

HEC Paris – ESCP Europe

Mathématiques approfondies – Mathématiques II

Corrigé du sujet adapté – 2011

Les questions et sous-questions conservent la numérotation de l'énoncé adapté.

Partie I – Quelques propriétés des statistiques d'ordre

Partie I – Statistiques d'ordre

$(X_1, \dots, X_n)$ i.i.d. de densité f_X , $Y_1 \leq \dots \leq Y_n$ réarrangement croissant, $J_k(x) = \mathbf{1}_{[X_k \leq x]}$, $S_n(x) = \sum_{k=1}^n J_k(x)$.

Corrigé – Partie I.

Comme X est à densité, F_X est continue sur $\mathbb{R}$ et il existe une partie finie D de $\mathbb{R}$ telle que F_X soit de classe C^1 sur $\mathbb{R} \setminus D$, avec $F'_X = f_X$ sur $\mathbb{R} \setminus D$.

1. Lois des statistiques d'ordre.

- a) Soit $x \in \mathbb{R}$. On a $[Y_n \leq x] = \bigcap_{k=1}^n [X_k \leq x]$ et $[Y_1 > x] = \bigcap_{k=1}^n [X_k > x]$. Par indépendance des X_k , qui ont tous la loi de X :

$$F_{Y_n}(x) = (F_X(x))^n, \quad F_{Y_1}(x) = 1 - \mathbb{P}(Y_1 > x) = 1 - (1 - F_X(x))^n.$$

Ces fonctions sont continues sur $\mathbb{R}$ et de classe C^1 sur $\mathbb{R} \setminus D$ (composées de F_X par des polynômes). On dérive sur $\mathbb{R} \setminus D$:

$$F'_{Y_n}(x) = n(F_X(x))^{n-1}f_X(x), \quad F'_{Y_1}(x) = n(1 - F_X(x))^{n-1}f_X(x).$$

Les fonctions f_{Y_1} et f_{Y_n} de l'énoncé sont positives et coïncident avec ces dérivées hors de l'ensemble fini D .

$$f_{Y_1} = n(1 - F_X)^{n-1}f_X \text{ et } f_{Y_n} = nF_X^{n-1}f_X \text{ sont des densités de } Y_1 \text{ et } Y_n.$$

- b) Chaque $J_k(x)$ suit la loi de Bernoulli de paramètre $\mathbb{P}(X_k \leq x) = F_X(x)$. Les variables $J_1(x), \dots, J_n(x)$ sont indépendantes, car $J_k(x)$ est une fonction de X_k seule (lemme des coalitions). $S_n(x)$ compte donc les succès dans n épreuves de Bernoulli indépendantes de même paramètre.

$$S_n(x) \hookrightarrow \mathcal{B}(n, F_X(x)).$$

- c) Soit $\omega \in \Omega$. $S_n(x)(\omega)$ est le nombre d'indices i tels que $X_i(\omega) \leq x$, qui est aussi le nombre d'indices j tels que $Y_j(\omega) \leq x$ (les $Y_j(\omega)$ sont un réarrangement des $X_i(\omega)$).
- Si $Y_k(\omega) \leq x$, alors $Y_1(\omega) \leq \dots \leq Y_k(\omega) \leq x$, donc $S_n(x)(\omega) \geq k$.
 - Si $Y_k(\omega) > x$, alors $Y_j(\omega) > x$ pour $j \geq k$, donc au plus $k - 1$ des $Y_j(\omega)$ sont inférieurs ou égaux à x , et $S_n(x)(\omega) \leq k - 1$.

$$[Y_k \leq x] = [S_n(x) \geq k].$$

d) Par c) et b) :

$$F_{Y_k}(x) = \mathbb{P}(S_n(x) \geq k) = \sum_{j=k}^n \binom{n}{j} (F_X(x))^j (1 - F_X(x))^{n-j}.$$

e) On pose, pour $p \in [0, 1]$, $G_k(p) = \sum_{j=k}^n \binom{n}{j} p^j (1-p)^{n-j}$, de sorte que $F_{Y_k} = G_k \circ F_X$. G_k est polynomiale. On utilise $j \binom{n}{j} = n \binom{n-1}{j-1}$ et $(n-j) \binom{n}{j} = n \binom{n-1}{j}$ pour $j \leq n-1$, et on pose $b_j = \binom{n-1}{j} p^j (1-p)^{n-1-j}$ pour $0 \leq j \leq n-1$, et $b_n = 0$. Alors

$$\begin{aligned} G'_k(p) &= \sum_{j=k}^n \binom{n}{j} (j p^{j-1} (1-p)^{n-j} - (n-j) p^j (1-p)^{n-j-1}) \\ &= n \sum_{j=k}^n (b_{j-1} - b_j) = n b_{k-1}. \end{aligned}$$

La somme est télescopique. Comme $n \binom{n-1}{k-1} = k \binom{n}{k}$, on obtient $G'_k(p) = k \binom{n}{k} p^{k-1} (1-p)^{n-k}$.

$F_{Y_k} = G_k \circ F_X$ est continue sur $\mathbb{R}$ et de classe C^1 sur $\mathbb{R} \setminus D$. Pour $x \notin D$, on a $F'_{Y_k}(x) = G'_k(F_X(x)) f_X(x)$. La fonction f_{Y_k} de l'énoncé est positive et égale à F'_{Y_k} hors de D .

$$f_{Y_k}(x) = k \binom{n}{k} (F_X(x))^{k-1} (1 - F_X(x))^{n-k} f_X(x) \text{ définit une densité de } Y_k.$$

f) Comme $0 \leq F_X \leq 1$, on a pour tout réel x

$$0 \leq |x|^r f_{Y_k}(x) \leq k \binom{n}{k} |x|^r f_X(x).$$

Comme X admet un moment d'ordre r , l'intégrale $\int_{-\infty}^{+\infty} |x|^r f_X(x) dx$ converge. Par comparaison d'intégrales de fonctions positives (sur $] -\infty, 0]$ et sur $[0, +\infty[$), $\int_{-\infty}^{+\infty} |x|^r f_{Y_k}(x) dx$ converge.

