ENSAI École nationale de la statistique et de l'analyse de l'information

Insee Mesurer pour comprendre

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 2021

ÉPREUVE DE MATHÉMATIQUES ET STATISTIQUES

Durée 4 heures

Coefficient 3

Sans documents – L'usage de la calculatrice est interdit

Le sujet comprend 7 pages

Le sujet est composé de 4 exercices. Les candidats sont invités à le parcourir dans son intégralité. Le barème est donné à titre indicatif.

Exercice I .... 3 pts

Soit f l'application de $\mathbb{R}^3$ dans $\mathbb{R}^2$ définie par

$$f(x, y, z) = (-2x - 2z, 5x + 2y + z).$$

1. Montrer que f est une application linéaire.
2. Vérifier que $e_1 = (1, 0, 0) \notin \ker(f)$ et que $e_2 = (0, 1, 0) \notin \ker(f)$.
3. (a) Montrer que le noyau $\ker(f)$ de f est engendré par le vecteur $e_0 = (-1, 2, 1)$.
   (b) Quelle est la dimension de $\ker(f)$?
4. On note $S = Vect(e_1, e_2)$ le sous-espace vectoriel de $\mathbb{R}^3$ engendré par e_1 et e_2.
   (a) Donner une base de S.
   (b) A-t-on $\ker(f) \cap S = \{0\}$?
   (c) Montrer que tout vecteur u de $\mathbb{R}^3$ s'écrit de manière unique sous la forme $u = v + s$ avec $v \in \ker(f)$ et $s \in S$.

   Dans la suite de cet exercice, pour tout $u \in \mathbb{R}^3$, on note $p(u)$ le vecteur $s \in S$ unique de la décomposition $u = v + s$ donnée à la question 4. (c).
5. Soit $g = f|_S$ la restriction de f à S.
   (a) Montrer que $f = g \circ p$.
   (b) Montrer que l'application $p : \mathbb{R}^3 \to S$ est linéaire.

Exercice II .... 7 pts

Partie A

Cet exercice porte sur deux exemples de calcul d'estimateur par la méthode dite du maximum de vraisemblance (MV).

1. On suppose dans cette question que X_1 et X_2 sont indépendantes et suivent la loi normale $\mathcal{N}(m, 1)$ dont on note f_m la densité. Soit L_1 la fonction définie sur $\mathbb{R}$ par

$$L_1(m) = f_m(x_1)f_m(x_2)$$

où x_1 et x_2 sont des réels donnés.

(a) Montrer que

$$L_1(m) = \frac{1}{2\pi}\,\mathrm{e}^{-\frac{1}{2}\left((x_1-m)^2+(x_2-m)^2\right)} \quad .$$

(b) Montrer l'égalité

$$(x_1 - m)^2 + (x_2 - m)^2 = 2m^2 - 2(x_1 + x_2)m + x_1^2 + x_2^2$$

puis déterminer le minimum de cette expression en m.

(c) Déduire du (b) que L_1 admet un unique maximum, atteint pour $m = \dfrac{x_1 + x_2}{2}$.

(d) On définit un estimateur de m par $\widehat{m}_{\mathrm{MV}} = \dfrac{X_1 + X_2}{2}$.
Montrer que $\mathbb{E}(\widehat{m}_{\mathrm{MV}}) = m$ et déterminer $\mathbb{V}(\widehat{m}_{\mathrm{MV}})$.

2. On suppose dans cette question que $X_1, \cdots, X_n$ sont indépendantes et suivent la loi exponentielle de paramètre λ ($\lambda > 0$). On note g_λ la densité de cette loi.

Soit L_2 la fonction définie sur $]0; +\infty[$ par

$$L_2(\lambda) = \prod_{i=1}^{n} g_\lambda(x_i) = g_\lambda(x_1)g_\lambda(x_2)\cdots g_\lambda(x_n)$$

où les $x_1, \cdots, x_n$ sont des réels strictement positifs donnés.

(a) Calculer l'espérance et la variance des $X_1, \cdots, X_n$.

(b) Montrer que $L_2(\lambda) = \lambda^n\,\mathrm{e}^{-\lambda s_n}$ avec $s_n = \displaystyle\sum_{i=1}^{n} x_i$.

