SESSION 2017

COMPOSITION DE MATHÉMATIQUES

Sujet : INSEE administrateur externe

DURÉE : 4 heures

L'énoncé comporte 5 pages, numérotées de 2 à 6.

L'épreuve est constituée de deux parties indépendantes, chacune comptant pour moitié dans la note finale.

L'usage de la calculatrice est interdit

Tournez la page S.V.P.

Partie 1 : Analyse-algèbre

On considère un entier n supérieur ou égal à 2, a_1,a_2,..., a_n, n nombres complexes donnés et la matrice A définie par :

$$A = \begin{pmatrix} 1+a_1 & 1 & \dots & \dots & 1 \\ 1 & 1+a_2 & 1 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 1 \\ 1 & \dots & \dots & 1 & 1+a_n \end{pmatrix}$$

c'est-à-dire la matrice dont les termes diagonaux sont $1+a_1$, $1+a_2$,..., $1+a_n$ et dont tous les autres termes valent 1.

Partie A

On suppose dans cette partie que les nombres a_i, $i \in [\![1,n]\!]$, sont réels.

1. Justifier, sans calcul, que A est diagonalisable.
2. On suppose dans cette question que, pour tout j de $[\![1,n]\!]$, $a_j = j-1$.

   (a) Montrer que le réel λ est valeur propre de A si et seulement si λ vérifie :

   $$\sum_{k=0}^{n-1} \frac{1}{\lambda - k} = 1.$$

   (b) On considère la fonction f_n définie sur $\mathbb{R} \setminus \{0, 1, \dots, n-1\}$ par :

   $$f_n(x) = \sum_{k=0}^{n-1} \frac{1}{x-k}.$$

   En étudiant la fonction f_n, montrer que A admet n valeurs propres réelles distinctes.

   (c) On note λ_n la plus grande valeur propre de A.

   i. Établir, pour tout réel y positif, l'inégalité suivante :

   $$\sum_{j=1}^{n} \frac{1}{y+j} \leqslant \int_0^n \frac{1}{t+y} dt.$$

   ii. En déduire que $f_n(n + \frac{n}{e-1}) \leqslant 1$.

   iii. Montrer de même que $f_n(n - 1 + \frac{n}{e-1}) \geqslant 1$.

   (d) Déduire de ce qui précède un équivalent simple de λ_n quand n tend vers $+\infty$:

   $$\lambda_n \underset{n \to +\infty}{\sim} \frac{ne}{e-1}.$$

Partie B

On revient dans cette partie au cas général où les a_i sont des nombres complexes et on se propose d'étudier l'inversibilité de A.

À cet effet, on considère les deux matrices-colonnes X et Y de $\mathcal{M}_{n,1}(\mathbb{C})$, $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ et $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ et le système (S) :

$AX = Y$.

On pose $s = \sum_{i=1}^{n} x_i$.

3. Écrire le système d'équations vérifié par les x_i.
4. On suppose dans cette question qu'aucun des a_i est nul.

(a) Montrer que, si $\sum_{j=1}^{n} \frac{1}{a_j} \neq -1$, la matrice A est inversible.

(b) Dans le cas où $\sum_{j=1}^{n} \frac{1}{a_j} = -1$, déterminer $\mathrm{Im}A$ et donner, lorsqu'elles existent, les solutions de l'équation $AX = Y$.

5. On suppose, dans cette question, que seul $a_1 = 0$, les autres a_i étant non nuls. Résoudre l'équation $AX = Y$. La matrice A est-elle inversible ?

6. Déduire de l'étude précédente une condition nécessaire et suffisante pour que A soit inversible.

Partie C

On s'intéresse dans cette partie à la diagonalisation de A.

7. (a) En utilisant les résultats de la partie B, donner une condition nécessaire et suffisante pour qu'un complexe λ, différent de chacun des a_i, soit valeur propre de A.
   (b) En déduire une équation polynômiale de degré n, fonction des a_i, satisfaite par les valeurs propres de A qui sont distinctes des a_i.
   (c) Donner, pour chacune de ces valeurs propres, la dimension et une base du sous-espace propre associé.

8. On étudie dans cette question le fait qu'un des a_i puisse être valeur propre de A.
   (a) Montrer, toujours à l'aide de la partie B, que si, pour tout i de $[\![2, n]\!]$, $a_i \neq a_1$, alors a_1 ne peut pas être valeur propre de A.
   (b) On suppose dans cette question que $a_1 = a_2 \ldots = a_p$ ($p \geqslant 2$) et que les autres a_i sont distincts de a_1. Montrer que a_1 est valeur propre de A et donner la dimension du sous-espace propre correspondant.

