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

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

AVRIL 2017

ÉPREUVE DE MATHÉMATIQUES ET STATISTIQUES

Durée 4 heures

Coefficient 3

Sans documents – L'usage de la calculatrice est interdit

Le sujet comprend 7 pages (y compris celle-ci)

Ce sujet se compose de 5 exercices. Le barème est donné à titre indicatif.

On pourra utiliser dans les exercices 1 et 5 la formule d'intégration par parties pour des fonctions continument dérivables f et g sur un intervalle de bornes a et b :

$$\int_a^b f'(t)g(t)dt = f(b)g(b) - f(a)g(a) - \int_a^b f(t)g'(t)dt.$$

Exercice 1. *(5 points)*

Partie A : Étude d'une fonction

On note f la fonction définie sur $]0;+\infty[$ par $f(x) = \dfrac{\ln(x)}{1+x^2}$.

1. (a) Justifier que f est dérivable sur son ensemble de définition.
   (b) Montrer que pour tout $x > 0$, $f'(x)$ est du signe de $g(x) = 1 + x^2 - 2x^2\ln(x)$.
   (c) Étudier les variations de g. On montrera en particulier que l'équation $g(x) = 0$ admet une et une seule solution sur $\mathbb{R}^{+*}$, cette solution sera notée m.
2. (a) Calculer les limites de f en 0 et $+\infty$ et montrer que $f(m) = \frac{1}{2m^2}$.
   (b) Dresser le tableau de variation de f.

Partie B : Étude d'une fonction intégrale

Dans cette partie la fonction F est définie par

$$\forall x > 0,\ F(x) = \int_1^x f(t)dt = \int_1^x \frac{\ln(t)}{1+t^2}dt$$

1. (a) Déterminer le signe de F sur $\mathbb{R}^{+*}$.
   (b) Justifier la continuité et la dérivabilité de F sur $\mathbb{R}^{+*}$.
   (c) Calculer $F'(x)$ pour $x > 0$.
2. Montrer que $\forall x > 0,\ F(x) = F\left(\frac{1}{x}\right)$.
3. (a) Soit φ la fonction définie sur $\mathbb{R}^{+*}$ par

$$\forall x > 0,\ \varphi(x) = \frac{\arctan(x)}{x}$$

   Montrer que φ est prolongeable par continuité en 0.

(b) Montrer avec une intégration par partie, que $\forall x > 0$, $F(x) = \arctan(x)\ln(x) - \int_1^x \varphi(t)dt$.

(c) En déduire que F est prolongeable par continuité en 0.
La nouvelle fonction ainsi obtenue sera encore notée F.
Que peut-on dire de F au voisinage de $+\infty$?

Partie C : Recherche d'une valeur approchée

Dans cette partie, on cherche à calculer une valeur approchée de $F(0)$.

1. Pour tout $k \in \mathbb{N}$ et $x > 0$, par intégration par partie, calculer $I_k(x) = \int_1^x t^k \ln(t)dt$.
Montrer que
$$\forall n \in \mathbb{N}, \forall x > 0,\ \frac{1}{1+x^2} = \sum_{k=0}^{n}(-1)^k x^{2k} + (-1)^{n+1}\frac{x^{2n+2}}{1+x^2}.$$
2. En déduire, pour $x \in \mathbb{N}$ et $x \in]0,1[$ la majoration suivante :
$$\left| F(x) - \sum_{k=0}^{n}(-1)^k I_{2k}(x) \right| \leq I_{2n+2}(x).$$
3. On pose pour $n \in \mathbb{N}$ $u_n = \sum_{k=0}^{n} \frac{(-1)^k}{(2k+1)^2}$, montrer que $|F(0) - u_n| \leq \frac{1}{(2n+3)^2}$.
4. Donner une méthode pour obtenir une valeur approchée de $F(0)$ à 10^{-2} près.

Exercice 2. *(4 points)*

Le but de ce problème est de déterminer toutes les matrices $X \in M_2(\mathbb{R})$ solutions de l'équation $X^2 = A$ avec $A = \begin{pmatrix} 5 & -4 \\ 2 & -1 \end{pmatrix}$.

On note f l'application linéaire de $\mathbb{R}^2$ dans $\mathbb{R}^2$ associée à la matrice A dans la base canonique. Id désignera l'application linéaire identité de $\mathbb{R}^2$ dans $\mathbb{R}^2$.

Partie A : Puissances de A

1. Montrer que $F_1 = \ker(f - Id)$ est un sous-espace propre de f de dimension 1. Donner un vecteur ε_1 non nul de F_1.
2. Montrer que 3 est une valeur propre de f et donner un vecteur ε_3 associé.
3. Donner, sans calculs, la matrice de l'application linéaire f dans la base $\mathcal{B}_\varepsilon = (\varepsilon_1, \varepsilon_3)$.
4. En déduire qu'il existe une matrice D, diagonale, et P, inversible tel que $A = PDP^{-1}$ (on précisera P et P^{-1}).
On rappelle que P ainsi définie est la matrice de l'application Id de $\mathbb{R}^2$ muni de la base canonique dans $\mathbb{R}^2$ muni de la base $(\varepsilon_1, \varepsilon_3)$.
5. Expliciter D^n et montrer que $A^n = PD^nP^{-1}$.

