

Chambre de Commerce et d'Industrie de Paris

DIRECTION DE L'ENSEIGNEMENT

SERVICE DES CONCOURS ET EXAMENS

Ecole des Hautes Etudes Commerciales**CONCOURS D'ADMISSION DE 1978****Composition de Mathématiques****2^e EPREUVE***Mercredi 24 Mai 1978, de 8 heures à midi*

Dans tout le problème, P est un plan affine réel ; $\vec{P}$ est le plan vectoriel associé ; on note $(A, B, M, \dots)$ des points de P , $(\vec{AB}, \dots)$ des vecteurs de $\vec{P}$. Dans la troisième partie, $\| \cdot \|$ est une norme euclidienne sur P .

PREMIÈRE PARTIE

1^o Si $(\vec{AB}, \vec{AC})$ est une base de $\vec{P}$, montrer que $(\vec{BC}, \vec{BA})$ et $(\vec{CA}, \vec{CB})$ sont des bases de $\vec{P}$.

2^o Soit A' (resp. B' , C') le milieu de BC (resp. CA , AB). Montrer que $(\vec{A'B'}, \vec{A'C'})$ est une base de $\vec{P}$. Quelle est la matrice de changement de base de $(\vec{AB}, \vec{AC})$ à $(\vec{A'B'}, \vec{A'C'})$?

3^o Soit $(\alpha, \beta, \gamma) \in \mathbb{R}^3$, $\alpha + \beta + \gamma = 1$, et M le barycentre de $(A(\alpha), B(\beta), C(\gamma))$.

Calculer $(\alpha', \beta', \gamma') \in \mathbb{R}^3$ tel que $\alpha' + \beta' + \gamma' = 1$ et que M soit barycentre de $(A'(\alpha'), B'(\beta'), C'(\gamma'))$. (On exprimera α' en fonction de α seul.)

4^o On considère les suites de points (A_n, B_n, C_n) , de nombres $(\alpha_n, \beta_n, \gamma_n)$, définis par $A_0 = A$ (resp. $B_0 = B$, $C_0 = C$), $A_1 = A'$ (resp. $B_1 = B'$, $C_1 = C'$), A_{n+1} (resp. B_{n+1} , C_{n+1}) étant le milieu de $B_n C_n$ (resp. $C_n A_n$, $A_n B_n$), $\alpha_n + \beta_n + \gamma_n = 1$ et telles que M soit barycentre de $(A_n(\alpha_n), B_n(\beta_n), C_n(\gamma_n))$. Déterminer une application f telle que, pour tout $n \in \mathbb{N}$:

$$\alpha_{n+1} = f(\alpha_n), \quad \beta_{n+1} = f(\beta_n), \quad \gamma_{n+1} = f(\gamma_n).$$

Déterminer l'ensemble des points M tels que $(\alpha_n, \beta_n, \gamma_n)$ soit indépendant de n ; interprétation géométrique. En déduire $(\alpha_n, \beta_n, \gamma_n)$ en fonction de n et de (α, β, γ) : on exprimera, d'une part α_n en fonction de n et de α seul, puis α_n sous la forme $\alpha_n = \alpha u_n + (\beta + \gamma) v_n$, où (u_n) et (v_n) sont deux suites ne dépendant que de n .

5^o Retrouver les résultats précédents en démontrant l'existence des suites (u_n) et (v_n) , en donnant des relations permettant de calculer (u_{n+1}, v_{n+1}) en fonction de (u_n, v_n) , puis en calculant, par la méthode de diagonalisation, la matrice $\begin{bmatrix} -1 & 2 \\ 1 & 0 \end{bmatrix}^n$. (Tous les calculs doivent figurer sur la copie.)

6^o a) Retrouver les résultats précédents en calculant H^n en fonction de (u_n, v_n) , où $H = \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$:

explicitant d'abord le lien entre $(\alpha_n, \beta_n, \gamma_n)$, (α, β, γ) et H^n , on déterminera l'équation caractéristique, les valeurs propres et les vecteurs propres d'un endomorphisme de $\mathbb{R}^3$ de matrice H dans la base canonique, puis on exhibera une matrice inversible P et une matrice diagonale D telles que $D = P^{-1} H P$; on montrera alors comment en déduire H^n . (Ce dernier calcul n'est pas demandé.)

b) Calculer H^n d'une autre manière en posant $E = H + 2I$, en calculant E^n pour $n \geq 1$, puis en explicitant H^n comme combinaison linéaire de E et de I .

.../...

DEUXIÈME PARTIE

1° Déterminer les coordonnées (x, y) du vecteur $\overrightarrow{AM}$ dans la base $(\overrightarrow{AB}, \overrightarrow{AC})$ en fonction de (α, β, γ) .

2° Pour tout $i \in \{1, 2, 3\}$, on se donne $(\alpha_i, \beta_i, \gamma_i) \in \mathbb{R}^3$ avec $\alpha_i + \beta_i + \gamma_i \neq 0$; si M_i est le barycentre de $(A(\alpha_i), B(\beta_i), C(\gamma_i))$, montrer que les points (M_1, M_2, M_3) appartiennent à une même droite si, et seulement si :

$$\begin{vmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{vmatrix} = 0.$$

3° Dessiner une figure formée de dix points, $A, B, C, S, Q, R, I, J, K, M$, deux à deux distincts de P , tels que les ensembles suivants $\{A, R, B\}$, $\{A, Q, C\}$, $\{A, M, S\}$, $\{R, M, J, C\}$, $\{Q, M, I, B\}$, $\{K, B, S, C\}$, $\{R, I, S\}$, $\{Q, J, S\}$, $\{K, R, Q\}$ soient inclus chacun dans une droite différente.

4° M étant barycentre de $(A(\alpha), B(\beta), C(\gamma))$, montrer que chacun des neuf autres points est barycentre de (A, B, C) affectés de coefficients dépendants de (α, β, γ) que l'on déterminera de manière aussi simple que possible (sans que leur somme, non nulle, soit nécessairement égale à 1). On donnera les résultats sous la forme d'un tableau :

M	A	B	C	S	Q	R	I	J	K
α									
β									
γ									

5° Montrer que I, J, K sont alignés.

TROISIÈME PARTIE

1° On pose $a = \|\overrightarrow{BC}\|$, $b = \|\overrightarrow{CA}\|$, $c = \|\overrightarrow{AB}\|$. Donner une condition nécessaire et suffisante portant sur (α, β, γ) pour que M se projette orthogonalement en A' sur BC .

2° Retrouver ainsi, en déterminant explicitement (α, β, γ) en fonction de (a, b, c) , l'existence et l'unicité d'un point M commun aux médiatrices de BC , CA et AB .

3° Si M est, de nouveau, un point quelconque de P , calculer $\|\overrightarrow{MA}\|^2$ en fonction de $(a, b, c, \alpha, \beta, \gamma)$. Retrouver ainsi la condition du 1°.

4° Donner une condition nécessaire et suffisante portant sur (a, b, c) pour que $\overrightarrow{MA} + \overrightarrow{MB} + \overrightarrow{MC} = \vec{0}$, où M est le point déterminé au 2°. Interprétation géométrique.

REMARQUE : Les trois parties du problème sont totalement indépendantes. Les notations des 1°, 2° et 3° de la Première Partie sont utilisées dans tout le problème.

DIRECTION DE L'ENSEIGNEMENT
SERVICE DES CONCOURS ET EXAMENS

Ecole des Hautes Etudes Commerciales

CONCOURS D'ADMISSION DE 1979

Composition de Mathématiques

2^e EPREUVE

Mercredi 9 Mai 1979, de 8 heures à midi

PREMIÈRE PARTIE

1^o Calculer les inverses des matrices réelles

$$A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 2 & 2 \\ 2 & 2 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -1 & 2 \\ 1 & -2 & 2 \\ 2 & -2 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} -2 & 1 & 2 \\ -1 & 2 & 2 \\ -2 & 2 & 3 \end{bmatrix}.$$

2^o On note tM la transposée d'une matrice M . Montrer qu'il existe une matrice unique J telle que ${}^tJ = J$ et que, pour tous les nombres réels x, y, z ,

$$x^2 + y^2 - z^2 = \begin{vmatrix} x & y & z \end{vmatrix} J \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

3^o Montrer que l'ensemble des matrices réelles carrées d'ordre trois M telles que

$${}^tMJM = J$$

est un groupe multiplicatif, que l'on notera G .

4^o Montrer que A, B et C appartiennent à G .

DEUXIÈME PARTIE

1^o Montrer que si x, y, z sont trois entiers rationnels, les nombres

$$x^2 + y^2 - z^2 + 2(x+z)(y+z)$$

et $x^2 + y^2 - z^2 + 2(x-z)(y-z)$

sont des carrés d'entiers naturels.

2^o Montrer que les relations

$$z_1 - x_1 = y + z, \quad z_1 - y_1 = x + z, \quad x_1 + y_1 - z_1 = x + y + z$$

sont équivalentes à une relation matricielle

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

où H est une matrice constante que l'on déterminera.

.../...

-2-

3° On appelle triplet rectangulaire tout triplet (x, y, z) d'entiers naturels sans diviseur commun autre que 1 et tels que

$$x^2 + y^2 = z^2.$$

Soit (x, y, z) un triplet rectangulaire tel que $z > 1$. Posant

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = A^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

montrer que l'un des triplets (x', y', z') , $(-x', y', z')$ ou $(x', -y', z')$ est rectangulaire et que l'on a

$$1 \leq z' \leq z - 1.$$

4° Imaginer un algorithme transformant, en un nombre fini d'étapes, par des produits matriciels comme ci-dessus, un triplet rectangulaire quelconque, en $(0, 1, 1)$.

5° Donner dix triplets rectangulaires distincts.

TROISIÈME PARTIE

1° Soit E l'ensemble des suites réelles (a_n) telles que, pour tout n :

$$a_{n+3} - 7a_{n+2} + 7a_{n+1} - a_n = 0.$$

Montrer que E est un espace vectoriel de dimension 3 dont on déterminera une base formée d'applications de la forme

$$n \longmapsto r^n.$$

2° Déterminer (a_n) sachant que $a_0 = 1$, $a_1 = a_2 = 0$. Calculer alors a_3 et a_4 .

3° Déterminer (a_n) sachant que $a_0 = a_1 = 0$, $a_2 = 1$.

4° Déterminer (a_n) sachant que $a_0 = 0$, $a_1 = 1$, $a_2 = 0$.

QUATRIÈME PARTIE

On considère toujours la matrice A définie dans la première partie.

1° Montrer que A^3 est combinaison linéaire de I , A et A^2 .

2° Montrer qu'il existe trois suites (u_n) , (v_n) , (w_n) telles que, pour tout n :

$$A^n = u_n I + v_n A + w_n A^2,$$

et que ces suites appartiennent à l'espace E défini au 1° de la troisième partie.

3° En déduire une méthode de calcul de A^n .

4° Déterminer une injection $f: \mathbb{N} \rightarrow \mathbb{N}^3$ telle que, pour tout n , $f(n)$ soit un triplet rectangulaire.

Ecole des Hautes Etudes Commerciales

CONCOURS D'ADMISSION DE 1980

Composition de Mathématiques (4 h)

2^e EPREUVE

INTRODUCTION

On étudie ci-dessous la fonction y d'une variable réelle, définie par :

$$y(x) = \frac{x}{x-1} \sqrt{x^2 + 1}.$$

Dans la première partie on en construit la courbe représentative. Dans la deuxième partie on résout certains problèmes relatifs à son intégration.

PREMIÈRE PARTIE

1^o a) Etudier les variations de la fonction u d'une variable réelle définie par :

$$u(x) = x^3 - 2x^2 - 1.$$

b) Démontrer que l'équation

$$u(x) = 0$$

a une et une seule racine réelle x_0 , et que

$$4/3 < x_0.$$

c) Démontrer que l'on a plus précisément :

$$2 < x_0 < 3.$$

2^o On construit par récurrence les suites réelles (x_n) et (y_n) en posant :

pour $n = 1$, $x_1 = 2$, $y_1 = 3$, et

pour $n > 1$, $x_n = \frac{1}{2}(x_{n-1} + y_{n-1})$, $y_n = y_{n-1}$, si $u[\frac{1}{2}(x_{n-1} + y_{n-1})] \leq 0$,

$x_n = x_{n-1}$, $y_n = \frac{1}{2}(x_{n-1} + y_{n-1})$, si $u[\frac{1}{2}(x_{n-1} + y_{n-1})] > 0$.

a) Calculer x_2 et y_2 , x_3 et y_3 , x_4 et y_4 .

b) Montrer que l'on a, pour tout n ,

$$x_n \leq x_0 \leq y_n.$$

c) Démontrer que les suites (x_n) et (y_n) sont adjacentes.

d) Donner une valeur approchée de x_0 à 10^{-1} près.

3^o a) Quel est le domaine de définition de la fonction y citée dans l'introduction ?

b) Calculer la dérivée $y'(x)$, et l'exprimer en fonction de $u(x)$.

c) Dresser le tableau de variation de y .

.../...

- 4° a) Calculer le développement limité à l'ordre 2 de la fonction y au voisinage de 0.
 b) Quelle est la tangente au point (0, 0) de la courbe représentative de y ? Quelle est la position de la courbe par rapport à cette tangente au voisinage de (0, 0) ?
 5° Etudier les quatre branches infinies de la courbe représentative de y : existence d'asymptotes et position de la courbe par rapport à celles-ci.
 6° Tracer la courbe représentative de la fonction y .

DEUXIÈME PARTIE

- 1° On considère les fonctions rationnelles A et B définies par

$$A(t) = \frac{t^6 + t^4 - t^2 - 1}{t^3(t^2 - 2t - 1)} \quad \text{et} \quad B(t) = \frac{6t^4 + 2t^3 - t^2 - 1}{t^3(t^2 - 2t - 1)}.$$

- a) Déterminer la partie entière de A .
 b) Quels sont les pôles de A ? Quels sont leurs ordres de multiplicité ? Donner la forme de la décomposition de A en éléments simples et la forme de la décomposition de B en éléments simples (sans calculer les numérateurs des éléments simples).
 c) Sachant que la partie de la décomposition de B relative au pôle 0 est

$$\frac{1}{t^3} - \frac{2}{t^2} + \frac{6}{t},$$
 donner la décomposition complète de B en éléments simples. En déduire la décomposition de A en éléments simples.
 d) Déduire des calculs précédents une primitive C de A .

- 2° On veut déterminer pour $x > 1$ une primitive de la fonction y citée dans l'introduction.

- a) Soit I la primitive de y définie, pour $a > 1$ et $x > 1$, par

$$I(x) = \int_a^x \frac{z}{z-1} \sqrt{z^2 + 1} \, dz.$$

Effectuer sur $I(x)$ le changement de variable :

$$2z = t - \frac{1}{t} \quad \text{avec } t > 0.$$

- b) Exprimer I en fonction de C définie en 1° d) de la deuxième partie.

- 3° L'intégrale $\int_1^2 y(x) \, dx$ est-elle convergente ? Pourquoi ?

- 4° On considère l'aire $J(m)$ de la région du plan définie par les inégalités suivantes :

$$x_0 \leq x \leq m \quad \text{et} \quad x + 1 \leq y \leq y(x),$$

où m est un paramètre supérieur ou égal à la racine x_0 définie en 1° b) de la première partie.

- a) Représenter $J(m)$ par des hachures sur le graphique de la courbe représentative de y .
 b) L'aire $J(m)$ a-t-elle une limite finie lorsque m tend vers l'infini ?
 c) Calculer une valeur approchée de $J(4)$ en utilisant la valeur approchée de x_0 calculée en 2° d) de la première partie.

Chambre de Commerce et d'Industrie de Paris

DIRECTION DE L'ENSEIGNEMENT

SERVICE DES CONCOURS ET EXAMENS

Ecole des Hautes Etudes Commerciales

CONCOURS D'ADMISSION DE 1981

Composition de Mathématiques

2^e EPREUVE

Samedi 16 Mai 1981, de 8 heures à midi

PREMIÈRE PARTIE

A tout a réel, on associe f_a , fonction réelle d'une variable réelle définie par : $f_a(x) = \frac{x^2}{\sqrt{x-a}}$.

Soit $\mathcal{C}_a$ la courbe représentative de f_a .

- 1° Indiquer pour quelles valeurs de x , $f_a(x)$ est défini ?
- 2° Dresser, en fonction de a , les différents tableaux de variation de f_a .
- 3° Etudier les branches infinies de $\mathcal{C}_a$.
- 4° Rechercher les éventuels points d'inflexion de $\mathcal{C}_a$.
- 5° Trouver la courbe Γ décrite par le point à tangente horizontale de $\mathcal{C}_a$ lorsque a varie.
- 6° Construire sur un même graphique, avec des échelles appropriées (non nécessairement les mêmes sur les axes ox et oy), les courbes Γ , $\mathcal{C}_{-1}$, $\mathcal{C}_0$ et $\mathcal{C}_1$.

DEUXIÈME PARTIE

On considère la fonction g_a , fonction réelle d'une variable réelle définie par : $g_a(x) = \frac{x^3}{\sqrt{|x-a|}}$.

Soit $\mathcal{D}_a$ la courbe représentative de g_a .

- 1° Reprendre, pour la famille des courbes $\mathcal{D}_a$, les questions 1° 2° 3° 4° de la 1^{re} partie.
Construire sur un même graphique les courbes $\mathcal{D}_{-1}$, $\mathcal{D}_0$, $\mathcal{D}_1$.
- 2° Déterminer le nombre de points d'intersection de la courbe $\mathcal{D}_1$ et de la droite d'équation $y = \lambda$.
Donner une valeur approchée, à 10^{-1} près par défaut, des abscisses des points de $\mathcal{D}_1$ d'ordonnées $\lambda = 1$ et $\lambda = 4$ respectivement.

TROISIÈME PARTIE

1° Démontrer que, pour tout $x > a$, l'intégrale $\int_a^x f_a(t) dt$ converge. Démontrer que l'intégrale $\int_a^{+\infty} f_a(t) dt$ diverge.

2° Déterminer deux réels α et β tels que

$$\lim_{x \rightarrow +\infty} \left[\frac{x^2}{\sqrt{x-a}} - \sqrt{x} (\alpha x + \beta) \right] = 0.$$

Quelle est la nature de l'intégrale $\int_A^{+\infty} \left[\frac{x^2}{\sqrt{x-a}} - \sqrt{x} (\alpha x + \beta) \right] dx$, où A est un réel strictement supérieur à a et à 0?

3° Déterminer un réel γ tel que l'intégrale $\int_A^{+\infty} \left[\frac{x^2}{\sqrt{x-a}} - \sqrt{x} (\alpha x + \beta + \frac{\gamma}{x}) \right] dx$ soit convergente.

4° Calculer $I_n = \int_a^n \frac{x^2}{\sqrt{x-a}} dx$, pour n entier strictement supérieur au paramètre a .

5° Soit $S_n = \sum_{k=N+1}^n \frac{k^2}{\sqrt{k-a}}$ où N est un entier strictement plus grand que $\left| \frac{4a}{3} \right|$. Montrer que $\lim_{n \rightarrow +\infty} S_n = +\infty$. Etablir la double inégalité :

$$I_{n+1} - I_{N+1} \geq S_n \geq I_n - I_N.$$

En déduire un équivalent de S_n quand n tend vers $+\infty$.

6° On pose $T_n = \sum_{k=1}^n \frac{1}{\sqrt{k}}$, et $u_k = \frac{1}{\sqrt{k}} - 2(\sqrt{k+1} - \sqrt{k})$.

Montrer que la série de terme général u_k est convergente.

Calculer $\sum_{k=1}^n u_k$ et en déduire que T_n est équivalent à $2\sqrt{n}$ quand n tend vers $+\infty$.

7° On pose $V_n = \sum_{k=1}^n k\sqrt{k}$ et $W_n = \sum_{k=1}^n \sqrt{k}$.

En utilisant la définition des intégrales $\int_0^1 x\sqrt{x} dx$ et $\int_0^1 \sqrt{x} dx$ par les sommes de Riemann, trouver

$$\lim_{n \rightarrow +\infty} \frac{V_n}{n^{5/2}} \quad \text{et} \quad \lim_{n \rightarrow +\infty} \frac{W_n}{n^{3/2}}.$$

8° Retrouver, à l'aide des questions 3° 6° 7° de la 3^{me} partie, un équivalent simple de S_n quand n tend vers $+\infty$.

Ecole des Hautes Etudes Commerciales

CONCOURS D'ADMISSION DE 1982

Mathématiques II

TOUTES OPTIONS

Samedi 15 Mai 1982, de 8 heures à midi

(La présentation, l'écriture et l'orthographe ont leur part dans la note.)

Dans tout le problème, on désigne par λ un nombre réel strictement positif. Le symbole $\ln$ représente le logarithme népérien.

PREMIÈRE PARTIE

1°) Soit φ la fonction numérique définie sur $]0, +\infty[$ par la relation

$$\varphi(x) = \ln x - \frac{1}{x^2}.$$

a) Etudier la variation de φ . Construire la courbe représentative de φ .

b) Montrer que l'équation $\varphi(x) = 0$ admet une solution et une seule. Soit α cette solution.

c) Montrer que $1 < \alpha < e$.

d) Déterminer l'entier naturel k tel que $10^{-3}k \leq \alpha < 10^{-3}(k+1)$.

2°) Soit f_λ la fonction numérique définie sur $]-\lambda, +\infty[$ par la relation

$$f_\lambda(x) = \frac{1}{\lambda + x} - \lambda \ln(\lambda + x).$$

a) Montrer que l'équation $f_\lambda(x) = 0$ admet une solution et une seule. Soit x_λ cette solution.

b) Montrer que $1 < \lambda + x_\lambda$.

c) Montrer que $\lambda + x_\lambda$ admet une limite lorsque λ tend vers $+\infty$, et calculer cette limite.

DEUXIÈME PARTIE

Soit g_λ la fonction numérique définie sur $]-\lambda, +\infty[$ par la relation

$$g_\lambda(x) = e^{-\lambda x} \ln(x + \lambda).$$

1°) a) Calculer la dérivée g'_λ de g_λ . A l'aide de la première partie, déterminer le signe de $g'_\lambda(x)$.

b) Dresser le tableau de variation de g_λ . On pose $y_\lambda = g_\lambda(x_\lambda)$.

c) Calculer $g'_\lambda(0)$. A l'aide de la première partie, déterminer le signe de x_λ suivant la position de λ par rapport à α .

... / ...

d) En calculant $g_\lambda(-\frac{\lambda}{2})$, trouver une minoration de y_λ .

e) Montrer que x_λ et y_λ admettent des limites (finies ou infinies) lorsque λ tend vers $+\infty$.

Déterminer ces limites.

2°) Dans chacun des cas $\lambda = 1$ et $\lambda = 5$, déterminer l'élément p_λ de $\mathbb{Z}$

tel que $\frac{p_\lambda}{100} \leq x_\lambda < \frac{p_\lambda + 1}{100}$, et calculer $g_\lambda\left(\frac{p_\lambda + \frac{1}{2}}{100}\right)$.

3°) a) Construire sur un même graphique les courbes représentatives de g_1 et g_5 . On prendra la même unité sur les deux axes et on ne dessinera que les points des courbes dont les coordonnées sont comprises entre -5 et 5 .

b) Montrer que ces courbes ont un point commun P et un seul. (On pourra considérer la fonction h définie sur $[0, +\infty[$ par la relation

$$h(x) = e^{4x} \ln(x+1) - \ln(x+5)$$

et étudier la variation de h .)

c) Déterminer des valeurs approchées des coordonnées de P à $0,1$ près.

TROISIÈME PARTIE

Pour tout entier naturel non nul n , on considère l'intégrale

$$I_n = \int_0^{+\infty} e^{-nx} \ln(n+x) dx.$$

1°) a) Montrer que l'intégrale I_n converge.

b) Montrer que

$$I_n = \frac{\ln n}{n} + \int_0^{+\infty} \frac{e^{-nx}}{n(n+x)} dx.$$

En déduire que

$$\frac{\ln n}{n} \leq I_n \leq \frac{\ln n}{n} + \frac{1}{n^3}.$$

c) Montrer que

$$I_n \geq \frac{\ln n}{n} + \int_0^{n/10} \frac{e^{-nx}}{n(n+x)} dx \geq \frac{\ln n}{n} + \frac{10}{11} \frac{1}{n^3} (1 - e^{-n^2/10}).$$

2°) En déduire une valeur approchée I'_5 de I_5 à $0,001$ près.

FIN

Ecole des Hautes Etudes Commerciales 1983

CONCOURS D'ADMISSION DE 1983

Mathématiques II

TOUTES OPTIONS

ATTENTION : La présentation, l'écriture et l'orthographe ont leur part dans la note.

Pour tout entier naturel non nul n , on désigne par f_n , g_n et h_n les fonctions numériques définies sur $\mathbb{R}$ par les relations :

$$f_n(x) = \frac{1}{(1+x+x^2)^n}, \quad g_n(x) = \frac{x}{(1+x+x^2)^n}, \quad h_n(x) = \frac{x^2}{(1+x+x^2)^n}.$$

I. 1. Etudier la variation des fonctions f_1 , g_1 et h_1 . Dresser les tableaux de variation de ces fonctions. Construire leurs courbes représentatives dans un même repère orthonormal.

2. Déterminer les nombres de points d'inflexion de ces courbes (ce qui revient à déterminer les nombres de points où la dérivée seconde s'annule en changeant de signe). Calculer les coordonnées des points d'inflexion et les pentes des tangentes en ces points à 10^{-2} près.

II. On pose :

$$I_n = \int_0^1 f_n(t) dt, \quad J_n = \int_0^1 g_n(t) dt, \quad K_n = \int_0^1 h_n(t) dt.$$

1. Montrer que, pour tout élément x de $[0,1]$,

$$1-x \leq \frac{1}{1+x+x^2} \leq \frac{1}{1+x}.$$

2. En déduire que, pour tout entier naturel n tel que $n \geq 2$,

$$\frac{1}{n+1} \leq I_n \leq \frac{1}{n-1}.$$

Trouver un équivalent simple de I_n . (On dit que les suites (u_n) et (v_n) de nombres réels strictement positifs sont équivalentes lorsque le rapport $\frac{u_n}{v_n}$ tend vers 1 si n tend vers $+\infty$).

3. Pour tout entier naturel n tel que $n \geq 3$, calculer les intégrales

$$\int_0^1 t(1-t)^n dt \quad \text{et} \quad \int_0^1 \frac{t}{(1+t)^n} dt.$$

Trouver un équivalent simple de J_n .

4. Pour tout entier naturel n tel que $n \geq 4$, calculer les intégrales

$$\int_0^1 t^2(1-t)^n dt \quad \text{et} \quad \int_0^1 \frac{t^2}{(1+t)^n} dt.$$

Trouver un équivalent simple de K_n .

.../...

DIRECTION DE L'ENSEIGNEMENT

SERVICE DES CONCOURS ET EXAMENS

ÉCOLE DES HAUTES ÉTUDES COMMERCIALES

CONCOURS D'ADMISSION DE 1984

Mathématiques II

TOUTES OPTIONS

Samedi 12 mai 1984, de 14 heures à 18 heures

ATTENTION! La présentation, l'écriture et l'orthographe ont leur part dans la note.

Soit a un élément de $\mathbb{R}_+$.1. On désigne par E_a l'ensemble des suites réelles $u = (u_n)_{n \in \mathbb{N}}$ satisfaisant à la relation de récurrence :

$$(1) \quad 4u_{n+3} = 4(1+a)u_{n+2} - (1+4a)u_{n+1} + u_n \quad \text{où } n \in \mathbb{N}.$$

1. a) Montrer que E_a est un sous-espace vectoriel de l'espace vectoriel des suites réelles.b) En considérant l'application de E_a dans $\mathbb{R}^3$ qui à toute suite u associe (u_0, u_1, u_2) , calculer la dimension de E_a .2. a) Vérifier que l'espace vectoriel K des suites constantes est inclus dans E_a .b) Soit u un élément de E_a . On considère la suite $v = (v_n)_{n \in \mathbb{N}}$ définie par la relation :

$$v_n = u_{n+1} - u_n \quad \text{où } n \in \mathbb{N}.$$

Établir une relation de récurrence (2) satisfaite par v , reliant v_{n+2} , v_{n+1} et v_n .c) On désigne par F_a l'ensemble des suites réelles satisfaisant à la relation (2). Montrer que F_a est un sous-espace vectoriel de E_a .Déterminer une base de F_a . On sera amené à distinguer trois cas : $0 < a < 1$, $a = 1$, $a > 1$.Dans le premier cas, on posera $a = \cos \theta$, avec $0 < \theta < \frac{\pi}{2}$; dans le dernier cas, on posera $a = \cosh \theta$, avec $\theta > 0$.d) Montrer qu'il existe une valeur a_0 de a et une seule, que l'on calculera, pour laquelle K est inclus dans F_a .e) On suppose que $a \neq a_0$. Montrer que E_a est somme directe de K et de F_a . En déduire, dans chacun des trois cas envisagés au c), une base de E_a .f) Montrer que E_{a_0} contient la suite de terme général $u_n = n$. En déduire une base de E_{a_0} .3. Soit u l'élément de E_a déterminé par les conditions initiales :

$$u_0 = 1 - \sqrt{|a^2 - 1|} \quad u_1 = 1 \quad u_2 = 1 + \frac{1}{4}\sqrt{|a^2 - 1|}.$$

Calculer u_n en fonction de n . (On discutera suivant la valeur de a .)Étudier la convergence de u , et calculer la limite de cette suite, lorsqu'elle existe.II. On désigne par $M_3(\mathbb{C})$ l'ensemble des matrices carrées d'ordre 3 à coefficients complexes et par I_3 la matrice unité de $M_3(\mathbb{C})$.On donne la matrice carrée suivante, considérée comme élément de $M_3(\mathbb{C})$:

$$M = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{4} & -a - \frac{1}{4} & a+1 \end{pmatrix}$$

1. Calculer le polynôme caractéristique de M :

$$\delta_M(X) = \det(XI_3 - M).$$

Calculer les valeurs propres, réelles ou complexes, de la matrice M . Lorsque $a \neq 1$, on exprimera ces valeurs propres à l'aide du nombre θ introduit dans la partie I.2. Déterminer les valeurs de a pour lesquelles la matrice M est diagonalisable sur le corps $\mathbb{C}$. Lorsque M est diagonalisable, trouver une matrice carrée inversible P telle que la matrice $D = P^{-1}MP$ soit diagonale. Expliciter D .3. Pour tout polynôme $Q = \sum_{k=0}^n a_k X^k$ on pose $Q(M) = \sum_{k=0}^n a_k M^k$ avec la convention $M^0 = I_3$.On admettra que, pour tout couple (Q_1, Q_2) de polynômes, $(Q_1 Q_2)(M) = Q_1(M) Q_2(M)$; on admettra aussi que $\delta_M(M) = 0$.Soit n un nombre entier naturel. Montrer qu'il existe un triplet $(\alpha_n, \beta_n, \gamma_n)$ de nombres réels et un seul tel que :

$$M^n = \alpha_n M^2 + \beta_n M + \gamma_n I_3.$$

Expliciter ce triplet dans chacun des cas $a = \frac{5}{4}$ et $a = 1$.

Dans ces deux cas, retrouver le résultat de la question I 3.

DIRECTION DE L'ENSEIGNEMENT

SERVICE DES CONCOURS ET EXAMENS

ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE SUPÉRIEURE DE COMMERCE DE PARIS
ÉCOLE EUROPÉENNE DES AFFAIRES
ÉCOLE SUPÉRIEURE DE COMMERCE DE LYON

CONCOURS D'ADMISSION DE 1985

Mathématiques II

TOUTES OPTIONS

Mardi 7 Mai 1985, de 8 h à 12 h

*La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
L'usage des instruments de calcul est autorisé.*

Pour tout couple (q, r) de nombres entiers naturels non nuls tel que $q \leq r$, on note :

$$\frac{1}{q} + \frac{1}{q+1} + \dots + \frac{1}{r}$$

la somme des inverses des nombres entiers naturels consécutifs de q à r (cette somme se réduisant à $1/q$ si $q = r$).

L. Soient $u = (u_n)_{n \geq 1}$ et $v = (v_n)_{n \geq 1}$ les suites définies par les relations :

$$u_n = 1 + \frac{1}{2} + \dots + \frac{1}{n} - \ln n$$

$$v_n = 1 + \frac{1}{2} + \dots + \frac{1}{n} - \ln(n+1).$$

1. Étudier le signe des fonctions φ et ψ définies sur l'intervalle $] -1, +\infty[$ par les relations :

$$\varphi(x) = \ln(1+x) - x \quad \text{et} \quad \psi(x) = \ln(1+x) - \frac{x}{1+x}.$$

2. Montrer que la suite u est décroissante et que la suite v est croissante.

3. Montrer que les suites u et v convergent vers une même limite a (qu'on ne demande pas de calculer).

4. a) Déterminer un nombre entier n_0 tel que la relation $n \geq n_0$ implique :

$$|a - v_{n-1}| \leq 10^{-2}.$$

b) Calculer une valeur approchée de a à la précision 10^{-1} .

II. Une entreprise doit recruter au plus un employé à choisir parmi n candidats, où $n \geq 5$. On suppose que ces candidats peuvent être classés, sans *ex aequo*, selon leur valeur. Ils sont entendus, l'un après l'autre, par le directeur, lequel doit décider immédiatement, à la fin de chaque entretien, d'engager ou de refuser le candidat (qu'il ne peut donc comparer qu'aux candidats préalablement entendus et refusés).

L'ordre de passage des candidats est supposé aléatoire, les différents ordres de passage possibles étant pris équiprobables.

Le problème a pour objet l'étude de la procédure de choix suivante: on commence par se donner un *seuil*, c'est-à-dire un nombre entier s tel que $2 \leq s \leq n$. On entend d'abord les $s - 1$ premiers candidats qu'on refuse systématiquement, et qui constituent un échantillon servant à fixer le niveau de qualité du recrutement: on engage le premier des candidats à se présenter ensuite qui se révèle meilleur que les $s - 1$ candidats de l'échantillon. (Au cas où il ne s'en présente pas, personne n'est engagé.)

On note $\theta_n(s)$ la probabilité pour que soit engagé de cette façon le meilleur des n candidats.

1. On note E_k , où $1 \leq k \leq n$, l'événement: «le meilleur des n candidats est le k -ième à se présenter», et $F_{k,s}$, où $s \leq k \leq n$, l'événement: «le meilleur des $k - 1$ premiers candidats est parmi les $s - 1$ premiers à se présenter».

a) Calculer les probabilités $P(E_k)$ et $P(F_{k,s})$ des événements E_k et $F_{k,s}$, où $s \leq k \leq n$.

b) Calculer $P(E_k \cap F_{k,s})$ et en déduire que:

$$\theta_n(s) = \frac{s-1}{n} \left(\frac{1}{s-1} + \frac{1}{s} + \dots + \frac{1}{n-1} \right).$$

c) À l'aide des résultats de la partie I, s étant fixé, trouver la limite de $\theta_n(s)$ lorsque n tend vers $+\infty$.

2. a) Calculer la probabilité pour qu'on ne recrute aucun candidat.

b) Pour tout nombre entier j tel que $s \leq j \leq n$, calculer la probabilité pour que le j -ième candidat qui se présente soit recruté.

c) En déduire la loi de la variable aléatoire X égale au nombre des candidats qui auront été entendus à la fin de la procédure, et montrer que X a pour espérance:

$$E(X) = (s-1) \left(\frac{1}{s-1} + \frac{1}{s} + \dots + \frac{1}{n-1} + 1 \right).$$

III. 1. a) Écrire l'expression de $\theta_n(s+1) - \theta_n(s)$, où $s \leq n-1$, à l'aide de la formule

II. 1. b) et en déduire qu'il existe un nombre entier naturel s_n et un seul appartenant à l'intervalle $[2, n-1]$ et tel que, pour tout nombre entier naturel s appartenant à $[2, n]$, $\theta_n(s_n) \geq \theta_n(s)$.

On admettra à cet effet que, pour tout couple (q, r) de nombres entiers naturels vérifiant $1 < q \leq r$,

$$\frac{1}{q} + \frac{1}{q+1} + \dots + \frac{1}{r} \neq 1.$$

On montrera que le nombre s_n est caractérisé par la conjonction des relations:

$$\frac{1}{s_n} + \frac{1}{s_n+1} + \dots + \frac{1}{n-1} < 1$$

$$\frac{1}{s_n-1} + \frac{1}{s_n} + \dots + \frac{1}{n-1} > 1.$$

b) Calculer s_5, s_6, s_7 et s_8 .

c) Montrer que, pour tout nombre entier naturel n tel que $n \geq 5$, soit $s_{n+1} = s_n$ soit $s_{n+1} = 1 + s_n$.

2. a) Montrer que, pour tout nombre entier q strictement supérieur à 1,

$$\frac{1}{q+1} + \frac{1}{q+2} + \dots + \frac{1}{2q} > \frac{1}{2}$$

et

$$\frac{1}{q+1} + \frac{1}{q+2} + \dots + \frac{1}{4q} > 1.$$

b) En déduire que:

$$s_n - 1 > \frac{n-1}{4}.$$

3. a) Montrer que:

$$\lim_{n \rightarrow +\infty} \left(\frac{1}{s_n - 1} + \frac{1}{s_n} + \dots + \frac{1}{n-1} \right) = 1.$$

En déduire la limite de $\frac{n}{s_n - 1} \theta_n(s_n)$ lorsque n tend vers $+\infty$.

b) À l'aide de la partie I, montrer que:

$$\lim_{n \rightarrow +\infty} \left(\frac{n}{s_n - 1} \theta_n(s_n) - \ln \frac{n}{s_n - 1} \right) = 0.$$

c) En déduire la limite de $\frac{n}{s_n}$ lorsque n tend vers $+\infty$.

En conclure que:

$$\lim_{n \rightarrow +\infty} \theta_n(s_n) = \frac{1}{e}.$$

4. *Application numérique.* On suppose que $n = 300$. À l'aide des résultats précédents, déterminer rapidement une valeur approchée (on ne demande pas d'évaluer l'erreur commise):

- Du nombre s qu'il faut choisir pour avoir la probabilité la plus grande de recruter le meilleur des 300 candidats;
- De cette probabilité;
- De l'espérance de la variable aléatoire X définie dans la question II.2 c).

5. On pose $\alpha_n = \frac{\theta_{n+1}(s_{n+1})}{\theta_n(s_n)}$.

a) On suppose que $s_{n+1} = s_n$. Montrer que:

$$\alpha_n < 1.$$

b) Soit f la fonction définie sur l'intervalle $[0, 1]$ par la relation:

$$f(x) = \frac{s_n}{n+1} \cdot \frac{n}{s_n-1} \cdot \frac{x + \frac{1}{n}}{x + \frac{1}{s_n-1}}.$$

Montrer que la fonction f est croissante. En déduire que le résultat de la question a) reste valable dans le cas où $s_{n+1} = 1 + s_n$.

6. À l'aide des résultats de la question III.3, montrer que, pour n suffisamment grand, l'égalité $s_{n+1} = 1 + s_n$ implique $s_{n+2} = s_n \circ 1$.

DIRECTION DE L'ENSEIGNEMENT

DIRECTION DES ADMISSIONS ET CONCOURS

ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
 ÉCOLE SUPÉRIEURE DE COMMERCE DE PARIS
 ÉCOLE EUROPÉENNE DES AFFAIRES
 ÉCOLE SUPÉRIEURE DE COMMERCE DE LYON

CONCOURS D'ADMISSION DE 1986

Mathématiques II 4 heures

OPTIONS : Générale, Économique et Technologique

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
L'usage des instruments de calcul est autorisé.

PROBLÈME

L'objet du problème est de décrire et comparer deux méthodes de détection de pannes. On se place dans la situation suivante : on considère un ordinateur comprenant un ensemble C de n circuits intégrés, où $n \geq 2$, et on suppose qu'une panne a endommagé un circuit et un seul. On note $c_1, c_2, \dots, c_n$ ces circuits, où c_n est le circuit défectueux.

Dans la partie I, on étudie une méthode de tests par paquets quelconques de circuits. Dans la partie II, on évalue l'espérance du nombre de tests nécessaires pour détecter la panne par cette méthode et on compare celle-ci avec une méthode de tests où les circuits sont pris un à un.

I. Tests par tirages de paquets de circuits, avec remise

On note Ω l'ensemble des parties de C . On tire au hasard et de façon équiprobable, par un procédé adéquat, des parties de C , c'est-à-dire des éléments de Ω (y compris la partie vide). Pour tout élément G de Ω , on note $\bar{G}$ le complémentaire de G dans C .

1. Soit A une partie de C de cardinal r , où $0 \leq r \leq n$.

- a) Déterminer le nombre de parties de C ne rencontrant pas A .
- b) Déterminer le nombre de parties de C contenant A .

2. On tire un élément G de Ω .

- a) Calculer la probabilité d'obtenir une partie donnée B de C .
- b) Soit A une partie de C de cardinal r , où $0 \leq r \leq n$. Calculer la probabilité pour que G contienne A .
- c) Soit c_i un élément de C distinct de c_n . On suppose que c_n appartient à G ; calculer la probabilité conditionnelle pour que c_i appartienne à G .

3. Soit Ω_n l'ensemble de parties de C contenant l'élément c_n . Soit h un nombre entier naturel non nul. On tire, successivement et avec remise, h éléments $B_1, B_2, \dots, B_h$ de Ω . Une technique permet de tester un ensemble de circuits et de savoir si le circuit défectueux se trouve parmi eux. Pour chaque entier i appartenant à l'intervalle $[1, h]$, on teste le paquet B_i . On pose $G_i = B_i$ si B_i contient le circuit défectueux; dans le cas contraire, on pose $G_i = \bar{B}_i$. Ainsi, G_i est un élément de Ω_n . On désigne enfin par D_h l'intersection des parties G_i .

a) Montrer que, pour tout nombre entier naturel j tel que $1 \leq j \leq n-1$, la probabilité pour que c_j appartienne à D_k est égale à $\frac{1}{2^k}$.

b) Plus généralement, soit E une partie de C de cardinal r ne contenant pas l'élément c_n . Calculer la probabilité pour que D_k contienne E .

c) En déduire que les événements $c_j \in D_k$, où $1 \leq j \leq n-1$, sont mutuellement indépendants et qu'il en est de même pour les événements $c_j \notin D_k$.

d) Prouver enfin que :

$$P(D_k = \{c_n\}) = \left(1 - \frac{1}{2^k}\right)^{n-1}.$$

II. Étude du nombre de tests nécessaires pour détecter la panne

1. Pour tout nombre entier naturel k , on pose :

$$a_k = 1 - \left(1 - \frac{1}{2^k}\right)^{n-1}.$$

a) Montrer que l'application qui à tout nombre entier naturel non nul k associe $a_{k-1} - a_k$ définit une loi de probabilité sur $\mathbb{N}^*$.

b) Déterminer la fonction de répartition associée à cette loi; construire sa courbe représentative lorsque $n=3$.

2. On conserve la procédure de la question 1.3, mais le nombre de tirages n est pas fixé; pour tout nombre entier naturel non nul k , on note D_k l'intersection des parties $G_1, G_2, \dots, G_k$.

On considère une variable aléatoire X_n , à valeurs dans $\mathbb{N}^*$, qui prend la valeur 1 si l'événement $D_1 = \{c_n\}$ est réalisé, la valeur k , où $k \geq 2$, si l'événement $D_k = \{c_n\}$ est réalisé et si l'événement $D_{k-1} = \{c_n\}$ ne l'est pas.

(Une telle variable aléatoire représente donc le nombre de tests nécessaires pour détecter le circuit défectueux.)

a) Pour tout nombre entier naturel non nul k , calculer la probabilité de l'événement $X_n \leq k$.

b) En déduire que la loi de probabilité de X_n est celle qui a été définie dans la question 1. a).

3. a) Montrer que, pour tout nombre entier naturel non nul s fixé, la série de terme général $\left(\frac{1}{2^k}\right)$ est convergente.

En déduire que la série de terme général (a_k) est convergente.

b) Montrer que la variable aléatoire X_n admet une espérance et que :

$$E(X_n) = \sum_{k=0}^{+\infty} a_k.$$

À cet effet, on pourra calculer :

$$\lim_{q \rightarrow +\infty} \left(\sum_{k=1}^q k(a_{k-1} - a_k) - \sum_{k=0}^{q-1} a_k \right).$$

4. On se propose d'évaluer $E(X_n)$ en comparant la série de terme général (a_k) à une intégrale. À cet effet, pour tout nombre entier naturel non nul p , on pose :

$$I_p(x) = \int_0^x f_p(t) dt, \quad \text{où } f_p(t) = 1 - (1 - e^{-t})^p.$$

a) Montrer que l'intégrale $I_p = \int_0^{+\infty} f_p(t) dt$ est convergente.

(On pourra développer $(1 - e^{-t})^p$.)

b) Calculer une primitive de la fonction $f_{p+1} - f_p$ sur l'intervalle $[0, +\infty[$. En déduire la valeur de $I_{p+1} - I_p$. Montrer finalement que :

$$I_{n-1} = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n-1}.$$

c) Montrer que, pour tout nombre entier naturel $m \geq 2$:

$$\int_m^{m+1} \frac{dt}{t} \leq \frac{1}{m} \leq \int_{m-1}^m \frac{dt}{t}.$$

En déduire que :

$$\ln n \leq I_{n-1} \leq 1 + \ln(n-1).$$

Calculer $\lim_{n \rightarrow +\infty} \frac{I_{n-1}}{\ln n}$.

d) Soit g_n la fonction numérique définie sur $[0, +\infty[$ par la relation :

$$g_n(u) = 1 - \left(1 - \frac{1}{2^u}\right)^{n-1}.$$

En étudiant la variation de g_n , montrer que, pour tout nombre entier naturel non nul k :

$$a_k \leq \int_{k-1}^k g_n(u) du \leq a_{k-1}.$$

e) Montrer que, pour tout nombre entier naturel $q \geq 2$:

$$\sum_{k=1}^q a_k \leq \int_0^q g_n(u) du \leq \sum_{k=0}^{q-1} a_k.$$

En effectuant le changement de variable $t = u \ln 2$ dans cette intégrale et en passant à la limite dans l'encadrement précédent lorsque q tend vers $+\infty$, montrer que :

$$E(X_n) - 1 \leq \frac{I_{n-1}}{\ln 2} \leq E(X_n).$$

f) Déduire des résultats des questions c) et e) un encadrement de $E(X_n)$ et déterminer la limite de $\frac{E(X_n)}{\ln n}$.

5. On teste maintenant les circuits c_j un par un, en les tirant de manière équiprobable et sans remise. On désigne par Y_n le nombre de tests nécessaires pour détecter le circuit défectueux.

a) Déterminer la loi de la variable aléatoire Y_n .

b) Calculer l'espérance de Y_n .

6. a) Calculer la limite de $\frac{E(X_n)}{E(Y_n)}$ lorsque n tend vers $+\infty$.

b) Grâce à l'encadrement obtenu dans la question 4. f), comparer $E(X_n)$ et $E(Y_n)$ lorsque $n=100$, puis lorsque $n=1000$.

DIRECTION DE L'ENSEIGNEMENT

DIRECTION DES ADMISSIONS ET CONCOURS

CONCOURS D'ADMISSION DE 1987

Mathématiques II

OPTIONS : Générale, Économique et Technologique

Lundi 4 Mai 1987, de 14 heures à 18 heures

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

L'usage des instruments de calcul est autorisé.

L'étude simultanée de deux variables statistiques X et Y a permis d'obtenir les trois observations suivantes :

$$(x_1 = 0, y_1 = 0), (x_2 = 1, y_2 = 1), (x_3 = -2, y_3 = 0).$$

Dans un plan rapporté à un repère orthonormal, on représente ces trois observations par le nuage de points : $M_1(0, 0)$, $M_2(1, 1)$, $M_3(-2, 0)$.

PARTIE I

L'objet de cette partie est d'étudier par différentes méthodes l'ajustement du nuage (M_1, M_2, M_3) par une droite.

On désigne par δ une direction donnée du plan et par D une droite non parallèle à δ d'équation $y = ax + b$.

On projette les points M_1, M_2, M_3 sur D dans la direction δ . On note m_1, m_2, m_3 les points obtenus (pour $1 \leq i \leq 3$, m_i est donc l'intersection de la droite D avec la droite de direction δ passant par M_i).

1) Dans cette question, la direction δ est celle de l'axe des ordonnées Oy . On cherche la droite D rendant minimale l'expression :

$$f(a, b) = M_1m_1 + M_2m_2 + M_3m_3.$$

a) Calculer les distances M_1m_1, M_2m_2, M_3m_3 .

b) Le nombre réel b est fixé. En distinguant trois cas suivant la position de b par rapport à $\frac{2}{3}$, représenter graphiquement la fonction φ définie par la relation :

$$\varphi(x) = |2x - b| + |x + b - 1|.$$

Montrer qu'elle passe par un minimum pour $x = \frac{b}{2}$.

c) En déduire l'existence et l'unicité d'un couple (a_1, b_1) conduisant à la plus petite valeur possible pour $f(a, b)$. Tracer la droite D_1 d'équation $y = a_1x + b_1$.

2) Dans cette question, la direction δ est encore celle de l'axe des ordonnées. On cherche la droite D rendant minimale l'expression :

$$g(a, b) = \sup (M_1m_1, M_2m_2, M_3m_3),$$

où $\sup (M_1m_1, M_2m_2, M_3m_3)$ désigne le plus grand des trois nombres réels M_1m_1, M_2m_2, M_3m_3 .

a) Représenter graphiquement sur une même figure :

— l'ensemble E_1 des points $M(x, y)$ du plan tels que $|y| \leq \frac{1}{3}$;

— l'ensemble E_2 des points $M(x, y)$ du plan tels que $|y - 2x| \leq \frac{1}{3}$;

— l'ensemble E_3 des points $M(x, y)$ du plan tels que $|x + y - 1| \leq \frac{1}{3}$.

En déduire l'ensemble $E_1 \cap E_2 \cap E_3$.

b) En déduire l'existence et l'unicité d'un couple (a_2, b_2) conduisant à la plus petite valeur possible pour $g(a, b)$. Tracer la droite D_2 d'équation $y = a_2x + b_2$.

3) Dans cette question, la direction δ est toujours celle de l'axe des ordonnées. On cherche la droite D rendant minimale l'expression :

$$h(a, b) = (M_1m_1)^2 + (M_2m_2)^2 + (M_3m_3)^2.$$

Le nombre réel a étant fixé, montrer que la fonction $b \mapsto h(a, b)$ admet un minimum en un point unique que l'on précisera.

En déduire l'existence et l'unicité d'un couple (a_3, b_3) conduisant à la plus petite valeur possible pour $h(a, b)$. Tracer la droite D_3 d'équation $y = a_3x + b_3$. Quelle est la droite D_3 ainsi obtenue ?

4) Dans cette question, m est un nombre réel donné et l'on suppose $a \neq m$. La direction δ est celle de la droite d'équation $y = mx$. On cherche la droite D rendant minimale l'expression :

$$h_m(a, b) = (M_1m_1)^2 + (M_2m_2)^2 + (M_3m_3)^2.$$

a) Calculer les coordonnées de m_1, m_2, m_3 . Exprimer $h_m(a, b)$ en fonction de a, b et m .

b) Le nombre m est fixé et différent de $\frac{2}{7}$.

Le nombre réel a étant donné, pour quelle valeur de b la fonction $b \mapsto h_m(a, b)$ est-elle minimale ?