9. On suppose que, parmi les a_i, il n'y a au total que q valeurs distinctes dont q_1 n'apparaissent qu'une seule fois et q_2 au moins deux fois. On renumérote les a_i et on note :

   - a'_i, $i \in [\![1, q_1]\!]$ les valeurs distinctes des a_i n'apparaissant qu'une seule fois dans la matrice A.
   - a''_j, $j \in [\![1, q_2]\!]$ les valeurs des a_i apparaissant chacune en nombre N_j dans la matrice A.

   On a donc :
$$n = q_1 + \sum_{j=1}^{q_2} N_j.$$
   (a) En utilisant les résultats de la question 7, montrer que toute valeur propre de A, distincte de chacun des a_i, satisfait une équation polynômiale de degré q et que, réciproquement, toute racine de cette équation est valeur propre de A.
   (b) On suppose que cette équation possède r racines distinctes. Calculer la somme des dimensions des sous-espaces propres de A identifiés dans les questions 7(c) et 8(b).
   (c) En déduire une condition nécessaire et suffisante pour que A soit diagonalisable.

Partie D

On revient dans cette partie au cas où les a_i sont réels et on se propose de redémontrer, par des considérations analytiques, que A est diagonalisable.

10. Montrer que toutes les valeurs propres de A sont réelles.

11. On considère une équation, d'inconnue complexe z, de la forme $\sum_{i=1}^{q} \frac{\alpha_i}{z - b_i} = 1$, où les α_i sont des réels strictement positifs et les b_i des complexes distincts.
    (a) Montrer que cette équation est équivalente à l'équation polynômiale $Q(z) = 0$, où :
$$Q(X) = P(X) - \sum_{j=1}^{q} \alpha_j P_j(X),$$
    avec : $P(X) = \prod_{i=1}^{q}(X - b_i)$ et $P(X) = (X - b_j)P_j(X)$, où $j \in [\![1, q]\!]$.

(b) On suppose que z est une racine double de Q. Établir l'égalité suivante :

$$P'(z) - \sum_{j=1}^{q} \alpha_j P_j'(z) = P'(z) \left(1 - \sum_{j=1}^{q} \frac{\alpha_j}{z - b_j}\right) + \sum_{j=1}^{q} \alpha_j \frac{P_j(z)}{z - b_j},$$

puis que :

$$0 = P(z) \sum_{j=1}^{q} \frac{\alpha_j}{(z - b_j)^2}.$$

(c) En déduire, que si les b_i sont réels, le polynôme Q ne peut admettre de racine double réelle.

12. Conclure.

Partie 2 : Probabilités-statistiques

Cette partie est constituée de deux exercices indépendants.

Exercice 1

On considère deux variables aléatoires X et Y, indépendantes et suivant toutes les deux la loi exponentielle de paramètre λ ($\lambda > 0$), c'est-à-dire qu'une densité de X et de Y est donnée par :

$$f_X(t) = f_Y(t) = \begin{cases} \lambda e^{-\lambda t} \text{ si } t \geqslant 0 \\ \\ 0 \text{ sinon} \end{cases}$$

On pose $T = X - Y$ et $Z = \min(X, Y)$.

1. Déterminer la fonction de répartition de Z et vérifier que Z suit une loi exponentielle dont on précisera le paramètre.
2. (a) Déterminer une densité de $-Y$.
   (b) À l'aide d'un produit de convolution, déterminer une densité de $X - Y$.
   (c) Donner la fonction de répartition de T.
3. On pose $W = (T, Z)$ et on note F_W la fonction de répartition de W.
   (a) i. Établir, pour tout couple (t, z) de réels positifs, la formule suivante :
   $$F_W(t,z) = \int_0^z \left(\int_0^{y+t} f_X(x)\mathrm{d}x\right) f_Y(y)\mathrm{d}y + \int_z^{+\infty} \left(\int_0^z f_X(x)\mathrm{d}x\right) f_Y(y)\mathrm{d}y.$$
   ii. En déduire, pour tout couple (t, z) de réels positifs, l'expression explicite de $F_W(t, z)$ en fonction de z et de t.
   (b) i. Établir, pour tout couple (t, z) de réels, avec z positif et t négatif, la formule suivante :
   $$F_W(t,z) = \int_0^z \lambda e^{-\lambda x} \left(\int_{x-t}^{+\infty} \lambda e^{-\lambda y}\mathrm{d}y\right) \mathrm{d}x.$$
   ii. En déduire l'expression de $F_W(t, z)$ pour z positif et t négatif.
4. Montrer que Z et T sont indépendantes.