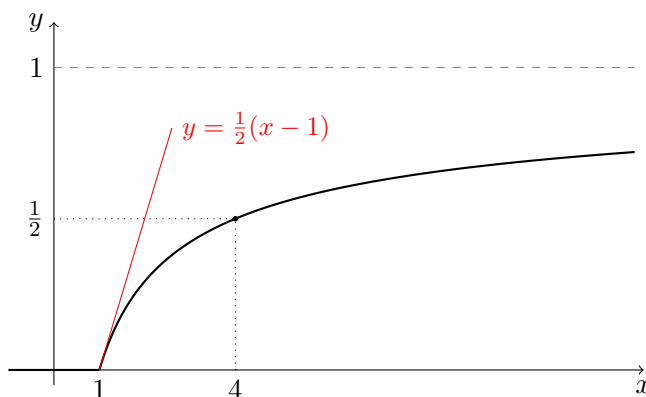
$$Y_k \text{ admet un moment d'ordre } r.$$

2. Exemple.

a) Sur $]1, +\infty[$, F_X est dérivable, avec $F'_X(x) = \frac{1}{2x^{3/2}} > 0$ et $F''_X(x) = -\frac{3}{4x^{5/2}} < 0$: F_X y est strictement croissante et concave. De plus, $F_X(1) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$ (asymptote horizontale $y = 1$). En 1, pour $h > 0$,

$$\frac{F_X(1+h) - F_X(1)}{h} = \frac{1 - (1+h)^{-1/2}}{h} \xrightarrow{h \rightarrow 0^+} \frac{1}{2},$$

car $(1+h)^{-1/2} = 1 - \frac{h}{2} + o(h)$. Donc $(F_X)'_d(1) = \frac{1}{2}$ et la demi-tangente à droite en 1 est la demi-droite d'équation $y = \frac{1}{2}(x-1)$, $x \geq 1$. Par ailleurs, $(F_X)'_g(1) = 0$.



F_X est continue sur $\mathbb{R}$ (en 1 : $F_X(1) = 0 = \lim_{x \rightarrow 1^-} F_X(x)$) et de classe C^1 sur $\mathbb{R} \setminus \{1\}$. X est donc une variable à densité et, en dérivant hors de 1 :

$$f_X(x) = \begin{cases} \frac{1}{2x\sqrt{x}} & \text{si } x \geq 1, \\ 0 & \text{si } x < 1. \end{cases}$$

- b) Soit $r \in \mathbb{N}^*$. f_X est nulle sur $] -\infty, 1[$ et, pour $x \geq 1$, $x^r f_X(x) = \frac{1}{2} x^{r-3/2} \geq \frac{1}{2\sqrt{x}} \geq 0$. Or $\int_1^{+\infty} \frac{dx}{2\sqrt{x}}$ diverge (intégrale de Riemann d'exposant $\frac{1}{2} \leq 1$). Par comparaison de fonctions positives, $\int_1^{+\infty} x^r f_X(x) dx$ diverge.

X n'admet de moment d'aucun ordre $r \in \mathbb{N}^*$ (en particulier, pas d'espérance).

- c) Sur $] -\infty, 1[$, $F_X = 0 \neq \frac{1}{2}$. Sur $[1, +\infty[$, F_X est continue et strictement croissante, et $1 - \frac{1}{\sqrt{x}} = \frac{1}{2} \iff \sqrt{x} = 2 \iff x = 4$.

X a une unique médiane théorique : $M = 4$.

- d) Pour $x \geq 1$, $F_X(x) = 1 - \frac{1}{\sqrt{x}}$ et $1 - F_X(x) = x^{-1/2}$. Par 1e) :

$$f_{Y_k}(x) = \begin{cases} \frac{k}{2} \binom{n}{k} \left(1 - \frac{1}{\sqrt{x}}\right)^{k-1} x^{-\frac{n-k+3}{2}} & \text{si } x \geq 1, \\ 0 & \text{si } x < 1. \end{cases}$$

Comme $\left(1 - \frac{1}{\sqrt{x}}\right)^{k-1} \rightarrow 1$ lorsque $x \rightarrow +\infty$:

$$f_{Y_k}(x) \underset{x \rightarrow +\infty}{\sim} \frac{k}{2} \binom{n}{k} x^{-\frac{n-k+3}{2}}.$$

3. Espérance des statistiques d'ordre ($n \geq 3$).

- a) Soit $k \in \llbracket 1, n-2 \rrbracket$. f_{Y_k} est nulle sur $] -\infty, 1[$ et continue sur $[1, +\infty[$. Il suffit d'étudier $\int_1^{+\infty} x f_{Y_k}(x) dx$, dont l'intégrande est positive. Par 2d),

$$x f_{Y_k}(x) \underset{x \rightarrow +\infty}{\sim} \frac{k}{2} \binom{n}{k} x^{-\frac{n-k+1}{2}}, \quad \text{avec } \frac{n-k+1}{2} \geq \frac{3}{2} > 1.$$

Par le critère de Riemann et le critère d'équivalence pour des fonctions positives, l'intégrale converge.