En déduire que $h_m(a, b)$ prend la plus petite valeur possible lorsque la fonction θ définie par :

$$\theta(a) = \frac{7a^2 - 4a + 1}{(a - m)^2}$$

est minimale. Étudier la variation de la fonction θ .

Tournez la page S. V. P.

En déduire l'existence et l'unicité d'un couple (a_4, b_4) conduisant à la plus petite valeur possible pour $h_m(a, b)$. (On explicitera a_4 et b_4 en fonction de m .)

c) Étudier l'existence d'une telle droite D_4 dans le cas où $m = \frac{2}{7}$.

PARTIE II

Dans toute la suite, $M_0(x_0, y_0)$ désigne un point du plan associé à une observation supplémentaire des variables statistiques X et Y , et l'on étudie les modifications apportées par ce nouveau point au coefficient de corrélation et à la droite de régression de Y par rapport à X du nuage (M_1, M_2, M_3) .

1) Soient ρ et $\rho(M_0)$ les coefficients de corrélation respectivement associés aux nuages (M_1, M_2, M_3) et (M_0, M_1, M_2, M_3) .

a) Calculer ρ en précisant les formules utilisées.

b) Calculer $\rho(M_0)$ en fonction de x_0 et y_0 .

c) Pour tout nombre réel k , on considère les points E_k et F_k de coordonnées respectives $(5+k, k)$ et $(5+k, -k)$. Soient α et β les fonctions qui associent respectivement au nombre réel k les nombres $\rho(E_k)$ et $\rho(F_k)$.

Montrer que α et β sont continues sur $\mathbb{R}$. Calculer $\alpha(0)$ et $\beta(0)$. Déterminer les limites lorsque k tend vers $+\infty$ de $\alpha(k)$ et de $\beta(k)$.

En déduire que, pour tout nombre réel r appartenant à l'intervalle $] -1, 1[$, il existe au moins un point M_0 tel que $\rho(M_0) = r$.

Existe-t-il un point M_0 tel que $\rho(M_0) = 1$, ou tel que $\rho(M_0) = -1$?

2) Soit D une droite d'équation $y = ax + b$.

a) On suppose que $3b - a - 1 \neq 0$. Montrer qu'il existe un point M_0 et un seul tel que la droite de régression Δ du nuage (M_0, M_1, M_2, M_3) de Y par rapport à X soit D .

b) On suppose que $3b - a - 1 = 0$ et que $(a, b) \neq \left(\frac{2}{7}, \frac{3}{7}\right)$. Montrer qu'il n'existe aucun point M_0 tel que $\Delta = D$.

c) On suppose que $(a, b) = \left(\frac{2}{7}, \frac{3}{7}\right)$. Déterminer l'ensemble des points M_0 tels que $\Delta = D$. Retrouver ce résultat sans calcul.

• • •
•

Chambre de Commerce et d'Industrie de Paris

CONCOURS D'ADMISSION DE 1988

DIRECTION DE L'ENSEIGNEMENT

DIRECTION DES ADMISSIONS ET CONCOURS

ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE SUPÉRIEURE DE COMMERCE DE PARIS
ÉCOLE EUROPÉENNE DES AFFAIRES
ÉCOLE SUPÉRIEURE DE COMMERCE DE LYON

Mathématiques II

OPTION : Générale

Mardi 3 Mai 1988, de 8 heures à 12 heures

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

sont autorisées : règles graduées, tables de valeurs numériques sans formulaire, calculatrices de poche, y compris les calculatrices programmables et alphanumériques, à fonctionnement autonome, sans imprimante, sans document d'accompagnement et de format maximum 21 cm de long x 15 cm de large.

Le but du problème est l'étude des réalisations de trois « face » consécutifs dans une suite de parties de pile ou face, ce qui fait l'objet de la partie III. À cet effet, on s'intéresse tout d'abord dans la partie I aux puissances de deux matrices, puis, dans la partie II, au comportement asymptotique d'une suite.

La répétition d'un même événement dans une suite de tirages indépendants constitue un modèle commode pour l'étude de certains problèmes de gestion, notamment de réapprovisionnement des stocks.

Partie I

On considère les deux matrices M et N suivantes :

$$M = \frac{1}{2} \begin{pmatrix} 1111 \\ 1001 \\ 0100 \\ 0010 \end{pmatrix} \quad N = \frac{1}{2} \begin{pmatrix} 1110 \\ 1000 \\ 0100 \\ 0012 \end{pmatrix}$$

1. Calculer M^2 , M^3 et M^4 . En déduire la relation : $4M^4 - 2M^3 - M^2 - M = 0$ (1)

2. Calculer N^2 , N^3 et N^4 . En déduire la relation : $8N^4 - 12N^3 + 2N^2 + N + I_4 = 0$ (2)

Partie II

On étudie dans cette partie la suite $(q_n)_{n \geq 1}$ telle que : $q_1 = q_2 = 0$, $q_3 = \frac{1}{8}$ et vérifiant pour $n \geq 1$ la relation :

$$8q_{n+3} = 4q_{n+2} + 2q_{n+1} + q_n \quad (3)$$

À cet effet, on considère la fonction g suivante, définie sur $\mathbb{R}$ par :

$$g(x) = 8x^3 - 4x^2 - 2x - 1$$

1. a) Étudier et représenter graphiquement la fonction g .

b) En déduire que l'équation $g(x) = 0$ admet une racine réelle r et une seule.

2. a) Donner l'équation $y = t(x)$ de la tangente en $x = 1$ à la courbe représentative de g . Calculer la racine ρ de l'équation

$$t(x) = 0.$$

b) Calculer $g(\rho)$ et $g(\rho - 0,01)$. Donner un encadrement de r d'amplitude 0,01.

3. a) Montrer qu'il existe un réel a et un seul, que l'on exprimera en fonction de r , tel que, pour tout réel x :

$$g(x) = (x - r) \left(8x^2 + ax + \frac{1}{r} \right)$$

b) Montrer que l'équation $g(x) = 0$ admet, outre la racine réelle r , deux racines complexes z et $\bar{z}$ (que l'on ne cherchera pas à expliciter). Évaluer le produit $z \cdot \bar{z}$ de ces deux racines. Exprimer $|z|$ en fonction de r et prouver que $|z| < r$.

TSVP

4. Soit E l'ensemble des suites réelles ou complexes $(u_n)_{n \geq 1}$ vérifiant la relation suivante, pour tout entier naturel non nul n :

$$8u_{n+3} = 4u_{n+2} + 2u_{n+1} + u_n$$

a) Montrer que les suites $(r^n)_{n \geq 1}$, $(z^n)_{n \geq 1}$, $(\bar{z}^n)_{n \geq 1}$ appartiennent à E .

b) On considère le système d'équations :

$$\begin{aligned} \alpha r + \beta z + \gamma \bar{z} &= 0 \\ \alpha r^2 + \beta z^2 + \gamma \bar{z}^2 &= 0 \\ \alpha r^3 + \beta z^3 + \gamma \bar{z}^3 &= \frac{1}{8} \end{aligned}$$

Montrer que : $\alpha = \frac{1}{8r^3 + ar^2 + 1}$, où a est le réel introduit dans la question précédente.

Prouver que $\beta = \gamma$. (On ne demande pas d'explicitier β et γ .)

c) En déduire que, pour tout entier naturel non nul n : $q_n = \alpha r^n + \beta z^n + \beta \bar{z}^n$ et $|q_n - \alpha r^n| \leq \frac{2|\beta|}{(8r)^{n/2}}$

d) Établir que : $q^n \sim \alpha r^n$ quand n tend vers l'infini. Calculer : $\lim_{n \rightarrow \infty} \frac{q_{n+1}}{q_n}$.

5. a) À l'aide de la relation (3) définissant (q_n) , donner un algorithme permettant de calculer q_k pour $k \leq n$ et obtenir ainsi des valeurs approchées de q_k pour $k \leq 15$ (que l'on donnera avec six décimales), puis du quotient $\frac{q_{k+1}}{q_k}$ (avec quatre décimales).

b) En procédant comme dans la question 2, donner un encadrement de r d'amplitude 0,000 1, puis une valeur approchée de α .

Partie III

A On effectue une suite infinie de pile ou face avec une pièce équilibrée. Pour tout entier naturel non nul n , on considère la variable aléatoire Y_n prenant pour valeur :

- 0 si le $n^{\text{ième}}$ jet a amené « pile » ;
- 1 si le $n^{\text{ième}}$ jet a amené « face », celui-ci étant le premier, ou le quatrième, ou le septième,...., ou le $(3k+1)^{\text{ième}}$,.... « face » obtenu depuis le précédent « pile » (ou depuis le début du jeu si « pile » n'est pas encore sorti) ;
- 2 si le $n^{\text{ième}}$ jet a amené « face », celui-ci étant le deuxième, ou le cinquième, ou le huitième,...., ou le $(3k+2)^{\text{ième}}$,.... « face » obtenu depuis le précédent « pile » (ou depuis le début du jeu si « pile » n'est pas encore sorti) ;
- 3 si le $n^{\text{ième}}$ jet a amené « face », celui-ci étant le troisième, ou le sixième, ou le neuvième,...., ou le $(3k+3)^{\text{ième}}$,.... « face » obtenu depuis le précédent « pile » (ou depuis le début du jeu si « pile » n'est pas encore sorti).

On dit qu'une série de trois « face » consécutifs s'achève à l'issue du $n^{\text{ième}}$ jet si et seulement si l'événement $[Y_n = 3]$ est réalisé.

Exemple. Soit la suite de résultats (F désigne « face » et P désigne « pile ») :

numéro du jet	1	2	3	4	5	6	7	8	9	10	11	12	13	14...
résultat	P	F	F	P	P	P	F	F	F	F	F	F	P	F...

Alors : $Y_1 = 0$, $Y_2 = 1$, $Y_3 = 2$, $Y_4 = 0$, $Y_5 = 0$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 2$, $Y_9 = 3$, $Y_{10} = 1$, $Y_{11} = 2$, $Y_{12} = 3$, $Y_{13} = 0$, $Y_{14} = 1$,.... et trois « face » consécutifs ont été obtenus aux neuvième, douzième,.... jets.

1. Exprimer pour $i = 0$, $i = 1$, $i = 2$ et $i = 3$ les probabilités $P([Y_{n+1} = i])$ en fonction des probabilités $P([Y_n = j])$ pour $j = 0$, $j = 1$, $j = 2$ et $j = 3$.

2. On pose : $a_n = P([Y_n = 0])$, $b_n = P([Y_n = 1])$, $c_n = P([Y_n = 2])$, $d_n = P([Y_n = 3])$. Pour tout entier naturel non nul n , soit V_n la matrice-colonne définie par :

$$V_n = \begin{pmatrix} a_n \\ b_n \\ c_n \\ d_n \end{pmatrix}$$

a) Montrer que, pour tout entier naturel non nul n : $V_{n+1} = M V_n$

où la matrice M est définie dans la partie I.

b) En multipliant à droite la relation (1) par V_{n-1} (où $n \geq 2$), établir une relation entre V_{n+3} , V_{n+2} , V_{n+1} et V_n , puis entre d_{n+3} , d_{n+2} , d_{n+1} et d_n .

c) Calculer d_1 , d_2 , d_3 , d_4 .

d) À l'aide de la relation définissant (d_n) , donner un algorithme permettant de calculer d_k pour $k \leq n$. Obtenir ainsi des valeurs approchées de d_k pour $k \leq 15$ (que l'on donnera avec six décimales). Que constate-t-on ?

B On s'intéresse maintenant à la variable aléatoire X indiquant le numéro du jet où, pour la première fois, on a obtenu trois fois de suite le résultat « face ». Pour tout entier naturel non nul n , on considère la variable aléatoire Z_n :

— prenant pour valeur 3 si trois « face » consécutifs sont obtenus à l'issue du $n^{\text{ième}}$ jet ou ont été déjà obtenus à l'issue d'un jet antérieur ;

— prenant sinon pour valeur :

0 si le $n^{\text{ième}}$ jet a amené « pile » ;

1 si le $n^{\text{ième}}$ jet a amené « face », ce « face » étant le premier obtenu depuis le précédent « pile » (ou depuis le début du jeu si « pile » n'est pas encore sorti) ;

2 si le $n^{\text{ième}}$ jet a amené « face », ce « face » étant le second obtenu depuis le précédent « pile » (ou depuis le début du jeu si « pile » n'est pas encore sorti).

L'événement $[X \leq n]$ est donc réalisé si et seulement si l'événement $[Z_n = 3]$ est réalisé.

Exemple. En reprenant la suite de pile ou face donnée dans la partie III.A, on a : $X = 9$, $Z_1 = 0$, $Z_2 = 1$, $Z_3 = 2$, $Z_4 = 0$, $Z_5 = 0$, $Z_6 = 0$, $Z_7 = 1$, $Z_8 = 2$ et, pour $n \geq 9$, $Z_n = 3$.

1. Exprimer pour $i = 0$, $i = 1$, $i = 2$ et $i = 3$ les probabilités $P([Z_{n+1} = i])$ en fonction des probabilités $P([Z_n = j])$ pour $j = 0$, $j = 1$, $j = 2$ et $j = 3$.

2. On pose : $a'_n = P([Z_n = 0])$, $b'_n = P([Z_n = 1])$, $c'_n = P([Z_n = 2])$, $d'_n = P([Z_n = 3])$. Pour tout entier naturel non nul n , soit W_n la matrice-colonne définie par :

$$W_n = \begin{pmatrix} a'_n \\ b'_n \\ c'_n \\ d'_n \end{pmatrix}$$

a) Montrer que, pour tout entier naturel non nul n :

$$W_{n+1} = N W_n$$

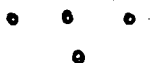
où la matrice N est définie dans la partie I.

b) En multipliant à droite la relation (2) par W_n , établir une relation entre W_{n+4} , W_{n+3} , W_{n+2} , W_{n+1} et W_n , puis entre d'_{n+4} , d'_{n+3} , d'_{n+2} , d'_{n+1} et d'_n .

c) Calculer $P([X = 1])$, $P([X = 2])$, $P([X = 3])$ et $P([X = 4])$. Montrer que, pour $n \geq 2$:

$$P([X = n]) = d'_n - d'_{n-1}$$

d) Comparer $P([X = n])$ à la suite (q_n) étudiée dans la partie II. En déduire un équivalent de $P([X = n])$ et la limite de la suite $\frac{P([X = n+1])}{P([X = n])}$.



ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE SUPÉRIEURE DE COMMERCE DE PARIS
ÉCOLE EUROPÉENNE DES AFFAIRES
ÉCOLE SUPÉRIEURE DE COMMERCE DE LYON

Mathématiques II

OPTION : Générale

4 heures

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

sont autorisées : règles graduées, tables de valeurs numériques sans formulaire, calculatrices de poche, y compris les calculatrices programmables et alphanumériques, à fonctionnement autonome, sans imprimante, sans document d'accompagnement et de format maximum 21 cm de long x 15 cm de large.

L'objet du problème est l'étude d'une promenade aléatoire (parties III et IV). Dans les parties I et II on établit des résultats liminaires qui seront utilisés dans la partie III.

PARTIE I

Soient y_1, y_2, y_3, y_4 des nombres réels. On considère le système d'équations :

$$\begin{pmatrix} -\frac{3}{4} & 1 & 0 & 0 \\ \frac{1}{4} & -1 & 1 & 0 \\ \frac{1}{4} & 0 & -1 & 1 \\ \frac{1}{4} & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}.$$

1. Déterminer une condition nécessaire et suffisante portant sur y_1, y_2, y_3, y_4 pour que ce système admette au moins une solution.

2. On suppose que cette condition est satisfaite. Soit a un nombre réel. Exprimer en fonction de y_2, y_3, y_4 l'unique solution (x_1, x_2, x_3, x_4) vérifiant la relation :

$$x_1 + x_2 + x_3 + x_4 = a.$$

PARTIE II

On considère une suite réelle (p_n) satisfaisant à la relation de récurrence :

$$(1) \quad p_{n+4} = \frac{1}{4} (p_{n+3} + p_{n+2} + p_{n+1} + p_n).$$

On lui associe les deux suites (m_n) et (M_n) définies par :

$$m_n = \min (p_n, p_{n+1}, p_{n+2}, p_{n+3}) ; \quad M_n = \max (p_n, p_{n+1}, p_{n+2}, p_{n+3}).$$

(m_n et M_n sont donc le plus petit et le plus grand des nombres réels $p_n, p_{n+1}, p_{n+2}, p_{n+3}$.)

1. Dans cette question, on établit la convergence des suites (m_n) et (M_n) .

a) Montrer que m_n est inférieur ou égal aux nombres $p_{n+1}, p_{n+2}, p_{n+3}$ et p_{n+4} . En déduire que la suite (m_n) est croissante. Établir de même que la suite (M_n) est décroissante.

b) Prouver que, pour tout nombre entier naturel n : $m_0 \leq m_n \leq p_n \leq M_n \leq M_0$.

c) Prouver que les suites (m_n) et (M_n) sont convergentes et que leurs limites respectives, notées m et M , vérifient :

$$m \leq M.$$

2. Dans cette question, on établit la convergence de la suite (p_n) .

a) Montrer que, pour tout nombre entier naturel n :

$$\begin{aligned} p_{n+4} &\leq \frac{3}{4} M_n + \frac{1}{4} m_n \\ p_{n+4} &\leq \frac{3}{4} M_n + \frac{1}{4} m. \end{aligned}$$

En appliquant la dernière inégalité à $p_{n+5}, p_{n+6}, p_{n+7}$, montrer que :

$$M_{n+4} \leq \frac{3}{4} M_n + \frac{1}{4} m.$$

b) En déduire que $M \leq m$, puis que $M = m$.

c) Établir la convergence de la suite (p_n) .

T. S. V. P.

3. Dans cette question, on étudie numériquement la suite (p_n) vérifiant la relation de récurrence (1) et les conditions initiales :

$$p_0 = 1 ; \quad p_1 = \frac{1}{4} ; \quad p_2 = \frac{1}{4} + \frac{1}{16} ; \quad p_3 = \frac{1}{4} + \frac{1}{8} + \frac{1}{64} .$$

a) Rédiger un algorithme (en français) permettant de calculer, pour tout nombre entier naturel N , les termes $p_0, p_1, \dots, p_N$ de la suite (p_n) .

b) Utiliser cet algorithme pour donner des valeurs décimales approchées (à la précision de la calculatrice employée) de $p_0, p_1, \dots, p_{10}$ et encadrer $\lim_{n \rightarrow +\infty} p_n$.

PARTIE III

Dans la suite du problème, on étudie la promenade aléatoire d'un jeton sur les quatre cases C_1, C_2, C_3, C_4 suivantes :

C_1	C_2	C_3	C_4
-------	-------	-------	-------

Au cours des instants successifs $0, 1, 2, \dots, n, \dots$, on y déplace un jeton de la manière suivante :

a) À l'instant 0, le jeton est placé sur C_1 .

b) Si, à l'instant n , le jeton est placé sur C_1 , on le place à l'instant $n+1$ sur l'une des cases C_1, C_2, C_3, C_4 , le choix d'une de ces cases s'effectuant de manière équiprobable (et indépendamment des positions du jeton aux instants antérieurs).

c) Si, à l'instant n , le jeton est placé sur C_i , où $2 \leq i \leq 4$, on le place à l'instant $n+1$ sur la case C_{i-1} .

Pour tout nombre entier naturel n et pour i égal à 1, 2, 3 ou 4, on note désormais :

$Z(n, i)$ la variable aléatoire prenant pour valeur 1 si le jeton est sur la case C_i à l'instant n , et 0 dans le cas contraire ;

$q(n, i)$ la probabilité pour que le jeton soit sur la case C_i à l'instant n .

On pose :

$$Q_n = \begin{pmatrix} q(n, 1) \\ q(n, 2) \\ q(n, 3) \\ q(n, 4) \end{pmatrix} .$$

1. Dans cette question, on étudie les suites $(q(n, i))$, où i est égal à 1, 2, 3 ou 4.

a) Expliciter la matrice A d'ordre 4 telle que, pour tout nombre entier naturel n :

$$Q_{n+1} = A Q_n .$$

b) En déduire les relations suivantes :

$$\begin{cases} q(n+1, 4) = \frac{1}{4} q(n, 1) \\ q(n+2, 3) = \frac{1}{4} [q(n, 1) + q(n+1, 1)] \\ q(n+3, 2) = \frac{1}{4} [q(n, 1) + q(n+1, 1) + q(n+2, 1)] \\ q(n+4, 1) = \frac{1}{4} [q(n, 1) + q(n+1, 1) + q(n+2, 1) + q(n+3, 1)] \end{cases}$$

c) Calculer $q(0, 1), q(1, 1), q(2, 1), q(3, 1)$. Comparer la suite $(q(n, 1))$ à la suite définie dans la question II.3. En déduire que les quatre suites $(q(n, i))$, où i est égal à 1, 2, 3 ou 4, sont convergentes ; exprimer leurs limites en fonction de la limite L de la suite $(q(n, 1))$.

d) Calculer $q(n, 1) + q(n, 2) + q(n, 3) + q(n, 4)$. En déduire les limites des quatre suites $(q(n, i))$ pour i égal à 1, 2, 3 ou 4.

T. S. V. P

2. Pour tout nombre entier naturel n et pour i égal à 1, 2, 3 ou 4, on pose :

$$Y(n, i) = Z(0, i) + Z(1, i) + \dots + Z(n, i).$$

$Y(n, i)$ est donc la variable aléatoire indiquant le nombre de passages du jeton sur la case C_i au cours des instants 0, 1, 2, ..., n . On détermine dans cette question le nombre moyen des passages du jeton sur l'une des cases au cours de ces instants, autrement dit les espérances des variables aléatoires $Y(n, i)$.

- Déterminer la somme $Y(n, 1) + Y(n, 2) + Y(n, 3) + Y(n, 4)$.
- Pour tout nombre entier naturel n , on pose :

$$E_n = \begin{pmatrix} E[Y(n, 1)] \\ E[Y(n, 2)] \\ E[Y(n, 3)] \\ E[Y(n, 4)] \end{pmatrix}.$$

Exprimer l'espérance $E[Z(n, i)]$ en fonction de $q(n, i)$. En déduire que :

$$E_n = Q_0 + Q_1 + \dots + Q_n.$$

- Déduire des résultats précédents que :

$$(A - I_4) E_n = Q_{n+1} - Q_0,$$

où I_4 est la matrice identité d'ordre 4, et que :

$$E[Y(n, 1)] + E[Y(n, 2)] + E[Y(n, 3)] + E[Y(n, 4)] = n + 1.$$

Déduire des résultats de la partie I l'expression de $E[Y(n, i)]$, où i est égal à 1, 2, 3 ou 4, en fonction de $q(n+1, j)$, où j est égal à 2, 3 ou 4.

- Pour i égal à 1, 2, 3 ou 4, expliciter des nombres réels f_i et g_i tels que :

$$E[Y(n, i)] = f_i n + g_i + \varepsilon_i(n) \quad \text{avec} \quad \lim_{n \rightarrow +\infty} \varepsilon_i(n) = 0.$$

PARTIE IV

On étudie dans cette partie les deux variables aléatoires suivantes :

U , indiquant le premier instant $n \geq 1$ où le jeton se trouve sur la case C_2 ;

V , indiquant le premier instant $n \geq 1$ où le jeton se trouve sur la case C_3 .

- Calculer les probabilités $P(U = 1)$ et $P(U = 2)$, puis $P(U = n)$ pour $n \geq 3$.
- Vérifier que :

$$\sum_{n=1}^{+\infty} P(U = n) = 1.$$

- Calculer l'espérance de U .

2. Pour tout nombre entier naturel n et pour i égal à 1, 2 ou 4, on note $\pi(n, i)$ la probabilité pour que le jeton soit placé à l'instant n sur la case C_i sans jamais avoir été placé au cours des instants 0, 1, 2, ..., n sur la case C_3 .

- Calculer $\pi(0, 1)$ et $\pi(1, 1)$. Montrer que, pour tout nombre $n \geq 2$:

$$\pi(n, 1) = \frac{1}{4} [\pi(n-1, 1) + \pi(n-2, 1)].$$

En déduire la valeur de $\pi(n, 1)$ en fonction de n .

- Calculer $\pi(n, 2)$ et $\pi(n, 4)$.
- Calculer la probabilité $P(V = n)$ et l'espérance de V .



CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS
DIRECTION DE L'ENSEIGNEMENT
Direction des Admissions et Concours

ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE SUPÉRIEURE DE COMMERCE DE PARIS
ÉCOLE EUROPÉENNE DES AFFAIRES
ÉCOLE SUPÉRIEURE DE COMMERCE DE LYON

CONCOURS D'ADMISSION DE 1990

PRÉAMBULE

Dans tout le problème, on désigne par a et b deux nombres réels tels que $a < b$ et par φ une fonction numérique continue et strictement croissante sur $[a, b]$, telle que $\varphi(a) \leq 0 \leq \varphi(b)$. On note ω l'unique nombre réel tel que $a \leq \omega \leq b$ et $\varphi(\omega) = 0$.

On appelle évaluation de φ le calcul d'une valeur de φ . L'objet du problème est l'étude de deux types d'algorithmes permettant d'obtenir à partir de l'encadrement initial $a \leq \omega \leq b$ un encadrement plus fin à l'aide d'évaluations successives de la fonction φ .

EG

Dans le cas du premier type (partie III), on fixe une précision et on estime le nombre moyen d'évaluations de φ à effectuer pour obtenir cette précision.

Dans le cas du second type (partie IV), on impose le nombre d'évaluations de φ à effectuer et l'on estime la précision moyenne obtenue.

On se propose notamment de montrer que, [pour chacun des deux types] il existe un algorithme optimal, en un sens que l'on précisera.

La partie IV est indépendante des autres.

PARTIE I : Étude d'une fonction

Soit f la fonction numérique définie sur $]0, +\infty[$ par les relations :

$$\begin{cases} f(x) = \frac{(x+2)(x-1)}{x \ln x} & \text{si } x \neq 1 \\ f(1) = 3 \end{cases}$$

1. a) Montrer que la fonction f est continue.

b) Déterminer la limite de f en 0.

c) Déterminer la limite de f en $+\infty$.

2. Montrer que f est dérivable en 1 et déterminer $f'(1)$.

3. a) Calculer la dérivée de f . Montrer que si $x \neq 1$, le signe de $f'(x)$ est celui de : $g(x) = \ln x - \frac{x^2 + x - 2}{x^2 + 2}$

b) Calculer la dérivée de g et montrer que le signe de $g'(x)$ est celui de : $h(x) = x^4 + x^3 - 4x^2 - 2x + 4$

c) En remarquant que 1 et -2 sont racines de la fonction polynomiale h , décomposer $h(x)$ en produit de facteurs du premier degré.

4. a) Dresser le tableau de variation de g .

b) Montrer que l'équation $g(x) = 0$ admet une solution t et une seule autre que 1.

Vérifier que $\sqrt{2} < t < 2$.

5. a) Dresser le tableau de variation de f .

b) Donner l'allure de la courbe représentative de f .

6. Application. Pour tout nombre entier $p \geq 2$, on pose :

Montrer que la suite $(c_p)_{p \geq 2}$ est strictement croissante.

$$c_p = \frac{(p+2)(p-1)}{p \ln p}$$

Dans toute la suite du problème, on désigne par p un nombre entier supérieur ou égal à 2.

.../...

Mathématiques II

OPTION GÉNÉRALE

OPTION ÉCONOMIQUE · OPTION TECHNOLOGIQUE

Vendredi 4 mai 1990, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entrent pour une part importante dans l'appréciation des copies.

sont autorisées : règles graduées, tables de valeurs numériques sans formulaire, calculatrices de poche, y compris les calculatrices programmables et alphanumériques, à fonctionnement autonome, sans imprimante, sans document d'accompagnement et de format maximum 21 cm de long x 15 cm de large, à raison d'une seule calculatrice par candidat.

N. B. : | OG : Option Générale Seule
ET : Options Économique et Techno. Seules

PARTIE II : Étude d'une variable aléatoire

1. On considère une variable aléatoire Z à valeurs dans l'ensemble $\{1, 2, \dots, p-1\}$ et telle que :

$$\begin{cases} P(Z = i) = \frac{1}{p} & \text{si } i \neq p-1 \\ P(Z = p-1) = \frac{2}{p} \end{cases}$$

Ainsi, lorsque $p = 2$, Z est une variable aléatoire certaine de valeur 1.

- Calculer l'espérance de Z .
- Calculer la variance de Z . Contrôler le résultat trouvé en prenant $p = 2$.

2. *Application à l'étude d'un algorithme.* On considère p boîtes $B_1, B_2, \dots, B_p$. Un objet Ω est caché dans l'une de ces boîtes ; on cherche à le localiser, c'est-à-dire à déterminer le numéro de la boîte qui le contient, grâce à l'algorithme suivant.

On ouvre successivement les boîtes $B_1, B_2, \dots, B_k$ jusqu'à ce que l'une ou l'autre des deux situations suivantes soit réalisée :

- l'objet Ω est découvert ;
- une seule boîte n'a pas encore été ouverte.

On remarque que cet algorithme permet effectivement de localiser l'objet Ω .

On suppose que la boîte dans laquelle est caché Ω a été choisie au hasard, de façon équiprobable, parmi les p boîtes. On note Y la variable aléatoire indiquant le nombre de boîtes ouvertes au cours de la mise en œuvre de l'algorithme. Montrer que Y suit la même loi que la variable aléatoire Z définie dans la question II.1.

En déduire le nombre moyen de boîtes que l'on ouvre pour localiser l'objet.

PARTIE III : Étude d'un premier type d'algorithme

On rappelle que a, b, φ et ω ont été définis dans le préambule.

1. *Algorithme $\mathcal{A}_p$.* On se propose d'étudier un algorithme $\mathcal{A}_p$ permettant d'obtenir à partir d'un encadrement $u \leq \omega \leq v$, où u et v sont des éléments distincts de $[a, b]$, un nouvel encadrement $u' \leq \omega \leq v'$ tel que :

$$u \leq u' \leq \omega \leq v' \leq v \quad \text{et} \quad v' - u' = \frac{v - u}{p} \quad (1)$$

On définit cet algorithme de la façon suivante. On effectue les évaluations successives :

$$\varphi\left(u + \frac{v-u}{p}\right), \quad \varphi\left(u + 2\frac{v-u}{p}\right), \dots, \varphi\left(u + k\frac{v-u}{p}\right)$$

jusqu'à ce que l'une ou l'autre des deux situations suivantes soit réalisée :

$$\varphi\left(u + k\frac{v-u}{p}\right) \geq 0 \quad \text{ou} \quad k = p-1$$

$$\begin{aligned} \text{Si } \varphi\left(u + k\frac{v-u}{p}\right) \geq 0, \text{ on pose : } & \begin{cases} u' = u + \frac{(k-1)(v-u)}{p} \\ v' = u + \frac{k(v-u)}{p} \end{cases} \\ \text{Si } \varphi\left(u + k\frac{v-u}{p}\right) < 0, \text{ on pose : } & \begin{cases} u' = u + \frac{(p-1)(v-u)}{p} \\ v' = u + \frac{p(v-u)}{p} \end{cases} \end{aligned}$$

Montrer que les conditions (1) sont effectivement satisfaites.

Suite de l'énoncé option générale : les deux pages suivantes
Suite de l'énoncé options Eco. et Tech. : troisième page qui suit

06

2. *Application répétée de l'algorithme $\mathcal{A}_p$.* Soit ε un nombre réel tel que $0 < \varepsilon < 1$. À partir de l'encadrement initial $a \leq \omega \leq b$, on cherche un encadrement $c \leq \omega \leq d$ tel que $d - c < \varepsilon(b - a)$. À cet effet, on itère l'algorithme $\mathcal{A}_p$ afin d'obtenir les encadrements successifs $a_1 \leq \omega \leq b_1$, $a_2 \leq \omega \leq b_2, \dots, a_n \leq \omega \leq b_n$, où l'encadrement $a_i \leq \omega \leq b_i$ est obtenu à partir de l'encadrement $a_{i-1} \leq \omega \leq b_{i-1}$ par application de l'algorithme $\mathcal{A}_p$. (On convient que $a_0 = a$ et $b_0 = b$.) On arrête les calculs dès que $b_n - a_n < \varepsilon(b - a)$. Ainsi, n est le nombre de mises en œuvre de l'algorithme nécessaires pour obtenir la précision souhaitée.

a) Exprimer $b_i - a_i$ en fonction de $b - a$.

b) Montrer que :

$$n = \left\lceil 1 - \frac{\ln \varepsilon}{\ln p} \right\rceil$$

Autrement dit, n est la partie entière de $1 - \frac{\ln \varepsilon}{\ln p}$.

3. *Estimation du nombre moyen d'évaluations à effectuer.* Pour évaluer la performance de la méthode décrite dans la question précédente, on suppose que ω suit une loi de probabilité uniforme sur l'intervalle $[a, b]$.

Pour tout nombre entier naturel i tel que $1 \leq i \leq n$, on note Y_i la variable aléatoire indiquant le nombre d'évaluations de φ effectuées au cours de la $i^{\text{ème}}$ mise en œuvre de l'algorithme $\mathcal{A}_p$.

a) Montrer que Y_i suit la même loi que la variable aléatoire Y définie dans la question II.2.

b) On note $V_p(\varepsilon) = Y_1 + Y_2 + \dots + Y_n$ la variable aléatoire qui indique le nombre total d'évaluations de φ effectuées au cours des n mises en œuvre de l'algorithme. Calculer l'espérance, notée $\overline{V_p(\varepsilon)}$, de $V_p(\varepsilon)$.

c) Montrer que, lorsque ε est au voisinage de 0 (l'entier p étant fixé) :

$$\overline{V_p(\varepsilon)} \sim \frac{c_p}{2} (-\ln \varepsilon), \quad \text{où } c_p = \frac{(p+2)(p-1)}{p \ln p}$$

d) Lorsque ε est suffisamment proche de 0, on assimile $\overline{V_p(\varepsilon)}$ à $\frac{c_p}{2} (-\ln \varepsilon)$. En utilisant cette approximation, déterminer la valeur optimale de p , c'est-à-dire celle qui minimise le nombre moyen d'évaluations de φ qu'on doit effectuer dans la procédure décrite dans la question III.2.

PARTIE IV : Étude d'un second type d'algorithme

Dans cette partie, on désigne par q un nombre réel tel que $0 < q < 1$ et par N un nombre entier supérieur ou égal à 2.

1. *Algorithme $\mathcal{B}_q$.* On se propose d'étudier un algorithme $\mathcal{B}_q$ permettant d'obtenir, à partir d'un encadrement $\alpha \leq \omega \leq \beta$, où α et β sont deux éléments distincts de $[a, b]$, un nouvel encadrement $\alpha' \leq \omega \leq \beta'$ tel que :

$$\alpha \leq \alpha' \leq \omega \leq \beta' \leq \beta \quad \text{et} \quad \beta' - \alpha' \leq (\beta - \alpha) \sup(q, 1 - q) \quad (2)$$

On définit cet algorithme de la façon suivante. On pose $\gamma = (1 - q)\alpha + q\beta$. On calcule $\varphi(\gamma)$.

Si $\varphi(\gamma) \geq 0$, on pose :

$$\begin{cases} \alpha' = \alpha \\ \beta' = \gamma \end{cases}$$

Si $\varphi(\gamma) < 0$, on pose :

$$\begin{cases} \alpha' = \gamma \\ \beta' = \beta \end{cases}$$

Montrer que les conditions (2) sont effectivement satisfaites, en précisant dans chaque cas la valeur de $\beta' - \alpha'$.

... / ...

06

2. *Étude d'une variable aléatoire.* On conserve les notations de la question IV.1. On suppose connu un encadrement $\alpha \leq \omega \leq \beta$ et on suppose que ω suit une loi de probabilité uniforme sur l'intervalle $[\alpha, \beta]$. On note R_q la variable aléatoire prenant la valeur $\beta' - \alpha'$.

- Déterminer l'ensemble des valeurs prises par R_q . Examiner le cas où $q = \frac{1}{2}$.
- Déterminer la loi de probabilité de R_q .
- Montrer que l'espérance de R_q est $(2q^2 - 2q + 1)(\beta - \alpha)$.
- Calculer la variance de R_q en distinguant les cas $q \neq \frac{1}{2}$ et $q = \frac{1}{2}$.

3. *Application répétée de l'algorithme B_q .* On itère N fois l'algorithme B_q afin d'obtenir, à partir de l'encadrement initial $a \leq \omega \leq b$, les encadrements successifs $\alpha_1 \leq \omega \leq \beta_1, \alpha_2 \leq \omega \leq \beta_2, \dots, \alpha_N \leq \omega \leq \beta_N$, où l'encadrement $\alpha_i \leq \omega \leq \beta_i$ est obtenu à partir de l'encadrement $\alpha_{i-1} \leq \omega \leq \beta_{i-1}$ par application de l'algorithme B_q . (On convient que $\alpha_0 = a$ et $\beta_0 = b$.)

Pour évaluer la performance de cette méthode, on suppose que ω suit une loi de probabilité uniforme sur l'intervalle $[a, b]$. On note T_q la variable aléatoire dont la valeur est $\beta_N - \alpha_N$.

- Pour tout nombre entier i tel que $1 \leq i \leq N$, on pose :

$$\gamma_i = (1 - q) \alpha_{i-1} + q\beta_{i-1}$$

On note X_i la variable aléatoire prenant la valeur 1 si $\varphi(\gamma_i) \geq 0$ et la valeur 0 dans le cas contraire.

Montrer que les variables aléatoires X_i suivent une même loi que l'on précisera.

- On pose $S_q = X_1 + X_2 + \dots + X_N$. Montrer que S_q suit la loi binomiale $\mathcal{B}(N, q)$.
- Montrer que si $S_q = k$, alors :

$$\beta_N - \alpha_N = q^k (1 - q)^{N-k} (b - a)$$

En déduire l'ensemble des valeurs que peut prendre T_q . Examiner le cas où $q = \frac{1}{2}$.

On suppose que $q \neq \frac{1}{2}$. Montrer que, pour tout nombre entier naturel k inférieur ou égal à N :

$$P(T_q = q^k (1 - q)^{N-k} (b - a)) = C_N^k q^k (1 - q)^{N-k}$$

- Calculer l'espérance et la variance de T_q lorsque $q \neq \frac{1}{2}$. Vérifier que les résultats obtenus restent valables lorsque $q = \frac{1}{2}$.

4. *Comparaison des algorithmes B_q .* Montrer que $\frac{1}{2}$ est la valeur de q qui minimise l'espérance de T_q lorsque N est fixé et que q parcourt $]0, 1[$. Conclure.

fin de l'énoncé option générale.

ET

Dans toute la suite, on désigne par m un nombre entier supérieur ou égal à 2.

2. Application répétée de l'algorithme $\mathcal{A}_p$. À partir de l'encadrement initial $a \leq \omega \leq b$, on cherche un encadrement $c \leq \omega \leq d$ tel que $d - c < \frac{b-a}{m}$. À cet effet, on itère l'algorithme $\mathcal{A}_p$ afin d'obtenir les encadrements successifs $a_1 \leq \omega \leq b_1$, $a_2 \leq \omega \leq b_2, \dots, a_n \leq \omega \leq b_n$, où l'encadrement $a_i \leq \omega \leq b_i$ est obtenu à partir de l'encadrement $a_{i-1} \leq \omega \leq b_{i-1}$ par application de l'algorithme $\mathcal{A}_p$. (On convient que $a_0 = a$ et $b_0 = b$.) On arrête les calculs dès que $b_n - a_n < \frac{b-a}{m}$. Ainsi, n est le nombre de mises en œuvre de l'algorithme nécessaires pour obtenir la précision souhaitée.

a) Exprimer $b_i - a_i$ en fonction de $b - a$.

b) Montrer que :

$$n = \left\lceil 1 + \frac{\ln m}{\ln p} \right\rceil$$

(On rappelle que, pour tout nombre réel t , le symbole $[t]$ désigne la partie entière de t , c'est-à-dire le plus grand des nombres entiers relatifs inférieurs ou égaux à t .)

3. Estimation du nombre moyen d'évaluations à effectuer. Pour évaluer la performance de la méthode décrite dans la question précédente, on suppose que ω suit une loi de probabilité uniforme sur l'intervalle $[a, b]$.

Pour tout nombre entier naturel i tel que $1 \leq i \leq n$, on note Y_i la variable aléatoire indiquant le nombre d'évaluations de φ effectuées au cours de la $i^{\text{ème}}$ mise en œuvre de l'algorithme $\mathcal{A}_p$.

a) Montrer que Y_i suit la même loi que la variable aléatoire Y définie dans la question II.2.

b) On note $V_p(m) = Y_1 + Y_2 + \dots + Y_n$ la variable aléatoire qui indique le nombre total d'évaluations de φ effectuées au cours des n mises en œuvre de l'algorithme. Calculer l'espérance, notée $\overline{V_p(m)}$, de $V_p(m)$.

c) On suppose $p \geq m + 1$. Montrer que :

$$\overline{V_p(m)} = \frac{(p+2)(p-1)}{2p}$$

d) En déduire que, pour m fixé, le meilleur choix de p , c'est-à-dire celui qui minimise $\overline{V_p(m)}$, se situe entre 2 et $m + 1$.

4. Comparaison des deux choix extrêmes ($p = 2$, $p = m + 1$)

a) Expliciter $\overline{V_2(m)}$ et $\overline{V_{m+1}(m)}$.

b) Déterminer le signe de $\overline{V_2(m)} - \overline{V_{m+1}(m)}$ pour m variant de 2 à 6.

c) Montrer que pour m suffisamment grand : $\overline{V_2(m)} < \overline{V_{m+1}(m)}$.

5. Comparaison des algorithmes $\mathcal{A}_p$

a) Montrer que lorsque x tend vers $+\infty$, $\frac{[x]}{x}$ tend vers 1.

b) L'entier p étant fixé et toujours supérieur ou égal à 2, montrer que, lorsque m tend vers $+\infty$, le rapport $\frac{\overline{V_p(m)}}{\ln m}$ tend vers $\frac{c_p}{2}$.

c) On prend cette fois $p \geq 3$. Montrer que, pour m suffisamment grand :

$$\overline{V_2(m)} < \overline{V_p(m)}$$

Conclure.

fin de l'énoncé options Economique et Technologique

ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE SUPÉRIEURE DE COMMERCE DE PARIS
ÉCOLE EUROPÉENNE DES AFFAIRES
ÉCOLE SUPÉRIEURE DE COMMERCE DE LYON

Mathématiques II

(4 heures)

OPTION GÉNÉRALE seule entre { }

OPTION ÉCONOMIQUE - OPTION TECHNOLOGIQUE seule entre []

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

sont autorisées : règles graduées, tables de valeurs numériques sans formulaire, calculatrices de poche, y compris les calculatrices programmables et alphanumériques, à fonctionnement autonome, sans imprimante, sans document d'accompagnement et de format maximum 21 cm de long X 15 cm de large.

L'objet du problème est une étude de gain associé à un jeu de pile ou face. La première partie permet d'établir quelques résultats liminaires d'analyse ; la seconde partie étudie la stratégie d'un joueur.

Dans tout le problème, on désigne par x un nombre réel appartenant à $]0, 1[$.

PARTIE I

1. Pour tout nombre entier naturel n , on pose : $s(n, 0) = 1 + x + x^2 + \dots + x^n$

Calculer $s(n, 0)$ et sa limite lorsque n tend vers $+\infty$.

2. Pour tout nombre entier naturel n , on pose : $s(n, 1) = 1 + 2x + 3x^2 + \dots + (n+1)x^n$

[a) Déterminer la limite de la suite (nx^n) lorsque n tend vers $+\infty$.]

b) Exprimer $(1-x)s(n, 1)$ à l'aide de $s(n, 0)$ et en déduire la limite de $s(n, 1)$ lorsque n tend vers $+\infty$.

c) Retrouver ce résultat à l'aide de la dérivation.

3. Plus généralement, pour tout couple (n, r) de nombres entiers naturels, on pose : $s(n, r) = \sum_{k=0}^n C_{r+k}^r x^k$

a) On suppose que n et r sont non nuls. On rappelle que, si $r \leq n$: $C_n^r = \frac{n}{r} C_{n-1}^{r-1}$

et que, pour tout nombre entier naturel non nul k : $C_{r+k}^r - C_{r+k-1}^r = C_{k+r-1}^{r-1}$

Déduire de ce dernier résultat que : $(1-x)s(n, r) = s(n, r-1) - C_{n+r}^r x^{n+1}$

b) Déterminer les limites des suites de termes généraux $n^r x^n$ et $C_{n+r}^r x^n$ lorsque n tend vers $+\infty$. En déduire par récurrence que, lorsque n tend vers $+\infty$, $s(n, r)$ tend vers la limite :

$$s(r) = \frac{1}{(1-x)^{r+1}}$$

[c) Soit y un nombre réel strictement positif. Déterminer suivant la valeur de y la nature de la série de terme général $(C_{r+k}^r y^k)$, le nombre r étant fixé.]

A SUIVRE

PARTIE II

On désigne par n et N des nombres entiers naturels non nuls. On considère une succession (éventuellement infinie) de jets d'une pièce. On suppose que la probabilité d'obtenir pile lors d'un jet est $1 - x$ et que la probabilité d'obtenir face est x . Les jets sont supposés indépendants.

On désigne enfin par S_n le nombre de fois où l'on a obtenu pile au cours des n premiers jets, par T_n le numéro du jet où l'on obtient pile pour la $n^{\text{ième}}$ fois.

A) 1. Préciser la loi de S_n . Calculer l'espérance et la variance de cette variable aléatoire.

2. Préciser la loi de T_1 . Calculer l'espérance et la variance de cette variable aléatoire.

3. L'objet de cette question est de calculer l'espérance et la variance de T_r .

Soient k un nombre entier naturel et r un nombre entier naturel non nul.

a) Montrer que l'événement $\{T_r = k + r\}$ est réalisé si et seulement si les événements $\{S_{k+r-1} = r - 1\}$ et "pile est obtenu au $(k + r)^{\text{ième}}$ jet" le sont. En déduire la loi de T_r . Vérifier que la somme des probabilités des événements $\{T_r = k + r\}$, où k appartient à $\mathbb{N}$, est égale à 1.

b) Calculer l'espérance de T_r en utilisant la limite $s(r)$ de la suite $s(n, r)$ introduite dans la partie I.

{ c) Calculer de même $E(T_r^2) + E(T_r)$. En déduire la variance de T_r . }

B) On décide que le jeu s'arrête dès que soit pile, soit face a été obtenu pour la $N^{\text{ième}}$ fois. Soit Z le nombre de jets nécessaires pour que le jeu s'arrête.

1. Donner l'ensemble des valeurs que peut prendre Z .

2. En utilisant une méthode analogue à celle de la question A) 3. a), déterminer, pour tout nombre entier naturel k , la probabilité pour que le jeu s'arrête au $(N + k)^{\text{ième}}$ jet, pile étant obtenu pour la $N^{\text{ième}}$ fois.

3. Donner la loi de probabilité de Z .

Options Eco - Techno

[C) Soit λ un nombre réel strictement supérieur à 1.

Un joueur parie de la façon suivante. Lors du $n^{\text{ième}}$ jet, il mise 1 franc.

– Si pile sort, il reçoit la somme λ (en francs), et il perd sa mise ;

– sinon, il perd sa mise.

On désigne par G_n la somme des profits et des pertes (celles-ci étant comptées négativement) du joueur après son $n^{\text{ième}}$ succès (qui survient donc à l'issue du jet ayant pour numéro T_n).

1. Montrer que $G_1 = \lambda - T_1$ et calculer l'espérance de G_1 .

2. Plus généralement, pour tout nombre entier naturel non nul r , exprimer G_r en fonction de T_r et en déduire l'espérance de G_r .

3. Étudier la limite de $E(G_r)$ lorsque r tend vers $+\infty$.

]

Tournez la page S.V.P.

Option Générale

C) Soient a un nombre réel strictement positif et λ un nombre réel strictement supérieur à 1.

Un joueur parie de la façon suivante. Lors du $n^{\text{ième}}$ jet, il mise la somme a^{n-1} (en francs).

- Si pile sort, il reçoit la somme λa^{n-1} et il perd sa mise ;
- sinon, il perd sa mise.

On désigne par G_n la somme des profits et des pertes (celles-ci étant comptées négativement) du joueur après son $n^{\text{ième}}$ succès (qui survient donc à l'issue du jet ayant pour numéro T_n).

1. Dans cette question, on suppose que $a = 1$ (le joueur parie donc un franc à chaque jet).

- a) Exprimer G_1 en fonction de T_1 et calculer l'espérance de G_1 .
- b) Plus généralement, pour tout nombre entier naturel non nul r , exprimer G_r en fonction de T_r et en déduire l'espérance de G_r .

2. Dans cette question, on suppose que $a > 1$.

- a) Exprimer G_1 en fonction de a^{T_1} .

Déterminer a en fonction de λ de telle sorte que G_1 ne dépende pas des valeurs prises par T_1 .

Dans le cas général, étudier l'existence des espérances de a^{T_1} et de G_1 . Lorsque ces espérances existent, les calculer.

- b) Exprimer G_2 en fonction de a^{T_1} et de a^{T_2} . Étudier l'existence et déterminer la valeur de l'espérance de G_2 .

- c) Soit, plus généralement, r un nombre entier naturel non nul. En utilisant la même méthode, étudier l'existence de l'espérance de G_r . Montrer que, si cette espérance existe, alors :

$$E(G_r) = \frac{1}{a-1} \left[1 - \frac{a^r(1-x)^r}{(1-ax)^r} \right] [1 - \lambda(1-x)]$$

- d) En déduire, si elles existent, la limite de $E(G_r)$ lorsque r tend vers $+\infty$ et la limite de $E(G_r)$ lorsque a tend vers 1 par valeurs supérieures.

3. Dans cette question, on suppose que $a < 1$.

- a) Les conditions d'existence de l'espérance de a^{T_r} sont-elles vérifiées ? La formule de la question 2. c) reste-t-elle valable ? En déduire la limite de $E(G_r)$ lorsque r tend vers $+\infty$.

- b) Soit g_k la somme des profits et des pertes réalisés lors du $k^{\text{ième}}$ jet. Exprimer g_k en fonction d'une variable de Bernoulli associée au $k^{\text{ième}}$ jet.

- c) Soit H_m le gain (algébrique) réalisé après m jets. Calculer l'espérance de H_m et la limite de $E(H_m)$ lorsque m tend vers $+\infty$.

}

FIN



CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS
DIRECTION DE L'ENSEIGNEMENT
Direction des Admissions et Concours

ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE SUPÉRIEURE DE COMMERCE DE PARIS
ÉCOLE EUROPÉENNE DES AFFAIRES
ÉCOLE SUPÉRIEURE DE COMMERCE DE LYON

CONCOURS D'ADMISSION DE 1992

Mathématiques II

OPTION GÉNÉRALE

Samedi 9 mai 1992, de 8 h à 12 h

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Sont autorisées:

- Règles graduées.
- Calculatrices de poche, y compris les calculatrices programmables et alphanumériques, à fonctionnement autonome, sans imprimante, sans document d'accompagnement et de format maximum 21 cm de long x 15 cm de large.

Dans tout le problème, on désigne par n , a et b des nombres entiers naturels non nuls.

PARTIE I

On effectue dans une urne contenant initialement b boules blanches et b boules noires une suite de tirages de la façon suivante : si les $n - 1$ premiers tirages ont tous donné une boule blanche (cette condition n'étant bien sûr pas à prendre en compte au premier tirage), on procède au $n^{\text{ième}}$ tirage.

