Exercices corrigés de statistiques mathématiques

Guillaume Lecué

14 octobre 2015

Table des matières

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 $\lim_{n\to+\infty} \mathbb{E}[e^{itZ_n}] = e^{-\frac{t^2}{2}}$. 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 surement* 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 $\epsilon > 0$, $\mathbb{P}\big[|X_n - X| \geq \epsilon\big] \to 0$ quand n tend vers $+\infty$.

— (X_n) *converge en loi* vers X quand pour toute fonction continue bornée f on a $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$.

— si $p \geq 1$, on dit que (X_n) *converge dans L_p* vers X quand $\mathbb{E}|X_n - X|^p \to 0$ quand n tend vers $+\infty$.

On a les implications suivantes :

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

Démo et contre-exemple de "$\underset{\Longrightarrow}{(1)}$" : Soit $\epsilon > 0$. On a $\{X_n \to X\} \subset \mathrm{liminf}_n\{|X_n - X| \leq \epsilon\}$. En passant, au complémentaire, on a :

$$\begin{aligned} 0 \leq \mathrm{limsup}_n\, \mathbb{P}\big[|X_n - X| > \epsilon\big] &\leq \mathbb{P}[\mathrm{limsup}_n\{|X_n - X| > \epsilon\}] \\ &= \mathbb{P}\big[\big(\mathrm{liminf}_n\{|X_n - X| \leq \epsilon\}\big)^c\big] \leq 0. \end{aligned}$$

Il n'y a pas équivalence dans "(1) $\Rightarrow$". Voici une exemple d'une suite qui converge en probabilité mais pas presque surement : (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 surement 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 $\mathbb{P}[\mathrm{limsup}_n\{X_n = 1\}] = 1$. Notamment, (X_n) ne converge pas presque surement vers 0.

Démo et contre-exemple de "$\underset{\Longrightarrow}{(2)}$" : Soit f une fonction continue bornée. Soit $\epsilon > 0$ et $N_\epsilon \in \mathbb{N}$ tel que $\mathbb{P}\big[|f(X_n) - f(X)| \geq \epsilon\big] \leq \epsilon$ (on rappel 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

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

La réciproque est trivialement fausse. Il suffit de prendre la suite stationnaire (X_n) où pour tout n, $X_n = g$ où g est une gaussienne. Comme g est symmétrique, $-g$ 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 par vers $-g$ en probabilité.

Démo et contre-exemple de "(3) $\Uparrow$" : D'après l'inégalité de Markov, $\mathbb{P}\big[|X_n - X \geq \epsilon|\big] \leq \epsilon^{-p}\mathbb{E}|X_n - X|^p$. 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

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

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 $t \mapsto \exp\big(-t^2(2-\sqrt{2})/2\big)$, 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 $\mathbb{P}[|Z_n - Z| > \epsilon] \to 0$. 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 :

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

et donc en passant à la limite, on obtient $\mathbb{P}\big[|Z_{2n}-Z_n| \geq 2\epsilon\big] \to 0$. 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 $X_n \overset{\text{loi}}{\to} X$ et $Y_n \overset{\text{loi}}{\to} Y$, 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) $X_n \overset{\text{loi}}{\to} X$ **et** $Y_n \overset{\mathbb{P}}{\to} Y$,
   (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 $c X$.**

****

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

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

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

$$\mathbb{E}\exp\Big(\frac{2it}{\sqrt{n}}\Big(\sum_{i=1}^{n}(\delta_i - 1/2)\Big)\Big) = \Big(\Big(\frac{1}{2}\Big)\Big(\exp\Big(\frac{-it}{\sqrt{n}}\Big) + \exp\Big(\frac{it}{\sqrt{n}}\Big)\Big)\Big)^n$$
$$= \Big(1 - \frac{t^2}{2n} + \mathcal{O}\Big(\frac{t^3}{n^{3/2}}\Big)\Big)^n \longrightarrow \exp\Big(\frac{-t^2}{2}\Big).$$

Soit g une variable Gaussienne standard. Comme g est symmétrique, $-g$ 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 $\mathbb{E}f(X_n)g(Y_n) \to \mathbb{E}f(X)g(Y)$ quand n tend vers l'infini. Par ailleurs, on sait que si (Y_n) converge en probablité 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 $n \geq N_\epsilon$,
$$\mathbb{P}\big[|g(Y_n) - g(Y)| \geq \epsilon\big] \leq \epsilon \text{ and } \big|\mathbb{E}f(X_n) - \mathbb{E}f(X)\big| \leq \epsilon.$$
On a pour tout $n \geq N_\epsilon$, par indépendance de $g(Y)$ avec $f(X_n)$ et $f(X)$,
$$\begin{aligned}&\big|\mathbb{E}f(X_n)g(Y_n) - \mathbb{E}f(X)g(Y)\big| \leq \big|\mathbb{E}f(X_n)(g(Y_n) - g(Y))I(|g(Y_n) - g(Y)| \geq \epsilon)\big|\\ &+ \big|\mathbb{E}f(X_n)(g(Y_n) - g(Y))I(|g(Y_n) - g(Y)| < \epsilon)\big| + \big|\mathbb{E}g(Y)(f(X_n) - f(X))\big|\\ &\leq 2\left\|f\right\|_\infty \left\|g\right\|_\infty \mathbb{P}\big[|g(Y_n) - g(Y)| \geq \epsilon\big] + \left\|f\right\|_\infty \epsilon + \big|\mathbb{E}g(Y)\mathbb{E}f(X_n) - \mathbb{E}f(X)\big|\\ &\leq \big(2\left\|f\right\|_\infty \left\|g\right\|_\infty + \left\|f\right\|_\infty + \left\|g\right\|_\infty\big)\epsilon.\end{aligned}$$