$$Y_k \text{ admet une espérance pour tout } k \in \llbracket 1, n-2 \rrbracket.$$

- b) On a $\mathbb{E}(Y_k) = \frac{k}{2} \binom{n}{k} \int_1^{+\infty} \left(1 - x^{-1/2}\right)^{k-1} x^{-\frac{n-k+1}{2}} dx$. On pose $x = \frac{1}{t^2}$: l'application $t \mapsto \frac{1}{t^2}$ est de classe C^1 et strictement décroissante, donc bijective, de $]0, 1]$ sur $[1, +\infty[$, avec $dx = -\frac{2}{t^3} dt$. Les deux intégrales sont donc de même nature (ici convergentes) et égales. Avec $x^{-1/2} = t$ et $x^{-\frac{n-k+1}{2}} = t^{n-k+1}$:

$$\mathbb{E}(Y_k) = \frac{k}{2} \binom{n}{k} \int_0^1 (1-t)^{k-1} t^{n-k+1} \frac{2}{t^3} dt.$$

$$\mathbb{E}(Y_k) = k \binom{n}{k} \int_0^1 t^{n-k-2} (1-t)^{k-1} dt.$$

- c) On raisonne par récurrence sur $s \geq 1$, avec la propriété $\mathcal{P}(s)$: « pour tout $r \in \mathbb{N}^*$, $I_{r,s} = \frac{(r-1)!(s-1)!}{(r+s-1)!}$ ».

Initialisation. $I_{r,1} = \int_0^1 t^{r-1} dt = \frac{1}{r} = \frac{(r-1)!0!}{r!}$.

Hérédité. Soit $s \geq 2$ tel que $\mathcal{P}(s-1)$ soit vraie, et soit $r \in \mathbb{N}^*$. On intègre par parties sur le segment $[0, 1]$ avec $u(t) = \frac{t^r}{r}$ et $v(t) = (1-t)^{s-1}$. Ces fonctions sont de classe C^1 sur $[0, 1]$, avec $u'(t) = t^{r-1}$ et $v'(t) = -(s-1)(1-t)^{s-2}$. Le crochet $\left[\frac{t^r(1-t)^{s-1}}{r}\right]_0^1$ est nul, car $r \geq 1$ et $s-1 \geq 1$. D'où

$$\begin{aligned} I_{r,s} &= \frac{s-1}{r} \int_0^1 t^r (1-t)^{s-2} dt = \frac{s-1}{r} I_{r+1,s-1} \\ &= \frac{s-1}{r} \cdot \frac{r!(s-2)!}{(r+s-1)!} = \frac{(r-1)!(s-1)!}{(r+s-1)!}. \end{aligned}$$

$$I_{r,s} = \frac{(r-1)!(s-1)!}{(r+s-1)!}.$$

- d) Par b), $\mathbb{E}(Y_k) = k \binom{n}{k} I_{n-k-1,k}$, avec $n-k-1 \geq 1$. Donc

$$\mathbb{E}(Y_k) = k \cdot \frac{n!}{k!(n-k)!} \cdot \frac{(n-k-2)!(k-1)!}{(n-2)!} = \frac{n!}{(n-2)!} \cdot \frac{(n-k-2)!}{(n-k)!}.$$

$$\mathbb{E}(Y_k) = \frac{n(n-1)}{(n-k)(n-k-1)} \quad (1 \leq k \leq n-2).$$

- e) Comme $n = 2\ell + 1$ est impair, $Y_{\ell+1}$ est la valeur centrale de l'échantillon ordonné : ℓ valeurs la précèdent et ℓ la suivent. De plus, $n \geq 5$ donne $\ell \geq 2$, donc $\ell+1 \leq 2\ell-1 = n-2$. Par a), la médiane empirique $Y_{\ell+1}$ admet une espérance. Par d), avec $n-k = \ell$ et $n-k-1 = \ell-1$:

$$\mathbb{E}(Y_{\ell+1}) = \frac{(2\ell+1)2\ell}{\ell(\ell-1)} = \frac{4\ell+2}{\ell-1} = \frac{4(\ell-1)+6}{\ell-1}.$$

$$\mathbb{E}(Y_{\ell+1}) = 4 + \frac{6}{\ell-1}.$$

Commentaire. $\mathbb{E}(Y_{\ell+1}) > 4 = M$: la médiane empirique est un estimateur biaisé de la médiane théorique M . Son biais $\frac{6}{\ell-1}$ tend vers 0 quand $\ell \rightarrow +\infty$: il est asymptotiquement sans biais. En revanche, X n'a pas d'espérance, donc la moyenne empirique n'en a pas non plus et ne peut pas servir ici.

4. Convergence en loi de $Z_n = Y_n/n^2$.

- a) Soit $x \in \mathbb{R}$. Par 1a), $F_{Z_n}(x) = \mathbb{P}(Y_n \leq n^2x) = (F_X(n^2x))^n$, et $n^2x \geq 1 \iff x \geq \frac{1}{n^2}$.

$$F_{Z_n}(x) = \begin{cases} \left(1 - \frac{1}{n\sqrt{x}}\right)^n & \text{si } x \geq \frac{1}{n^2}, \\ 0 & \text{si } x < \frac{1}{n^2}. \end{cases}$$

- b) φ_Z est nulle sur $]-\infty, 0]$, et de classe C^1 sur $]0, +\infty[$ (composée de fonctions C^1), avec

$$\varphi'_Z(x) = \frac{1}{2x\sqrt{x}} \exp\left(-\frac{1}{\sqrt{x}}\right) > 0 \quad (x > 0).$$

En 0 : $-\frac{1}{\sqrt{x}} \rightarrow -\infty$ quand $x \rightarrow 0^+$, donc $\varphi_Z(x) \rightarrow 0 = \varphi_Z(0)$. Ainsi φ_Z est continue sur $\mathbb{R}$, de classe C^1 sur $\mathbb{R}^*$ et croissante sur $\mathbb{R}$. Enfin, $\lim_{-\infty} \varphi_Z = 0$ et $\lim_{+\infty} \varphi_Z = \exp(0) = 1$. φ_Z est donc une fonction de répartition (croissante, continue à droite, de limites 0 et 1). Comme elle est continue sur $\mathbb{R}$ et de classe C^1 sauf en 0, la variable associée est à densité.