- Si la boule obtenue est noire, ce $n^{\text{ième}}$ tirage est le tirage final ;
- si la boule obtenue est blanche, elle est remplacée dans l'urne avec, en plus, a autres boules blanches (et l'urne est alors prête pour le $(n + 1)^{\text{ième}}$ tirage).

1. Dans cette question, on étudie le cas particulier où $a = b$ et l'on désigne par X la variable aléatoire associant à toute suite de tirages :

- la valeur 0 si, à chaque tirage, une boule blanche a été obtenue ;
- le numéro du tirage final où apparaît une boule noire, sinon.

a) Déterminer la probabilité p_n d'obtenir une boule noire au $n^{\text{ième}}$ tirage, autrement dit, la probabilité p_n pour que $X = n$.

- b) Déterminer des nombres réels α et β tels que, pour tout nombre entier naturel non nul n , on ait :

$$\frac{1}{n(n+1)} = \frac{\alpha}{n} + \frac{\beta}{n+1}$$

- c) Calculer la somme $p_1 + p_2 + \dots + p_n$ et en déduire que :

$$\sum_{n=1}^{+\infty} p_n = 1$$

- d) Étudier l'existence de l'espérance de X .

2. On revient au cas général. On note $P(A_n)$ la probabilité de l'événement :

$A_n =$ "une boule blanche apparaît à chacun des n premiers tirages"

Montrer que :

$$P(A_n) = \prod_{k=0}^{n-1} \frac{ka+b}{ka+2b} = \frac{b}{2b} \cdot \frac{a+b}{a+2b} \cdots \frac{(n-1)a+b}{(n-1)a+2b}$$

3. Dans cette question, on étudie la convergence de la suite $(P(A_n))_{n \geq 1}$.

- a) Soient $x_0, x_1, \dots, x_{n-1}$ des nombres réels positifs. Démontrer que :

$$\prod_{k=0}^{n-1} (1+x_k) \geq 1 + \sum_{k=0}^{n-1} x_k$$

En déduire une minoration de $\frac{1}{P(A_n)}$. En conclure que la limite de $P(A_n)$ lorsque n tend vers $+\infty$ est nulle.

- b) On désigne par X_a la variable aléatoire associant à toute suite de tirages :

- la valeur 0 si, à chaque tirage, une boule blanche a été obtenue ;
- le numéro du tirage final où apparaît une boule noire, sinon.

Exprimer en fonction de $P(A_n)$ la probabilité pour qu'une boule noire soit obtenue à l'un des n premiers tirages. En déduire que, si $P(X_a = k)$ désigne la probabilité pour que $X_a = k$, on a :

$$\sum_{k=1}^{+\infty} P(X_a = k) = 1$$

4. Dans cette question, on désigne par Y une variable aléatoire quelconque à valeurs dans $\mathbb{N}^*$. Pour tout nombre entier naturel non nul n , on pose $p_n = P(Y = n)$.

- a) Exprimer en fonction des termes de la suite (p_n) le nombre réel :

$$E_n = \sum_{k=0}^n P(Y > k)$$

- b) On suppose que Y admet une espérance $E(Y)$. Préciser le sens de variation de la suite (E_n) , prouver que $E_n \leq E(Y)$. En déduire que la suite (E_n) converge.

- c) On suppose que la suite (E_n) converge. Prouver que :

$$p_1 + 2p_2 + \dots + np_n \leq E_n$$

En déduire que Y admet une espérance.

- d) Sous l'une de ces hypothèses équivalentes, établir que :

$$(1) \quad E(Y) = \sum_{n=0}^{+\infty} P(Y > n)$$

À l'aide de la relation (1), on se propose d'étudier dans la fin de cette partie l'espérance de la variable aléatoire X_a définie dans la question 3. Pour tout nombre entier naturel n , on note $r_n(a)$ la probabilité pour que $X_a > n$.

5. On suppose dans cette question que $a \geq b$.

a) Soient a_1 et a_2 des nombres entiers naturels non nuls tels que $a_1 \leq a_2$. Montrer que, pour tout nombre entier naturel n :

$$r_n(a_1) \leq r_n(a_2)$$

En déduire la nature de la série de terme général $r_n(a)$ pour $a \geq b$.

b) Étudier l'existence de l'espérance de X_a pour $a \geq b$.

6. On suppose dans cette question que $1 \leq a < b$.

a) Exprimer $r_{n+1}(a)$ en fonction de $r_n(a)$. Établir que la suite $(nr_n(a))$ est décroissante à partir d'un certain rang ; en déduire qu'elle est convergente. On note ℓ sa limite.

b) Établir que si $\ell \neq 0$, alors $r_n(a) \geq \frac{\ell}{n}$ à partir d'un certain rang. En déduire que si $\ell \neq 0$, alors X_a n'admet pas d'espérance.

c) Montrer que, pour tout nombre entier naturel k :

$$(2) \quad a(k+1) r_{k+1}(a) - a k r_k(a) = (a-2b) r_{k+1}(a) + b r_k(a)$$

d) En sommant les relations (2) pour k variant de 0 à n , puis en faisant tendre n vers $+\infty$, montrer que X_a admet une espérance. En déduire que $\ell = 0$, puis que :

$$E(X_a) = \frac{2b-a}{b-a}$$

PARTIE II

Dans cette partie, on considère une suite croissante (u_n) de nombres entiers naturels telle que $u_0 = b$. On généralise la situation étudiée dans la partie I.

On effectue dans une urne contenant initialement b boules blanches et b boules noires une suite de tirages de la façon suivante : si les $n-1$ premiers tirages ont tous donné une boule blanche (cette condition n'étant bien sûr pas à prendre en compte au premier tirage), on procède au $n^{\text{ième}}$ tirage.

- Si la boule obtenue est noire, ce $n^{\text{ième}}$ tirage est le tirage final ;
- si la boule obtenue est blanche, elle est remplacée dans l'urne avec, en plus, $u_n - u_{n-1}$ autres boules blanches (et l'urne est alors prête pour le $(n+1)^{\text{ième}}$ tirage).

Ainsi, l'urne contient b boules noires et u_n boules blanches au moment où l'on procède au $(n+1)^{\text{ième}}$ tirage, dans la mesure où celui-ci a lieu.

1. Dans cette question, on étudie la probabilité d'obtention d'une boule noire.

a) Exprimer en fonction de b et des nombres u_k , où $1 \leq k < n$, la probabilité q_n de l'événement :
 $A_n = \text{"une boule blanche apparaît à chacun des } n \text{ premiers tirages"}$

b) Étudier le sens de variation de la suite (q_n) . En déduire que la suite (q_n) converge vers un nombre réel L appartenant à l'intervalle $\left[0, \frac{1}{2}\right]$.

On ne cherchera pas à expliciter ce nombre réel.

c) Soit p_n la probabilité d'obtenir une boule noire au $n^{\text{ième}}$ tirage. Exprimer la somme $p_1 + p_2 + \dots + p_n$ en fonction de q_n . En déduire que :

$$\sum_{n=1}^{+\infty} p_n = 1 - L$$

d) Comparer les natures de la suite de terme général $\ln\left(\frac{1}{q_n}\right)$ et de la série de terme général $\frac{1}{u_n}$. En déduire que la probabilité de ne pas obtenir de boule noire dans la suite des tirages est nulle si et seulement si cette dernière série diverge.

e) Décrire le mode de tirage et étudier l'éventualité de ne pas obtenir de boule noire dans la suite des tirages, dans les deux cas suivants :

- la suite (u_n) est une suite arithmétique, définie par $u_n = b + na$;
- la suite (u_n) est une suite géométrique, définie par $u_n = ba^n$ (avec $a \geq 2$).

2. Dans cette question, on suppose que la suite (u_n) est définie par $u_n = ba^n$ (avec $a \geq 2$).

a) Préciser l'expression de q_n (que l'on notera $q_n(a)$ dans cette question).

On ne cherchera pas à expliciter la limite $L(a)$ de la suite $(q_n(a))$.

b) On étudie la vitesse de convergence de la suite $(q_n(a))_{n \geq 1}$.

Soient n et p des nombres entiers naturels non nuls. Établir que :

$$\sum_{k=n}^{n+p-1} \frac{1}{a^k} \leq \frac{1}{a^{n-1}(a-1)}$$

Établir, pour tout nombre réel x appartenant à $[0, 1]$, l'inégalité :

$$1 + x \leq e^x \leq 1 + 2x$$

Déduire de ces résultats l'inégalité suivante :

$$1 \leq \frac{q_n(a)}{q_{n+p}(a)} \leq \exp\left(\frac{1}{a^{n-1}(a-1)}\right)$$

En faisant tendre p vers $+\infty$, en conclure que :

$$(3) \quad 0 \leq q_n(a) - L(a) \leq \frac{1}{(a-1)a^{n-1}}$$

c) On donne un nombre réel strictement positif ε . Écrire en langage PASCAL un algorithme calculant les valeurs de $q_n(a)$ tant que :

$$\frac{1}{(a-1)a^{n-1}} < \varepsilon$$

À l'aide de cet algorithme, donner une valeur approchée de $L(2)$ à 10^{-4} près.

d) On étudie enfin la suite associant à tout nombre entier $a \geq 2$ le nombre réel $L(a)$.

Soient a_1 et a_2 des nombres entiers naturels non nuls tels que $a_1 \leq a_2$. Montrer que, pour tout nombre entier naturel n :

$$q_n(a_1) \leq q_n(a_2)$$

En déduire que la suite $(L(a))_{a \geq 1}$ est croissante. En remplaçant n par 1 dans l'inégalité (3), déterminer la limite de $L(a)$ lorsque l'entier a tend vers $+\infty$.



CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS
DIRECTION DE L'ENSEIGNEMENT
Direction des Admissions et Concours

ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE SUPÉRIEURE DE COMMERCE DE PARIS
ÉCOLE EUROPÉENNE DES AFFAIRES
ÉCOLE SUPÉRIEURE DE COMMERCE DE LYON

CONCOURS D'ADMISSION DE 1993

Mathématiques II

OPTION GÉNÉRALE

lundi 17 mai 1993, de 8 h à 12 h

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Sont autorisées:

- Règles graduées.
- Calculatrices de poche, y compris les calculatrices programmables et alphanumériques, à fonctionnement autonome, sans imprimante, sans document d'accompagnement et de format maximum 21 cm de long x 15 cm de large.

Partie I

1. a) Calculer $\lim_{x \rightarrow 0^+} \frac{\ln(1-x)}{x}$

b) En déduire que : $e^{-1} = \lim_{n \rightarrow +\infty} \left(1 - \frac{1}{n}\right)^n$

2. a) Montrer que, pour tout nombre entier naturel n :

$$e^{-1} = \sum_{k=0}^n \frac{(-1)^k}{k!} + (-1)^{n+1} \int_0^1 \frac{(1-u)^n}{n!} e^{-u} du$$

b) Établir l'encadrement : $0 \leq \int_0^1 \frac{(1-u)^n}{n!} e^{-u} du \leq \frac{1}{n!}$

c) En conclure que :

$$e^{-1} = \lim_{n \rightarrow +\infty} \sum_{k=0}^n \frac{(-1)^k}{k!}$$

Dans toute la suite du problème, on désigne par N un nombre entier naturel non nul.

Partie II

On considère une urne contenant N boules numérotées de 1 à N : $B_1, B_2, \dots, B_N$. On effectue N tirages *avec remise*, en supposant l'équiprobabilité des résultats. Pour tout entier naturel i compris entre 1 et N , on note X_i la variable aléatoire prenant la valeur 1 si la boule B_i sort lors du i ème tirage et la valeur 0 dans le cas contraire. On pose :

$$S_N = X_1 + X_2 + \dots + X_N$$

1. Déterminer l'espérance de X_i ; en déduire celle de S_N .
2. Pour tout nombre entier naturel k , on note $p(N, k)$ la probabilité de l'événement $S_N = k$.

a) Calculer $p(N, k)$ lorsque $k > N$. Montrer que :

$$p(N, k) = C_N^k \left(1 - \frac{1}{N}\right)^{N-k} \frac{1}{N^k} \quad \text{si } 0 \leq k \leq N$$

b) Le nombre entier naturel k étant fixé, déterminer la limite de $p(N, k)$ lorsque N tend vers $+\infty$.

Partie III

On considère un ensemble E_N à N éléments. Soit s une permutation de E_N , c'est-à-dire une bijection de E_N sur lui-même. On appelle point fixe de s tout élément a de E_N tel que $s(a) = a$.

Pour tout nombre entier p tel que $0 \leq p \leq N$, on note $F(N, p)$ le nombre de permutations de E_N qui ont exactement p point fixes. On convient que $F(0, 0) = 1$.

1. a) Montrer que $F(N, N) = 1$ et $F(N, N-1) = 0$

b) Montrer que $\sum_{k=0}^N F(N, k) = N!$

2. On pose $\omega_N = F(N, 0)$. Ainsi, ω_N est le nombre de permutations de E_N qui n'ont aucun point fixe. On convient que $\omega_0 = 1$.

a) Montrer que, pour tout nombre entier k tel que $0 \leq k \leq N$:

$$F(N, k) = C_N^k \omega_{N-k}$$

b) En déduire que :

$$\sum_{k=0}^N \frac{1}{k!} \frac{\omega_{N-k}}{(N-k)!} = 1$$

c) En raisonnant par récurrence, établir la relation :

$$\frac{\omega_N}{N!} = \sum_{k=0}^N \frac{(-1)^k}{k!}$$

d) Déterminer la limite de $\frac{\omega_N}{N!}$ lorsque N tend vers $+\infty$.

On suppose désormais $N \geq 2$.

Partie IV

On considère à nouveau une urne contenant N boules numérotées de 1 à N : $B_1, B_2, \dots, B_N$. On effectue N tirages aléatoires, cette fois *sans remise*. Pour tout entier naturel i compris entre 1 et N , on note Y_i la variable aléatoire prenant la valeur 1 si la boule B_i sort lors du $i^{\text{ème}}$ tirage et la valeur 0 dans le cas contraire.

1. a) Déterminer l'espérance de Y_i et celle de $Y_i Y_j$, où $1 \leq i < j \leq N$.

b) Déterminer la covariance de Y_i et Y_j , où $1 \leq i < j \leq N$.

c) Les variables aléatoires Y_i et Y_j sont-elles indépendantes ?

2. On pose $T_N = Y_1 + Y_2 + \dots + Y_N$

Déterminer l'espérance et la variance de T_N .

3. Pour tout nombre entier naturel k , on note $q(N, k)$ la probabilité de l'événement $T_N = k$.

a) Calculer $q(N, k)$ lorsque $k > N$. Montrer que :

$$q(N, k) = \frac{\omega_{N-k}}{k! (N-k)!} \quad \text{si } k \leq N$$

b) Le nombre entier naturel k étant fixé, déterminer la limite de $q(N, k)$ lorsque N tend vers $+\infty$. Comparer ce résultat à celui de la question II.2 b).

Partie V

1. À l'aide de la relation établie dans la question III.2 c), montrer que, pour tout nombre entier naturel $n \geq 2$:

$$\omega_n = (n-1)(\omega_{n-1} + \omega_{n-2})$$

2. Pour tout nombre entier k tel que $0 \leq k \leq N$, on pose :

$$\bar{q}(N, k) = q(N, N - k)$$

a) Calculer $\bar{q}(N, 0)$ et $\bar{q}(N, 1)$.

b) Pour $2 \leq k \leq N$, exprimer $\bar{q}(N, k)$ en fonction de $\bar{q}(N, k-1)$ et de $\bar{q}(N, k-2)$.

3. Écrire un algorithme prenant N comme donnée et construisant la liste des valeurs de $\bar{q}(N, k)$, k variant de 0 à N . On s'attachera à minimiser le nombre d'opérations.



CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS
DIRECTION DE L'ENSEIGNEMENT
Direction des Admissions et Concours

ECOLE DES HAUTES ETUDES COMMERCIALES
ECOLE SUPERIEURE DE COMMERCE DE PARIS
ECOLE SUPERIEURE DE COMMERCE DE LYON

CONCOURS D'ADMISSION SUR CLASSES PREPARATOIRES

MATHEMATIQUES II
OPTION GENERALE

mardi 17 mai 1994, de 8 h à 12 h

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Sont autorisées :

- . Règles graduées.
- . Calculatrices de poche, y compris les calculatrices programmables et alphanumériques, à fonctionnement autonome, sans imprimante, sans document d'accompagnement et de format maximum 21 cm de long x 15 cm de large.

L'objet du problème est l'étude d'une méthode de test de prélèvements sanguins dite du poolage ; elle est présentée dans la partie II, où on l'étudie d'un point de vue probabiliste. Dans tout le problème, a désigne un nombre réel strictement positif. Dans la partie I, on étudie, à titre préliminaire, la fonction f_a définie sur $]0, +\infty[$ par la relation :

$$f_a(x) = \frac{1}{x} - e^{-ax}$$

PARTIE I. Étude de f_a

1. Soit g la fonction définie sur $]0, +\infty[$ par la relation :

$$g(x) = \frac{\ln x}{x}$$

On considère l'équation $(E_a) : g(x) = a$ (où l'inconnue x appartient à $]0, +\infty[$).

a) Étudier la variation de la fonction g . (On dressera le tableau de variation et on tracera la représentation graphique de g .)

b) On suppose que : $0 < a < \frac{1}{e}$.

Montrer que (E_a) admet exactement deux solutions ; on les note $u(a)$ et $v(a)$, en convenant que $u(a) < v(a)$.

Établir que : $1 < u(a) < e < v(a)$.

c) Discuter suivant les valeurs de a le nombre de solutions de (E_a) .

2. Soit h_a la fonction définie sur $]0, +\infty[$ par la relation :

$$h_a(x) = 2 \ln x + \ln a - ax$$

On considère l'équation $(F_a) : h_a(x) = 0$ (où l'inconnue x appartient à $]0, +\infty[$).

a) Étudier la variation de la fonction h_a . (On ne demande pas la représentation graphique de h_a .)

b) On suppose que : $0 < a < \frac{4}{e^2}$.

Montrer que (F_a) admet exactement deux solutions ; on les note $r(a)$ et $s(a)$, en convenant que $r(a) < s(a)$.

Établir que : $0 < r(a) < \frac{2}{a} < s(a)$.

c) Discuter suivant les valeurs de a le nombre de solutions de (F_a) .

3. Soient a et b des nombres réels tels que $0 < a < b$ et x élément de $]0, +\infty[$. Montrer que :

$$\frac{1}{x} - 1 < f_a(x) < f_b(x) < \frac{1}{x}$$

4. Comportement asymptotique de f_a

a) Calculer $\lim_{x \rightarrow 0} f_a(x)$.

b) Calculer $\lim_{x \rightarrow +\infty} f_a(x)$.

5. Signe de f_a

a) Comparer les signes de $f_a(x)$ et de $a - g(x)$.

b) En déduire le tableau de signes de $f_a(x)$ lorsque x décrit $]0, +\infty[$, a étant fixé. (On sera amené à distinguer trois cas suivant la position de a par rapport à $1/e$.)

6. Variation de f_a

a) Comparer les signes de $f'_a(x)$ et de $h_a(x)$.

b) Dresser le tableau de variation de f_a . On distinguera deux cas :

$$a \geq \frac{4}{e^2} \quad \text{et} \quad 0 < a < \frac{4}{e^2}$$

Dans ce dernier cas, on ne cherchera pas à préciser les valeurs de $f_a(r(a))$ et de $f_a(s(a))$.

7. On suppose dans cette question que $0 < a < \frac{1}{e}$.

a) Établir que $u(a) < r(a) < v(a) < s(a)$ et que f_a présente un minimum en $r(a)$.

b) Donner l'allure du graphe de f_a .

8. On suppose encore que $0 < a < \frac{1}{e}$.

On pose $m(a) = f_a(r(a))$.

a) Établir que : $r(a) = \frac{e^{ar(a)/2}}{\sqrt{a}}$.

b) Déterminer $\lim_{a \rightarrow 0} \sqrt{a} r(a)$.

On pourra écrire : $ar(a) = \sqrt{a} e^{ar(a)/2}$; on utilisera alors la question 2. b) pour obtenir la limite de $ar(a)$ lorsque a tend vers 0.

c) Déterminer $\lim_{a \rightarrow 0} m(a)$.

d) Calculer un équivalent simple de $m(a) + 1$ lorsque a tend vers 0.

PARTIE II. Étude du poolage

On étudie dans cette partie une méthode de détection des porteurs d'un parasite au sein d'un ensemble donné de N individus tirés au sort de façons indépendantes dans une population très vaste par rapport à N . La proportion de porteurs du parasite dans la population est p ($0 < p < 1$).

On dispose d'un test permettant d'établir de façon certaine qu'un échantillon de sang contient ou non le parasite, le résultat de ce test étant dit positif dans le premier cas et négatif dans le second.

Pour chacun des N individus, on possède un prélèvement sanguin. On envisage alors deux méthodes de détection.

Première méthode : on teste un à un les N prélèvements, effectuant ainsi N tests.

Seconde méthode (poolage) : on fixe un entier naturel non nul l . On suppose que N est un multiple de l et on pose $N = n l$. On répartit les N prélèvements en n groupes $G_1, G_2, \dots, G_n$, chaque groupe G_i contenant l prélèvements. Pour chacun des groupes G_i , on extrait une quantité de sang de chacun des l prélèvements qu'il contient, puis on mélange ces extraits, obtenant ainsi un échantillon de sang H_i , caractéristique du groupe G_i .

On teste alors H_i :

— si le test de H_i est négatif, aucun des individus au sein du groupe G_i n'est porteur du parasite. Le travail sur le groupe G_i est alors terminé ;

— si le test de H_i est positif, on teste un à un les prélèvements de G_i pour détecter les porteurs du parasite au sein du groupe G_i .

Soient X la variable aléatoire égale au nombre de groupes G_i pour lesquels le test de H_i a été positif et T la variable aléatoire égale au nombre total de tests effectués dans la réalisation de la méthode du poolage.

1. a) Exprimer T à l'aide de n , l et X .

b) Pour tout nombre entier naturel i compris entre 1 et n , calculer la probabilité de l'événement : « le test de H_i est négatif ».

c) Déterminer la loi de probabilité et l'espérance de X .

On pose désormais et pour toute la fin du problème : $a = -\ln(1-p)$.

d) Montrer que $E(T) = N(1 + f_a(l))$.

On suppose en outre, dans toute la fin du problème, que : $0 < p < 1 - \frac{1}{3^{1/3}}$.

2. a) Montrer que $f_a(3) < 0$. Comparer les deux méthodes pour $l = 3$.

b) Établir que : $a < \frac{1}{e}$.

On cherche maintenant à optimiser la méthode du poolage, c'est-à-dire choisir, en fonction de p , la valeur de l qui minimise $E(T)$.

3. Soit l un nombre entier naturel non nul. On dit que l vérifie la propriété (MIN) si, pour tout nombre entier naturel non nul l' , $f_a(l) \leq f_a(l')$.

a) Montrer qu'il existe au moins un entier naturel non nul l qui vérifie la propriété (MIN) et qu'un tel entier est égal soit à $[r(a)]$, soit à $[r(a)] + 1$, où $[r(a)]$ désigne la partie entière de $r(a)$.

On note désormais l_0 le plus petit des entiers naturels non nuls l qui vérifient la propriété (MIN).

b) Montrer que $f_a(l_0) < 0$. En déduire que $l_0 \geq 2$.

c) Montrer que $f_a(3) < f_a(2)$. Que peut-on en déduire pour l_0 ?

4. Proposer un algorithme, qu'on pourra écrire en TURBO-PASCAL, qui prend p en donnée et fournit la valeur de l_0 .

5. *Exemple* : on suppose $p = 0,01$.

a) Déterminer la valeur de l_0 .

b) Déterminer le rapport $\frac{E(T)}{N}$ lorsque $l = l_0$.

6. On note $\rho(p)$ la valeur du rapport $\frac{E(T)}{N}$ lorsque $l = l_0$.

a) Donner un équivalent de l_0 lorsque p tend vers 0.

b) Donner un équivalent de $\rho(p)$ lorsque p tend vers 0.

c) Le résultat du 5. b) est-il conforme à celui de 6. b) ?



CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS
DIRECTION DE L'ENSEIGNEMENT
Direction des Admissions et Concours

ECOLE DES HAUTES ETUDES COMMERCIALES
ECOLE SUPERIEURE DE COMMERCE DE PARIS
ECOLE SUPERIEURE DE COMMERCE DE LYON

CONCOURS D'ADMISSION SUR CLASSES PREPARATOIRES

MATHEMATIQUES II
OPTION GENERALE

vendredi 19 mai 1995, de 8h à 12h

*La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.*

Sont autorisées: – Règles graduées.
Calculatrices de poche, y compris les calculatrices programmables et alphanumériques, à fonctionnement autonome, sans imprimante, sans document d'accompagnement et de format maximum 21 cm x 15 cm de large, sans limitation de nombre.

Ce problème est consacré à l'étude de la loi de Pareto (Vilfredo Pareto 1848-1923).

La première partie étudie les propriétés de cette loi. On montre ensuite, sur un exemple, comment elle permet de modéliser de façon très satisfaisante des phénomènes rencontrés en économie.

La seconde partie est l'étude de l'indice d'inégalité de Gini, d'abord dans un cadre général, puis dans le cas particulier d'une loi de Pareto.

Toutes les variables aléatoires considérées dans ce problème sont à valeurs réelles.

PARTIE I. La loi de Pareto.

Dans tout le problème x_0 , C et α désignent trois nombres réels vérifiant $\alpha > 0$ et $x_0 + C > 0$.

On dit qu'une variable aléatoire réelle X suit une loi de Pareto de paramètres α , x_0 et C si X , à valeurs dans $[x_0, +\infty[$, admet pour densité la fonction f définie par :

$$\begin{cases} f(x) = 0 & \forall x < x_0 \\ f(x) = \alpha \frac{(x_0 + C)^\alpha}{(x + C)^{\alpha+1}} & \forall x \geq x_0 \end{cases}$$

On dit alors que X suit $VP(\alpha, x_0, C)$ (loi de Pareto à trois paramètres).

Lorsque $C = 0$, on dit que X suit $VP(\alpha, x_0)$ (loi de Pareto à deux paramètres) au lieu de $VP(\alpha, x_0, 0)$. Dans ce cas x_0 est strictement positif et X admet pour densité la fonction définie par :

$$\begin{cases} f(x) = 0 & \forall x < x_0 \\ f(x) = \alpha \frac{x_0^\alpha}{x^{\alpha+1}} & \forall x \geq x_0 \end{cases}$$

A. Quelques résultats probabilistes.

1. Soit X une variable aléatoire suivant $VP(\alpha, x_0, C)$.
 - a) Vérifier que f est bien une fonction de densité.
 - b) Déterminer la fonction de répartition de X
2. Soit X une variable aléatoire suivant $VP(\alpha, x_0)$.
 - a) Déterminer les valeurs de α pour lesquelles X admet une espérance et la calculer dans ce cas.
 - b) Déterminer les valeurs de α pour lesquelles X admet une variance et la calculer dans ce cas.
3. Soient une variable aléatoire X suivant $VP(\alpha, x_0)$ et un réel strictement positif λ . Déterminer la fonction de répartition de la variable aléatoire $Y = \lambda X$. Quelle loi reconnaît-on ?
4. a) Soient une variable aléatoire X suivant $VP(\alpha, x_0)$ et un réel μ . Montrer que la variable aléatoire $U = X + \mu$ suit une loi de Pareto à trois paramètres que l'on déterminera.
b) Réciproquement, soit une variable aléatoire Z suivant $VP(\alpha, x_0, C)$.
Quelle est la loi de la variable aléatoire $V = Z + C$?
c) Soit une variable aléatoire Z suivant $VP(\alpha, x_0, C)$.
Déduire des questions 2 et 4.b les valeurs de α pour lesquelles Z admet une espérance et la déterminer dans ce cas.
Déterminer les valeurs de α pour lesquelles Z admet une variance et la calculer dans ce cas.
5. Soit une variable aléatoire W qui suit une loi exponentielle de paramètre $\beta > 0$ c'est à dire que W admet une densité de probabilité g définie par :

$$\begin{cases} g(x) = 0 & \forall x < 0 \\ g(x) = \beta e^{-\beta x} & \forall x \geq 0 \end{cases}$$

Soient k un réel strictement supérieur à 1 et x_0 un réel strictement positif.

Déterminer la fonction de répartition de la variable aléatoire $T = x_0 k^W$? Quelle loi reconnaît-on ?

6. Soit une variable aléatoire X suivant $VP(\alpha, x_0)$. Déterminer la loi de la variable aléatoire $\sqrt{X}$.

B. Propriété caractéristique de la loi de Pareto.

Soient une variable aléatoire X de densité f , admettant une espérance, et un nombre réel x tel que $\int_x^{+\infty} f(t)dt \neq 0$.

On appelle moyenne de X sur $[x, +\infty[$ le nombre réel $M_X(x)$ égal à
$$\frac{\int_x^{+\infty} t f(t) dt}{\int_x^{+\infty} f(t) dt}.$$

1. Montrer que $M_X(x) \geq x$.
2. Dans cette question on suppose que la variable X suit $VP(\alpha, x_0)$ avec $\alpha > 1$. Calculer $M_X(x)$ pour $x \geq x_0$.
3. Réciproquement soient un réel x_0 strictement positif et une variable aléatoire X à valeurs dans $[x_0, +\infty[$, de densité f continue et à valeurs strictement positives sur $[x_0, +\infty[$, admettant une espérance et telle qu'il existe un réel $k > 1$ tel que :

$$\forall x \geq x_0 \quad M_X(x) = kx.$$

On se propose d'établir que X suit une loi de Pareto à deux paramètres.

On pose, pour $x \geq x_0$, $G(x) = \int_x^{+\infty} f(t)dt$ et $H(x) = \int_x^{+\infty} t f(t)dt$.

a) Montrer que G et H sont dérivables sur $[x_0, +\infty[$ et calculer leur dérivée.

b) Prouver que $G(x) = \frac{1-k}{k} x G'(x)$ pour $x \geq x_0$.

c) Pour $x \geq x_0$, on pose $I(x) = x^{\frac{k}{k-1}} G(x)$. Calculer $I'(x)$ et en déduire la valeur de $G(x)$.

d) En déduire que X suit $VP(\frac{k}{k-1}, x_0)$.

4. Soient un nombre réel x_0 et une variable aléatoire X à valeurs dans $[x_0, +\infty[$, de densité f continue et à valeurs strictement positives sur $[x_0, +\infty[$, admettant une espérance.

a) Soient un réel λ et la variable aléatoire $Y = X + \lambda$.

Prouver que : $M_Y(y) = M_X(y - \lambda) + \lambda \quad \forall y \geq x_0 + \lambda$.

b) On suppose dans cette question que X suit $VP(\alpha, x_0, C)$ avec $\alpha > 1$.

Exprimer $M_X(x)$ en fonction de x pour $x \geq x_0$.

c) Réciproquement on suppose qu'il existe deux réels h et k vérifiant $k > 1$ et $x_0 + \frac{h}{k-1} > 0$ tels que :

$M_X(x) = kx + h$ pour tout réel $x \geq x_0$.

Soit la variable aléatoire $Y = X + \frac{h}{k-1}$. Calculer $M_Y(y)$ pour $y \geq x_0 + \frac{h}{k-1}$.

En déduire la loi de Y .

Montrer enfin que la variable X suit $VP(\frac{k}{k-1}, x_0, \frac{h}{k-1})$.

C. Un exemple statistique : la répartition des revenus.

Des statistiques provenant de la Direction Générale des Impôts indiquent, pour l'année 1988, la répartition des revenus d'environ 25 000 000 de contribuables. On note :

x : Niveau de revenu en KF (milliers de francs),

$1 - F(x)$: Proportion de contribuables ayant un revenu strictement supérieur à x ,

$M(x)$: Revenu moyen des contribuables ayant un revenu strictement supérieur à x ,

et on dispose du tableau suivant :

x	$1 - F(x)$	$M(x)$
500	0.006	860
250	0.027	434
200	0.046	345
150	0.095	256
125	0.145	215
100	0.226	177
80	0.323	151
60	0.458	127
40	0.658	104

Ces données sont représentées dans les graphiques 1 et 2 figurant sur la page 6 de l'énoncé. Ces graphiques ne sont pas à reproduire sur la copie.

- Le graphique 1 représente les points d'abscisse x et d'ordonnée $M(x)$ pour les valeurs x du tableau.
Une étude statistique permet d'estimer que ce nuage de points peut être modélisé par une droite D (figurant sur le graphique).
 - Lire, sur le graphique, le coefficient directeur de D .
 - Expliquer pourquoi on peut modéliser la distribution des revenus par une loi de Pareto à trois paramètres $VP(\alpha, x_0, C)$.
 - Donner, d'après le graphique, une valeur approchée de α et de C .
 - Sachant que le revenu moyen de tous les contribuables est $\bar{M} = 75\text{KF}$, donner, à l'aide du graphique, une valeur approchée de x_0 .
- Le graphique 2 représente les points d'abscisse $\ln(x + C)$ et d'ordonnée $\ln(1000(1 - F(x)))$ pour les valeurs x du tableau.
(On rappelle que $\ln$ désigne la fonction logarithme népérien).
Une étude statistique permet d'estimer que ce nuage de points peut être modélisé par une droite Δ (figurant sur le graphique).
 - Lire, sur le graphique, le coefficient directeur de Δ .
 - Expliquer pourquoi cela confirme la modélisation de la distribution des revenus par une loi de Pareto à trois paramètres $VP(\alpha, x_0, C)$.
 - Retrouver ainsi une valeur approchée de α et de x_0 .

PARTIE II. Courbe de concentration et inégalité des revenus.

La courbe de concentration est une courbe statistique introduite par Lorentz et développée par Gini pour rendre compte de l'inégalité de la distribution des revenus.

On désigne par F_x la proportion des individus d'une population donnée ayant un revenu inférieur ou égal à x et par Q_x le quotient de la masse des revenus de ces mêmes individus par la masse totale des revenus de la population.

On appelle courbe de concentration la représentation graphique de la fonction donnant Q_x en fonction de F_x .

Dans toute la suite du problème, X est une variable aléatoire qui représente le revenu d'un individu de cette population.

Dans la partie A, on montre, dans le cas général, l'existence de la courbe de concentration et on étudie ses propriétés.

Dans la partie B, on étudie le cas particulier où la variable aléatoire X suit une loi de Pareto.

A. Courbe de concentration et indice de Gini.

On désigne par x_0 un nombre réel strictement positif et on suppose que la variable aléatoire X , à valeurs dans $[x_0, +\infty[$, admet une densité f , continue et à valeurs strictement positives sur $[x_0, +\infty[$, et possède une espérance $E(X)$.

On pose, pour $x \geq x_0$, $F(x) = \int_{x_0}^x f(t)dt$ et $Q(x) = \frac{1}{E(X)} \int_{x_0}^x tf(t)dt$.

1. a) Montrer que la restriction de F à $[x_0, +\infty[$ admet une application réciproque notée F^{-1} .
Quel est le domaine de définition de F^{-1} ?

b) On pose $C = Q \circ F^{-1}$.

Montrer que C se prolonge en une fonction continue strictement croissante de $[0, 1]$ dans lui-même.

Déterminer $C(0)$ et $C(1)$.

Ainsi la courbe de concentration de X existe bien. C'est la courbe représentative de C dans un repère orthonormé du plan.

2. a) Prouver que C est dérivable sur $[0, 1[$ et que $C'(t) = \frac{F^{-1}(t)}{E(X)}$ pour $t \in [0, 1[$.

b) En déduire sans calcul que C est convexe et que la courbe de concentration de X est située en dessous de la première bissectrice.

c) Déterminer $C'(0)$ et $\lim_{t \rightarrow 1} C'(t)$.

d) Tracer l'allure de la courbe de concentration en précisant les tangentes aux deux extrémités.

On appelle indice d'inégalité de Gini de la variable X le réel $I(X)$ qui est égal à deux fois l'aire située entre la courbe de concentration de X et la première bissectrice.

C'est à dire : $I(X) = 2 \int_0^1 (t - C(t))dt$.

On estime que plus $I(X)$ est grand, plus l'inégalité des revenus est grande.

B. Application : comparaison de quelques procédures d'imposition des revenus.

On suppose, dans toute cette partie, que la variable aléatoire X suit $VP(\alpha, x_0)$ avec $\alpha > 1$.

1. a) Calculer $Q(x)$ pour $x \geq x_0$.

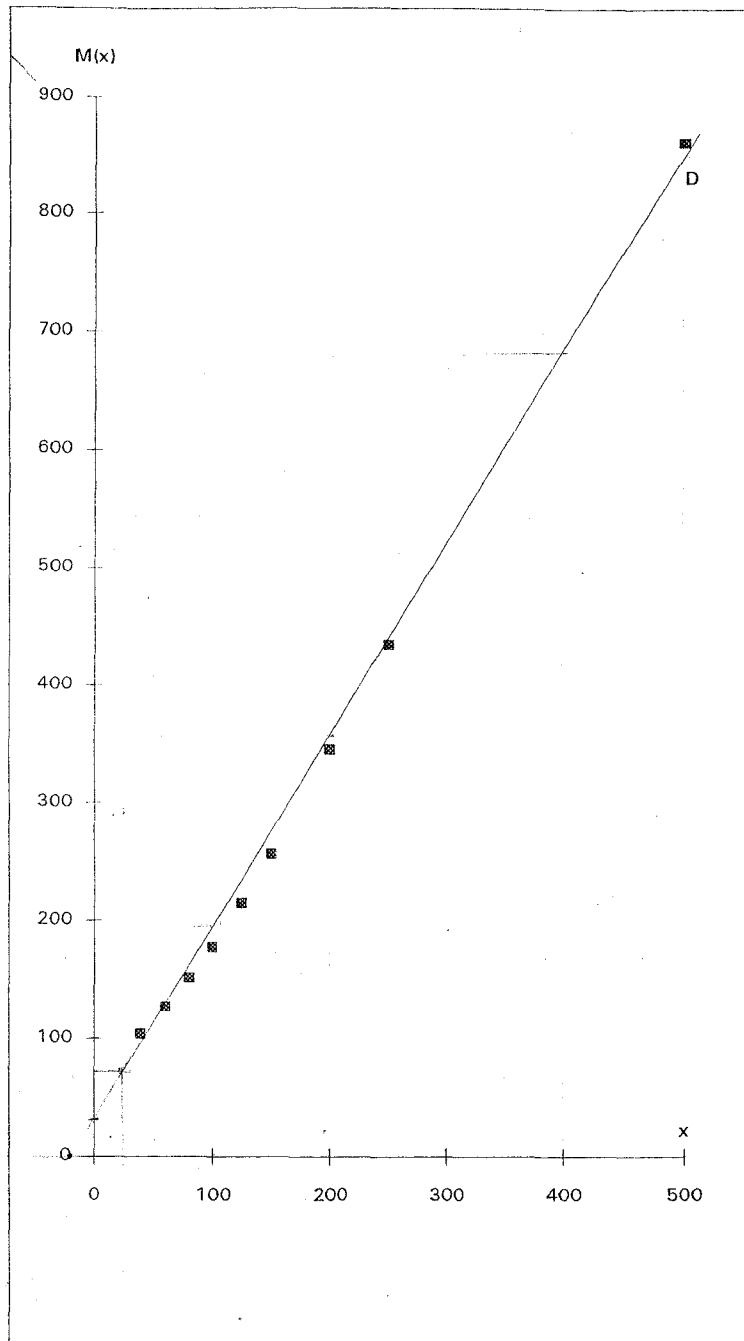
b) Prouver que $C(t) = 1 - (1 - t)^{\frac{\alpha-1}{\alpha}}$ $\forall t \in [0, 1[$.

c) Déterminer α si on sait que 30% des individus ayant les plus hauts revenus se partagent 60% de la masse des revenus. Dans la suite α n'est plus supposé égal à cette valeur.

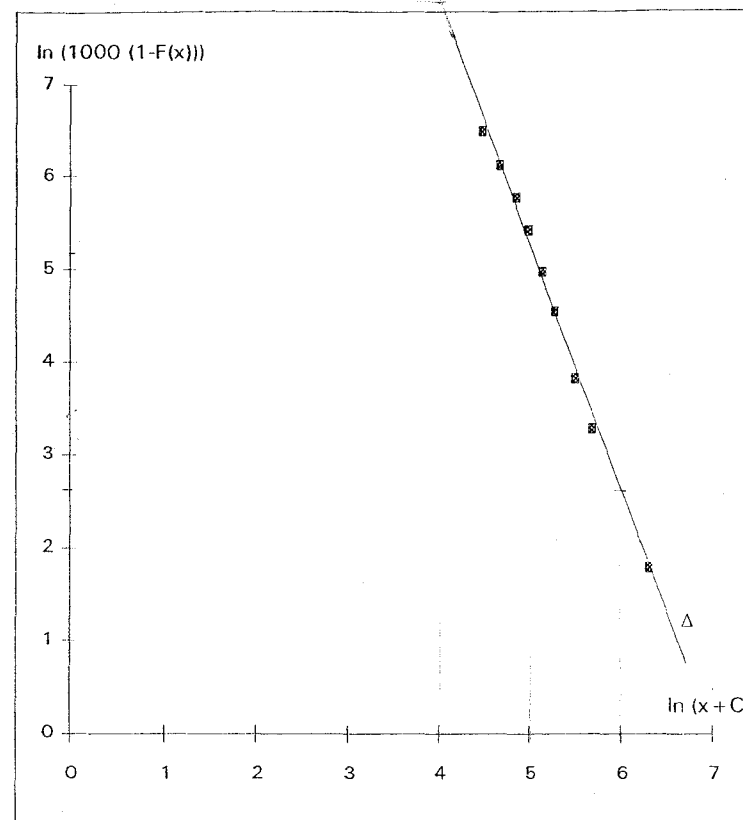
d) Déterminer l'indice d'inégalité de Gini $I(X)$.

2. On prélève sur tous les revenus un impôt proportionnel au revenu c'est à dire que, si Y désigne le revenu disponible après imposition, $Y = (1 - \lambda)X$ avec $\lambda \in]0, 1[$.
 - a) Calculer $I(Y)$. On pourra utiliser les résultats de I.A.3.
 - b) Quel est l'effet de cette imposition sur l'inégalité des revenus ?
3. On prélève sur tous les revenus un impôt progressif tel que, si T désigne le revenu disponible après imposition, on ait $T = h\sqrt{X}$ (h désignant un réel positif).
 - a) À l'aide de la partie I, déterminer la loi de T et calculer $I(T)$.
 - b) Quel est l'effet de cette imposition sur l'inégalité des revenus ?
4. On prélève sur tous les revenus un impôt constant a , c'est à dire que, si Z désigne le revenu disponible après imposition, $Z = X - a$ avec $a \in]0, x_0[$.
 - a) Déterminer la loi de Z .
 - b) On note C_Z la fonction dont le graphe est la courbe de concentration de Z . Montrer que :

$$t - C_Z(t) = \frac{E(X)}{E(X) - a} (t - C(t)) \quad \forall t \in [0, 1]$$
 - c) En déduire $I(Z)$.
 - d) Quel est l'effet de cette imposition sur l'inégalité des revenus ?



Revenus 1988 - Graphique n°1



Revenus 1988 - Graphique n°2

**ECOLE DES HAUTES ETUDES COMMERCIALES
ECOLE SUPERIEURE DE COMMERCE DE PARIS
ECOLE SUPERIEURE DE COMMERCE DE LYON**

CONCOURS D'ADMISSION SUR CLASSES PREPARATOIRES

MATHEMATIQUES II

OPTIONS SCIENTIFIQUE

mardi 14 mai 1996, de 8 h à 12 h

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Seules sont autorisées :

Règles graduées.

Calculatrices de poche, y compris les calculatrices programmables et alphanumériques, à fonctionnement autonome, sans imprimante, sans document d'accompagnement et de format maximum 21 cm x 15 cm de large, sans limitation de nombre.

Un banquier s'est imposé de vendre une action en dix jours ouvrables. Chaque jour, suivant le cours du jour, il décide de vendre ou d'attendre dans l'espoir de vendre mieux plus tard. S'il n'a pas réalisé la vente au neuvième jour, il s'impose de vendre son action au dixième jour. Quelle stratégie va-t-il choisir?

Le problème ci-dessous propose, dans un cadre théorique précis, d'évaluer diverses stratégies pour de tels choix en chaîne.

On considère une suite d'expériences aléatoires identiques et indépendantes, à laquelle on associe une suite $(X_i)_{i \geq 1}$ de variables aléatoires, définies sur un espace de probabilité $(\Omega, \mathcal{A}, P)$, indépendantes et toutes de même loi.

On considère un entier naturel non nul n .

Si n est égal à 1, on définit le gain G_1 par : $G_1 = X_1$.

Si n est supérieur ou égal à 2, on se donne pour chaque $i \in \{1, \dots, n-1\}$, un seuil σ_i et on définit le gain G_n par :

si, pour tout i strictement inférieur à n , $X_i < \sigma_i$, alors $G_n = X_n$

et sinon, $G_n = X_k$ où k est le plus petit rang i tel que $X_i \geq \sigma_i$.

Le gain G_n est une variable aléatoire dont l'espérance est notée g_n .

(Dans l'exemple introductif du banquier, n est égal à 10, X_i représente le cours de l'action au jour de rang i et G_{10} est égal au prix de la vente).

On étudie en partie I, trois stratégies dans le cas d'expériences aléatoires discrètes et en partie II, trois stratégies dans le cas d'expériences aléatoires continues. On étudie dans le préliminaire une suite numérique que l'on retrouve à la fin de la partie II.

Les parties I et II sont dans une large mesure indépendantes.

Préliminaire

On définit la suite $(u_n)_{n \geq 1}$ par : $u_1 = \frac{1}{2}$ et $\forall n \in \mathbb{N}^* \quad u_{n+1} = \frac{1+u_n^2}{2}$.

1. a. Ecrire un programme en Pascal qui calcule et affiche les 100 premiers termes de la suite.
b. A l'aide de la calculatrice, donner une valeur approchée de u_{100} à 10^{-2} près.

2. a. Etudier la fonction h définie par : $\forall x \in [0,1] \quad h(x) = \frac{1+x^2}{2}$.

b. Montrer que la suite $(u_n)_{n \geq 1}$ est croissante.

c. Montrer que $(u_n)_{n \geq 1}$ est convergente et déterminer sa limite.

3. On pose pour tout n , entier naturel non nul : $v_n = 1 - u_n$.

a. Montrer que : $\forall n \geq 1 \quad \frac{1}{v_{n+1}} - \frac{1}{v_n} = \frac{1}{2 - v_n}$. En déduire : $\forall n \geq 1 \quad \frac{1}{v_n} \geq \frac{n+3}{2}$.

b. Montrer que : $\forall x \in [0,1] \quad \frac{1}{2-x} \leq \frac{1}{2} + \frac{x}{2}$. En déduire : $\forall n \geq 1 \quad \frac{1}{v_{n+1}} - \frac{1}{v_n} \leq \frac{1}{2} + \frac{1}{n+3}$.

Montrer que : $\forall n \geq 2 \quad \sum_{k=2}^n \frac{1}{k} \leq \ln n$ où $\ln n$ désigne le logarithme népérien de n .

En déduire : $\forall n \geq 1 \quad \frac{1}{v_n} \leq \frac{n+2}{2} + \ln(n+2)$.

c. Déterminer un équivalent de v_n quand n tend vers $+\infty$.

I. Exemples d'expériences aléatoires discrètes.

Dans cette partie r est un entier impair, supérieur ou égal à 3, et on suppose que, pour tout i entier naturel non nul, la variable aléatoire X_i est discrète et équirépartie sur l'ensemble $\left\{0, \frac{1}{r}, \frac{2}{r}, \dots, \frac{r}{r}\right\}$ (chacune des $r+1$ valeurs étant prise avec la même probabilité).

1. Première stratégie.

On pose, pour tout entier n supérieur ou égal à 2, $\sigma_1 = 0$. On a donc, pour tout entier naturel n non nul, $G_n = X_1$.

Calculer l'espérance g_n de la variable aléatoire G_n . (On rappelle que : $\sum_{j=1}^k j = \frac{k(k+1)}{2}$).

2. Deuxième stratégie.

On pose, pour tout entier n supérieur ou égal à 2 et pour tout $i \in \{1, \dots, n-1\}$, $\sigma_i = 0,5$.

a. Calculer $P(X_1 < 0,5)$.

b. Exprimer, en fonction des variables $X_1, \dots, X_n$, l'événement $(G_n = \frac{j}{r})$ pour $j \in \left\{0, \dots, \frac{r-1}{2}\right\}$ puis pour

$j \in \left\{\frac{r+1}{2}, \dots, r\right\}$.

En déduire que la loi de G_n est donnée par :

$$\forall j \in \left\{0, \dots, \frac{r-1}{2}\right\} \quad P(G_n = \frac{j}{r}) = \frac{2}{r+1} \frac{1}{2^n}$$

$$\forall j \in \left\{\frac{r+1}{2}, \dots, r\right\} \quad P(G_n = \frac{j}{r}) = \frac{2}{r+1} \left(1 - \frac{1}{2^n}\right).$$

Mathématiques II 3/4

c. Calculer g_n . Montrer que la suite $(g_n)_{n \geq 1}$ est croissante.

Déterminer la limite de g_n quand n tend vers $+\infty$ avec r fixé. Pouvait-on prévoir ce résultat ?

Déterminer la limite de g_n quand r tend vers $+\infty$ avec n fixé.

3. *Troisième stratégie.*

On pose, pour tout entier n supérieur ou égal à 2 et pour tout $i \in \{1, \dots, n-1\}$, $\sigma_i = 1$.

a. Exprimer, en fonction des variables $X_1, \dots, X_n$, l'événement $(G_n = \frac{j}{r})$ pour $j \in \{0, \dots, r-1\}$.

En déduire la loi de G_n .

b. Calculer g_n . Montrer que la suite $(g_n)_{n \geq 1}$ est croissante.

Déterminer la limite de g_n quand n tend vers $+\infty$ avec r fixé. Pouvait-on prévoir ce résultat ?

Déterminer la limite de g_n quand r tend vers $+\infty$ avec n fixé.

4. Comparer brièvement les trois stratégies de la partie I.

II. Exemples d'expériences aléatoires continues.

Dans cette partie, on suppose que, pour tout entier naturel non nul i , la variable aléatoire X_i suit une loi de probabilité uniforme sur $[0, 1]$; elle admet donc une densité φ définie par :

$$\forall t \in [0, 1] \quad \varphi(t) = 1 \quad \text{et} \quad \text{sinon} \quad \varphi(t) = 0.$$

On dit que les variables $(X_i)_{i \geq 1}$ sont indépendantes si et seulement si, pour tout entier naturel n non nul et pour tout $(t_1, \dots, t_n) \in \mathbb{R}^n$, les événements $(X_1 \leq t_1), \dots, (X_n \leq t_n)$ sont mutuellement indépendants ; on a alors,

$$\text{pour tout } (t_1, \dots, t_n) \in \mathbb{R}^n, \quad P\left(\bigcap_{i=1}^n (X_i \leq t_i)\right) = \prod_{i=1}^n P(X_i \leq t_i). \quad (\text{Les variables étant des variables à densité, les}$$

égalités ci-dessus sont encore vraies si on remplace $(X_i \leq t_i)$ par $(X_i < t_i)$).

On note F_n la fonction de répartition de G_n .

1. Déterminer F , la fonction de répartition de X_1 .

2. *Première stratégie.*

On pose, pour tout entier n supérieur ou égal à 2, $\sigma_1 = 0$.