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. : $\big((X_n, Y_n)\big)$ 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 rappel qu'une suite (X_n) est *équi-intégrable* quand
$$\lim_{a \to +\infty} \sup_{n \in \mathbb{N}} \mathbb{E}\big[|X_n| I(|X_n| > a)\big] = 0.$$

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 $\sup_{n\in\mathbb{N}} \mathbb{E}\big[|Y_n|\mathbb{1}_A\big] \leq \epsilon$. Soit $\epsilon > 0$ et $a_0 > 0$ tel que pour tout $a \geq a_0$ et tout $n \in \mathbb{N}$, $\mathbb{E}\big[|X_n|I(|X_n| > a)\big] \leq \epsilon$. On a pour tout ensemble mesurable A, tout $n \in \mathbb{N}$ et tout $a \geq a_0$,

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

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

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

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

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

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

Comme (X_n) converge en probabilité, il existe un N_ϵ tel que pour tout $r, q \geq N_\epsilon$, $\mathbb{P}\big[|X_r - X_q| \geq \epsilon^{1/p}\big] \leq \eta$. On en déduit, que $\text{limsup}_{r,q} \mathbb{E}|X_r - X_q|^p \leq 2\epsilon$ 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 $\epsilon > 0$,

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

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

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

où B majore uniformément la suite $(\mathbb{E}|X_n|^p)$ (qui est bien bornée vue 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 $\mathbb{E}\big[|X|^p\mathbb{1}_A\big] \leq \epsilon/2^{p-1}$ 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 $a \geq a_0$,

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

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 (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 1.4

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

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

   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 bornée. Un changement de variable $x - y \to t$ donne

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

   (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 1.5 (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 1.5 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 symmétrique, 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 :

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

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

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

$$\begin{aligned}\mathbb{E}\hat{\sigma}_n^2 &= \mathbb{E}X^2 - \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big)^2 = \mathbb{E}X^2 - \frac{1}{n^2}\sum_{i,j}\mathbb{E}X_iX_j \\ &= \Big(1-\frac{1}{n}\Big)\big(\mathbb{E}X^2 - \big(\mathbb{E}X\big)^2\big) = \frac{n-1}{n}\mathrm{var}(X).\end{aligned}$$

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

On considère la décomposition suivante :

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

Par le TCL, on a :

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

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

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

****

Exercice 1.6 (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 $z_\alpha, n, \bar{X}_n$ tel que $\lim_{n\to\infty}\mathbb{P}(\theta \in \hat{I}_{n,\alpha}) = 1-\alpha$.**

****

Correction de l'exercice 1.6

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

2. D'après le TCL :
$$\sqrt{n}\big(\bar{X}_n - \theta\big) \rightsquigarrow \sigma g.$$
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 :
$$\sqrt{n}\big(g(\zeta_n) - g(\theta)\big) \rightsquigarrow \mathcal{N}(0, \sigma^2\big(g'(\theta)\big)^2). \tag{1}$$
Dans notre cas, on cherche à trouver g tel que $(g(\bar{X}_n))$ est asymptotiquement normal de moyenne asymptotique 0 et de variance asymptotique $\theta(1-\theta)\big(g'(\theta)\big)^2 = 1$. On est donc amener 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 $g : \theta \in [0,1] \mapsto 2\text{arcsin}(\sqrt{x})$ (on rappel que $(\text{arcsinx})' = (1-x^2)^{-1/2}, \forall x \in [-1,1]$). Cette fonction est continûment différentiable en tout $\theta \in (0,1)$, alors d'après Proposition 1.10 (voir (1)), on a
$$\sqrt{n}\big(g(\bar{X}_n) - g(\theta)\big) \rightsquigarrow \mathcal{N}(0,1).$$
(On rappelle que g a été choisit tel que $\theta(1-\theta)\big(g'(\theta)\big)^2 = 1$ pour tout $\theta \in (0,1)$).

3. Pour tout $\alpha[0,2]$, 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
$$\mathbb{P}\big[\theta \in \hat{I}_{n,\alpha}\big] = \mathbb{P}\Big[\Big|\sqrt{n}\big(g(\bar{X}_n) - g(\theta)\big)\Big| \leq z_\alpha\Big] \longrightarrow \mathbb{P}[g \in [-z_\alpha, z_\alpha]] = 1 - \alpha$$
pour
$$\hat{I}_{n,\alpha} = \Big[\sin^2\Big(g(\bar{X}_n) - \frac{z_\alpha}{\sqrt{n}}\Big), \sin^2\Big(g(\bar{X}_n) + \frac{z_\alpha}{\sqrt{n}}\Big)\Big]$$

****

Exercice 1.7 (Les statistiques d'ordre)

Soient $X_1, \ldots, X_n$ 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 $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ les variables aléatoires $X_1, \ldots, X_n$ réordonnées par ordre croissant.

1. **Donner l'expression de la loi de la statistique d'ordre $(X_{(1)}, \ldots, X_{(n)})$ 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 1.7

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

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

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

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

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

$$f(x_1,\ldots,x_n) = n!\Big(\Pi_{i=1}^n f(x_i)\Big)I(x_1<\cdots<x_n).$$

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

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

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

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

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

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

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

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

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

$$F_{X_{(n)}}(t) = \mathbb{P}[X_{(n)}\leq t] = \mathbb{P}[X_1\leq t,\ldots,X_n\leq t] = \Big(\mathbb{P}[X_1\leq t]\Big)^n = \big(F(t)\big)^n.$$

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 :

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

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,

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

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

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

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

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

Alors W a pour densité

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

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

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

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

****

Exercice 1.8 (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**
$$X \overset{Loi}{=} \mathcal{E}(\lambda) \Leftrightarrow \forall x > 0, \mathbb{P}(X > x) = \mathbf{exp}(-\lambda x).$$

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 $Z_1, \ldots, Z_n$.**
   (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 $X_1^1, \ldots, X_1^{n_1}$ et d'un stock de n_2 machines de type 2, de durées de vie $X_2^1, \ldots, X_2^{n_2}$. 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 $x \geq 0$:
   $$f_U(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} \mathbf{exp}(-\lambda x).$$
   (b) Écrire Z en fonction des X_i^j et en déduire $\mathbb{P}(Z \geq t)$ en fonction n_1, n_2, λ_1, λ_2 et t.

