



PRÉPASIMPLE

prepasimple.fr

HEC — 2011

Corrigé de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Corrigé

Corrigé

Rappel utilisé dans tout le corrigé : si une fonction G est continue sur $\mathbb{R}$ et de classe C^1 sur $\mathbb{R}$ privé d'un nombre fini de points, et si G est la fonction de répartition d'une variable aléatoire Z , alors Z est à densité et toute fonction positive égale à G' sauf en un nombre fini de points est une densité de Z . Pour la variable X de densité f_X , la fonction F_X est continue sur $\mathbb{R}$ et de classe C^1 sauf en un nombre fini de points, avec $F'_X = f_X$ sauf peut-être en un nombre fini de points.

Partie I : Quelques propriétés des statistiques d'ordre

- 1) a) Pour x réel, $[Y_n \leq x] = \bigcap_{i=1}^n [X_i \leq x]$ (le plus grand des X_i est inférieur à x si et seulement si tous le sont) et $[Y_1 > x] = \bigcap_{i=1}^n [X_i > x]$. Par indépendance des X_i , qui ont tous la loi de X :

$$F_{Y_n}(x) = (F_X(x))^n \quad \text{et} \quad F_{Y_1}(x) = 1 - (1 - F_X(x))^n$$

Ces fonctions sont continues sur $\mathbb{R}$, et de classe C^1 là où F_X l'est, c'est-à-dire sauf en un nombre fini de points ; en ces points de dérivabilité, par dérivation d'une composée :

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

Les fonctions f_{Y_1} et f_{Y_n} de l'énoncé sont positives et coïncident avec ces dérivées sauf en un nombre fini de points.

Conclusion

f_{Y_1} et f_{Y_n} sont des densités de Y_1 et Y_n .

- b) Les variables $J_1(x), \dots, J_n(x)$ sont mutuellement indépendantes (chaque $J_k(x)$ est une fonction de X_k), et suivent la loi de Bernoulli de paramètre $P([X_k \leq x]) = F_X(x)$.

Conclusion

$S_n(x)$ suit la loi binomiale de paramètres n et $F_X(x)$.

- c) Soit $\omega \in \Omega$. Comme $(Y_1(\omega), \dots, Y_n(\omega))$ est un réarrangement de $(X_1(\omega), \dots, X_n(\omega))$, le nombre $S_n(x)(\omega)$ d'indices i tels que $X_i(\omega) \leq x$ est égal au nombre d'indices j tels que $Y_j(\omega) \leq x$. Comme la suite $(Y_j(\omega))_j$ est croissante, ces indices j forment un ensemble $\llbracket 1, m \rrbracket$, avec $m = S_n(x)(\omega)$ (ensemble vide si $m = 0$). Donc
- $$Y_k(\omega) \leq x \iff k \in \llbracket 1, m \rrbracket \iff S_n(x)(\omega) \geq k.$$

Conclusion

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

- d) D'après c) et b) :

Conclusion

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

- e) La fonction F_{Y_k} est un polynôme en $F_X(x)$: elle est continue sur $\mathbb{R}$ et de classe C^1 là où F_X l'est. En un tel point, en notant $F = F_X(x)$ et $f = f_X(x)$:

$$F'_{Y_k}(x) = f \sum_{j=k}^n \binom{n}{j} [jF^{j-1}(1-F)^{n-j} - (n-j)F^j(1-F)^{n-j-1}]$$

où le terme en $(n-j)$ est nul pour $j = n$. Avec $j \binom{n}{j} = n \binom{n-1}{j-1}$ et $(n-j) \binom{n}{j} = n \binom{n-1}{j}$, et en posant $a_j = \binom{n-1}{j} F^j (1-F)^{n-1-j}$ pour $0 \leq j \leq n-1$, $a_n = 0$:

$$F'_{Y_k}(x) = nf \sum_{j=k}^n (a_{j-1} - a_j) = nf a_{k-1} = n \binom{n-1}{k-1} F^{k-1} (1-F)^{n-k} f$$

(somme télescopique). Enfin $n \binom{n-1}{k-1} = \frac{n!}{(k-1)!(n-k)!} = k \binom{n}{k}$. La fonction f_{Y_k} de l'énoncé est positive et égale à F'_{Y_k} sauf en un nombre fini de points.

Conclusion

f_{Y_k} est une densité de Y_k .

f) Supposons que X admet un moment d'ordre r : l'intégrale $\int_{-\infty}^{+\infty} |x|^r f_X(x) dx$ converge. Comme $0 \leq F_X \leq 1$, pour tout réel x :

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

Par comparaison d'intégrales de fonctions positives, $\int_{-\infty}^{+\infty} |x|^r f_{Y_k}(x) dx$ converge.

Conclusion

pour tout $k \in \llbracket 1, n \rrbracket$, Y_k admet un moment d'ordre r .

2) a) Étude de F_X . Sur $] -\infty, 1[$, F_X est nulle. Sur $[1, +\infty[$, $F_X(x) = 1 - x^{-1/2}$ est dérivable, de dérivée $\frac{1}{2}x^{-3/2} > 0$: F_X y est strictement croissante, de $F_X(1) = 0$ à la limite 1 en $+\infty$ (asymptote horizontale $y = 1$). Sa dérivée seconde $-\frac{3}{4}x^{-5/2} < 0$ montre que F_X est concave sur $[1, +\infty[$. En 1, F_X est continue (limite à gauche $0 = F_X(1)$), la dérivée à gauche vaut 0 et la dérivée à droite vaut $\frac{1}{2}$: la demi-tangente à droite au point $(1, 0)$ a pour équation $y = \frac{1}{2}(x - 1)$, et la courbe présente un point anguleux.