Calculer l'espérance g_n de la variable aléatoire G_n .

3. *Deuxième stratégie.*

Soit un réel $\alpha \in [0, 1[$. On pose, pour tout entier n supérieur ou égal à 2 et pour tout $i \in \{1, \dots, n-1\}$, $\sigma_i = \alpha$.

a. Que vaut $F_n(t)$ pour t n'appartenant pas à $[0, 1[$?

Pour $t \in [0, \alpha[$, décrire l'événement $(G_n \leq t)$ et en déduire $F_n(t)$.

Pour $t \in [\alpha, 1[$, décrire l'événement $(G_n > t)$ et en déduire $F_n(t)$.

Montrer que G_n admet une densité f_n définie par :

$$\forall t \in [0, \alpha[\quad f_n(t) = \alpha^{n-1}, \quad \forall t \in [\alpha, 1[\quad f_n(t) = \frac{1 - \alpha^n}{1 - \alpha}, \quad \text{et} \quad f_n(t) = 0 \quad \text{sinon.}$$

b. Calculer g_n en fonction de α . Montrer que la suite $(g_n)_{n \geq 1}$ est croissante.

Déterminer la limite de g_n quand n tend vers $+\infty$. Pouvait-on prévoir ce résultat ?

c. Déterminer la valeur de α telle que g_2 soit maximum.

d. Dans cette question, $\alpha = 0,5$.

Donner la valeur de g_n . Quelle remarque peut-on faire en comparant ce résultat avec celui de la deuxième stratégie de la partie I ?

4. *Troisième stratégie.*

Soit $(a_k)_{k \geq 1}$ une suite de réels telle que, pour tout entier k non nul, $0 \leq a_k < 1$.

Pour tout entier n supérieur ou égal à 2 et pour tout $i \in \{1, \dots, n-1\}$, on pose $\sigma_i = a_{n-i}$.

a. En utilisant les résultats de II.2., montrer que G_1 admet une densité et une espérance.

Donner la valeur de g_1 .

b. En utilisant les résultats de II.3., montrer que G_2 admet une densité et une espérance.

Donner la valeur de g_2 en fonction de a_1 .

Déterminer la valeur de a_1 qui maximise g_2 .

c. Soit n un entier supérieur ou égal à 2.

On suppose que G_n admet une densité notée f_n .

On suppose de plus que, pour tout entier $k \in \{1, \dots, n\}$, G_k admet une espérance g_k vérifiant $g_k \in [0, 1[$ et que, pour tout entier $k \in \{1, \dots, n-1\}$, $a_k = g_k$.

Et donc la variable aléatoire G_{n+1} est associée aux seuils : $\sigma_1 = a_n, \sigma_2 = g_{n-1}, \dots, \sigma_n = g_1$.

Montrer que : $\forall t \in [0, a_n[\quad F_{n+1}(t) = a_n F_n(t)$

$$\forall t \in [a_n, 1[\quad F_{n+1}(t) = t - a_n + a_n F_n(t)$$

En déduire que G_{n+1} admet une densité f_{n+1} et donner une relation entre f_{n+1} et f_n .

En déduire que G_{n+1} admet une espérance g_{n+1} vérifiant $g_{n+1} = a_n g_n + \frac{1}{2}(1 - a_n^2)$.

Déterminer la valeur de a_n qui maximise g_{n+1} .

Montrer que, pour cette valeur de a_n , $g_{n+1} = \frac{1+g_n^2}{2}$ et que $g_{n+1} \in [0, 1[$.

d. On construit ainsi, par récurrence, une suite de variables aléatoires $(G_n)_{n \geq 1}$ dont les espérances

vérifient : $\forall n \in \mathbb{N}^* \quad g_{n+1} = \frac{1+g_n^2}{2}$.

Quelle est alors la limite de la suite $(g_n)_{n \geq 1}$?

5. Comparer les trois stratégies de la partie II.

- FIN -



CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS
DIRECTION DE L'ENSEIGNEMENT
Direction des Admissions et Concours

**ECOLE DES HAUTES ETUDES COMMERCIALES
ECOLE SUPERIEURE DE COMMERCE DE PARIS
ECOLE SUPERIEURE DE COMMERCE DE LYON**

CONCOURS D'ADMISSION SUR CLASSES PREPARATOIRES

OPTION SCIENTIFIQUE

MATHEMATIQUES II

jeudi 24 avril 1997, de 8 h à 12 h

*La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.*

Seules sont autorisées:

Une règle graduée.

Une calculatrice de poche pouvant être programmable et /ou alphanumérique, à fonctionnement autonome, sans imprimante, sans document d'accompagnement et de surface de base maximum de 21 cm de long sur 15 cm de large.

Une éprouvette contient 10 bactéries, 4 sont des bactéries de type A, 6 de type B. On les laisse se reproduire en milliers d'exemplaires, la proportion de bactéries de chaque type restant inchangée. On prélève alors, au hasard, 10 bactéries que l'on met dans une autre éprouvette. On les laisse se reproduire en milliers d'exemplaires dans les mêmes conditions que précédemment, et on recommence l'expérience.
Que se passe-t-il après un grand nombre d'expériences?

L'énoncé théorique ci-dessous propose un modèle probabiliste pour répondre à cette question.

Définitions. Soit une variable aléatoire X ; on note $E(X)$ l'espérance de X si celle-ci existe.

On note N un entier supérieur ou égal à 2 et k_0 un entier de $\{0, \dots, N\}$. On pose $p = \frac{k_0}{N}$ et $q = 1-p$.

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires définies sur un espace de probabilité $(\Omega, \mathcal{A}, P)$, à valeurs dans $\{0, \dots, N\}$, dont les lois de probabilité sont définies de la manière suivante:

X_0 est la variable certaine égale à k_0 .

X_1 suit la loi binomiale de paramètres N et p . (Par convention, on dit que la loi binomiale de paramètres N et 0 est la loi de la variable certaine égale à 0 et que la loi binomiale de paramètres N et 1 est la loi de la variable certaine égale à N).

Pour tout entier n non nul et tout entier k de $\{0, \dots, N\}$ tel que $P(X_n = k) \neq 0$, la loi conditionnelle de X_{n+1} sachant

$(X_n = k)$ est la loi binomiale de paramètres N et $\frac{k}{N}$. En d'autres termes:

pour tout entier n non nul et tout entier k de $\{0, \dots, N\}$ tel que $P(X_n = k) \neq 0$ et pour tout entier i de $\{0, \dots, N\}$,

$$P(X_{n+1} = i / X_n = k) = C_N^i \left(\frac{k}{N}\right)^i \left(1 - \frac{k}{N}\right)^{N-i} \quad (\text{avec la convention habituelle que } 0^0 = 1).$$

On a de plus l'hypothèse (H): pour tout entier n non nul, pour tout n -uplet $(k_1, \dots, k_n)$ de $\{0, \dots, N\}^n$ tel que

$$P(X_n = k_n, \dots, X_1 = k_1) \neq 0, \text{ pour tout entier } i \text{ de } \{0, \dots, N\}, \quad P(X_{n+1} = i / X_n = k_n, \dots, X_1 = k_1) = P(X_{n+1} = i / X_n = k_n).$$

Cette hypothèse n'est utile que pour la question 6. de la partie 1.

On définit la suite de variables aléatoires $(F_n)_{n \geq 0}$ par $F_n = \frac{X_n}{N}$.

PRELIMINAIRE

Dans l'exemple ci-dessus, en appelant N le nombre de bactéries prélevées à chaque expérience, k_0 le nombre de bactéries de type A dans la première éprouvette au début de la première expérience et n le numéro de l'expérience, donner une interprétation de la variable X_n et justifier par des arguments tirés du cours, l'utilisation de la loi binomiale. Comment interpréter l'hypothèse (H)?

PARTIE 1

Dans cette partie, $N = 3$.

1) Que dire de la suite $(X_n)_{n \geq 0}$ si $k_0 = 0$? Si $k_0 = 3$?

On suppose désormais, dans la suite de cette partie, que $k_0 = 1$.

$$2) \text{ Pour tout entier } n, \text{ on pose } U_n = \begin{pmatrix} P(X_n = 0) \\ P(X_n = 1) \\ P(X_n = 2) \\ P(X_n = 3) \end{pmatrix}. \text{ Montrer que } U_{n+1} = A U_n \text{ où } A = \begin{pmatrix} 1 & \frac{8}{27} & \frac{1}{27} & 0 \\ 0 & \frac{4}{9} & \frac{2}{9} & 0 \\ 0 & \frac{2}{9} & \frac{4}{9} & 0 \\ 0 & \frac{1}{27} & \frac{8}{27} & 1 \end{pmatrix}.$$

3) a) Soit le vecteur ligne $V = (0, 1, 2, 3)$. Calculer VA .

b) Montrer que $E(X_n) = V U_n$, pour tout entier n . En déduire la valeur de $E(X_n)$.

4) a) On pose $Y_2 = \begin{pmatrix} 1 \\ -3 \\ 3 \\ -1 \end{pmatrix}$ et $Y_3 = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$. Calculer AY_2 et AY_3 en fonction de Y_2 et Y_3 .

- b) Montrer que A est diagonalisable. Donner ses valeurs propres et une base de vecteurs propres.
c) Calculer A^n pour tout entier n .

5) a) Montrer que la loi de X_n est donnée par:

$$P(X_n = 0) = \frac{2}{3} - \frac{1}{2} \left(\frac{2}{3} \right)^n - \frac{1}{6} \left(\frac{2}{9} \right)^n, \quad P(X_n = 1) = \frac{1}{2} \left(\left(\frac{2}{3} \right)^n + \left(\frac{2}{9} \right)^n \right)$$

$$P(X_n = 2) = \frac{1}{2} \left(\left(\frac{2}{3} \right)^n - \left(\frac{2}{9} \right)^n \right), \quad P(X_n = 3) = \frac{1}{3} - \frac{1}{2} \left(\frac{2}{3} \right)^n + \frac{1}{6} \left(\frac{2}{9} \right)^n.$$

b) Montrer que la suite $(F_n)_{n \geq 0}$ converge en loi. Quelle est la loi limite?

6) Pour tout entier n non nul, on définit l'événement $B_n = \bigcap_{k=1}^n (X_k \leq 1)$.

On pose $x_1 = P(X_1 = 0)$, pour tout entier k supérieur ou égal à 2, $x_k = P(X_1 = 1, \dots, X_{k-1} = 1, X_k = 0)$, et pour tout entier k non nul, $y_k = P(X_1 = 1, \dots, X_k = 1)$.

a) Exprimer pour tout entier k non nul, x_{k+1} et y_{k+1} en fonction de y_k . En déduire les valeurs de x_n et y_n pour tout entier n non nul.

b) Montrer que $P(B_n) = \sum_{k=1}^n x_k + y_n$. En déduire $P(B_n)$ et la limite de la suite $(P(B_n))_{n \geq 1}$.

c) En déduire la probabilité qu'il existe n vérifiant $F_n > 0,5$.

PARTIE 2

Dans cette partie, N est un entier supérieur ou égal à 2 et k_0 un entier de $\{1, \dots, N-1\}$.

On pose pour tout entier n , $u_n = P(X_n = 0) + P(X_n = N)$ et $v_n = 1 - u_n$.

A. Loi de X_n

1) Montrer que pour tout entier i de $\{0, \dots, N\}$, $P(X_{n+1} = i) = \sum_{k=0}^N P(X_n = k) C_N^i \left(\frac{k}{N} \right)^i \left(1 - \frac{k}{N} \right)^{N-i}$.

2) Montrer que pour tout entier n , $E(X_{n+1}) = E(X_n)$.

3) a) Montrer que pour tout entier n , $E[X_{n+1}(N - X_{n+1})] = \frac{N-1}{N} E[X_n(N - X_n)]$.

b) En déduire la valeur de $E[X_n(N - X_n)]$ en fonction de n , N et k_0 .

4) Montrer que la suite $(u_n)_{n \geq 0}$ est croissante et convergente.

5) a) Soit X une variable aléatoire à valeurs dans $\mathbb{R}_+$, prenant un nombre fini de valeurs.

Montrer que pour tout réel a strictement positif, $P(X \geq a) \leq \frac{E(X)}{a}$.

b) Etudier sur $[1, N-1]$ la fonction f définie par: $\forall x \in [1, N-1], f(x) = x(N-x)$.

c) En utilisant la valeur de $E[X_n(N - X_n)]$, montrer que pour tout entier n , $0 \leq v_n \leq pq \frac{N^2}{N-1} \left(1 - \frac{1}{N} \right)^n$.

- 6) a) Quelle est la limite de la suite $(v_n)_{n \geq 0}$?
 b) En déduire que pour tout entier k de $\{1, \dots, N-1\}$, $\lim_{n \rightarrow +\infty} P(X_n = k) = 0$.
 c) En utilisant le résultat du 2), montrer que $\lim_{n \rightarrow +\infty} P(X_n = N) = p$. Déterminer $\lim_{n \rightarrow +\infty} P(X_n = 0)$.
 d) Montrer que la suite $(F_n)_{n \geq 0}$ converge en loi. Quelle est la loi limite?

B. Temps d'arrêt

On définit la variable aléatoire T par:

si pour tout entier n , $(X_n \neq 0)$ et $(X_n \neq N)$, alors $T = 0$

sinon, $T = n$ où n est le plus petit entier k tel que $(X_k = 0)$ ou $(X_k = N)$.

- 1) Que vaut $P(T=0)$? Montrer que pour tout entier n non nul, $P(T=n) = v_{n-1} - v_n$.

- 2) a) Montrer que $\sum_{k=1}^n kP(T=k) = \sum_{k=0}^{n-1} v_k - nv_n$.

- b) En déduire que T admet une espérance et que $E(T) \leq pq \frac{N^3}{N-1}$ (on ne cherchera pas à calculer $E(T)$).

C. Retour aux bactéries

Dans l'exemple des bactéries, on a posé la question : « Que se passe-t-il après un grand nombre d'expériences ? ». Pouvez-vous maintenant y répondre ?

ECOLE DES HAUTES ETUDES COMMERCIALES
ECOLE SUPERIEURE DE COMMERCE DE PARIS
ECOLE SUPERIEURE DE COMMERCE DE LYON

CONCOURS D'ADMISSION SUR CLASSES PREPARATOIRES

OPTION SCIENTIFIQUE

MATHEMATIQUES II

Samedi 25 Avril 1998, de 8 h à 12 h

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document; l'utilisation de toute calculatrice et de tout matériel électronique est interdite.

Seule l'utilisation d'une règle graduée est autorisée.

Notations

Dans tout le problème, n désigne un entier supérieur ou égal à 2.

Toutes les variables aléatoires, considérées dans chaque partie de ce problème, sont des variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathcal{A}, P)$.

X étant une variable aléatoire admettant des moments d'ordre 1 et 2, $E(X)$ désigne l'espérance de X , $V(X)$ sa variance.

Tous les couples (X, Y) de variables aléatoires à densité considérés dans ce problème sont tels que X et Y admettent des moments d'ordre 1 et 2 et le produit XY est une variable aléatoire à densité admettant un moment d'ordre 1. On définit alors la covariance de X et Y par:

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

et on admet que cette covariance vérifie les propriétés suivantes :

- 1) $\text{cov}(\alpha X + \beta Y, Z) = \alpha \text{cov}(X, Z) + \beta \text{cov}(Y, Z)$
- 2) $\text{cov}(X, Y) = \text{cov}(Y, X)$
- 3) $|\text{cov}(X, Y)| \leq \sqrt{V(X)V(Y)}$
- 4) si X et Y sont indépendantes, alors $\text{cov}(X, Y) = 0$

où α et β sont des réels, X , Y et Z des variables aléatoires à densité.

Ces propriétés ne doivent pas être démontrées.

Un gestionnaire investit un capital parmi n actifs, notés $A_1, A_2, \dots, A_n$ (par exemple des actions), disponibles sur le Marché Boursier. Les rendements à un an de ces actifs, exprimés en pourcentage, sont des variables aléatoires $R_1, R_2, \dots, R_n$, admettant des moments d'ordre 1 et 2. Par exemple, si l'actif A_1 a rapporté 6%, R_1 prend la valeur 6.

Le gestionnaire constitue un portefeuille, c'est-à-dire un n -uplet $(x_1, x_2, \dots, x_n)$ tel que, pour tout i de $\{1, \dots, n\}$, x_i est un

réel positif ou nul et tel que : $\sum_{i=1}^n x_i = 1$. Chaque coefficient x_i représente la proportion du capital investie dans l'actif

A_i . Par exemple, si n vaut 3 et si le gestionnaire investit un quart du capital dans l'actif A_1 , la moitié du capital dans

l'actif A_2 et le quart du capital dans l'actif A_3 , le portefeuille vaut $\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$.

Si $(x_1, x_2, \dots, x_n)$ est un portefeuille donné, le rendement (en pourcentage) de ce portefeuille est la variable aléatoire :

$$R = \sum_{i=1}^n x_i R_i.$$

Notre gestionnaire prudent désire minimiser les risques et recherche pour cela les portefeuilles dont le rendement R est de variance minimale, sous certaines hypothèses.

Préliminaire

On considère le portefeuille : $Q = (x_1, \dots, x_n)$ et son rendement : $R = \sum_{i=1}^n x_i R_i$. On rappelle que la variance de R est :

$$V(R) = \sum_{i=1}^n x_i^2 V(R_i) + 2 \sum_{1 \leq i < j \leq n} x_i x_j \text{cov}(R_i, R_j).$$

1) On suppose, dans un programme Pascal, avoir défini :

Const $n = 5$;

Type $\text{Tab} = \text{Array}[1..n] \text{ of real}$;

var $A : \text{Array}[1..n, 1..n] \text{ of real}$;

où $A[i, j]$ représente $\text{cov}(R_i, R_j)$.

Ecrire une fonction V de type *real* de paramètre le portefeuille Q de type *Tab* qui renvoie la valeur de $V(R)$.

2) On considère l'ensemble : $H = \left\{ (x_1, \dots, x_n) \in (\mathbb{R}^+)^n, \sum_{i=1}^n x_i = 1 \right\}$. On admet que H est fermé.

On définit sur $\mathbb{R}^n$, la fonction $F : F(x_1, \dots, x_n) = \sum_{i=1}^n x_i^2 V(R_i) + 2 \sum_{1 \leq i < j \leq n} x_i x_j \text{cov}(R_i, R_j)$.

Montrer que F admet un minimum sur H .

La suite du problème consiste à déterminer ce minimum et les points de H où ce minimum est atteint, pour répondre ainsi à la demande du gestionnaire prudent.

Mathématiques II 3/4

Partie 1

Dans cette partie, l'entier n vaut 2 et les rendements des actifs A_1 et A_2 sont notés respectivement X et Y .

On suppose : $V(X) = 12$, $V(Y) = 3$, $\text{cov}(X, Y) = c$, où c est un réel donné.

Pour un réel a de $[0, 1]$, on considère le portefeuille $(a, 1 - a)$ dont le rendement est la variable aléatoire :

$$R = aX + (1 - a)Y.$$

1) Montrer que l'on a : $|c| \leq 6$.

2) a) Montrer que l'on a : $V(R) = (15 - 2c)a^2 + 2(c - 3)a + 3$.

b) On définit sur $\mathbb{R}$ la fonction h par : $h(x) = (15 - 2c)x^2 + 2(c - 3)x + 3$. Etudier les variations de h sur $[0, 1]$, en distinguant les deux cas : $c \in [-6, 3[$ et $c \in [3, 6]$. Montrer qu'il existe un unique portefeuille dont le rendement est de variance minimale. On déterminera ce portefeuille en fonction de c .

3) a) On suppose : $c = -6$.

Déterminer le portefeuille dont le rendement R est de variance minimale et cette variance minimale. Que peut-on dire de la variable aléatoire R dans ce cas?

b) On suppose que les variables X et Y sont indépendantes. Montrer que $\left(\frac{1}{5}, \frac{4}{5}\right)$ est le portefeuille de rendement de variance minimale.

4) On suppose dans cette question que X et Y sont des variables gaussiennes indépendantes, X de moyenne égale à 6 et de variance égale à 12, Y de moyenne égale à 3 et de variance égale à 3.

Soit R le rendement du portefeuille $\left(\frac{1}{5}, \frac{4}{5}\right)$. Quelle est la loi de R ? Calculer la probabilité que R soit supérieur ou égal

à 4. (On donne : $\Phi\left(\frac{1}{\sqrt{15}}\right) = 0,60$, où Φ est la fonction de répartition de la loi normale centrée réduite).

5) On suppose dans cette question que les variables à densité X et Y sont indépendantes. On suppose de plus que X suit la loi uniforme sur $[0, 12]$ et que Y suit la loi uniforme sur $[0, 6]$.

a) Donner les valeurs des espérances de ces variables et vérifier : $V(X) = 12$, $V(Y) = 3$.

b) Déterminer la loi de la variable $4Y$, puis la densité de la variable $X + 4Y$. Tracer le graphe de cette densité.

c) Soit R le rendement du portefeuille $\left(\frac{1}{5}, \frac{4}{5}\right)$. Calculer la probabilité que R soit supérieur ou égal à 4.

Partie 2

Dans cette partie, n vaut 3 et les rendements des actifs A_1 , A_2 et A_3 sont notés respectivement X , Y et Z .

On suppose de plus : $V(X) = 2$, $V(Y) = V(Z) = 6$, $\text{cov}(X, Y) = -1$, $\text{cov}(X, Z) = 2$, $\text{cov}(Y, Z) = 1$.

On considère le portefeuille (x, y, z) dont le rendement est la variable : $R = xX + yY + zZ$.

La fonction f , l'ensemble K et l'ensemble K_0 sont définis comme suit :

pour tout (x, y) de $\mathbb{R}^2$, $f(x, y) = 4x^2 + 10y^2 + 4xy - 8x - 10y + 6$,

$K = \{(x, y) \in (\mathbb{R}^+)^2, x + y \leq 1\}$, $K_0 = \{(x, y) \in (\mathbb{R}^+)^2, x + y < 1\}$. On admet que K_0 est ouvert.

1) Montrer que le problème du gestionnaire revient à déterminer le minimum de f sur K . Dessiner le domaine K .

2) a) Montrer que f admet un minimum local sur $\mathbb{R}^2$ atteint au point (x_0, y_0) que l'on déterminera.

b) En déduire que f n'admet pas de minimum sur K_0 .

3) a) On pose : $K_1 = \{(0, y), y \in [0, 1]\}$, $K_2 = \{(x, 0), x \in [0, 1]\}$, $K_3 = \{(x, 1 - x), x \in [0, 1]\}$.

Déterminer les minimums de f respectivement sur K_1 , K_2 et K_3 .

b) En déduire l'unique portefeuille dont le rendement est de variance minimale.

Partie 3

Dans cette partie, n vaut 3 et les rendements des actifs A_1 , A_2 et A_3 sont notés respectivement X , Y et Z .

1) On suppose : $V(X) = V(Y) = V(Z) = 1$, $cov(X, Y) = cov(X, Z) = cov(Y, Z) = c$, où c est un réel donné.

a) Calculer $V(X + Y + Z)$. Montrer que l'on a : $c \in \left[-\frac{1}{2}, 1\right]$.

b) On suppose : $c \neq 1$. On considère un portefeuille (x, y, z) de rendement R .

Montrer que l'on a : $V(R) = (1 - c) \left[\left(x - \frac{1}{3}\right)^2 + \left(y - \frac{1}{3}\right)^2 + \left(z - \frac{1}{3}\right)^2 \right] + \frac{1 + 2c}{3}$.

Déterminer le portefeuille dont le rendement est de variance minimale.

2) Soit A , B , C des variables aléatoires indépendantes à valeurs dans $\mathbb{N}^*$ suivant toutes la loi géométrique de paramètre $\frac{1}{2}$; on a donc : pour tout k de $\mathbb{N}^*$, $P(A = k) = P(B = k) = P(C = k) = \left(\frac{1}{2}\right)^k$.

On suppose que les variables X , Y et Z vérifient :

$$\begin{cases} X = B + C \\ Y = A + C + 1 \\ Z = A + B + 2 \end{cases}$$

a) Déterminer les espérances, variances et covariances des variables X , Y et Z .

b) Montrer que le portefeuille de rendement de variance minimale est $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

c) Déterminer la loi de $A+B+C$. En déduire la probabilité que le rendement R du portefeuille ci-dessus soit supérieur ou égal à 5.

Partie 4

Dans cette partie, n est un entier supérieur ou égal à 2.

On note M la matrice de covariance de $R_1, \dots, R_n$, matrice carrée d'ordre n dont le coefficient de la ligne i et de la colonne j est égal à $cov(R_i, R_j)$.

On note $M_{n,1}(\mathbb{R})$ l'espace vectoriel des matrices réelles ayant n lignes et 1 colonne.

1) On considère un élément de $M_{n,1}(\mathbb{R})$: $U = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ et la variable aléatoire : $T = \sum_{i=1}^n x_i R_i$.

Vérifier que l'on a : $V(T) = {}^t U M U$, où ${}^t U$ désigne la transposée de U .

2) a) Montrer que M est diagonalisable.

b) A l'aide du 1), montrer que les valeurs propres de M sont positives ou nulles.

3) On suppose M inversible. Pour tout (U, W) de $(M_{n,1}(\mathbb{R}))^2$, on pose $\varphi(U, W) = {}^t U M W$.

a) Montrer : si ${}^t U M U = 0$, alors $U = 0$.

b) Montrer que φ est un produit scalaire. On note N la norme associée à ce produit scalaire.

c) Montrer : pour tout (U, W) de $(M_{n,1}(\mathbb{R}))^2$, $\left[N\left(\frac{U+W}{2}\right)\right]^2 = \frac{1}{2} \left[(N(U))^2 + (N(W))^2 \right] - \left[N\left(\frac{U-W}{2}\right) \right]^2$.

d) En déduire l'unicité du portefeuille dont le rendement est de variance minimale (l'existence ayant été montrée dans le préliminaire).



ECOLE DES HAUTES ETUDES COMMERCIALES
ECOLE SUPERIEURE DE COMMERCE DE PARIS
ECOLE SUPERIEURE DE COMMERCE DE LYON

CONCOURS D'ADMISSION SUR CLASSES PREPARATOIRES

OPTION SCIENTIFIQUE

MATHEMATIQUES II

Samedi 15 Mai 1999, de 8h. à 12h.

*La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.
Ils ne doivent faire usage d'aucun document ; l'utilisation de toute calculatrice et de tout matériel électronique est interdite.
Seule l'utilisation d'une règle graduée est autorisée.*

La partie 1 présente la loi du χ^2 (lire khi-deux) et certaines de ses propriétés. La partie 2 présente une application de la loi du χ^2 . La partie 3 considère, sur un exemple, un test statistique, reposant sur les résultats des parties 1 et 2.

Notation

Si X et Y sont deux variables aléatoires discrètes, on note $cov(X, Y)$ leur covariance, si celle-ci existe.

Partie 1

Soit r un entier non nul. On dit qu'une variable aléatoire X suit la loi du χ^2 à r degrés de liberté si et seulement si X suit la loi Γ de paramètres 2 et $\frac{r}{2}$, c'est-à-dire si X admet pour densité la fonction f_r définie par :

$$\forall x > 0, \quad f_r(x) = \frac{1}{\Gamma\left(\frac{r}{2}\right) 2^{\frac{r}{2}}} x^{\frac{r}{2}-1} e^{-\frac{x}{2}} \quad \forall x \leq 0, \quad f_r(x) = 0.$$

1) Déterminer l'espérance et la variance d'une variable X suivant la loi du χ^2 à r degrés de liberté.

2) a) Montrer que pour tout $\lambda > 0$, pour tout entier n non nul :
$$e^\lambda = \sum_{k=0}^{n-1} \frac{\lambda^k}{k!} + \int_0^\lambda e^{\lambda-t} \frac{t^{n-1}}{(n-1)!} dt.$$

b) Soit Y_λ une variable aléatoire suivant la loi de Poisson de paramètre λ et X_{2n} une variable aléatoire suivant la loi du χ^2 à $2n$ degrés de liberté. Montrer que $P(X_{2n} > 2\lambda) = P(Y_\lambda < n)$.

c) Ecrire une fonction en langage Pascal de paramètres n entier et x réel qui retourne la valeur de $P(X_{2n} > x)$.

d) A l'aide de la table des lois de Poisson donnée en annexe, donner l'allure de la fonction F_6 , fonction de répartition d'une variable suivant la loi du χ^2 à 6 degrés de liberté en précisant les valeurs de $F_6(0)$, $F_6(4)$, $F_6(8)$.

3) Soit k un entier non nul et $X_1, \dots, X_k$ des variables indépendantes gaussiennes centrées réduites.

a) Déterminer la loi de X_1^2 .

b) En déduire la loi de $X_1^2 + \dots + X_k^2$ (on admettra que les variables $X_1^2, \dots, X_k^2$ sont indépendantes).

c) Tracer sur un même graphique l'allure des fonctions de répartition de deux variables aléatoires, l'une suivant la loi du χ^2 à r degrés de liberté et l'autre suivant la loi du χ^2 à r' degrés de liberté, avec r et r' deux entiers non nuls tels que $r < r'$.

Partie 2

Soit n et s des entiers supérieurs ou égaux à 2. On considère une urne contenant des boules de couleurs $C_1, \dots, C_s$. Les boules de couleur C_i sont en proportion p_i . On a donc $\sum_{i=1}^s p_i = 1$ et on suppose que, pour tout i , $p_i > 0$.

On effectue n tirages successifs d'une boule avec remise.

Pour tout i de $\{1, \dots, s\}$, on note X_i la variable aléatoire égale au nombre de boules de couleur C_i obtenues à l'issue des n

tirages. On remarque que la variable X_i dépend de n et que $\sum_{i=1}^s X_i = n$.

On définit la variable aléatoire U_n par : $U_n = \sum_{i=1}^s \frac{(X_i - np_i)^2}{np_i}$.

A. Etude des variables X_i .

1) Déterminer la loi de X_i , son espérance et sa variance.

2) Soit $(i, j) \in \{1, \dots, s\}^2$ tel que $i \neq j$. Déterminer la loi de $X_i + X_j$ et sa variance.

En déduire que $\text{cov}(X_i, X_j) = -np_i p_j$.

B. On suppose dans cette partie que $s = 2$.

1) Montrer que $U_n = Z_1^2$ où $Z_1 = \frac{X_1 - np_1}{\sqrt{np_1 p_2}}$.

2) a) Par quelle loi peut-on approcher la loi de Z_1 lorsque n est grand?

b) Montrer que la suite (U_n) converge en loi vers la loi du χ^2 à 1 degré de liberté, quand n tend vers l'infini (on demande une démonstration précise).

C. On suppose dans cette partie que $s = 3$ et que $p_1 = p_2 = \frac{1}{4}$ et $p_3 = \frac{1}{2}$.

On pose $Z_1 = \frac{2}{\sqrt{n}} \left(X_3 - \frac{n}{2} \right)$ et $Z_2 = \sqrt{\frac{2}{n}} (X_1 - X_2)$.

1) Montrer que $U_n = Z_1^2 + Z_2^2$.

2) Déterminer les espérances et variances de Z_1 et Z_2 et $\text{cov}(Z_1, Z_2)$.

3) Par quelle loi peut-on approcher la loi de Z_1 lorsque n est grand?

4) Pour i élément de $\{1, \dots, n\}$, on définit la variable T_i par : $T_i = 1$ si au i ème tirage on a obtenu une boule de couleur C_1 , $T_i = -1$ si au i ème tirage on a obtenu une boule de couleur C_2 , $T_i = 0$ si au i ème tirage on a obtenu une boule de couleur C_3 .

a) Exprimer $X_1 - X_2$ à l'aide des variables T_i .

b) En déduire que l'on peut approcher la loi de Z_2 , quand n est grand, par la loi normale centrée réduite.

5) On admettra l'approximation suivante : n est supposé grand et sous cette hypothèse, Z_1 et Z_2 sont des variables indépendantes gaussiennes centrées réduites.

Quelle est la loi de U_n ?

6) On suppose avoir défini dans un programme Pascal :

`Type Tableau = Array[1..100] of integer;`

a) Ecrire une procédure *procedure Tirage*(var C : Tableau); permettant de simuler le tirage avec remise de 100 boules dans une urne contenant des boules de couleur C_1 ou C_2 ou C_3 en proportion respectivement $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$.

L'élément $C[i]$ vaut 1, 2 ou 3 et représente la couleur de la i ème boule tirée (C_1 , C_2 ou C_3). On utilisera la fonction *random* : *random*(4) retourne un entier aléatoire compris entre 0 et 3.

b) Ecrire une fonction *Difference* de paramètre C qui retourne la valeur de $X_1 - X_2$.

D. On suppose désormais s entier quelconque supérieur ou égal à 2.

1) On pose $Y_i = \frac{X_i - np_i}{\sqrt{np_i}}$. On note M la matrice de covariance des variables $Y_1, \dots, Y_s$.

Montrer que $M = I - N$ où I est la matrice unité et N la matrice dont le terme en ligne i et colonne j vaut $\sqrt{p_i p_j}$.

2) Montrer que $N^2 = N$ et déterminer le rang de N .

3) Montrer qu'il existe une matrice Q telle que ${}^t Q Q = I$ et telle que $M = Q J_{s-1} {}^t Q$ où J_{s-1} est la matrice carrée d'ordre s , diagonale, dont les $s-1$ premiers éléments diagonaux sont égaux à 1 et le dernier est nul. (On ne demande pas de calculer explicitement la matrice Q).

4) On définit les variables $Z_1, \dots, Z_s$ par :
$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_s \end{pmatrix} = {}^t Q \begin{pmatrix} Y_1 \\ \vdots \\ Y_s \end{pmatrix}.$$

On note a_{ij} l'élément de la ligne i et de la colonne j de ${}^t Q$. Exprimer chaque Z_i en fonction de $Y_1, \dots, Y_s$ et des a_{ij} .

Montrer que les variables Z_i sont centrées.

En utilisant la bilinéarité de la covariance, déterminer la matrice de covariance de $Z_1, \dots, Z_s$.

Qu'en déduit-on pour la variable Z_s ?

5) On admettra l'approximation suivante : n est supposé grand et sous cette hypothèse, $Z_1, \dots, Z_{s-1}$ sont des variables indépendantes gaussiennes centrées réduites.

Montrer que U_n suit la loi du χ^2 à $s-1$ degrés de liberté.

Partie 3

On s'intéresse à la répartition le long de l'année des naissances en Suède dans les années 1930.

Une statistique a été réalisée pour les 4 périodes dont les proportions p_i de journées dans l'année sont les suivantes :

période T_1 : Avril - Juin $p_1 = 0,250$

période T_2 : Juillet-Août $p_2 = 0,169$

période T_3 : Septembre - Octobre $p_3 = 0,167$

période T_4 : Novembre - Mars $p_4 = 0,414$.

On dispose d'un échantillon de 8000 naissances en Suède dans les années 1930, regroupées suivant les quatre périodes définies ci-dessus, selon les proportions f_i suivantes :

période T_1 : $f_1 = 0,264$

période T_2 : $f_2 = 0,173$

période T_3 : $f_3 = 0,159$

période T_4 : $f_4 = 0,404$.

On fait l'hypothèse que, parmi la population observée, le jour (aléatoire) de naissance dans l'année d'un enfant suit une loi uniforme. On confronte cette hypothèse à l'échantillon.

On donne les valeurs numériques suivantes :

- $0,99 = F_3(11,3)$ où F_3 désigne la fonction de répartition de la loi du χ^2 à 3 degrés de liberté,

- $8000 \sum_{i=1}^4 \frac{(f_i - p_i)^2}{p_i} = 12,03$.

En introduisant des variables aléatoires convenables X_1, X_2, X_3, X_4 et la variable U_n associée, justifier le rejet de l'hypothèse de répartition uniforme des naissances au cours de l'année.

ANNEXE

Extraits de la table cumulée des lois de Poisson $P(\lambda)$: valeurs de $P(X_\lambda \leq x)$.

$\lambda \backslash x$	1	2	3	4	5	6	7
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009
1	0,7358	0,4060	0,1991	0,0916	0,0404	0,0174	0,0073
2	0,9197	0,6767	0,4232	0,2381	0,1247	0,0620	0,0296
3	0,9810	0,8571	0,6472	0,4335	0,2650	0,1512	0,0818
4	0,9963	0,9473	0,8153	0,6288	0,4405	0,2851	0,1730
5	0,9994	0,9834	0,9161	0,7851	0,6160	0,4457	0,3007



CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS
DIRECTION DE L'ENSEIGNEMENT
Direction des Admissions et Concours

**ECOLE DES HAUTES ETUDES COMMERCIALES
E.S.C.P. - E.A.P.
ECOLE SUPERIEURE DE COMMERCE DE LYON**
CONCOURS D'ADMISSION SUR CLASSES PREPARATOIRES

OPTION SCIENTIFIQUE

MATHEMATIQUES II

Mardi 16 Mai 2000, de 8h. à 12h.

*La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.
Ils ne doivent faire usage d'aucun document ; l'utilisation de toute calculatrice et de tout matériel électronique est interdite.
Seule l'utilisation d'une règle graduée est autorisée.*

Le problème consiste en l'analyse d'un algorithme de tri et l'étude de sa complexité.

On note $\ln x$ le logarithme népérien d'un réel strictement positif x et $\log_2 x$ son logarithme en base 2. On rappelle que $\log_2 x = \frac{\ln x}{\ln 2}$.

Partie I : Étude d'une suite réelle

Dans cette partie la lettre n désignera toujours un entier naturel **au moins égal à 2**.

On considère la suite $(u_k)_{k \in \mathbb{N}^*}$ définie par son premier terme u_1 et vérifiant, pour tout n , la relation de récurrence : $u_n = n - 1 + \frac{2}{n} \sum_{i=1}^{n-1} u_i$.

- 1) a) Calculer u_2 et u_3 en fonction de u_1 .
b) Montrer que, pour tout n au moins égal à 3, on a : $nu_n - (n+1)u_{n-1} = 2n - 2$.
- 2) Pour tout entier naturel k non nul, on pose : $v_k = \frac{u_k}{k+1}$.
a) Pour tout n au moins égal à 3, exprimer $v_n - v_{n-1}$ en fonction de n .
b) Déterminer deux réels α et β vérifiant, pour tout réel x non nul et distinct de -1 , l'égalité :
$$\frac{2x-2}{x(x+1)} = \frac{\alpha}{x} + \frac{\beta}{x+1}$$

c) Pour tout n , établir l'égalité : $v_n = 2 \sum_{k=2}^n \frac{1}{k} + \frac{u_1}{3} - 2 + \frac{4}{n+1}$.
- 3) Pour tout n , on pose $h_n = \sum_{k=2}^n \frac{1}{k}$ et $z_n = \frac{1}{n} - \ln\left(\frac{n}{n-1}\right)$.
a) Calculer u_n en fonction de h_n , u_1 et n .
b) Prouver l'égalité : $h_n = \sum_{k=2}^n z_k + \ln n$.
c) Déterminer la nature de la série de terme général z_n .
d) En déduire un équivalent de h_n quand n tend vers l'infini.
e) Déterminer un équivalent de u_n quand n tend vers l'infini.

Partie II : Étude d'une suite de variables aléatoires

A. On considère un espace probabilisé dont la probabilité est notée $\mathbf{P}$, une variable aléatoire Z , définie sur cet espace, prenant un nombre fini de valeurs réelles notées $z_1, z_2, \dots, z_p$ et un événement A de probabilité non nulle.

On note $\mathbf{E}(Z/A)$ l'espérance de la variable aléatoire Z pour la probabilité conditionnelle sachant A , i.e.

$$\mathbf{E}(Z/A) = \sum_{i=1}^p z_i \mathbf{P}([Z = z_i] / A)$$

Soit $(A_1, A_2, \dots, A_q)$ un système complet d'événements tous de probabilité non nulle. Prouver l'égalité :

$$\mathbf{E}(Z) = \sum_{j=1}^q \mathbf{P}(A_j) \mathbf{E}(Z/A_j)$$

B. Toutes les variables aléatoires considérées dans cette sous-partie sont définies sur un même espace probabilisé dont la probabilité est notée $\mathbf{P}$.

On considère une suite $(I_n)_{n \in \mathbb{N}^*}$ de variables aléatoires telle que, pour tout n non nul, I_n suit la loi uniforme sur l'ensemble $\{1, 2, \dots, n\}$ des entiers compris, au sens large, entre 1 et n .

D'autre part, on considère une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires ayant les propriétés suivantes :

- X_1 est la variable constante égale à 0
- pour tout entier naturel n au moins égal à 2, les lois conditionnelles de X_n sachant $[I_n = 1]$ et de X_n sachant $[I_n = n]$ sont toutes deux égales à la loi de $n - 1 + X_{n-1}$
- pour tout entier naturel n au moins égal à 3 et tout entier i tel que $2 \leq i \leq n - 1$, la loi conditionnelle de X_n sachant $[I_n = i]$ est égale à la loi de $n - 1 + Z_{n,i} + T_{n,i}$ où $Z_{n,i}$ et $T_{n,i}$ sont deux variables aléatoires indépendantes, $Z_{n,i}$ ayant même loi que X_{i-1} et $T_{n,i}$ ayant même loi que X_{n-i} .

Par exemple, on a :

$\mathbf{P}([X_6 = 9]/[I_6 = 1]) = \mathbf{P}([X_5 = 4])$ et aussi $\mathbf{P}([X_6 = 9]/[I_6 = 3]) = \mathbf{P}([Z_{6,3} + T_{6,3} = 4])$ ce qui, compte tenu des hypothèses, s'écrit : $\mathbf{P}([X_6 = 9]/[I_6 = 3]) = \sum_j \mathbf{P}([X_2 = j])\mathbf{P}([X_3 = 4 - j])$, la somme étant étendue aux

valeurs convenables de l'entier j .

- 1) a) Montrer que X_2 est une variable aléatoire presque sûrement constante égale à 1.
b) Établir les égalités : $\mathbf{P}([X_3 = 2]) = \frac{1}{3}$ et $\mathbf{P}([X_3 = 3]) = \frac{2}{3}$. Calculer l'espérance de X_3 qu'on notera U_3 .
- 2) Déterminer la loi de X_4 et calculer son espérance qu'on notera U_4 .
- 3) En procédant par récurrence, montrer que, pour tout entier naturel n non nul, X_n prend, presque sûrement, des valeurs entières inférieures ou égales à $\frac{n(n-1)}{2}$.
- 4) Soit n un entier naturel au moins égal à 2. On note U_n l'espérance de X_n .
a) À l'aide des résultats de la sous-partie A, établir l'égalité : $U_n = n - 1 + \frac{2}{n} \sum_{i=1}^{n-1} U_i$.
b) À l'aide de la partie I, donner l'expression de U_n en fonction de n ainsi qu'un équivalent de U_n quand n tend vers l'infini.
- 5) Pour tout entier naturel n non nul, on note α_n la plus petite valeur (entière) prise par la variable X_n avec une probabilité non nulle.
a) Soit n et k deux entiers naturels, l'entier n étant au moins égal à 3.
Montrer que $\mathbf{P}([X_n = k])$ est nul si et seulement si les nombres $\mathbf{P}([n - 1 + X_{n-1} = k])$ et $\mathbf{P}([n - 1 + Z_{n,i} + T_{n,i} = k])$ (l'entier i variant de 2 à $n - 1$) sont nuls.
En déduire que α_n est au moins égal au minimum des nombres $n - 1 + \alpha_{n-1}$, $n - 1 + \alpha_1 + \alpha_{n-2}$, $n - 1 + \alpha_2 + \alpha_{n-3}$, ..., $n - 1 + \alpha_{n-2} + \alpha_1$.
b) On considère la fonction g définie, pour tout x strictement positif, par $g(x) = x \log_2 x - 2x + 2$.
i) Montrer que g est convexe. Pour tout couple d'entiers (i, n) tel que $2 \leq i \leq n - 1$, en déduire l'inégalité : $g(i) + g(n + 1 - i) \geq 2g\left(\frac{n+1}{2}\right)$.
ii) Pour tout entier naturel n non nul, établir l'inégalité : $g(n+1) - g(n) \leq \log_2(n+1)$. En traitant à part les cas $n = 1$ et $n = 2$, montrer que, pour tout entier naturel n non nul, on a : $g(n+1) - g(n) \leq n - 1$.
- 6) En procédant par récurrence, établir, pour tout entier naturel n non nul, l'inégalité :
$$\alpha_n \geq (n+1) \log_2(n+1) - 2n$$

Partie III : Étude d'un algorithme de tri

A. On considère un entier naturel n non nul et un ensemble $E = \{e_1, e_2, \dots, e_n\}$ où $e_1, e_2, \dots, e_n$ sont des réels vérifiant $e_1 < e_2 < \dots < e_n$. On munit l'ensemble des permutations de E de la probabilité **uniforme** notée $\mathbf{P}$. On considère les n variables aléatoires $T_1, T_2, \dots, T_n$ qui, à toute permutation σ de E , associent les images des éléments de E par σ , i.e. $T_1(\sigma) = \sigma(e_1)$, $T_2(\sigma) = \sigma(e_2)$, ..., $T_n(\sigma) = \sigma(e_n)$, et on note T le vecteur aléatoire $(T_1, T_2, \dots, T_n)$. Pour toute liste $(\alpha_1, \alpha_2, \dots, \alpha_n)$ d'éléments **distincts** de E on a donc

$$\mathbf{P}([T = (\alpha_1, \alpha_2, \dots, \alpha_n)]) = \frac{1}{n!}$$

1) Déterminer la loi de la variable aléatoire T_1 .

2) On suppose n au moins égal à 2. Pour toute permutation σ de E on note $T'_1(\sigma)$ le premier élément de la liste $(\sigma(e_1), \sigma(e_2), \dots, \sigma(e_n))$ inférieur à $\sigma(e_1) = T_1(\sigma)$ si un tel élément existe et $T'_1(\sigma) = 0$ sinon, $T'_2(\sigma)$ le deuxième élément inférieur à $\sigma(e_1)$ si un tel deuxième élément existe et $T'_2(\sigma) = 0$ sinon, etc., $T'_{n-1}(\sigma)$ le $(n-1)$ -ième élément inférieur à $\sigma(e_1)$ si un tel $(n-1)$ -ième élément existe et $T'_{n-1}(\sigma) = 0$ sinon.

Par exemple, si

$n = 4$, $(e_1, e_2, e_3, e_4) = (3, 5, 7, 10)$ et $(\sigma(e_1), \sigma(e_2), \sigma(e_3), \sigma(e_4)) = (7, 10, 5, 3)$

alors $T'_1(\sigma) = 5$, $T'_2(\sigma) = 3$ et $T'_3(\sigma) = 0$.

Soit k un entier vérifiant $1 \leq k \leq n-1$.

a) Combien y a-t-il de listes $(i_1, i_2, \dots, i_k)$ d'entiers vérifiant $2 \leq i_1 < i_2 < \dots < i_k \leq n$?

b) Soit $(\alpha_1, \alpha_2, \dots, \alpha_k)$ une liste d'éléments **distincts** de $\{e_1, e_2, \dots, e_k\}$. Établir l'égalité :

$$\mathbf{P}\left(\bigcap_{j=1}^k [T'_j = \alpha_j] \cap [T_1 = e_{k+1}]\right) = \frac{1}{n \cdot k!}$$

c) En déduire l'égalité :

$$\mathbf{P}\left(\bigcap_{j=1}^k [T'_j = \alpha_j] / [T_1 = e_{k+1}]\right) = \frac{1}{k!}$$

où $\mathbf{P}(A/B)$ désigne la probabilité conditionnelle de A sachant B .

Ainsi la loi conditionnelle de $(T'_1, T'_2, \dots, T'_k)$ sachant $[T_1 = e_{k+1}]$ est uniforme.

B. Dans un programme écrit en langage Pascal on fait les déclarations suivantes :

```
const      n = «entier naturel non nul fixé par l'utilisateur» ;
type tableau = array [1..n] of integer ;
```

Soit T une variable de type tableau qu'on suppose constituée d'éléments distincts. Si deb et fin sont deux entiers tels que $1 \leq deb < fin \leq n$ on dira qu'un élément $T[k]$ de la liste $(T[deb], T[deb+1], \dots, T[fin])$ est à sa place dans le sous-tableau $[T[deb], \dots, T[fin]]$ si les éléments $T[deb], T[deb+1], \dots, T[k-1]$ sont inférieurs à $T[k]$ et si les éléments $T[k+1], T[k+2], \dots, T[fin]$ sont supérieurs à $T[k]$ (bien sûr, si $k = deb$ ou $k = fin$, une seule de ces conditions subsiste).

Ainsi, si $n = 4$ et $T = [3, 1, 7, 5]$ alors $T[3] (=7)$ est à sa place dans le sous-tableau $[T[1], T[2], T[3]] (= [3, 1, 7])$

mais n'est pas à sa place dans le sous-tableau $[T[2], T[3], T[4]] (= [1, 7, 5])$.

On dira que le tableau T est **trié** si chacun de ses éléments est à sa place dans T .

On suppose qu'on dispose d'une procédure (écrite en langage Pascal) dont l'en-tête est

```
Placer(var T : tableau ; deb, fin : integer ; var pl : integer) ;
```

qui ne fait rien si $fin \leq deb$ et qui, si $1 \leq deb < fin \leq n$, effectue, à l'aide de $(fin - deb)$ comparaisons, les opérations suivantes :

i) d'une part, elle ne modifie que le sous-tableau $[T[deb], \dots, T[fin]]$ de sorte que les éléments de ce sous-tableau plus petits que $T[deb]$ «ont glissé» (sans permutation entre-eux) à gauche de $T[deb]$ et les éléments de ce sous-tableau plus grands que $T[deb]$ «ont glissé» (sans permutation entre-eux) à droite de $T[deb]$.

Ainsi $T[deb]$ se retrouve à sa place dans le sous-tableau modifié.

ii) d'autre part, elle met dans la variable pl l'indice i tel que $T[i]$ reçoit, au cours de la procédure, la valeur qui était stockée dans $T[deb]$ avant l'exécution de la procédure.

Par exemple, si $T = [12, 3, 8, 10, 6, 4, 5]$ l'instruction $Placer(T, 3, 6, pl)$ une fois exécutée aura changé T en $[12, 3, 6, 4, 8, 10, 5]$ et affecté la valeur 5 à la variable pl , alors que l'instruction $Placer(T, 3, 4, pl)$ aura laissé T inchangé et affecté la valeur 3 à la variable pl .

