INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

ÉCOLE NATIONALE DE LA STATISTIQUE ET DE L'ANALYSE DE L'INFORMATION

Concours interne d'attaché statisticien de l'Insee

SESSION 2020

ÉPREUVE DE MATHÉMATIQUES ET STATISTIQUES

Durée 4 heures

Coefficient 3

Sans documents – L'usage de la calculatrice est interdit

Le sujet comprend 8 pages

Le sujet se compose de 4 exercices. Le barème est donné à titre indicatif.

Exercice 1 .... 6 pts

Depuis le début de l'année (les 20 jours ouvrés du mois de janvier) Alice a relevé le temps qu'elle devait attendre son bus. Elle suppose que ce temps d'attente peut être représenté par une variable aléatoire A qui suit une loi exponentielle de paramètre λ.

Les 20 relevés du mois de janvier d'Alice figure dans le tableau ci-dessous (le temps est en minute) :

jour	1	2	3	4	5	6	7	8	9	10
Temps	2	1	7	3	1	8	7	17	8	2

Jour	11	12	13	14	15	16	17	18	19	20
Temps	1	4	2	2	11	4	10	2	2	6

Alice a comptabilisé qu'au mois de janvier elle avait attendu au total 1h40 min.

1. (a) Rappeler la densité, l'espérance et la variance d'une loi exponentielle de paramètre λ.

   (b) Retrouver l'expression de la fonction de répartition de A, c'est-à-dire la fonction F définie par $F_A(t) = P(A \leqslant t)$ pour $t \in \mathbb{R}$.

   (c) À partir des relevés d'Alice, donner une estimation pour la valeur de λ.
   λ sera identifié à cette estimation dans la suite de l'exercice.

2. En février, Alice fait ses trajets en bus avec son ami Bob.

   (a) Montrer que la probabilité d'attendre moins de 8 minutes est de $p = 0,798$. *On pourra utiliser la table située en page 3.*

   (b) Bob vient la rejoindre alors qu'elle attend déjà depuis 3 minutes.
   On note B la variable aléatoire représentant le temps d'attente de Bob. Calculer la fonction de répartition de la variable aléatoire B, c'est-à dire

   $$F_B(t) = P(B \leqslant t) = P(A - 3 \leqslant t | A > 3)$$

   pour $t \in \mathbb{R}$.

   (c) Reconnaître la loi de B.

   (d) Quelle est l'espérance de B ?

   (e) Quand Bob est arrivé, Alice lui a annoncé qu'ils n'auront plus que 2 minutes à attendre en moyenne puisqu'elle a déjà attendu 3 minutes. Alice a-t-elle raison ?

3. On s'intéresse aux 20 jours ouvrés du mois de mars. Sur ces 20 jours, on note N le nombre de jours pour lesquels Alice a attendu moins de 8 min.
   Les jours ouvrés du mois de mars étant numérotés de 1 à 20, R désigne la variable aléatoire définie comme le numéro du premier jour pour lequel Alice a attendu (strictement) plus de 8 minutes son bus.
   Par convention, si pendant tout le mois Alice a attendu moins de 8 minutes, alors $R = 21$. On suppose que les temps d'attente journaliers sont indépendants.

(a) Donner la loi de N.

(b) Donner sans calcul l'espérance et la variance de N.

(c) Donner la loi de R, c'est-à-dire que l'on donnera une expression de $P(R = k)$ pour k entier entre 1 et 21.

(d) On définit, sur $[0, 1[$, la fonction S_n par $S_n(x) = \sum_{i=1}^{n} x^i$.

i. Calculer $S_n(x)$.

ii. En dérivant S_n, montrer que

$$\sum_{i=1}^{n} i x^{i-1} = \frac{n x^{n+1} - (n+1)x^n + 1}{(1-x)^2}.$$

(e) En déduire l'espérance de R.

4. Alice rencontre Chloé qui lui conseille de prendre le métro. On note C la variable aléatoire du temps d'attente à la station de métro. On suppose que C suit une loi uniforme sur $[0, 10]$. Alice décide de prendre le métro les 20 jours ouvrés du mois d'avril.

(a) Calculer l'espérance et la variance de la variable aléatoire C.

(b) Comparer et interpréter les variances des variables A et C.

(c) Calculer la probabilité qu'Alice attende moins de 8 minutes.