Exercice 2

<u>Préliminaires</u>

$\boxed{\text{P1}}$ On considère deux variables aléatoires indépendantes X_1 et X_2, de carré intégrable et non constantes. Donner une condition nécessaire et suffisante pour que $\mathrm{Var}(X_1X_2) = \mathrm{Var}(X_1)\mathrm{Var}(X_2)$.

$\boxed{\text{P2}}$ Soit (X, U) un couple de variables aléatoires réelles suivant la loi normale $\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\mu^2 & 0\\ 0 & \sigma^2\end{pmatrix}\right)$ avec $\mu > 0$ et $\sigma > 0$. On pose $Y = a + bX + U$, où a et b sont des paramètres réels.
Déterminer la loi du couple (X, Y) et donner sa densité. On notera $\mathcal{L}(a, b, \mu, \sigma)$ la loi trouvée.

■ ■ ■ ■ ■ ■

On considère maintenant, dans la suite du problème, une suite de couples de variables aléatoires réelles, (X_i, Y_i), indépendants entre eux, de même loi $\mathcal{L}(a, b, \mu, \sigma)$. On suppose que l'on dispose de N observations de ces couples, pour $i \in [\![1, N]\!]$. Sauf mention explicite du contraire, les quatre paramètres a, b, μ, σ sont inconnus et feront l'objet d'estimations.

1. (a) Construire un estimateur de a, sans biais, fondé sur les seules observations de Y_i, que l'on notera $\hat{a}_Y$.
   (b) Calculer la variance de $\hat{a}_Y$ et montrer que cet estimateur est convergent.
   (c) Dans le cas où b, μ, σ sont connus, proposer un test de l'hypothèse nulle $H_0 : \{a = 0\}$, fondé sur cet estimateur, avec un risque de première espèce (*probabilité de rejeter l'hypothèse nulle alors qu'elle vraie*) de valeur donnée $\alpha \in]0, 1[$.
   On exprimera le résultat au moyen de la fonction de répartition Φ de la loi $\mathcal{N}(0, 1)$.

2. On suppose dans cette question $\boxed{\mu \text{ connu}}$.
   (a) À partir de la considération de $\mathbb{E}(X_iY_i)$, utiliser la méthode des moments pour déterminer un estimateur sans biais et convergent de b (*on ne demande pas de démontrer explicitement les propriétés demandées*).
   (b) Donner la variance de cet estimateur (*on rappelle que, si* $T \hookrightarrow \mathcal{N}(0,1)$, $\mathrm{Var}(T^2) = 2$).

3. (a) Dans le modèle linéaire $Y_i = a + bX_i + U_i$, donner la valeur de l'estimateur des moindres carrés ordinaires de b, noté $\hat{b}_0$ (*on l'exprimera en fonction de b et des variables X_i et U_i*).
   (b) Étudier sa convergence lorsque N tend vers $+\infty$.
   (c) Calculer sa variance.
   On fera ici l'approximation $\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}_N\right)^2 \approx \mu^2$, *valable pour N assez grand, avec* $\overline{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$.

4. Toujours dans ce même modèle linéaire, on considère l'estimateur des moindres carrés ordinaires de a, soit $\hat{a}_0$.
   (a) Donner l'expression de $\hat{a}_0$ en fonction des paramètres, de $\hat{b}_0$ et des variables X_i et U_i.
   (b) En déduire la variance de $\hat{a}_0$.
   (c) Comparer cette variance à celle de $\hat{a}_Y$.

5. On suppose que, pour des raisons de confidentialité, les variables X_i ont été « floutées », c'est-à-dire que l'on ne peut observer que les variables $X_i^* = X_i + \varepsilon_i$, où les ε_i sont des variables aléatoires indépendantes entre elles, indépendantes des (X_i, U_i) et suivant toutes la loi $\mathcal{N}(0, \beta^2)$ avec $\beta > 0$.
   (a) Déterminer la loi conditionnelle de X_i sachant $[X_i^* = x^*]$.
   (b) Montrer que le modèle linéaire de la question 3(*a*) peut alors s'écrire sous la forme $Y_i = a + bX_i^* + U_i^*$, où les U_i^* sont d'espérance nulle et à définir en fonction des variables et des paramètres préexistants.
   (c) Montrer que l'estimateur des moindres carrés ordinaires de b dans ce nouveau modèle, noté $\hat{b}_0^*$, n'est pas convergent. Quelles conclusions pratiques en tirez-vous vis-à-vis des utilisateurs de ces données ?