La courbe est donc : la demi-droite $y = 0$ pour $x < 1$, puis un arc croissant et concave partant de $(1, 0)$ avec la pente $\frac{1}{2}$, passant par $(4, \frac{1}{2})$ et $(9, \frac{2}{3})$, et s'approchant de l'asymptote $y = 1$ par en dessous. Le code suivant, exécuté, trace cette courbe avec la demi-tangente et l'asymptote :

Python

```
import numpy as np

import matplotlib.pyplot as plt

x = np.linspace(-2, 30, 2000)

F = np.zeros(len(x))

for i in range(len(x)):
    if x[i] >= 1:
        F[i] = 1 - 1 / np.sqrt(x[i])

plt.plot(x, F)                                # courbe de F_X
plt.plot([1, 3], [0, 1], '--')               # demi-tangente y = (x-1)/2
plt.plot([-2, 30], [1, 1], ':')              # asymptote y = 1
plt.show()
```

Caractère à densité. F_X est croissante, continue sur $\mathbb{R}$, de limites 0 en $-\infty$ et 1 en $+\infty$: c'est la fonction de répartition de X . Elle est de classe C^1 sur $\mathbb{R} \setminus \{1\}$. Donc X est à densité.

Conclusion

la demi-tangente à droite en 1 est $y = \frac{x-1}{2}$, X est à densité et une densité est $f_X(x) = \frac{1}{2x^{3/2}}$ si $x > 1$, $f_X(x) = 0$ si $x \leq 1$.

b) Soit $r \in \mathbb{N}^*$. La fonction $x \mapsto |x|^r f_X(x)$ est nulle sur $] -\infty, 1]$ et vaut $\frac{1}{2}x^{r-3/2}$ sur $]1, +\infty[$. Comme $r - \frac{3}{2} \geq -\frac{1}{2} > -1$, l'intégrale de Riemann $\int_1^{+\infty} x^{r-3/2} dx$ diverge.

Conclusion

X n'admet aucun moment (pas même d'espérance).

c) Pour $x < 1$, $F_X(x) = 0 \neq \frac{1}{2}$. Sur $]1, +\infty[$, $F_X(x) = \frac{1}{2} \iff \frac{1}{\sqrt{x}} = \frac{1}{2} \iff x = 4$.

Conclusion

X admet une unique médiane théorique, $M = 4$.

d) D'après 1) e), pour $x > 1$:

$$f_{Y_k}(x) = k \binom{n}{k} \left(1 - \frac{1}{\sqrt{x}}\right)^{k-1} \left(\frac{1}{\sqrt{x}}\right)^{n-k} \frac{1}{2x^{3/2}} = \frac{k}{2} \binom{n}{k} \left(1 - \frac{1}{\sqrt{x}}\right)^{k-1} x^{-\frac{n-k+3}{2}}$$

et $f_{Y_k}(x) = 0$ pour $x \leq 1$. Lorsque $x \rightarrow +\infty$, $\left(1 - \frac{1}{\sqrt{x}}\right)^{k-1} \rightarrow 1$.

Conclusion

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

3) a) Soit $k \in \llbracket 1, n-2 \rrbracket$. La fonction $x \mapsto x f_{Y_k}(x)$ est positive, nulle sur $] -\infty, 1]$, continue sur $]1, +\infty[$ et prolongeable par continuité en 1. D'après 2) d),

$x f_{Y_k}(x) \sim \frac{k}{2} \binom{n}{k} x^{-\frac{n-k+1}{2}}$ en $+\infty$, avec $\frac{n-k+1}{2} \geq \frac{3}{2} > 1$. Par comparaison à une intégrale de Riemann convergente, $\int_1^{+\infty} x f_{Y_k}(x) dx$ converge (absolument, la fonction étant positive).

Conclusion

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

b) La fonction $t \mapsto t^{-2}$ est de classe C^1 et strictement décroissante sur $]0, 1[$, et réalise une bijection de $]0, 1[$ sur $]1, +\infty[$; sa réciproque est $x \mapsto \frac{1}{\sqrt{x}}$. On justifie le changement de variable sur un segment puis par passage à la limite. Pour $1 < a < A$, avec $x = t^{-2}$, $dx = -2t^{-3} dt$, en notant $c = \frac{k}{2} \binom{n}{k}$:

$$\int_a^A x f_{Y_k}(x) dx = c \int_a^A (1 - x^{-1/2})^{k-1} x^{-\frac{n-k+1}{2}} dx = c \int_{1/\sqrt{A}}^{1/\sqrt{a}} (1-t)^{k-1} t^{n-k+1} \cdot 2t^{-3} dt = k \binom{n}{k} \int_{1/\sqrt{A}}^{1/\sqrt{a}} t^{n-k-2} (1-t)^{k-1} dt$$

(l'échange des bornes compense le signe de dx). Quand $a \rightarrow 1^+$ et $A \rightarrow +\infty$, le membre de gauche tend vers $E(Y_k)$ (question a)) ; dans le membre de droite, la fonction $t \mapsto t^{n-k-2}(1-t)^{k-1}$ est continue sur $[0, 1]$ car $n-k-2 \geq 0$, et les bornes tendent vers 0 et 1.