(d) Calculer la probabilité qu'Alice attende 8 minutes sachant qu'elle a déjà attendu 3 minutes.

(e) Comparez le résultat de la question précédente avec le résultat de la question 2.(b).

5. A partir du mois de mai, Alice adopte la stratégie suivante :

— le premier jour, elle prend le bus,

— si le jour n elle attend strictement plus de 8 minutes le bus alors le jour $n + 1$ elle prend le métro, sinon elle continue de prendre le bus.

— si le jour n elle attend strictement plus de 8 minutes le métro alors le jour $n + 1$ elle prend le bus, sinon elle continue de prendre le métro.

On note A_n l'évènement « Alice prend le bus le jour n » et $a_n = P(A_n)$.
Pour rappel le nombre p est la probabilité qu'Alice attende son bus moins de 8 minutes, défini en 2.(a).

(a) Montrer que $a_{n+1} = (p - 0,2)a_n + 0,2$. *On pourra utliser la formule des probabilités totales.*

(b) Soit l le nombre solution de l'équation $l = (p - 0,2)l + 0,2$. Donner une expression de l en fonction de p.

(c) Montrer que la suite $(u_n)_{n\geqslant 1}$ définie par $u_n = a_n - l$ est une suite géométrique de raison que l'on déterminera.

(d) En déduire une expression de a_n en fonction de n.

(e) La suite $(a_n)_{n\geqslant 1}$ converge-t-elle ? Si oui, déterminer sa limite.

Table : fonction $x \mapsto e^{-x}$

x	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	1.000	0.905	0.819	0.741	0.670	0.607	0.549	0.497	0.449	0.407
1	0.368	0.333	0.301	0.273	0.247	0.223	0.202	0.183	0.165	0.150
2	0.135	0.122	0.111	0.100	0.091	0.082	0.074	0.067	0.061	0.055
3	0.050	0.045	0.041	0.037	0.033	0.030	0.027	0.025	0.022	0.020
4	0.018	0.017	0.015	0.014	0.012	0.011	0.010	0.009	0.008	0.007
5	0.007	0.006	0.006	0.005	0.005	0.004	0.004	0.003	0.003	0.003

Lecture : $e^{-2,5} = 0.082$

Exercice 2 .... 3 pts

Soit f l'application linéaire de $\mathbb{R}^3$ dans $\mathbb{R}^3$ définie par

$$f(x, y, z) = (-x + y + z, -6x + 4y + 2z, 3x - y + z).$$

1. Écrire la matrice de f dans la base canonique de $\mathbb{R}^3$.
2. (a) Montrer que $f \circ f = 2f$.
   (b) En déduire que si $v \in \mathrm{Im}(f)$ alors $f(v) = 2v$.
   (c) Calculer le noyau $\mathrm{Ker}\,(f)$.
3. Montrer que $\mathrm{Ker}\,(f)$ et $\mathrm{Im}(f)$ vérifient (on dit alors que $\mathrm{Ker}\,(f)$ et $Im(f)$ sont des sous-espaces vectoriels supplémentaires de $\mathbb{R}^3$) :
   (i) $\mathrm{Ker}\,(f) \cap \mathrm{Im}(f) = \{0\}$
   (ii) Tout élément v de $\mathbb{R}^3$ s'écrit $v = w + t$ avec $w \in \mathrm{Ker}\,(f)$ et $t \in \mathrm{Im}(f)$.
4. (a) Trouver une base $\mathscr{B} = (b_1, b_2, b_3)$ de $\mathbb{R}^3$ telle que $b_1 \in \ker(f)$ et $b_2, b_3 \in \mathrm{Im}(f)$.
   (b) L'application f est-elle diagonalisable ?

Exercice 3 .... 5 pts

On définit la fonction f par :

$$f_{\alpha,\beta}(t) = \begin{cases} t^{\alpha-1}(1-t)^{\beta-1} & \text{si } t \in [0,1] \\ 0 & \text{sinon.} \end{cases}$$

pour $\alpha \geqslant 1$ et $\beta \geqslant 1$. On note $I(\alpha,\beta) = \int_0^1 f_{\alpha,\beta}(t)\,\mathrm{d}t$.