Par ailleurs on considère la procédure suivante :

```

procedure   Tri (var  $T$  : tableau ;  $deb, fin$  : integer) ;
var  $pl$  : integer ;
begin
    if  $fin > deb$  then   begin   Placer( $T, deb, fin, pl$ ) ;
                           if  $pl > deb$  then   Tri( $T, deb, pl - 1$ ) ;
                           if  $pl < fin$  then   Tri( $T, pl + 1, fin$ ) ;
                           end ;
    end ;

```

- 1) a) L'entier i étant compris entre 1 et n , quel est l'effet sur la variable T de l'instruction $Tri(T, i, i)$?
b) Dans cette sous-question on suppose qu'initialement $T = [2, 9, 6, 1, 5]$.
Déterminer la «trace» de l'instruction $Tri(T, 1, 5)$ en donnant la liste des procédures **successives** (avec les valeurs de leurs paramètres) qui sont effectuées et en indiquant à chaque fois les affectations des variables pl et T .
c) Expliquer succinctement l'effet et le principe de fonctionnement de la procédure Tri en indiquant, en particulier, pourquoi l'algorithme s'arrête.
- 2) On se place à nouveau dans le contexte probabiliste de la sous-partie A et, si σ est une permutation de E , on affecte la valeur $[\sigma(e_1), \sigma(e_2), \dots, \sigma(e_n)]$ à la variable T de type tableau, i.e. $T[1] := \sigma(e_1)$, $T[2] := \sigma(e_2), \dots$, $T[n] := \sigma(e_n)$. On note $X_n(\sigma)$ le nombre de comparaisons faites lors des différentes exécutions de la procédure *Placer* (et seulement au cours de celles-ci) quand on effectue la procédure $Tri(T, 1, n)$.
a) À l'aide de la sous-partie A, montrer que la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ vérifie les hypothèses de II.B. En déduire un équivalent de l'espérance de X_n lorsque l'entier naturel n tend vers l'infini.
b) Dans le cas où $E = \{1, 2, \dots, n\}$, donner (en le commentant de manière succincte) un exemple de tableau nécessitant $\frac{n(n-1)}{2}$ comparaisons pour être trié.
c) Dans le cas où $E = \{1, 2, \dots, 7\}$, donner de même un exemple de tableau nécessitant 10 comparaisons pour être trié.
d) Déterminer une suite d'entiers n pour lesquels, dans le cas où $E = \{1, 2, \dots, n\}$, il existe un tableau d'éléments nécessitant $g(n+1)$ comparaisons pour être trié.



CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS
DIRECTION DE L'ENSEIGNEMENT
Direction des Admissions et Concours

**ECOLE DES HAUTES ETUDES COMMERCIALES
E.S.C.P. - E.A.P.
ECOLE SUPERIEURE DE COMMERCE DE LYON**

CONCOURS D'ADMISSION SUR CLASSES PREPARATOIRES

OPTION SCIENTIFIQUE

MATHEMATIQUES II

Mercredi 9 Mai 2001, de 8h. à 12h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document ; l'utilisation de toute calculatrice et de tout matériel électronique est interdite.

Seule l'utilisation d'une règle graduée est autorisée.

L'objet du problème est l'étude de quelques aspects de la théorie classique du risque dont le contexte et les notations sont introduits au fur et à mesure.

Dans tout le problème, on considère deux suites de variables aléatoires réelles $(\Delta_n)_{n \in \mathbb{N}^*}$ et $(C_n)_{n \in \mathbb{N}^*}$, définies sur un même espace probabilisé $(\Omega, \mathcal{B}, \mathbf{P})$, vérifiant les conditions suivantes :

- i) les variables aléatoires $\Delta_1, \Delta_2, \dots, \Delta_n, \dots, C_1, C_2, \dots, C_n, \dots$ sont indépendantes,
- ii) les variables aléatoires $\Delta_1, \Delta_2, \dots, \Delta_n, \dots$ sont **strictement** positives et ont toutes la même densité égale sur $]0, +\infty[$ à la densité d'une variable aléatoire exponentielle d'espérance égale à 1,
- iii) les variables aléatoires $C_1, C_2, \dots, C_n, \dots$ ont toutes la même densité qu'une variable aléatoire exponentielle d'espérance égale à c .

On pose $T_0 = 0$ et, pour tout entier naturel n non nul, on note T_n la variable aléatoire définie par :

$$T_n = \sum_{i=1}^n \Delta_i$$

On observera que la suite $(T_n)_{n \in \mathbb{N}}$ est strictement croissante et que, pour tout entier naturel n , on a l'égalité : $\Delta_{n+1} = T_{n+1} - T_n$.

On notera $E(X)$ l'espérance d'une variable aléatoire X définie sur $(\Omega, \mathcal{B}, \mathbf{P})$.

Partie I Étude d'une variable aléatoire

1) Pour tout entier naturel n , déterminer l'espérance et la variance de la variable aléatoire T_n .

2) Soit t un réel positif ou nul.

a) Pour tout entier naturel n strictement supérieur à t , justifier l'inclusion entre événements :

$$[T_n < t] \subset [|T_n - n| \geq n - t]$$

b) À l'aide de l'inégalité de Bienaymé-Tchebychev, en déduire la valeur de $\lim_{n \rightarrow +\infty} \mathbf{P}([T_n < t])$.

c) En déduire que l'événement $\bigcap_{k=1}^{+\infty} [T_k < t]$ est de probabilité nulle.

3) Soit t un réel positif ou nul. Étant donné un élément ω de Ω , on note $N(t)(\omega)$ le plus grand élément de l'ensemble $\{n \in \mathbb{N}; T_n(\omega) \leq t\}$ (qui contient 0) si cet ensemble est fini, et $N(t)(\omega) = 0$ sinon.

On observera que, pour tout entier naturel n non nul, $N(t)$ est égal à n si et seulement si : $T_n \leq t < T_{n+1}$.
Montrer que l'application $N(t)$ est une variable aléatoire réelle vérifiant : $\mathbf{P}([N(t) = 0]) = \mathbf{P}([T_1 > t])$.

4) a) Pour tout entier naturel n non nul, reconnaître la loi de la variable aléatoire T_n .

b) Soit t un réel strictement positif. Pour tout entier naturel n non nul, justifier l'égalité :

$$1 = \sum_{k=0}^n \frac{t^k e^{-t}}{k!} + e^{-t} \int_0^t \frac{(t-u)^n}{n!} e^u du$$

$$\text{En déduire l'égalité : } \mathbf{P}([N(t) \leq n]) = \sum_{k=0}^n \frac{t^k e^{-t}}{k!}.$$

c) Pour tout réel t positif ou nul, reconnaître la loi de la variable aléatoire $N(t)$.

Partie II Étude de la probabilité d'être en déficit après le premier ou le second sinistre

Dans cette partie on considère deux réels a et r , r étant strictement positif et, pour tout réel positif t , on note $K_a(t)$ la variable aléatoire définie par l'égalité : $K_a(t) = a + rt - \sum_{i=1}^{N(t)} C_i$ en convenant que la somme $\sum_{i=1}^{N(t)} C_i$ est nulle lorsque $N(t)$ est nul.

En particulier, $K_a(T_0) = K_a(0) = a$ et, pour tout entier naturel n non nul, puisque $N(T_n) = n$,

$$K_a(T_n) = a + rT_n - \sum_{i=1}^n C_i$$

Par exemple, $K_a(t)$ pourrait représenter le capital (aléatoire) au temps t d'une compagnie d'assurance disposant d'un capital initial (de montant a éventuellement négatif), percevant des primes (de montant égal à r par unité de temps), et indemnisant des assurés victimes de sinistres de coûts aléatoires (les C_i) survenant à des dates elles-mêmes aléatoires (les T_i).

Dans cette partie, le réel a étant fixé, la variable aléatoire $K_a(t)$ sera notée plus simplement $K(t)$.

1) a) Déterminer une densité de probabilité de la variable aléatoire $-r\Delta_1$.

b) Déterminer une densité de probabilité f , continue sur $\mathbb{R}$, de la variable aléatoire $L_1 = C_1 - r\Delta_1$.

c) En déduire l'expression de la fonction de répartition F de la variable L_1 puis l'égalité :

$$\mathbf{P}([K(T_1) < 0]) = \begin{cases} 1 - \frac{r}{c+r} \exp\left(\frac{a}{r}\right) & \text{si } a \leq 0 \\ \frac{c}{c+r} \exp\left(\frac{-a}{c}\right) & \text{si } a > 0 \end{cases}$$

2) On pose $L_2 = C_2 - r\Delta_2$ et on considère la fonction g associant à tout réel x le réel

$$g(x) = \mathbf{P}([L_1 \leq x] \cap [L_1 + L_2 \leq a])$$

a) Pour tout réel h strictement positif, justifier les inégalités :

$$g(x+h) - g(x) \geq \mathbf{P}([x < L_1 \leq x+h])\mathbf{P}([L_2 \leq a-x-h])$$

et

$$g(x+h) - g(x) \leq \mathbf{P}([x < L_1 \leq x+h])\mathbf{P}([L_2 < a-x])$$

b) En déduire que la fonction g est dérivable à droite sur $\mathbb{R}$ avec, pour tout réel x , $g'_d(x) = f(x)F(a-x)$.

On admet que g est dérivable sur $\mathbb{R}$ avec, pour tout réel x , $g'(x) = f(x)F(a-x)$.

3) a) Prouver l'égalité

$$\mathbf{P}([L_1 \leq a] \cap [L_1 + L_2 \leq a]) = \lim_{n \rightarrow +\infty} \mathbf{P}([-n < L_1 \leq a] \cap [L_1 + L_2 \leq a])$$

b) En déduire l'égalité :

$$\mathbf{P}([L_1 \leq a] \cap [L_1 + L_2 \leq a]) = \int_{-\infty}^a f(x)F(a-x) dx$$

c) Établir les égalités :

$$\mathbf{P}([K(T_1) < 0] \cup [K(T_2) < 0]) = 1 - \int_{-\infty}^a f(x) F(a-x) dx$$

et

$$\mathbf{P}([K(T_1) < 0] \cup [K(T_2) < 0]) = \mathbf{P}([L_1 > a]) + \int_{-\infty}^a f(x) \mathbf{P}([L_2 > a-x]) dx$$

4) En déduire, dans le cas où a est un réel positif ou nul, l'égalité :

$$\mathbf{P}([K(T_1) < 0] \cup [K(T_2) < 0]) = \frac{c}{c+r} \left(1 + \frac{a}{c+r} + \frac{rc}{(c+r)^2} \right) \exp\left(\frac{-a}{c}\right)$$

Partie III Étude de la probabilité d'être en déficit au cours du temps : deux premiers cas

Dans cette partie, le réel a n'étant plus nécessairement fixé, on utilisera la notation $K_a(t)$.

Pour tout réel a , on note $\Pi(a)$ la probabilité suivante :

$$\Pi(a) = \mathbf{P}\left(\bigcup_{n=0}^{+\infty} [K_a(T_n) < 0]\right)$$

Dans le contexte décrit plus haut, $\Pi(a)$ représenterait la probabilité que la compagnie d'assurance (disposant d'un capital initial de montant a) soit en déficit après un sinistre. En particulier $\Pi(a) = 1$ si $a < 0$.

- 1) Montrer que la fonction Π est décroissante.
- 2) Pour tout réel a , quelles minoration de $\Pi(a)$ peut-on déduire de la partie II ?
- 3) On admet que la fonction Π est continue sur $\mathbb{R}_+$ et vérifie, pour tout réel a positif ou nul l'égalité :

$$\Pi(a) = \mathbf{P}([L_1 > a]) + \int_{-\infty}^a f(x) \Pi(a-x) dx$$

Pourquoi, intuitivement, peut-on conjecturer cette égalité ?

- 4) Soit a un réel et n un entier naturel.
 - a) Calculer l'espérance de $K_a(T_n)$ en fonction de n , a , c et r . Trouver sa limite quand n tend vers l'infini, selon les valeurs comparées de c et r .
 - b) Calculer la variance de $K_a(T_n)$ en fonction de n , r et c .
- 5) Dans cette question, on suppose que c est strictement plus grand que r et on considère un réel a positif ou nul.
 - a) Pour tout entier n strictement supérieur à $\frac{a}{c-r}$, établir l'inégalité :

$$\mathbf{P}([K_a(T_n) < 0]) \geq 1 - \frac{n(c^2 + r^2)}{(a + nr - nc)^2}$$

- b) En déduire l'égalité : $\Pi(a) = 1$.
- 6) Dans cette question, on suppose que c est égal à r et on considère un réel a positif ou nul.
 - a) Soit y un nombre réel. En remarquant que, pour tout entier naturel n non nul, on a l'égalité

$$K_a(T_n) = a - \sum_{i=1}^n (C_i - r\Delta_i)$$

et, à l'aide du théorème de la limite centrée, exprimer le réel $\lim_{n \rightarrow +\infty} \mathbf{P}([K_a(T_n) \leq a + y\sqrt{n}])$, en utilisant la fonction de répartition de la loi normale centrée réduite.

- b) Pour tout nombre réel y strictement positif fixé, établir, pour tout entier naturel n assez grand, la double inégalité :

$$\mathbf{P}([K_a(T_n) \leq a - y\sqrt{n}]) \leq \mathbf{P}([K_a(T_n) < 0]) \leq \mathbf{P}([K_a(T_n) \leq a + y\sqrt{n}])$$

- c) En déduire la limite de la probabilité $\mathbf{P}([K_a(T_n) < 0])$ quand n tend l'infini puis l'inégalité : $\Pi(a) \geq \frac{1}{2}$.

Partie IV Étude de la probabilité d'être en déficit au cours du temps : le dernier cas

Dans cette partie, on suppose que c est strictement plus petit que r .

1) a) En procédant par récurrence, établir, pour tout entier naturel n et tout réel a positif ou nul, l'inégalité :

$$\Pi(a) \geq \frac{c}{c+r} \exp\left(\frac{-a}{c}\right) \sum_{k=0}^n \frac{a^k}{k!(c+r)^k}$$

b) En déduire, pour tout réel a positif ou nul, la minoration :

$$\Pi(a) \geq \frac{c}{c+r} \exp\left(\frac{-ar}{c(c+r)}\right)$$

2) a) Montrer que pour tout réel positif λ vérifiant $\lambda < \frac{1}{c}$, la variable $\exp(\lambda L_1)$ possède une espérance qu'on calculera.

b) Soit n un entier naturel non nul. On pose $S_n = \sum_{k=1}^n (C_k - r\Delta_k)$. Pour tout réel positif λ vérifiant $\lambda < \frac{1}{c}$, justifier l'égalité :

$$\mathbf{E}(\exp(\lambda S_n)) = \frac{1}{(1+r\lambda)^n(1-c\lambda)^n}$$

3) a) Pour tout réel positif λ vérifiant $\lambda < \frac{1}{c}$, tout réel a positif ou nul et tout entier naturel n non nul, établir l'inégalité :

$$\mathbf{P}([S_n > a]) \leq e^{-\lambda a} \mathbf{E}(\exp(\lambda S_n))$$

b) En déduire que tout réel λ élément de $]0, \frac{1}{c} - \frac{1}{r}[$, la série de terme général $\frac{1}{(1+r\lambda)^n(1-c\lambda)^n}$ converge et qu'on a l'inégalité :

$$\mathbf{P}\left(\bigcup_{n=1}^{+\infty} [S_n > a]\right) \leq e^{-\lambda a} \sum_{n=1}^{+\infty} \frac{1}{(1+r\lambda)^n(1-c\lambda)^n}$$

4) En remarquant que, pour tout réel a positif ou nul, $\Pi(a) = \mathbf{P}\left(\bigcup_{n=1}^{+\infty} [S_n > a]\right)$, établir les résultats suivants :

i) $\lim_{a \rightarrow +\infty} \Pi(a) = 0$,

ii) Pour tout réel λ vérifiant $0 < \lambda < \frac{1}{c} - \frac{1}{r}$, et tout réel a assez grand, on a l'inégalité : $\Pi(a) \leq e^{-\lambda a}$
(on introduira un réel μ vérifiant $\lambda < \mu < \frac{1}{c} - \frac{1}{r}$).

5) a) Montrer que si une fonction ψ est continue sur $\mathbb{R}_+$ et de limite nulle en $+\infty$, alors la fonction $|\psi|$ a un maximum sur $\mathbb{R}_+$.

b) Soit Π_1 et Π_2 deux fonctions continues sur $\mathbb{R}_+$, toutes deux de limite nulle en $+\infty$, vérifiant pour tout réel a positif ou nul, les égalités :

$$\Pi_1(a) = \mathbf{P}([L_1 > a]) + \int_{-\infty}^a f(x) \Pi_1(a-x) dx \quad \text{et} \quad \Pi_2(a) = \mathbf{P}([L_1 > a]) + \int_{-\infty}^a f(x) \Pi_2(a-x) dx$$

Montrer que les fonctions Π_1 et Π_2 coïncident sur $\mathbb{R}_+$.

c) Établir, pour tout réel a positif ou nul, l'égalité suivante :

$$\Pi(a) = \frac{c}{r} \exp\left(-a\left(\frac{1}{c} - \frac{1}{r}\right)\right)$$



CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS
DIRECTION DE L'ENSEIGNEMENT
Direction des Admissions et Concours

**ECOLE DES HAUTES ETUDES COMMERCIALES
E.S.C.P. – E.A.P.
ECOLE SUPERIEURE DE COMMERCE DE LYON**

CONCOURS D'ADMISSION SUR CLASSES PREPARATOIRES

OPTION SCIENTIFIQUE

MATHEMATIQUES II

Lundi 13 Mai 2002, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite.

Seule l'utilisation d'une règle graduée est autorisée.

Les variables aléatoires qui interviennent dans ce problème sont toutes définies sur un même espace probabilisé $(\Omega, \mathcal{B}, \mathbf{P})$ et à valeurs réelles. L'espérance d'une variable aléatoire X est notée $\mathbf{E}(X)$.

On admet les résultats suivants :

- i) si X et Y sont deux variables aléatoires possédant une espérance et vérifiant l'inégalité $X \leq Y$ (c'est-à-dire vérifiant $X(\omega) \leq Y(\omega)$ pour tout élément ω de Ω) alors on a l'inégalité : $\mathbf{E}(X) \leq \mathbf{E}(Y)$.
- ii) Étant donné une fonction f continue sur $[0, +\infty[$ et une variable aléatoire Y possédant une densité φ continue sur $[0, +\infty[$ et nulle sur $]-\infty, 0[$, si l'intégrale $\int_0^{+\infty} f(u) \varphi(u) du$ converge absolument alors la variable aléatoire $f(Y)$ possède une espérance vérifiant $\mathbf{E}(f(Y)) = \int_0^{+\infty} f(u) \varphi(u) du$.

Partie I Définition de l'application L

On note E l'ensemble des fonctions f réelles définies, continues sur $[0, +\infty[$ et telles que, pour tout réel x strictement positif, l'intégrale $\int_0^{+\infty} e^{-xt} f(t) dt$ converge absolument.

- 1) a) Vérifier que E est un espace vectoriel réel.
b) Vérifier que E contient les fonctions continues et bornées sur $[0, +\infty[$.
- 2) Pour tout élément f de E on note $L(f)$ la fonction définie, pour tout réel x strictement positif, par :

$$L(f)(x) = \int_0^{+\infty} e^{-xt} f(t) dt$$

- a) Vérifier que L est une application linéaire de E dans l'espace vectoriel des fonctions de $]0, +\infty[$ dans $\mathbb{R}$.
- b) Pour tout réel λ positif ou nul, on note ε_λ la fonction réelle définie par $\varepsilon_\lambda(t) = e^{-\lambda t}$ pour tout réel t positif ou nul. Vérifier que, pour tout réel λ positif ou nul, la fonction ε_λ est dans E et, pour tout réel x strictement positif, calculer $L(\varepsilon_\lambda)(x)$.
- c) Montrer que, pour tout réel λ positif ou nul et toute fonction f de E , la fonction $\varepsilon_\lambda f$ est aussi dans E et vérifie, pour tout réel x strictement positif, l'égalité : $L(\varepsilon_\lambda f)(x) = L(f)(x + \lambda)$.
- 3) On considère une fonction H élément de E , de classe C^1 , croissante et bornée sur $[0, +\infty[$. Montrer que la fonction H' est aussi dans E et, pour tout réel x strictement positif, justifier l'égalité :

$$L(H')(x) = -H(0) + xL(H)(x)$$

- 4) Soit une fonction f élément de E . Pour tout entier naturel n , montrer que la fonction qui à tout réel t positif ou nul associe $t^n f(t)$ est aussi élément de E .

Partie II Dérivabilité de la fonction $L(f)$

Dans toute cette partie on considère un réel x strictement positif et une fonction f élément de E .

1) Soit h un réel non nul vérifiant l'inégalité $|h| < \frac{x}{2}$.

- a) Pour tout réel t strictement positif, justifier l'inégalité : $\left| e^{-(x+h)t} - e^{-xt} + h t e^{-xt} \right| \leq \frac{h^2 t^2}{2} e^{-xt/2}$.
- b) Pour tout réel T strictement positif, justifier l'inégalité :

$$\left| \int_0^T \left(\frac{e^{-(x+h)t} - e^{-xt}}{h} f(t) + t e^{-xt} f(t) \right) dt \right| \leq \frac{|h|}{2} \int_0^{+\infty} t^2 |f(t)| e^{-xt/2} dt$$

c) En déduire que $L(f)$ est dérivable en x et que son nombre dérivé en x vaut :

$$(L(f))'(x) = - \int_0^{+\infty} t f(t) e^{-xt} dt$$

d) Montrer que la fonction $L(f)$ est indéfiniment dérivable sur $]0, +\infty[$ et, pour tout entier naturel k , donner à l'aide d'une intégrale la valeur de la dérivée k -ième de $L(f)$ en x .

Partie III Injectivité de l'application $L: f \mapsto L(f)$

Dans toute cette partie on considère un réel x strictement positif et une fonction f continue et bornée sur l'intervalle $]0, +\infty[$. Ainsi f est élément de E .

1) Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes suivant toutes la loi exponentielle de paramètre égal à $\frac{1}{x}$ (donc d'espérance x). Pour tout entier naturel n , on pose $S_n = \sum_{k=1}^n X_k$.

- a) Donner une densité de la variable aléatoire S_n .
- b) Donner une densité, qu'on notera φ_n , de la variable aléatoire $\frac{S_n}{n}$.

2) a) Soit α un réel strictement positif. Prouver l'égalité :

$$\lim_{n \rightarrow +\infty} \mathbf{P} \left(\left[\left| \frac{S_n}{n} - x \right| > \alpha \right] \right) = 0$$

b) En utilisant la continuité de la fonction f en x , pour tout réel ε strictement positif, justifier l'existence d'un réel α strictement positif tel que, pour tout entier naturel n non nul, on a :

$$\left[\left| f\left(\frac{S_n}{n}\right) - f(x) \right| > \varepsilon \right] \subset \left[\left| \frac{S_n}{n} - x \right| > \alpha \right]$$

c) Soit ε un réel strictement positif. Prouver l'égalité :

$$\lim_{n \rightarrow +\infty} \mathbf{P} \left(\left[\left| f\left(\frac{S_n}{n}\right) - f(x) \right| > \varepsilon \right] \right) = 0$$

3) On note M un majorant de $|f|$ sur $]0, +\infty[$.

a) Soit ε un réel strictement positif. Pour tout entier naturel n , on note A_n l'événement :

$$A_n = \left[\left| f\left(\frac{S_n}{n}\right) - f(x) \right| \leq \varepsilon \right]$$

et $\mathbf{1}_{A_n}$ son indicatrice. Justifier l'inégalité suivante entre variables aléatoires :

$$\left| f\left(\frac{S_n}{n}\right) - f(x) \right| \leq \varepsilon \mathbf{1}_{A_n} + 2M (1 - \mathbf{1}_{A_n})$$

b) En déduire l'égalité :

$$\lim_{n \rightarrow +\infty} \mathbf{E} \left(f\left(\frac{S_n}{n}\right) \right) = f(x)$$

4) a) Dédurre des questions précédentes l'égalité :

$$f(x) = \lim_{n \rightarrow +\infty} \frac{n^n}{(n-1)!x^n} \int_0^{+\infty} t^{n-1} f(t) e^{-nt/x} dt$$

puis l'égalité :

$$f(x) = \lim_{n \rightarrow +\infty} \frac{n^n (-1)^{n-1}}{(n-1)!x^n} \left(L(f) \right)^{(n-1)} \left(\frac{n}{x} \right)$$

- b) Montrer que si deux fonctions f et g continues et bornées sur l'intervalle $[0, +\infty[$ vérifient $L(f) = L(g)$ alors f et g sont égales.
- c) Montrer, plus précisément, que si deux fonctions f et g continues et bornées sur l'intervalle $[0, +\infty[$ vérifient $L(f)(x) = L(g)(x)$ seulement pour tout x dans $]a, +\infty[$ (où a est positif ou nul) alors f et g sont encore égales.

Partie IV Étude du régime permanent d'une file d'attente

Un certain jour des clients arrivent dans une poste ne possédant qu'un seul guichet. Un client qui arrive dans la poste soit se fait servir tout de suite si le guichet est libre, soit prend place dans la file d'attente si le guichet est occupé, se fait servir dès que tous ses prédécesseurs dans la file ont été servis et quitte aussitôt la poste. On modélise cette situation en notant, pour tout entier naturel n non nul, T_n l'instant (aléatoire) d'arrivée dans la poste du n -ième client, U_n sa durée d'attente (aléatoire) dans la file ($U_n = 0$ si le guichet est libre), S_n la durée (aléatoire) de son service au guichet et $W_n = U_n + S_n$ la durée de présence dans la poste.

On pose $T_0 = 0$ et, pour tout entier naturel n non nul, on note $\Delta_n = T_n - T_{n-1}$ et on a alors $T_n = \sum_{k=1}^n \Delta_k$.

On fait les hypothèses suivantes :

- les variables aléatoires $\Delta_1, \Delta_2, \dots, \Delta_n, \dots, S_1, S_2, \dots, S_n, \dots$ sont indépendantes ;
- les variables aléatoires $S_1, S_2, \dots, S_n, \dots$ suivent toutes la loi exponentielle de paramètre μ (d'espérance égale à $\frac{1}{\mu}$) ;
- les variables aléatoires $\Delta_1, \Delta_2, \dots, \Delta_n, \dots$ sont **strictement** positives et ont toutes la même densité égale sur $]0, +\infty[$ à la densité d'une variable aléatoire exponentielle de paramètre λ (d'espérance égale à $\frac{1}{\lambda}$) ;
- l'espérance commune des Δ_i est supérieure à celle des S_i c'est-à-dire : $\mu > \lambda$.

Pour tout entier naturel n non nul, on note F_n la fonction de répartition de U_n et G_n celle de W_n . On **admet** que F_n et G_n sont continues sur $[0, +\infty[$.

Dans les trois premières questions de cette partie on considère un entier n au moins égal à 2, un réel x positif ou nul et un réel h strictement positif.

1) Justifier les égalités : $U_n = 0$ si $W_{n-1} - \Delta_n < 0$ et $U_n = W_{n-1} - \Delta_n$ sinon.

2) Justifier l'indépendance des variables aléatoires W_{n-1} et Δ_n .

3) a) Pour tout entier naturel k , justifier l'inégalité :

$$\mathbf{P}\left(\left[W_{n-1} - \Delta_n \leq x\right] \cap \left[kh \leq \Delta_n < (k+1)h\right]\right) \leq \mathbf{P}\left(\left[W_{n-1} \leq x + (k+1)h\right] \cap \left[kh \leq \Delta_n < (k+1)h\right]\right)$$

puis l'inégalité :

$$\mathbf{P}\left(\left[W_{n-1} - \Delta_n \leq x\right] \cap \left[kh \leq \Delta_n < (k+1)h\right]\right) \leq e^{\lambda h} \int_{(k+1)h}^{(k+2)h} \lambda e^{-\lambda s} G_{n-1}(x+s) ds$$

b) Pour tout entier naturel k non nul, justifier l'inégalité :

$$\mathbf{P}\left(\left[W_{n-1} \leq x + kh\right] \cap \left[kh \leq \Delta_n < (k+1)h\right]\right) \leq \mathbf{P}\left(\left[W_{n-1} - \Delta_n \leq x\right] \cap \left[kh \leq \Delta_n < (k+1)h\right]\right)$$

puis l'inégalité :

$$e^{-\lambda h} \int_{(k-1)h}^{kh} \lambda e^{-\lambda s} G_{n-1}(x+s) ds \leq \mathbf{P}\left(\left[W_{n-1} - \Delta_n \leq x\right] \cap \left[kh \leq \Delta_n < (k+1)h\right]\right)$$

c) En déduire l'encadrement :

$$e^{-\lambda h} \int_0^{+\infty} \lambda e^{-\lambda s} G_{n-1}(x+s) \, ds \leq F_n(x) \leq e^{\lambda h} \int_h^{+\infty} \lambda e^{-\lambda s} G_{n-1}(x+s) \, ds$$

4) Soit x un réel positif ou nul. En utilisant l'encadrement précédent, établir l'égalité :

$$F_n(x) = \int_0^{+\infty} \lambda e^{-\lambda s} G_{n-1}(x+s) \, ds$$

En raisonnant de la même façon on montrerait et on **admettra** l'égalité :

$$G_n(x) = \int_0^x \mu e^{-\mu s} F_{n-1}(x-s) \, ds$$

5) On fait désormais, et jusqu'à la fin du problème, l'hypothèse que les fonctions F_n et G_n sont indépendantes de n et on note F et G les fonctions vérifiant, pour tout entier naturel n non nul, $F = F_n$ et $G = G_n$.
On dit alors qu'on étudie la file d'attente en régime permanent.

a) Pour tout réel x positif ou nul, établir l'égalité :

$$F(x) = \lambda e^{\lambda x} \left(\int_0^{+\infty} e^{-\lambda t} G(t) \, dt - \int_0^x e^{-\lambda t} G(t) \, dt \right)$$

b) En déduire, pour tout réel x positif ou nul, l'égalité :

$$e^{-\lambda x} F(x) = F(0) - \lambda \int_0^x e^{-\lambda t} G(t) \, dt$$

6) a) Montrer que la fonction $H: x \mapsto \int_0^x e^{-\lambda t} G(t) \, dt$ est de classe C^1 , croissante et bornée sur $[0, +\infty[$.

Pour tout réel x strictement positif, établir l'égalité : $xL(H)(x) = L(G)(x + \lambda)$.

b) Montrer que pour tout réel x vérifiant $x > \lambda$, on a l'égalité :

$$L(F)(x) = \frac{F(0)}{x - \lambda} - \frac{\lambda}{x - \lambda} L(G)(x)$$

7) Montrer que la fonction G est de classe C^1 , croissante et bornée sur $[0, +\infty[$. Pour tout réel x strictement positif, établir successivement les égalités :

$$G'(x) = -\mu G(x) + \mu F(x) \quad \text{et} \quad L(G)(x) = \frac{\mu}{x + \mu} L(F)(x)$$

8) a) Pour tout réel x vérifiant $x > \lambda$, justifier l'égalité :

$$L(F)(x) = \frac{F(0)}{\mu - \lambda} \left(\frac{\mu}{x} - \frac{\lambda}{x + \mu - \lambda} \right)$$

b) Pour tout réel x positif ou nul, en déduire l'égalité :

$$F(x) = \frac{F(0)}{\mu - \lambda} \left(\mu - \lambda e^{-(\mu - \lambda)x} \right)$$

c) Justifier que la fonction F admet la limite 1 en $+\infty$ et en déduire, pour tout réel x positif ou nul, l'égalité :

$$F(x) = 1 - \frac{\lambda}{\mu} e^{-(\mu - \lambda)x}$$

9) a) Montrer que, en régime permanent, le temps passé dans la poste suit une loi exponentielle de paramètre égal à $\mu - \lambda$.

b) On suppose qu'un autre jour les arrivées des clients sont en moyenne deux fois plus fréquentes et la durée de service deux fois plus rapide. Que deviennent, en régime permanent, le temps moyen passé dans la poste par un client et la probabilité d'être servi tout de suite ?



CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS
Direction de l'Enseignement

DIRECTION DES ADMISSIONS ET CONCOURS

H.E.C.
E.S.C.P. – E.A.P.
E.M. LYON

OPTION SCIENTIFIQUE

MATHEMATIQUES II

Mardi 13 Mai 2003, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite.

Seule l'utilisation d'une règle graduée est autorisée.

Toutes les variables aléatoires qui interviennent dans ce problème sont définies sur un même espace probabilisé $(\Omega, \mathcal{B}, \mathbf{P})$ et à valeurs réelles. L'espérance d'une variable aléatoire X est notée $\mathbf{E}(X)$. Si A est un événement de probabilité non nulle on note $\mathbf{P}(E|A)$ la probabilité conditionnelle sachant A de l'événement E .

Si n est un entier naturel non nul et si $x_1, \dots, x_n$ sont n réels on note $\min(x_1, \dots, x_n)$ ou $\min_{1 \leq i \leq n} x_i$ le plus petit d'entre eux.

On rappelle que deux variables aléatoires X et Y prenant des valeurs **positives ou nulles** sont indépendantes si et seulement si, pour tout couple (a, b) de réels positifs ou nuls, on a :

$$\mathbf{P}([X \leq a] \cap [Y \leq b]) = \mathbf{P}([X \leq a])\mathbf{P}([Y \leq b])$$

On rappelle qu'une variable aléatoire X prenant des valeurs positives ou nulles suit une loi exponentielle si et seulement si elle vérifie la propriété, dite d'absence de mémoire :

$$\forall (x, y) \in \mathbb{R}_+^2 \quad \mathbf{P}([X > x + y] | [X > x]) = \mathbf{P}([X > y])$$

L'objet du problème est l'obtention de diverses caractérisations de la loi exponentielle.

Partie I Un résultat d'analyse

On considère une fonction réelle φ continue sur $[0, 1]$. On note M le maximum de la fonction $|\varphi|$ sur $[0, 1]$.

Pour tout entier naturel n non nul et tout réel v de $[0, 1]$, on note $Y_{n,v}$ une variable aléatoire suivant la loi binomiale de paramètres n et v .

1) Soit n un entier naturel non nul, x un réel de $]0, 1[$, ε un réel strictement positif vérifiant les inégalités :

$$0 < x - \varepsilon < x < x + \varepsilon < 1$$

a) Comparer, pour tout réel v de $[x + \varepsilon, 1]$, les événements $[Y_{n,v} \leq nx]$ et $[|Y_{n,v} - nv| \geq n(v - x)]$ et en déduire les inégalités :

$$\mathbf{P}([Y_{n,v} \leq nx]) \leq \frac{v(1-v)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

b) Justifier d'une façon analogue, pour tout réel v de $[0, x - \varepsilon]$, l'inégalité :

$$\mathbf{P}([Y_{n,v} > nx]) \leq \frac{1}{4n\varepsilon^2}$$

c) Établir les inégalités :

$$\left| \int_{x+\varepsilon}^1 \varphi(v) P\left([Y_{n,v} \leq nx]\right) dv \right| \leq \frac{M(1-x)}{4n\varepsilon^2} \quad \text{et} \quad \left| \int_0^{x-\varepsilon} \varphi(v) \left(1 - P\left([Y_{n,v} \leq nx]\right)\right) dv \right| \leq \frac{Mx}{4n\varepsilon^2}$$

d) En déduire l'inégalité :

$$\left| \int_0^x \varphi(v) dv - \int_0^1 \varphi(v) P\left([Y_{n,v} \leq nx]\right) dv \right| \leq \left(\frac{1}{4n\varepsilon^2} + 2\varepsilon \right) M$$

2) Établir que, pour tout réel x de $]0, 1[$, on a, pour tout entier naturel n assez grand, l'inégalité :

$$\left| \int_0^x \varphi(v) dv - \int_0^1 \varphi(v) P\left([Y_{n,v} \leq nx]\right) dv \right| \leq \frac{9M}{4\sqrt[3]{n}}$$

3) On suppose maintenant que la fonction φ vérifie, pour tout entier naturel n , $\int_0^1 \varphi(v) v^n dv = 0$.

a) Justifier, pour tout polynôme P à coefficient réels, l'égalité : $\int_0^1 \varphi(v) P(v) dv = 0$.

b) Dédurre des questions précédentes que, pour tout réel x de $]0, 1[$, on a l'égalité : $\int_0^x \varphi(v) dv = 0$.

c) Montrer que la fonction φ est nulle.

Ainsi, on a montré dans cette partie que si φ est une fonction continue sur $[0, 1]$ vérifiant pour tout entier naturel n , $\int_0^1 \varphi(v) v^n dv = 0$, alors φ est nulle.

Dans toute la suite du problème, on considère une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires indépendantes, positives ou nulles, admettant toutes la même densité (nulle sur l'intervalle $] -\infty, 0[$) dont on note f la restriction à l'intervalle $[0, +\infty[$. On suppose que la fonction f est continue et strictement positive sur $[0, +\infty[$. On note F la restriction à l'intervalle $[0, +\infty[$ de la fonction de répartition commune à toutes ces variables. On suppose de plus que X_1 (et donc chaque variable X_i) admet une espérance.

Partie II Caractérisations de la loi exponentielle à l'aide du minimum d'un n -échantillon

Pour tout entier naturel n non nul, on note I_n l'application définie, pour tout ω de Ω , par $I_n(\omega) = \min_{1 \leq i \leq n} X_i(\omega)$ et on admet que I_n est une variable aléatoire qui admet une espérance.

1) Déterminer à l'aide de F , pour tout entier naturel n non nul, la fonction de répartition de I_n .

2) Dans cette question, on suppose que la loi de X_1 (qui est la loi commune à tous les X_i) est exponentielle de paramètre λ strictement positif.

a) Montrer que pour tout entier naturel n non nul, la variable nI_n a même loi que X_1 .

b) Déterminer, pour tout entier naturel n non nul, l'espérance de I_n .

L'objet des questions suivantes est d'établir que chacune de ces propriétés est caractéristique de la loi exponentielle.

3) Dans cette question, on suppose que, pour tout entier naturel n non nul, nI_n a même loi que X_1 .

a) Établir, pour tout entier naturel n non nul et tout réel x positif ou nul, l'égalité :

$$F(x) = 1 - \left(1 - F\left(\frac{x}{n}\right) \right)^n$$

b) Déterminer, pour tout réel x positif ou nul, la valeur de : $\lim_{n \rightarrow +\infty} n \ln \left(1 - F\left(\frac{x}{n}\right) \right)$.

c) Montrer que la loi de X_1 est exponentielle de paramètre $F'(0)$.

4) On revient au cas général.

- a) Montrer que la fonction F réalise une bijection de $[0, +\infty[$ sur $[0, 1[$. On note F^{-1} sa réciproque.
 b) À l'aide d'un changement de variable, établir, pour tout entier naturel n non nul, l'égalité :

$$E(I_n) = n \int_0^1 F^{-1}(u) (1-u)^{n-1} du$$

c) Établir, pour tout réel u de $[0, 1[$, les inégalités :

$$0 \leq (1-u) F^{-1}(u) \leq \int_u^1 F^{-1}(t) dt$$

En déduire que la fonction G définie sur $[0, 1]$ par $G(u) = (1-u) F^{-1}(u)$ si u est élément de $[0, 1[$ et par $G(1) = 0$ est continue.

Établir, pour tout entier n au moins égal à 2, les égalités :

$$E(I_n) = n \int_0^1 G(u) (1-u)^{n-2} du \quad \text{et} \quad E(I_n) = n \int_0^1 G(1-v) v^{n-2} dv$$

d) On suppose maintenant qu'il existe un réel λ strictement positif tel que, pour tout entier naturel n non nul, l'espérance de I_n est égale à $\frac{1}{n\lambda}$.

On note F_λ la restriction à $[0, +\infty[$ de la fonction de répartition de la loi exponentielle de paramètre λ et G_λ la fonction définie sur $[0, 1]$ par $G_\lambda(u) = (1-u) F_\lambda^{-1}(u)$ si u est élément de $[0, 1[$ et par $G_\lambda(1) = 0$.

i) Quelle est, pour tout entier naturel n au moins égal à 2, la valeur de $n \int_0^1 G_\lambda(1-v) v^{n-2} dv$?

ii) À l'aide du résultat de la partie I, montrer que G et G_λ sont égales.

iii) En déduire que la loi de X_1 est exponentielle de paramètre λ .

Partie III Caractérisation de la loi exponentielle à l'aide des deux premiers records

On pose $R_1 = X_1$. On note R_2 l'application définie, pour tout élément ω de Ω , par :

$$R_2(\omega) = \begin{cases} X_n(\omega) & \text{si } n \text{ est le plus petit des entiers } k \text{ tels que } X_k(\omega) > X_1(\omega) \\ X_1(\omega) & \text{si un tel entier n'existe pas.} \end{cases}$$

On admet que R_2 est une variable aléatoire.

A. Préliminaire

1) Exprimer l'événement $[R_2 = R_1]$ à l'aide de la suite d'événements $([X_k \leq X_1])_{k \in \mathbb{N}^*}$.

2) Établir, pour tout réel t positif ou nul et pour tout entier naturel n non nul, l'inégalité :

$$P\left(\bigcap_{k=2}^{n+1} [X_k \leq X_1]\right) \leq (F(t))^{n+1} + 1 - F(t)$$

3) Soit ε un réel strictement positif. En choisissant un réel t de façon convenable et à l'aide de l'inégalité précédente, montrer que, pour tout entier n assez grand, on a : $P\left(\bigcap_{k=2}^{n+1} [X_k \leq X_1]\right) \leq 2\varepsilon$.

Comment énoncer le résultat obtenu ?

4) En déduire que, presque sûrement, $R_2 > R_1$.

B. La caractérisation

Pour tout couple (x, y) de réels positifs ou nuls on pose : $\varphi(x, y) = P\left(\left[R_1 \leq x\right] \cap \left[R_2 - R_1 > y\right]\right)$.

1) Soit (x, y) un couple de réels positifs ou nuls et h un réel strictement positif.

a) Justifier l'égalité :

$$\varphi(x+h, y) - \varphi(x, y) = \sum_{j=1}^{+\infty} P\left(\left[x < X_1 \leq x+h\right] \cap \left(\bigcap_{i=2}^j \left[X_i \leq X_1\right]\right) \cap \left[X_{j+1} > y + X_1\right]\right)$$

b) En déduire les inégalités :

$$\frac{F(x+h) - F(x)}{1 - F(x)} (1 - F(x+y+h)) \leq \varphi(x+h, y) - \varphi(x, y) \leq \frac{F(x+h) - F(x)}{1 - F(x+h)} (1 - F(x+y))$$

2) Calculer, pour tout couple (x, y) de réels positifs ou nuls, la limite de $\frac{\varphi(x+h, y) - \varphi(x, y)}{h}$ quand h tend vers 0 par valeurs supérieures et, en admettant que le résultat tient encore pour la limite quand h tend vers 0 par valeurs inférieures, en déduire l'égalité :

$$\frac{\partial \varphi}{\partial x}(x, y) = \frac{f(x)}{1 - F(x)} (1 - F(x+y))$$

3) Dans cette question on suppose que la loi de X_1 est exponentielle de paramètre λ strictement positif.

a) Établir, pour tout couple (x, y) de réels positifs ou nuls, l'égalité : $\varphi(x, y) = (1 - e^{-\lambda x}) e^{-\lambda y}$.

b) En déduire la loi de $R_2 - R_1$ puis l'indépendance des variables R_1 et $R_2 - R_1$.

4) Réciproquement, dans cette question, on suppose que les variables R_1 et $R_2 - R_1$ sont indépendantes et on note G la fonction de répartition de $R_2 - R_1$.

a) Établir, pour tout couple (x, y) de réels positifs ou nuls, l'égalité : $\frac{1 - F(x+y)}{1 - F(x)} = 1 - G(y)$.

b) En déduire que les fonctions G et F sont égales puis, à l'aide de la propriété d'absence de mémoire, montrer que la loi de X_1 est exponentielle.



CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS
Direction de l'Enseignement

**BANQUE COMMUNE D'ÉPREUVES ÉCRITES
POUR LE HAUT ENSEIGNEMENT COMMERCIAL**

Concepteurs :

H.E.C.

E.S.C.P. – E.A.P.

OPTION SCIENTIFIQUE

MATHÉMATIQUES II

Lundi 10 Mai 2004, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite.

Seule l'utilisation d'une règle graduée est autorisée.

Toutes les variables aléatoires qui interviennent dans ce problème sont définies sur un même espace probabilisé $(\Omega, \mathcal{F}, P)$.

L'objet de ce problème est la recherche et l'étude de lois possédant une propriété, dite de *stabilité*, qui intervient dans la modélisation de nombreux phénomènes satisfaisant une certaine invariance d'échelle.

• Soit X une variable aléatoire réelle. On dit qu'une suite $(X_k)_{k \geq 1}$ de variables aléatoires est une *suite de copies* de X si $(X_k)_{k \geq 1}$ est une suite de variables *indépendantes* ayant toutes *même loi* que X .

• On dit qu'une variable aléatoire réelle X suit une loi *stable* si il existe une suite réelle strictement positive $(a_n)_{n \geq 1}$ telle que, pour toute suite $(X_k)_{k \geq 1}$ de copies de X et pour tout entier n supérieur ou égal 1, $X_1 + \dots + X_n$ et $a_n X$ ont même loi. On vérifie facilement l'unicité de la suite $(a_n)_{n \geq 1}$ si X n'est pas nulle presque sûrement. On dira alors que $(a_n)_{n \geq 1}$ est la *suite associée* à la loi de X .

On note $\mathbb{N}^*$ l'ensemble des entiers supérieurs ou égaux à 1 (i.e. $\{1, 2, 3, \dots\}$).

On admettra que

$$\forall A > 0, \quad \arctan A + \arctan \frac{1}{A} = \frac{\pi}{2}$$

où l'expression $\arctan$ désigne la *fonction réciproque* de la restriction de la fonction tangente à $]-\frac{\pi}{2}, \frac{\pi}{2}[$.

I. Un résultat sur certaines suites positives

Soit $(u_n)_{n \geq 1}$ une suite de réels *strictement positifs* vérifiant les deux propriétés suivantes:

- pour tout couple d'entiers $(m, n) \in \mathbb{N}^{*2}$, $u_{mn} = u_m u_n$,
- il existe un réel strictement positif A tel que, pour tout couple $(m, n) \in \mathbb{N}^{*2}$, si $m \leq n$, alors $u_m \leq A u_n$.

On veut montrer qu'il existe un réel positif α tel que, pour tout entier $n \in \mathbb{N}^*$, $u_n = n^\alpha$.

1) Montrer que $u_1 = 1$.

2) Montrer que, pour tout couple $(r, k) \in \mathbb{N}^* \times \mathbb{N}$, $u_{r^k} = u_r^k$.

4) Soient $a, a' > 0$, et $y \in \mathbb{R}^*$. Soient u, u', v, v' quatre réels tels que

$$u + iv = \frac{a'}{\pi((y - ia)^2 + a'^2)} \quad \text{et} \quad u' + iv' = \frac{a}{\pi((y + ia')^2 + a^2)}$$

où i désigne le nombre complexe de module 1 et d'argument $\frac{\pi}{2}$.

a) Montrer que :

$$\forall x \in \mathbb{R}, \quad \frac{aa'}{\pi^2(x^2 + a^2)((x - y)^2 + a'^2)} = \frac{vx + au}{\pi(x^2 + a^2)} + \frac{v'(x - y) + a'u'}{\pi((x - y)^2 + a'^2)} \quad (*)$$

(On multipliera les deux membres de $(*)$ par leur dénominateur commun et on appliquera la question précédente en prenant $z_1 = ia$ et $z_2 = y + ia'$.)

b) On admet les égalités suivantes :

$$u + iv = \frac{a'(y^2 + a'^2 - a^2) + 2iaa'y}{\pi(y^2 + (a + a')^2)(y^2 + (a - a')^2)}$$

$$u' + iv' = \frac{a(y^2 + a^2 - a'^2) - 2iaa'y}{\pi(y^2 + (a + a')^2)(y^2 + (a - a')^2)}$$

Montrer que :

$$u + u' = \frac{a + a'}{\pi(y^2 + (a + a')^2)}.$$

5) Soit $B > 0$. Calculer $\int_{-B}^B \frac{x}{x^2 + a^2} dx$ et $\int_{-B}^B \frac{x - y}{(x - y)^2 + a^2} dx$.

6) Soient Z_a et $Z_{a'}$ deux variables aléatoires *indépendantes* suivant des lois de Cauchy de paramètres respectifs a et a' . Montrer que la valeur de la densité de la loi de $Z_a + Z_{a'}$ au point y est égale à $u + u'$ (cf. question 4). En déduire la loi de $Z_a + Z_{a'}$.

7) En déduire que Z suit une loi stable. Quelle est la suite associée à la loi de Z ?

IV. Les événements exceptionnels

Du fait de la décroissance rapide à l'infini de la fonction densité des variables gaussiennes, celles-ci n'accordent que peu d'importance aux valeurs extrêmes. Aussi, pour inclure, dans un modèle mathématique, l'éventualité de phénomènes extrêmes, on est amené à privilégier des lois dont la fonction densité décroît moins vite à l'infini. Le but de cette partie est d'étudier ce qu'il en est pour la loi de Cauchy.

Dans cette partie, $(X_i)_{i \geq 1}$ est une suite de variables aléatoires *indépendantes* suivant la loi de Cauchy de paramètre 1.

On dira qu'un événement exceptionnel s'est produit avant l'instant n , si il existe un entier k inférieur ou égal à n tel que, pour tout entier i inférieur ou égal à n et différent de k , $|X_k| > 2|X_i|$. Autrement dit, à l'instant n , la variable la plus forte de l'histoire (en valeur absolue) est supérieure au double de chacune des autres variables. On appellera E_n un tel événement. Ainsi,

$$E_n = \bigcup_{k=1}^n \left(\bigcap_{\substack{1 \leq i \leq n \\ i \neq k}} (|X_k| > 2|X_i|) \right)$$

1) Montrer que :

$$P(E_n) = nP\left(\bigcap_{i=2}^n (|X_1| > 2|X_i|)\right).$$

3) Soit $r \in \mathbb{N}^*, r \geq 2$. Montrer qu'il existe un réel α_r tel que, pour tout entier n de la forme r^k , où k est un entier positif, $u_n = n^{\alpha_r}$. Exprimer α_r en fonction de r et de u_r .

4) Soit $(r_1, r_2) \in \mathbb{N}^{*2}$, $r_2 > r_1 \geq 2$. On introduit alors les réels α_{r_1} et α_{r_2} définis selon la question précédente.

a) Soit $k \in \mathbb{N}^*$. Montrer qu'il existe un entier ℓ tel que $r_2^k \leq r_1^\ell < r_2^{k+1}$.

b) En déduire que $(r_2^k)^{\alpha_{r_2}} \leq A(r_2^{k+1})^{\alpha_{r_1}}$ et $(r_2^k)^{\alpha_{r_1}} \leq A(r_2^{k+1})^{\alpha_{r_2}}$.

c) En faisant tendre k vers l'infini, déduire l'égalité $\alpha_{r_1} = \alpha_{r_2}$. Conclure.

II. La loi gaussienne

A. On rappelle l'expression de la densité d'une variable gaussienne de moyenne m et de variance σ^2 :

$$f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

1) Soit a un réel *strictement positif* et b et c deux réels quelconques.

Trouver trois réels α, m, σ , que l'on exprimera en fonction de a, b, c tels que

$$\forall x \in \mathbb{R}, \quad \frac{(x-m)^2}{2\sigma^2} + \alpha = ax^2 + bx + c$$

2) En déduire que $\int_{-\infty}^{+\infty} \exp(-(ax^2 + bx + c)) dx = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2 - 4ac}{4a}\right)$.

3) Soient G et G' deux variables aléatoires gaussiennes *centrées indépendantes* de variances respectives σ^2 et σ'^2 . **Redémontrer** en calculant la densité de la loi de $G + G'$, que $G + G'$ est une variable gaussienne dont on donnera l'espérance et la variance.

4) Montrer que G suit une loi stable. Quelle est la suite associée à la loi de G ?

B. Dans cette section, X est une variable aléatoire qui suit une loi *stable* et qui admet une espérance m et une variance σ^2 *strictement positive*. On ne suppose pas que X suit une loi gaussienne. Soit $(X_k)_{k \geq 1}$ une suite de copies de X et $(a_k)_{k \geq 1}$ la suite associée à la loi de X .

1) En considérant les variances de $X_1 + \dots + X_n$ et de $a_n X$, donner, pour tout entier naturel n non nul, la valeur de a_n . Montrer que $m = 0$.