****

Correction de l'exercice 1.8

1. Par définition, une v.a.r. suit une loi exponentielle $\mathcal{E}(\lambda)$ quand elle admet une densité de la form $f_\lambda : x \in \mathbb{R} \mapsto \lambda \exp(-\lambda x) I(x > 0)$. 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 systéme. On a donc $Z > t$ ssi $X_1 > t$ et $X_2 > t$ et donc par indépendance
   $$\mathbb{P}[Z > t] = \mathbb{P}[\{X_1 > t\} \cap \{X_2 > t\}] = \mathbb{P}[X_1 > t]\mathbb{P}[X_2 > t] = \exp\big(-(\lambda_1 + \lambda_2)t\big).$$
   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 $\mathbb{P}[X_1 < X_2]$:
   $$\begin{aligned} \mathbb{P}[X_1 < X_2] &= \mathbb{E} I(X_1 < X_2) = \int_{\mathbb{R}_+^2} I(x_1 < x_2) f_{\lambda_1}(x_1) f_{\lambda_2}(x_2) dx_1 dx_2 \\ &= \int_0^\infty f_{\lambda_1}(x_1) \Big(\int_{x_1}^\infty f_{\lambda_2}(x_2) dx_2 \Big) dx_1 = \int_0^\infty f_{\lambda_1}(x_1) \exp(-\lambda_2 x_1) dx_1 \\ &= \int_0^\infty \lambda_1 \exp\big(-(\lambda_1 + \lambda_2) x_1\big) dx_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2}. \end{aligned}$$

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

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

Par symmétrie,

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

On en déduit que Z et I sont indépendantes.

4. a) Le modéle statistique est $\{\mathcal{E}(\lambda_1 + \lambda_2)^{\otimes n} : \lambda_1, \lambda_2 > 0\}$. Ce modèle n'est pas identifiable en le paramétre (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 1.9 (Lemme de Fatou)
si (f_n) est une suite de fonctions measurables alors

$$\int \text{liminf}_n f_n \leq \text{liminf}_n \int f_n.$$

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

$$\text{limsup}_{\mathbf{n}} \mathbb{P}(\mathbf{A_n}) \leq \mathbb{P}(\text{limsup}_{\mathbf{n}} \mathbf{A_n}),$$

où on rappelle que $\text{limsup}_n A_n = \cap_N \cup_{n \geq N} A_n$.

****

Correction de l'exercice 1.9

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 surement vers $\text{liminf}_n f_n$. Le théorème de convergence monotone donne :

$$\lim_n \int g_n = \int \lim_n g_n = \int \text{liminf}_n f_n.$$

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

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

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 $\text{liminf}_n f_n = \mathbb{1}_{\text{liminf}_n A_n^c}$ et $\big(\text{liminf}_n A_n^c\big)^c = \text{limsup}_n A_n$ donc

$$1 - \mathbb{P}[\text{limsup}_n A_n] = \mathbb{P}[\text{liminf}_n A_n^c] \leq \text{liminf}_n \mathbb{P}[A_n^c].$$

****

Exercice 1.10 (la loi du $0-1$ de Kolmogorov)
Soit (σ_n) une suite de tribus indépendantes. La tribu asymptotique est $\sigma_\infty = \cap_n \sigma\Big(\cup_{p\geq n}\sigma_p\Big)$. 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 1.10 On note $\alpha_n = \sigma\Big(\cup_{p\geq n}\sigma_p\Big)$ et $\beta_n = \sigma\Big(\cup_{p<n}\sigma_p\Big)$. Les deux tribus α_n et β_n sont indépendantes. Comme $\sigma_\infty \subset \alpha_n$ alors σ_∞ est indépendantes de β_n pour tout n. Notamment, σ_∞ est indépendante de $\cup_{n\in\mathbb{N}}\beta_n$ et donc de $\sigma\Big(\cup_n \beta_n\Big) = \sigma\Big(\cup_n \sigma_n\Big) = \alpha_0$. 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 1.11 (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 1.11 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 $\mathbb{P}^{Y_n}[A] \to \mathbb{P}^Y[A]$.

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

****

Exercice 1.12 (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 1.12

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

pour $x \in [0,1)$,

$$\begin{aligned}\mathbb{P}[B_n] &= \mathbb{P}[\cap_{p\geq n} A_p^c] = \Pi_{p\geq n}\mathbb{P}[A_p^c] = \Pi_{p\geq n}\big(1-\mathbb{P}[A_p]\big)\\ &= \exp\Big(\sum_{p\geq n}\log\big(1-\mathbb{P}[A_p]\big)\Big) \leq \exp\Big(-\sum_{p\geq n}\mathbb{P}[A_p]\Big) = 0.\end{aligned}$$

On en déduit le résultat.

****

Exercice 1.13 (L'asymptotique normalité implique la converge 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 1.13 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 1.14 (quartile)
Soit la loi de probabilité de densité $f(x) = 2xI\{0 \leq x \leq 1\}$.