Conclusion

$$E(Y_k) = k \binom{n}{k} \int_0^1 t^{n-k-2} (1-t)^{k-1} dt \text{ pour tout } k \in \llbracket 1, n-2 \rrbracket.$$

c) On montre par récurrence sur $s \in \mathbb{N}^*$ la propriété : pour tout $r \in \mathbb{N}^*$,

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

- Pour $s = 1$: $I_{r,1} = \int_0^1 t^{r-1} dt = \frac{1}{r} = \frac{(r-1)! 0!}{r!}.$
- Si la propriété est vraie au rang s , soit $r \in \mathbb{N}^*$. Par intégration par parties sur le segment $[0, 1]$, avec $u(t) = (1-t)^s$ et $v(t) = \frac{t^r}{r}$, fonctions de classe C^1 :

$$I_{r,s+1} = \int_0^1 t^{r-1} (1-t)^s dt = \left[\frac{t^r}{r} (1-t)^s \right]_0^1 + \frac{s}{r} \int_0^1 t^r (1-t)^{s-1} dt = \frac{s}{r} I_{r+1,s}$$

car le crochet est nul ($s \geq 1, r \geq 1$). Par hypothèse de récurrence appliquée à $r+1$:

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

ce qui est la propriété au rang $s+1$.

Conclusion

$$I_{r,s} = \frac{(r-1)!(s-1)!}{(r+s-1)!} \text{ pour tout } (r, s) \in (\mathbb{N}^*)^2.$$

d) Pour $k \in \llbracket 1, n-2 \rrbracket$, l'intégrale de b) est $I_{n-k-1,k}$ avec $n-k-1 \geq 1$ et $k \geq 1$. Donc :

$$\begin{aligned} E(Y_k) &= k \binom{n}{k} \frac{(n-k-2)!(k-1)!}{(n-2)!} = \frac{n!}{(k-1)!(n-k)!} \cdot \frac{(n-k-2)!(k-1)!}{(n-2)!} \\ &= \frac{n(n-1)}{(n-k)(n-k-1)} \end{aligned}$$

Conclusion

$$E(Y_k) = \frac{n(n-1)}{(n-k)(n-k-1)} \text{ pour tout } k \in \llbracket 1, n-2 \rrbracket.$$

e) Comme $n = 2\ell + 1$, l'indice $\ell + 1$ appartient à $\llbracket 1, n \rrbracket$: la médiane empirique $Y_{\ell+1}$ est bien définie ; c'est la valeur centrale de l'échantillon ordonné, avec ℓ valeurs de part et d'autre. Comme $n \geq 5$, $\ell \geq 2$ et $\ell + 1 \leq 2\ell - 1 = n - 2$: d'après a) et d), $Y_{\ell+1}$ admet une espérance, avec $n - k = \ell$ et $n - k - 1 = \ell - 1$:

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

Commentaire : la variable X n'a pas d'espérance, de sorte que la moyenne empirique n'a pas d'espérance non plus et ne peut pas servir à estimer un « centre » de la loi. La médiane empirique, elle, admet une espérance dès que $n \geq 5$, et $E(Y_{\ell+1}) \rightarrow 4 = M$ quand $\ell \rightarrow +\infty$: c'est un estimateur asymptotiquement sans biais de la médiane théorique (biaisé pour chaque ℓ , avec un biais $\frac{6}{\ell-1} > 0$).

Conclusion

$$E(Y_{\ell+1}) = 4 + \frac{6}{\ell-1}, \text{ qui tend vers la médiane théorique } M = 4.$$

- 4) a) Pour x réel, $[Z_n \leq x] = [Y_n \leq n^2 x]$, donc $F_{Z_n}(x) = F_{Y_n}(n^2 x) = (F_X(n^2 x))^n$ (question 1) a)).

Conclusion

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

- b) Soit f la fonction définie par $f(x) = \frac{1}{2}x^{-3/2} \exp\left(-\frac{1}{\sqrt{x}}\right)$ pour $x > 0$ et $f(x) = 0$ pour $x \leq 0$. Elle est positive, continue sur $\mathbb{R}^*$. Pour $0 < a < A$, comme $\frac{d}{dx} \exp(-x^{-1/2}) = \frac{1}{2}x^{-3/2} \exp(-x^{-1/2})$:

$$\int_a^A f(x) dx = \exp\left(-\frac{1}{\sqrt{A}}\right) - \exp\left(-\frac{1}{\sqrt{a}}\right)$$

qui tend vers $1 - 0 = 1$ quand $a \rightarrow 0^+$ et $A \rightarrow +\infty$. Donc f est une densité de probabilité, et d'après le cours il existe une variable aléatoire Z de densité f . Sa fonction de répartition vaut 0 sur $]-\infty, 0]$ et, pour $x > 0$,

$$\int_0^x f = \lim_{a \rightarrow 0^+} (\exp(-x^{-1/2}) - \exp(-a^{-1/2})) = \exp\left(-\frac{1}{\sqrt{x}}\right).$$

Conclusion

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

- c) La fonction φ_Z est continue sur $\mathbb{R}$ (en 0, $\exp(-x^{-1/2}) \rightarrow 0$ quand $x \rightarrow 0^+$). Il suffit donc de montrer que $F_{Z_n}(x) \rightarrow \varphi_Z(x)$ pour tout réel x .

- Si $x \leq 0$: $F_{Z_n}(x) = 0 = \varphi_Z(x)$ pour tout n .
- Si $x > 0$: pour $n \geq \frac{1}{\sqrt{x}}$, on a $x \geq \frac{1}{n^2}$ 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 $\ln(1+h) \sim h$ en 0, $n \ln\left(1 - \frac{1}{n\sqrt{x}}\right) \sim n \cdot \left(-\frac{1}{n\sqrt{x}}\right) = -\frac{1}{\sqrt{x}}$. Par continuité de l'exponentielle, $F_{Z_n}(x) \rightarrow \exp\left(-\frac{1}{\sqrt{x}}\right)$.

