CONCOURS ADMINISTRATEUR EXTERNE DE L'INSEE

SESSION 2020

ÉPREUVE DE MATHÉMATIQUES

DURÉE : 4 heures

L'énoncé comporte 6 pages, numérotées de 1 à 6.

Tous documents et appareils électroniques interdits.

Tournez la page S.V.P.

Partie 1 : Algèbre-Analyse

Cette partie est constituée de deux exercices indépendants

Exercice 1

On considère $\mathbb{R}^n$ muni de son produit scalaire canonique, noté $\langle\ \rangle$ et de sa base canonique, $\mathcal{B} = (e_1, \ldots, e_n)$.
On note $\|\ \|$ la norme associée au produit scalaire $\langle\ ,\ \rangle$.
Si x est un vecteur de $\mathbb{R}^n$, on note X la matrice colonne de ses coordonnées dans $\mathcal{B}$.
On note $\mathrm{Sp}(A)$ l'ensemble des valeurs propres d'une matrice A de $\mathcal{M}_n(\mathbb{R})$ et ${}^tA$ sa transposée.
On dit qu'une matrice symétrique A est définie positive si, pour tout vecteur X non nul de $\mathbb{R}^n$, ${}^tXAX > 0$. Dans tout le problème A désigne une matrice de $\mathcal{M}_n(\mathbb{R})$ symétrique définie positive.

1. (a) Justifier qu'il existe une base orthonormale de $\mathbb{R}^n$, notée $\mathcal{B}' = (\varepsilon_1, \ldots, \varepsilon_n)$, constituée de vecteurs propres de A associés à des valeurs propres réelles $\lambda_1, \ldots, \lambda_n$.
   (b) Montrer l'équivalence suivante : A est définie positive $\iff \mathrm{Sp}(A) \subset \mathbb{R}_+^*$
2. On note $0 < \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$ les valeurs propres de A ; pour tout x non nul de $\mathbb{R}^n$, on définit la fonction r_A, appelée quotient de Rayleigh, par :
$$\forall x \in \mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}, \quad r_A(x) = \frac{\langle Ax, x\rangle}{\|x\|^2}$$
Établir, pour tout vecteur x non nul, l'encadrement suivant :
$$\lambda_1 \leqslant r_A(x) \leqslant \lambda_n$$
3. On conserve les notations des questions précédentes et on appelle conditionnement de A le réel noté C_A défini par : $C_A = \dfrac{\lambda_n}{\lambda_1}$.
On se propose dans cette question de démontrer la formule suivante, appelée inégalité de Kantorovitch :
$$\forall x \in \mathbb{R}^n, \quad \|x\|^4 \leqslant \langle Ax, x\rangle\langle A^{-1}x, x\rangle \leqslant \frac{1}{4}\left(\frac{1}{\sqrt{C_A}} + \sqrt{C_A}\right)^2 \|x\|^4 \tag{1}$$
   (a) i. Montrer que l'application suivante définit un produit scalaire sur $\mathbb{R}^n$:
$$\forall (x, y) \in (\mathbb{R}^n)^2, \quad (x|y) = \langle Ax, y\rangle$$
   On note $\|\ \|_A$ la norme associée.
   ii. Exprimer $\langle A^{-1}x, x\rangle$ et $\langle Ax, x\rangle$ à l'aide de $\|\ \|_A$.
   iii. En déduire l'inégalité suivante : $\|x\|^4 \leqslant \langle Ax, x\rangle\langle A^{-1}x, x\rangle$
   (b) Montrer que, pour établir la relation (1), il suffit de la vérifier pour un vecteur x vérifiant $\|x\|^2 = 1$.
   (c) On note donc un vecteur x de $\mathbb{R}^n$ de norme 1 et qui s'écrit $x = \displaystyle\sum_{k=1}^{n} x_k \varepsilon_k$.
   On considère un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et on définit la variable aléatoire Z par : $Z(\Omega) = \{\lambda_1, \ldots, \lambda_n\}$ et, pour tout i de $[\![1, n]\!]$, $\mathbb{P}([Z = \lambda_i]) = x_i^2$.
   i. Justifier que la relation précédente définit bien une loi de probabilité.
   ii. Calculer $\mathbb{E}(Z)$ et $\mathbb{E}\left(\dfrac{1}{Z}\right)$ en fonction de $\langle Ax, x\rangle$ et de $\langle A^{-1}x, x\rangle$.
   iii. Établir l'inégalité suivante : $\dfrac{1}{Z} \leqslant \dfrac{\lambda_1 + \lambda_n - Z}{\lambda_1\lambda_n}$.
   iv. En déduire que :
$$\mathbb{E}(Z)\mathbb{E}\left(\frac{1}{Z}\right) \leqslant -\frac{1}{\lambda_1\lambda_n}\left(\mathbb{E}(Z) - \frac{\lambda_1 + \lambda_n}{2}\right)^2 + \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}$$
   v. Déduire de ce qui précède l'inégalité de Kantorovitch.