2) En appliquant le théorème de la limite centrée, montrer que X suit une loi gaussienne.

III. La loi de Cauchy

1) Soit $a > 0$. Vérifier que la fonction $f_a : x \mapsto \frac{a}{\pi(x^2 + a^2)}$ est bien une densité de probabilité. (On utilisera le changement de variable $x = a \tan t$).

On dit qu'une variable aléatoire suit la loi de Cauchy de paramètre a si elle admet la fonction f_a pour densité.

2) Soit Z une variable aléatoire suivant la loi de Cauchy de paramètre *égal* à 1.

a) La variable Z admet-elle une espérance ?

b) Soit $\lambda > 0$. Quelle est la loi de λZ ?

3) Soit P un polynôme de degré inférieur ou égal à 3 à *coefficients réels*. Soient z_1 et z_2 deux nombres complexes *distincts* de partie imaginaire *strictement positive*. Montrer que si z_1 et z_2 sont des racines de P , alors $P = 0$. (On remarquera que $\bar{z}_1$ et $\bar{z}_2$ sont également racines de P .)

2) En déduire que :

$$\forall A > 0, \quad P(E_n) \geq nP\left(|X_1| > 2A \cap \left(\bigcap_{i=2}^n (|X_i| < A)\right)\right).$$

3) Montrer que : $\forall A > 0, P(|X_1| > A) = \frac{2}{\pi} \arctan \frac{1}{A}$.

4) Soit $\lambda > 0$, et n assez grand pour que $\frac{\pi\lambda}{2n} < \frac{\pi}{2}$. En choisissant $A = \frac{1}{\tan \frac{\pi\lambda}{2n}}$, montrer que

$$P(E_n) \geq nP\left(|X_1| > \frac{2}{\tan \frac{\pi\lambda}{2n}}\right)\left(1 - \frac{\lambda}{n}\right)^{n-1}.$$

5) Soit $\varepsilon > 0$ et $\lambda > 0$. Montrer que, pour tout entier n assez grand, $P(E_n) > \frac{\lambda}{2}e^{-\lambda} - \varepsilon$.

6) En déduire que, pour tout entier n assez grand, $P(E_n) > \frac{1}{6}$.

V. Le nombre a_n est une puissance de n

Soit X une variable aléatoire suivant une loi *stable*. Soit $(X_k)_{k \geq 1}$ une suite de copies de X et $(a_k)_{k \geq 1}$ la suite associée à la loi de X .

A. Une variable aléatoire X est dite *symétrique* si elle a la même loi que la variable $-X$. Autrement dit, pour tout intervalle $I \subset \mathbb{R}$, $P(X \in I) = P(-X \in I)$ (exemple : une variable gaussienne centrée).

Dans cette section, on suppose X non nulle et symétrique.

1) Montrer que $P(X > 0) = \frac{1}{2}(1 - P(X = 0))$.

2) Montrer qu'il existe $\mu > 0$ tel que $P(X > \mu) > 0$.

3) a) Montrer que, pour tout couple $(m, n) \in \mathbb{N}^{*2}$, $a_{m+n}X$ a même loi que $a_mX_1 + a_nX_2$.

b) En déduire que, pour tout k -uplet d'entiers $(m_1, \dots, m_k)$, $a_{m_1+\dots+m_k}X$ a même loi que $a_{m_1}X_1 + \dots + a_{m_k}X_k$.

c) En prenant tous les entiers m_i égaux à un même entier ℓ , montrer que $a_{k\ell} = a_k a_\ell$.

4) En considérant l'événement $(X_1 \geq 0) \cap (X_2 > t)$, montrer en utilisant la question V.A.3.a, que pour tout couple $(m, n) \in \mathbb{N}^{*2}$, et pour tout $t > 0$,

$$P(X > \frac{a_n}{a_{m+n}}t) \geq \frac{1}{2}P(X > t).$$

5) En utilisant la question V.A.2., montrer que l'ensemble $\left\{\frac{a_n}{a_{n+m}} : (m, n) \in \mathbb{N}^{*2}\right\}$ est majoré. En déduire l'existence d'un réel α tel que, pour tout entier naturel non nul n , $a_n = n^\alpha$.

B. On suppose que X suit une loi stable à densité, mais on ne suppose plus que X est symétrique.

1) Montrer que la variable $X_1 - X_2$ est symétrique.

2) Montrer que $X_1 - X_2$ suit une loi stable. Soit $(b_n)_{n \geq 1}$ la suite associée à la loi de $X_1 - X_2$. Montrer que, pour tout entier naturel n non nul, $a_n = b_n$. Conclure.



BANQUE COMMUNE D'ÉPREUVES

Concepteurs : H.E.C. – E.S.C.P. – E.A.P.

CODE ÉPREUVE :

283

CCIP_M2_S

OPTION : SCIENTIFIQUE

MATHÉMATIQUES II

Mardi 10 Mai 2005, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite.

Seule l'utilisation d'une règle graduée est autorisée.

Dans tout le problème, n et r désignent des entiers strictement positifs. On note $\mathcal{M}_{n,r}(\mathbb{R})$, l'ensemble des matrices rectangulaires à n lignes et r colonnes à coefficients réels. Pour $n = r$, on pose $\mathcal{M}_n(\mathbb{R}) = \mathcal{M}_{n,n}(\mathbb{R})$. Pour tout entier n de $\mathbb{N}^*$, on identifie $\mathbb{R}^n$ et $\mathcal{M}_{n,1}(\mathbb{R})$.

La transposée d'une matrice A appartenant à $\mathcal{M}_{n,r}(\mathbb{R})$ est notée tA . On pourra également la noter A^T .

On étudie dans ce problème, quelques propriétés du *modèle linéaire*, qui constitue l'instrument de base de l'économétrie.

Partie I. Trace et matrices aléatoires

Pour toute matrice M appartenant à $\mathcal{M}_n(\mathbb{R})$, on appelle *trace de M* , notée $\text{tr}(M)$, la somme de ses coefficients diagonaux ; ainsi, si $M = (m_{i,j})_{1 \leq i,j \leq n}$, $\text{tr}(M) = \sum_{i=1}^n m_{i,i}$.

On rappelle les trois résultats suivants (que les candidats n'ont pas à démontrer) :

- l'application tr qui, à toute matrice M de $\mathcal{M}_n(\mathbb{R})$, associe sa trace, est une application linéaire de $\mathcal{M}_n(\mathbb{R})$ dans $\mathbb{R}$;
- si A est une matrice de $\mathcal{M}_{n,r}(\mathbb{R})$ et B une matrice de $\mathcal{M}_{r,n}(\mathbb{R})$, alors $\text{tr}(AB) = \text{tr}(BA)$;
- si M et N sont deux matrices semblables de $\mathcal{M}_n(\mathbb{R})$, alors $\text{tr}(M) = \text{tr}(N)$.

1. Soit M une matrice de $\mathcal{M}_n(\mathbb{R})$ possédant q valeurs propres ($1 \leq q \leq n$) notées $\lambda_1, \lambda_2, \dots, \lambda_q$. Pour tout entier i de $[1, q]$, on désigne par n_i la dimension du sous-espace propre associé à la valeur propre λ_i .

a) On suppose que la matrice M est diagonalisable sur $\mathbb{R}$. Montrer que $\text{tr}(M) = \sum_{i=1}^q n_i \lambda_i$.

b) On suppose que la matrice $M = (m_{i,j})_{1 \leq i,j \leq n}$ de $\mathcal{M}_n(\mathbb{R})$ est symétrique. Montrer les égalités suivantes :

$$\text{tr}({}^tMM) = \text{tr}(M^2) = \sum_{i=1}^q n_i \lambda_i^2 = \sum_{i=1}^n \sum_{j=1}^n m_{i,j}^2$$

2. Pour tout entier i de $[1, n]$ et pour tout entier j de $[1, r]$, on considère des variables aléatoires réelles $Z_{i,j}$ définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mu)$. On définit la *matrice aléatoire* Z , à n lignes et r colonnes, en associant à tout ω de Ω , la matrice :

$$Z(\omega) = \begin{pmatrix} Z_{1,1}(\omega) & \dots & Z_{1,r}(\omega) \\ \vdots & \ddots & \vdots \\ Z_{n,1}(\omega) & \dots & Z_{n,r}(\omega) \end{pmatrix} = (Z_{i,j}(\omega))_{\substack{1 \leq i \leq n \\ 1 \leq j \leq r}}$$

On suppose que les nr variables aléatoires $Z_{i,j}$ admettent une espérance $E(Z_{i,j})$, et on définit l'espérance de la matrice Z , notée $E(Z)$, comme la matrice de $\mathcal{M}_{n,r}(\mathbb{R})$ dont les éléments sont les espérances $E(Z_{i,j})$, soit $E(Z) = (E(Z_{i,j}))_{\substack{1 \leq i \leq n \\ 1 \leq j \leq r}}$.

Si Z et W sont deux matrices aléatoires à n lignes et r colonnes admettant chacune une espérance, et si λ est réel, on remarquera que $E(\lambda Z + W) = \lambda E(Z) + E(W)$.

Dans le cas où $n = r$, on appelle trace de Z , notée $\text{tr}(Z)$, la variable aléatoire définie par $\text{tr}(Z) = \sum_{i=1}^n Z_{i,i}$ et si $n = r = 1$, la matrice aléatoire Z coïncide avec la variable aléatoire Z et on a $\text{tr}(Z) = Z$.

Dans le cas où $r = 1$ et n est quelconque, si $T = {}^t(T_1 \dots T_n)$ et $W = {}^t(W_1 \dots W_n)$ sont deux vecteurs aléatoires de $\mathbb{R}^n$, et si λ est un réel quelconque, on définit le vecteur aléatoire $\lambda T + W$ de $\mathbb{R}^n$ par

$$\lambda T + W = {}^t(\lambda T_1 + W_1 \dots \lambda T_n + W_n)$$

a) Soit Z une matrice aléatoire à n lignes et r colonnes admettant une espérance $E(Z)$. On considère une matrice A de $\mathcal{M}_{r,n}(\mathbb{R})$. Montrer que $E(AZ) = AE(Z)$. Soit B un élément de $\mathcal{M}_{r,q}(\mathbb{R})$, avec $q \in \mathbb{N}^*$. Montrer que $E(ZB) = E(Z)B$.

b) Soit Z une matrice aléatoire à n lignes et n colonnes admettant une espérance $E(Z)$. Établir les deux égalités : $E({}^tZ) = {}^t(E(Z))$ et $E(\text{tr}(Z)) = \text{tr}(E(Z))$.

3. Dans cette question, Y désigne un vecteur aléatoire de $\mathbb{R}^n$, noté $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$, admettant une espérance $E(Y)$

et une matrice de variance-covariance notée $V(Y)$.

On rappelle que $V(Y) = E[(Y - E(Y)) \times {}^t(Y - E(Y))]$.

On admet que la définition et les propriétés de la matrice de variance-covariance $V(Y)$ d'un vecteur aléatoire discret restent valables pour un vecteur aléatoire dont les composantes sont des variables aléatoires quelconques (discrètes ou à densité).

Ainsi, en supposant que pour tout i de $\llbracket 1, n \rrbracket$ et pour tout j de $\llbracket 1, n \rrbracket$, la variable aléatoire $Y_i Y_j$ possède un moment d'ordre 1 au moins, on définit la covariance de Y_i et Y_j par $\text{cov}(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i)E(Y_j)$, et si Y_i et Y_j sont indépendantes, alors $\text{cov}(Y_i, Y_j) = 0$.

a) Montrer que, pour tout vecteur aléatoire Y de $\mathbb{R}^n$, $V(Y) = E(Y {}^tY) - E(Y)E({}^tY)$.

b) Soit B une matrice de $\mathcal{M}_{r,n}(\mathbb{R})$. Justifier l'égalité $V(BY) = BV(Y){}^tB$.

c) Soit A une matrice de $\mathcal{M}_n(\mathbb{R})$. On pose $m = E(Y)$ et $J = V(Y)$. Établir les égalités : $E({}^tYAY) = \text{tr}(A \cdot E(Y {}^tY))$ et $E({}^tYAY) = \text{tr}(AJ) + {}^t m A m$.

Partie II. Le modèle linéaire

Dans les parties II.A et II.B, n et k sont deux entiers donnés qui vérifient $1 \leq k < n$. L'espace vectoriel $\mathbb{R}^n$ est muni de sa structure euclidienne canonique. Toutes les variables aléatoires considérées sont définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mu)$ et admettent des moments d'ordre au moins 2.

On considère un échantillon de n individus extrait d'une population donnée. Ces individus sont décrits à l'aide de k variables statistiques réelles (caractères) $C_1, C_2, \dots, C_k$.

Pour tout entier j de $\llbracket 1, k \rrbracket$, chaque caractère C_j fait l'objet de n observations notées $x_{1,j}, \dots, x_{n,j}$. On définit ainsi une application linéaire f de $\mathbb{R}^k$ dans $\mathbb{R}^n$, dont la matrice dans les bases canoniques de $\mathbb{R}^k$ et $\mathbb{R}^n$ est la matrice $X = (x_{i,j})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq k}}$ de $\mathcal{M}_{n,k}(\mathbb{R})$. On suppose que le rang de X est égal à k .

Soit $U = \begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix}$ un vecteur aléatoire de $\mathbb{R}^n$, dont les composantes $U_1, \dots, U_n$ sont des variables aléatoires réelles

définies sur $(\Omega, \mathcal{A}, \mu)$, mutuellement indépendantes et de même loi. On suppose que $E(U) = 0_n$ et $V(U) = \sigma^2 I_n$, où 0_n désigne le vecteur nul de $\mathbb{R}^n$, I_n la matrice identité de $\mathcal{M}_n(\mathbb{R})$ et σ un réel strictement positif inconnu.

Soit $\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix}$ un vecteur non nul de $\mathbb{R}^k$ dont les composantes $\alpha_1, \dots, \alpha_k$ sont inconnues (α est un paramètre vectoriel)

On considère un *vecteur aléatoire* non nul, $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$ de $\mathbb{R}^n$ tel que, pour tout i de $\llbracket 1, n \rrbracket$, la variable aléatoire

$$Y_i \text{ définie sur } (\Omega, \mathcal{A}, \mu) \text{ s'écrit } Y_i = \sum_{j=1}^k x_{i,j} \alpha_j + U_i.$$

Sous forme matricielle, le *modèle linéaire* s'écrit $Y = X\alpha + U$. On s'intéresse dans cette partie II, à l'étude de quelques propriétés de ce modèle, liées à l'estimation des paramètres inconnus $\alpha_1, \alpha_2, \dots, \alpha_k$ et σ^2 .

Pour cela, on désigne par y et on note $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, le vecteur non nul de $\mathbb{R}^n$ qui représente la réalisation sur l'échantillon considéré du vecteur aléatoire Y ; ainsi, pour tout i de $\llbracket 1, n \rrbracket$, y_i est la réalisation de la variable aléatoire Y_i .

Soit $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$ le vecteur de $\mathbb{R}^n$, dit *vecteur d'écart*, défini par $u = y - X\alpha$.

A. Quelques résultats algébriques

1. On considère l'endomorphisme h de $\mathbb{R}^k$ dont la matrice H , dans la base canonique de $\mathbb{R}^k$, est définie par $H = {}^tXX$.

a) Montrer que H est une matrice symétrique réelle de $\mathcal{M}_k(\mathbb{R})$.

b) En étudiant le noyau de h , montrer que le rang de h est égal à k . En déduire que la matrice H est inversible. On notera H^{-1} son inverse.

2. Dans cette question, on veut trouver, en fonction de y et X , les vecteurs α de $\mathbb{R}^k$ qui minimisent $\|u\|$.

Montrer que ce problème admet une unique solution $\hat{\alpha}$ définie par $\hat{\alpha} = \begin{pmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_n \end{pmatrix} = H^{-1} {}^tXy$.

3. Soit p le projecteur orthogonal de $\mathbb{R}^n$ sur le sous-espace vectoriel engendré par les colonnes de la matrice X . On note P la matrice de p dans la base canonique de $\mathbb{R}^n$.

a) Montrer que $p(y) = X\hat{\alpha}$. En déduire que $P = XH^{-1} {}^tX$. Vérifier que $P = P^2 = {}^tP$.

b) Établir que le rang de P et la trace de P sont égaux. Quelle est leur valeur commune ?

c) Montrer que les colonnes de X constituent une base de vecteurs propres de la matrice P , associés à la valeur propre 1.

d) Montrer qu'il existe une matrice S de $\mathcal{M}_n(\mathbb{R})$, orthogonale, telle que $P = SD^tS$, où $D = (d_{i,j})_{1 \leq i,j \leq n}$ est la matrice diagonale définie par :

$$\begin{cases} d_{i,i} = 1 & \text{si } 1 \leq i \leq k \\ d_{i,j} = 0 & \text{sinon} \end{cases}$$

Préciser les k premières colonnes de S .

4. Soit $\hat{u}$ le vecteur de $\mathbb{R}^n$ défini par $\hat{u} = y - X\hat{\alpha}$.

a) On pose $Q = I_n - P$. Montrer que $\hat{u} = Qu$. Vérifier que $Q = Q^2 = {}^tQ$. Calculer la trace de Q .

b) Exprimer ${}^t\hat{u}\hat{u}$ et tyQy en fonction de Q et u .

5. Par définition, on dit qu'une matrice A symétrique réelle d'ordre n est *positive*, si pour tout vecteur z de $\mathbb{R}^n$, on a ${}^tAz \geq 0$.

a) Montrer que A , symétrique réelle, est positive si et seulement si ses valeurs propres sont positives ou nulles.

b) Soit L une matrice appartenant à $\mathcal{M}_{n,k}(\mathbb{R})$. Établir que tLL est symétrique réelle positive.

B. Estimation des paramètres $\alpha_1, \alpha_2, \dots, \alpha_k$ et σ^2

1. Soit $\hat{G}$ le vecteur aléatoire de $\mathbb{R}^k$ défini par : $\hat{G} = H^{-1} {}^tXY$.

a) Établir que $E(Y) = X\alpha$, et que $V(Y) = \sigma^2 I_n$. En déduire que $E(\hat{G}) = \alpha$ ($\hat{G}$ est un *estimateur* sans biais de α , tandis que $\hat{\alpha}$ est une *estimation* sans biais de α).

b) Montrer que $V(\hat{G}) = \sigma^2 H^{-1}$.

2. On veut montrer dans cette question, que dans l'ensemble des estimateurs sans biais du paramètre α , de la forme tBY , où B est une matrice quelconque, non nulle de $\mathcal{M}_{n,k}(\mathbb{R})$, l'estimateur $\widehat{G}$ est *optimal* dans le sens suivant : tout autre estimateur G^* sans biais du paramètre α , de la forme tBY est tel que la matrice $V(G^*) - V(\widehat{G})$ est positive.

Soit B une matrice non nulle de $\mathcal{M}_{n,k}(\mathbb{R})$. On considère le vecteur aléatoire $\widehat{C} = {}^tBY$

a) Quelle condition doit satisfaire la matrice B pour que, pour tout vecteur α de $\mathbb{R}^k$, $\widehat{C}$ soit un estimateur sans biais de α ?

b) En supposant cette condition vérifiée, on pose ${}^tF = {}^tB - H^{-1} {}^tX$. Calculer tFX , et montrer que la matrice $V(\widehat{C}) - V(\widehat{G})$ est positive.

3. On désigne par $\widehat{U}$ le vecteur aléatoire de $\mathbb{R}^n$ défini par $\widehat{U} = \begin{pmatrix} \widehat{U}_1 \\ \vdots \\ \widehat{U}_n \end{pmatrix} = Y - X\widehat{G}$.

a) Montrer que $\widehat{U} = QU$.

b) Déterminer $E(\widehat{U})$ et $V(\widehat{U})$. Les variables aléatoires $\widehat{U}_1, \dots, \widehat{U}_n$ sont-elles indépendantes ?

c) Montrer que ${}^t\widehat{U}\widehat{U} = \sum_{i=1}^n \widehat{U}_i^2 = {}^tUQU = {}^tYQY$.

d) Calculer $E({}^t\widehat{U}\widehat{U})$. En déduire que la variable aléatoire s_n définie par $s_n = \frac{{}^t\widehat{U}\widehat{U}}{n-k} = \frac{{}^tYQY}{n-k}$ est un estimateur sans biais de σ^2 .

C. Étude d'une suite d'estimateurs

Dans cette partie, k est fixé dans $\mathbb{N}^*$. On veut montrer que la suite d'estimateurs $(s_n)_{n \geq k+1}$ de σ^2 , est convergente.

On suppose que, pour tout i de $\llbracket 1, n \rrbracket$, la variable aléatoire U_i possède des moments d'ordre 3 et 4 avec $E(U_i^3) = 0$ et $E(U_i^4) = 3\sigma^4$. On pose $Q = (q_{i,j})_{1 \leq i,j \leq n}$.

1. Établir que ${}^tUQU = \sum_{i=1}^n \sum_{j=1}^n q_{i,j} U_i U_j$.

2. Montrer que $E[({}^tUQU)^2] = \sigma^4[(\text{tr}(Q))^2 + 2 \text{tr}(Q^2)]$. En déduire que $E[({}^tUQU)^2] = \sigma^4(n-k)(n-k+2)$.

3. Calculer la variance $V(s_n)$ de la variable aléatoire s_n . Conclure.



BANQUE COMMUNE D'ÉPREUVES

Concepteurs : H.E.C. – E.S.C.P. – E.A.P.

OPTION : SCIENTIFIQUE

MATHÉMATIQUES II

CODE ÉPREUVE :

283

CCIP_M2_S

Mercredi 10 Mai 2006, de 14 h. à 18 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite.

Seule l'utilisation d'une règle graduée est autorisée.

Le problème a pour objet l'étude de quelques propriétés concernant le nombre de racines réelles d'un polynôme de degré n , ($n \geq 1$), à coefficients réels fixés ou aléatoires.

Dans les parties II et III, les polynômes considérés sont à coefficients réels et on pourra confondre polynôme et fonction polynomiale associée.

Pour toute fonction Ψ dérivable sur son domaine de définition, la dérivée de Ψ est notée Ψ' .

Les quatre parties du problème sont, dans une large mesure, indépendantes.

Partie I. Nombre de racines réelles d'un polynôme du second degré à coefficients aléatoires

On considère dans cette partie, deux variables aléatoires réelles X_0 et X_1 définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$, indépendantes et de même loi.

Pour tout ω de Ω , on considère le polynôme Q_ω d'indéterminée y , défini par :

$$Q_\omega(y) = y^2 + X_1(\omega)y + X_0(\omega)$$

On désigne par $M(\omega)$ le nombre de racines réelles de Q_ω .

1. Montrer que l'application M qui, à tout ω de Ω associe $M(\omega)$, est une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$.

2. Soit Z une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$, qui suit une loi de Bernoulli de paramètre p ($p \in]0, 1[$). On suppose dans cette question que X_0 et X_1 suivent la même loi que $2Z - 1$.

a) Déterminer la loi de X_0 .

b) Déterminer la loi de M et calculer son espérance $E(M)$.

Dans les questions suivantes, on suppose que X_0 et X_1 suivent une même loi exponentielle de paramètre 1/2. On pose : $Y_0 = -4X_0$, $Y_1 = X_1^2$, $Y = Y_1 + Y_0$, et on note F_{Y_0} , F_{Y_1} et F_Y , les fonctions de répartition de Y_0 , Y_1 et Y , respectivement.

3. Montrer que l'on a, pour tout x réel :

$$F_{Y_1}(x) = \begin{cases} 1 - e^{-\sqrt{x}/2} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} \quad \text{et} \quad F_{Y_0}(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ e^{x/8} & \text{si } x < 0 \end{cases}$$

En déduire l'expression d'une densité f_{Y_0} de Y_0 et d'une densité f_{Y_1} de Y_1 .

4. Soit g la fonction définie sur $\mathbb{R}^{+*}$ par $g(t) = \frac{1}{\sqrt{t}} \times \exp \left[-\frac{1}{2} \left(\frac{t}{4} + \sqrt{t} \right) \right]$, où $\exp$ désigne la fonction exponentielle.

a) Établir la convergence de l'intégrale impropre $\int_0^{+\infty} g(t) dt$.

b) En déduire qu'une densité f_Y de la variable aléatoire Y est donnée, pour tout x réel, par :

$$f_Y(x) = \begin{cases} \frac{1}{32} e^{x/8} \int_0^{+\infty} g(t) dt & \text{si } x < 0 \\ \frac{1}{32} e^{x/8} \int_x^{+\infty} g(t) dt & \text{si } x \geq 0 \end{cases}$$

5. On désigne par Φ la fonction de répartition d'une variable aléatoire qui suit la loi normale centrée, réduite.

a) Justifier la validité du changement de variable $u = \sqrt{t}$ dans l'intégrale impropre $\int_0^{+\infty} g(t) dt$.

b) En déduire que $\int_0^{+\infty} g(t) dt = 4\sqrt{e} \int_1^{+\infty} e^{-v^2/2} dv$, et donner, pour tout réel x négatif, l'expression de $f_Y(x)$ en fonction de Φ .

c) Montrer que, pour tout réel x positif, on a : $f_Y(x) = \frac{\sqrt{2\pi e}}{8} e^{x/8} \left[1 - \Phi \left(\frac{\sqrt{x}}{2} + 1 \right) \right]$.

d) Déterminer la loi de M et son espérance $E(M)$ (on fera intervenir le nombre $\Phi(1)$).

Partie II. Suites de Sturm

Soit n un entier supérieur ou égal à 1, et soit $P(X) = X^n + a_{n-1}X^{n-1} + \dots + a_1X + a_0$ un polynôme normalisé ($a_n = 1$) donné, à coefficients réels. On suppose que toutes les racines réelles de P sont simples.

L'objectif de cette partie est de décrire un algorithme permettant de déterminer le nombre de racines réelles de P appartenant à un intervalle donné $[a, b]$.

On associe au polynôme P , la suite $(R_i)_{i \geq 0}$ de polynômes définie de la manière suivante : $R_0 = P, R_1 = -P'$, et pour tout entier j tel que $R_{j+1} \neq 0$, le polynôme R_{j+2} est l'opposé du reste de la division euclidienne de R_j par R_{j+1} . Si $R_{j+1} = 0$, on pose $R_{j+2} = 0$.

1. Montrer qu'il existe un entier k ($k \geq 2$), tel que $R_k = 0$. On note R_m , ($m \geq 1$), le dernier polynôme non nul de la suite $(R_i)_{i \geq 0}$.

Dans toute cette partie, on pose :

$$\begin{cases} R_0 = S_1 R_1 - R_2 \\ R_1 = S_2 R_2 - R_3 \\ \vdots \\ R_{m-2} = S_{m-1} R_{m-1} - R_m \\ R_{m-1} = S_m R_m \end{cases}$$

2. a) Montrer que s'il existe un entier j de $\llbracket 0, m-1 \rrbracket$ et un réel x_0 tels que $R_j(x_0) = R_{j+1}(x_0) = 0$, alors $P(x_0) = P'(x_0) = 0$.

b) En déduire que le polynôme R_m n'admet pas de racine réelle.

c) Soit j un entier de $\llbracket 1, m-1 \rrbracket$. Montrer que si x_0 est une racine réelle de R_j , alors $R_{j-1}(x_0) \times R_{j+1}(x_0) < 0$.

3. Soit $s = (s_1, s_2, \dots, s_t)$ une t -liste ($t \geq 2$) de nombres réels non tous nuls. On ôte de s tous les éléments nuls en préservant l'ordre, et on obtient ainsi une p -liste ($p \leq t$) $\hat{s} = (\hat{s}_1, \hat{s}_2, \dots, \hat{s}_p)$. On appelle *nombre de changements de signe de s* , le nombre d'éléments de l'ensemble $\mathcal{E}$ défini par : $\mathcal{E} = \{i \in \llbracket 1, p-1 \rrbracket \mid \widehat{s_i s_{i+1}} < 0\}$. Si $p = 1$, on dit que le nombre de changements de signe est nul.

Par exemple, si $s = (0, 3, 0, 5, -3, 2)$, on a : $\hat{s} = (3, 5, -3, 2)$, et le nombre de changements de signe est égal à 2.

Pour tout réel x , on note respectivement $C_1(x)$, $C_2(x)$ et $C(x)$, le nombre de changements de signe du couple $(R_0(x), R_1(x))$, de la m -liste $(R_1(x), R_2(x), \dots, R_m(x))$, et de la $(m+1)$ -liste $(R_0(x), R_1(x), R_2(x), \dots, R_m(x))$. On désigne par x_0 une racine réelle du polynôme P .

- a) En étudiant les variations de P au voisinage de x_0 , montrer qu'il existe un réel $\delta_1 > 0$ tel que, si $h \in]0, \delta_1[$, on a : $C_1(x_0 + h) - C_1(x_0 - h) = 1$.
- b) À l'aide de la question 2. c), montrer qu'il existe un réel $\delta_2 > 0$ tel que, si $h \in]0, \delta_2[$, on a : $C_2(x_0 + h) = C_2(x_0 - h)$ (on distinguera les deux éventualités : soit, x_0 n'est racine d'aucun des polynômes $R_1, R_2, \dots, R_m$, soit, il existe un entier j de $[1, m-1]$ tel que $R_j(x_0) = 0$).
- c) Dédire des deux questions précédentes que pour $\delta = \min(\delta_1, \delta_2)$ et $h \in]0, \delta[$, on a $C(x_0 + h) - C(x_0 - h) = 1$, et que si a et b sont deux réels qui ne sont pas racines de P et qui vérifient $a < b$, alors le nombre de racines réelles de P dans $[a, b]$ est égal à $C(b) - C(a)$.

4. a) Soit α une racine (réelle ou complexe) de P . Montrer que si $|\alpha| \geq 1$, alors $|\alpha|^n \leq |\alpha|^{n-1} \times \sum_{k=0}^{n-1} |a_k|$. En

dédire, pour toute racine α de P , l'inégalité : $|\alpha| \leq 1 + \sum_{k=0}^{n-1} |a_k|$.

b) Écrire en français, un algorithme permettant de déterminer le nombre de racines réelles de P .

5. On définit en Pascal

`const n = ... ;`

`Type tab = array[1..n] of real ;`

`Var T : tab ;`

Écrire une fonction Pascal dont l'en-tête est `Function nbchgs(T : tab) : integer` qui donne le nombre de changements de signe dans la suite de réels $(T[1], T[2], \dots, T[n])$.

On tiendra compte du fait que le tableau T peut contenir des éléments nuls. La fonction `nbchgs` n'utilisera que le tableau T et aucun autre tableau auxiliaire. On expliquera en français la démarche utilisée.

Partie III. Un majorant du nombre de racines réelles de P

Soit V un polynôme de $\mathbb{R}[X]$ tel que $V(X) = v_m X^m + v_{m-1} X^{m-1} + \dots + v_1 X + v_0$, avec $v_m \neq 0$ et $m \in \mathbb{N}^*$. On note V^* le polynôme réciproque du polynôme V , défini par : $V^*(X) = v_0 X^m + v_1 X^{m-1} + \dots + v_{m-1} X + v_m$. Soit n un entier de $\mathbb{N}^*$. On considère l'application T qui, à tout polynôme P de degré n , normalisé, à coefficients réels, $P(X) = X^n + a_{n-1} X^{n-1} + \dots + a_1 X + a_0$, associe le polynôme $T(P)$ défini par $T(P)(X) = X P'(X)$.

On désigne par $N_0(P)$ le nombre de racines non nulles de P dans l'intervalle $[-1, 1]$ comptées avec leurs ordres de multiplicité, par $N_1(P)$ le nombre de racines de P dans $] -\infty, -1] \cup [1, +\infty[$ comptées avec leurs ordres de multiplicité, et par $N(P)$ le nombre de racines réelles de P comptées avec leurs ordres de multiplicité.

1. a) Établir, à l'aide du théorème de Rolle, l'inégalité : $N_1(P) \leq N_1(T(P)) + 2$.

b) Pour tout k de $\mathbb{N}^*$, on pose $T^k = T \circ T \circ \dots \circ T$ (k fois). Montrer que $N_1(P) \leq N_1(T^k(P)) + 2k$.

2. a) Montrer que pour tout réel x non nul, on a $P^*(x) = x^n P\left(\frac{1}{x}\right)$.

b) Montrer que $N_1(P) = N_0(P^*)$.

3. Pour tout réel x et pour tout entier naturel k non nul, on pose :

$$Q_k(x) = 1 + a_{n-1} \left(1 - \frac{1}{n}\right)^k x + a_{n-2} \left(1 - \frac{2}{n}\right)^k x^2 + \dots + a_1 \left(1 - \frac{n-1}{n}\right)^k x^{n-1}. \text{ Montrer que } (T^k(P))^* = n^k Q_k.$$

4. a) Établir, pour tout réel y de $[0, 1]$, l'inégalité : $(1 - y)e^y \leq 1$.

b) On **admet** la propriété suivante : soit r et ρ deux réels tels que $0 < r < \rho$. On note $D_\rho = \{z \in \mathbb{C} / |z| \leq \rho\}$. Soit U un polynôme de $\mathbb{R}[X]$ tel que $U(0) \neq 0$. Soit μ un réel strictement positif tel que pour tout z de D_ρ , $|U(z)| \leq \mu$. Alors, le nombre de racines réelles de U comptées avec leurs ordres de multiplicité, dans l'intervalle $[-r, r]$, est majoré par le réel : $\frac{1}{\ln\left(\frac{\rho}{r}\right)} \times \ln\left(\frac{\mu}{|U(0)|}\right)$.

En appliquant cette propriété au polynôme Q_k avec $r = 1$ et $\rho = e^{k/n}$, ($k \in \mathbb{N}^*$), déduire des questions précédentes que pour tout k de $\mathbb{N}^*$, on a : $N_1(P) \leq 2k + \frac{n}{k} \ln(L(P))$, avec $L(P) = 1 + \sum_{i=0}^{n-1} |a_i|$.

- c) Soit ψ la fonction définie sur $\mathbb{R}^{+*}$ par : $\psi(x) = 2x + \frac{\theta}{x}$, où θ est un paramètre réel positif.
- i) Étudier les variations de ψ .
 - ii) Montrer que $\psi(\sqrt{\theta/2} + 1) \leq 2 + 2\sqrt{2\theta}$.
 - iii) En déduire l'inégalité : $N_1(P) \leq 2 + 2\sqrt{2n \ln(L(P))}$.
- d) En supposant $a_0 \neq 0$, on démontrerait de même (et on admettra dans la suite du problème) que :

$$N_0(P) \leq 2 + 2\sqrt{2n \ln \left(\frac{L(P)}{|a_0|} \right)}$$

Conclure en donnant un majorant de $N(P)$, fonction des coefficients $a_0, a_1, \dots, a_{n-1}$.

Partie IV. Nombre de racines réelles d'un polynôme de degré n à coefficients aléatoires

Pour n entier supérieur ou égal à 2, on considère dans cette partie, les variables aléatoires réelles $X_1, X_2, \dots, X_{n-1}$ définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$, indépendantes et de même loi de Poisson de paramètre λ , strictement positif.

Pour tout ω de Ω , on considère le polynôme Q_ω d'indéterminée y , défini par :

$$Q_\omega(y) = y^n + X_{n-1}(\omega)y^{n-1} + \dots + X_1(\omega)y + 1$$

Soit $M_n(\omega)$ le nombre de racines réelles de Q_ω . On admet que l'application $M_n : \omega \mapsto M_n(\omega)$ est une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$.

1. On définit la variable aléatoire L_n par : $L_n = 2 + \sum_{i=1}^{n-1} X_i$. Soit $Z_n = L_n - 2$. Rappeler la loi de Z_n .

2. À l'aide des résultats de la partie III, montrer que pour tout ω de Ω , on a :

$$M_n(\omega) \leq 4 + 4\sqrt{2n} \times \sqrt{\ln(Z_n(\omega) + 2)}$$

3. Soit h une fonction de classe C^2 , concave sur $\mathbb{R}^+$. Soit W une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $\mathbb{N}$. On suppose l'existence des espérances $E(W)$ et $E(h(W))$.

- a) Montrer que, pour tout couple (x_0, x) de réels positifs, on a : $h(x) \leq h'(x_0)(x - x_0) + h(x_0)$.
- b) En prenant $x_0 = E(W)$, établir l'inégalité suivante : $E(h(W)) \leq h(E(W))$.

4. a) Montrer que la fonction φ définie sur $\mathbb{R}^+$ par $\varphi(x) = \sqrt{\ln(x+2)}$ est concave sur $\mathbb{R}^+$.

- b) Soit a un réel positif. Montrer que la série de terme général $\sqrt{\ln(k+2)} \times \frac{a^k}{k!}$ est convergente.

5. a) Prouver l'existence de l'espérance $E(M_n)$.

- b) Montrer que, pour tout réel β strictement supérieur à $1/2$, on a :

$$\lim_{n \rightarrow +\infty} \frac{E(M_n)}{n^\beta} = 0$$

* FIN *



BANQUE COMMUNE D'ÉPREUVES

CODE SUJET :

283

CCIP_M2_S

Conceptions : H.E.C. – E.S.C.P. – E.A.P.

OPTION : SCIENTIFIQUE

MATHÉMATIQUES II

Mercredi 9 Mai 2007, de 14 h. à 18 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite.

Seule l'utilisation d'une règle graduée est autorisée.

Pour toute variable aléatoire réelle Y définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ et possédant une espérance mathématique, on note $E(Y)$ cette espérance pour la probabilité P .

Pour tout événement C de $\mathcal{A}$ tel que $P(C) > 0$, on note, sous réserve d'existence, $E(Y/C)$ l'espérance de Y pour la probabilité conditionnelle P_C (espérance conditionnelle de Y sachant C).

Partie I.

Cette partie constitue une application particulière des résultats généraux étudiés dans la suite du problème.

On possède n urnes ($n \geq 3$) numérotées de 1 à n , dans lesquelles on répartit au hasard et de façon indépendante, m boules indiscernables ($m \geq 4$), de sorte que, pour tout i de $\llbracket 1, n \rrbracket$, la probabilité pour chaque boule d'être placée dans l'urne numéro i soit égale à $1/n$.

On suppose que cette expérience est modélisée par un espace probabilisé $(\Omega, \mathcal{A}, P)$.

À l'issue de cette expérience, on pose pour tout i de $\llbracket 1, n \rrbracket$:

$$X_i = \begin{cases} 1 & \text{si l'urne n°} i \text{ est vide} \\ 0 & \text{sinon} \end{cases}$$

On pose $W_n = \sum_{i=1}^n X_i$.

1. a) Déterminer pour tout i de $\llbracket 1, n \rrbracket$, la loi de la variable aléatoire X_i .
b) Pour tout couple (i, j) d'entiers de $\llbracket 1, n \rrbracket$ distincts, calculer $P([X_i = 1] \cap [X_j = 1])$, ainsi que la covariance de X_i et X_j . Les variables aléatoires X_i et X_j sont-elles indépendantes ?
2. a) Exprimer l'espérance $E(W_n)$ de W_n en fonction de n et m .
b) On note $V(W_n)$ la variance de W_n . Calculer $V(W_n)$ en fonction de n et m .

c) Vérifier l'égalité : $E(W_n) - V(W_n) = n^2 \left(1 - \frac{1}{n}\right)^{2m} - n(n-1) \left(1 - \frac{2}{n}\right)^m$.

En déduire que $E(W_n) - V(W_n) \geq 0$.

3. Dans cette question, l'entier m vérifie $m = \lfloor n \ln n + \theta n \rfloor$, où θ est une constante réelle positive et $\lfloor x \rfloor$ désigne la partie entière de x .

a) Calculer $\lim_{n \rightarrow +\infty} E(W_n)$.

b) Montrer que $\lim_{n \rightarrow +\infty} (E(W_n) - V(W_n)) = 0$.

c) Soit T_n une variable aléatoire qui suit une loi de Poisson de paramètre $\mu_n = E(W_n)$.

On admet que pour tout k de $\mathbb{N}$, on a :

$$|P([W_n = k]) - P([T_n = k])| \leq \min\left(1, \frac{1}{\mu_n}\right) \times (\mu_n - V(W_n))$$

Quelle est la limite en loi de la suite de variables aléatoires $(W_n)_{n \geq 3}$?

4. On pose $\mu = e^{-\theta}$, et on suppose que le paramètre μ est inconnu. Dans cette question, on veut estimer μ . Pour p entier de $\mathbb{N}^*$, on considère un p -échantillon indépendant, identiquement distribué $(T_1, T_2, \dots, T_p)$ de la loi de Poisson de paramètre μ . On pose :

$$\overline{T}_p = \frac{1}{p} \sum_{i=1}^p T_i \text{ et } U_p = \sqrt{p} \frac{\overline{T}_p - \mu}{\sqrt{\mu}}$$

a) Montrer que $\overline{T}_p$ est un estimateur sans biais et convergent du paramètre μ .

b) Quelle est la limite en loi de la suite de variables aléatoires $(U_p)_{p \geq 1}$?

c) On veut construire, pour p assez grand, un intervalle de confiance du paramètre μ au risque α donné. Soit u le réel strictement positif tel que $P(|U| \geq u) = \alpha/2$, où U est une variable aléatoire qui suit la loi normale centrée réduite.

Justifier que pour p assez grand, on peut écrire : $P(|U_p| \leq u) = 1 - \alpha$, et déterminer alors un intervalle de confiance $[I_p, J_p]$ pour μ au risque α .

Partie II

Dans cette partie, λ désigne un réel strictement positif.

Soit M une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ qui suit une loi de Poisson de paramètre λ .

Soit A une partie quelconque de $\mathbb{N}$ et $\overline{A}$ son complémentaire dans $\mathbb{N}$. On rappelle que si A est non vide, alors, $P([M \in A]) = \sum_{i \in A} e^{-\lambda} \frac{\lambda^i}{i!}$, et on pose par convention $[M \in \emptyset] = \emptyset$.

On considère la fonction f_A définie sur $\mathbb{N}$ par $f_A(0) = 0$, et pour tout k de $\mathbb{N}$:

$$f_A(k+1) = \frac{k!}{\lambda^{k+1}} e^\lambda (P([M \in A] \cap [M \leq k]) - P([M \in A]) \times P([M \leq k]))$$

1. a) Déterminer la fonction f_A dans les cas particuliers $A = \emptyset$ et $A = \mathbb{N}$.

b) Donner l'expression de $f_A(1)$ en fonction de λ et de $P([M \in A])$ dans les deux cas suivants : $0 \in A$ et $0 \in \overline{A}$. Exprimer $f_A(2)$ en fonction de λ et de $P([M \in A])$ dans le cas où 0 et 1 appartiennent à A .

2. Soit A et B deux parties de $\mathbb{N}$ disjointes.

a) Montrer que $f_{A \cup B} = f_A + f_B$.

b) En déduire que $f_{\overline{A}} = -f_A$.

3. a) Montrer que pour tout k de $\mathbb{N}$, la fonction f_A vérifie la relation suivante :

$$\lambda f_A(k+1) - k f_A(k) = \begin{cases} P([M \in \overline{A}]) & \text{si } k \in A \\ -P([M \in A]) & \text{si } k \in \overline{A} \end{cases}$$

b) En déduire que si A est non vide et distincte de $\mathbb{N}$, la fonction f_A n'est pas identiquement nulle.

4. Dans cette question, j est un entier naturel non nul, et A est le singleton $\{j\}$. On pose $f_{\{j\}} = f_j$.

a) Pour tout k de $\mathbb{N}^*$, montrer l'égalité suivante :

$$f_j(k+1) = \begin{cases} \frac{k!}{j! \lambda^{k-j+1}} P([M \geq k+1]) & \text{si } k \geq j \\ -\frac{k!}{j! \lambda^{k-j+1}} P([M \leq k]) & \text{si } k < j \end{cases}$$

b) Calculer $f_j(j+1) - f_j(j)$, et déterminer son signe.

c) Calculer pour tout k de $\mathbb{N}^*$, différent de j , $f_j(k+1) - f_j(k)$ en distinguant les deux cas : $k > j$ et $k < j$. En déduire que la différence $f_j(k+1) - f_j(k)$ est positive si et seulement si $k = j$.

d) Établir les inégalités suivantes : $f_j(j+1) - f_j(j) \leq \frac{1 - e^{-\lambda}}{\lambda} \leq \min\left(1, \frac{1}{\lambda}\right)$.

5. On considère le singleton $\{0\}$ et on pose $f_{\{0\}} = f_0$. Montrer, pour tout k de $\mathbb{N}^*$, l'inégalité suivante : $f_0(k+1) - f_0(k) \leq 0$.

6. a) Établir pour tout k de $\mathbb{N}$, l'inégalité suivante : $f_A(k+1) - f_A(k) \leq f_k(k+1) - f_k(k)$.
(on distinguera les deux cas : $k \in A$ et $k \in \overline{A}$)

b) En déduire, pour toute partie A de $\mathbb{N}$, l'inégalité suivante :

$$\sup_{k \geq 0} |f_A(k+1) - f_A(k)| \leq \min\left(1, \frac{1}{\lambda}\right)$$

Partie III.

Soit n un entier supérieur ou égal à 2. On considère n variables aléatoires discrètes indépendantes $X_1, X_2, \dots, X_n$ définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$, telles que pour tout i de $\llbracket 1, n \rrbracket$, la variable aléatoire X_i suit une loi de Bernoulli de paramètre p_i strictement positif.

On pose $\lambda_n = \sum_{i=1}^n p_i$, $W_n = \sum_{i=1}^n X_i$ et, pour tout i de $\llbracket 1, n \rrbracket$, $R_i = W_n - X_i$.

On note M_n une variable aléatoire qui suit une loi de Poisson de paramètre λ_n . Soit A une partie quelconque de $\mathbb{N}$, et f_A la fonction définie dans la partie II, dans l'expression de laquelle on remplace M par M_n et λ par λ_n . On pose $f = f_A$.

1. a) Établir pour tout i de $\llbracket 1, n \rrbracket$, l'égalité des deux variables aléatoires $X_i f(W_n)$ et $X_i f(1 + R_i)$.

b) En déduire pour tout i de $\llbracket 1, n \rrbracket$, l'égalité : $E(X_i f(W_n)) = p_i E(f(1 + R_i))$.

2. Pour tout i de $\llbracket 1, n \rrbracket$, on pose : $Y_i = f(1 + W_n) - f(1 + R_i)$.

Établir la relation suivante : $E(\lambda_n f(1 + W_n) - W_n f(W_n)) = \sum_{i=1}^n p_i E(Y_i)$.

3. a) Établir pour tout i de $\llbracket 1, n \rrbracket$, la formule suivante :

$$E(Y_i / [X_i = 1]) = E(f(2 + R_i) - f(1 + R_i))$$

b) Calculer pour tout i de $\llbracket 1, n \rrbracket$, $E(Y_i / [X_i = 0])$.

c) Dédurre des questions précédentes l'égalité suivante :

$$E(\lambda_n f(1 + W_n) - W_n f(W_n)) = \sum_{i=1}^n p_i^2 E(f(2 + R_i) - f(1 + R_i))$$

4. Établir l'inégalité suivante :

$$|E(\lambda_n f(1 + W_n) - W_n f(W_n))| \leq \min\left(1, \frac{1}{\lambda_n}\right) \times \sum_{i=1}^n p_i^2$$

5. À l'aide de la question II.3.a, montrer, pour toute partie A de $\mathbb{N}$, l'égalité suivante :

$$E(\lambda_n f(1 + W_n) - W_n f(W_n)) = P([W_n \in A]) - P([M_n \in A])$$

En déduire, pour toute partie A de $\mathbb{N}$, la majoration suivante :

$$|P([W_n \in A]) - P([M_n \in A])| \leq \min\left(1, \frac{1}{\lambda_n}\right) \times \sum_{i=1}^n p_i^2$$

6. Dans cette question uniquement, on suppose que pour tout i de $\llbracket 1, n \rrbracket$, X_i suit la loi de Bernoulli de paramètre $p_i = \frac{1}{n+i}$.

a) Déterminer $\lim_{n \rightarrow +\infty} \lambda_n$. Montrer que $\lim_{n \rightarrow +\infty} \sum_{i=1}^n p_i^2 = 0$.

b) Déterminer la limite en loi de la suite $(W_n)_{n \geq 2}$.

Partie IV.

Les notations sont identiques à celles de la partie III, mais les variables aléatoires $X_1, X_2, \dots, X_n$ définies sur $(\Omega, \mathcal{A}, P)$, **ne sont pas nécessairement indépendantes**.

1. a) Montrer que pour tout i de $\llbracket 1, n \rrbracket$, on a : $E(X_i f(W_n)) = p_i E(f(1 + R_i) / [X_i = 1])$.

b) En déduire l'égalité suivante : $P([W_n \in A]) - P([M_n \in A]) = \sum_{i=1}^n p_i \left[E(f(1 + W_n)) - E(f(1 + R_i) / [X_i = 1]) \right]$.

2. On suppose que pour tout i de $\llbracket 1, n \rrbracket$, il existe une variable aléatoire Z_i définie sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $\mathbb{N}$, telle que la loi de Z_i soit identique à la loi conditionnelle de R_i sachant $[X_i = 1]$.

a) Justifier, pour tout couple (ℓ, j) d'entiers naturels, l'inégalité : $|f(\ell) - f(j)| \leq |\ell - j| \times \min\left(1, \frac{1}{\lambda_n}\right)$, et en déduire la majoration suivante : $|P([W_n \in A]) - P([M_n \in A])| \leq \min\left(1, \frac{1}{\lambda_n}\right) \times \sum_{i=1}^n p_i E(|W_n - Z_i|)$.

b) On suppose de plus que pour tout ω de Ω , pour tout i de $\llbracket 1, n \rrbracket$, on a $W_n(\omega) \geq Z_i(\omega)$. Établir l'égalité : $\sum_{i=1}^n p_i E(|W_n - Z_i|) = \lambda_n - V(W_n)$, où $V(W_n)$ désigne la variance de W_n .

En déduire, pour toute partie A de $\mathbb{N}$, l'inégalité suivante :

$$|P([W_n \in A]) - P([M_n \in A])| \leq \min\left(1, \frac{1}{\lambda_n}\right) \times (\lambda_n - V(W_n))$$



BANQUE COMMUNE D'ÉPREUVES

CODE ÉPREUVE :

283

CCIP_M2_S

Concepteur : H.E.C. – E.S.C.P. – E.A.P.

OPTION SCIENTIFIQUE

MATHÉMATIQUES II

Mercredi 7 mai 2008, de 14 h. à 18 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite.

Seule l'utilisation d'une règle graduée est autorisée.

Toutes les variables aléatoires qui interviennent dans ce problème sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$. Sous réserve d'existence, on note $E(X)$ et $V(X)$ respectivement, l'espérance et la variance d'une variable aléatoire réelle X quelconque. Pour toute variable aléatoire réelle X admettant une densité sur $\mathbb{R}$, notée f_X , on note $\mathcal{D}_X$ l'ensemble des réels s pour lesquels la variable aléatoire e^{sX} admet une espérance, et on note Φ_X la fonction définie sur $\mathcal{D}_X$ par : $\Phi_X(s) = E(e^{sX})$.

On **admet** les résultats suivants :

- si deux variables aléatoires X et Y sont telles que Φ_X et Φ_Y coïncident sur un intervalle ouvert non vide, alors X et Y ont la même loi ;
- si n est un entier naturel non nul, et $X_1, X_2, \dots, X_n$ des variables aléatoires réelles quelconques, mutuellement indépendantes, alors, pour tout entier p de $\llbracket 1, n-1 \rrbracket$ et pour toutes fonctions réelles continues φ_1 et φ_2 , les variables aléatoires $\varphi_1(X_1, \dots, X_p)$ et $\varphi_2(X_{p+1}, \dots, X_n)$ sont indépendantes ;
- si X et Y sont des variables aléatoires indépendantes admettant une espérance, alors XY admet une espérance, et $E(XY) = E(X)E(Y)$.