6. Montrer que $A^n = 3^n B + C$ avec $B \in M_2(\mathbb{R})$ et $C \in M_2(\mathbb{R})$ deux matrices que l'on explicitera.

Partie B : Résolution de l'équation

1. Soit $M \in M_2(\mathbb{R})$ tel que $M^2 = D$.
   Montrer que $MD = DM$ et en déduire que M est diagonale.
   Expliciter les différentes solutions que l'on notera M_1, M_2 ...
2. Soit $X \in M_2(\mathbb{R})$. En étudiant $M = P^{-1}XP$, déterminer une écriture des matrices X solutions de l'équation $X^2 = A$ (on demande d'exprimer ces solutions à l'aide des matrices M_i, P et P^{-1}).
3. On note $X_1, \dots X_m$ les solutions de l'équation $X^2 = A$.
   Sans calculer explicitement les m solutions, déterminer leur somme $S = X_1 + \dots + X_m$ et leur produit $\Pi = X_1 X_2 \dots X_m$.

Exercice 3. *(4 points)*

Dans une ville, deux voyageurs, Albert et Béatrice, savent que les bus passent à intervalle de temps régulier. Ils souhaiteraient connaître cet intervalle de temps, mais ils n'observent que le temps d'attente du bus. Chacun propose sa méthode d'estimation.

Le premier voyageur, Albert, pense qu'en moyenne, il arrive à l'arrêt de bus à mi-temps entre deux bus. Il calcule la moyenne de ses observations et la double.

Comme Béatrice trouve que parfois elle attend vraiment longtemps, elle propose de prendre le temps maximal qu'elle a pu observer.

On note X_i le $i^{\text{ème}}$ temps d'attente à l'arrêt de bus. On suppose que X_i suit une loi uniforme sur $[0; a]$ où a est un réel strictement positif, a est donc le paramètre que les voyageurs souhaiteraient connaître.

Numéro de l'observation	1	2	3	4	5	6	7	8	9	10
temps d'attente en min	7	9	7	2	8	3	6	0	4	5
Numéro de l'observation	11	12	13	14	15	16	17	18	19	20
temps d'attente en min	3	6	8	11	9	5	8	2	7	10

Ils disposent d'un jeu de n observations indépendantes X_1, ... X_n identiquement distribuées selon la loi uniforme. On rappelle que la densité de cette loi est $f_a(x) = \dfrac{1}{a}$ pour $x \in [0, a]$ et $f_a(x) = 0$ pour $x \notin [0, a]$.

On rappelle que le biais d'un estimateur $\hat{\theta}$ cherchant à estimer la valeur du paramètre θ est : $\text{Biais}(\hat{\theta}) = E(\hat{\theta}) - \theta$.

1. On note $A_n = \frac{2}{n}(X_1 + X_2 + \dots + X_n)$ l'estimateur d'Albert.
   (a) Calculer l'espérance et la variance de X_i.
   (b) Calculer l'espérance et la variance de l'estimateur A_n. Est-ce un estimateur sans biais de a (c'est-à-dire que le biais est nul) ?

2. Soit $B_n = \max(X_1, X_2, \dots, X_n)$ l'estimateur de Béatrice.
   (a) Calculer la fonction de répartition de la variable aléatoire X_i. et en déduire la probabilité $P(B_n \leq t)$ pour $t > 0$.
   (b) Calculer la dérivée, notée f_{B_n}, de la fonction de répartition de B_n sur $]0, a[$.
   On admet que la fonction f_{B_n} calculée précédemment sur $]0, a[$ et prolongée par 0 en dehors de $]0, a[$ est la densité de B_n sur $\mathbb{R}$.
   (c) Calculer le biais de B_n. Pourriez-vous proposer à Béatrice l'expression d'un autre estimateur B_n^* sans biais ?
   (d) Calculer la variance de B_n^*.
3. Quel estimateur conseillez-vous d'utiliser ? Justifier votre choix.
4. On donne $\sum_{i=1}^{20} X_i = 120$ et $\sum_{i=1}^{20} X_i^2 = 886$.
   Donner une estimation de la valeur de a ainsi qu'une estimation de la variance pour chacun des deux estimateurs.
   Les valeurs ainsi calculées confirment-elles les arguments théoriques avancés précédement ?

Exercice 4. *(4 points)*

Les deux parties peuvent être traitées indépendamment l'une de l'autre. Dans la seconde partie, on pourra utiliser les résultats de la première.