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 $n \to \infty$., où $\widehat{F}_n$ est la fonction de répartition empirique.**

****

Correction de l'exercice 1.14

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

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

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

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

2 Vraisemblance, EMV, IC, Information de Fisher

****

Exercice 2.1 (Modèle probit)

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

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

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 :

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

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 2.1

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 $(x - m)/\sigma \to t$ donne

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

   La loi de Y est donc $(1 - \Phi(-m/\sigma^2))\delta_1 + \Phi(-m/\sigma^2)\delta_0$.

2. Les paramétres 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 2.2 (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 2.2

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

   Modèle 1 : On modèlise ce problème par une famille de n couples $(\delta_1^{(1)}, \delta_1^{(2)}), \ldots, (\delta_n^{(1)}, \delta_n^{(2)})$ où les $\delta_i^{(j)}, i = 1, \ldots, n, j = 1, 2$ sont i.i.d. Bernoulli sur $\{A, a\}$ de paramétre θ. On dit que $\delta_i^{(j)} = a$ quand l'alléle a est présent chez l'individu i au gène numéro 2. On a donc bien le 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'alléle a pour chacun des deux gènes.

   Modèle 2 : On peut modèliser ce problème par une famille de n variables aléatoires $X_1, \ldots, X_n$ 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'alléle 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

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

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

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

   Alors l'estimateur du maximum de vraisemblance est donné par

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

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

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

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

   car

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

   et

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

   alors

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

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

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

alors $g(\theta) = 2\sqrt{2}\text{arcsin}(\sqrt{\theta})$. 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 ∞,

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

****

Exercice 2.3 (Modèle d'autorégression)

On considère les observations $X_1, \ldots, X_n$, où les X_i sont issus du *modèle d'autorégression d'ordred* 1 **:**

$$X_i = \theta X_{i-1} + \xi_i, \quad i = 1, \ldots, n, \qquad X_0 = 0, \tag{2}$$

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

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

****

Correction de l'exercice 2.3

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

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

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 $\mathbb{P}_\theta$ est donc la loi de $(X_1, X_2, \ldots, X_n)$ sous l'hypothèse "AR(1)" de l'équation 2. On a $\mathbb{P}_\theta^{X_i|X_{i-1},\ldots,X_1} = \mathbb{P}_\theta^{X_i|X_{i-1}} \sim \mathcal{N}(\theta X_i, \sigma^2)$. On montre par récurrence que

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

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 $(x_1, \ldots, x_n) \in \mathbb{R}^n$ par :

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

où f est la densité d'une Gaussienne de moyenne nulle et de variance σ^2. Pour la recurrence, 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 :

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

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

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

et aussi

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

Alors l'EMV est donné par :

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

****

Exercice 2.4 (Paramètre vectoriel - vitesses de convergence différentes)

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

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

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 $\sqrt{n}(\hat{\theta}_n - \theta)$.**

****

Correction de l'exercice 2.4

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

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

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 $\theta \mapsto V(\theta, \hat{\alpha}_n)$ 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

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

Par indépendance, on a

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

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

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

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

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

****

Exercice 2.5 (EMV Gaussienne tronquée)

Soit une loi de probabilité de densité (par rapport à la mesure de Lebesgue sur $\mathbb{R}$)

$$f(x, \theta) = (2/\sqrt{\pi\theta}) \exp(-x^2/\theta) I(x > 0), \quad \forall x \in \mathbb{R},$$

où $\theta > 0$ est un paramètre inconnu. On dispose d'un n-échantillon $X_1, \ldots, X_n$ de cette loi.

1. **Expliciter l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ. Montrer qu'il est consistant.**
2. **Calculer la variance de l'estimateur $\hat{\theta}_n$ (on rappelle que $\mathbb{E}g^4 = 3$ pour $g \sim \mathcal{N}(0,1)$).**
3. **On admettra que le modèle statistique en question est régulier. Calculer l'information de Fisher associée à ce modèle. Comparer la avec la variance de $\hat{\theta}_n$. Conclusion ?**
4. **Déterminer la loi limite quand $n \to \infty$ de $\sqrt{n}(\hat{\theta}_n - \theta)$.**
5. **Fixons $\alpha \in]0,1[$. En utilisant le résultat de 4), proposer un test de niveau asymptotique α de l'hypothèse $H_0 : \ \theta < 3$ contre l'alternative $H_1 : \ \theta > 3$.**

****

Correction de l'exercice 2.5

1. La fonction de vraisemblance est définie pour tout $\theta > 0$ par

$$V(\theta, (X_i)_i) = \prod_{i=1}^n f(X_i, \theta) = \Big(\frac{2}{\sqrt{\pi\theta}}\Big)^n \exp\Big(-\frac{1}{\theta}\sum_{i=1}^n X_i^2\Big) I\big(\min_i X_i > 0\big).$$

Pour tout $\theta > 0$, les X_i sont presque sûrement positifs sous $\mathbb{P}_\theta$ alors $\min_i X_i > 0$ p.s. et donc La log-vraisemblance est ici :

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

En étudiant, cette fonction en $\theta > 0$, on voit que la vraisemblance est maximale en $\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} = \frac{2}{n}\sum_i X_i^2$. De plus, si X a pour densité $f(\cdot,\theta)$ pour un certain $\theta > 0$ alors, après un changement de variable ($u = x/\sqrt{\theta}$), on voit que

$$\mathbb{E}_\theta X^2 = \int_{x>0} \frac{2x^2}{\sqrt{\pi\theta}}\exp\big(-x^2/\theta\big)dx = \frac{\theta}{2}.$$

Alors $\mathbb{E}_\theta\,\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} = 2(\theta/2) = \theta$ et donc l'EMV est sans biais.