Conclusion

$(Z_n)_{n \in \mathbb{N}^*}$ converge en loi vers Z .

Partie II : Existence et unicité d'un estimateur optimal

- 5) Les X_k sont indépendantes et de loi $\mathcal{N}(\theta, 1)$; par stabilité des lois normales indépendantes,

$$\sum_{k=1}^n X_k$$

suit la loi $\mathcal{N}(n\theta, n)$, puis par transformation affine, $\bar{X}_n$ suit la loi $\mathcal{N}\left(\theta, \frac{1}{n}\right)$.

$\bar{X}_n$ est une fonction de l'échantillon ne dépendant pas de θ : c'est un estimateur de θ , et $E(\bar{X}_n) = \theta$: il est sans biais. Pour $\varepsilon > 0$, l'inégalité de Bienaymé-Tchebychev donne

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

Conclusion

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

- 6) a) Φ est continue sur $\mathbb{R}$, de classe C^1 de dérivée $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2} > 0$, donc strictement croissante, avec $\Phi(x) \rightarrow 0$ en $-\infty$ et $\Phi(x) \rightarrow 1$ en $+\infty$. Par le théorème de la bijection :

Conclusion

Φ réalise une bijection de $\mathbb{R}$ sur $]0, 1[$, et Φ^{-1} existe (continue, strictement croissante de $]0, 1[$ sur $\mathbb{R}$).

- b) D'après 5), $\sqrt{n}(\bar{X}_n - \theta)$ suit la loi normale centrée réduite. Pour $c > 0$, par parité de la densité, $P(|\sqrt{n}(\bar{X}_n - \theta)| \leq c) = \Phi(c) - \Phi(-c) = 2\Phi(c) - 1$. Cette probabilité vaut $1 - \alpha$ si et seulement si $\Phi(c) = 1 - \frac{\alpha}{2}$, c'est-à-dire $c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$. Comme $\Phi(-y) = 1 - \Phi(y)$, on a $\Phi(-\Phi^{-1}(\alpha/2)) = 1 - \frac{\alpha}{2}$, donc $c = -\Phi^{-1}(\alpha/2)$, qui est strictement positif car $\frac{\alpha}{2} < \frac{1}{2} = \Phi(0)$. Ainsi :

$$P\left(\left[\bar{X}_n - \frac{c}{\sqrt{n}} \leq \theta \leq \bar{X}_n + \frac{c}{\sqrt{n}}\right]\right) = 1 - \alpha$$

Conclusion

$\left[\bar{X}_n + \frac{\Phi^{-1}(\alpha/2)}{\sqrt{n}}, \bar{X}_n - \frac{\Phi^{-1}(\alpha/2)}{\sqrt{n}}\right]$ est un intervalle de confiance de θ au risque α , de milieu $\bar{X}_n$, et $\mu(\alpha) = -\frac{\Phi^{-1}(\alpha/2)}{\sqrt{n}}$.

- c) La condition $\mu(\beta) = b\mu(\alpha)$ s'écrit $\Phi^{-1}(\beta/2) = b\Phi^{-1}(\alpha/2)$, soit $\beta = 2\Phi(b\Phi^{-1}(\alpha/2))$. Comme $\Phi^{-1}(\alpha/2) < 0$ et $0 < b < 1$, on a $\Phi^{-1}(\alpha/2) < b\Phi^{-1}(\alpha/2) < 0$, puis par stricte croissance de Φ : $\frac{\alpha}{2} < \frac{\beta}{2} < \Phi(0) = \frac{1}{2}$.

Commentaire : à taille d'échantillon fixée, réduire la marge d'erreur (intervalle plus précis) se paie par une augmentation du risque : on ne peut pas gagner simultanément en précision et en confiance sans augmenter n .

Conclusion

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

- 7) Dans cette question, toutes les covariances sont prises entre variables admettant une variance ; la covariance est alors définie et bilinéaire (préambule), et une

combinaison linéaire de telles variables admet une variance.

a) $\bar{X}_n = g_n(X_1, \dots, X_n)$ avec

$$g_n(x_1, \dots, x_n) = \frac{1}{n} \sum x_i,$$

sans biais (question 5)) et de variance $\frac{1}{n}$.

Conclusion

$\bar{X}_n \in \mathcal{E}_\theta$, donc $\mathcal{E}_\theta$ n'est pas vide.

b) Soit $U_n \in \mathcal{E}_\theta$. En écrivant $U_n = \bar{X}_n + (U_n - \bar{X}_n)$ et avec le résultat admis :

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

Conclusion

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

c) $A_n(\lambda) = (1 - \lambda)Z_n + \lambda U_n = h(X_1, \dots, X_n)$ avec $h = (1 - \lambda)z + \lambda g$ si $Z_n = z(X_1, \dots, X_n)$ et $U_n = g(X_1, \dots, X_n)$. Il admet une variance et $E(A_n(\lambda)) = (1 - \lambda)\theta + \lambda\theta = \theta : A_n(\lambda) \in \mathcal{E}_\theta$. Par bilinéarité :

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

Notons $c = \operatorname{Cov}(Z_n, U_n - Z_n)$ et $v = V(U_n - Z_n) \geq 0$. Comme Z_n est optimal, $V(A_n(\lambda)) \geq V(Z_n)$, soit $2\lambda c + \lambda^2 v \geq 0$ pour tout réel λ . Si $v > 0$ et $c \neq 0$, le choix $\lambda = -\frac{c}{v}$ donne $-\frac{c^2}{v} < 0$: contradiction. Si $v = 0$, $2\lambda c \geq 0$ pour tout λ impose $c = 0$.