φ_Z est la fonction de répartition d'une variable aléatoire Z à densité.

- c) Soit $x \in \mathbb{R}$.

- Si $x \leq 0$: pour tout n , $F_{Z_n}(x) = 0 = \varphi_Z(x)$.
- Si $x > 0$: pour $n > \frac{1}{\sqrt{x}}$, on a $x > \frac{1}{n^2}$ et $F_{Z_n}(x) = \exp\left(n \ln\left(1 - \frac{1}{n\sqrt{x}}\right)\right)$. Comme $\frac{1}{n\sqrt{x}} \rightarrow 0$ et $\ln(1-v) \underset{v \rightarrow 0}{\sim} -v$, on obtient $n \ln\left(1 - \frac{1}{n\sqrt{x}}\right) \sim -\frac{1}{\sqrt{x}}$. Par continuité de $\exp$, $F_{Z_n}(x) \rightarrow \exp\left(-\frac{1}{\sqrt{x}}\right) = \varphi_Z(x)$.

Ainsi $F_{Z_n}(x) \rightarrow \varphi_Z(x)$ pour tout réel x . Par définition de la convergence en loi :

$$Z_n \xrightarrow{\mathcal{L}} Z.$$

Partie II – Existence et unicité d'un estimateur optimal

Partie II – Estimation de la moyenne d'une loi normale

$X \hookrightarrow \mathcal{N}(\theta, 1)$, Φ fonction de répartition de $\mathcal{N}(0, 1)$.

Corrigé – Partie II.

5. Loi de $\bar{X}_n$.

Les X_k sont indépendantes, de loi $\mathcal{N}(\theta, 1)$. Par stabilité des lois normales, $\sum_{k=1}^n X_k \hookrightarrow$

$\mathcal{N}(n\theta, n)$, puis, par transformation affine, $\bar{X}_n \hookrightarrow \mathcal{N}\left(\theta, \frac{1}{n}\right)$. Ainsi $\mathbb{E}(\bar{X}_n) = \theta$: l'estimateur est sans biais. Pour $\varepsilon > 0$, l'inégalité de Bienaymé-Tchebychev donne

$$\mathbb{P}(|\bar{X}_n - \theta| \geq \varepsilon) \leq \frac{\mathbb{V}(\bar{X}_n)}{\varepsilon^2} = \frac{1}{n\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0.$$

$$\boxed{\bar{X}_n \hookrightarrow \mathcal{N}\left(\theta, \frac{1}{n}\right) \text{ est un estimateur sans biais et convergent de } \theta.}$$

6. Intervalle de confiance.

- a) Φ est dérivable sur $\mathbb{R}$, de dérivée $t \mapsto \frac{1}{\sqrt{2\pi}} e^{-t^2/2} > 0$: elle est continue et strictement croissante sur $\mathbb{R}$, avec $\lim_{-\infty} \Phi = 0$ et $\lim_{+\infty} \Phi = 1$. Par le théorème de la bijection, Φ réalise une bijection de $\mathbb{R}$ sur $]0, 1[$.

$$\boxed{\Phi^{-1} :]0, 1[\rightarrow \mathbb{R} \text{ existe (continue, strictement croissante).}}$$

- b) $T_n = \sqrt{n}(\bar{X}_n - \theta) \hookrightarrow \mathcal{N}(0, 1)$. On pose $c = \Phi^{-1}(1 - \frac{\alpha}{2}) > 0$, car $1 - \frac{\alpha}{2} > \frac{1}{2} = \Phi(0)$. Comme $\Phi(-c) = 1 - \Phi(c)$,

$$\mathbb{P}(-c \leq T_n \leq c) = 2\Phi(c) - 1 = 1 - \alpha,$$

et $[-c \leq T_n \leq c] = \left[\bar{X}_n - \frac{c}{\sqrt{n}} \leq \theta \leq \bar{X}_n + \frac{c}{\sqrt{n}}\right]$. Un intervalle de confiance de θ au risque α centré en $\bar{X}_n$ est donc :

$$\boxed{\left[\bar{X}_n - \frac{\Phi^{-1}(1 - \alpha/2)}{\sqrt{n}}, \bar{X}_n + \frac{\Phi^{-1}(1 - \alpha/2)}{\sqrt{n}}\right]}.$$

Il est centré en $\bar{X}_n$, de demi-longueur $\mu(\alpha) = \frac{c}{\sqrt{n}}$. Or $\Phi(-\Phi^{-1}(\alpha/2)) = 1 - \frac{\alpha}{2}$, donc $c = -\Phi^{-1}(\alpha/2)$.

$$\boxed{\mu(\alpha) = -\frac{\Phi^{-1}(\alpha/2)}{\sqrt{n}}}.$$

- c) Par b), $\mu(\beta) = b\mu(\alpha) \iff \Phi^{-1}(\beta/2) = b\Phi^{-1}(\alpha/2)$. Comme Φ est bijective, cela équivaut à :

$$\boxed{\beta = 2\Phi(b\Phi^{-1}(\alpha/2))}.$$

On a $\Phi^{-1}(\alpha/2) < 0$, car $\frac{\alpha}{2} < \frac{1}{2}$, et $0 < b < 1$, donc $b\Phi^{-1}(\alpha/2) > \Phi^{-1}(\alpha/2)$. La stricte croissance de Φ donne $\frac{\beta}{2} > \frac{\alpha}{2}$.

$$\boxed{\beta > \alpha}.$$

Commentaire. À taille d'échantillon fixée, on ne peut réduire la marge d'erreur (gagner en précision) qu'en acceptant un risque plus grand : précision et risque varient en sens contraire.