2. La variance de $\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}}$ est donnée, sous $\mathbb{P}_\theta$, par

$$\mathrm{Var}_\theta(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}}) = \mathbb{E}_\theta\Big(\frac{2}{n}\sum_{i=1}^n X_i^2 - \theta\Big)^2 = (4/n)\,\mathrm{Var}_\theta(X^2) = (4/n)\big(\mathbb{E}_\theta X^4 - (\mathbb{E}_\theta X^2)^2\big) = \frac{4(3\theta^2/4 - \theta^2/4)}{n} = \frac{2\theta^2}{n}.$$

3. L'information de Fisher contenue dans un $n-$échantillon s'obtient à partir de la formule $I_n(\theta) = nI_1(\theta)$ et pour une observation, on a

$$I_1(\theta) = -\mathbb{E}_\theta\partial_\theta^2\log f(X,\theta) = \mathbb{E}_\theta\Big(\frac{2X^2}{\theta^3} - \frac{1}{2\theta^2}\Big) = \frac{1}{2\theta^2}.$$

Finalement, l'information de Fisher du n-échantillon est $n/(2\theta^2)$. On obtient donc que (pour ce modèle) $\mathrm{Var}_\theta(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}}) = I_n(\theta)^{-1}$. Autrement dit, la variance de l'EMV vaut de manière <u>non-asymptotique</u> l'inverse de l'information de Fisher. Ce résultat est à mettre en parallèle avec le résultat sur la normalité asymptotique des EMV dans les modèles réguliers qui assure que la variance asymptotique des EMV vaut l'inverse de l'information de Fisher :

$$\sqrt{n}\big(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} - \theta\big) \rightsquigarrow \mathcal{N}(0, I_1(\theta)^{-1})$$

quand $n \to \infty$. Ici, on a

$$n\mathbb{E}_\theta\big(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} - \theta\big)^2 = I_1(\theta)^{-1}$$

<u>pour tout n</u> (càd de manière non-asymptotique).

4. On a rappelait précédemment que, dans les modèles réguliers (ce qui est admis ici), l'EMV est asymptotiquement normal de variance asymptotique donnée par l'information de Fisher. On a donc ici :

$$\sqrt{n}\big(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} - \theta\big) \rightsquigarrow \mathcal{N}(0, I_1(\theta)^{-1})$$

quand $n \to \infty$ où $I_1(\theta)^{-1} = 2\theta^2$.

5. Soit $\alpha \in]0,1[$. On utilise l'EMV comme statistique de test. On considère alors un test de la forme

$$\varphi_\alpha(X_1,\ldots,X_n) = \begin{cases} H_0 & \text{quand } \sqrt{n}\big(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} - 3\big) \le t_\alpha \\ H_1 & \text{sinon.}\end{cases}$$

On fixe le seuil t_α en fonction du niveau asymptotique :

$$\sup_{\theta<3} \limsup_{n\to+\infty} \mathbb{P}_\theta[\sqrt{n}\big(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} - 3\big) > t_\alpha] = \alpha.$$

Étant donné la normalité asymptotique de l'EMV énoncé dans la question précédente, on voit que le $\sup_{\theta<3}$ est obtenu en $\theta = 3$ et donc

$$\sup_{\theta<3} \limsup_{n\to+\infty} \mathbb{P}_\theta[\sqrt{n}\big(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} - 3\big) > t_\alpha] = \limsup_{n\to+\infty} \mathbb{P}_3[\sqrt{n}\big(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} - 3\big) > t_\alpha] = \mathbb{P}[I_1(3)^{-1/2} g > t_\alpha]$$

où $g \sim \mathcal{N}(0,1)$ et $I_1(3) = 18$. Il suffit alors de prendre $t_\alpha = q_{1-\alpha}/\sqrt{18}$, où $q_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ d'une $\mathcal{N}(0,1)$. Pour conclure, un test de niveau asymptotique α est donné par

$$\varphi_\alpha(X_1,\dots,X_n) = \begin{cases} H_0 & \text{quand } \sqrt{n}\big(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} - 3\big) \leq \frac{q_{1-\alpha}}{3\sqrt{2}} \\ H_1 & \text{sinon.} \end{cases}$$

****

Exercice 2.6 (Durées de connection)

On peut modéliser la durée d'une connection sur le site `www.Cpascher.com` par une loi gamma$(2, 1/\theta)$ 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 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 2.6

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 $L : \theta \in \mathbb{R}_+^* \longmapsto \sum_{i=1}^n \log f_\theta(X_i)$. 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 $\hat{\theta}_n \in \mathrm{argmax}_{\theta>0} L(\theta) = \{(1/2)\bar{X}_n\}$.
2. $\mathbb{E}\hat{\theta}_n = \theta$. Pour la variance, on a
$$\mathrm{var}(\hat{\theta}_n) = \frac{1}{4n}\mathrm{var}(X_1) = \frac{\mathbb{E}X^2 - (\mathbb{E}X)^2}{4n} = \frac{2\theta^2}{4n}.$$

****

Exercice 2.7 (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 défaillants 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 $\mathbb{P}(\theta \geq \hat{\theta}_n + \epsilon_n) \overset{n\to\infty}{\to} 5\%$.

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 $\mathbb{P}(\theta \geq \hat{\theta}_n + \epsilon'_n) \overset{n\to\infty}{\to} 5\%$.