Conclusion

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

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

- d) Comme $\bar{X}_n$ et Z_n sont optimaux, $V(Z_n) \leq V(\bar{X}_n)$ et $V(\bar{X}_n) \leq V(Z_n)$, donc $V(Z_n) = V(\bar{X}_n)$. La question c) avec $U_n = \bar{X}_n$ donne $\operatorname{Cov}(Z_n, \bar{X}_n) = V(Z_n)$. Alors :

$$V(\bar{X}_n - Z_n) = V(\bar{X}_n) - 2 \operatorname{Cov}(\bar{X}_n, Z_n) + V(Z_n) = V(\bar{X}_n) - V(Z_n) = 0$$

La variable $D = \bar{X}_n - Z_n$ a une espérance nulle et une variance nulle. Pour tout entier $m \geq 1$, l'inégalité de Bienaymé-Tchebychev donne $P(|D| \geq \frac{1}{m}) \leq m^2 V(D) = 0$. La suite d'événements $(|D| \geq \frac{1}{m})_m$ est croissante, de réunion $[D \neq 0]$; par continuité monotone, $P([D \neq 0]) = \lim_m P(|D| \geq \frac{1}{m}) = 0$.

Conclusion

$Z_n = \bar{X}_n$ presque sûrement : $\bar{X}_n$ est, presque sûrement, l'unique estimateur optimal dans $\mathcal{E}_\theta$.

- 8) a) Comme $X - \theta$ suit la loi normale centrée réduite, $F_X(x) = \Phi(x - \theta)$. D'après 6) a), F_X est une bijection strictement croissante de $\mathbb{R}$ sur $]0, 1[$: l'équation $F_X(x) = \frac{1}{2}$ a une unique solution, et $\Phi(0) = \frac{1}{2}$ (parité de la densité).

Conclusion

X admet une unique médiane théorique, $M = \theta$.

- b) $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}$, donc $f_X(M) = \frac{1}{\sqrt{2\pi}}$. Pour x réel, $F_X(2M - x) = \Phi(\theta - x) = 1 - \Phi(x - \theta) = 1 - F_X(x)$.

Méthode 1 — Par dérivation.

Les deux membres sont dérivables sur $\mathbb{R}$; en dérivant $F_X(2M - x) = 1 - F_X(x)$:
 $-f_X(2M - x) = -f_X(x)$.

Méthode 2 — Par l'expression de la densité.

$$f_X(2M - x) = \frac{1}{\sqrt{2\pi}} e^{-(\theta - x)^2/2} = \frac{1}{\sqrt{2\pi}} e^{-(x - \theta)^2/2} = f_X(x).$$

La méthode 1 montre que la symétrie de la fonction de répartition entraîne celle de toute densité continue; la méthode 2 est une vérification directe.

Conclusion

$$f_X(M) = \frac{1}{\sqrt{2\pi}}, F_X(2M - x) = 1 - F_X(x) \text{ et } f_X(2M - x) = f_X(x) \text{ pour tout réel } x.$$

- c) X admet des moments de tout ordre (loi normale), donc d'après 1) f), chaque Y_k admet une espérance. Notons $F = F_X$, $f = f_X$, $c_k = k \binom{n}{k} = \frac{n!}{(k-1)!(n-k)!}$ et remarquons que $c_{n-k+1} = \frac{n!}{(n-k)!(k-1)!} = c_k$.

Méthode 1 — Changement de variable affine.

Avec le changement de variable $x = 2M - y$ (affine), et b) :

$$\begin{aligned} E(Y_k) &= \int_{-\infty}^{+\infty} x c_k F(x)^{k-1} (1 - F(x))^{n-k} f(x) dx \\ &= \int_{-\infty}^{+\infty} (2M - y) c_k (1 - F(y))^{k-1} F(y)^{n-k} f(y) dy \end{aligned}$$

Or

$$c_k F^{n-k} (1 - F)^{k-1} f = c_{n-k+1} F^{(n-k+1)-1} (1 - F)^{n-(n-k+1)} f = f_{Y_{n-k+1}}$$

(question 1) e)). Comme $f_{Y_{n-k+1}}$ est une densité et que Y_{n-k+1} a une espérance :

$$E(Y_k) = 2M - E(Y_{n-k+1})$$

Méthode 2 — Par symétrie de l'échantillon.

Les variables $X'_i = 2M - X_i$ sont indépendantes et, d'après b), de même fonction de répartition que X : $P([X'_i \leq x]) = P([X_i \geq 2M - x]) = 1 - F(2M - x) = F(x)$. Le réarrangement croissant de $(X'_1, \dots, X'_n)$ est $(2M - Y_n, \dots, 2M - Y_1)$, donc sa k -ième valeur est $Y'_k = 2M - Y_{n-k+1}$. La formule de 1) d), qui ne dépend que de la fonction de répartition commune, montre que Y'_k et Y_k ont la même fonction de répartition, donc la même loi et la même espérance : $E(Y_k) = 2M - E(Y_{n-k+1})$.

La méthode 1 est un calcul d'intégrale; la méthode 2 explique le résultat par la symétrie de la loi autour de M .