(c) On appelle log la fonction logarithme népérien. Justifier que

$$\log L_2(\lambda) = n \log \lambda - \lambda s_n \quad .$$

(d) Justifier que la dérivée de $\log L_2$, notée $(\log L_2)'$ est définie sur $]0; +\infty[$ par

$$(\log L_2)'(\lambda) = \frac{n - \lambda s_n}{\lambda} \quad .$$

(e) Déduire du (d) que L_2 admet un maximum en $\lambda = \dfrac{n}{s_n}$.

(f) Soit $\widehat{\gamma}_{\mathrm{MV}}$ l'estimateur de λ^{-1} défini par $\widehat{\gamma}_{\mathrm{MV}} = \dfrac{S_n}{n}$ avec $S_n = \sum_{i=1}^{n} X_i$.
Déterminer $\mathbb{E}\left(\widehat{\gamma}_{\mathrm{MV}}\right)$ et $\mathbb{V}\left(\widehat{\gamma}_{\mathrm{MV}}\right)$.

Partie B

Des clients passent régler leurs achats en caisse dans un magasin. On s'intéresse alors à la durée de passage en caisse, que l'on suppose suivre la même loi pour tous les clients.

1. Dans cette question, les durées de passage en caisse sont exprimées en minutes et suivent la loi exponentielle de paramètre λ ($\lambda > 0$). On note T une variable aléatoire suivant cette loi.

   (a) Quelle est la probabilité que T soit inférieure à 3 ?

   (b) Sachant que la durée de passage est supérieure à 3 minutes, montrer que la probabilité qu'elle soit supérieure à 4 minutes est égale à $\mathrm{e}^{-\lambda}$.

   (c) Déterminer la médiane de la loi de T, c'est-à-dire le réel t tel que

   $$\mathbb{P}(T \leq t) = 0,5 \quad .$$

   (d) Un logiciel enregistre les durées de passage en caisse de 100 clients, que l'on suppose indépendantes, et donne un temps total de passage de 300 minutes. Donner l'estimation de λ par la méthode du maximum de vraisemblance (Partie A).

2. (a) Soit A un réel positif. Justifier l'égalité

   $$\int_0^A x\,\mathrm{e}^{-2x}\,\mathrm{d}x = \frac{1-(1+2A)\,\mathrm{e}^{-2A}}{4} \quad .$$

   Indication : on pourra effectuer une intégration par parties.

   (b) En déduire que la limite de $\int_0^A x\,\mathrm{e}^{-2x}\,\mathrm{d}x$ quand A tend vers $+\infty$ existe et donner sa valeur C.

   On notera dans la suite $\displaystyle\int_0^{+\infty} x\,\mathrm{e}^{-2x}\,\mathrm{d}x$ l'intégrale dite *généralisée* et dont la valeur est égale à C calculée précédemment.

3. Soit h la fonction définie sur $\mathbb{R}$ par $h(x) = 4x\,\mathrm{e}^{-2x}$ si $x \geq 0$ et 0 sinon.
   Justifier que h est une densité de probabilité.

   On suppose à présent et jusqu'à la fin de cet exercice que les durées de passage en caisse des clients sont indépendantes et suivent la loi de probabilité dont la densité est h. On note $\widetilde{T}$ une variable aléatoire suivant cette loi.

4. On définit la fonction de survie de $\widetilde{T}$ notée $H_{\widetilde{T}}$ par

   $$H_{\widetilde{T}}(t) = 1 - \mathbb{P}(\widetilde{T} \leq t) \quad .$$

   (a) Montrer que $H_{\widetilde{T}}(t) = 1$ si $t < 0$ et $H_{\widetilde{T}}(t) = (1+2t)\,\mathrm{e}^{-2t}$ si $t \geq 0$.

   (b) Soit u et v deux réels tels que $u < v$. Montrer que

   $$\mathbb{P}(\widetilde{T} \leq v \mid \widetilde{T} > u) = 1 - \frac{H_{\widetilde{T}}(v)}{H_{\widetilde{T}}(u)} \quad .$$

5. Deux caisses du magasin se libèrent et trois clients s'y rendent simultanément. L'un d'entre d'eux, moins pressé, cède sa place. On note $\widetilde{T}_1$ et $\widetilde{T}_2$ les durées de passage en caisse des clients, que l'on suppose indépendantes, et M le temps d'attente du troisième client.

