X}_{n_1} - \overline{Y}_{n_2} \sim \mathcal{N}(-\Delta, \sigma^2(n_1^{-1} + n_2^{-1}))$.

2. On peut voir U^2 comme étant la norme Euclidienne au carré de la projection du vecteur aléatoire Gaussien $G := (X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2})^\top$ sur l'espace orthogonal à $V := \text{vect}(v_1, v_2)$ où

$$v_1 = \frac{1}{\sqrt{n_1}} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{n_1+n_2} \text{ et } v_2 = \frac{1}{\sqrt{n_2}} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^{n_1+n_2}. \tag{10}$$

En effet, v_1 et v_2 sont deux vecteurs unitaires orthogonaux alors

$$P_V G = \langle G, v_1 \rangle v_1 + \langle G, v_2 \rangle v_2 = \begin{pmatrix} \overline{X}_{n_1} \\ \vdots \\ \overline{X}_{n_1} \\ \overline{Y}_{n_2} \\ \vdots \\ \overline{Y}_{n_2} \end{pmatrix}$$

et donc $P_{V^\perp} G = G - P_V G = (X_1 - \overline{X}_{n_1}, \dots, X_{n_1} - \overline{X}_{n_1}, Y_1 - \overline{Y}_{n_2}, \dots, Y_{n_2} - \overline{Y}_{n_2})$. On a donc bien $U^2 = \|P_{V^\perp} G\|_2^2$. Par ailleurs, V^\perp est un espace vectoriel de dimension $n_1 + n_2 - 2$ donc, d'après le théorème de Cochran, U^2/σ^2 est distribuée selon une χ_2 à $n_1 + n_2 - 2$ degrés de liberté.

3. On peut aussi voir U^2 comme étant la norme Euclidienne au carré de la projection du vecteur aléatoire Gaussien $G' := (X_1, \dots, X_{n_1}, -Y_1 + \Delta, \dots, -Y_{n_2} + \Delta)^\top$ sur l'espace orthogonal à $V = \text{vect}(v_1, v_2)$ (défini dans (10)). On en déduit par le Théorème de Cochran que :

- $P_V G'$ et $P_{V^\perp} G'$ sont indépendantes,
- $U^2/\sigma^2 = \|G'\|_2^2/\sigma^2$ est distribuée selon une χ_2 à $n_1 + n_2 - 2$ degrés de liberté.

Par ailleurs,

$$Z = \frac{\langle P_V G', e_1 \rangle + \langle P_{V^\perp} G', e_{n_1+1} \rangle}{\sigma \sqrt{n_1^{-1} + n_2^{-1}}}$$

donc Z est indépendant de U^2 et c'est un vecteur Gaussien (en tant que combinaison linéaire de Gaussiennes) de moyenne nulle et de variance 1 (cf. Question 1). On en déduit que $\sqrt{n_1 + n_2 - 2}Z/(U/\sigma)$ suit une loi de Student à $n_1 + n_2 - 2$ degrés de liberté.

4. On considère la statistique de test

$$T = \sqrt{\frac{n_1 + n_2 - 2}{n_1^{-1} + n_2^{-1}}} \frac{(\bar{X}_{n_1} - \bar{Y}_{n_2})}{\sqrt{(n_1 - 1)\hat{\sigma}_{X,n_1}^2 + (n_2 - 1)\hat{\sigma}_{Y,n_2}^2}}.$$

Sous H_0 , T suit une loi de student à $n_1 + n_2 - 2$ degrés de liberté. On construit donc un test de niveau α avec

$$\varphi_\alpha((X_i)_i, (Y_i)_i) = \begin{cases} H_0 & \text{quand } |T| \leq t_\alpha \\ H_1 & \text{sinon} \end{cases}$$

où t_α est le quantile d'ordre $1 - \alpha/2$ d'une loi de Student à $n_1 + n_2 - 2$ degrés de liberté.

5. La p -value du test est le plus petit α pour lequel le test rejette. Ici la p -value est donnée par $\hat{\alpha} \in (0, 1)$ tel que $t_{\hat{\alpha}} = T$. Comme

$$T = \sqrt{\frac{13 + 14 - 2}{1/13 + 1/14}} \frac{25.97 - 25.38}{\sqrt{12 \times 1.36 + 13 \times 1.77}} = 1.222,$$

on cherche $\hat{\alpha}$ tel que $q_{1-\hat{\alpha}/2}^{S(25)} = 1.222$. On obtient $\hat{\alpha} \approx 0.24$. En particulier, pour les niveaux 1%, 5% et 10%, le test va accepter.