Exercice 2:

Le préambule et la première partie de ce problème sont indépendantes. La deuxième partie combine des résultats des parties précédentes.

Préambule

Soient α un ***irrationnel*** >0 et x un réel >0. On s'intéresse à la série $\sum \frac{x^n}{\sin(\pi\alpha n)}$.

1.
   a) Montrer que les termes de cette série sont bien définis pour tout entier naturel $n \geq 1$.
   b) Montrer que la série diverge pour tout $x \geq 1$.

1ère partie

2. On s'intéresse ici à la suite $\{u_n\}$ définie par $u_0 > 0$ et la relation de récurrence $u_{n+1} = u_n^{u_n}$
   a) Montrer que, si $u_0 < \frac{1}{e}$, alors : $u_1 > \frac{1}{e}$.
   b) Montrer que la suite est convergente pour $0 < u_0 \leq 1$.
   c) Montrer que la suite tend vers $+\infty$ pour $u_0 > 1$.

3. On se place dorénavant dans le cas $\boxed{u_0 > 1.}$
   a) Montrer que la série $\sum \frac{1}{u_n}$ est convergente.
   b) Montrer que : $\exists\, N \in \mathbb{N}, \forall\, k \in \mathbb{N} : u_{N+k} \geq k+2$.
   c) Montrer qu'il existe $C > 0$ tel que : $\forall\, n \geq N : \sum_{k=n+1}^{+\infty} \frac{1}{u_k} \leq \frac{C}{u_{n+1}}$
   d) En déduire que : $\forall\, n \geq N : u_n \sum_{k=n+1}^{+\infty} \frac{1}{u_k} \leq \frac{C}{u_n^{u_n - 1}}$.

2ème partie

4. On se restreint maintenant au cas où $\boxed{u_0 \in \mathbb{N},\, u_0 \geq 2.}$ On pose : $\boxed{\alpha = \sum_{n=0}^{+\infty} \frac{1}{u_n}.}$ et on admet dans un premier temps que α est ***irrationnel***.
   a) Montrer que : $\forall\, n \in \mathbb{N} : u_n \sum_{k=0}^{n} \frac{1}{u_k} \in \mathbb{N}$.
   b) En déduire que : $\forall\, n \in \mathbb{N} : \frac{x^{u_n}}{|\sin(u_n \pi \alpha)|} \geq \frac{1}{\pi C} x^{u_n} u_n^{u_n - 1}$
   c) En déduire que la série $\sum \frac{x^n}{\sin(\pi\alpha n)}$ diverge pour $0 < x < 1$.

5. **On va démontrer que le α défini à la question 4 est bien irrationnel**. On raisonne par l'absurde en supposant que $\alpha = \dfrac{p}{q}$ avec *p* et *q* entiers naturels non nuls.

a) Montrer que : $\forall\, n \in \mathbb{N} : q\, u_n \displaystyle\sum_{k=n+1}^{+\infty} \frac{1}{u_k} \in \mathbb{N}$.

b) En déduire une contradiction (*on fera tendre n vers* $+\infty$).