   (a) Soit $t > 0$. Justifier successivement les égalités

   $$\{M > t\} = \{\widetilde{T}_1 > t\} \cap \{\widetilde{T}_2 > t\}$$

   puis

   $$\mathbb{P}(M > t) = \left(H_{\widetilde{T}}(t)\right)^2 \quad .$$

   (b) On note F_M la fonction de répartition F_M de M, c'est-à-dire la fonction définie pour tout t réel par $F_M(t) = \mathbb{P}(M \leq t)$. Déduire du (a) l'égalité

   $$F_M(t) = 1 - (1+2t)^2\, \mathrm{e}^{-4t} \text{ si } t \geq 0 \text{ et } 0 \text{ sinon} \quad .$$

   (c) Calculer $F_M'(t)$ pour $t > 0$.

   (d) On admet dans cette question que M admet une densité. Soient a et b des réels tels que $0 < a < b$. Justifier que

   $$\mathbb{P}(a \leq M \leq b) = \int_a^b F_M'(u)\,\mathrm{d}u \quad .$$

   En déduire la définition d'une densité de M.

Exercice III .... 4 pts

Dans l'exercice si B est une matrice, ${}^tB$ désigne la matrice transposée de B.
On note $< u, v >= {}^tu.v = u_1v_1 + u_2v_2$ le produit scalaire standard de deux vecteurs u et v de $\mathbb{R}^2$.
On note $\|u\| = \sqrt{< u, u >}$ la norme euclidienne.

Partie A - Algèbre linéaire

Soit

$$M = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}.$$

1. (a) Montrer que $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ est un vecteur propre de la matrice M pour une valeur propre α_1 que l'on précisera.
   (b) On admet que M admet 4 comme valeur propre. Déterminer un vecteur propre v_2 de M pour cette valeur propre.
2. (a) Soit v un vecteur non nul de $\mathbb{R}^2$. On pose $w = \dfrac{1}{\|v\|}v$. Montrer que $\|w\| = 1$.
   (b) En déduire des vecteurs propres w_1 et w_2 de M de norme 1.
3. On note $P = \dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$.
   (a) Montrer que ${}^tPP = P\,{}^tP = I_2$. Qu'en déduisez-vous ?
   (b) Calculer $D = P^{-1}MP$. Que dire de la matrice M ?
4. (a) Si $u = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$, montrer que $\|Du\| = \sqrt{4x^2 + 16y^2}$.
   (b) En déduire que pour tout $u \in \mathbb{R}^2$, on a $\|Du\| \leqslant 4\|u\|$.
   (c) De manière similaire, montrer que si $u \in \mathbb{R}^2$, on a $\|Pu\| = \|u\|$.
   (d) En posant $v = P^{-1}u$, prouver que $\|P^{-1}u\| = \|u\|$.
   (e) En déduire que pour tout $u \in \mathbb{R}^2$, on a $\|Mu\| \leqslant 4\|u\|$.
5. Établir que pour $u \in \mathbb{R}^2$ le produit scalaire $< Du, u >$ est supérieur ou égal à $2\|u\|^2$.

Partie B - Optimisation

On définit la fonction f sur $\mathbb{R}^2$ par

$$f(x, y) = 1 - 2x - 2y + \frac{1}{2}\left(3x^2 + 3y^2 - 2xy\right).$$

1. (a) Pour $(x, y) \in \mathbb{R}^2$, calculer les dérivées partielles $\frac{\partial f}{\partial x}(x, y)$ et $\frac{\partial f}{\partial y}(x, y)$.
   (b) On note $\nabla_f(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x}(x, y) \\ \frac{\partial f}{\partial y}(x, y) \end{pmatrix}$ le gradient de f en (x, y).
   Montrer que $\nabla_f(x, y) = M\begin{pmatrix} x \\ y \end{pmatrix} + b$ pour un vecteur $b \in \mathbb{R}^2$ que l'on précisera, M étant la matrice de la partie A.
   (c) On note $H_f(x, y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x, y) & \frac{\partial^2 f}{\partial x \partial y}(x, y) \\ \frac{\partial^2 f}{\partial y \partial x}(x, y) & \frac{\partial^2 f}{\partial y^2}(x, y) \end{pmatrix}$ la Hessienne de f en (x, y).
   Calculer $H_f(x, y)$.
2. Montrer que l'équation $\nabla_f(u) = 0$ admet une unique solution sur $\mathbb{R}^2$ que l'on note u^*.
3. On suppose que $u \in \mathbb{R}^2$ est un extremum local de f. Justifier que $u = u^*$.
4. Prouver que u^* est un minimum local de f.

