



PRÉPASIMPLE

prepasimple.fr

HEC / ESSEC

Mathématiques — Mathématiques appliquées

2023

Mathématiques appliquées

Mathématiques appliquées — Prépa ECG

Annales de concours · Sujet

Dans tout le problème, les variables aléatoires considérées sont définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$. Les notations $E(X)$ et $V(X)$ désignent, lorsqu'elles existent, l'espérance et la variance de X .

On désigne par W l'ensemble des fonctions h de classe C^1 sur $\mathbb{R}$ telles que, pour tout $x \in \mathbb{R}$, $|h'(x)| \leq 1$. La variable N suit la loi normale centrée réduite. On admet que, si X admet une espérance et $h \in W$, alors $h(X)$ admet une espérance. On note $c_h = E(h(N))$.

On note Φ la fonction de répartition de N et φ sa densité :

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

La fonction Φ est la primitive de φ qui vaut $\frac{1}{2}$ en 0. Les parties 1 et 2 sont indépendantes. La partie 3 peut être traitée indépendamment des deux premières. La partie 4 utilise les parties 1 et 2 et est indépendante de la partie 3, sauf sa dernière question.

Partie 1 - Transformation de Stein

Soit $h \in W$. On définit sur $\mathbb{R}$ la fonction $\theta : x \mapsto \frac{\Phi(x)}{\varphi(x)}$ et, lorsque les intégrales convergent,

$$f_h(x) = \theta(-x) \int_{-\infty}^x h'(t) \Phi(t) dt + \theta(x) \int_x^{+\infty} h'(t) (1 - \Phi(t)) dt.$$

L'objectif est d'obtenir, pour X admettant une espérance, une expression de $|E(h(X)) - E(h(N))|$ ne faisant pas intervenir directement N .

1. a) Montrer que, pour tout $x \geq 0$ et tout $t \in [x, +\infty[$, $0 \leq x\varphi(t) \leq t\varphi(t)$. En déduire :

$$\forall x \geq 0, \quad 0 \leq x(1 - \Phi(x)) \leq \varphi(x).$$

On remarquera que $\varphi'(t) = -t\varphi(t)$.

1. b) Procéder de façon analogue pour montrer que, pour tout $x \leq 0$,

$$-\varphi(x) \leq x\Phi(x) \leq 0.$$

1. c) En déduire, à l'aide d'une intégration par parties, la convergence des intégrales suivantes et montrer, pour tout $x \in \mathbb{R}$:

$$\int_{-\infty}^x \Phi(t) dt = x\Phi(x) + \varphi(x), \quad \int_x^{+\infty} (1 - \Phi(t)) dt = -x(1 - \Phi(x)) + \varphi(x). \quad (R_1)$$

2. a) Montrer, pour tous réels x, y , $|h(x) - h(y)| \leq |x - y|$, puis $|h(x)| \leq |x| + |h(0)|$.

2. b) Pour tout $x \in \mathbb{R}$, justifier la convergence de $\int_{-\infty}^x h'(t)\Phi(t) dt$ et montrer :

$$\int_{-\infty}^x h'(t)\Phi(t) dt = h(x)\Phi(x) - \int_{-\infty}^x h(t)\varphi(t) dt.$$

On admet de même la convergence et l'égalité :

$$\int_x^{+\infty} h'(t)(1 - \Phi(t)) dt = -h(x)(1 - \Phi(x)) + \int_x^{+\infty} h(t)\varphi(t) dt.$$

2. c) En déduire, pour tout $x \in \mathbb{R}$:

$$-\int_{-\infty}^x h'(t)\Phi(t) dt + \int_x^{+\infty} h'(t)(1 - \Phi(t)) dt = c_h - h(x).$$

3. a) Établir, pour tout $x \in \mathbb{R}$:

$$\theta'(x) = 1 + x\theta(x), \quad \theta''(x) = x + (1 + x^2)\theta(x), \quad \theta(-x)\Phi(x) = \theta(x)(1 - \Phi(x)).$$

3. b) En déduire que f_h est de classe C^1 sur $\mathbb{R}$ et vérifie :

$$f_h'(x) - xf_h(x) = c_h - h(x).$$

Pourquoi peut-on alors affirmer que f_h est de classe C^2 sur $\mathbb{R}$?