Exercice 7.3 (Variables uniformes)

Soient X_1, \dots, X_n des variables i.i.d. uniformes sur $[a, b]$ où a et b sont des paramètres inconnus tels que $a < b$. On note par $\mathbb{P}_{(a,b)}$ la loi des observations.

1. Estimateur par la méthode des moments.

- (a) Soit X une variable de loi uniforme sur $[a, b]$. Calculer $\mathbb{E}X$ et $\mathbb{E}X^2$ en fonction de a et b .
- (b) Proposer un estimateur (\hat{a}, \hat{b}) par la méthode des moments.
- (c) Etudier la consistance de (\hat{a}, \hat{b}) .
- (d) Etudier la normalité asymptotique de \hat{a} .

(e) Etudier la normalité asymptotique de (\hat{a}, \hat{b}) . Cette question étant très calculatoire, on ne s'intéressera qu'à la démarche à suivre plus qu'au calcul effectif de la matrice de covariance asymptotique. Pour les plus courageux, on ne demande de calculer que la forme de la matrice de covariance asymptotique et de préciser uniquement l'élément en haut à gauche de cette matrice.

2. Estimateur par maximum de vraisemblance.

- (a) Donner l'estimateur (\tilde{a}, \tilde{b}) maximisant la vraisemblance en (a, b) .
- (b) Soit $\epsilon > 0$. Majorer $\mathbb{P}_{(a,b)}[|\tilde{a} - a| \geq \epsilon]$. En déduire la consistance de \tilde{a} .
- (c) Etudier la normalité asymptotique de \tilde{a} .

3. Comparer les estimateurs \hat{a} et \tilde{a} de a .

Correction de l'exercice 7.3

1. Estimateur par la méthode des moments.

- (a) Soit X une variable uniformément distribuée sur $[a, b]$. On a $\mathbb{E}X = (a + b)/2$ et $\mathbb{E}X^2 = (b^2 + ab + a^2)/3$.
- (b) Un estimateur des moments de (a, b) est donné par (\hat{a}, \hat{b}) solution de

$$\begin{cases} \overline{X}_n := \frac{1}{n} \sum_{i=1}^n X_i &= \frac{\hat{b} + \hat{a}}{2} \\ \overline{X}_n^2 := \frac{1}{n} \sum_{i=1}^n X_i^2 &= \frac{\hat{b}^2 + \hat{a}\hat{b} + \hat{a}^2}{3}. \end{cases}$$

C'est-à-dire

$$\begin{cases} \hat{a} &= \overline{X}_n - \sqrt{3(\overline{X}_n^2 - (\overline{X}_n)^2)} = \overline{X}_n - \sqrt{3}\hat{\sigma}_n \\ \hat{b} &= \overline{X}_n + \sqrt{3(\overline{X}_n^2 - (\overline{X}_n)^2)} = \overline{X}_n + \sqrt{3}\hat{\sigma}_n. \end{cases}$$

où $\hat{\sigma}_n^2 = \overline{X}_n^2 - (\overline{X}_n)^2$.

- (c) Par la LFGN, $(\overline{X}_n)_n$ converge p.s. vers $\mathbb{E}X$ et $(\overline{X}_n^2)_n$ converge p.s. vers $\mathbb{E}X^2$. Par le "continuous map theorem", on en déduit que (\hat{a}, \hat{b}) converge p.s. vers (a, b) quand n tend vers $+\infty$. Donc (\hat{a}, \hat{b}) est un estimateur fortement consistant de (a, b) .
- (d) Par le TCL (en dimension 2), on voit que $(\overline{X}_n, \overline{X}_n^2)_n$ converge en loi vers une Gaussienne $\mathcal{N}(0, \Sigma)$ où

$$\begin{aligned} \Sigma &= \begin{pmatrix} \text{var}(X) & \text{cov}(X, X^2) \\ \text{cov}(X, X^2) & \text{var}(X^2) \end{pmatrix} = \begin{pmatrix} \mathbb{E}X^2 - (\mathbb{E}X)^2 & \mathbb{E}X^3 - \mathbb{E}X\mathbb{E}X^2 \\ \mathbb{E}X^3 - \mathbb{E}X\mathbb{E}X^2 & \mathbb{E}X^4 - (\mathbb{E}X^2)^2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{(a-b)^2}{12} & \frac{b^3 - b^2a - ba^2 + a^3}{12} \\ \frac{b^3 - b^2a - ba^2 + a^3}{12} & \frac{4b^4 - b^3a - 6b^2a^2 - ba^3 + 4a^4}{45} \end{pmatrix} \end{aligned}$$

