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 2022

ÉPREUVE DE MATHÉMATIQUES ET STATISTIQUES

Durée 4 heures

Coefficient 3

Sans documents – L'usage de la calculatrice est interdit

Le sujet comprend 9 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 1 .... 4 pts

Partie A

Soient α et λ des réels strictement positifs. On note f la fonction définie sur $\mathbb{R}$ par

$$f(x) = \alpha\lambda x^{\alpha-1}\,\mathrm{e}^{-\lambda x^{\alpha}} \quad \text{si} \quad x \geq 0 \quad \text{et} \quad 0 \quad \text{si} \quad x < 0 \quad .$$

1. On suppose dans cette question *uniquement* que $\alpha = 2$ et $\lambda = 1$. On se place de plus sur $I = [0; +\infty[$.
   (a) Calculer la dérivée de f sur I.
   (b) Étudier la limite de f en $+\infty$.
   (c) Dresser le tableau de variations de f sur I.
   (d) Déterminer le ou les sous-intervalles de I sur lesquels la fonction f est convexe.
2. Soit M un nombre réel strictement positif.
   (a) Prouver que $\int_0^M f(x)\,\mathrm{d}x = 1 - \mathrm{e}^{-\lambda M^{\alpha}}$.
   (b) En déduire que f est une densité de probabilité.

Partie B

Soit X une variable aléatoire de densité f définie dans la partie A . On dit que X suit la loi de Weibull de paramètres α et λ, loi que l'on note $\mathcal{W}(\alpha, \lambda)$.

Partie B.1 - Étude de X^{α}

Soit Y définie par $Y = X^{\alpha}$. On admet qu'ainsi définie, Y est une variable aléatoire dont on note F_Y la fonction de répartition.

1. Justifier l'égalité $\mathbb{P}(Y \leq y) = \mathbb{P}\left(X \leq y^{\frac{1}{\alpha}}\right)$ pour $y > 0$.
2. Montrer que $F_Y(y) = 1 - \mathrm{e}^{-\lambda y}$ pour $y > 0$.
   On pourra utiliser la fonction de répartition F_X de X.
3. Identifier la loi suivie par Y.
   On pourra admettre que la fonction de répartition d'une variable aléatoire réelle caractérise sa loi, c'est-à-dire que si $F_U = F_V$ avec U et V variables aléatoires réelles, alors U et V suivent la même loi.
4. Donner l'espérance et la variance de Y.

Partie B.2 - Estimation des paramètres α et λ

Dans la suite, on définit la fonction de survie S_X pour tout réel x par $S_X(x) = 1 - F_X(x)$. De plus, log désigne la fonction logarithme népérien et afin d'alléger les notations, on note plus simplement S la fonction S_X.

Soit $X_1, X_2, ..., X_N$ un échantillon de variables aléatoires indépendantes de même loi que X. On désigne par $S_N^{ech}(x)$ la variable aléatoire égale à la proportion des N variables $X_1, X_2, ..., X_N$ supérieurs strictement à x.

La fonction S_N^{ech} ainsi définie est aléatoire et converge vers S (résultat admis) au sens où pour tout x,

$$\mathbb{P}\left(\lim_{N\to\infty} S_N^{ech}(x) = S(x)\right) = 1 \quad .$$

1. Pour $x > 0$, justifier l'égalité

$$\log S(x) = -\lambda x^{\alpha}$$

2. En déduire que pour $x > 0$, il existe des réels A et B tels

$$\log(-\log S(x)) = A \log x + B \quad .$$

   On donnera les expressions de A et B en fonction de α et λ.

3. Soit $x_1, x_2, ..., x_N$ un échantillon de réalisations des X_i et s_N^{ech} la réalisation de S_N^{ech} correspondant à cet échantillon. Les figures 1 et 2 ci-dessous représentent respectivement les points de coordonnées $\left(x_i, s_N^{ech}(x_i)\right)$ et $\left(\log x_i, \log(-\log s_N^{ech}(x_i))\right)$ avec $N = 100$.