Partie A : Étude d'un ensemble de suites

Dans cette partie on cherche à étudier l'ensemble $\mathcal{S}$ des suites réelles définies par la relation de récurrence (R) et les conditions initiales $(u_0, u_1) \in \mathbb{R}^2$.

$$(R) : u_{n+2} = pu_{n+1} + (1-p)u_n \quad \text{où} \quad p \in]0,1[\text{ est paramètre fixé.}$$

On notera $U = (u_n)_{n\in\mathbb{N}}$ une suite et λU, $\lambda \in \mathbb{R}$ la suite $(\lambda u_n)_{n\in\mathbb{N}}$.

1. (a) Démontrer que $\mathcal{S}$ est un sous-espace vectoriel de l'espace vectoriel des suites réelles.
   (b) Soit U (respectivement V) la suite de $\mathcal{S}$ avec les conditions initiales $u_0 = 1$ et $u_1 = 0$ (resp. $v_0 = 0$ et $v_1 = 1$).
   Montrer que (U, V) forme un système libre.
   (c) Montrer que toute suite W de $\mathcal{S}$ peut s'écrire comme combinaison linéaire de U et V.
   Quelle est la dimension de $\mathcal{S}$?
2. (a) On s'intéresse dans cette question aux suites de $\mathcal{S}$, s'exprimant sous la forme $u_n = Q^n$.
   Montrer que Q satisfait l'équation $Q^2 = pQ + (1-p)$.
   Déterminer Q_0 et Q_1 les deux racines de cette dernière équation.
   (b) Montrer que les éléments de $\mathcal{S}$ s'écrivent $u_n = \lambda + \mu(p-1)^n$ avec $(\lambda, \mu) \in \mathbb{R}^2$.

Partie B : Nombre de cartouches d'encre

La maintenance d'une imprimante intervient tous les jours de la semaine pour remplacer les cartouches d'encre. On impose que chaque cartouche soit remplacée au moins une fois tous les deux jours. Ainsi, par exemple, si le lundi la cartouche n'est pas remplacée, elle le sera nécessairement le mardi.

On note T_i le nombre de jours d'utilisation de la $i^{\text{ème}}$ cartouche. On suppose que les T_i sont des variables aléatoires indépendantes et identiquement distribuées telles que $P(T_i = 1) = p$ et $P(T_i = 2) = 1 - p$ avec $p \in]0, 1[$.

Soit N le nombre de jour de fonctionnement de l'imprimante. Pour $n \in [\![1, N]\!]$, R_n désigne la variable aléatoire prenant la valeur 1 si une cartouche est remplacée le jour n et 0 sinon. On note $r_n = E(R_n)$ l'espérance de R_n.

1. (a) Calculer $E(T_i)$.

   (b) Calculer r_1 et r_2.

2. Soit n un entier strictement positif.

   (a) À l'aide de la formule des probabilités totales, écrire une relation donnant $P(R_{n+2} = 1)$ en fonction de $P(R_{n+1} = 1)$ et $P(R_n = 1)$.

   (b) En déduire l'égalité $r_{n+2} = pr_{n+1} + (1 - p)r_n$.

3. (a) Établir une expression de r_n en fonction de n et p.

   (b) En déduire que $\lim\limits_{n \to +\infty} r_n = \dfrac{1}{E(T_i)}$.

4. Soit $N \in \mathbb{N}^*$. On note C_N le nombre de cartouches qui ont été remplacées après N jours de fonctionnement. Exprimer C_N en fonction des variables aléatoires R_n. Calculer l'espérance $E(C_N)$.

Exercice 5. *(3 points)*

L'objectif de ce problème est de proposer une méthode pour calculer :

$$m = \min_{(a,b) \in \mathbb{R}^2} \int_{-1}^{1} (e^t - at - b)^2 dt.$$

On note $E = \mathcal{C}([-1, 1], \mathbb{R})$ l'ensemble des fonctions continues de $[-1, 1]$ dans $\mathbb{R}$ et φ l'application de $E \times E$ dans $\mathbb{R}$ définie par :

$$\varphi(f, g) = \int_{-1}^{1} f(t)g(t)dt$$

$\mathbb{R}_n[X]$ est l'ensemble des polynômes réels de degré inférieur ou égal à n. On note $p_0(x) = 1$ et $p_1(x) = x$. On note $exp(x) = e^x$ la fonction exponentielle définie sur $[-1; 1]$.

1. Montrer que φ est un produit scalaire sur E.

   On notera $\|f\| = \sqrt{\varphi(f, f)}$ la norme associée.

2. Montrer que (p_0, p_1) forme une base φ-orthogonale de $\mathbb{R}_1[X]$.

3. Calculer a et b tel que $\varphi(exp - ap_1 - bp_0, p_0) = 0$ et $\varphi(exp - ap_1 - bp_0, p_1) = 0$ (on pourra utiliser une intégration par partie).
   On note $p(exp) = ap_1 + bp_0$ et $u = exp - p(exp)$ avec a et b les solutions déterminées précédemment.
4. Que dire de u et $p(exp)$ par rapport à $\mathbb{R}_1[X]$?
5. Montrer que $\forall p \in \mathbb{R}_1[X]$, $\|u - p\|^2 = \|u\|^2 + \|p\|^2$.
6. Déterminer m. Justifier votre réponse.