3. c) En conclure que, si X admet une espérance,

$$|E(h(X)) - E(h(N))| = |E(f'_h(X) - X f_h(X))|.$$

4. Majoration de $|f_h|$.

4. a) Montrer, en utilisant (R_1) , que, pour tout $x \in \mathbb{R}$:

$$\theta(-x) \int_{-\infty}^x \Phi(t) dt + \theta(x) \int_x^{+\infty} (1 - \Phi(t)) dt = 1.$$

4. b) En déduire que $|f_h(x)| \leq 1$ pour tout $x \in \mathbb{R}$.

5. Majoration de $|f''_h|$.

5. a) Montrer, pour tout $x \in \mathbb{R}$:

$$\theta''(-x) \int_{-\infty}^x \Phi(t) dt + \theta''(x) \int_x^{+\infty} (1 - \Phi(t)) dt = 1.$$

5. b) Établir, pour tout $x \in \mathbb{R}$:

$$f''_h(x) = -h'(x) + \theta''(-x) \int_{-\infty}^x h'(t) \Phi(t) dt + \theta''(x) \int_x^{+\infty} h'(t) (1 - \Phi(t)) dt.$$

5. c) Étudier les variations sur $\mathbb{R}$ de $x \mapsto \Phi(x) + \frac{x}{1+x^2} \varphi(x)$. En déduire son signe et celui de θ'' , puis montrer que $|f''_h(x)| \leq 2$ pour tout réel x .

Partie 2 - Majoration uniforme de la distance de Kolmogorov

Si X a pour fonction de répartition F_X , on définit, pour tout réel x , sa distance de Kolmogorov au point x à la loi normale centrée réduite par :

$$d_X(x) = |F_X(x) - \Phi(x)|.$$

On définit $h_x(t) = 1$ si $t \leq x$ et $h_x(t) = 0$ si $t > x$. On définit aussi :

$$\gamma(t) = \begin{cases} 1 & \text{si } t < 0, \\ 0 & \text{si } t > 1, \\ 1 - 3t^2 + 2t^3 & \text{si } t \in [0, 1]. \end{cases}$$

Soit X une variable aléatoire.

6. Pour tout réel x , quelle est la loi de $h_x(X)$? En déduire que son espérance existe et vaut $F_X(x)$.
7. a) Écrire une fonction Python `gamma(t)` qui calcule et renvoie $\gamma(t)$, t étant donné.
7. b) Utiliser cette fonction pour écrire un script qui affiche le graphe de γ sur $[-1, 2]$ dans un repère. On pourra utiliser `np.linspace(a, b, m)`, qui renvoie un tableau de m valeurs régulièrement espacées entre a et b , bornes comprises. Après `import matplotlib.pyplot as plt`, `plt.plot(T, Y)` relie les points dont les abscisses et ordonnées sont dans T et Y , et `plt.show()` affiche le graphique.
8. a) Montrer que γ est continue sur $\mathbb{R}$ et de classe C^1 sur $\mathbb{R} \setminus \{0, 1\}$.
8. b) Étudier les variations de γ sur $[0, 1]$ et montrer que $\gamma(t) \in [0, 1]$ pour tout réel t .
8. c) Établir que γ est dérivable en 1 et que $\gamma'(1) = 0$. On admet de même sa dérivabilité en 0 et l'égalité $\gamma'(0) = 0$.
8. d) Justifier que γ est de classe C^1 sur $\mathbb{R}$ et que $|\gamma'(t)| \leq \frac{3}{2}$ pour tout réel t .

On suppose dans la suite de cette partie que X admet une espérance. On considère un réel M_X tel que, pour tout $h \in W$,

$$|E(h(X)) - E(h(N))| \leq M_X.$$

9. Soient $t > 0$ et $x \in \mathbb{R}$. Pour tout $y \in \mathbb{R}$, on pose $k_x(y) = \gamma\left(\frac{y-x}{t}\right)$.

9. a) Montrer que $h_x(y) \leq k_x(y)$ pour tout réel y .