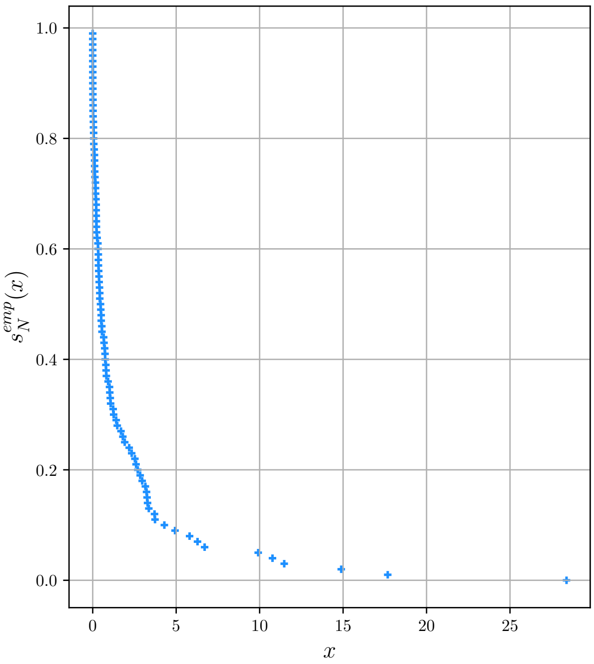

Figure 1

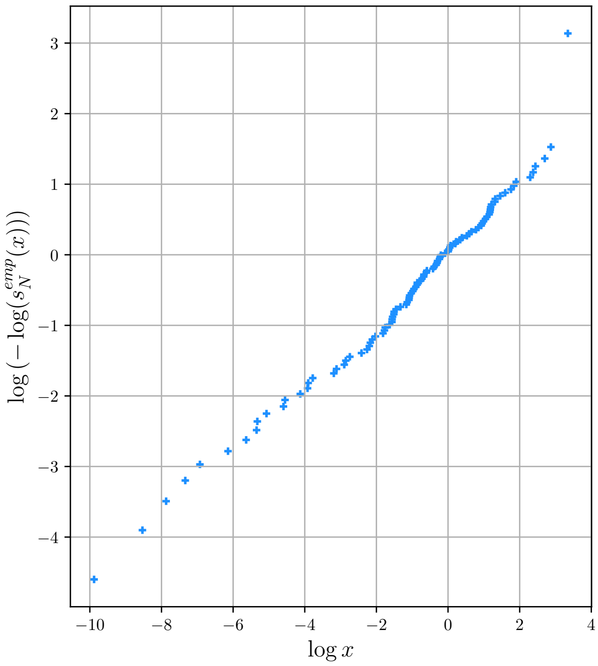

Figure 2

(a) Quelle caractéristique présente le nuage de points représenté par la figure 2 ? Expliquer ce résultat.

(b) En détaillant la méthode utilisée, déterminer graphiquement des estimations de A et B.

(c) En déduire des estimations de α et λ.

Exercice 2 .... 5 pts

Partie A - Étude de deux exemples en dimension 3

Soit E un espace vectoriel sur $\mathbb{R}$ de dimension finie $n \geq 2$. On note Id_E l'application identité de E. Pour tout endomorphisme f de E, on note $f^0 = Id_E$, pour k et m entiers naturels, $f^{k+1} = f \circ f^k$ et plus généralement $f^{m+k} = f^m \circ f^k$.

Les questions 1 et 2 sont indépendantes.

Dans cette partie, l'espace E est de dimension 3 et est rapporté à sa base canonique (e_1, e_2, e_3).

1. On considère l'endomorphisme f dont la matrice A est

$$A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

dans la base (e_1, e_2, e_3).

   (a) Exprimer $f(e_1), f(e_2), f(e_3)$ et $f^2(e_1)$ en fonction de e_1, e_2, e_3.
   En déduire que $(e_1, f(e_1), f^2(e_1))$ est une base de E.

   (b) Déterminer les valeurs propres de f.

   (c) Déterminer pour chacune des valeurs propres le sous-espace propre associé.
   En déduire une matrice P inversible telle que $P^{-1}AP$ soit une matrice diagonale que l'on explicitera.

2. On considère l'endomorphisme g dont la matrice B est

$$B = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}$$

dans la base (e_1, e_2, e_3).

   (a) La matrice B est-elle inversible ?

   (b) Déterminer les valeurs propres de g.

   (c) Déterminer pour chacune des valeurs propres le sous-espace propre associé.
   L'endomorphisme g est-il diagonalisable ?

Partie B - Étude en dimension n

Dans cette partie E un espace vectoriel sur $\mathbb{R}$ de dimension finie $n \geq 2$.
On dit qu'un endomorphisme f de E est **cyclique** s'il existe un vecteur x_0 de E tel que $E = \text{Vect}(f^k(x_0), k \in \mathbb{N})$
c'est à dire : si x est un élément de E, il existe un entier p et des réels $\alpha_0, ..., \alpha_p$ tels que $x = \alpha_0 x_0 + \alpha_1 f(x_0) + ... + \alpha_p f^p(x_0)$.

1. Soit f un endomorphisme cyclique et soit m le plus grand entier naturel tel que $(x_0, f(x_0), ..., f^{m-1}(x_0))$ soit libre et $(x_0, f(x_0), ..., f^m(x_0))$ soit liée
(*L'existence de m n'est pas à justifier*).