Exercice IV .... 6 pts

Partie A

Vous organisez une soirée lors de laquelle chacun viendra avec un cadeau. Afin de gagner du temps, vous choisissez de procéder par avance à un tirage au sort pour savoir à qui chaque personne donnera son cadeau. Le problème apparaît rapidement : en faisant comme cela, il est possible de récupérer son propre cadeau !

On appelle n le nombre total de personnes. En supposant le tirage au sort sans remise et équiprobable, X_n désigne la variable aléatoire égale au nombre de personnes ayant récupéré le cadeau qu'elles avaient apporté.

1. *Étude du cas* $n = 2$
   (a) Déterminer la loi de X_2.
   (b) Calculer $\mathbb{E}(X_2)$ et $\mathbb{V}(X_2)$.
2. *Étude du cas* $n = 3$
   (a) Montrer que le nombre total de cas pour la distribution des cadeaux est égal à 6.
   (b) Justifier les éléments de la loi de X_3 donnés dans le tableau ci-dessous.

k	0	1	2	3
$\mathbb{P}(X_3 = k)$	$\bullet$	$\frac{1}{2}$	0	$\frac{1}{6}$

   (c) Déduire du tableau donné dans le (b) la probabilité que personne ne récupère son propre cadeau.

Partie B

Soit n un entier naturel non nul. On pose $E_n = \{1, \cdots, n\}$ et on désigne par $\mathcal{P}_n$ l'ensemble des permutations sur E_n. Un dérangement est une permutation sans point fixe. On note $\mathcal{D}_n$ l'ensemble des dérangements de E_n et D_n le cardinal de $\mathcal{D}_n$ c'est-à-dire le nombre de dérangements de E_n. Par exemple, les dérangements de E_3 correspondent aux résultats 312 et 231. On a donc $D_3 = 2$. Par convention, on pose $D_0 = 1$ et $0! = 1$.

1. (a) Donner cinq éléments de $\mathcal{D}_4$.
   (b) On admet que $D_4 = 9$. Lister tous les éléments de $\mathcal{D}_4$.
2. Soit k un entier compris entre 0 et n. On appelle $\mathcal{I}_{k,n}$ l'ensemble des permutations de E_n laissant invariant k points. Par exemple, un élément de $\mathcal{I}_{1,4}$ est 3241.
   (a) Justifier les égalités $\operatorname{Card} \mathcal{I}_{0,n} = D_n$ et pour n supérieur ou égal à 1,
   $$\operatorname{Card} \mathcal{I}_{1,n} = nD_{n-1} \quad .$$
   (b) Pour k entier entre 0 et n, montrer que
   $$\operatorname{Card} \mathcal{I}_{k,n} = \binom{n}{k} D_{n-k} \quad .$$
   (c) Déduire du (b) l'égalité
   $$\sum_{k=0}^{n} \binom{n}{k} D_{n-k} = n! \quad .$$

(d) En justifiant l'égalité $\binom{n}{n-k} = \binom{n}{k}$, montrer que l'égalité donnée à la question précédente peut aussi s'écrire

$$\sum_{k=0}^{n} \binom{n}{k} D_k = n! \quad .$$

3. En appliquant la formule d'inversion de PASCAL donnée ci-dessous et que l'on admettra, justifier que

$$D_n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} \quad .$$

Formule d'inversion de PASCAL

Soit $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ deux suites de nombres réels tels que $u_n = \sum_{k=0}^{n} \binom{n}{k} v_k$. On a alors l'égalité

$$v_n = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} u_k \quad .$$

Partie C

On revient au problème des cadeaux abordé dans la partie A et on se place dans le cas général. Le nombre de personnes est à nouveau noté n et X_n désigne toujours le nombre de personnes ayant récupéré son propre cadeau après le tirage au sort.

On cherche dans cette partie à déterminer la loi de X_n, c'est-à-dire à déterminer les probabilités $\mathbb{P}(X_n = k)$ pour k entier compris entre 0 et n, en utilisant les résultats de la partie B.

1. Justifier que $\mathbb{P}(X_n = 0) = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$.
2. Montrer l'égalité

$$\mathbb{P}(X_n = k) = \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!} \quad .$$

pour tout k entier compris entre 0 et n.

3. Utiliser le 2. pour calculer $\mathbb{P}(X_4 = 2)$.