La fonction exponentielle est également notée $\exp$. On rappelle que : $\int_{-\infty}^{+\infty} \exp\left(\frac{-x^2}{2}\right) dx = \sqrt{2\pi}$.

Dans tout le problème, U désigne une variable aléatoire qui suit la loi normale centrée réduite.

Préliminaire

On rappelle que, pour tout s de $\mathcal{D}_X$, on a : $\Phi_X(s) = \int_{-\infty}^{+\infty} \exp(sx) f_X(x) dx$.

1. Soit a un réel non nul et b un réel quelconque.

a) Montrer que l'intégrale $\int_{-\infty}^{+\infty} \exp(-ax^2) dx$ est convergente si et seulement si $a > 0$, et vaut alors $\sqrt{\frac{\pi}{a}}$.

b) En déduire que l'intégrale $\int_{-\infty}^{+\infty} \exp(-ax^2 + bx)dx$ est convergente si et seulement si $a > 0$, puis montrer que dans ces conditions, on a : $\int_{-\infty}^{+\infty} \exp(-ax^2 + bx)dx = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a}\right)$.

2. a) Déterminer $\mathcal{D}_U$; pour tout s de $\mathcal{D}_U$, calculer $\Phi_U(s)$.

b) On pose : $Z = U^2$. Établir que : $\mathcal{D}_Z =]-\infty, \frac{1}{2}[$; montrer, à l'aide du théorème de transfert, que pour tout réel s de $\mathcal{D}_Z$, on a : $\Phi_Z(s) = (1 - 2s)^{-1/2}$.

3. Soit X une variable aléatoire réelle à densité, et soit μ et β deux réels quelconques.

a) Montrer qu'un réel s appartient à $\mathcal{D}_{\mu X + \beta}$ si et seulement si μs appartient à $\mathcal{D}_X$, et que dans ce cas, on a : $\Phi_{\mu X + \beta}(s) = \exp(\beta s) \Phi_X(\mu s)$.

b) On suppose que X suit une loi γ de paramètre ν , où ν est un réel strictement positif.

Montrer que : $\mathcal{D}_X =]-\infty, 1[$; pour tout s de $\mathcal{D}_X$, établir la formule : $\Phi_X(s) = (1 - s)^{-\nu}$. De même, déterminer $\mathcal{D}_{2X}$; pour tout s de $\mathcal{D}_{2X}$, calculer $\Phi_{2X}(s)$.

Partie I. Loi du χ^2 centré

Soit r un entier supérieur ou égal à 1. On considère une variable aléatoire X_r suivant la loi Γ de paramètres $(2, \frac{r}{2})$, c'est-à-dire que X_r possède une densité f_{X_r} donnée par :

$$f_{X_r}(x) = \begin{cases} \frac{1}{2^{\frac{r}{2}} \times \Gamma(\frac{r}{2})} \times x^{\frac{r}{2}-1} \times \exp\left(-\frac{x}{2}\right) & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

On dit que X_r suit une loi du χ^2 (« chi deux ») centré à r degrés de liberté, et on note : $X_r \hookrightarrow \chi^2(r)$.

1. Étudier les variations de f_{X_r} et tracer sa courbe représentative dans un repère orthogonal du plan.

2. a) Montrer que la variable aléatoire $\frac{X_r}{2}$ suit une loi γ de paramètre $\frac{r}{2}$. En déduire $E(X_r)$ et $V(X_r)$.

b) Déterminer $\mathcal{D}_{X_r}$; pour tout s de $\mathcal{D}_{X_r}$, calculer $\Phi_{X_r}(s)$.

3. Soit n un entier de $\mathbb{N}^*$. On considère n variables aléatoires indépendantes $U_1, U_2, \dots, U_n$ de même loi que U . Pour tout i de $[1, n]$, on pose : $Z_i = U_i^2$.

a) Vérifier que X_1 et U^2 sont de même loi.

b) On pose : $W_n = \sum_{i=1}^n Z_i$. Quelle est la loi de probabilité de W_n ?

c) Déterminer $\mathcal{D}_{W_n}$, et pour tout s de $\mathcal{D}_{W_n}$, exprimer $\Phi_{W_n}(s)$ en fonction de s et de n . Établir une relation entre $\Phi_{W_n}(s)$ et $\Phi_{Z_1}(s), \Phi_{Z_2}(s), \dots, \Phi_{Z_n}(s)$.

4. Soit T une variable aléatoire qui suit la loi normale centrée, de variance σ^2 inconnue, σ étant un réel strictement positif. Pour n entier supérieur ou égal à 2, on dispose d'un n -échantillon indépendant, identiquement distribué (i.i.d.), $T_1, T_2, \dots, T_n$ de la loi de T . On considère la variable aléatoire S_n définie

$$\text{par : } S_n = \frac{1}{n} \sum_{i=1}^n T_i^2.$$

a) Montrer que S_n est un estimateur sans biais et convergent du paramètre σ^2 .

b) Soit α un réel vérifiant : $0 < \alpha < 1$, et soit k_α le réel strictement positif tel que : $P([W_n \geq k_\alpha]) = 1 - \alpha$. Montrer que l'intervalle $]0, \frac{nS_n}{k_\alpha}]$ est un intervalle de confiance de σ^2 au risque α .

Partie II. Loi du χ^2 décentré

On considère une suite $(M_j)_{j \geq 1}$ de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, mutuellement indépendantes, telles que pour tout j de $\mathbb{N}^*$, M_j suive la loi normale d'espérance m_j ($m_j \in \mathbb{R}$) et de variance égale à 1.

Pour n entier de $\mathbb{N}^*$, on pose : $Y_n = \sum_{j=1}^n M_j^2$ et $\lambda_n = \sum_{j=1}^n m_j^2$.

On dit que Y_n suit une loi du χ^2 décentré à n degrés de liberté, de paramètre de décentrage λ_n , et on note : $Y_n \hookrightarrow \chi^2(n, \lambda_n)$.

1. Dans cette question *uniquement*, on suppose que l'entier n est égal à 1.

a) Montrer les deux égalités suivantes : $E(U^3) = 0$ et $E(U^4) = 3$.

b) En déduire, en fonction de λ_1 , les valeurs respectives de $E(Y_1)$ et de $V(Y_1)$.

c) Montrer que : $\mathcal{D}_{Y_1} =]-\infty, \frac{1}{2}[$ et établir, pour tout réel s de $\mathcal{D}_{Y_1}$, la formule suivante :

$$\Phi_{Y_1}(s) = (1 - 2s)^{-1/2} \times \exp\left(\frac{s\lambda_1}{1 - 2s}\right)$$

2. Soit n un entier de $\mathbb{N}^*$.

a) Calculer $E(Y_n)$ et $V(Y_n)$ en fonction de n et λ_n .

b) On admet que l'on a : $\mathcal{D}_{Y_n} =]-\infty, \frac{1}{2}[$. Pour tout s de $\mathcal{D}_{Y_n}$, exprimer $\Phi_{Y_n}(s)$ en fonction de s, n et λ_n .

Partie III. Nombre aléatoire de degrés de liberté

Sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, on considère une variable aléatoire X à valeurs dans $\mathbb{R}$ admettant une espérance $E(X)$, et une variable aléatoire K à valeurs dans $\mathbb{N}$. On note N_K l'ensemble des entiers naturels k vérifiant $P([K = k]) > 0$, et on suppose que pour tout entier k de N_K , la variable aléatoire X admet une espérance pour la probabilité conditionnelle $P_{[K=k]}$, notée $E(X/K = k)$.

On admet alors l'égalité suivante : $(\star) \quad E(X) = \sum_{k \in N_K} E(X/K = k)P([K = k])$

Soit g l'application définie sur $\mathbb{N}$ par : $g(k) = \begin{cases} E(X/K = k) & \text{si } k \in N_K \\ 0 & \text{sinon} \end{cases}$.

1. *Vérification de la formule $(\star)$ sur un exemple.*

Soit $(J_i)_{i \geq 1}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$, indépendantes et de même loi uniforme sur l'intervalle $[0, 1]$. Pour tout k de $\mathbb{N}^*$, on pose : $X_k = \sup_{1 \leq i \leq k} (J_i)$; autrement dit, pour tout ω de Ω ,

$X_k(\omega) = \max_{1 \leq i \leq k} J_i(\omega)$. Soit K une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$ qui suit la loi uniforme discrète sur $[1, n]$ ($n \in \mathbb{N}^*$). On suppose que K est indépendante des variables aléatoires de la suite $(J_i)_{i \geq 1}$.

Pour tout ω de Ω , on pose : $X(\omega) = \max_{1 \leq i \leq K(\omega)} J_i(\omega)$, et on admet que X est une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$.

a) Établir, pour tout entier k de $[1, n]$ et pour tout réel x , la relation : $P_{[K=k]}([X \leq x]) = P([X_k \leq x])$.

b) Déterminer la fonction de répartition F_X de la variable aléatoire X .

c) En déduire que X est une variable aléatoire à densité, qui admet une espérance $E(X)$ que l'on exprimera en fonction de $\sum_{k=1}^n \frac{k}{k+1}$.

d) Vérifier l'égalité $(\star)$: $E(X) = E(g(K))$.

2. Soit $(U_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes qui suivent la loi normale centrée réduite. Soit K une variable aléatoire indépendante des variables aléatoires de la suite $(U_i)_{i \geq 1}$, qui suit la loi de Poisson de paramètre $\frac{\lambda}{2}$ strictement positif.

Pour n entier de $\mathbb{N}^*$, on pose : $H_n = U_1^2 + U_2^2 + \dots + U_{n+2K}^2$. On admet que H_n est une variable aléatoire à densité à valeurs positives, et que $\mathcal{D}_{H_n} =]-\infty, \frac{1}{2}[$.

Soit s un réel de $] -\infty, \frac{1}{2}[$.

a) Montrer que pour tout k de $\mathbb{N}$, la loi conditionnelle de H_n sachant $[K = k]$ est la loi de la variable aléatoire W_{n+2k} définie dans la question I.3.b.

b) En posant : $X = e^{sH_n}$, déterminer, pour tout entier k de $\mathbb{N}$, l'expression de $g(k)$ en fonction de k .

c) Établir la formule suivante :

$$E(g(K)) = (1 - 2s)^{-n/2} \times \exp\left(\frac{\lambda s}{1 - 2s}\right)$$

d) En utilisant l'égalité $(\star)$, admise au début de cette partie, avec $X = e^{sH_n}$, déterminer la loi de H_n .

e) À l'aide de la question III.2.a, montrer que pour tout entier n supérieur ou égal à 3, on a :

$$E\left(\frac{1}{H_n}\right) = E\left(\frac{1}{n - 2 + 2K}\right)$$

Partie IV. Estimateur de James-Stein

Soit p un entier supérieur ou égal à 3. On suppose qu'un modèle aléatoire défini sur $(\Omega, \mathcal{A}, P)$ comporte p paramètres réels inconnus $\theta_1, \dots, \theta_p$ non tous nuls. Un échantillon d'observations statistiques permet d'exhiber des estimateurs $\hat{\theta}_1, \dots, \hat{\theta}_p$ sans biais des paramètres $\theta_1, \dots, \theta_p$ respectivement. On suppose que les variables aléatoires $\hat{\theta}_1, \dots, \hat{\theta}_p$ sont indépendantes et suivent une loi normale de variance égale à 1.

On pose : $\theta = (\theta_1, \dots, \theta_p)$, $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)$, $B_p = \sum_{j=1}^p \hat{\theta}_j^2$ et $b_p = \sum_{j=1}^p \theta_j^2$.

On dit que le vecteur aléatoire $\hat{\theta}$ est un estimateur sans biais du paramètre vectoriel θ , et $E(\hat{\theta})$ est alors le vecteur θ .

On définit le risque quadratique scalaire d'un estimateur θ^* de θ , noté $R(\theta^*, \theta)$, par :

$$R(\theta^*, \theta) = E\left(\sum_{j=1}^p (\theta_j^* - \theta_j)^2\right)$$

Dans cette partie, on cherche un estimateur θ^* de θ , représenté par un vecteur aléatoire $(\theta_1^*, \dots, \theta_p^*)$, dont le risque $R(\theta^*, \theta)$ est strictement inférieur à $R(\hat{\theta}, \theta)$.

1. Justifier que la variable aléatoire B_p suit la loi $\chi^2(p, b_p)$, et qu'elle constitue un estimateur biaisé de b_p .

2. On pose : $\theta^* = \left(1 - \frac{c}{B_p}\right) \times \hat{\theta}$, où c est un paramètre réel strictement positif. Soit K une variable aléatoire qui suit la loi de Poisson de paramètre $\frac{b_p}{2}$.

a) En admettant que l'on a : $E\left(\frac{1}{B_p} \sum_{j=1}^p \theta_j \hat{\theta}_j\right) = E\left(\frac{2K}{p - 2 + 2K}\right)$, établir l'égalité suivante :

$$R(\theta^*, \theta) - R(\hat{\theta}, \theta) = (c^2 - 2c(p - 2)) \times E\left(\frac{1}{p - 2 + 2K}\right)$$

b) Montrer que l'inégalité : $R(\theta^*, \theta) < R(\hat{\theta}, \theta)$, est vérifiée si et seulement si : $0 < c < 2(p - 2)$.

Déterminer en fonction de p , la valeur de c pour laquelle $R(\theta^*, \theta) - R(\hat{\theta}, \theta)$ est minimale.

Comment s'écrit alors l'estimateur θ^* ?



BANQUE COMMUNE D'ÉPREUVES

CONCOURS D'ADMISSION DE 2009

Conceptions : H.E.C. – E.S.C.P. – E.A.P

283

OPTION SCIENTIFIQUE

CCIP_M2_S

MATHEMATIQUES II

Mardi 5 mai 2009, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite. Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre

Dans tout le problème, N désigne un entier supérieur ou égal à 1.

On note $E(X)$ et $V(X)$ respectivement, l'espérance et la variance lorsqu'elles existent, de toute variable aléatoire réelle X définie sur un espace probabilisé.

Soit $(U_n)_{n \geq 1}$ une suite de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, mutuellement indépendantes et de même loi uniforme discrète sur $[1, N]$.

On pose, pour tout n de $\mathbb{N}^*$: $T_n = \sup(U_1, U_2, \dots, U_n)$ et $Z_n = \inf(U_1, U_2, \dots, U_n)$. On admet que T_n et Z_n sont des variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$. Ainsi, pour tout ω de Ω , on a :

$$T_n(\omega) = \max(U_1(\omega), U_2(\omega), \dots, U_n(\omega)) \text{ et } Z_n(\omega) = \min(U_1(\omega), U_2(\omega), \dots, U_n(\omega))$$

On rappelle que si C désigne un élément de $\mathcal{A}$, on note 1_C la variable aléatoire indicatrice de l'événement C , définie sur $(\Omega, \mathcal{A}, P)$ par :

$$1_C(\omega) = \begin{cases} 1 & \text{si } \omega \in C \\ 0 & \text{si } \omega \notin C \end{cases}$$

$$\text{On pose, pour tout } n \text{ de } \mathbb{N}^* : d_n(N) = \begin{cases} \sum_{k=1}^{N-1} \left(\frac{k}{N}\right)^n & \text{si } N \geq 2 \\ 0 & \text{si } N = 1 \end{cases}$$

Préliminaire

1. Soit Y une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $[1, N]$. Établir les deux relations suivantes :

$$E(Y) = \sum_{k=0}^{N-1} P([Y > k]) \quad \text{et} \quad E(Y^2) = \sum_{k=0}^{N-1} (2k+1)P([Y > k])$$

Partie I. Inf et Sup

2. Rappeler, sans démonstration, les valeurs respectives de $E(U_1)$ et de $V(U_1)$.
3. a) Calculer, pour tout k de $\llbracket 1, N \rrbracket$, $P([T_n \leq k])$.
b) En déduire la loi de probabilité de T_n .
4. a) Montrer que la suite $(d_n(N))_{n \geq 1}$ est convergente et calculer sa limite.
b) Exprimer $E(T_n)$ en fonction de N et $d_n(N)$. En déduire la valeur de $\lim_{n \rightarrow +\infty} E(T_n)$.
c) Établir la formule suivante : $V(T_n) = (2N - 1)d_n(N) - 2Nd_{n+1}(N) - d_n^2(N)$.
En déduire la valeur de $\lim_{n \rightarrow +\infty} V(T_n)$.
- d) Montrer que si $N \geq 2$, on a : $\lim_{n \rightarrow +\infty} \frac{d_{n+1}(N)}{d_n(N)} = 1 - \frac{1}{N}$; en déduire que, lorsque n tend vers $+\infty$, on a : $V(T_n) \sim d_n(N)$.
5. Déterminer la loi de Z_n . Calculer $E(Z_n)$ et $V(Z_n)$.
6. On rappelle que la fonction Pascal `random(N)` permet de simuler une variable aléatoire suivant la loi uniforme sur $\llbracket 0, N - 1 \rrbracket$. Écrire une fonction Pascal d'en-tête `simulmax(n : integer) : integer` qui simule la variable aléatoire T_n .

Partie II. Couple (Inf, Sup)

7. On pose, pour tout n de $\mathbb{N}^*$ et pour tout couple (k, ℓ) de $\mathbb{N}^2$: $\phi_n(k, \ell) = P([T_n \leq k] \cap [Z_n \leq \ell])$.
a) Montrer, pour tout (k, ℓ) de $\llbracket 1, N \rrbracket^2$, la relation suivante :

$$\phi_n(k, \ell) = \begin{cases} \left(\frac{k}{N}\right)^n & \text{si } k \leq \ell \\ \left(\frac{k}{N}\right)^n - \left(\frac{k-\ell}{N}\right)^n & \text{si } k > \ell \end{cases}$$

- b) Établir, pour tout (k, ℓ) de $\llbracket 1, N \rrbracket^2$, la formule suivante :

$$P([T_n = k] \cap [Z_n = \ell]) = \phi_n(k, \ell) + \phi_n(k-1, \ell-1) - \phi_n(k-1, \ell) - \phi_n(k, \ell-1)$$

- c) En déduire, en distinguant les trois cas $k < \ell$, $k = \ell$ et $k > \ell$, l'expression de $P([T_n = k] \cap [Z_n = \ell])$ en fonction de k et ℓ .

8. On donne, pour tout couple (m, n) de $(\mathbb{N}^*)^2$, les deux relations suivantes :

$$\begin{aligned} \text{i) } \sum_{j=1}^m [(j+1)^n - 2j^n + (j-1)^n] &= (m+1)^n - m^n - 1 ; \\ \text{ii) } \sum_{j=1}^m j[(j+1)^n - 2j^n + (j-1)^n] &= m(m+1)^n - (m+1)m^n. \end{aligned}$$

- a) En déduire, pour tout n de $\mathbb{N}^*$, la formule suivante : $E(T_n Z_n) = N(1 + d_{n+1}(N))$.
b) On note, pour tout n de $\mathbb{N}^*$, ρ_n le coefficient de corrélation linéaire entre T_n et Z_n .
Calculer $\lim_{n \rightarrow +\infty} \rho_n$ lorsque $N \geq 2$.

9. a) Pour tout n de $\mathbb{N}^*$ et pour tout couple (k, ℓ) de $\llbracket 1, N \rrbracket^2$, calculer la probabilité conditionnelle $P_{[T_n=k]}([Z_n = \ell])$.
b) En déduire, pour tout n de $\mathbb{N}^*$ et pour tout k de $\llbracket 1, N \rrbracket$, l'expression de l'espérance conditionnelle $E(Z_n/[T_n = k])$ de Z_n sachant $[T_n = k]$.

Partie III. Pr vision

Pour n entier de $\mathbb{N}^*$, on dispose d'un $(n+1)$ - chantillon ind pendant identiquement distribu  (i.i.d.) $(U_1, U_2, \dots, U_{n+1})$ de la loi uniforme sur $\llbracket 1, N \rrbracket$.

On pose : $T_n = \sup(U_1, U_2, \dots, U_n)$ et $T_{n+1} = \sup(U_1, U_2, \dots, U_{n+1}) = \sup(T_n, U_{n+1})$.

Pour tout $t = (t_1, t_2, \dots, t_N)$ de $\mathbb{R}^N$, on pose : $W_t(T_n) = \sum_{k=1}^N t_k \times \mathbf{1}_{[T_n=k]}$.

Dans cette partie, on se propose de d terminer la valeur de t pour laquelle les deux conditions suivantes sont v rifi es :

- i) $E(W_t(T_n)) = E(T_{n+1})$;
- ii) $E[(T_{n+1} - W_t(T_n))^2]$ est minimale.

10. Montrer, pour tout k de $\llbracket 1, N \rrbracket$, la relation : $P([W_t(T_n) = t_k]) = P([T_n = k])$.

11.  tablir, pour tout k de $\llbracket 1, N \rrbracket$, la formule suivante :

$$E(T_{n+1} \times \mathbf{1}_{[T_n=k]}) = E(T_{n+1}/[T_n = k]) \times P([T_n = k])$$

12. a) Calculer, pour tout couple (k, j) de $\llbracket 1, N \rrbracket^2$, $P([T_n = k] \cap [T_{n+1} = j])$.

b) En d duire, pour tout couple (k, j) de $\llbracket 1, N \rrbracket^2$, la probabilit  conditionnelle $P_{[T_n=k]}([T_{n+1} = j])$.

c) D terminer, pour tout k de $\llbracket 1, N \rrbracket$, l'expression de l'esp rance conditionnelle $E(T_{n+1}/[T_n = k])$ de T_{n+1} sachant $[T_n = k]$.

d) En appliquant la formule de l'esp rance totale, d duire de la question pr c dente la relation suivante :

$$E(T_{n+1}) = \frac{N+1}{2} + \frac{1}{2N} (E(T_n^2) - E(T_n))$$

13.  tablir l' galit  suivante : $(W_t(T_n))^2 = \sum_{k=1}^N t_k^2 \times \mathbf{1}_{[T_n=k]}$.

14. Soit g la fonction d finie sur $\mathbb{R}^N$   valeurs r elles par :

$$g(t_1, t_2, \dots, t_N) = E[(T_{n+1} - W_t(T_n))^2]$$

a)   l'aide des r sultats des questions 11, 12 et 13, expliciter g en fonction des variables $t_1, t_2, \dots, t_N$.

b) Montrer que g admet un minimum global sur $\mathbb{R}^N$ atteint en un point $\theta = (\theta_1, \theta_2, \dots, \theta_N)$ que l'on d terminera en fonction de $E(T_{n+1}/[T_n = 1]), E(T_{n+1}/[T_n = 2]), \dots, E(T_{n+1}/[T_n = N])$.

15.  tablir les deux relations suivantes :

$$E(W_\theta(T_n)) = E(T_{n+1}) \quad \text{et} \quad V(W_\theta(T_n)) \leq V(T_{n+1})$$

16. a)  tablir, pour tout i de $\mathbb{N}^*$, l' galit  suivante : $\sum_{k=1}^N k^i \times \mathbf{1}_{[T_n=k]} = (T_n)^i$.

b) En d duire la relation suivante : $W_\theta(T_n) = \frac{N+1}{2} + \frac{1}{2N} (T_n^2 - T_n)$.

Partie IV. Estimation

Soit U une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, de loi uniforme discrète sur $\llbracket 1, N \rrbracket$. On suppose que le paramètre N est inconnu.

Cette partie a pour objet la détermination d'un estimateur ponctuel de N , sans biais et de variance minimale.

Pour n entier supérieur ou égal à 1, soit $(U_1, U_2, \dots, U_n)$ un n -échantillon i.i.d. de la loi de U .

17. Soit ε un réel strictement positif. On pose :

$$A_n(\varepsilon) = [|T_n - N| \geq \varepsilon] \text{ et } B_n(\varepsilon) = [|T_n - E(T_n)| + |d_n(N)| \geq \varepsilon]$$

- a) Peut-on dire que $T_n + d_n(N)$ est un estimateur sans biais de N ?
- b) Montrer que la suite $(T_n)_{n \geq 1}$ est une suite d'estimateurs asymptotiquement sans biais du paramètre N .
- c) Montrer que $A_n(\varepsilon) \subset B_n(\varepsilon)$ et qu'il existe un entier naturel n_0 tel que, pour tout $n > n_0$, on a : $B_n(\varepsilon) \subset [|T_n - E(T_n)| \geq \varepsilon/2]$.
- d) En déduire que la suite d'estimateurs $(T_n)_{n \geq 1}$ est convergente.

18. a) Calculer, pour tout n -uplet $(u_1, u_2, \dots, u_n)$ de $\llbracket 1, N \rrbracket^n$, $P\left(\bigcap_{i=1}^n [U_i = u_i]\right)$.

b) En déduire que, pour tout k de $\llbracket 1, N \rrbracket$, la loi conditionnelle du vecteur aléatoire $(U_1, U_2, \dots, U_n)$ sachant $[T_n = k]$ est donnée par :

$$P_{[T_n=k]}\left(\bigcap_{i=1}^n [U_i = u_i]\right) = \begin{cases} \frac{1}{k^n - (k-1)^n} & \text{si pour tout } i \text{ de } \llbracket 1, n \rrbracket, 1 \leq u_i \leq N \text{ et } \max_{1 \leq i \leq n} (u_i) = k \\ 0 & \text{sinon} \end{cases}$$

On remarquera que cette loi conditionnelle ne dépend pas du paramètre N .

19. On pose, pour n entier de $\mathbb{N}^*$: $S_n = T_n + Z_n - 1$ et, pour tout k de $\llbracket 1, N \rrbracket$: $\psi_n(k) = \frac{k^{n+1} - (k-1)^{n+1}}{k^n - (k-1)^n}$.

- a) Montrer que S_n est un estimateur sans biais de N .
- b) Établir, pour tout k de $\llbracket 1, N \rrbracket$, l'égalité : $\psi_n(k) = E(S_n/[T_n = k])$.
- c) En déduire que $\psi_n(T_n)$ est un estimateur sans biais de N .
- d) On pose, pour tout k de $\llbracket 1, N \rrbracket$: $\varphi_n(k) = E(S_n^2/[T_n = k])$. Établir, pour tout k de $\llbracket 1, N \rrbracket$, l'inégalité : $\psi_n^2(k) \leq \varphi_n(k)$ (on pourra utiliser la fonction définie sur $\mathbb{R}$ par : $\lambda \mapsto E((S_n - \lambda)^2/[T_n = k])$). En déduire que $V(\psi_n(T_n)) \leq V(S_n)$.
- e) Calculer $V(S_n)$. En déduire que $\psi_n(T_n)$ est un estimateur convergent de N .

20. Soit, pour n entier de $\mathbb{N}^*$, un estimateur sans biais R_n du paramètre N .

On pose, pour tout k de $\llbracket 1, N \rrbracket$: $f_n(k) = E(R_n/[T_n = k])$.

- a) En utilisant une méthode analogue à celle de la question 19.d, montrer que : $V(f_n(T_n)) \leq V(R_n)$.
- b) Soit F une fonction réelle. Montrer que, pour n fixé dans $\mathbb{N}^*$, la condition « pour tout N de $\mathbb{N}^*$, $E(F(T_n)) = N$ » est vérifiée, si et seulement si, pour tout k de $\llbracket 1, N \rrbracket$, on a : $F(k) = \psi_n(k)$.
- c) En déduire que dans l'ensemble des estimateurs sans biais de N , l'estimateur $\psi_n(T_n)$ est optimal, dans le sens où $V(\psi_n(T_n))$ est minimale.

La partie IV constitue une démonstration du théorème de Lehmann-Scheffé dans le cas particulier d'une loi uniforme sur $\llbracket 1, N \rrbracket$, avec N inconnu.



BANQUE COMMUNE D'ÉPREUVES

CONCOURS D'ADMISSION DE 2010

Conceptions : H.E.C. – E.S.C.P. / EUROPE

283

OPTION SCIENTIFIQUE

MATHÉMATIQUES II

CCIP_M2_S

Lundi 10 mai 2010, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite. Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre

Dans tout le problème :

- toutes les variables aléatoires sont supposées définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$;
- pour tout réel $t > 0$, X_t désigne une variable aléatoire à valeurs strictement positives qui suit la loi gamma de paramètre t , notée $\gamma(t)$;
- U désigne une variable aléatoire qui suit la loi uniforme sur $]0, 1]$;
- l'espérance et la variance d'une variable aléatoire A sont notées respectivement $E(A)$ et $V(A)$;
- la notation $\exp$ désigne la fonction exponentielle de base e .

On rappelle ou on admet sans démonstration les résultats suivants :

- la fonction Γ définie pour tout réel $t > 0$ par $\Gamma(t) = \int_0^{+\infty} e^{-u} u^{t-1} du$, est de classe C^∞ sur $\mathbb{R}^{+*}$; on note Γ' et Γ'' les dérivées première et seconde de la fonction Γ . Pour tout réel $t > 0$, les intégrales $\int_0^{+\infty} (\ln u) e^{-u} u^{t-1} du$ et $\int_0^{+\infty} (\ln u)^2 e^{-u} u^{t-1} du$ sont convergentes et valent respectivement $\Gamma'(t)$ et $\Gamma''(t)$;
- on a pour tout réel $t > 0$: $\Gamma(t+1) = t\Gamma(t)$;
- pour tout réel $t > 0$, une densité f_{X_t} de X_t est donnée par : $f_{X_t}(x) = \begin{cases} \frac{1}{\Gamma(t)} e^{-x} x^{t-1} & \text{si } x > 0 ; \\ 0 & \text{si } x \leq 0 \end{cases}$;
- $\Gamma(1/2) = \sqrt{\pi}$.

L'objet du problème est de démontrer quelques propriétés de la fonction Γ en utilisant des méthodes essentiellement probabilistes.

Partie I. Quelques résultats préliminaires

1. On pose pour tout n de $\mathbb{N}^*$: $h_n = \sum_{k=1}^n \frac{1}{k}$. On considère les deux suites $(\gamma_n)_{n \geq 1}$ et $(v_n)_{n \geq 1}$ définies par : pour tout n de $\mathbb{N}^*$, $\gamma_n = h_n - \ln n$ et $v_n = \gamma_{n+1} - \gamma_n$.

- a) Montrer que la série de terme général v_n est convergente.
- b) En déduire la convergence de la suite $(\gamma_n)_{n \geq 1}$; on note γ sa limite.
- c) On pose pour tout réel $t > 0$: $d_{n,t} = \gamma + \ln(t+n) - h_n$. Déterminer $\lim_{n \rightarrow +\infty} d_{n,t}$.
2. a) Rappeler sans démonstration les valeurs respectives de $E(X_t)$ et $V(X_t)$.
- b) On note pour tout réel $t > 0$, $\psi(t) = \Gamma'(t)/\Gamma(t)$, et ψ' la dérivée de ψ . Montrer que $E(\ln(X_t)) = \psi(t)$ et $V(\ln(X_t)) = \psi'(t)$.
3. a) Montrer que pour tout réel $t > 1$, $E(1/X_t)$ existe et calculer sa valeur.
- b) Établir pour tout réel $x > 0$, l'encadrement : $1 - \frac{1}{x} \leq \ln x \leq x - 1$.
- En déduire que l'on a : $(\ln x)^2 \leq \left(1 - \frac{1}{x}\right)^2 + (x-1)^2$.
- c) À l'aide des questions précédentes, établir les inégalités suivantes : pour tout réel $t > 0$, $E\left(\ln\left(\frac{X_t}{t}\right)\right) \leq 0$;
pour tout réel $t > 1$, $E\left(\ln\left(\frac{X_t}{t}\right)\right) \geq -\frac{1}{t-1}$; pour tout réel $t > 2$, $E\left(\ln^2\left(\frac{X_t}{t}\right)\right) \leq \frac{2t}{(t-2)^2}$.
- d) Soit t un réel fixé strictement positif. Montrer que la suite de variables aléatoires $\left(\ln\left(\frac{X_{t+n}}{t+n}\right)\right)_{n \geq 1}$ converge en probabilité vers 0.
4. Soit $(A_n)_{n \geq 1}$, $(B_n)_{n \geq 1}$ et $(C_n)_{n \geq 1}$ trois suites de variables aléatoires à densité qui convergent en probabilité vers 0. On pose pour tout n de $\mathbb{N}^*$: $D_n = A_n + B_n + C_n$. Soit $(u_n)_{n \geq 1}$ une suite réelle qui converge vers u . On considère deux variables aléatoires réelles à densité M et N telles que pour tout n de $\mathbb{N}^*$, M est de même loi que $N + D_n + u_n$.
- a) Montrer pour tout réel $\varepsilon > 0$, l'inclusion : $\{|D_n| > \varepsilon\} \subset \{|A_n| > \varepsilon/3\} \cup \{|B_n| > \varepsilon/3\} \cup \{|C_n| > \varepsilon/3\}$.
En déduire que la suite $(D_n)_{n \geq 1}$ converge en probabilité vers 0.
- b) On pose pour tout n de $\mathbb{N}^*$: $V_n = D_n + u_n - u$. Montrer que la suite de variables aléatoires $(V_n)_{n \geq 1}$ converge en probabilité vers 0. En déduire la limite en probabilité de la suite $((N + u) + V_n)_{n \geq 1}$.
- c) On admet sans démonstration que la convergence en probabilité entraîne la convergence en loi.
Montrer que les variables aléatoires M et $N + u$ sont de même loi.
5. Soit (α, β) un couple de réels strictement positifs, et X_α et X_β deux variables aléatoires indépendantes de lois respectives $\gamma(\alpha)$ et $\gamma(\beta)$. On pose : $T_{\alpha,\beta} = \frac{X_\alpha}{X_\beta}$, $Q_{\alpha,\beta} = \ln(T_{\alpha,\beta})$ et $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.
- a) Préciser $Q_{\alpha,\beta}(\Omega)$. Déterminer une densité de $\ln(X_\alpha)$ et de $-\ln(X_\beta)$ respectivement.
- b) En déduire qu'une densité $f_{Q_{\alpha,\beta}}$ de $Q_{\alpha,\beta}$ est donnée par : pour tout réel x ,

$$f_{Q_{\alpha,\beta}}(x) = \frac{e^{-\beta x}}{\Gamma(\alpha)\Gamma(\beta)} \int_{-\infty}^{+\infty} e^{(\alpha+\beta)y} \exp(-e^y(1+e^{-x})) dy$$

- c) À l'aide du changement de variable $u = e^y(1+e^{-x})$, dont on justifiera la validité, établir la formule suivante : pour tout x réel, $f_{Q_{\alpha,\beta}}(x) = \frac{1}{B(\alpha, \beta)} \times \frac{e^{\alpha x}}{(1+e^x)^{\alpha+\beta}}$.

- d) En déduire une densité $f_{T_{\alpha,\beta}}$ de $T_{\alpha,\beta}$.

- e) On pose : $J_{\alpha,\beta} = \frac{X_\alpha}{X_\alpha + X_\beta}$. Montrer qu'une densité $f_{J_{\alpha,\beta}}$ de $J_{\alpha,\beta}$ est donnée par :

$$f_{J_{\alpha,\beta}}(z) = \begin{cases} 0 & \text{si } z \notin]0, 1[\\ \frac{1}{B(\alpha, \beta)} z^{\alpha-1} (1-z)^{\beta-1} & \text{si } z \in]0, 1[\end{cases}$$

Partie II. Étude de la variable aléatoire $\ln(X_t)$

Soit $(Y_k)_{k \geq 1}$ une suite de variables aléatoires indépendantes de même loi exponentielle de paramètre 1. On suppose que pour tout réel $\alpha > 0$, X_α est indépendante de chacune des variables aléatoires de la suite $(Y_k)_{k \geq 1}$.

On pose : $S_0 = 0$ et pour tout k de $\mathbb{N}^*$, $S_k = \sum_{i=1}^k Y_i$.

6. Rappeler sans démonstration la loi de S_k ainsi que les valeurs respectives de $E(S_k)$ et $V(S_k)$.

7. a) Justifier pour tout n de $\mathbb{N}^*$, l'égalité suivante : $\ln(X_t) = \ln\left(\prod_{k=1}^n \frac{X_t + S_{k-1}}{X_t + S_k}\right) + \ln(X_t + S_n)$.

b) Montrer que pour tout entier m de $\mathbb{N}^*$, la loi de $X_t + S_m$ est celle de X_{t+m} .

On admet jusqu'à la fin du problème les résultats suivants :

- soit n un entier de $\mathbb{N}^*$ et $A_1, \dots, A_n$ des variables aléatoires réelles mutuellement indépendantes. Alors, pour tout p de $[1, n-1]$, pour toutes fonctions réelles φ_1 et φ_2 continues, les variables aléatoires $\varphi_1(A_1, \dots, A_p)$ et $\varphi_2(A_{p+1}, \dots, A_n)$ sont indépendantes ;
- si A et B sont deux variables aléatoires indépendantes admettant une espérance, alors AB admet une espérance et $E(AB) = E(A)E(B)$;
- pour tout couple (α, β) de réels strictement positifs, si les variables aléatoires X_α et X_β sont indépendantes, de lois respectives $\gamma(\alpha)$ et $\gamma(\beta)$, alors les variables aléatoires $\frac{X_\alpha}{X_\alpha + X_\beta}$ et $X_\alpha + X_\beta$ sont indépendantes.

8. On pose pour tout k de $\mathbb{N}^*$: $R_{t,k} = \frac{X_t + S_{k-1}}{X_t + S_k}$.

a) Montrer que $R_{t,1}$ et $R_{t,2}$ sont indépendantes. On admet dans la suite que les variables aléatoires $R_{t,k}$ ($k \geq 1$) sont indépendantes.

b) En déduire à l'aide des questions précédentes que pour tout n de $\mathbb{N}^*$, les variables aléatoires $\ln(X_t)$ et $\sum_{k=1}^n \ln(R_{t,k}) + \ln(X_{t+n})$ sont de même loi.

9. a) Déterminer la loi de la variable aléatoire $-\ln U$.

b) À l'aide de la question 5.e, calculer une densité $f_{R_{t,k}}$ de la variable aléatoire $R_{t,k}$.

c) Soit $(U_k)_{k \geq 1}$ une suite de variables aléatoires indépendantes de même loi que U . Montrer que pour tout k de $\mathbb{N}^*$, les variables aléatoires $R_{t,k}$ et $U_k^{\frac{1}{t+k-1}}$ sont de même loi.

d) Déduire des questions précédentes que pour tout n de $\mathbb{N}^*$, la variable aléatoire $\ln(X_t)$ est de même loi que la variable aléatoire $\sum_{k=0}^{n-1} \left(\frac{1}{k+1} - \frac{Y_{k+1}}{t+k} \right) + \ln\left(\frac{X_{t+n}}{t+n}\right) + d_{n,t} - \gamma$.

10. En utilisant les résultats des questions 1.c, 2.b, 3.c et 9.d, montrer que pour tout réel $t > 0$, on a :

$$E(\ln(X_t)) = \psi(t) = -\gamma + \sum_{k=0}^{\infty} \left(\frac{1}{k+1} - \frac{1}{t+k} \right) \text{ et } V(\ln(X_t)) = \psi'(t) = \sum_{k=0}^{\infty} \frac{1}{(t+k)^2}$$

11. En utilisant la question 3.c, calculer $\lim_{t \rightarrow +\infty} (\psi(t) - \ln t)$.

12. On pose : $W = U^{1/\mu}$, où μ désigne un paramètre réel strictement positif inconnu. Afin d'estimer μ , on considère pour p supérieur ou égal à 3, un p -échantillon $(W_1, W_2, \dots, W_p)$ i.i.d. de la loi de W .

On pose : $G_p = -p \left(\sum_{i=1}^p \ln W_i \right)^{-1}$. Justifier que la variable aléatoire G_p est un estimateur du paramètre μ .

Est-il sans biais ? Est-il convergent ?

13. On rappelle que l'appel à la fonction Pascal `random` a pour résultat un nombre de type `real` pris au hasard dans l'intervalle $[0, 1[$.

a) Soit `X` la fonction Pascal suivante :

```
function X(lambda : real) : real ;
begin
  X := -ln(1-random)/lambda
end ;
```

Cette fonction simule une variable aléatoire réelle. Donner sa loi. Justifier votre réponse.

b) Écrire une fonction Pascal d'en-tête `g(n : integer) : real` simulant une variable aléatoire de loi $\gamma(n)$.

c) Soit `m` la fonction Pascal suivante :

```
function m(p : integer) : real;
begin
  m := p/g(p)
end;
```

On appelle la fonction `m` pour différentes valeurs de p de plus en plus grandes. Que devrait-on constater ?

Partie III. Quelques propriétés de la fonction Γ

Les notations sont celles des parties I et II.

14. Première application : les formules de Wilks et Legendre.

a) Soit $(a_k)_{k \in \mathbb{N}}$ une suite réelle. Pour tout n de $\mathbb{N}^*$ et tout réel $t > 0$, établir l'égalité :

$$2 \sum_{k=0}^{2n-1} \left(\frac{1}{k+1} - \frac{a_{k+1}}{t+k} \right) = \sum_{k=0}^{n-1} \left(\frac{1}{k+1} - \frac{a_{2k+1}}{\frac{t}{2}+k} \right) + \sum_{k=0}^{n-1} \left(\frac{1}{k+1} - \frac{a_{2k+2}}{\frac{t+1}{2}+k} \right) + 2 \sum_{k=0}^{n-1} \left(\frac{1}{2k+1} - \frac{1}{2k+2} \right)$$

b) Exprimer $w_n = \sum_{k=0}^{n-1} \left(\frac{1}{2k+1} - \frac{1}{2k+2} \right)$ en fonction de deux termes de la suite $(h_n)_{n \geq 1}$. En déduire que $\lim_{n \rightarrow +\infty} w_n = \ln 2$.

c) Pour $t > 0$, soit X_t et $X_{t+\frac{1}{2}}$ deux variables aléatoires indépendantes de lois respectives $\gamma(t)$ et $\gamma(t+\frac{1}{2})$.

En utilisant les questions 4 et 9.d, montrer que la variable aléatoire $2 \ln(X_t)$ est de même loi que la variable aléatoire $\ln(X_{\frac{t}{2}}) + \ln(X_{\frac{t+1}{2}}) + 2 \ln 2$.

d) On pose : $t = 2s$. Déduire de la question précédente que pour tout réel $r > 0$, $(X_{2s})^{2r}$ et $2^{2r}(X_s)^r(X_{s+\frac{1}{2}})^r$ sont de même loi.

e) En choisissant une valeur particulière de s , établir pour tout $r > 0$, la formule :

$$2^{2r-1} \Gamma(r) \Gamma(r + \frac{1}{2}) = \Gamma(2r) \sqrt{\pi}$$

15. Deuxième application : la formule de Stirling.

a) Déterminer quatre réels a, b, c, d tels que pour tout réel $u > 0$, on a : $\frac{1}{u^2(u+1)^2} = \frac{a}{u^2} + \frac{b}{(u+1)^2} + \frac{c}{u} + \frac{d}{u+1}$.

En déduire pour tout $t > 0$, la relation : $\psi'(t) = \frac{1}{t} + \frac{1}{2t^2} + \frac{1}{2} \sum_{k=0}^{+\infty} \frac{1}{(t+k)^2(t+k+1)^2}$.

On admet sans démonstration que pour tout $u > 0$, on a :

$$\frac{1}{3} \left(\frac{1}{(u + \frac{1}{14})^3} - \frac{1}{(u + \frac{15}{14})^3} \right) \leq \frac{1}{u^2(u+1)^2} \leq \frac{1}{3} \left(\frac{1}{u^3} - \frac{1}{(u+1)^3} \right)$$

b) Déduire des deux résultats précédents, pour tout $t > 0$, les deux encadrements :

$$\frac{1}{2t^2} + \frac{1}{6(t + \frac{1}{14})^3} \leq \psi'(t) - \frac{1}{t} \leq \frac{1}{2t^2} + \frac{1}{6t^3} \text{ et } \ln t - \frac{1}{2t} - \frac{1}{12t^2} \leq \psi(t) \leq \ln t - \frac{1}{2t} - \frac{1}{12(t + \frac{1}{14})^2}$$

En déduire un équivalent de $E(\ln(X_t))$ et de $V(\ln(X_t))$ respectivement, lorsque t tend vers $+\infty$.

c) Calculer pour tout y vérifiant $y > t > 0$, l'intégrale : $\int_t^y \left(\psi(x) - \ln(x) + \frac{1}{2x} \right) dx$.

Montrer pour t fixé, l'existence de $\lim_{y \rightarrow +\infty} \ln \left(\frac{\Gamma(y)}{y^{y-\frac{1}{2}} e^{-y}} \right)$; on note θ cette limite.

d) En utilisant la question 14.e et l'identité : $x^x = \left(x + \frac{1}{2}\right)^x \left(1 + \frac{1}{2x}\right)^{-x}$, valable pour $x > 0$, calculer e^θ .

En déduire que $\Gamma(x)$ est équivalent à $\sqrt{2\pi} x^{x-\frac{1}{2}} e^{-x}$, lorsque x tend vers $+\infty$.



BANQUE COMMUNE D'ÉPREUVES

CONCOURS D'ADMISSION DE 2011

Conceptions : H.E.C. – E.S.C.P. / EUROPE

Code épreuve :

283

OPTION SCIENTIFIQUE

MATHEMATIQUES II

Lundi 9 mai 2011, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite. Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre

Toutes les variables aléatoires qui apparaissent dans ce problème sont supposées définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$.

On note F_Z la fonction de répartition d'une variable aléatoire Z et, si cette variable aléatoire admet une densité, on note f_Z une densité de Z .

Sous réserve d'existence, on note $E(Z)$ et $V(Z)$ respectivement, l'espérance et la variance d'une variable aléatoire réelle Z , et $\text{Cov}(Z_1, Z_2)$ la covariance de deux variables aléatoires Z_1 et Z_2 .

La fonction exponentielle est notée $\exp$ et la partie entière d'un réel x est notée $[x]$.

On admet les résultats suivants :

- la définition et les propriétés de la covariance et du coefficient de corrélation linéaire de deux variables aléatoires discrètes, s'appliquent au cas de variables aléatoires à densité ;
- si Z_1 et Z_2 sont deux variables aléatoires indépendantes admettant une espérance, alors la variable aléatoire $Z_1 Z_2$ admet une espérance et $E(Z_1 Z_2) = E(Z_1)E(Z_2)$.

Dans tout le problème, on considère une variable aléatoire X de fonction de répartition F_X et admettant une densité f_X .

Les solutions éventuelles de l'équation $F_X(x) = \frac{1}{2}$ s'appellent les *médianes théoriques* de X .

Pour n entier de $\mathbb{N}^*$, on considère un n -échantillon $(X_1, X_2, \dots, X_n)$ i.i.d. (indépendant, identiquement distribué) de la loi de X et on définit la variable aléatoire : $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$, qui est la *moyenne empirique* de l'échantillon $(X_1, X_2, \dots, X_n)$.

On admet l'existence de variables aléatoires à densité $Y_1, Y_2, \dots, Y_n$ telles que, pour tout ω de Ω , les réels $Y_1(\omega), Y_2(\omega), \dots, Y_n(\omega)$ constituent un réarrangement par ordre croissant des réels $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$, de telle sorte que, pour tout ω de Ω : $Y_1(\omega) \leq Y_2(\omega) \leq \dots \leq Y_n(\omega)$.

En particulier, $Y_1 = \inf(X_1, X_2, \dots, X_n)$ et $Y_n = \sup(X_1, X_2, \dots, X_n)$. Plus généralement, pour tout k de $[1, n]$, il existe une fonction ψ_k définie et continue sur $\mathbb{R}^n$ à valeurs réelles, telle que $Y_k = \psi_k(X_1, X_2, \dots, X_n)$.

Si n est un entier impair ($n = 2\ell + 1$, avec $\ell \in \mathbb{N}$), alors la variable aléatoire $Y_{\ell+1}$ est appelée la médiane empirique de l'échantillon $(X_1, X_2, \dots, X_n)$.

La partie II du problème est indépendante de la partie I.

Partie I. Quelques propriétés des statistiques d'ordre

Pour tout réel x et tout entier k de $[1, n]$, on note $J_k(x)$ la variable aléatoire de Bernoulli définie par :

$$J_k(x) = \begin{cases} 1 & \text{si } [X_k \leq x] \text{ est réalisé} \\ 0 & \text{si } [X_k > x] \text{ est réalisé} \end{cases} ; \quad \text{on pose : } S_n(x) = \sum_{k=1}^n J_k(x)$$

1. a) Montrer que les fonctions f_{Y_1} et f_{Y_n} définies pour tout x réel par : $f_{Y_1}(x) = n(1 - F_X(x))^{n-1}f_X(x)$ et $f_{Y_n}(x) = n(F_X(x))^{n-1}f_X(x)$, sont des densités de Y_1 et Y_n respectivement.

b) Quelle est la loi de probabilité de la variable aléatoire $S_n(x)$?

c) Justifier l'égalité entre événements suivante : $[Y_k \leq x] = [S_n(x) \geq k]$.

d) Établir la relation : pour tout x réel, $F_{Y_k}(x) = \sum_{j=k}^n \binom{n}{j} (F_X(x))^j (1 - F_X(x))^{n-j}$.

e) En déduire que pour tout k de $[1, n]$, la fonction f_{Y_k} définie pour tout x réel par :

$$f_{Y_k}(x) = k \binom{n}{k} (F_X(x))^{k-1} (1 - F_X(x))^{n-k} f_X(x)$$

est une densité de Y_k .

f) Montrer que si X admet un moment d'ordre r ($r \in \mathbb{N}^*$), alors pour tout k de $[1, n]$, Y_k admet un moment d'ordre r .

Exemple. Dans les questions 2 à 4, on suppose que la fonction de répartition F_X est donnée par :

$$F_X(x) = \begin{cases} 1 - \frac{1}{\sqrt{x}} & \text{si } x \geq 1 \\ 0 & \text{si } x < 1 \end{cases}$$

2. a) Tracer la courbe représentative de F_X dans le plan rapporté à un repère orthonormé.

Préciser la demi-tangente à droite au point d'abscisse $x = 1$.

Justifier que X est une variable aléatoire à densité et préciser une densité f_X de X .

b) Montrer que X n'admet aucun moment.

c) Établir l'unicité de la médiane théorique M de X . Calculer M .

d) Expliciter, pour tout k de $[1, n]$ et pour tout x réel, l'expression $f_{Y_k}(x)$ d'une densité de Y_k .

En déduire un équivalent de $f_{Y_k}(x)$ lorsque x tend vers $+\infty$.

3. On suppose dans cette question que $n \geq 3$.

a) Montrer que pour tout k de $[1, n-2]$, Y_k admet une espérance.

b) En justifiant l'emploi du changement de variable $t = \frac{1}{\sqrt{x}}$, établir pour tout k de $[1, n-2]$,

$$\text{la formule : } E(Y_k) = k \binom{n}{k} \int_0^1 t^{n-k-2} (1-t)^{k-1} dt.$$

c) Pour tout couple (r, s) de $(\mathbb{N}^*)^2$, on pose : $I_{r,s} = \int_0^1 t^{r-1} (1-t)^{s-1} dt$.

$$\text{Montrer que pour tout couple } (r, s) \text{ de } (\mathbb{N}^*)^2, \text{ on a : } I_{r,s} = \frac{(r-1)!(s-1)!}{(r+s-1)!}.$$

d) En déduire l'expression de $E(Y_k)$ pour tout k de $[1, n-2]$.

e) On suppose que n est impair et supérieur ou égal à 5, et on pose $n = 2\ell + 1$. Justifier la définition de

la médiane empirique $Y_{\ell+1}$ d'un échantillon, et établir l'égalité : $E(Y_{\ell+1}) = 4 + \frac{6}{\ell-1}$. Commenter.