   (a) Montrer que $f^m(x_0)$ est combinaison linéaire de $x_0, f(x_0), ..., f^{m-1}(x_0)$.

(b) Montrer par récurrence sur k que pour tout entier naturel k, $f^{m+k}(x_0)$ est combinaison linéaire de $x_0, f(x_0), ..., f^{m-1}(x_0)$.

(c) Montrer que $(x_0, f(x_0), ..., f^{m-1}(x_0))$ est une famille génératrice de E.

(d) En déduire que $m = n$.

2. Dans cette question, on pose $f^n(x_0) = q_0 x_0 + q_1 f(x_0) + ... + q_{n-1} f^{n-1}(x_0)$.
Écrire la matrice M de f dans la base $(x_0, f(x_0), ..., f^{n-1}(x_0))$.

Exercice 3 .... 5 pts

Partie A

1. Soit X une variable aléatoire prenant des valeurs entières, éventuellement négatives. La loi de probabilité de X est symétrique ce qui signifie que pour tout i entier,
$$\mathbb{P}(X = i) = \mathbb{P}(X = -i) \quad .$$
   (a) On suppose que X peut valoir *uniquement* -1 ou 1 et on pose $\mathbb{P}(X = 1) = p$.
   - Prouver que la loi de X est symétrique pour une unique valeur de p que l'on précisera.
   - Représenter graphiquement la fonction de répartition F de X dans le cas symétrique

   (b) On suppose qu'il existe un entier naturel N tel que $\mathbb{P}(-N \leq X \leq N) = 1$. Montrer que l'espérance de X est nulle.

   Dans les questions 2. , 3. et 4., X désigne une variable aléatoire symétrique telle que $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 0,5$.
2. Calculer la variance de X.
3. Soit Y la variable aléatoire définie telle que $X = 2Y - 1$.
   (a) Déterminer la loi de probabilité de Y.
   (b) Comment s'appelle cette loi de probabilité ?
   (c) Donner la variance de Y puis en déduire une autre méthode qu'au 2. pour déterminer la variance de X.
4. Soit X_n une suite de variables aléatoires indépendantes de même loi que X (on rappelle que $p = 0,5$). Pour tout entier naturel n, on définit la variable Z_n par $Z_n = \prod_{k=0}^{n} X_k$ c'est-à-dire le produit des variables aléatoires $X_0, X_1, \ldots, X_n$. par exemple, $Z_2 = X_0 \times X_1 \times X_2$.
   (a) Déterminer l'image de Z_n c'est-à-dire l'ensemble des valeurs que peut prendre Z_n.
   (b) Montrer que Z_n est centrée.
   (c) Calculer la covariance $\text{Cov}(Z_n, Z_{n+1})$ de Z_n et Z_{n+1}.

Partie B

Dans cette partie, on se place de nouveau dans le cas général : X désigne une variable aléatoire symétrique telle que $\mathbb{P}(X = 1) = p$ et $\mathbb{P}(X = -1) = 1 - p$ avec p réel strictement compris entre 0 et 1. On reprend de plus la définition de Z_n introduite au 4. .

1. (a) Justifier l'égalité
$$\mathbb{P}(Z_{n+1} = 1) = \mathbb{P}(Z_{n+1} = 1 \mid Z_n = 1)\mathbb{P}(Z_n = 1) + \mathbb{P}(Z_{n+1} = 1 \mid Z_n = -1)\mathbb{P}(Z_n = -1)$$
   dans laquelle $\mathbb{P}(B \mid A)$ désigne la probabilité conditionnelle de l'événement B sachant l'événement A.
   (b) Montrer que $\mathbb{P}(Z_{n+1} = 1 \mid Z_n = 1) = \mathbb{P}(X = 1)$ et $\mathbb{P}(Z_{n+1} = 1 \mid Z_n = -1) = \mathbb{P}(X = -1)$.
2. On pose $u_n = \mathbb{P}(Z_n = 1)$.
   (a) Montrer que pour tout n, $u_{n+1} = au_n + b$ avec $a = 2p - 1$ et $b = 1 - p$.
   (b) Soit l le réel tel que $l = al + b$. Vérifier que la suite $(v_n)_{n\in\mathbb{N}}$ définie par $v_n = u_n - l$ est géométrique et convergente.
   (c) Déduire du b. que la limite de u_n est $0,5$.

Exercice 4 .... 6 pts

Partie A

$\mathbb{N}^*$ désigne l'ensemble des entiers naturels non nuls.
Pour tout réel x, $|x|$ désigne la valeur absolue de x.

