Concours externe de recrutement d'administrateurs stagiaires de l'INSEE

SESSION 2019

ÉPREUVE DE MATHÉMATIQUES

DURÉE : 4 heures

L'énoncé comporte 5 pages, numérotées de 1 à 5.

Tous documents et appareils électroniques interdits.

Partie 1 : analyse-algèbre

Cette partie est constituée de deux exercices indépendants

Exercice 1

Dans tout l'exercice, n désigne un entier supérieur ou égal à 1.
Une matrice M de $\mathcal{M}_n(\mathbb{R})$ est dite **à diagonale propre** si les valeurs propres de M sont toutes réelles et si ses termes diagonaux sont ses valeurs propres.
On note $\mathcal{E}_n$ l'ensemble des matrices de $\mathcal{M}_n(\mathbb{R})$ à diagonale propre, $\mathcal{A}_n(\mathbb{R})$ le sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$ constitué des matrices antisymétriques (c'est-à-dire vérifiant ${}^tA = -A$).

1. (a) Montrer que $\mathcal{E}_n$ n'est pas vide.
   (b) L'ensemble $\mathcal{E}_n$ est-il un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$?
   (c) Montrer que toute matrice de $\mathcal{M}_n(\mathbb{R})$ est somme de deux matrices de $\mathcal{E}_n(\mathbb{R})$.
   (d) Caractériser les matrices de $\mathcal{E}_2$.
2. Soit $A = (a_{i,j})_{1 \leqslant i,j \leqslant n}$ une matrice symétrique de $\mathcal{M}_n(\mathbb{R})$ et $\lambda_1, \ldots, \lambda_n$ ses valeurs propres non nécessairement distinctes.
   (a) Établir l'égalité suivante :
   $$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}^2 = \sum_{i=1}^{n} \lambda_i^2$$
   (b) Déterminer l'ensemble des matrices symétriques réelles à diagonale propre.
3. Soit A une matrice antisymétrique à diagonale propre.
   (a) Montrer que ${}^tAA$ est nilpotente.
   (b) En déduire que $A = 0$.
   (*On pourra utiliser entre autres le résultat suivant : toute matrice M de $\mathcal{M}_n(\mathbb{R})$ est trigonalisable dans $\mathbb{C}$, c'est-à-dire semblable à une matrice triangulaire dont les termes diagonaux sont les valeurs propres de M*).
4. (a) Quelle est la dimension de $\mathcal{A}_n(\mathbb{R})$?
   (b) Soit F un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$ tel que $F \subset \mathcal{E}_n$.
   Montrer que $\dim(F) \leqslant \dfrac{n(n+1)}{2}$.
   (c) Quelle est la dimension maximale d'un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$, inclus dans $\mathcal{E}_n$?

Exercice 2

On considère la suite de fonctions définie par : $u_n(x) = \prod_{k=1}^{n}\left(1+\frac{x}{n}f\left(\frac{k}{n}\right)\right)$, où f est une fonction continue de [0, 1] dans $\mathbb{R}$, non identiquement nulle, et x appartient à $[-A,\, A]$, avec $A > 0$.

On se propose d'étudier la limite de cette suite et d'en trouver un développement asymptotique quand n tend vers $+\infty$.

Première partie

1. Montrer que, pour n assez grand, pour tout $x \in [-A,\, A]$ et pour tout entier $k \in \{1, \ldots, n\}$: $\left|\frac{x}{n} f\left(\frac{k}{n}\right)\right| < 1$.

2. Montrer que, si $|u| \leq \frac{1}{2}$, alors : $0 \leq u - \ln(1+u) \leq u^2$.

3. En déduire qu'il existe K tel que, pour n assez grand et pour tout $x \in [-A,\, A]$:

$$0 \leq \sum_{k=1}^{n} \frac{x}{n} f\left(\frac{k}{n}\right) - \sum_{k=1}^{n} \ln\left(1+\frac{x}{n} f\left(\frac{k}{n}\right)\right) \leq \frac{K}{n}.$$

4. En déduire la limite $u(x)$ de la suite $\{u_n(x)\}$, quand n tend vers $+\infty$, qu'on écrira sous la forme $u(x) = e^{Lx}$, en précisant la valeur de L.

Deuxième partie

On étudie ici la convergence *uniforme* de la suite $\{u_n(x)\}$ sur $[-A, A]$.

5.

a) Montrer qu'il existe une constante B telle que, pour n assez grand et pour tout $x \in [-A,\, A]$: $\begin{cases} |\ln(u_n(x))| \leq B \\ |\ln(u(x))| \leq B \end{cases}$.

b) En déduire que, dans les mêmes conditions : $|u_n(x) - u(x)| \leq |\ln(u_n(x)) - L\,x|\, e^{B}$.