Partie 2 : Probabilités-Statistiques

Cette partie est constituée de deux exercices indépendants

Exercice 1

Dans tout l'exercice, X est une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, admettant une densité f nulle sur $]-\infty, 0[$. On suppose que la restriction de f à $[0, +\infty[$ est continue et strictement positive.

On note F la fonction de répartition de X.

On considère une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, indépendantes et de même loi que X.

On pose $Z_1 = X_1$ et on note Z_2 l'application définie sur Ω par :

$$\forall \omega \in \Omega, \quad Z_2(\omega) = \begin{cases} X_n(\omega) \text{ si } n \text{ est le plus petit des entiers } k \text{ tels que } X_k(\omega) > X_1(\omega) \\ \\ X_1(\omega) \text{ si un tel entier n'existe pas} \end{cases}$$

On admet que Z_2 est une variable aléatoire, définie sur $(\Omega, \mathcal{A}, \mathbb{P})$.

1. (a) Établir, pour tout entier supérieur ou égal à 2 et tout réel t positif, l'égalité suivante :

$$\mathbb{P}\left(\bigcap_{k=2}^{n} [X_k \leqslant X_1]\right) \leqslant \big(F(t)\big)^n + 1 - F(t)$$

   (b) En déduire que, presque sûrement, $Z_2 > Z_1$.

2. On considère dans cette question un couple (x, y) de réels positifs et h un réel srictement positif. On pose :

$$\varphi(x, y) = \mathbb{P}\big([Z_1 \leqslant x] \cap [Z_2 - Z_1 > y]\big)$$

   (a) Justifier l'égalité suivante :

$$[Z_1 \leqslant x + h] = [Z_1 \leqslant x] \cup [x < Z_1 \leqslant x + h]$$

   (b) En déduire que :

$$\varphi(x + h, y) - \varphi(x, y) = \mathbb{P}\big([x < Z_1 \leqslant x + h] \cap [Z_2 - Z_1 > y]\big)$$

   (c) Établir la formule suivante :

$$\varphi(x + h, y) - \varphi(x, y) = \sum_{j=2}^{+\infty} \mathbb{P}\left([x \leqslant X_1 \leqslant x + h] \cap \left[\bigcap_{i=2}^{j-1} [X_i \leqslant X_1]\right] \cap [X_j > y + X_1]\right)$$

   (d) En déduire l'encadrement suivant :

$$\frac{F(x+h) - F(x)}{1 - F(x)}\big(1 - F(x+y+h)\big) \leqslant \varphi(x+h, y) - \varphi(x, y) \leqslant \frac{F(x+h) - F(x)}{1 - F(x+h)}\big(1 - F(x+y)\big)$$

   (e) Calculer $\displaystyle\lim_{h\to 0^+} \frac{\varphi(x+h, y) - \varphi(x, y)}{h}$.

   (f) En admettant que le résultat précédent soit encore valable quand h tend vers 0 par valeurs inférieures, calculer $\dfrac{\partial \varphi}{\partial x}(x, y)$ en fonction de f et de F.

3. On suppose dans cette question que X suit la loi exponentielle de paramètre λ.

   (a) Montrer, pour tout couple (x, y) de réels positifs, que : $\varphi(x, y) = (1 - e^{-\lambda x})e^{-\lambda y}$.

   (b) Déterminer la fonction de répartition de $Z_2 - Z_1$.

   (c) Montrer que Z_1 et $Z_2 - Z_1$ sont indépendantes.

Exercice 2 :

On considère une suite de variables aléatoires indépendantes, $\{Z_i\}$, suivant chacune la loi de BERNOULLI $\mathcal{B}\ (1, p_i(\theta))$, où les p_i sont des fonctions de classe C^1 et θ un paramètre réel.

On dispose de *n* observations de ces variables.

1.

a) Donner l'expression de $f_i(q, \theta) = P\{Z_i = q\}$.

b) En déduire la *vraisemblance* du modèle dont les observations sont les valeurs de $(Z_1, \ldots, Z_n)$.