9. b) On admet l'existence de $E(k_x(X))$ et $E(k_x(N))$. Justifier :

$$E(h_x(X)) - E(h_x(N)) \leq E(k_x(X)) - E(k_x(N)) + E(k_x(N)) - E(h_x(N)).$$

9. c) Montrer que $E(k_x(N)) - E(h_x(N)) = \int_x^{x+t} k_x(u) \varphi(u) du$.

9. d) Établir que $g : u \mapsto \frac{2t}{3} k_x(u)$ appartient à W . En déduire :

$$E(h_x(X)) - E(h_x(N)) \leq \frac{3M_X}{2t} + \frac{t}{\sqrt{2\pi}} \leq \frac{3M_X}{2t} + \frac{t}{2}.$$

On admet de même, en utilisant k_{x-t} :

$$E(h_x(N)) - E(h_x(X)) \leq \frac{3M_X}{2t} + \frac{t}{2}.$$

10. En étudiant $t \mapsto \frac{3M_X}{2t} + \frac{t}{2}$ sur $]0, +\infty[$, en déduire, pour tout réel x :

$$|E(h_x(X)) - E(h_x(N))| \leq \sqrt{3M_X}, \quad d_X(x) \leq \sqrt{3M_X}. \quad (R_2)$$

Partie 3 - Estimation d'une densité

On considère une variable aléatoire à densité X , de fonction de répartition F et de densité f , dont la loi dépend d'un paramètre inconnu $\theta \in \Theta$, où Θ est un intervalle de $\mathbb{R}$. Soit a un point de continuité fixé de f . On souhaite estimer $f(a)$. Par exemple, si X suit la loi exponentielle de paramètre θ et $a > 0$, on souhaite estimer $\theta e^{-\theta a}$.

Pour tout $\theta \in \Theta$, on dispose d'une suite $(X_i)_{i \geq 1}$ de variables indépendantes de même loi que X . On choisit une suite $(h_n)_{n \geq 1}$ de réels strictement positifs telle que $h_n \rightarrow 0$ et

$$nh_n \rightarrow +\infty.$$

Pour $n \geq 1$ et $\omega \in \Omega$, on définit $C_n(\omega)$ comme le nombre d'indices $i \in \{1, \dots, n\}$ tels que $X_i(\omega) \in]a - h_n, a + h_n]$, puis :

$$f_n(\omega) = \frac{C_n(\omega)}{2nh_n}.$$

- 11.** On dispose d'un fichier `stats.csv` comportant une colonne salaire. Ses valeurs sont la réalisation d'un échantillon de la loi de X , de taille supérieure à 10000. On suppose cette colonne numérique et sans valeur manquante.

Documentation : après `import pandas as pd`, `pd.read_csv("nom.csv")` lit un fichier CSV dans une table ; `table["colonne"]` en extrait une série ; `série.count()` donne le nombre de valeurs non manquantes ; `série.iloc[i]` renvoie la valeur de position i , les positions commençant à zéro. `input("a=")` lit une chaîne et `float` la convertit en réel. Après `import numpy as np`, `np.sqrt(x)` renvoie la racine carrée de x .

- 11. a)** Après `import pandas as pd`, quelles instructions permettent de lire les valeurs de la colonne salaire et d'affecter la série obtenue à une variable `echantillon` ? Le fichier est dans le répertoire de travail.

- 11. b)** On souhaite calculer et afficher $f_n(\omega)$ pour a donné, avec $h_n = \frac{1}{\sqrt{n}}$. La réalisation de l'échantillon est représentée par `echantillon`. Compléter le script suivant :

Python

```
import numpy as np

a = float(input("a="))

n = echantillon.count()

h = 1 / np.sqrt(n)

C = 0

for i in range(n):
    if ... and ...:
        ... += 1

print(C / ...)
```

12. Montrer que C_n suit une loi binomiale de paramètres (n, p_n) en précisant p_n en fonction de a et h_n . En déduire l'existence de $E(f_n)$ et l'égalité :

$$E(f_n) = \frac{F(a + h_n) - F(a - h_n)}{2h_n}.$$

13. a) En utilisant la dérivabilité de F en a , montrer que $E(f_n) \rightarrow f(a)$.

13. b) Montrer que $\mathbb{V}(f_n)$ existe et tend vers zéro.