6.

a) Montrer qu'il existe une constante C telle que, pour n assez grand et pour tout $x \in [-A,\, A]$: $|\ln(u_n(x)) - L\,x| \leq \frac{C}{n} + A\left[\frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k}{n}\right) - L\right]$.

b) En déduire la convergence *uniforme* de la suite $\{u_n(x)\}$ sur $[-A, A]$.

Troisième partie

On va chercher ici un équivalent de $u_n(x)-L\,x$ quand $n\to+\infty$. Dans toute la suite, on considère un **x fixé non nul dans** $[-A, A]$.

7. Montrer que : $u_n(x)-e^{Lx}\sim e^{Lx}\left[\ln(u_n(x))-L\,x\right]$ quand $n\to+\infty$.

8. Montrer que, si $|u|\le\dfrac{2}{3}$, alors : $\left|\ln(1+u)-u+\dfrac{u^2}{2}\right|\le|u|^3$.

9. Montrer qu'il existe une constante D telle que, pour n assez grand :

$$\ln(u_n(x))-L\,x=x\left[\frac{1}{n}\sum_{k=1}^{n}f\left(\frac{k}{n}\right)-L\right]-\frac{x^2}{2n^2}\sum_{k=1}^{n}f^2\left(\frac{k}{n}\right)+D\frac{|x|^3}{n^2}.$$

10. En décomposant le terme $\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}f\left(\frac{k}{n}\right)-L$ sous forme d'une somme d'intégrales sur les intervalles $\left[\dfrac{k-1}{n},\dfrac{k}{n}\right]$, donner un équivalent de $\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}f\left(\frac{k}{n}\right)-L$ quand $n\to+\infty$.

11. En déduire un équivalent de $u_n(x)-L\,x$ quand $n\to+\infty$.

Partie 2 : probabilités-statistiques

Cette partie est constituée de deux exercices indépendants

Exercice 1

On considère une suite $(X_k)_{k\in\mathbb{N}^*}$ de variables aléatoires indépendantes, définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, de même loi, admettant toutes un moment d'ordre 2. On suppose de plus que les variables X_k sont centrées et réduites, c'est-à-dire que $\mathbb{E}(X_k) = 0$ et que $\mathrm{Var}(X_k) = 1$.
On note Φ la fonction de répartition de la loi normale centrée réduite.

Pour tout entier naturel n non nul et tout réel α strictement positif, on pose :

$$S_n = \sum_{k=1}^{n} X_k, \quad U_n(\alpha) = \frac{1}{n^\alpha} S_n, \quad T_n(\alpha) = \frac{1}{n^\alpha} \sum_{k=1}^{n} S_k$$

1. (a) Montrer que si $\alpha > \frac{1}{2}$, alors la suite $(U_n(\alpha))_{n\geqslant 1}$ converge en probabilité vers une variable certaine que l'on précisera.

   (b) Montrer que si $\alpha > \frac{3}{2}$, alors la suite $(T_n(\alpha))_{n\geqslant 1}$ converge en probabilité vers une variable certaine que l'on précisera.

Dans la suite de l'exercice, on pose, pour tout entier naturel n non nul :

$$\begin{cases} U_n = U_n(\frac{1}{2}) = \dfrac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}} \\[2ex] V_n = \dfrac{X_{n+1} + X_{n+2} + \ldots + X_{2n}}{\sqrt{n}} \\[2ex] Y_n = (\sqrt{2} - 1)U_n - V_n \end{cases}$$

2. Soit ε un réel strictement positif.

   (a) Montrer que les suites $\big(\mathbb{P}([U_n \geqslant \varepsilon])\big)_{n\in\mathbb{N}^*}$ et $\big(\mathbb{P}([V_n \leqslant -\varepsilon])\big)_{n\in\mathbb{N}^*}$ admettent des limites finies quand n tend vers$+\infty$, limites que l'on exprimera à l'aide de Φ.

   (b) On suppose qu'il existe un réel ℓ tel que $\lim\limits_{n\to+\infty} \mathbb{P}([Y_n \geqslant \sqrt{2}\varepsilon]) = \ell$.

   Montrer que $\ell \geqslant \big(1 - \Phi(\varepsilon)\big)^2$.

   (c) En déduire que Y_n ne converge pas en probabilité vers 0.

3. (a) Montrer que si la suite $(U_n)_{n\in\mathbb{N}^*}$ convergeait en probabilité vers une variable aléatoire U, alors la suite $U_{2n} - U_n$ convergerait en probabilité vers 0.

   (b) En déduire qu'il n'existe pas de variable Z telle que U_n converge en probabilité vers Z.

   (c) Quel résultat vient-on de montrer concernant le théorème de la limite centrée ?

4. Dans cette question, on suppose que α est strictement inférieur à $\frac{1}{2}$ et on admet le résultat suivant :
   Si deux suites de variables aléatoires (A_n) et (B_n) sont telles que (A_n) converge en loi vers la loi d'une variable A et (B_n) converge en probabilité vers une constante c, alors la suite (A_nB_n) converge en loi vers la loi de la variable cA et la suite $(A_n + B_n)$ converge en loi vers la loi de la variable $A + c$.