1. (a) Établir que $I(\alpha,\beta+1) = \frac{\beta}{\alpha} I(\alpha+1,\beta)$.
   (b) Montrer $I(\alpha+1,\beta) + I(\alpha,\beta+1) = I(\alpha,\beta)$
   (c) Calculer $I(m,1)$ pour $m \in \mathbb{N}^*$.
   (d) Montrer par récurrence (sur n) que $I(m+1,n+1) = \dfrac{m!n!}{(m+n+1)!}$, pour $(m,n) \in \mathbb{N} \times \mathbb{N}$.
2. À propos de la densité
   (a) Calculer la dérivée de $f_{\alpha,\beta}$ sur $]0,1[$.
   (b) Déterminer le signe de $f'_{\alpha,\beta}$. *On pourra faire plusieurs cas suivant si α et β valent 1 ou pas.*
   (c) Dresser le tableau de variation de $f_{\alpha,\beta}$.
   (d) On note $g_{\alpha,\beta} = \dfrac{f_{\alpha,\beta}}{I(\alpha,\beta)}$. Montrer que $g_{\alpha,\beta}$ définit une densité de probabilité.
   Cette loi s'appelle loi Beta et se note $\mathcal{B}(\alpha,\beta)$.
3. Quelques caractéristiques de la loi $\mathcal{B}(m,n)$ pour $m \in \mathbb{N}^*$ et $n \in \mathbb{N}^*$
   (a) Calculer l'espérance de la loi $\mathcal{B}(m,n)$.
   (b) Calculer le moment d'ordre 2 de $\mathcal{B}(m,n)$.
   (c) En déduire la variance de $\mathcal{B}(m,n)$.
4. Statistique d'ordre
   Soient $X_1, X_2, \ldots, X_n$ des variables indépendantes suivant toutes la loi uniforme sur $[0,1]$.
   On note h la densité de la loi uniforme sur $[0,1]$ et H sa fonction de répartition, c'est-à-dire $H(t) = P(X_1 \leqslant t)$.
   On note $Y = \max\limits_{1 \leqslant i \leqslant n}(X_i)$.
   On appelle F la fonction de répartition de Y, c'est-à-dire que $F(t) = P(Y \leqslant t)$. On admet que Y admet une densité et on note f sa densité.
   (a) Montrer que $F(t) = (H(t))^n$.
   (b) Donner une expression de $H(t)$ en fonction de t.
   (c) On admet que pour $t \in]0,1[$ on a $f(t) = F'(t)$. Calculer f sur $]0,1[$.
   (d) Que vaut f sur $]-\infty,0[$? Sur $]1,+\infty[$?
   (e) Montrer que Y suit une loi Beta dont on déterminera les paramètres.

Exercice 4 .... 6 pts

La partie II est indépendante des deux autres parties. Dans la partie III on pourra admettre des résultats des parties I et II.
Dans l'exercice si A est une matrice, ${}^tA$ désigne la matrice transposée de A. De même, si $u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ est un vecteur, ${}^tu = (u_1, u_2, u_3)$ désigne le vecteur transposé.

Partie I

On note $M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{pmatrix}$, et $< u, v >= {}^tu.v = u_1v_1 + u_2v_2 + u_3v_3$ désigne le produit scalaire standard de deux vecteurs u et v de $\mathbb{R}^3$. On note $\|u\| = \sqrt{< u, u >}$ la norme euclidienne.

1. (a) Vérifier que $a = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ est un vecteur propre de M pour la valeur propre 4.
   (b) Montrer que le noyau de M est de dimension 1, et exhiber un vecteur non-nul b de ce noyau.
   (c) Montrer que 1 est valeur propre de M et exhiber un vecteur propre c pour cette valeur propre.
2. (a) Calculer $\|a\|$, $\|b\|$ et $\|c\|$.
   On pose $a' = \frac{1}{\|a\|}a$, $b' = \frac{1}{\|b\|}b$ et $c' = \frac{1}{\|c\|}c$.
   On note A la matrice carrée de taille 3 dont les colonnes sont $\left(\begin{array}{c|c|c} a' & b' & c' \end{array} \right)$.
   (b) Montrer que les vecteurs a', b', c' sont deux à deux orthogonaux, pour le produit scalaire $<,>$.
   (c) Que vaut ${}^tAA$?
3. Montrer que $M = AD\,{}^tA$ pour une matrice diagonale D à préciser.
4. Trouver toutes les matrices diagonales S solution de l'équation $S^2 = D$. Dans la suite on note T la solution ayant tous ses coefficients positifs.
5. Soit $u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \in \mathbb{R}^3$. On souhaite montrer que ${}^tuMu \geqslant 0$.
   (a) Montrer que $< T\,{}^tAu, T\,{}^tAu >= {}^tuMu$.
   (b) Conclure.
6. Soit $H = \left\{ v \in \mathbb{R}^3 ; < \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, v >= 0 \right\}$ l'orthogonal de Vect $\left(\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right)$. Trouver une base de H. Quelle est la dimension de H ?