7. Estimateur optimal.

- a) $\bar{X}_n = g_n(X_1, \dots, X_n)$ avec $g_n(x_1, \dots, x_n) = \frac{1}{n} \sum x_k$. C'est un estimateur sans biais de θ qui admet une variance (question 5).

$$\boxed{\bar{X}_n \in \mathcal{E}_\theta, \text{ donc } \mathcal{E}_\theta \neq \emptyset.}$$

- b) Soit $U_n \in \mathcal{E}_\theta$. On écrit $U_n = \bar{X}_n + (U_n - \bar{X}_n)$. Par le résultat admis, $\text{Cov}(\bar{X}_n, U_n - \bar{X}_n) = 0$, donc

$$\begin{aligned}\mathbb{V}(U_n) &= \mathbb{V}(\bar{X}_n) + \mathbb{V}(U_n - \bar{X}_n) + 2 \text{Cov}(\bar{X}_n, U_n - \bar{X}_n) \\ &= \mathbb{V}(\bar{X}_n) + \mathbb{V}(U_n - \bar{X}_n) \geq \mathbb{V}(\bar{X}_n).\end{aligned}$$

$\bar{X}_n$ est optimal dans $\mathcal{E}_\theta$.

- c) On écrit $Z_n = z_n(X_1, \dots, X_n)$ et $U_n = g_n(X_1, \dots, X_n)$. Alors $A_n(\lambda) = ((1 - \lambda)z_n + \lambda g_n)(X_1, \dots, X_n)$ est une statistique. Par linéarité, $\mathbb{E}(A_n(\lambda)) = (1 - \lambda)\theta + \lambda\theta = \theta$. Enfin, $A_n(\lambda)$ admet une variance, comme combinaison linéaire de variables qui admettent un moment d'ordre 2. Donc $A_n(\lambda) \in \mathcal{E}_\theta$, et, par bilinéarité de la covariance :

$$\mathbb{V}(A_n(\lambda)) = \mathbb{V}(Z_n) + 2\lambda \text{Cov}(Z_n, U_n - Z_n) + \lambda^2 \mathbb{V}(U_n - Z_n).$$

On note $c = \text{Cov}(Z_n, U_n - Z_n)$ et $v = \mathbb{V}(U_n - Z_n)$. Par optimalité de Z_n , $\mathbb{V}(A_n(\lambda)) \geq \mathbb{V}(Z_n)$, soit $\lambda(2c + \lambda v) \geq 0$ pour tout réel λ . Pour $\lambda > 0$, on a $2c + \lambda v \geq 0$, et $\lambda \rightarrow 0^+$ donne $c \geq 0$. Pour $\lambda < 0$, on a $2c + \lambda v \leq 0$, et $\lambda \rightarrow 0^-$ donne $c \leq 0$.

$$\text{Cov}(Z_n, U_n - Z_n) = 0.$$

- d) Les deux estimateurs étant optimaux, $\mathbb{V}(Z_n) \leq \mathbb{V}(\bar{X}_n)$ et $\mathbb{V}(\bar{X}_n) \leq \mathbb{V}(Z_n)$, donc $\mathbb{V}(Z_n) = \mathbb{V}(\bar{X}_n)$. Par c) avec $U_n = \bar{X}_n$, $\text{Cov}(Z_n, \bar{X}_n) = \mathbb{V}(Z_n)$. On pose $D_n = Z_n - \bar{X}_n$. Alors

$$\mathbb{V}(D_n) = \mathbb{V}(Z_n) - 2 \text{Cov}(Z_n, \bar{X}_n) + \mathbb{V}(\bar{X}_n) = \mathbb{V}(\bar{X}_n) - \mathbb{V}(Z_n) = 0.$$

Une variable de variance nulle est presque sûrement égale à son espérance, ici $\mathbb{E}(D_n) = \theta - \theta = 0$. Donc $\mathbb{P}(D_n = 0) = 1$.

$$Z_n = \bar{X}_n \text{ presque sûrement.}$$

Conclusion. $\bar{X}_n$ est, à une égalité presque sûre près, l'unique estimateur optimal dans $\mathcal{E}_\theta$.

8. La médiane empirique comme estimateur de θ .

- a) $X - \theta \hookrightarrow \mathcal{N}(0, 1)$, donc $F_X(x) = \Phi(x - \theta)$. F_X est une bijection continue strictement croissante de $\mathbb{R}$ sur $]0, 1[$ (6a), donc $F_X(x) = \frac{1}{2}$ a une unique solution. Or $\Phi(0) = \frac{1}{2}$.

$$X \text{ a une unique médiane théorique : } M = \theta.$$

- b) $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}$, d'où :

$$f_X(M) = \frac{1}{\sqrt{2\pi}}.$$

Pour tout réel x , par parité de la densité de $\mathcal{N}(0, 1)$, $\Phi(-y) = 1 - \Phi(y)$ pour tout réel y , donc

$$F_X(2M - x) = \Phi(2\theta - x - \theta) = \Phi(-(x - \theta)) = 1 - \Phi(x - \theta).$$

$$F_X(2M - x) = 1 - F_X(x).$$

Les deux membres sont de classe C^1 sur $\mathbb{R}$; en dérivant, $-f_X(2M - x) = -f_X(x)$.