   (a) Justifier la convergence en loi de la suite de variables aléatoires $(Z_n)_{n\geqslant 1}$ définie par $Z_n = \dfrac{1}{|U_n| + n^{\alpha - \frac{1}{2}}}$.

   (b) En déduire que la suite $(W_n)_{n\geqslant 1}$, où $W_n = \dfrac{1}{1 + |U_n(\alpha)|}$, converge en probabilité vers 0.

Exercice 2

On dispose d'observations d'une variable d'intérêt X, soit x_i, $i = 1, \ldots, n$, s'interprétant comme les réalisations de variables aléatoires X_i, indépendantes et de même loi que celle de X.

Le statisticien envisage deux spécifications possibles de cette loi :

- soit la loi normale $\mathcal{N}(\theta, \theta)$, de densité f_1
- soit la loi normale $\mathcal{N}(\theta, \theta^2)$ de densité f_2.

Dans les deux cas, θ est un **paramètre strictement positif** .

Préambule

On considère la fonction g définie par : $g(x) = e^{-ax^2+bx+c}$ où $a > 0$. Montrer que cette fonction est, à une constante multiplicative près, la densité d'une loi normale dont on précisera les paramètres.

Problème

Afin de choisir entre les deux spécifications des lois ci-dessus, le statisticien construit un modèle mixte, où les observations suivent la loi de densité $f = A(\lambda, \theta) f_1^{\lambda} f_2^{1-\lambda}$, où λ est un paramètre de $[0, 1]$ et f_1 et f_2 les densités respectives des lois $\mathcal{N}(\theta, \theta)$ et $\mathcal{N}(\theta, \theta^2)$.

1. (a) Montrer que la loi des observations (de densité f) est une loi normale dont on précisera les paramètres.
   (b) Calculer la fonction $A(\lambda, \theta)$.
2. On rappelle que la *vraisemblance du modèle* est la fonction (qui dépend par ailleurs de θ et de λ) :
$$(x_1, \ldots, x_n) \longmapsto L(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$$
   Les *estimateurs du maximum de vraisemblance* de θ et de λ, notés $\hat{\theta}_n$ et $\hat{\lambda}_n$, sont les quantités (dépendants des x_i) maximisant la vraisemblance, considérée comme dépendant de θ et de λ, à x_i **fixés**, ou, ce qui est équivalent, son logarithme. Ils peuvent être considérés chacun comme des réalisations d'une fonction des variables aléatoires X_i. *Les équations de vraisemblance* sont les conditions du premier ordre que doivent satisfaire ces estimateurs (on ne demande pas de vérifier que ces conditions caractérisent bien un maximum).
   (a) Écrire les équations de vraisemblance permettant de déterminer les estimateurs $\hat{\theta}_n$ et $\hat{\lambda}_n$.
   (b) Calculer explicitement ces estimateurs, qu'on exprimera en fonction des moments empiriques $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ et $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$.
3. Étudier la convergence en probabilité de $\hat{\lambda}_n$, quand n tend vers $+\infty$.
4. (a) Étudier la convergence en probabilité et la *normalité asymptotique* de S_n^2, quand n tend vers $+\infty$.
   [On rappelle que la normalité asymptotique consiste à étudier la convergence en loi vers une loi normale de $\sqrt{n}(S_n^2 - s^2)$ où s^2 est la limite en probabilité de S_n^2.]
   (b) En déduire la normalité asymptotique de $\hat{\lambda}_n$, sous l'hypothèse $H_0 : \lambda = 0$, quand n tend vers $+\infty$.
5. On suppose que l'on sait que θ ne peut prendre ses valeurs que dans $]0, \varepsilon]$, avec $0 < \varepsilon < 1$. Proposer un test de l'hypothèse $H_0 : \lambda = 0$ contre $H' : \lambda \neq 0$ avec une région critique indépendante de θ, pour *un risque de première espèce* valant au plus α.
   [On rappelle que le risque de première espèce est la probabilité de refuser (à tort) l'hypothèse nulle alors qu'elle est vraie]
6. (a) Étudier la normalité asymptotique de $\hat{\lambda}_n$ sous l'hypothèse $H_1 : \lambda = 1$, quand n tend vers $+\infty$.
   (b) En déduire la puissance asymptotique (en utilisant les approximations normales démontrées ci-dessus, lorsque n est assez grand) du test visé en 5, lorsque $\lambda = 1$, pour un risque de première espèce fixé au plus à α.
   On l'exprimera en introduisant la fonction de répartition Φ de la loi normale centrée réduite.
   [On rappelle que la puissance est la probabilité de refuser (à raison) l'hypothèse nulle lorsqu'elle est fausse, ici dans le cas particulier $\lambda = 1$.]