****

Correction de l'exercice 2.7

1. On modélise ce problème par une famille de n variables de Bernoulli $\delta_1, \ldots, \delta_n$ 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-iéme produit prélevé est défaillant et $\delta_i = 0$ signifie qu'il n'est pas défaillant. 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 $0 \leq k \leq n$,
$$\mathbb{P}\big[n\hat{\theta}_n = k\big] = \begin{pmatrix} n \\ k \end{pmatrix} \theta^k (1-\theta)^{n-k}.$$
On reconnaît la loi d'une multinomiale de paramétre n, θ.
2. La loi des grands nombres assure que $(\hat{\theta}_n)$ converge presque surement vers $\mathbb{E}\delta = \theta$. Comme $\mathrm{var}(\delta) = \mathbb{E}\delta^2 - (\mathbb{E}\delta)^2 = \theta - \theta^2$, 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,
$$\mathbb{P}\Big[\theta \geq \hat{\theta}_n + \epsilon\sqrt{\frac{\theta - \theta^2}{n}}\Big] = \mathbb{P}\Big[\sqrt{\frac{n}{\theta - \theta^2}}(\theta - \hat{\theta}_n) \geq \epsilon\Big] \longrightarrow \mathbb{P}[g \geq \epsilon].$$
Si on choisit $\epsilon > 0$ tel que $\mathbb{P}[g \geq \epsilon] = 5\%$, on obtient le résultat pour $\varepsilon_n = \sqrt{(\theta - \theta^2)/n}\epsilon$.
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 surement 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 surement. 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 surement vers $f(\theta)$, on en déduit par le lemme de Slutsky que $\big(f(\hat{\theta}_n)\sqrt{n}(\theta - \hat{\theta}_n)\big)_{n\geq N}$ 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
$$\varepsilon'_n = \sqrt{\frac{\hat{\theta}_n - \hat{\theta}_n^2}{n}}\epsilon.$$
On a alors :
$$\mathbb{P}\Big[\theta \geq \hat{\theta}_n + \epsilon'_n\Big] = \mathbb{P}\Big[f(\hat{\theta}_n)\sqrt{n}(\theta - \hat{\theta}_n) \geq \epsilon\Big] \longrightarrow \mathbb{P}[g \geq \epsilon] = 5\%.$$

****

Exercice 2.8 (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 soir 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 $n \to +\infty$? 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 2.8

1. On rappel qu'une variable de Poisson Z de paramétre 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 paramétre $\theta = \lambda/n$. Pour tout $0 \leq k \leq n$, on a

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

   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 $+\infty$, $n! \sim \sqrt{2\pi n}\Big(\frac{n}{e}\Big)^n$. Alors, on a

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

   qui converge bien vers 1. Donc $\hat{\lambda}_n$ converge en loi vers une variable de Poisson de paramétre λ.
2. Comme $\hat{\lambda}_n$ converge en loi vers une Poisson de paramétre 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 2.9 (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. $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{B}(\theta)$.
2. $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{N}(m, v)$.
3. $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{U}[0, \theta]$.

****

Correction de l'exercice 2.9 On rappelle les formules du cours pour le calcul de l'information de Fisher :

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

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 $(\overline{X}_n, \hat{\sigma}_n^2)$ où $\overline{X}_n = \frac{1}{n}\sum_i X_i$ et $\hat{\sigma}_n^2 = \frac{1}{n}\sum_i (X_i - \overline{X}_n)^2$. 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 $\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} = \max_i X_i$ et son comportement asymptotique en étudiant sa fonction de répartition :
$$\mathbb{P}_\theta\Big[\frac{n(\theta - \widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}})}{\theta} > x\Big] = \mathbb{P}_\theta\Big[\forall i = 1, \ldots, n : \quad \frac{n(\theta - X_i)}{\theta} > x\Big] = \begin{cases} 1 & \text{si } x < 0 \\ \left(1 - \frac{x}{n}\right)^n & \text{quand } 0 < x < n \\ 0 & \text{sinon} \end{cases}$$
car pour tout i, sous $\mathbb{P}_\theta$, $(\theta - X_i)/\theta \sim \mathcal{U}([0,1])$. Alors quand $n \to \infty$, on obtient :
$$\lim_{n\to+\infty} \mathbb{P}_\theta\Big[\frac{n(\theta - \widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}})}{\theta} > x\Big] = \begin{cases} 1 & \text{si } x < 0 \\ \exp(-x) & \text{quand } x > 0 \end{cases}$$
donc
$$\frac{n(\theta - \widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}})}{\theta} \rightsquigarrow \mathcal{E}(1)$$
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
$$L(\theta, (X_i)_i) = \prod_{i=1}^{n} \theta \exp(-\theta X_i) = \theta^n \exp\Big(- \theta \sum_i X_i\Big).$$
Le rapport de vraisemblance est
$$\frac{L(\theta_0, (X_i)_i)}{L(\theta_1, (X_i)_i)} = \exp\Big(- (\theta_0 - \theta_1) \sum_i X_i\Big).$$
Le rapport de vraisemblance est donc une fonction croissante de $\overline{X}_n$ (on a ici $\theta_1 > \theta_0$). Alors, le test de Neyman-Pearson de niveau α est de la forme :
$$\varphi_\alpha((X_i)_i) = \begin{cases} H_0 & \text{quand } \overline{X}_n \geq t_\alpha \\ H_1 & \text{sinon} \end{cases}$$
où t_α est un seuil à choisir tel que
$$\mathbb{P}_{\theta_0}[\overline{X}_n < t_\alpha] = \alpha.$$
On sait qu'une telle solution existe car $\overline{X}_n$ est une v.a.r. admettant une densité par rapport à la mesure de Lebesgue. Néanmoins, cette quantité reste difficile à calculer, on préféra alors fixer le seuil de manière asymptotique vue que $\overline{X}_n$ 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
$$\frac{L(\theta_0, (X_i)_i)}{L(\theta_1, (X_i)_i)} = \Big(\frac{\theta_0}{\theta_1}\Big)^{\sum_i X_i} \Big(\frac{1-\theta_0}{1-\theta_1}\Big)^{\sum_i (1-X_i)}.$$
Le rapport de vraisemblance est donc une fonction décroissante de $\overline{X}_n$ (on a ici $\theta_1 > \theta_0$). Alors, le test de Neyman-Pearson de niveau α est de la forme :
$$\varphi_\alpha((X_i)_i) = \begin{cases} H_0 & \text{quand } \overline{X}_n \leq t_\alpha \\ H_1 & \text{sinon} \end{cases}$$
où t_α est un seuil à choisir tel que
$$\mathbb{P}_{\theta_0}[\overline{X}_n > t_\alpha] = \alpha.$$
Ici, cette équation n'admet pas nécessairement de solution car $\overline{X}_n$ est une variable discrète. Dans ce cas, on peut avoir recours à des tests "randomizés" (hors programme), mais on préférera fixer le seuil t_α de manière asymptotique vue que $\overline{X}_n$ 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 défaillants 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 $n \to \infty$.**
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 `défaillant`.