On considère la fonction

$$\phi : \begin{cases} \{(x, y) \in \mathbb{R}^2 : y > x^2\} & \longrightarrow \mathbb{R} \\ \begin{pmatrix} x \\ y \end{pmatrix} & \longrightarrow x - \sqrt{3(y - x^2)} \end{cases}$$

Comme

$$\hat{a} = \phi \left(\begin{array}{c} \overline{X_n} \\ \overline{X_n^2} \end{array} \right),$$

la méthode Delta permet d'obtenir

$$\sqrt{n}(\hat{a} - a) = \sqrt{n} \left(\phi \left(\begin{array}{c} \overline{X_n} \\ \overline{X_n^2} \end{array} \right) - \phi \left(\begin{array}{c} \mathbb{E}X \\ \mathbb{E}X^2 \end{array} \right) \right) \xrightarrow{d} \mathcal{N}(0, \bar{\Sigma})$$

où

$$\bar{\Sigma} = \nabla \phi \left(\begin{array}{c} \mathbb{E}X \\ \mathbb{E}X^2 \end{array} \right)^\top \Sigma \nabla \phi \left(\begin{array}{c} \mathbb{E}X \\ \mathbb{E}X^2 \end{array} \right)$$

et

$$\nabla \phi \left(\begin{array}{c} x \\ y \end{array} \right) = \frac{1}{\sqrt{y-x^2}} \left(\begin{array}{c} \sqrt{y-x^2} + \sqrt{3}x \\ -\sqrt{3}/2 \end{array} \right).$$

On pose $\sigma^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2 = (b-a)^2/12$. On a

$$\nabla \phi \left(\begin{array}{c} \mathbb{E}X \\ \mathbb{E}X^2 \end{array} \right) = \frac{1}{2\sigma} \left(\begin{array}{c} 2\sigma + \sqrt{3}(a+b) \\ -\sqrt{3} \end{array} \right) = \frac{1}{b-a} \left(\begin{array}{c} 2a+4b \\ -3 \end{array} \right).$$

On obtient donc

$$\begin{aligned} \bar{\Sigma} &= \frac{1}{(b-a)^2} \left(\begin{array}{cc} 2a+4b & -3 \end{array} \right) \times \left(\begin{array}{cc} \frac{(a-b)^2}{12} & \frac{b^3-b^2a-ba^2+a^3}{12} \\ \frac{b^3-b^2a-ba^2+a^3}{12} & \frac{4b^4-b^3a-6b^2a^2-ba^3+4a^4}{45} \end{array} \right) \\ &\times \left(\begin{array}{c} 2a+4b \\ -3 \end{array} \right) = \frac{2}{15(b-a)^2} (b^4 - 4b^3a + 6b^2a^2 - 4ba^3 + a^4) \end{aligned}$$

(e) Par le TCL (en dimension 2), on voit que $(\overline{X_n}, \overline{X_n^2})_n$ converge en loi vers une Gaussienne $\mathcal{N}(0, \Sigma)$ où

$$\begin{aligned} \Sigma &= \left(\begin{array}{cc} \text{var}(X) & \text{cov}(X, X^2) \\ \text{cov}(X, X^2) & \text{var}(X^2) \end{array} \right) = \left(\begin{array}{cc} \mathbb{E}X^2 - (\mathbb{E}X)^2 & \mathbb{E}X^3 - \mathbb{E}X\mathbb{E}X^2 \\ \mathbb{E}X^3 - \mathbb{E}X\mathbb{E}X^2 & \mathbb{E}X^4 - (\mathbb{E}X^2)^2 \end{array} \right) \\ &= \left(\begin{array}{cc} \frac{(a-b)^2}{12} & \frac{b^3-b^2a-ba^2+a^3}{12} \\ \frac{b^3-b^2a-ba^2+a^3}{12} & \frac{4b^4-b^3a-6b^2a^2-ba^3+4a^4}{45} \end{array} \right) \end{aligned}$$