$$f_X(2M - x) = f_X(x) \quad \text{pour tout réel } x.$$

- c) X admet des moments de tout ordre, donc, par 1f), Y_k et Y_{n-k+1} admettent une espérance. On note $W = 2M - Y_{n-k+1}$. Par la formule de la densité d'une transformation affine, W admet la densité $x \mapsto f_{Y_{n-k+1}}(2M - x)$. Par 1e) et b) :

$$f_{Y_{n-k+1}}(2M - x) = (n - k + 1) \binom{n}{n - k + 1} (1 - F_X(x))^{n-k} (F_X(x))^{k-1} f_X(x).$$

Or $(n - k + 1) \binom{n}{n - k + 1} = \frac{n!}{(k-1)!(n-k)!} = k \binom{n}{k}$. Donc $f_{Y_{n-k+1}}(2M - x) = f_{Y_k}(x)$: W et Y_k ont la même densité, donc la même loi et la même espérance. Ainsi $\mathbb{E}(Y_k) = 2M - \mathbb{E}(Y_{n-k+1})$.

$$\boxed{\mathbb{E}(Y_k - M) = \mathbb{E}(M - Y_{n-k+1})}.$$

- d) Pour $n = 2\ell + 1$ et $k = \ell + 1$, on a $n - k + 1 = \ell + 1$. Par c), $\mathbb{E}(Y_{\ell+1}) - M = M - \mathbb{E}(Y_{\ell+1})$.

$$\boxed{\mathbb{E}(Y_{\ell+1}) = M = \theta}.$$

$Y_{\ell+1} = \psi_{\ell+1}(X_1, \dots, X_n)$ est une statistique, sans biais pour θ . Elle admet une variance par 1f) avec $r = 2$, car X admet un moment d'ordre 2. Donc $Y_{\ell+1} \in \mathcal{E}_\theta$, et, par 7b),

$$\boxed{\mathbb{V}(Y_{\ell+1}) \geq \mathbb{V}(\bar{X}_n) = \frac{1}{n}}.$$

Commentaire. La médiane empirique est, comme $\bar{X}_n$, un estimateur sans biais de θ , mais elle est moins précise : pour estimer θ , on lui préfère $\bar{X}_n$.

Partie III – Résultats asymptotiques

Partie III – Loi asymptotique de la médiane empirique

$T \hookrightarrow \mathcal{N}(0, 1)$, $L_U(s) = \mathbb{E}(\exp(sU))$, $y_n = M + \frac{x}{\sqrt{n}}$, $q_n = F_X(y_n)$, $k(n) = [n/2] + 1$.

Corrigé – Partie III.

9. Calculs de L_U .

- a) e^{sJ} prend la valeur 1 avec probabilité $1 - p$ et la valeur e^s avec probabilité p .

$$\boxed{L_J(s) = 1 - p + p e^s}.$$

- b) Par le théorème de transfert, $\mathbb{E}(e^{sT})$ existe si et seulement si $\int_{-\infty}^{+\infty} e^{st} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$ converge (intégrande positive et continue). Par croissances comparées, $t^2 e^{st-t^2/2} \rightarrow 0$ quand $t \rightarrow \pm\infty$. Donc $e^{st-t^2/2} = o\left(\frac{1}{t^2}\right)$ en $\pm\infty$, et le critère de Riemann donne la convergence.

$$\boxed{L_T(s) \text{ existe pour tout réel } s}.$$

- c) On a $st - \frac{t^2}{2} = -\frac{(t-s)^2}{2} + \frac{s^2}{2}$, donc

$$L_T(s) = e^{s^2/2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-(t-s)^2/2} dt = e^{s^2/2},$$

car on intègre sur $\mathbb{R}$ une densité de la loi $\mathcal{N}(s, 1)$.

$$L_T(s) = \exp\left(\frac{s^2}{2}\right).$$

Pour $(\theta, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$, $\exp(s(\sigma T + \theta)) = e^{\theta s} \exp((\sigma s)T)$ admet une espérance par b). On a donc $L_{\sigma T + \theta}(s) = e^{\theta s} L_T(\sigma s)$.

$$L_{\sigma T + \theta}(s) = \exp\left(\frac{\sigma^2 s^2}{2} + \theta s\right).$$

10. Développements asymptotiques.

- a) Par définition de la partie entière, $[n/2] \leq \frac{n}{2} < [n/2] + 1$, donc $0 < k(n) - \frac{n}{2} \leq 1$.
On en déduit $0 < \frac{k(n) - n/2}{\sqrt{n}} \leq \frac{1}{\sqrt{n}}$. Par le théorème d'encadrement, $\frac{k(n) - n/2}{\sqrt{n}} \rightarrow 0$.

$$k(n) = \frac{n}{2} + o(\sqrt{n}).$$

- b) F_X est de classe C^1 sur $\mathbb{R}$, de dérivée f_X , avec $F_X(M) = \frac{1}{2}$ et $f_X(M) = \frac{1}{\sqrt{2\pi}}$ (8a, 8b). Par la formule de Taylor-Young à l'ordre 1 en M , il existe une fonction ε telle que $\varepsilon(h) \rightarrow 0$ quand $h \rightarrow 0$ et, pour tout réel h ,

$$F_X(M + h) = \frac{1}{2} + \frac{h}{\sqrt{2\pi}} + h \varepsilon(h).$$

Avec $h = \frac{x}{\sqrt{n}} \rightarrow 0$, on obtient $q_n = \frac{1}{2} + \frac{x}{\sqrt{2\pi n}} + \frac{x}{\sqrt{n}} \varepsilon\left(\frac{x}{\sqrt{n}}\right)$. De plus, $\sqrt{n} \cdot \frac{x}{\sqrt{n}} \varepsilon\left(\frac{x}{\sqrt{n}}\right) = x \varepsilon\left(\frac{x}{\sqrt{n}}\right) \rightarrow 0$.