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
$$\begin{aligned} L(p,(X_i)_i) &= \prod_{i=1}^n \Big[p^2 I(X_i = N) + 2p(1-p)I(X_i = MF) + (1-p)^2 I(X_i = D)\Big] \\ &= (p^2)^{\#N}[2p(1-p)]^{\#MF}[(1-p)^2]^{\#D}. \end{aligned}$$
   et la log-vraisemblance est
$$\ell_n(p,(X_i)_i) = \log\Big(\frac{p}{1-p}\Big)[\#N - \#D] + (\#D - \#N)\log 2 + n\log[2p(1-p)].$$
   En étudiant la fonction de log-vraisemblance, on voit que la vraisemblance est maximale en
$$\hat{p}_n = \frac{1}{2}\Big(\frac{\#N - \#D}{n} + 1\Big),$$
   qui est donc l'estimateur du maximum de vraisemblance. Par la loi forte des grands nombres, on a :
$$\frac{\#N}{n} \underset{n\to+\infty}{\longrightarrow} p^2, \text{ et } \frac{\#D}{n} \underset{n\to+\infty}{\longrightarrow} (1-p)^2$$

et donc $\hat{p}_n \underset{n\to+\infty}{\longrightarrow} p$, 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 $\sqrt{n}\big(\bar{Z}_n - \mathbb{E}Z_1\big) \rightsquigarrow \mathcal{N}(0, \mathrm{Var}\, Z_1)$. On obtient alors :

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

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 $T_n \underset{n\to+\infty}{\longrightarrow} +\infty$ 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 $X_1, \ldots, X_n$ 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" : $\theta \in \Theta_1 \to \mathbb{P}_\theta[rejet]$. 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 modèlise ce problème par le modèle d'échantillonnage $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{B}(p)$ où

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

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 décés 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 $\overline{X}_n$. 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 $T_n = \sqrt{n}(\overline{X}_n - 1/2)$.

— 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 :

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

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 $2T_n = 2\sqrt{1919}\big(922/1919 - 1/2\big) \approx -1.712$. 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

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

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 $\bar{X} = \frac{1}{n}e^T X = \big(\bar{X}_1, \ldots, \bar{X}_k\big)$.**

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 Var($\hat{\theta}_j$), pour $j = 0, \ldots, p$?**

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

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

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

5. **A-t-on la relation $\bar{\hat{y}} = \bar{y}$? 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 :

$$(\hat{\theta}_0, \hat{\theta})^\top \in \text{argmin}_{(\theta_0', \theta')^\top \in \mathbb{R}\times\mathbb{R}^k} \left\|y - \theta_0' e - X\theta'\right\|_2.$$

Alors $\hat{y} = \hat{\theta}_0 e + X\hat{\theta}$ est la projection orthogonale de y sur $\text{vect}(e, X^{(1)}, \cdots, X^{(k)})$ où $X^{(1)}, \cdots, X^{(k)}$ sont les vecteurs colonnes de X. En particulier, pour tout $\theta_0' \in \mathbb{R}, \theta' \in \mathbb{R}^k$, on a

$$\left\langle y - \hat{y}, \theta_0' e + X\theta' \right\rangle = 0.$$

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 $\bar{X} = \left(\bar{X}^{(1)}, \cdots, \bar{X}^{(k)}\right)$.

2. $\bar{y}e$ est un élément de $\text{vect}(e, X^{(1)}, \cdots, X^{(k)})$. 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 $\text{vect}(e, X^{(1)}, \cdots, X^{(k)})$ (modèle sans bruit).
   2. $R^2 = 0$ signifie que $\hat{y} = \bar{y}e$. Donc y est orthogonal à $\text{vect}(X^{(1)}, \cdots, X^{(k)})$. Alors $X^{(1)}, \cdots, X^{(k)}$ sont des mauvaises variables pour expliquer ou prédire y.

3. Soit Proj l'opérateur de projection sur $\text{vect}(e, X^{(1)}, \cdots, X^{(k)})$. On a $Z\left(\hat{\theta}_0, \hat{\theta}\right)^\top = \text{Proj}(y)$. On a pour tout $\theta_0' \in \mathbb{R}, \theta' \in \mathbb{R}^k$, $\left\langle y - Z\left(\hat{\theta}_0, \hat{\theta}\right)^\top, Z\left(\theta_0', \theta'\right)^\top \right\rangle = 0$. Par ailleurs,

$$\left\langle y - Z\left(\hat{\theta}_0, \hat{\theta}\right)^\top, Z\left(\theta_0', \theta'\right)^\top \right\rangle = \left\langle Z^\top y - Z^\top Z\left(\hat{\theta}_0, \hat{\theta}\right)^\top, \left(\theta_0', \theta'\right)^\top \right\rangle.$$

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

On peux aussi voir que

$$(\hat{\theta}_0, \hat{\theta})^\top \in \text{argmin}_{\theta_0' \in \mathbb{R}, \theta' \mathbb{R}^k} \left\|y - \theta_0' e - X\theta'\right\|_2.$$