On considère la fonction

$$\Phi : \left\{ \begin{array}{l} \{(x, y) \in \mathbb{R}^2 : y > x^2\} \longrightarrow \mathbb{R}^2 \\ \left(\begin{array}{c} x \\ y \end{array} \right) \longrightarrow \left(\begin{array}{c} x - \sqrt{3(y-x^2)} \\ x + \sqrt{3(y-x^2)} \end{array} \right) \end{array} \right.$$

Comme

$$\left(\begin{array}{c} \hat{a} \\ \hat{b} \end{array} \right) = \Phi \left(\begin{array}{c} \overline{X_n} \\ \overline{X_n^2} \end{array} \right),$$

la méthode Delta permet d'obtenir

$$\sqrt{n} \left(\left(\begin{array}{c} \hat{a} \\ \hat{b} \end{array} \right) - \left(\begin{array}{c} a \\ b \end{array} \right) \right) = \sqrt{n} \left(\Phi \left(\begin{array}{c} \overline{X_n} \\ \overline{X_n^2} \end{array} \right) - \Phi \left(\begin{array}{c} \mathbb{E}X \\ \mathbb{E}X^2 \end{array} \right) \right) \xrightarrow{d} \mathcal{N}(0, \bar{\Sigma})$$

où

$$\tilde{\Sigma} = \nabla\Phi \begin{pmatrix} \mathbb{E}X \\ \mathbb{E}X^2 \end{pmatrix}^\top \Sigma \nabla\Phi \begin{pmatrix} \mathbb{E}X \\ \mathbb{E}X^2 \end{pmatrix}$$

et

$$\nabla\Phi \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{y-x^2}} \begin{pmatrix} \sqrt{y-x^2} + \sqrt{3}x & \sqrt{y-x^2} - \sqrt{3}x \\ -\sqrt{3}/2 & \sqrt{3}/2 \end{pmatrix}.$$

On pose $\sigma^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2 = (b-a)^2/12$. On a

$$\nabla\Phi \begin{pmatrix} \mathbb{E}X \\ \mathbb{E}X^2 \end{pmatrix} = \frac{1}{\sigma} \begin{pmatrix} \sigma + \sqrt{3}(a+b) & \sigma - \sqrt{3}(a+b) \\ -\sqrt{3} & \sqrt{3} \end{pmatrix} = \frac{1}{b-a} \begin{pmatrix} 2a+4b & -4a-2b \\ -3 & 3 \end{pmatrix}.$$

On obtient donc

$$\begin{aligned} \tilde{\Sigma} &= \frac{1}{(b-a)^2} \begin{pmatrix} 2a+4b & -3 \\ -4a-2b & 3 \end{pmatrix} \times \begin{pmatrix} \frac{(a-b)^2}{12} & \frac{b^3-b^2a-ba^2+a^3}{12} \\ \frac{b^3-b^2a-ba^2+a^3}{12} & \frac{4b^4-b^3a-6b^2a^2-ba^3+4a^4}{45} \end{pmatrix} \\ &\times \begin{pmatrix} 2a+4b & -4a-2b \\ -6 & 6 \end{pmatrix} = \frac{2}{15(b-a)^2} \begin{pmatrix} b^4-4b^3a+6b^2a^2-4ba^3+a^4 & \star \\ \star & \star \end{pmatrix}. \end{aligned}$$

2. Estimateur par maximum de vraisemblance.

(a) La vraisemblance du modèle d'échantillonnage à n observations est

$$\mathcal{L}_n \left(\begin{pmatrix} a \\ b \end{pmatrix}, X_1, \dots, X_n \right) = \prod_{i=1}^n \frac{I(a \leq X_i \leq b)}{b-a} = \frac{1}{(b-a)^n} I(a \leq \min(X_i)) I(\max X_i \leq b).$$

La vraisemblance est donc maximale pour

$$\tilde{a} = \min_{1 \leq i \leq n} X_i \text{ et } \tilde{b} = \max_{1 \leq i \leq n} X_i$$

et donc l'EMV est ici $(\tilde{a}, \tilde{b}) = (\min_{1 \leq i \leq n} X_i, \max_{1 \leq i \leq n} X_i)$.