1. On considère la fonction f définie sur $\left[0, \frac{\pi}{2}\right]$ par :

$$\begin{cases} f(x) & = & \dfrac{-x+\dfrac{1}{\pi}x^2}{\sin(x)} \quad \text{si} \quad x \in \left]0, \dfrac{\pi}{2}\right] \\ f(0) & = & -1 \end{cases} .$$

   (a) Donner le développement limité en 0 de $\sin(x)$ à l'ordre 3 .

   (b) Justifier que f est continue sur $\left]0, \frac{\pi}{2}\right]$.

   (c) Démontrer que f est continue en 0.

   (d) Démontrer que f est dérivable sur $\left[0, \frac{\pi}{2}\right]$.

   On désigne par f' la dérivée de f et on pose $f'(0) = \frac{1}{\pi}$.
   On admet pour la suite que f' est continue sur $\left[0, \frac{\pi}{2}\right]$ et que f' est bornée sur $\left[0, \frac{\pi}{2}\right]$, c'est à dire :
   il existe un réel A tel que pour tout réel x de $\left[0, \frac{\pi}{2}\right]$, $|f'(x)| \leqslant A$

2. On note pour tout entier naturel non nul n : $F(n) = \displaystyle\int_0^{\frac{\pi}{2}} f(x)\sin(nx)\mathrm{d}x$

   (a) Justifier l'existence de $F(n)$.

   (b) A l'aide d'une intégration par parties, démontrer qu'il existe un réel M indépendant de n tel que pour tout $n \in \mathbb{N}^*$, $|F(n)| \leq \frac{M}{n}$.
   En déduire la limite de $F(n)$ quand n tend vers $+\infty$.

Partie B

Dans cette partie, on considère pour tout entier naturel $n \geq 1$, $u_n = \sum_{k=1}^{n} \frac{1}{k^2}$.

1. On note pour tout $k \in \mathbb{N}^*$, $I_k = \int_0^{\frac{\pi}{2}} \left(-x + \frac{1}{\pi}x^2\right) \cos(2kx)\mathrm{d}x$.

   A l'aide de deux intégrations par parties successives, montrer que
   $$I_k = \frac{1}{4k^2}$$
   pour tout entier naturel non nul k.

2. On rappelle que pour tout réel x, $e^{ix} = \cos(x) + i\sin(x)$.

   On note pour tout nombre complexe z , $\mathcal{R}e(z)$ la partie réelle de z.

   Soit n un entier naturel non nul.
   Pour $x \in \left[0, \frac{\pi}{2}\right]$, on considère la somme $S_n(x) = \sum_{k=1}^{n} \mathrm{e}^{2ikx} = \mathrm{e}^{2ix} + ... + \mathrm{e}^{2inx}$.

   (a) Montrer que pour tout nombre complexe z et pour tout $n \in \mathbb{N}^*$ on a :
   $$(1 + z + z^2 + ... + z^{n-1})(1 - z) = 1 - z^n \quad .$$
   En déduire que pour tout $x \in \left]0, \frac{\pi}{2}\right]$, $S_n(x) = \mathrm{e}^{2ix}\left(\frac{1 - \mathrm{e}^{2inx}}{1 - \mathrm{e}^{2ix}}\right)$

   (b) En déduire que $S_n(x)$ vaut
   $$\begin{cases} S_n(x) &= \frac{i}{2}\left(\frac{\mathrm{e}^{ix} - \mathrm{e}^{i(2n+1)x}}{\sin(x)}\right) \quad \text{si} \quad x \in \left]0, \frac{\pi}{2}\right] \\ S_n(0) &= n \end{cases} \quad .$$
   Soit $r_n(x) = \sum_{k=1}^{n} \cos(2kx)$.

   (c) Justifier que $r_n(x) = \mathcal{R}e(S_n(x))$.

   (d) Déduire de la question précédente que
   $$\begin{cases} r_n(x) &= \frac{\sin\left((2n+1)x\right) - \sin(x)}{2\sin(x)} \quad \text{si} \quad x \in \left]0, \frac{\pi}{2}\right] \\ r_n(0) &= n \end{cases}$$
   (la fonction r_n ainsi définie est continue sur $\left[0, \frac{\pi}{2}\right]$.)

   (e) Démontrer que pour tout entier naturel n, $n \geq 1$:
   $$u_n = 4\int_0^{\frac{\pi}{2}} \left(-x + \frac{1}{\pi}x^2\right) r_n(x)\mathrm{d}x \quad .$$

   (f) Démontrer que pour tout entier naturel n, $n \geq 1$:
   $$u_n = 2F(2n+1) - 2\int_0^{\frac{\pi}{2}} \left(-x + \frac{1}{\pi}x^2\right) \mathrm{d}x$$
   où F est la fonction définie A-2.

(g) Déduire des résultats précédents que la suite $(u_n)_{n\in\mathbb{N}^*}$ est convergente et préciser sa limite.