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 $\big(Z^\top Z\big)^{-1} Z^\top y = \big(\hat{\theta}_0, \hat{\theta}\big)^\top$.
La matrice de covariance de $\hat{\Theta} := (\hat{\theta}_0, \hat{\theta})^\top$ est donnée par

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

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

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

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

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

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

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

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

****

Exercice 4.2 (Régression Ridge)
On considère le modèle de regression

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

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

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**
$$\hat{\theta}_\lambda = \arg\min_{\theta \in \mathbb{R}^k} \left\{\|Y - X\theta\|^2 + \lambda\|\theta\|^2\right\}.$$
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 $\theta^{MC} = \left(X^\top X\right)^{-1} X^\top Y$. Comme $\ker(X^\top X) = \ker X$, on a voit 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 conditionement (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*

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

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

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

On doit aussi faire en sorte de bien choisir $\lambda > 0$. Ceci introduit le problème de la sélection de paramétre 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 $-2X^\top(Y - X\hat{\theta}_\lambda) + 2\lambda\theta = 0$. On a donc

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

3. Le biais de l'ER est donné par :

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

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

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

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

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

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

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

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

$$\Big(\frac{1}{1+\mu} - 1\Big)^2\theta^2 + \frac{\big(\sigma^2/\|X\|_2^2\big)}{\big(1+\mu\big)^2} < \big(\sigma^2/\|X\|_2^2\big) \tag{3}$$

càd $\mu\big(\theta^2 - \big(\sigma^2/\|X\|_2^2\big)\big) < 2\big(\sigma^2/\|X\|_2^2\big)$. 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

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

On suppose que X est une matrice déterministe, $\mathbb{E}[\xi] = 0$, $\mathbb{E}[\xi\xi^T] = \sigma^2 I_n$, 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 linÈaire en Y, i.e., $\tilde{\theta} = LY$ pour une matrice $L \in \mathbb{R}^{k\times n}$ déterministe. Donner une condition nÈcessaire et suffisante sur L pour que $\tilde{\theta}$ soit sans biais. On supposera maintenant cette hypothËse vÈrifiÈe.**

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 cov$(\tilde{\theta})$ = cov$(\hat{\theta}) + \sigma^2 \Delta \Delta^T$. En déduire que

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

4. En passant au risques quadratiques $\mathbb{E}\big[\|\tilde{\theta}-\theta\|^2\big]$ et $\mathbb{E}\big[\|\hat{\theta}-\theta\|^2\big]$, 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 $\hat{\theta} = \left(X^\top X\right)^\top X^\top y$. On remarque que rang$(X) = k$ donc $n \geq k$ et X est injective (donc $X^\top X$ esy inversible : en effet, $X^\top X$ est symmétrique 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 $X^\top X$ est inversible).
   On a donc $\mathbb{E}\hat{\theta} = \left(X^\top X\right)^{-1} X^\top \mathbb{E}y = \left(X^\top X\right)^{-1} X^\top X\theta = \theta$. Donc $\hat{\theta}$ est bien un estimateur sans biais. La matrice de covariance de $\hat{\theta}$ est donnée par
   $$\Sigma := \mathbb{E}\big(\hat{\theta} - \mathbb{E}\hat{\theta}\big)\big(\hat{\theta} - \mathbb{E}\hat{\theta}\big)^\top = \left(X^\top X\right)^{-1} X^\top \mathbb{E}\zeta\zeta^\top X \left(X^\top X\right)^\top = \sigma^2 \left(X^\top X\right)^{-1}$$
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. $\Sigma = \mathbb{E}\big((\tilde{\theta}-\theta)(\tilde{\theta}-\theta)^\top\big) = L\mathrm{var}(Y)L^\top = \sigma^2 LL^\top$. Comme $LX = I_k$, on a :
   $$\Delta X = LX - \left(X^\top X\right)^{-1} X^\top X = Ik - I_k = 0$$
   et la covariance de $\tilde{\theta}$ est donnée par :
   $$\begin{aligned} &\mathrm{var}(\tilde{\theta}) = \mathrm{var}(\Delta Y + \hat{\theta}) = \mathrm{var}(\Delta Y) + \mathrm{var}(\hat{\theta}) + \mathrm{cov}(\hat{\theta}, \Delta Y) + \mathrm{cov}(\Delta Y, \hat{\theta}) \\ &\sigma^2 \Delta\Delta^\top + \mathrm{var}(\hat{\theta}) + \mathrm{cov}(\hat{\theta}, \Delta Y) + \mathrm{cov}(\Delta Y, \hat{\theta}). \end{aligned}$$
   Par ailleurs, comme $\Delta X = 0$, on a $\mathbb{E}\Delta Y = 0$ et
   $$\mathrm{cov}(\Delta Y, \hat{\theta}) = \mathbb{E}\big[\Delta Y \hat{\theta}^\top\big] = \Delta \mathbb{E}\big[(X\theta + \zeta)\zeta^\top X (X^\top X)^{-1}\big] = 0$$
   car $\mathbb{E}\zeta\zeta^\top = \sigma^2 I_n$. De même cov$(\hat{\theta}, \Delta Y) = 0$. On en déduit que
   $$\mathrm{var}(\tilde{\theta}) = \mathrm{var}(\hat{\theta}) + \sigma^2 \Delta\Delta^\top \succeq \mathrm{var}(\hat{\theta}).$$
4. On a
   $$\left\|\tilde{\theta} - \theta\right\|_2^2 = \sum_{j=1}^k (\tilde{\theta}_j - \theta_j)^2 = \sum_{j=1}^k e_j^\top (\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^\top e_j$$
   alors
   $$\mathbb{E}\left\|\tilde{\theta} - \theta\right\|_2^2 = \sum_{j=1}^k e_j \mathrm{var}(\tilde{\theta}) e_j$$

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

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