(b) Soit $0 < \epsilon < b-a$. On a

$$\begin{aligned} \mathbb{P}_{(a,b)}[|\tilde{a}-a| \geq \epsilon] &= \mathbb{P}_{(a,b)}[\min_{1 \leq i \leq n} X_i \geq a + \epsilon] = \prod_{i=1}^n \mathbb{P}_{(a,b)}[X_i \geq a + \epsilon] = (\mathbb{P}_{(a,b)}[X_1 \geq a + \epsilon])^n \\ &= \left(\frac{1}{b-a} \int_a^b I(x \geq a + \epsilon) dx \right)^n = \left(\frac{b-(a+\epsilon)}{b-a} \right)^n = \left(1 - \frac{\epsilon}{b-a} \right)^n. \end{aligned}$$

On en déduit que \tilde{a} est un estimateur consistant de a .

(c) On sait que (Z_n) tend en loi vers Z si et seulement si la suite des fonctions de répartition des Z_n converge simplement vers la fonction de répartition de Z en tout point de continuité de cette dernière. On voit ici que la suite des fonctions de répartition de $(n(\tilde{a}-a))_n$ tend vers la fonction de répartition d'une loi exponentielle de paramètre $1/(b-a)$ càd

$$\lim_{n \rightarrow +\infty} \mathbb{P}_{(a,b)}[n(\tilde{a}-a) \leq t] = \begin{cases} 0 & \text{si } t \leq 0 \\ 1 - \exp\left(\frac{-t}{b-a}\right) & \text{quand } t > 0 \end{cases}$$

donc

$$n(\tilde{a}-a) \xrightarrow{d} \mathcal{E}(1/(b-a))$$

où $\mathcal{E}(1/(b-a))$ est une loi exponentielle de paramètre $1/(b-a)$.

3. la vitesse de convergence de \hat{a} est en $1/\sqrt{n}$ alors que celle de \tilde{a} est en $1/n$. Ce dernier estimateur est donc préférable à \hat{a} .

8 Rattrapage 2016-2017

Exercice 8.1 (Test dans une urne)

Nous disposons d'une urne contenant 5 boules. Ces boules peuvent être rouges ou blanches. On souhaite tester si les boules de l'urne sont toutes blanches ou toutes rouges. On note par r le nombre de boules rouges dans l'urne ; on a donc $5 - r$ boules blanches. On considère alors le problème de test

$$H_0 : r \in \{0, 5\} \text{ contre } H_1 : r \in \{1, 2, 3, 4\}.$$

On tire deux boules de l'urne (avec ou sans remise) et on prend la décision suivante :

1. on accepte si les deux boules ont la même couleur
2. on rejette sinon.

On propose dans la suite d'étudier les propriétés de ce test dans les deux cas d'avec et sans remise.

1. Quel est le risque de première espèce du test précédent dans les deux cas d'avec et sans remise ?
2. quelle est le puissance du test quand le tirage se fait avec remise ?
3. quelle est le puissance du test quand le tirage se fait sans remise ?

Correction de l'exercice 8.1

1. Sous H_0 , toutes le boules ont la même couleur. Alors si on effectue deux tirages avec ou sans remise, les deux boules tirées seront de la même couleur et donc l'hypothèse H_0 est acceptée. En particulier, la probabilité de rejeter à tort est nulle. Donc le niveau du test est dans les deux cas (avec ou sans remise) de zéro.
2. On rappelle que la puissance d'un test est la fonction qui à tout paramètre r dans l'alternative associe la probabilité de rejeter à raison : $r \in \{1, 2, 3, 4\} \rightarrow \mathbb{P}_r[\text{test} = H_1]$.

On considère dans cette question un tirage avec remise.