On suppose désormais $f(a) > 0$, $p_n \in]0, 1[$ pour tout $n \geq 1$, F de classe C^2 au voisinage de a , et $nh_n^3 \rightarrow 0$. On pose :

$$\begin{aligned} \sigma_n &= \sqrt{np_n(1-p_n)}, & \theta_n &= \sqrt{2h_n f(a)}, & D_n &= \frac{C_n - np_n}{\sigma_n}, & \hat{f}_n \\ &= \frac{\theta_n \sqrt{n}}{f(a)} (f_n - f(a)). \end{aligned}$$

14. a) En appliquant directement Taylor-Young à F au point a , à l'ordre deux, pour les accroissements h_n et $-h_n$, puis en soustrayant les deux égalités, montrer :

$$p_n = 2h_n f(a) + o(h_n^2) = \theta_n^2 + o(h_n^2).$$

14. b) En déduire $p_n \sim 2h_n f(a)$, puis $np_n \rightarrow +\infty$.

14. c) Montrer :

$$\hat{f}_n = \frac{\sigma_n}{\theta_n \sqrt{n}} D_n + \sqrt{n} \left(\frac{p_n}{\theta_n} - \theta_n \right),$$

puis établir :

$$\frac{\sigma_n}{\theta_n \sqrt{n}} \rightarrow 1, \quad \sqrt{n} \left(\frac{p_n}{\theta_n} - \theta_n \right) \rightarrow 0.$$

On admet dans la suite de cette partie que $(\hat{f}_n)$ converge en loi vers N , ce qui implique, pour tous réels $x \leq y$:

$$P(x \leq \hat{f}_n \leq y) \rightarrow \Phi(y) - \Phi(x).$$

15. Soit $\alpha \in]0, 1[$. On pose $\eta_\alpha = t_\alpha^2$, où t_α est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite.

15. a) Montrer :

$$P\left(f(a)^2 - \left(2f_n + \frac{\eta_\alpha}{2nh_n}\right)f(a) + f_n^2 \leq 0\right) \rightarrow 1 - \alpha.$$

15. b) Pour $n \geq 1$, on note $\Delta_n = \sqrt{\left(f_n + \frac{\eta_\alpha}{4nh_n}\right)^2 - f_n^2}$. Montrer :

$$P\left(f(a) \in \left[f_n + \frac{\eta_\alpha}{4nh_n} - \Delta_n, f_n + \frac{\eta_\alpha}{4nh_n} + \Delta_n\right]\right) \rightarrow 1 - \alpha.$$

Partie 4 - Convergence « uniforme » en loi vers la loi normale

Soit $n \geq 1$. On considère n variables aléatoires indépendantes centrées $X_1, \dots, X_n$ possédant un moment d'ordre trois. On admet qu'elles possèdent alors une variance. On pose :

$$v_k = E(X_k^2), \quad S_n = \sum_{k=1}^n X_k, \quad Y_k = S_n - X_k, \quad \sum_{k=1}^n v_k = 1.$$

Soit f de classe C^2 sur $\mathbb{R}$ telle que $|f(x)| \leq 1$ et $|f''(x)| \leq 2$ pour tout réel x . On admet que, si X, Y admettent une espérance, alors $Yf(X)$ et $f'(X)$ admettent une espérance.

16. a) Montrer :

$$\sum_{k=1}^n v_k E(f'(S_n) - f'(Y_k)) + \sum_{k=1}^n E(X_k^2 f'(Y_k)) = E(f'(S_n)).$$

16. b) Montrer :

$$\sum_{k=1}^n E(X_k(f(S_n) - f(Y_k))) = \sum_{k=1}^n E(X_k f(S_n)) = E(S_n f(S_n)).$$

16. c) En déduire :

$$\begin{aligned} E(f'(S_n) - S_n f(S_n)) &= \sum_{k=1}^n v_k E(f'(S_n) - f'(Y_k)) \\ &+ \sum_{k=1}^n E(X_k [X_k f'(Y_k) - (f(S_n) - f(Y_k))]). \end{aligned}$$

17. Soient $a, b \in \mathbb{R}$.

17. a) Montrer :

$$b f'(a) - (f(a+b) - f(a)) = \int_0^1 b(f'(a) - f'(a+tb)) dt.$$

17. b) En déduire $|b f'(a) - (f(a+b) - f(a))| \leq b^2$.