$$q_n = \frac{1}{2} + \frac{x}{\sqrt{2\pi n}} + o\left(\frac{1}{\sqrt{n}}\right).$$

- c) Par a) et b),

$$u_n = \frac{k(n) - \frac{n}{2}}{\sqrt{n}} - \sqrt{n} \left(q_n - \frac{1}{2} \right) = o(1) - \left(\frac{x}{\sqrt{2\pi}} + o(1) \right).$$

$$(u_n) \text{ converge vers } u = -\frac{x}{\sqrt{2\pi}}.$$

11. Fonction L_{W_n} .

- a) Par 1b), $S_n(y_n) = \sum_{k=1}^n J_k(y_n)$, où les $J_k(y_n)$ sont indépendantes, de loi de Bernoulli de paramètre $q_n = F_X(y_n) \in]0, 1[$. W_n est une variable finie, donc $\exp(sW_n)$ admet une espérance, et

$$\exp(sW_n) = \exp(-s\sqrt{n} q_n) \prod_{k=1}^n \exp\left(\frac{s}{\sqrt{n}} J_k(y_n)\right).$$

Les variables $\exp\left(\frac{s}{\sqrt{n}} J_k(y_n)\right)$ sont indépendantes (lemme des coalitions). L'espérance de leur produit est donc le produit de leurs espérances (résultat admis, étendu par récurrence à n variables). Chacune vaut $L_{J_k(y_n)}\left(\frac{s}{\sqrt{n}}\right) = 1 - q_n + q_n \exp\left(\frac{s}{\sqrt{n}}\right)$ par 9a).

$$L_{W_n}(s) = \exp(-s\sqrt{n} q_n) \left(1 + q_n \exp\left(\frac{s}{\sqrt{n}}\right) - q_n \right)^n.$$

- b) On pose $v_n = q_n \left(\exp \left(\frac{s}{\sqrt{n}} \right) - 1 \right) > -1$, de sorte que $\ln L_{W_n}(s) = -s\sqrt{n} q_n + n \ln(1 + v_n)$. Comme (q_n) est bornée et $e^w = 1 + w + \frac{w^2}{2} + o(w^2)$ quand $w \rightarrow 0$:

$$v_n = \frac{q_n s}{\sqrt{n}} + \frac{q_n s^2}{2n} + o\left(\frac{1}{n}\right), \quad v_n^2 = \frac{q_n^2 s^2}{n} + o\left(\frac{1}{n}\right).$$

En particulier, $v_n \rightarrow 0$ et $v_n^2 = O\left(\frac{1}{n}\right)$. Le développement limité à l'ordre 2, $\ln(1 + v) = v - \frac{v^2}{2} + o(v^2)$, donne

$$n \ln(1 + v_n) = q_n s \sqrt{n} + \frac{q_n s^2}{2} - \frac{q_n^2 s^2}{2} + o(1),$$

donc

$$\ln L_{W_n}(s) = \frac{q_n(1 - q_n)s^2}{2} + o(1).$$

Par 10b), $q_n \rightarrow \frac{1}{2}$, donc $q_n(1 - q_n) \rightarrow \frac{1}{4}$.

$$\ln L_{W_n}(s) = \frac{s^2}{8} + o(1).$$

Par continuité de $\exp$, $L_{W_n}(s) \rightarrow \exp\left(\frac{s^2}{8}\right)$, et par 9c) avec $\sigma = \frac{1}{2}$, $\theta = 0$, $L_{T/2}(s) = \exp\left(\frac{s^2}{8}\right)$.

$$\lim_{n \rightarrow +\infty} L_{W_n}(s) = L_{T/2}(s).$$

12. Le cas $x = 0$.

Si $x = 0$, alors $y_n = M$ et $q_n = F_X(M) = \frac{1}{2}$ pour tout n . Ainsi $S_n(M) = \sum_{k=1}^n J_k(M)$ est la somme de n variables indépendantes de même loi de Bernoulli de paramètre $\frac{1}{2}$, d'espérance $\frac{1}{2}$ et de variance $\frac{1}{4}$. On a

$$2W_n = \frac{S_n(M) - n \cdot \frac{1}{2}}{\sqrt{n} \cdot \frac{1}{2}}.$$

Par le théorème central limite, $2W_n \xrightarrow{\mathcal{L}} T$. Par composition par la fonction continue $t \mapsto \frac{t}{2}$, $W_n \xrightarrow{\mathcal{L}} \frac{T}{2}$.

Indépendance des $J_k(M)$, de même loi $\mathcal{B}(1/2)$, théorème central limite et composition par une fonction continue donnent directement $W_n \xrightarrow{\mathcal{L}} \frac{T}{2}$.

(Le calcul de L_{W_n} est aussi immédiat : $L_{W_n}(s) = \left(\frac{e^{-s/(2\sqrt{n})} + e^{s/(2\sqrt{n})}}{2} \right)^n$, sans recours à la formule de Taylor-Young pour q_n .)