Soit $r \in \{1, 2, 3, 4\}$ (un paramètre dans l'alternative). La probabilité de rejeter sous \mathbb{P}_r est

$$\mathbb{P}_r[\text{test} = H_1] = \mathbb{P}_r[\text{tirer deux boules de couleur différentes}] = 2\left(\frac{r}{5} \times \frac{5-r}{5}\right) = \frac{2r(5-r)}{25}.$$

La puissance du test est donc la fonction

$$r \in \{1, 2, 3, 4\} \mapsto \frac{2r(5-r)}{25}$$

3. On considère maintenant le problème sans remise. Soit $r \in \{1, 2, 3, 4\}$, on a

$$\mathbb{P}_r[\text{test} = H_1] = \mathbb{P}_r[\text{tirer deux boules de couleur différentes}] = \left(\frac{r}{5} \times \frac{5-r}{4}\right) + \left(\frac{5-r}{5} \times \frac{r}{4}\right) = \frac{r(5-r)}{10}.$$

La puissance du test est donc la fonction

$$r \in \{1, 2, 3, 4\} \mapsto \frac{r(5-r)}{10}.$$

Exercice 8.2 (Paramètre vectoriel - vitesses de convergence différentes)

Soient X_1, \dots, X_n des variables aléatoires i.i.d. de loi exponentielle translatée dont la densité est de la forme :

$$f(x, \theta, \alpha) = \frac{1}{\theta} \exp\left[-\frac{(x-\alpha)}{\theta}\right] I_{[\alpha, +\infty[}(x),$$

où $\theta > 0$ et $\alpha \in \mathbb{R}$ sont deux paramètres inconnus.

1. Donner les estimateurs du maximum de vraisemblance $(\hat{\alpha}_n, \hat{\theta}_n)$ du paramètre (bidimensionnel) (α, θ) .
2. Quelle est la loi de $X_i - \alpha$? Calculer la loi (exacte) de $n(\hat{\alpha}_n - \alpha)$.
3. Déterminer la loi limite de $\sqrt{n}(\hat{\theta}_n - \theta)$.
4. L'objectif de cette question est de montrer que $\hat{\alpha}_n$ et $\hat{\theta}_n$ sont indépendants.
 - (a) On rappelle que $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)}$ sont les statistiques d'ordres de l'échantillon. Monter que

$$(nX_{(1)}, (n-1)(X_{(2)} - X_{(1)}), \dots, 2(X_{(n-1)} - X_{(n-2)}), X_{(n)} - X_{(n-1)})$$

est un vecteur de \mathbb{R}^n ayant pour densité par rapport à la mesure de Lebesgue

$$y = (y_i)_{i=1}^n \in \mathbb{R}^n \rightarrow \exp\left(-\sum_{i=1}^n y_i\right) I(0 < y_i : i = 1, \dots, n).$$

- (b) En déduire que $\hat{\alpha}_n$ et $\hat{\theta}_n$ sont indépendants pour tout n .

Correction de l'exercice 8.2

1. La fonction de vraisemblance est donnée pour tout $\theta > 0$ et $\alpha \in \mathbb{R}$ par

$$V(\theta, \alpha) = \prod_{i=1}^n f(X_i, \theta, \alpha) = \frac{1}{\theta^n} \exp\left[-\sum_{i=1}^n \frac{(X_i - \alpha)}{\theta}\right] I(\alpha \leq \min_i X_i).$$

On voit déjà que l'EMV pour α est

$$\hat{\alpha}_n = \min_{1 \leq i \leq n} X_i.$$

On en déduit par dérivation de $\theta \mapsto \log V(\theta, \hat{\alpha}_n) = -n \log \theta - \sum_i X_i/\theta + n\hat{\alpha}_n/\theta$ que l'EMV pour θ est

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i - \hat{\alpha}_n = \bar{X}_n - \min_i X_i.$$

2. On voit que X suit une loi exponentielle de paramètre θ et de translatée α si et seulement si sa fonction de répartition F_X est donnée par

$$F_X(t) = \left[1 - \exp\left(\frac{-(t - \alpha)}{\theta}\right)\right] I(t \geq \alpha).$$

Par indépendance, on a

$$\mathbb{P}[\hat{\alpha}_n \geq t] = (\mathbb{P}[X_1 \geq t])^n = \exp\left(-\frac{n(t - \alpha)}{\theta}\right) I(t \geq \alpha).$$

Alors $\hat{\alpha}_n$ suit une loi expo de paramètre θ/n et translatée α . Donc $n(\hat{\alpha}_n - \alpha)$ est une loi expo de paramètre θ (et translatée nulle).