17. c) En conclure :

$$|E(f'(S_n) - S_n f(S_n))| \leq 2 \sum_{k=1}^n v_k E(|X_k|) + \sum_{k=1}^n E(|X_k|^3),$$

puis, grâce à (R_2) , pour tout réel x :

$$d_{S_n}(x) \leq \sqrt{3 \left(2 \sum_{k=1}^n v_k E(|X_k|) + \sum_{k=1}^n E(|X_k|^3) \right)}. \quad (R_3)$$

Dans la suite, si $(U_n)_{n \geq 1}$ est une suite de variables aléatoires réelles et (δ_n) une suite réelle tendant vers zéro telle que $d_{U_n}(x) \leq \delta_n$ pour tous $n \geq 1$ et $x \in \mathbb{R}$, on dit que (U_n) converge uniformément en loi vers N . On admet qu'une telle suite converge en loi vers N .

18. Une première application. On suppose que $(Z_k)_{k \geq 1}$ est une suite de variables indépendantes de même loi, admettant des moments d'ordre un à trois. On note, pour $i \in \{1, 2, 3\}$, $s_i = E(|Z_k - E(Z_k)|^i)$, puis $\sigma = \sqrt{s_2} \neq 0$ et $X_k = \frac{Z_k - E(Z_k)}{\sigma \sqrt{n}}$ pour $1 \leq k \leq n$. On utilise les notations précédentes.

18. a) Montrer que l'on peut appliquer (R_3) , qui donne ici :

$$d_{S_n}(x) \leq \sqrt{3 \frac{2\sigma^2 s_1 + s_3}{\sigma^3 \sqrt{n}}}.$$

18. b) En déduire la convergence uniforme en loi de (S_n) vers N , donc sa convergence en loi. Quel résultat du cours permettrait d'obtenir directement cette dernière convergence ?

19. Une deuxième application. On suppose que $Z_1, \dots, Z_n$ sont indépendantes et de même loi de Bernoulli de paramètre $p_n \in]0, 1[$. On pose $\sigma_n = \sqrt{np_n(1 - p_n)}$ et $X_k = \frac{Z_k - p_n}{\sigma_n}$.

19. a) Montrer $E(|X_k|) = \frac{2\sigma_n}{n}$ et $E(|X_k|^3) \leq \frac{2}{n\sigma_n}$.

19. b) En déduire, pour tout réel x :

$$d_{S_n}(x) \leq 2\sqrt{3 \left(\frac{\sigma_n}{n} + \frac{1}{2\sigma_n} \right)}.$$

19. c) Justifier : si T_n suit la loi binomiale (n, p_n) , si $p_n \rightarrow 0$ et $np_n \rightarrow +\infty$, alors $\left(\frac{T_n - np_n}{\sqrt{np_n(1 - p_n)}} \right)$ converge uniformément en loi vers N .

20. Un petit lemme. Soit (V_n) une suite qui converge uniformément en loi vers N . Soient (a_n) une suite de réels strictement positifs convergeant vers $a > 0$, et (b_n) une suite réelle convergeant vers b .

20. a) Soient X une variable aléatoire et $\alpha > 0, \beta \in \mathbb{R}$. Montrer, pour tout réel x :

$$|F_{\alpha X + \beta}(x) - F_{\alpha N + \beta}(x)| = d_X \left(\frac{x - \beta}{\alpha} \right).$$

20. b) Montrer, pour tout réel x :

$$P(a_n V_n + b_n \leq x) - P(a_n N + b_n \leq x) \rightarrow 0.$$

20. c) Établir $P(a_n N + b_n \leq x) \rightarrow P(aN + b \leq x)$ pour tout réel x . En déduire que $(a_n V_n + b_n)$ converge en loi vers $aN + b$. Quelle est sa loi ?

21. On reprend les notations de la partie 3.

21. a) Justifier que (D_n) converge uniformément en loi vers N .

21. b) En utilisant les questions 14 et 20, en déduire que $(\hat{f}_n)$ converge en loi vers N .