Conclusion

$$E(Y_k - M) = E(M - Y_{n-k+1}) \text{ pour tout } k \in \llbracket 1, n \rrbracket.$$

d) Avec $n = 2\ell + 1$ et $k = \ell + 1$, on a $n - k + 1 = \ell + 1$, donc c) donne

$$E(Y_{\ell+1}) - M = M - E(Y_{\ell+1}), \text{ soit } E(Y_{\ell+1}) = M = \theta.$$

Ainsi $Y_{\ell+1} = \psi_{\ell+1}(X_1, \dots, X_n)$ est un estimateur sans biais de θ , qui admet une variance (1) f) avec $r = 2$) : $Y_{\ell+1} \in \mathcal{E}_\theta$. Par 7) b) et 5) :

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

Commentaire : la médiane empirique est, comme la moyenne empirique, un estimateur sans biais de θ , mais elle est moins précise ; la partie III montrera que $V(Y_{\ell+1}) \sim \frac{\pi}{2n}$, soit environ 1,57 fois la variance de $\bar{X}_n$.

Conclusion

$$E(Y_{\ell+1}) = \theta \text{ et } V(Y_{\ell+1}) \geq \frac{1}{n}.$$

Partie III : Résultats asymptotiques

9) a) $\exp(sJ)$ prend les valeurs 1 et e^s avec les probabilités $1 - p$ et p .

Conclusion

$$L_J(s) = 1 - p + pe^s \text{ pour tout réel } s.$$

b) Par le théorème de transfert, $\exp(sT)$ admet une espérance si et seulement si

$\int_{-\infty}^{+\infty} e^{sx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ converge (la fonction est positive). Or $sx - \frac{x^2}{2} = -\frac{(x-s)^2}{2} + \frac{s^2}{2}$, donc l'intégrande vaut $e^{s^2/2} \times \frac{1}{\sqrt{2\pi}} e^{-(x-s)^2/2}$, qui est $e^{s^2/2}$ fois une densité de la loi $\mathcal{N}(s, 1)$: l'intégrale converge.

Conclusion

$L_T(s)$ existe pour tout réel s .

- c) D'après le calcul de b), $L_T(s) = e^{s^2/2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-s)^2/2} dx = e^{s^2/2}$. Ensuite, $\exp(s(\sigma T + \theta)) = e^{\theta s} \exp((\sigma s)T)$, qui admet une espérance par b), et par linéarité :

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

Conclusion

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

- 10) a) Par définition de la partie entière, $\frac{n}{2} - 1 < \lfloor n/2 \rfloor \leq \frac{n}{2}$, donc $\frac{n}{2} < k(n) \leq \frac{n}{2} + 1$. Ainsi $0 < \frac{k(n) - n/2}{\sqrt{n}} \leq \frac{1}{\sqrt{n}} \rightarrow 0$.

Conclusion

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

- b) $F_X(x) = \Phi(x - \theta)$ est de classe C^1 sur $\mathbb{R}$, de dérivée f_X . La formule de Taylor-Young à l'ordre 1 en M donne, avec 8) :

$$F_X(M + h) = F_X(M) + h f_X(M) + h \varepsilon(h) = \frac{1}{2} + \frac{h}{\sqrt{2\pi}} + h \varepsilon(h) \quad \text{où} \quad \lim_{h \rightarrow 0} \varepsilon(h) = 0$$

Avec $h = \frac{x}{\sqrt{n}} \rightarrow 0$: $q_n = \frac{1}{2} + \frac{x}{\sqrt{2\pi n}} + \frac{x}{\sqrt{n}} \varepsilon\left(\frac{x}{\sqrt{n}}\right)$, et $\varepsilon\left(\frac{x}{\sqrt{n}}\right) \rightarrow 0$ (si $x = 0$, ce reste est nul).

Conclusion

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

c) D'après a) et b) :

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

Le premier et le dernier termes tendent vers 0.

Conclusion

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

11) a) D'après 1) b), $S_n(y_n)$ suit la loi binomiale de paramètres n et $q_n = F_X(y_n) \in]0, 1[$.
Soit s réel et $h = \frac{s}{\sqrt{n}}$. On a $\exp(sW_n) = e^{-s\sqrt{n}q_n} \exp(h S_n(y_n))$.

Méthode 1 — Formule du binôme.

Par transfert, puis par la formule du binôme :

$$E(\exp(hS_n(y_n))) = \sum_{j=0}^n e^{hj} \binom{n}{j} q_n^j (1 - q_n)^{n-j} = (q_n e^h + 1 - q_n)^n$$

Méthode 2 — Produit de variables indépendantes.

$$\exp(hS_n(y_n)) = \prod_{k=1}^n \exp(hJ_k(y_n)),$$

où les variables $\exp(hJ_k(y_n))$ sont indépendantes (fonctions des X_k), finies, d'espérance $L_{J_k(y_n)}(h) = 1 - q_n + q_n e^h$ par 9) a). Par le résultat admis du préambule (étendu à n facteurs par récurrence et lemme des coalitions), l'espérance du produit est le produit des espérances.

La méthode 1 utilise la loi binomiale déjà identifiée ; la méthode 2 montre le principe général : la fonction L d'une somme de variables indépendantes est le produit des fonctions L .