3. Si $X \sim f$ alors $\mathbb{E}X = \theta + \alpha$ et $\mathbb{E}X^2 = \alpha^2 + 2\theta\alpha + 2\theta^2$. On a

$$\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}(\bar{X}_n - (\theta + \alpha)) - \left(\frac{1}{\sqrt{n}}\right) \left[n(\min_i X_i - \alpha)\right].$$

On voit que $(n^{-1/2}[n(\min_i X_i - \alpha)])$ converge en probabilité vers 0 et par le TCL ($\sqrt{n}(\bar{X}_n - (\theta + \alpha))$) converge en loi vers $\mathcal{N}(0, \text{var}(X_1))$ où $\text{var}(X_1) = \theta^2$. Par Slutsky, on en déduit que

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, \theta^2).$$

4. Sans perte de généralité, on peut supposer les X_i centrés et réduits, càd $\alpha = 0$ et $\theta = 1$.

(a) On note \mathcal{S}_n l'ensemble de toutes les permutations de $\{1, \dots, n\}$. Soit B un borélien de \mathbb{R}^n .

On a

$$\begin{aligned} & \mathbb{P}[(nX_{(1)}, (n-1)(X_{(2)} - X_{(1)}), \dots, 2(X_{(n-1)} - X_{(n-2)}), X_{(n)} - X_{(n-1)}) \in B] \\ &= \mathbb{P}[\exists \tau \in \mathcal{S}_n : (nX_{\tau(1)}, \dots, X_{\tau(n)} - X_{\tau(n-1)}) \in B \text{ et } X_{\tau(1)} < \dots < X_{\tau(n)}] \\ &= \sum_{\tau \in \mathcal{S}_n} \mathbb{P}[(nX_{\tau(1)}, \dots, X_{\tau(n)} - X_{\tau(n-1)}) \in B \text{ et } X_{\tau(1)} < \dots < X_{\tau(n)}] \\ &= n! \mathbb{P}[(nX_1, \dots, X_n - X_{n-1}) \in B \text{ et } X_1 < \dots < X_n] \\ &= n! \int_{\mathbb{R}^n} I((nx_1, \dots, x_n - x_{n-1}) \in B) I(0 < x_1 < \dots < x_n) \exp\left(-\sum_{i=1}^n x_i\right) dx_1 \cdots dx_n. \end{aligned}$$

On considère le changement de variable

$$\Phi : \begin{cases} \{(x_1, \dots, x_n) : 0 < x_1 < \dots < x_n\} & \longrightarrow & (\mathbb{R}_+^*)^n \\ (x_1, \dots, x_n)^\top & \longrightarrow & (nx_1, (n-1)(x_2 - x_1), \dots, 2(x_{n-1} - x_{n-2}), x_n - x_{n-1}) \end{cases}$$

On voit facilement que le déterminant du gradient (Jacobien) de Φ vaut $n!$. De plus $\sum x_i = \sum y_i$ pour $(y_i) = \Phi(x_i)$. On en déduit donc, par la formule de changement de variable en dimension n que

$$\begin{aligned} & n! \int_{\mathbb{R}^n} I((nx_1, \dots, x_n - x_{n-1}) \in B) I(x_1 < \dots < x_n) \exp\left(-\sum_{i=1}^n x_i\right) dx_1 \cdots dx_n \\ &= \int_{\mathbb{R}_+^n} I((y_1, \dots, y_n) \in B) \exp\left(-\sum_{i=1}^n y_i\right) dy_1 \cdots dy_n \end{aligned}$$

Ainsi la loi de

$$(nX_{(1)}, (n-1)(X_{(2)} - X_{(1)}), \dots, 2(X_{(n-1)} - X_{(n-2)}), X_{(n)} - X_{(n-1)}) \quad (11)$$

a pour densité $(y_1, \dots, y_n) \rightarrow \exp\left(-\sum_{i=1}^n y_i\right) I(0 < y_i : i = 1, \dots, n)$.

(b) On déduit de la question précédente que les coordonnées de (11) sont indépendante et comme $\hat{\alpha}_n = X_{(1)}$ et

$$\sum_{k=2}^{n-1} (n-k)((X_{(k+1)} - X_{(k)}) = X_{(n)} + \dots + X_{(2)} - (n-1)X_{(1)} = n(\bar{X}_n - X_{(1)}) = n\hat{\theta}_n$$

on a bien que $\hat{\alpha}_n$ et $\hat{\theta}_n$ sont indépendantes.