Partie II

Dans les questions 1 et 2 on redémontre des résultats sur la covariance, utiles dans la suite. La question 3 étudie le lien entre covariance et indépendance.

On rappelle que si X_1, X_2 sont deux variables aléatoires réelles, la covariance de X_1 et X_2, notée $\mathrm{Cov}(X_1, X_2)$ est le nombre $\mathbb{E}\left((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))\right)$.
La variance de X_1 est $\mathbb{V}(X_1) = \mathrm{Cov}(X_1, X_1)$.
Dans la suite on pourra utiliser la propriété suivante : si X_1 et X_2 sont indépendantes, alors $\mathbb{E}(X_1 X_2) = \mathbb{E}(X_1).\mathbb{E}(X_2)$.

1. Montrer que $\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_1) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$.
2. En déduire que si X_1' est une autre variable aléatoire réelle et si $a \in \mathbb{R}$,
$$\mathrm{Cov}(aX_1 + X_1', X_2) = a.\mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(X_1', X_2).$$
3. (a) On suppose que X_1 et X_2 sont indépendantes. Que vaut dans ce cas $\mathrm{Cov}(X_1, X_2)$?
   (b) On suppose que X_1 suit la loi donnée par le tableau suivant

i	-2	-1	0	1	2
$P(X_1 = i)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

   et que Z est une variable aléatoire suivant la loi donnée par le tableau

i	-1	1
$P(Z = i)$	$\frac{1}{2}$	$\frac{1}{2}$

   .

   On suppose X_1 et Z indépendantes.
   i. Calculer $\mathbb{E}(Z)$.
   ii. On pose $X_2 = X_1 Z$. Calculer $\mathrm{Cov}(X_1, X_2)$.
   iii. Montrer que X_1 et X_2 ne sont pas indépendantes. Commenter.

Partie III

On note $X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ un vecteur aléatoire à valeurs dans $\mathbb{R}^3$ (ses composantes, les X_i, sont donc des variables aléatoires réelles, supposées discrètes). On suppose que

$$\begin{pmatrix} \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \\ \mathbb{E}(X_3) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Au vecteur X on associe sa matrice de covariance

$$\Sigma_X = \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \mathrm{Cov}(X_1, X_3) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) & \mathrm{Cov}(X_2, X_3) \\ \mathrm{Cov}(X_3, X_1) & \mathrm{Cov}(X_3, X_2) & \mathrm{Cov}(X_3, X_3) \end{pmatrix}.$$

1. On fixe un vecteur $u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ de $\mathbb{R}^3$. On note $Y = < u, X > = u_1 X_1 + u_2 X_2 + u_3 X_3$.
   (a) Calculer $\mathbb{E}(Y)$.

(b) Exprimer $\mathbb{V}(Y)$ en terme de covariance.

(c) Montrer que $\mathbb{V}(Y) = {}^t u \Sigma_X u$. Est-ce cohérent avec le résultat de la question 5 de la partie I ?

2. Dans la suite de l'exercice on suppose que $\Sigma_X = M$. On prend aussi $u = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$.

   On rappelle (voir la partie I) que $u \in \ker(M)$.
   On suppose aussi que les X_i sont à valeurs dans $\{0, 1, 2\}$.

   (a) Que vaut $\mathbb{V}(Y)$?

   (b) En déduire que $\mathbb{E}(Y^2) = 0$.

   (c) Exprimer Y en fonction des X_i, puis montrer que Y^2 est à valeurs dans $\{0, 1, 4\}$.

   (d) Déduire que $P(Y = 0) = P(Y^2 = 0) = 1$.

   (e) Conclure que le vecteur X prend ses valeurs dans l'hyperplan H avec une probabilité 1.