Conclusion

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

b) Si $s = 0$, $L_{W_n}(0) = 1$ et $\ln(L_{W_n}(0)) = 0 = \frac{0^2}{8}$. Soit $s \neq 0$ et $h_n = \frac{s}{\sqrt{n}} \rightarrow 0$. On utilise les développements limités à l'ordre 2 en 0, écrits avec des fonctions de reste :

$$e^h = 1 + h + \frac{h^2}{2} + h^2\varepsilon_1(h), \quad \ln(1+v) = v - \frac{v^2}{2} + v^2\varepsilon_2(v)$$

où $\varepsilon_1(h) \rightarrow 0$ quand $h \rightarrow 0$ et $\varepsilon_2(v) \rightarrow 0$ quand $v \rightarrow 0$. Posons

$v_n = q_n(e^{h_n} - 1) = q_n h_n + q_n h_n^2 \left(\frac{1}{2} + \varepsilon_1(h_n) \right)$. Comme $0 < q_n < 1$ et $h_n \rightarrow 0$, $v_n \rightarrow 0$, $v_n > -1$, et :

$$v_n^2 = q_n^2 h_n^2 \left(1 + h_n \left(\frac{1}{2} + \varepsilon_1(h_n) \right) \right)^2 = q_n^2 h_n^2 (1 + \eta_n) \quad \text{avec} \quad \eta_n \rightarrow 0$$

Avec $nh_n = s\sqrt{n}$ et $nh_n^2 = s^2$:

$$\ln(L_{W_n}(s)) = -s\sqrt{n}q_n + n \ln(1 + v_n) = -s\sqrt{n}q_n + n \left(v_n - \frac{v_n^2}{2} + v_n^2 \varepsilon_2(v_n) \right)$$

$$= -s\sqrt{n}q_n + s\sqrt{n}q_n + s^2 q_n \left(\frac{1}{2} + \varepsilon_1(h_n) \right) - \frac{s^2 q_n^2}{2} (1 + \eta_n) + s^2 q_n^2 (1 + \eta_n) \varepsilon_2(v_n)$$

Comme $q_n \rightarrow \frac{1}{2}$ (question 10) b)), $\varepsilon_1(h_n) \rightarrow 0$, $\eta_n \rightarrow 0$ et $\varepsilon_2(v_n) \rightarrow 0$ (limites de fonctions le long de suites tendant vers 0) :

$$\ln(L_{W_n}(s)) = \frac{s^2}{2} q_n (1 - q_n) + o(1) = \frac{s^2}{2} \cdot \frac{1}{4} + o(1) = \frac{s^2}{8} + o(1)$$

Par continuité de l'exponentielle, $L_{W_n}(s) \rightarrow e^{s^2/8}$, et par 9) c) avec $\sigma = \frac{1}{2}$ et $\theta = 0$, $L_T(s) = e^{s^2/8}$.

Conclusion

$$\ln(L_{W_n}(s)) = \frac{s^2}{8} + o(1) \text{ et } \lim_{n \rightarrow +\infty} L_{W_n}(s) = L_T(s).$$

12) Pour $x = 0$, on a $y_n = M$ et $q_n = F_X(M) = \frac{1}{2}$ pour tout n . Alors

$$S_n(M) = \sum_{k=1}^n J_k(M)$$

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 $\frac{1}{4} > 0$, et $W_n = \frac{S_n(M) - \frac{n}{2}}{\sqrt{n}}$.

- Le théorème limite central s'applique directement : $\frac{S_n(M) - n/2}{\sqrt{n}/2}$ converge en loi vers T . Par composition par la fonction continue $y \mapsto \frac{y}{2}$, W_n converge en loi vers $\frac{T}{2}$, ce qui est le résultat admis après 11) b), sans passer par les fonctions L .
- La limite des fonctions L s'obtient aussi sans le terme $\frac{x}{\sqrt{2\pi n}}$:

$$L_{W_n}(s) = \left(\frac{e^{-s/(2\sqrt{n})} + e^{s/(2\sqrt{n})}}{2} \right)^n$$
, et un développement limité à l'ordre 2 de cette moyenne donne $n \ln \left(1 + \frac{s^2}{8n} + o\left(\frac{1}{n}\right) \right) \rightarrow \frac{s^2}{8}$.

Conclusion

pour $x = 0$, la probabilité q_n ne dépend plus de n , et le théorème limite central (avec la composition par une fonction continue) donne directement la convergence en loi de (W_n) vers $\frac{T}{2}$.

13) a) Soit $n \in \mathbb{N}^*$ et x réel. Comme $\sqrt{n} > 0$:

$\sqrt{n}(Y_{k(n)} - M) \leq x \iff Y_{k(n)} \leq M + \frac{x}{\sqrt{n}} = y_n$. Comme $k(n) = \lfloor n/2 \rfloor + 1 \in \llbracket 1, n \rrbracket$, la question 1) c) donne $[Y_{k(n)} \leq y_n] = [S_n(y_n) \geq k(n)]$. Enfin :

$$\begin{aligned} S_n(y_n) \geq k(n) &\iff S_n(y_n) - nq_n \geq k(n) - nq_n \iff \frac{S_n(y_n) - nq_n}{\sqrt{n}} \geq \frac{k(n) - nq_n}{\sqrt{n}} \\ &\iff W_n \geq u_n \end{aligned}$$

Conclusion

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

b) On évite tout théorème général sur les convergences en loi et on procède par encadrement. Notons $G(c) = P\left(\left[\frac{T}{2} \geq c\right]\right) = 1 - \Phi(2c)$, fonction continue sur $\mathbb{R}$.

Étape 1 : pour tout réel c , $P([W_n \geq c]) \rightarrow G(c)$. La fonction de répartition de $\frac{T}{2}$,

$t \mapsto \Phi(2t)$, est continue partout, donc par la convergence en loi admise, $F_{W_n}(t) \rightarrow \Phi(2t)$ pour tout réel t . Soit $\varepsilon > 0$; par continuité de Φ , il existe $\delta > 0$ tel que $\Phi(2c) - \Phi(2(c - \delta)) \leq \varepsilon$. Comme $[W_n \leq c - \delta] \subset [W_n < c] \subset [W_n \leq c]$:

$$F_{W_n}(c - \delta) \leq P([W_n < c]) \leq F_{W_n}(c)$$

Pour n assez grand, $F_{W_n}(c - \delta) \geq \Phi(2(c - \delta)) - \varepsilon \geq \Phi(2c) - 2\varepsilon$ et $F_{W_n}(c) \leq \Phi(2c) + \varepsilon$. Donc $|P([W_n < c]) - \Phi(2c)| \leq 2\varepsilon$ pour n assez grand : $P([W_n < c]) \rightarrow \Phi(2c)$ et $P([W_n \geq c]) \rightarrow G(c)$.