Exercice 8.3 (Test du signe)

Soient X_1, \dots, X_n un n échantillon de variables i.i.d. de fonction de répartition F continue et Y_1, \dots, Y_n un n -échantillon de variables i.i.d. de fonction de répartition G continue qui sont aussi indépendantes des X_i . On considère le problème de test

$$H_0 : F = G \text{ contre } H_1 : F \neq G \quad (12)$$

1. Montrer que $\mathbb{P}[X_i = Y_i] = 0$ et en déduire que si $F = G$ alors $\mathbb{P}[X_i > Y_i] = 1/2$.
2. On pose $\hat{N} = \sum_{i=1}^n I(X_i > Y_i)$. Quelle est la loi de \hat{N} sous H_0 ?
3. Soit $\alpha \in (0, 1)$. Construire un test de niveau asymptotique α ayant une zone de rejet de la forme

$$\mathcal{R}(c_\alpha) = \{z = ((x_i, y_i))_{i=1}^n : \sqrt{n}|\hat{N}(z) - 1/2| > c_\alpha\}$$

pour un bon choix de c_α .

4. On dit qu'un test est *consistant* quand sa fonction puissance tend vers 1 quand n tend vers l'infini. Montrer que le test construit dans la question précédente n'est pas consistant.

Correction de l'exercice 8.3

1. Soit $i \in \{1, \dots, n\}$. Comme les fonctions de répartition de X_i et Y_i sont continues et que X_i et Y_i sont indépendantes, la fonction de répartition du couple (X_i, Y_i) est le produit tensoriel des fonctions de répartition de X_i et Y_i qui est aussi une fonction continue sur \mathbb{R}^2 ; et en particulier, la loi de (X, Y) ne charge pas les sous-ensembles de mesure nulle pour le mesure de Lebesgue.

On a alors

$$\mathbb{P}[X_i = Y_i] = \int_{\{(x,y) \in \mathbb{R}^2 : x=y\}} d\mathbb{P}^{(X,Y)}(x, y) = 0$$

car l'ensemble $\{(x, y) \in \mathbb{R}^2 : x = y\}$ est de mesure nulle pour la mesure de Lebesgue. On a

$$1 = \mathbb{P}[X_i > Y_i] + \mathbb{P}[X_i = Y_i] + \mathbb{P}[X_i < Y_i] = \mathbb{P}[X_i > Y_i] + \mathbb{P}[X_i < Y_i]$$

et comme X_i et Y_i sont i.i.d. sous H_0 , on a $\mathbb{P}[X_i > Y_i] = \mathbb{P}[X_i < Y_i]$. On en déduit que $\mathbb{P}[X_i > Y_i] = 1/2$ sous H_0 .

2. Sous H_0 , $(I(X_i) > Y_i)_i$ est une famille de n variables de Bernoulli i.i.d. de paramètre $1/2$. Donc \hat{N} est une variable aléatoire multinomiale de paramètre $1/2$ sous H_0 .
3. Sous H_0 , le TCL donne

$$2\sqrt{n} \left(\frac{\hat{N}}{n} - \frac{1}{2} \right) \xrightarrow{d} \mathcal{N}(0, 1). \quad (13)$$

Ainsi pour $c_\alpha = q_{1-\alpha/2}^{\mathcal{N}(0,1)}/2$, le test de zone de rejet $\mathcal{R}(c_\alpha)$ est de niveau asymptotique α .

4. L'alternative est l'ensemble des couples de fonctions de répartition (F, G) telles que F et G sont continues et $F \neq G$. Prenons pour F la fonction de répartition d'une variable aléatoire X uniformément distribuée sur $[-1, 0] \cup [1, 2]$ et pour G la fonction de répartition d'une variable aléatoire Y uniformément distribuée sur $[0, 1]$. On a donc bien $F \neq G$, F et G sont continues et comme $\mathbb{P}[X > Y] = \mathbb{P}[X > 1] = 1/2$, on aura aussi (13) dans ce cas et en particulier,

$$\lim_{n \rightarrow +\infty} \mathbb{P}[\{(X_i, Y_i)\}_{i=1}^n \in R(c_\alpha)] = \alpha < 1.$$

Donc le test n'est pas consistant : en au moins un point de l'alternative, la puissance ne tend pas vers 1 quand n tend vers l'infini.