13. Loi limite de la médiane empirique.

- a) Soit $n \in \mathbb{N}^*$. On a $1 \leq k(n) \leq n$, car $[n/2] \leq n - 1$. De plus, $\sqrt{n}(Y_{k(n)} - M) \leq x \iff Y_{k(n)} \leq y_n$. Par 1c), $[Y_{k(n)} \leq y_n] = [S_n(y_n) \geq k(n)]$. Enfin, en retranchant nq_n et en divisant par $\sqrt{n} > 0$, on a $S_n(y_n) \geq k(n) \iff W_n \geq u_n$.

$$[\sqrt{n}(Y_{k(n)} - M) \leq x] = [S_n(y_n) \geq k(n)] = [W_n \geq u_n].$$

- b) On note G la fonction de répartition de $T/2 : G(t) = \Phi(2t)$, continue sur $\mathbb{R}$. Par hypothèse, $F_{W_n}(t) \rightarrow G(t)$ pour tout réel t . Soit $\varepsilon > 0$. Par continuité de G en u , il existe $\delta > 0$ tel que $G(u + \delta) \leq G(u) + \varepsilon$ et $G(u - \delta) \geq G(u) - \varepsilon$. Comme $u_n \rightarrow u$ (10c), il existe N_1 tel que $|u_n - u| < \delta$ pour $n \geq N_1$. Alors $[W_n \leq u - \delta] \subset [W_n < u_n] \subset [W_n \leq u + \delta]$, donc

$$F_{W_n}(u - \delta) \leq \mathbb{P}(W_n < u_n) \leq F_{W_n}(u + \delta).$$

Par convergence en loi, il existe N_2 tel que, pour $n \geq N_2$, $F_{W_n}(u + \delta) \leq G(u + \delta) + \varepsilon$ et $F_{W_n}(u - \delta) \geq G(u - \delta) - \varepsilon$. Pour $n \geq \max(N_1, N_2)$, on obtient $G(u) - 2\varepsilon \leq \mathbb{P}(W_n < u_n) \leq G(u) + 2\varepsilon$. Ainsi $\mathbb{P}(W_n < u_n) \rightarrow G(u)$, et

$$\mathbb{P}(W_n \geq u_n) \rightarrow 1 - \Phi(2u) = \Phi(-2u) = \Phi\left(\frac{2x}{\sqrt{2\pi}}\right) = \Phi\left(x\sqrt{\frac{2}{\pi}}\right) = \mathbb{P}\left(\sqrt{\frac{\pi}{2}}T \leq x\right).$$

$$\lim_{n \rightarrow +\infty} \mathbb{P}([W_n \geq u_n]) = \mathbb{P}\left(\left[\sqrt{\frac{\pi}{2}}T \leq x\right]\right).$$

- c) Soit $t \in \mathbb{R}$. On applique a) et b) avec $x = t\sqrt{\frac{\pi}{2}}$:

$$\begin{aligned} \mathbb{P}\left(\sqrt{\frac{2n}{\pi}}(Y_{k(n)} - M) \leq t\right) &= \mathbb{P}\left(\sqrt{n}(Y_{k(n)} - M) \leq t\sqrt{\frac{\pi}{2}}\right) \\ &\xrightarrow{n \rightarrow +\infty} \mathbb{P}\left(\sqrt{\frac{\pi}{2}}T \leq t\sqrt{\frac{\pi}{2}}\right) = \Phi(t). \end{aligned}$$

Par définition de la convergence en loi :

$$\sqrt{\frac{2n}{\pi}}(Y_{k(n)} - M) \xrightarrow{\mathcal{L}} T \hookrightarrow \mathcal{N}(0, 1).$$

14. Corrélation entre médiane et moyenne empiriques.

- a) $\frac{n}{2} = \ell + \frac{1}{2}$, donc $[n/2] = \ell$.

$$k(n) = \ell + 1 : Y_{k(n)} \text{ est la médiane empirique.}$$

- b) Par 8d), $\mathbb{E}(Y_{k(n)}) = M = \theta$ pour tout n impair.

$$\lim_{n \rightarrow +\infty} \mathbb{E}(Y_{k(n)}) = \theta.$$

- c) Comme $\mathbb{E}(Y_{k(n)}) = M$, on a $\mathbb{E}\left(\left(\sqrt{\frac{2n}{\pi}}(Y_{k(n)} - M)\right)^2\right) = \frac{2n}{\pi}\mathbb{V}(Y_{k(n)})$. Cette quantité tend vers $\mathbb{E}(T^2) = \mathbb{V}(T) = 1$.

$$\mathbb{V}(Y_{k(n)}) \underset{n \rightarrow +\infty}{\sim} \frac{\pi}{2n}.$$

- d) Par 8d), $Y_{k(n)} \in \mathcal{E}_\theta$. Le résultat admis à la question 7 donne $\text{Cov}(\bar{X}_n, Y_{k(n)} - \bar{X}_n) = 0$, d'où, avec la question 5, $\text{Cov}(Y_{k(n)}, \bar{X}_n) = \mathbb{V}(\bar{X}_n) = \frac{1}{n}$. Donc

$$\rho_n = \frac{\text{Cov}(Y_{k(n)}, \bar{X}_n)}{\sqrt{\mathbb{V}(Y_{k(n)})\mathbb{V}(\bar{X}_n)}} = \frac{1/n}{\sqrt{\mathbb{V}(Y_{k(n)})/n}} = \frac{1}{\sqrt{n\mathbb{V}(Y_{k(n)})}}.$$

Par c), $n\mathbb{V}(Y_{k(n)}) \rightarrow \frac{\pi}{2}$.

$$\lim_{n \rightarrow +\infty} \rho_n = \sqrt{\frac{2}{\pi}} \approx 0,80.$$