Étape 2 : seuil variable. Soit $\varepsilon > 0$. Par continuité de G en u , il existe $\eta > 0$ tel que $|G(u + \eta) - G(u)| \leq \varepsilon$ et $|G(u - \eta) - G(u)| \leq \varepsilon$. Comme $u_n \rightarrow u$ (question 10) c)), il existe n_1 tel que $u - \eta \leq u_n \leq u + \eta$ pour $n \geq n_1$; alors $[W_n \geq u + \eta] \subset [W_n \geq u_n] \subset [W_n \geq u - \eta]$, d'où :

$$P([W_n \geq u + \eta]) \leq P([W_n \geq u_n]) \leq P([W_n \geq u - \eta])$$

D'après l'étape 1, il existe n_2 tel que pour $n \geq n_2$, $P([W_n \geq u \pm \eta])$ soit à moins de ε de $G(u \pm \eta)$. Pour $n \geq \max(n_1, n_2)$:

$$G(u) - 2\varepsilon \leq P([W_n \geq u_n]) \leq G(u) + 2\varepsilon$$

Donc $P([W_n \geq u_n]) \rightarrow G(u)$. Enfin, avec $u = -\frac{x}{\sqrt{2\pi}}$ et $1 - \Phi(-y) = \Phi(y)$:

$$G(u) = 1 - \Phi\left(-\frac{2x}{\sqrt{2\pi}}\right) = \Phi\left(x\sqrt{\frac{2}{\pi}}\right) = P\left(\left[T \leq x\sqrt{\frac{2}{\pi}}\right]\right) = P\left(\left[\sqrt{\frac{\pi}{2}}T \leq x\right]\right)$$

Conclusion

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

c) Soit z réel et $x = z\sqrt{\frac{\pi}{2}}$. D'après a) et b) (appliquées à ce x) :

$$\begin{aligned} P\left(\left[\sqrt{\frac{2n}{\pi}}(Y_{k(n)} - M) \leq z\right]\right) &= P\left([\sqrt{n}(Y_{k(n)} - M) \leq x]\right) \\ &= P([W_n \geq u_n]) \rightarrow P\left(\left[\sqrt{\frac{\pi}{2}} T \leq z\sqrt{\frac{\pi}{2}}\right]\right) = \Phi(z) \end{aligned}$$

La convergence a lieu en tout réel z , et Φ est continue.

Conclusion

$$\left(\sqrt{\frac{2n}{\pi}}(Y_{k(n)} - M)\right)_{n \in \mathbb{N}^*} \text{ converge en loi vers } T, \text{ de loi normale centrée réduite.}$$

14) a) Avec $n = 2\ell + 1$, $\lfloor n/2 \rfloor = \lfloor \ell + \frac{1}{2} \rfloor = \ell$.

Conclusion

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

b) D'après 8) d), $E(Y_{k(n)}) = E(Y_{\ell+1}) = \theta$ pour tout n impair.

Conclusion

$$\lim_{n \rightarrow +\infty} E(Y_{k(n)}) = \theta \text{ (la suite est constante).}$$

- c) Comme $E(Y_{k(n)}) = M$, on a $E((Y_{k(n)} - M)^2) = V(Y_{k(n)})$, et le résultat admis s'écrit
- $$\frac{2n}{\pi} V(Y_{k(n)}) \rightarrow E(T^2) = 1.$$

Conclusion

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

- d) D'après 8) d), $Y_{k(n)} \in \mathcal{E}_\theta$, donc par le résultat admis en 7),
- $$\text{Cov}(\bar{X}_n, Y_{k(n)} - \bar{X}_n) = 0, \text{ soit } \text{Cov}(\bar{X}_n, Y_{k(n)}) = V(\bar{X}_n) = \frac{1}{n} \text{ (question 5)}. \text{ Les}$$
- variances sont strictement positives ($V(Y_{k(n)}) \geq \frac{1}{n}$), donc :

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

Par c), $nV(Y_{k(n)}) \rightarrow \frac{\pi}{2}$, et par continuité de la racine carrée :

Conclusion

$$\lim_{n \rightarrow +\infty} \rho_n = \sqrt{\frac{2}{\pi}} \approx 0,80 \text{ (valeur confirmée par simulation pour } n = 21 \text{ et } n = 101, \text{ sans valeur de preuve).}$$

Relevé des adaptations

- Énoncé : aucune question n'a dû être modifiée ; le sujet est conforme au référentiel.
- 13) b) : la question est conservée, mais la preuve évite tout théorème général sur les convergences en loi combinées à des suites convergentes (type Slutsky).
Justification : le référentiel exclut Slutsky ; la limite est obtenue par un encadrement élémentaire des probabilités, à partir de la convergence en loi admise de (W_n) et de la continuité de Φ .

- 11) b) : le développement limité est mené en évaluant des développements de $\exp$ et de $\ln(1 + \cdot)$ le long de suites tendant vers 0, avec des fonctions de reste explicites.
Justification : la composition des développements limités est hors programme.