On rappelle que la *vraisemblance d'un modèle* dont les observations Z_i sont discrètes, indépendantes et prennent les valeurs q_i, est la fonction :

$$(q_1, \ldots, q_n, \theta) \rightarrow L(q_1, \ldots, q_n, \theta) = \prod_{i=1}^{n} f_i(q_i, \theta).$$

c) En déduire l'équation du maximum de vraisemblance pour l'estimation du paramètre θ. *On ne cherchera pas à résoudre cette équation.*

L'*estimateur du maximum de vraisemblance* de θ , noté $\hat{\theta}_n$, est la quantité (dépendant des q_i) maximisant la vraisemblance (considérée comme dépendant de θ , les q_i étant **fixés**), ou, ce qui est équivalent, son logarithme. Cet estimateur peut être considéré comme une variable aléatoire dont la réalisation est fonction de celles des variables aléatoires Z_i. L'*équation de vraisemblance* est la condition du 1er ordre que doit vérifier cet estimateur (on ne demande pas de vérifier que cette condition caractérise bien un maximum).,

2.

a) Déterminer les fonctions $p_i(\theta)$ telles que, pour tout θ : $\dfrac{p'_i(\theta)}{p_i(\theta)[1-p_i(\theta)]}$ soit une constante x_i.

b) Que deviennent alors les équations de la question 1 ?

On suppose maintenant qu'on dispose d'une suite de couples de variables aléatoires indépendantes, $\{(Z_i, X_i)\}$, tels que, pour tout i :

- la loi de X_i est une loi discrète définie par : $P\{X_i = x_k\} = \pi_k$ pour $k = 1, \ldots, K$, les x_k (deux à deux distincts) et les π_k étant fixés et connus ;
- la loi conditionnelle de Z_i sachant $X_i = x_k$ est une loi de BERNOULLI $\mathcal{B}$(1, $p(\theta, k)$), où la fonction $\theta \rightarrow p(\theta, k)$ est de classe C^1.

3. Les équations du maximum de vraisemblance pour l'estimation du paramètre θ sont-elles modifiées dans ce cas par rapport à celles de la question 1 (toujours lorsqu'on dispose de *n* observations des variables (Z_i, X_i)) ?

4. Pour un couple générique (Z, X) correspondant à une valeur quelconque de l'indice *i* ci-dessus, on note : $Y = \begin{pmatrix} Z\, 1_{X=x_1} \\ \cdot \\ \cdot \\ Z\, 1_{X=x_K} \end{pmatrix}$.

   a) Calculer la matrice de variance-covariance de Y.

   b) On note, pour tout entier naturel n et pour tout entier $k \in \{1, ..., K\}$:
   $$\hat{p}_{k,n} = \frac{1}{n} \sum_{i=1}^{n} Z_i\, 1_{X_i = x_k}.$$
   Pour n fixé, les $\hat{p}_{k,n}$ sont-ils indépendants ?

   c) On note enfin $\hat{p}_n$ le vecteur de composantes $\hat{p}_{k,n}$. Étudier la convergence en probabilité de $\hat{p}_n$ et sa normalité asymptotique quand $n \to +\infty$.

5. On se place ici dans le cas où : $p(\theta, k) = \dfrac{1}{1+e^{-\theta\, x_k}}$.

   a) Déterminer l'estimateur des moindres carrés ordinaires de θ, soit $\hat{\theta}_n$, dans le modèle linéaire :
   $$\ln \frac{\hat{p}_{k,n}/\pi_k}{1-\hat{p}_{k,n}/\pi_k} = \theta\, x_k + u_k,\, k = 1, ..., K.$$

   b) Pour *k* fixé, étudier la normalité asymptotique de $\ln \dfrac{\hat{p}_{k,n}/\pi_k}{1-\hat{p}_{k,n}/\pi_k}$ quand $n \to +\infty$.

   c) Étudier la convergence en probabilité de cet estimateur $\hat{\theta}_n$, quand $n \to +\infty$.

   d) Étudier sa normalité asymptotique.