4. On pose pour tout n de $\mathbb{N}^*$: $Z_n = \frac{1}{n^2} \sup(X_1, X_2, \dots, X_n) = \frac{Y_n}{n^2}$.

a) Calculer pour tout x réel, $F_{Z_n}(x)$.

b) On définit la fonction φ_Z par : $\varphi_Z(x) = \begin{cases} \exp\left(-\frac{1}{\sqrt{x}}\right) & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$.

Montrer que φ_Z est la fonction de répartition d'une variable aléatoire Z à densité.

c) Montrer que la suite de variables aléatoires $(Z_n)_{n \in \mathbb{N}^*}$ converge en loi vers Z .

Partie II. Existence et unicité d'un estimateur optimal

Dans cette partie, X suit la loi normale d'espérance θ et de variance égale à 1. On suppose que le paramètre réel θ est inconnu.

On note Φ la fonction de répartition de la loi normale centrée réduite.

On rappelle que pour n entier de $\mathbb{N}^*$, $(X_1, X_2, \dots, X_n)$ est un n -échantillon i.i.d. de la loi de X .

5. Quelle est la loi de $\bar{X}_n$? Montrer que $\bar{X}_n$ est un estimateur sans biais et convergent du paramètre θ .

6. Soit α un réel tel que $0 < \alpha < 1$. On appelle *marge d'erreur* associée à un intervalle de confiance de θ au risque α , le réel positif noté $\mu(\alpha)$, égal à la demi-longueur de cet intervalle.

a) Justifier l'existence de la fonction réciproque Φ^{-1} de la fonction Φ .

b) Déterminer un intervalle de confiance du paramètre θ au risque α dont le milieu est $\bar{X}_n$.

Vérifier que $\mu(\alpha) = -\frac{\Phi^{-1}(\alpha/2)}{\sqrt{n}}$.

c) On considère un risque β ($\beta \neq \alpha$) tel que $\mu(\beta) = b\mu(\alpha)$, avec $0 < b < 1$. Exprimer β en fonction de α .

Comparer α et β . Commenter.

7. On note $\mathcal{E}_\theta$ l'ensemble des statistiques $U_n = g_n(X_1, X_2, \dots, X_n)$, où g_n est une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$, qui sont des estimateurs sans biais de θ et qui admettent une variance.

Sous réserve d'existence, on dit qu'un élément Z_n de $\mathcal{E}_\theta$ est un estimateur optimal dans $\mathcal{E}_\theta$, si pour tout U_n de $\mathcal{E}_\theta$, on a : $V(Z_n) \leq V(U_n)$.

On admet que pour tout U_n de $\mathcal{E}_\theta$, on a : $\text{Cov}(\bar{X}_n, U_n - \bar{X}_n) = 0$.

a) Montrer que $\mathcal{E}_\theta$ n'est pas vide.

b) Montrer que $\bar{X}_n$ est optimal dans $\mathcal{E}_\theta$.

c) Soit Z_n un estimateur optimal dans $\mathcal{E}_\theta$. On pose pour tout U_n de $\mathcal{E}_\theta$ et pour tout λ réel :

$$A_n(\lambda) = Z_n + \lambda(U_n - Z_n).$$

Montrer que $A_n(\lambda)$ est un élément de $\mathcal{E}_\theta$. Calculer $V(A_n(\lambda))$. En déduire que $\text{Cov}(Z_n, U_n - Z_n) = 0$.

d) On suppose l'existence de deux estimateurs optimaux $\bar{X}_n$ et Z_n dans $\mathcal{E}_\theta$.

Montrer que $Z_n = \bar{X}_n$ presque sûrement. Conclure.

8. a) Justifier l'existence et l'unicité de la médiane théorique M de X , et exprimer M en fonction de θ .

b) Calculer $f_X(M)$. Montrer que pour tout réel x , on a : $F_X(2M - x) = 1 - F_X(x)$.

En déduire une relation entre $f_X(2M - x)$ et $f_X(x)$.

c) Établir pour tout k de $[1, n]$, la relation : $E(Y_k - M) = E(M - Y_{n-k+1})$.

d) En supposant que $n = 2\ell + 1$ ($\ell \in \mathbb{N}$), calculer $E(Y_{\ell+1})$, puis justifier que $V(Y_{\ell+1}) \geq \frac{1}{n}$. Commenter.

Partie III. Résultats asymptotiques

Le contexte de cette partie est identique à celui de la partie II.

Dans cette partie, on note T une variable aléatoire qui suit la loi normale centrée réduite.

Si U est une variable aléatoire et s un réel tels que la variable aléatoire $\exp(sU)$ admette une espérance, on pose : $L_U(s) = \exp(sU)$.

9. Soit J une variable aléatoire qui suit la loi de Bernoulli de paramètre p ($0 < p < 1$).

- Calculer pour tout s réel, $L_J(s)$.
- Établir pour tout s réel, l'existence de $L_T(s)$.
- Calculer pour tout s réel, $L_T(s)$. En déduire que pour tout couple (θ, σ) de $\mathbb{R} \times \mathbb{R}^{++}$ et pour tout s réel, on a : $L_{\sigma T + \theta}(s) = \exp\left(\sigma^2 \frac{s^2}{2} + \theta s\right)$.

Dans les questions 10 et 11, x est un réel fixé.

10. On pose pour tout n de $\mathbb{N}^*$: $y_n = M + \frac{x}{\sqrt{n}}$, $q_n = F_X(y_n)$ et $k(n) = \lfloor n/2 \rfloor + 1$.

- Montrer que l'on a : $k(n) = \frac{n}{2} + o(\sqrt{n})$ lorsque n tend vers $+\infty$.
- En appliquant la formule de Taylor-Young à la fonction F_X au voisinage de M , justifier la relation :

$$q_n = \frac{1}{2} + \frac{x}{\sqrt{2\pi n}} + o\left(\frac{1}{\sqrt{n}}\right)$$

- Soit $(u_n)_{n \in \mathbb{N}^*}$ la suite définie par : pour tout n de $\mathbb{N}^*$, $u_n = \frac{1}{\sqrt{n}}(k(n) - nq_n)$.

Montrer que la suite $(u_n)_{n \in \mathbb{N}^*}$ est convergente et déterminer sa limite u .

11. On pose pour tout n de $\mathbb{N}^*$: $W_n = \frac{1}{\sqrt{n}}(S_n(y_n) - nq_n)$, où $S_n(y_n)$ a été définie dans le préambule de la partie I.

- Établir pour tout réel s , la relation :

$$L_{W_n}(s) = \exp(-s\sqrt{n}q_n) \left(1 + q_n \exp\left(\frac{s}{\sqrt{n}}\right) - q_n\right)^n$$

- En utilisant un développement limité à l'ordre 2, montrer que l'on a : $\ln(L_{W_n}(s)) = \frac{s^2}{8} + o(1)$ lorsque n tend vers $+\infty$. En déduire l'égalité : $\lim_{n \rightarrow +\infty} L_{W_n}(s) = L_{\frac{T}{2}}(s)$.

On admet alors que la suite de variables aléatoires $(W_n)_{n \in \mathbb{N}^*}$ converge en loi vers la variable aléatoire $\frac{T}{2}$.

12. On suppose que $x = 0$. Quels sont les arguments qui permettent d'obtenir directement le résultat final de la question 11.b) ?

13. a) Établir pour tout n de $\mathbb{N}^*$ et pour tout x réel, les égalités d'événements suivantes :

$$[\sqrt{n}(Y_{k(n)} - M) \leq x] = [S_n(y_n) \geq k(n)] = [W_n \geq u_n]$$

- Montrer l'égalité : $\lim_{n \rightarrow +\infty} P([W_n \geq u_n]) = P\left(\left[\sqrt{\frac{\pi}{2}}T \leq x\right]\right)$.

- En déduire que la suite de variables aléatoires $\left(\sqrt{\frac{2n}{\pi}}(Y_{k(n)} - M)\right)_{n \in \mathbb{N}^*}$ converge en loi et préciser sa limite.

14. On suppose dans cette question que n est impair et on pose $n = 2\ell + 1$ ($\ell \in \mathbb{N}$).

On note ρ_n le coefficient de corrélation linéaire de $Y_{k(n)}$ et $\bar{X}_n$.

- Que vaut $k(n)$?
- Préciser la valeur de $\lim_{n \rightarrow +\infty} E(Y_{k(n)})$.

- On admet sans démonstration que la suite réelle de terme général $E\left(\left(\sqrt{\frac{2n}{\pi}}(Y_{k(n)} - M)\right)^2\right)$

converge vers $E(T^2)$. En déduire un équivalent de $V(Y_{k(n)})$ lorsque n tend vers $+\infty$.

- À l'aide des questions 5,7,8 et 14.c), déterminer la limite de ρ_n lorsque n tend vers $+\infty$.



BANQUE COMMUNE D'ÉPREUVES

CONCOURS D'ADMISSION DE 2012

Conception : C.C.I.P.

OPTION SCIENTIFIQUE

Code épreuve : 283

MATHEMATIQUES II

Mercredi 9 mai 2012, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite. Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre

- Toutes les variables aléatoires qui interviennent dans ce problème sont réelles et définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$, où P peut dépendre de paramètres réels inconnus a, b, σ etc ; elles admettent toutes une espérance et une variance : si J désigne l'une de ces variables aléatoires, on note $E(J)$ son espérance et $V(J)$ sa variance.

Si J_1, J_2 et $J_1 + J_2$ sont des variables aléatoires à densité, on admet alors l'existence de la covariance de J_1 et J_2 , notée $\text{Cov}(J_1, J_2)$, qui est définie par la formule : $\text{Cov}(J_1, J_2) = \frac{1}{2}(V(J_1 + J_2) - V(J_1) - V(J_2))$.

On admet que les covariances de variables aléatoires à densité vérifient les mêmes règles de calcul que celles des variables aléatoires discrètes.

- Pour tout (k, ℓ) de $(\mathbb{N}^*)^2$, on note $\mathcal{M}_{k, \ell}(\mathbb{R})$ l'ensemble des matrices à k lignes et ℓ colonnes à coefficients réels ; on note $\mathcal{M}_k(\mathbb{R})$ l'ensemble des matrices carrées d'ordre k .
- On note tQ la transposée d'une matrice Q .
- Dans tout le problème, n désigne un entier supérieur ou égal à 3.

L'objet du problème est l'étude de quelques propriétés du modèle de régression linéaire élémentaire.

Partie I. Quelques résultats statistiques et algébriques

On considère une population d'individus statistiques dans laquelle on étudie deux caractères quantitatifs $\mathcal{X}$ et $\mathcal{Y}$. On extrait de cette population, un échantillon de n individus sélectionnés selon des valeurs choisies du caractère $\mathcal{X}$ et numérotés de 1 à n .

Pour tout i de $\llbracket 1, n \rrbracket$, les réels x_i et y_i sont les observations respectives de $\mathcal{X}$ et de $\mathcal{Y}$ pour l'individu i de l'échantillon. On suppose que les réels $x_1, x_2, \dots, x_n$ ne sont pas tous égaux.

Soit a et b deux paramètres réels. On pose pour tout i de $\llbracket 1, n \rrbracket$: $u_i = y_i - (ax_i + b)$. (*)

1. On note $\bar{x}$ (resp. $\bar{y}$) et s_x^2 (resp. s_y^2), la moyenne empirique et la variance empirique de la série statistique

$$(x_i)_{1 \leq i \leq n} \text{ (resp. } (y_i)_{1 \leq i \leq n}) ; \text{ on rappelle que : } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \text{ et } s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

- a) Montrer que $s_x^2 > 0$.

- b) Établir les formules : $\sum_{i=1}^n (x_i - \bar{x})y_i = \sum_{i=1}^n (x_i y_i) - n\bar{x}\bar{y}$ et $\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i^2) - n\bar{x}^2$.

- c) On pose pour tout i de $\llbracket 1, n \rrbracket$: $\alpha_i = \frac{(x_i - \bar{x})}{ns_x^2}$. Montrer que : $\sum_{i=1}^n \alpha_i = 0$, $\sum_{i=1}^n \alpha_i x_i = 1$ et $\sum_{i=1}^n \alpha_i^2 = \frac{1}{ns_x^2}$.
2. On pose : $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$, $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$, $\theta = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathcal{M}_{2,1}(\mathbb{R})$ et $M = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \in \mathcal{M}_{n,2}(\mathbb{R})$.
- Les n relations (*) s'écrivent sous la forme matricielle suivante : $y = M\theta + u$.
- a) Quel est le rang de la matrice M ?
- b) Calculer la matrice tMM et justifier son inversibilité.
3. L'espace vectoriel $\mathbb{R}^n$ est muni de sa structure euclidienne canonique. Soit $\mathcal{F}$ le sous-espace vectoriel engendré par les vecteurs $(x_1, x_2, \dots, x_n)$ et $(1, 1, \dots, 1)$ de $\mathbb{R}^n$. On note K la matrice du projecteur orthogonal de $\mathbb{R}^n$ sur $\mathcal{F}$ dans la base canonique de $\mathbb{R}^n$ et $G = I - K$, où I désigne la matrice identité de $\mathcal{M}_n(\mathbb{R})$.
- a) On cherche les matrices $\theta = \begin{pmatrix} a \\ b \end{pmatrix}$ de $\mathcal{M}_{2,1}(\mathbb{R})$ qui minimisent $\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - (ax_i + b))^2$.
- Montrer que ce problème admet une unique solution $\hat{\theta} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$ et qu'elle vérifie la relation : ${}^tMM\hat{\theta} = {}^tMy$.
- b) Montrer que : $\hat{a} = \sum_{i=1}^n \alpha_i y_i$ et $\hat{b} = \bar{y} - \hat{a}\bar{x}$.
- c) Exprimer K en fonction de M et tM .
- d) Soit $\hat{u}$ la matrice-colonne de $\mathcal{M}_{n,1}(\mathbb{R})$ de composantes $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_n$ définie par $\hat{u} = y - M\hat{\theta}$.
Montrer que : $\hat{u} = Gy = Gu$.
- e) En déduire les égalités : ${}^t\hat{u}\hat{u} = \sum_{i=1}^n \hat{u}_i^2 = {}^t_y Gy = {}^t_u Gu$.

Partie II. Le modèle de régression linéaire

Le contexte et les notations sont ceux de la partie I. Dans cette partie, on cherche à modéliser les fluctuations aléatoires du caractère $\mathcal{Y}$ sur l'échantillon.

Les hypothèses du modèle de régression linéaire élémentaire sont les suivantes :

- les réels a et b sont des paramètres inconnus ;
- pour tout i de $\llbracket 1, n \rrbracket$, la valeur x_i du caractère $\mathcal{X}$ est connue et la valeur y_i du caractère $\mathcal{Y}$ est la réalisation d'une variable aléatoire Y_i ;
- pour tout i de $\llbracket 1, n \rrbracket$, Y_i est la somme d'une composante déterministe $ax_i + b$, fonction affine de la valeur choisie x_i , et d'une composante aléatoire U_i ;
- les variables aléatoires $U_1, U_2, \dots, U_n$ sont mutuellement indépendantes, de même loi, possèdent une densité, et pour tout i de $\llbracket 1, n \rrbracket$: $E(U_i) = 0$ et $V(U_i) = \sigma^2$, où le paramètre inconnu σ est strictement positif.

Le modèle de régression linéaire s'écrit alors : pour tout i de $\llbracket 1, n \rrbracket$, $Y_i = ax_i + b + U_i$ (1).

L'objectif consiste à estimer les paramètres inconnus a , b et σ^2 du modèle (1).

On pose pour tout $n \geq 3$: $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$ et $\bar{U}_n = \frac{1}{n} \sum_{i=1}^n U_i$.

4. On note A_n et B_n les deux variables aléatoires définies par : $A_n = \sum_{i=1}^n \alpha_i Y_i$ et $B_n = \bar{Y}_n - A_n \bar{x}$, où le réel α_i

a été défini dans la question 1.c).

a) Montrer que A_n et B_n sont des estimateurs sans biais de a et b respectivement.

b) Établir les formules suivantes : $V(A_n) = \frac{\sigma^2}{ns_x^2}$ et $V(B_n) = \left(1 + \frac{\bar{x}^2}{s_x^2}\right) \frac{\sigma^2}{n}$.

c) Calculer $\text{Cov}(A_n, B_n)$.

5. Dans cette question uniquement, l'entier n n'est plus fixé. On suppose l'existence de $\lambda = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n x_i$ et

$$\mu^2 = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \text{ avec } (\lambda, \mu) \in \mathbb{R} \times \mathbb{R}_+^*.$$

Montrer que les deux suites $(A_n)_{n \geq 3}$ et $(B_n)_{n \geq 3}$ convergent en probabilité vers a et b respectivement.

6.a) On pose pour tout i de $\llbracket 1, n \rrbracket$: $\hat{U}_i = Y_i - A_n x_i - B_n$. Calculer $E(\hat{U}_i)$.

b) Établir l'égalité : $\sum_{i=1}^n \hat{U}_i^2 = \sum_{i=1}^n (U_i - \bar{U}_n)^2 - n s_x^2 (A_n - a)^2$.

c) Calculer $E\left(\sum_{i=1}^n \hat{U}_i^2\right)$. En déduire un estimateur sans biais de σ^2 .

Partie III. Hypothèse de normalité et prévision

Le contexte et les notations de cette partie sont ceux des parties I et II. De plus, on suppose dans cette partie que pour tout i de $\llbracket 1, n \rrbracket$, la variable aléatoire U_i suit une loi normale $\mathcal{N}(0, \sigma^2)$.

On pose : $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$ et $U = \begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix}$. Le modèle (1) de la partie II s'écrit alors matriciellement : $Y = M\theta + U$.

Soit $W_1, W_2, \dots, W_q$ ($q \in \mathbb{N}^*$), q variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, P)$. On définit le vecteur aléatoire $(W_1, W_2, \dots, W_q)$ à valeurs dans $\mathbb{R}^q$, en associant à tout ω de Ω le vecteur $(W_1(\omega), W_2(\omega), \dots, W_q(\omega))$ de $\mathbb{R}^q$.

On dit que le vecteur aléatoire $(W_1, W_2, \dots, W_q)$ est *normal* si pour tout q -uplet $(\rho_1, \rho_2, \dots, \rho_q)$ de nombres réels, différent de $(0, 0, \dots, 0)$, la variable aléatoire $\sum_{i=1}^q \rho_i W_i$ suit une loi normale de variance non nulle.

Dans le cas où le vecteur $(W_1, W_2, \dots, W_q)$ est normal, on admet que les variables aléatoires $W_1, W_2, \dots, W_q$ sont mutuellement indépendantes si et seulement si pour tout (i, j) de $\llbracket 1, q \rrbracket^2$ avec $i \neq j$, $\text{Cov}(W_i, W_j) = 0$.

7.a) Montrer que le vecteur aléatoire $(Y_1, Y_2, \dots, Y_n)$ est normal mais que le vecteur $(Y_1 - \bar{Y}_n, Y_2 - \bar{Y}_n, \dots, Y_n - \bar{Y}_n)$ ne l'est pas.

b) Déterminer la loi de chacune des variables aléatoires A_n et B_n . Le vecteur aléatoire (A_n, B_n) est-il normal ?

8. Soit S une matrice inversible de $\mathcal{M}_n(\mathbb{R})$. On note T la matrice-colonne des composantes du vecteur aléatoire $(T_1, T_2, \dots, T_n)$ telle que $T = SU$.

a) Montrer que le vecteur $(T_1, T_2, \dots, T_n)$ est normal.

b) On suppose que la matrice S est orthogonale. Montrer que $T_1, T_2, \dots, T_n$ sont mutuellement indépendantes.

9. Soit $\hat{U}_1, \hat{U}_2, \dots, \hat{U}_n$ les variables aléatoires qui ont été définies dans la question 6.

On note $\hat{U}$ la matrice-colonne de composantes $\hat{U}_1, \hat{U}_2, \dots, \hat{U}_n$ définie par $\hat{U} = Y - M \begin{pmatrix} A_n \\ B_n \end{pmatrix}$.

a) Montrer que $\hat{U} = GU$, où la matrice G a été définie dans la question 3.

b) Justifier l'existence d'une matrice orthogonale R de $\mathcal{M}_n(\mathbb{R})$ et d'une matrice diagonale D de $\mathcal{M}_n(\mathbb{R})$, telles que $G = RD^t R$. Quels sont les éléments diagonaux de D ?

c) Soit Z la matrice-colonne de composantes $Z_1, Z_2, \dots, Z_n$ définie par $Z = {}^t R U$. Quelle est la loi de $\sum_{i=1}^{n-2} Z_i^2$?

d) En déduire que la variable aléatoire $\sum_{i=1}^n \hat{U}_i^2$ suit la loi $\Gamma\left(2\sigma^2, \frac{n-2}{2}\right)$.

e) Soit p un réel donné vérifiant $0 < p < 1$. Établir l'existence d'un réel c_n ne dépendant pas des paramètres inconnus a, b et σ^2 , tel que $P\left(\left[\sum_{i=1}^n \hat{U}_i^2 \geq c_n \sigma^2\right]\right) = p$.

Dans les questions 10 et 11, on suppose qu'une $(n+1)$ -ième valeur de $\mathcal{X}$, notée x_{n+1} , est choisie mais que la valeur correspondante y_{n+1} de $\mathcal{Y}$ est inconnue. On suppose que y_{n+1} est la réalisation d'une variable aléatoire Y_{n+1} qui vérifie $Y_{n+1} = ax_{n+1} + b + U_{n+1}$, où les variables aléatoires $U_1, U_2, \dots, U_{n+1}$ sont mutuellement indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$.

10. On pose pour tout n -uplet $r = (r_1, r_2, \dots, r_n)$ de $\mathbb{R}^n$: $\hat{Y}_{n+1}^{(r)} = \sum_{i=1}^n r_i Y_i$.

L'ensemble $\{\hat{Y}_{n+1}^{(r)} ; r \in \mathbb{R}^n\}$ est l'ensemble des "prédicteurs linéaires" de Y_{n+1} .

a) Soit g la fonction définie sur $\mathbb{R}^n$ à valeurs réelles, telle que pour tout $r = (r_1, r_2, \dots, r_n)$ de $\mathbb{R}^n$,

$$g(r_1, r_2, \dots, r_n) = \sum_{i=1}^n r_i^2. \text{ On rappelle que pour tout } i \text{ de } \llbracket 1, n \rrbracket : \alpha_i = \frac{(x_i - \bar{x})}{ns_x^2}.$$

Montrer que la fonction g admet un minimum absolu sous les contraintes $\sum_{i=1}^n r_i = 1$ et $\sum_{i=1}^n x_i r_i = x_{n+1}$,

atteint en l'unique point $r^* = (r_1^*, r_2^*, \dots, r_n^*)$, où pour tout i de $\llbracket 1, n \rrbracket$, $r_i^* = \frac{1}{n} + (x_{n+1} - \bar{x})\alpha_i$.

b) Montrer que parmi les prédicteurs linéaires $\hat{Y}_{n+1}^{(r)}$ de Y_{n+1} , qui vérifient $E(\hat{Y}_{n+1}^{(r)}) = E(Y_{n+1})$ pour tout (a, b) de $\mathbb{R}^2$, $\hat{Y}_{n+1}^{(r^*)}$ est celui qui a la plus petite variance.

Vérifier que $\hat{Y}_{n+1}^{(r^*)} = A_n x_{n+1} + B_n$.

11.a) Déterminer la loi de la variable aléatoire $Y_{n+1} - (A_n x_{n+1} + B_n)$.

b) On note Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$. Soit p un réel donné vérifiant $\frac{1}{2} < p < 1$.

Justifier l'existence d'un réel d_n , que l'on exprimera à l'aide de Φ^{-1} , ne dépendant pas de a, b et σ^2 , tel que $P(|Y_{n+1} - (A_n x_{n+1} + B_n)| \leq d_n \sigma) = p$.

c) En déduire, à l'aide de la question 9.e), un intervalle dont les bornes ne dépendent que des $(Y_i)_{1 \leq i \leq n}$, des $(x_i)_{1 \leq i \leq n+1}$, de c_n et d_n , qui contienne Y_{n+1} avec une probabilité supérieure ou égale à $2p - 1$.

S'agit-il d'un intervalle de confiance au sens usuel du terme ?



Code épreuve : 283

BANQUE COMMUNE D'ÉPREUVES

CONCOURS D'ADMISSION DE 2013

Conceptions : H.E.C. – E.S.C.P. / EUROPE

OPTION SCIENTIFIQUE

MATHEMATIQUES II

Mardi 7 mai 2013, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite. Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre

Toutes les variables aléatoires qui interviennent dans ce problème sont définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$ et à valeurs réelles.

L'objectif du problème est d'introduire une famille de lois de probabilité discrètes, dites lois de Poisson mélangées, qui jouent un rôle important en mathématique de l'assurance, car elles permettent de modéliser le nombre d'apparitions d'événements aléatoires provenant de sources hétérogènes.

Partie I. Polynômes factoriels ascendants et lois binomiales négatives

Pour tout réel x et tout entier naturel n , on pose : $x^{<n>} = \begin{cases} \prod_{k=1}^n (x+k-1) & \text{si } n \geq 1 \\ 1 & \text{si } n = 0 \end{cases}$.

On associe aux fonctions polynomiales $x \mapsto x^{<n>}$, les polynômes $X^{<n>}$ ($n \in \mathbb{N}$) de $\mathbb{R}[X]$, dits polynômes factoriels ascendants.

1.a) Vérifier les égalités : $X^2 = X^{<2>} - X^{<1>}$ et $X^3 = X^{<3>} - 3X^{<2>} + X^{<1>}$.

b) Exprimer X^4 comme combinaison linéaire des polynômes $X^{<4>}$, $X^{<3>}$, $X^{<2>}$ et $X^{<1>}$.

c) Justifier plus généralement que les polynômes X^n ($n \in \mathbb{N}$) de la base canonique de $\mathbb{R}[X]$ sont tous des combinaisons linéaires de polynômes factoriels ascendants.

2. Soit (r, s) un couple de réels et n un entier naturel.

a) Exprimer $r^{<n+1>}$ en fonction de $r^{<n>}$.

b) En déduire pour tout $k \in \llbracket 0, n \rrbracket$, la relation : $(r+s+n)r^{<k>}s^{<n-k>} = r^{<k+1>}s^{<n-k>} + r^{<k>}s^{<n-k+1>}$.

c) Montrer par récurrence que pour tout $n \in \mathbb{N}$, on a : $(r+s)^{<n>} = \sum_{k=0}^n \binom{n}{k} r^{<k>}s^{<n-k>}$.

3. Dans cette question, r est un réel strictement positif et x est un réel fixé de $]0, 1[$. Pour tout $n \in \mathbb{N}$, on pose :

$$u_n = \frac{r^{<n+1>} x^{n+1}}{n!(1-x)^{r+1}} \quad \text{et} \quad R_n = \frac{r^{<n+1>}}{n!} \int_0^x \left(\frac{x-t}{1-t} \right)^n \frac{1}{(1-t)^{r+1}} dt.$$

a) Calculer $\lim_{n \rightarrow +\infty} \frac{u_{n+1}}{u_n}$.

b) En déduire l'existence d'un entier naturel N tel que pour tout $n \in \mathbb{N}$, on ait : $u_{N+n} \leq u_N \left(\frac{1+x}{2} \right)^n$.

c) Quelle est la limite de la suite $(u_n)_{n \in \mathbb{N}}$?

d) À l'aide d'une formule de Taylor que l'on citera avec ses hypothèses, justifier pour tout $n \in \mathbb{N}$, l'égalité :

$$(1-x)^{-r} = \sum_{k=0}^n \frac{r^{<k>}}{k!} x^k + R_n.$$

e) Montrer que pour tout $t \in [0, x]$, on a : $0 \leq \frac{x-t}{1-t} \leq x$. En déduire pour tout $n \in \mathbb{N}$, l'inégalité : $R_n \leq u_n$.

f) Déduire des résultats précédents que la série de terme général $\frac{r^{<n>}}{n!} x^n$ est convergente et que :

$$\sum_{n=0}^{+\infty} \frac{r^{<n>}}{n!} x^n = \frac{1}{(1-x)^r}.$$

4. Vérifier que pour tout couple (r, p) de réels tels que $r > 0$ et $0 < p < 1$, on a : $\sum_{k=0}^{+\infty} \frac{r^{<k>}}{k!} p^r (1-p)^k = 1$.

Dans toute la suite du problème, on dit qu'une variable aléatoire X définie sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{N}$ suit la loi binomiale négative de paramètres r ($r > 0$) et p ($0 < p < 1$), si pour tout entier naturel k , on a :

$$P([X = k]) = \frac{r^{<k>}}{k!} p^r (1-p)^k. \text{ Cette loi est notée } BN(r, p).$$

5. Soit X une variable aléatoire qui suit la loi $BN(r, p)$.

a) Soit n un entier supérieur ou égal à 1. Montrer que la variable aléatoire X_n définie par $X_n = \prod_{k=1}^n (X - k + 1)$

admet une espérance égale à $r^{<n>} \left(\frac{1-p}{p} \right)^n$.

b) En déduire que X admet des moments de tous ordres. Calculer l'espérance et la variance de X .

6. Soit X et Y deux variables aléatoires indépendantes suivant les lois $BN(r, p)$ et $BN(s, p)$, respectivement.

On pose : $Z = X + Y$. Montrer que Z suit la loi $BN(r + s, p)$.

Partie II. Inégalités stochastiques

Soit X et Y deux variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$. On dit que X est stochastiquement inférieure à Y lorsque pour tout réel x , on a : $P([X \geq x]) \leq P([Y \geq x])$.

7. Montrer que si les deux variables aléatoires X et Y vérifient pour tout $\omega \in \Omega$ l'inégalité $X(\omega) \leq Y(\omega)$, alors X est stochastiquement inférieure à Y .

8. On suppose que X suit la loi normale d'espérance égale à -1 et de variance égale à 1, que Y suit la loi normale d'espérance égale à 1 et de variance égale à 1 et que X et Y sont indépendantes.

a) Exprimer $P([X \geq 0] \cap [Y < 0])$ à l'aide de la fonction de répartition Φ de la loi normale centrée réduite.

b) Montrer que X est stochastiquement inférieure à Y .

c) A-t-on pour tout $\omega \in \Omega$, l'inégalité $X(\omega) \leq Y(\omega)$?

9. On suppose que X et Y sont discrètes à valeurs dans $\mathbb{N}$. Montrer que X est stochastiquement inférieure à Y si et seulement si, pour tout $k \in \mathbb{N}$, on a : $P([X \leq k]) \geq P([Y \leq k])$.

10. Soit θ et λ deux réels vérifiant $0 < \theta < \lambda$, et soit X et Z deux variables aléatoires indépendantes suivant respectivement la loi de Poisson de paramètre θ et la loi de Poisson de paramètre $\lambda - \theta$.

a) Rappeler la loi de $X + Z$ en citant précisément le résultat de cours utilisé.

b) Soit Y une variable aléatoire suivant la loi de Poisson de paramètre λ . Montrer que X est stochastiquement inférieure à Y .

11. Pour tout $k \in \mathbb{N}$ et pour tout réel $t > 0$, on pose : $F(t, k) = \sum_{j=0}^k \frac{t^j}{j!} e^{-t}$.

a) Écrire en Pascal une fonction d'en-tête `function suite (k : integer ; t : real) : real` ; qui permet de calculer $F(t, k)$.

b) Établir pour tout $k \in \mathbb{N}$ et pour tout réel $\beta \in]0, 1[$, l'existence d'un unique réel strictement positif $M(\beta, k)$ vérifiant l'égalité suivante : $F(M(\beta, k), k) = \beta$.

12. Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$ de fonction de répartition G .

On note V et W les deux applications de Ω dans $[0, 1]$, définies par : $\forall \omega \in \Omega, \begin{cases} V(\omega) = G(X(\omega) - 1) \\ W(\omega) = G(X(\omega)) \end{cases}$.

Soit α un réel vérifiant $0 < \alpha < 1$.

a) Justifier l'existence de $L_\alpha = \min\{k \in \mathbb{N}; G(k) \geq \alpha\}$ et comparer les réels α , $G(L_\alpha - 1)$ et $G(L_\alpha)$.

b) Montrer que $[W < \alpha]$ et $[V \geq \alpha]$ sont des événements. Qu'en déduit-on pour les applications V et W ?

c) Exprimer $P([W < \alpha])$ et $P([V \geq \alpha])$ à l'aide de G et L_α .

d) Soit U une variable aléatoire à densité qui suit la loi uniforme sur le segment $[0, 1]$.

Montrer que V est stochastiquement inférieure à U et que U est stochastiquement inférieure à W .

13. Pour n entier supérieur ou égal à 2, soit $(X_1, X_2, \dots, X_n)$ un n -échantillon i.i.d. (indépendant, identiquement distribué) de la loi de Poisson de paramètre inconnu $\theta > 0$. On pose pour tout $n \geq 2$: $S_n = \sum_{k=1}^n X_k$.

Les notations F et M sont celles de la question 11.

a) Proposer un estimateur sans biais de θ .

b) Soit α un réel tel que $0 < \alpha < 1$.

À l'aide de la question 12, établir les deux inégalités suivantes :

$$P\left(\left[F(n\theta, S_n) < \frac{\alpha}{2}\right]\right) \leq \frac{\alpha}{2} \quad \text{et} \quad P\left(\left[F(n\theta, S_n - 1) \geq 1 - \frac{\alpha}{2}\right]\right) \leq \frac{\alpha}{2}.$$

c) On pose : $J(X_1, X_2, \dots, X_n) = \frac{1}{n} M\left(\frac{\alpha}{2}, S_n\right)$ et $I(X_1, X_2, \dots, X_n) = \begin{cases} \frac{1}{n} M\left(1 - \frac{\alpha}{2}, S_n - 1\right) & \text{si } S_n \geq 1 \\ 0 & \text{si } S_n = 0 \end{cases}$.

Déduire des questions précédentes que $I(X_1, X_2, \dots, X_n)$ et $J(X_1, X_2, \dots, X_n)$ sont les bornes d'un intervalle de confiance de risque inférieur ou égal à α pour le paramètre inconnu θ .

Partie III. Lois de Poisson mélangées

Dans cette partie, T est une variable aléatoire à densité dont une densité f est nulle sur $\mathbb{R}_-$ et continue sur $\mathbb{R}_+^*$.

14. Justifier pour tout $n \in \mathbb{N}$, la convergence de l'intégrale $\int_0^{+\infty} t^n e^{-t} f(t) dt$.

15. On pose pour tout $n \in \mathbb{N}$: $z_n = \int_0^{+\infty} \frac{t^n}{n!} e^{-t} f(t) dt$ et $v_n = \sum_{k=0}^n z_k$.

a) Soit A un réel strictement positif et X_A une variable aléatoire suivant la loi de Poisson de paramètre A .

Établir pour tout $n \in \mathbb{N}$, l'encadrement suivant : $0 \leq 1 - v_n \leq P([X_A > n]) + \int_A^{+\infty} f(t) dt$.

b) Montrer que la série $\sum_{n \geq 0} z_n$ est convergente et que $\sum_{n=0}^{+\infty} z_n = 1$.

On dit qu'une variable aléatoire X à valeurs dans $\mathbb{N}$ suit la loi de Poisson mélangée associée à la densité f , notée $\mathcal{P}_f$, si pour tout entier naturel n , on a : $P([X = n]) = \int_0^{+\infty} \frac{t^n}{n!} e^{-t} f(t) dt$.

16. La notation $\exp$ désigne la fonction exponentielle. Soit r un réel strictement positif et p un réel vérifiant $0 < p < 1$. On suppose dans cette question qu'une densité f de T est donnée par :

$$f(t) = \begin{cases} \frac{1}{\Gamma(r)} \left(\frac{p}{1-p} \right)^r t^{r-1} \exp \left(-\frac{pt}{1-p} \right) & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}.$$

a) Reconnaître la loi de T . Donner (sans démonstration) l'espérance et la variance de T .

b) Montrer que $\mathcal{P}_f$ est la loi binomiale négative $BN(r, p)$.

17. Soit (r, p) et (s, q) deux couples de réels vérifiant $0 < r < s$ et $0 < q < p < 1$. On note Y, Z et W trois variables aléatoires qui suivent les lois $BN(r, p)$, $BN(s, p)$ et $BN(s, q)$, respectivement.

a) Montrer que Y est stochastiquement inférieure à Z .

b) À l'aide de la question 16, en déduire que Y est stochastiquement inférieure à W .

Conceptions : HEC Paris – ESCP Europe

MATHEMATIQUES II

OPTION SCIENTIFIQUE

Mercredi 7 mai 2014, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à **encadrer** dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : **l'utilisation de toute calculatrice et de tout matériel électronique est interdite**. Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre

Dans tout le problème, k désigne un entier supérieur ou égal à 2.

Notations algébriques

- Pour tout $n \in \mathbb{N}^*$, on note $\mathcal{M}_{n,1}(\mathbb{R})$ l'ensemble des matrices-colonnes à n lignes à coefficients réels et $\mathcal{M}_n(\mathbb{R})$ l'ensemble des matrices carrées à n lignes et n colonnes à coefficients réels. On identifie les ensembles $\mathcal{M}_1(\mathbb{R})$ et $\mathbb{R}$ en assimilant une matrice de $\mathcal{M}_1(\mathbb{R})$ à son unique coefficient.
- La base canonique de $\mathcal{M}_{k,1}(\mathbb{R})$ est notée $\mathcal{C}_k = (e_1, e_2, \dots, e_k)$ et l'espace vectoriel $\mathcal{M}_{k,1}(\mathbb{R})$ est muni de sa structure euclidienne usuelle pour laquelle la base $\mathcal{C}_k$ est orthonormale. On note $\langle u, v \rangle$ le produit scalaire de deux vecteurs u et v de $\mathcal{M}_{k,1}(\mathbb{R})$ et $\|u\| = \sqrt{\langle u, u \rangle}$ la norme du vecteur u .
- Pour toute matrice-colonne d de $\mathcal{M}_{n,1}(\mathbb{R})$ de composantes $d_1, d_2, \dots, d_n$, on note $\text{Diag}(d)$ la matrice diagonale de $\mathcal{M}_n(\mathbb{R})$ définie par :

$$\text{Diag}(d) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}.$$

- La transposée d'une matrice M est notée tM et I_k désigne la matrice identité de $\mathcal{M}_k(\mathbb{R})$.

Notations probabilistes

- Toutes les variables aléatoires et tous les vecteurs aléatoires qui interviennent dans ce problème sont définis sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$.
- On dit qu'un vecteur aléatoire discret $(Y_1, Y_2, \dots, Y_k)$, à valeurs dans $\mathbb{R}^k$, admet une espérance lorsque chacune de ses composantes en admet une.

On note Y la matrice-colonne de $\mathcal{M}_{k,1}(\mathbb{R})$ de composantes $Y_1, Y_2, \dots, Y_k$ et $\mathcal{E}(Y)$ la matrice-colonne de $\mathcal{M}_{k,1}(\mathbb{R})$ dont les composantes sont les espérances $E(Y_1), E(Y_2), \dots, E(Y_k)$.

Lorsque chacune des composantes Y_i ($i \in \llbracket 1, k \rrbracket$) admet une variance, on appelle matrice de variance-covariance de Y , notée $\mathcal{V}(Y)$, la matrice symétrique de $\mathcal{M}_k(\mathbb{R})$ dont les coefficients diagonaux sont les variances $V(Y_i)$ et les coefficients non diagonaux les covariances $\text{Cov}(Y_i, Y_j)$ pour tout $(i, j) \in \llbracket 1, k \rrbracket^2$ avec $i \neq j$.

En résumé, on pose sous réserve d'existence :

$$\mathcal{E}(Y) = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_k) \end{pmatrix} \quad \text{et} \quad \mathcal{V}(Y) = \begin{pmatrix} V(Y_1) & \text{Cov}(Y_1, Y_2) & \cdots & \text{Cov}(Y_1, Y_k) \\ \text{Cov}(Y_2, Y_1) & V(Y_2) & \cdots & \text{Cov}(Y_2, Y_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(Y_k, Y_1) & \text{Cov}(Y_k, Y_2) & \cdots & V(Y_k) \end{pmatrix}.$$

- Dans tout le problème, on note $p = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix}$ une matrice-colonne de $\mathcal{M}_{k,1}(\mathbb{R})$ vérifiant $\sum_{i=1}^k p_i = 1$ et pour

tout $i \in \llbracket 1, k \rrbracket$, $p_i \geq 0$.

L'objet du problème est l'étude des propriétés des matrices de variance-covariance en liaison avec la loi des vecteurs aléatoires correspondants.

Partie I. Loix généralisées de Bernoulli

Dans cette partie, on note u la matrice-colonne de $\mathcal{M}_{k,1}(\mathbb{R})$ dont tous les coefficients valent 1.

1. Soit $a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}$ une matrice-colonne non nulle de $\mathcal{M}_{k,1}(\mathbb{R})$ et $\alpha = \sum_{i=1}^k a_i$. On pose : $M = a^t u$.

- a) Calculer la matrice M et préciser son rang.
- b) Calculer la matrice Ma et en déduire une valeur propre de M .
- c) Montrer que $M^2 = \alpha M$. Que peut-on en déduire sur les valeurs propres de M ?
- d) Montrer que M est diagonalisable si et seulement si $\alpha \neq 0$.
- e) Pour quelles valeurs de α la matrice $I_k - M$ est-elle inversible ?
- f) On suppose que $\alpha = 1$. Montrer que M est la matrice dans la base canonique de $\mathbb{R}^k$ d'un projecteur dont on précisera l'image et le noyau. Dans quel cas ce projecteur est-il orthogonal ?

On dit qu'un vecteur aléatoire $(X_1, X_2, \dots, X_k)$ suit la loi généralisée de Bernoulli de paramètre p , notée $\mathcal{B}_k(p)$, si on a :

$$\forall i \in \llbracket 1, k \rrbracket, P([X = e_i]) = p_i, \text{ avec } X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix}.$$

2. Soit $(X_1, X_2, \dots, X_k)$ un vecteur aléatoire suivant la loi $\mathcal{B}_k(p)$.
 - a) Pour $i \in \llbracket 1, k \rrbracket$, comparer les événements $[X = e_i]$ et $[X_i = 1]$; en déduire que chaque variable aléatoire X_i suit une loi de Bernoulli de paramètre p_i et écrire la matrice $\mathcal{E}(X)$.
 - b) Quelle est la loi de la variable aléatoire $X_1 + X_2$?
 - c) Montrer que $\text{Cov}(X_1, X_2) = -p_1 p_2$.
 - d) Écrire la matrice $\mathcal{V}(X)$.
3. Soit $M(p)$ la matrice de $\mathcal{M}_k(\mathbb{R})$ définie par : $M(p) = p^t u$.
 - a) Vérifier l'égalité : $\mathcal{V}(X) = (I_k - M(p)) \text{Diag}(p)$.
 - b) Montrer que si $p_1, p_2, \dots, p_k$ sont différents de 0, le rang de $\mathcal{V}(X)$ est égal à $k - 1$.
 - c) Soit σ une permutation de $\llbracket 1, k \rrbracket$ et p_σ la matrice-colonne de $\mathcal{M}_{k,1}(\mathbb{R})$ de composantes $p_{\sigma(1)}, p_{\sigma(2)}, \dots, p_{\sigma(k)}$. Montrer que $\mathcal{V}(X)$ est semblable à $(I_k - p_\sigma^t u) \text{Diag}(p_\sigma)$.

- d) Exprimer le rang de $\mathcal{V}(X)$ en fonction du nombre d'éléments i de $\llbracket 1, k \rrbracket$ pour lesquels on a $p_i \neq 0$.

Partie II. Tirages avec remise dans une population stratifiée

Dans cette partie, on suppose que pour tout $i \in \llbracket 1, k \rrbracket$, on a $p_i > 0$ et que $p_1, p_2, \dots, p_k$ sont les proportions d'individus appartenant aux diverses catégories d'une population statistique scindée en k catégories distinctes. Pour modéliser une suite illimitée de tirages équiprobables avec remise effectués dans cette population, on utilise des variables aléatoires $X_i^{(n)}$ définies par :

$$\forall n \in \mathbb{N}^*, \forall i \in \llbracket 1, k \rrbracket, X_i^{(n)} = \begin{cases} 1 & \text{si l'individu extrait au } n\text{-ième tirage appartient à la } i\text{-ème catégorie} \\ 0 & \text{sinon} \end{cases}$$

On suppose que les vecteurs aléatoires $(X_1^{(n)}, X_2^{(n)}, \dots, X_k^{(n)})$ ($n \in \mathbb{N}^*$) suivent chacun la loi $\mathcal{B}_k(p)$ (partie I) et sont mutuellement indépendants. Cette indépendance mutuelle signifie que pour tout entier $n \geq 2$ et pour toutes fonctions $\varphi_1, \varphi_2, \dots, \varphi_n$ définies sur $\mathbb{R}^k$ à valeurs réelles, les variables aléatoires $\varphi_1(X_1^{(1)}, X_2^{(1)}, \dots, X_k^{(1)})$, $\varphi_2(X_1^{(2)}, X_2^{(2)}, \dots, X_k^{(2)})$, $\dots$, $\varphi_n(X_1^{(n)}, X_2^{(n)}, \dots, X_k^{(n)})$ sont indépendantes.

Pour tout $n \in \mathbb{N}^*$, on note $X^{(n)}$ la matrice-colonne de $\mathcal{M}_{k,1}(\mathbb{R})$ de composantes $X_1^{(n)}, X_2^{(n)}, \dots, X_k^{(n)}$ et $S^{(n)}$ la matrice-colonne de $\mathcal{M}_{k,1}(\mathbb{R})$ de composantes $S_1^{(n)}, S_2^{(n)}, \dots, S_k^{(n)}$, où pour tout $i \in \llbracket 1, k \rrbracket$, on a $S_i^{(n)} = \sum_{j=1}^n X_i^{(j)}$.

- 4.a) Préciser l'ensemble N_n des matrices-colonnes s de $\mathcal{M}_{k,1}(\mathbb{R})$ pour lesquelles on a $P([S^{(n)} = s]) > 0$.
- b) Déterminer les lois respectives des deux variables aléatoires $S_1^{(n)}$ et $S_1^{(n)} + S_2^{(n)}$. Sont-elles indépendantes ?
- c) Montrer que $\mathcal{V}(S^{(n)}) = n \mathcal{V}(X^{(1)})$.
5. Soit H un élément de $\mathcal{A}$ vérifiant $0 < P(H) < 1$, $\overline{H}$ l'événement contraire de H et W une variable aléatoire discrète admettant une variance.
 - a) Justifier l'existence de $E(W^2|H)$, espérance de W^2 pour la probabilité conditionnelle P_H .
 - b) On pose : $V(W|H) = E(W^2|H) - (E(W|H))^2$ (variance de W pour la probabilité conditionnelle P_H).
En utilisant le système complet d'événements $(H, \overline{H})$ et la formule de l'espérance totale pour W et W^2 , établir l'inégalité : $V(W) \geq P(H) V(W|H)$.
6. Pour tout $i \in \llbracket 1, k \rrbracket$, on note T_i le temps d'attente du premier tirage d'un individu de la i -ème catégorie et on note T la matrice-colonne de $\mathcal{M}_{k,1}(\mathbb{R})$ de composantes $T_1, T_2, \dots, T_k$.
 - a) Soit $i \in \llbracket 1, k \rrbracket$. Justifier que la probabilité que T_i soit infini est nulle. Quelle est la loi de T_i ?
 - b) On pose : $H_k = \bigcap_{i=1}^{k-1} [T_i = i]$. Calculer $P(H_k)$. Préciser la loi conditionnelle de $T_k - (k-1)$ sachant H_k .
En déduire $E(T_k|H_k)$ et $V(T_k|H_k)$.
 - c) En exploitant le résultat de la question 5.b), établir pour tout vecteur $v = (v_1, v_2, \dots, v_k)$ de $\mathbb{R}^k$, l'inégalité :

$$V\left(\sum_{i=1}^k v_i T_i\right) \geq \frac{v_k^2(1-p_k)}{p_k^2} \times \prod_{i=1}^{k-1} p_i.$$

- d) Montrer plus généralement que pour tout $j \in \llbracket 1, k \rrbracket$, on a : $V\left(\sum_{i=1}^k v_i T_i\right) \geq \frac{v_j^2(1-p_j)}{p_j^2} \times \prod_{\substack{i \in \llbracket 1, k \rrbracket \\ i \neq j}} p_i$.

Partie III. Support et rang stochastiques d'un vecteur aléatoire

Dans toute cette partie, $(Y_1, Y_2, \dots, Y_k)$ désigne un vecteur aléatoire discret, à valeurs dans $\mathbb{R}^k$, dont chaque composante admet une espérance et une variance. On rappelle que Y est la matrice-colonne de $\mathcal{M}_{k,1}(\mathbb{R})$ de composantes $Y_1, Y_2, \dots, Y_k$.

7. On appelle *support vectoriel* de Y , tout sous-espace vectoriel F de $\mathcal{M}_{k,1}(\mathbb{R})$ tel que $P([Y - \mathcal{E}(Y) \in F]) = 1$. On note $\mathcal{S}(Y)$ l'ensemble des supports vectoriels de Y .

a) Justifier l'existence d'un plus petit élément de l'ensemble des dimensions des éléments de $\mathcal{S}(Y)$.

Ce plus petit élément est appelé le *rang stochastique* de Y et noté $R_s(Y)$.

b) Dans quels cas le rang stochastique $R_s(Y)$ est-il nul ?

c) Montrer que l'intersection de deux supports vectoriels F_1 et F_2 de Y est un support vectoriel de Y .

d) En déduire l'existence d'un unique élément F de $\mathcal{S}(Y)$ tel que la dimension de F soit égale à $R_s(Y)$.

L'espace vectoriel F est appelé le *support stochastique* de Y .

8. Soit u une matrice-colonne de $\mathcal{M}_{k,1}(\mathbb{R})$ de composantes $u_1, u_2, \dots, u_k$.

a) Montrer que la variable aléatoire $\sum_{i=1}^k u_i Y_i$ admet une variance, égale à ${}^t u \mathcal{V}(Y) u$.

b) Établir l'existence d'un unique vecteur $(\lambda_1, \lambda_2, \dots, \lambda_k)$ de $\mathbb{R}^k$ tel que $\mathcal{V}(Y)$ soit semblable à la matrice $\text{Diag}(\lambda)$ et pour lequel $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$ (on note $\text{Diag}(\lambda)$ la matrice diagonale de $\mathcal{M}_k(\mathbb{R})$ de coefficients diagonaux $\lambda_1, \lambda_2, \dots, \lambda_k$).

c) On pose : $\|Y - \mathcal{E}(Y)\|^2 = \sum_{i=1}^k (Y_i - E(Y_i))^2$. Montrer que $E(\|Y - \mathcal{E}(Y)\|^2) = \sum_{i=1}^k \lambda_i$.

9. Soit $q \in \llbracket 1, k \rrbracket$, F un sous-espace vectoriel de $\mathcal{M}_{k,1}(\mathbb{R})$ de dimension q et $(f^{(1)}, f^{(2)}, \dots, f^{(q)})$ une base orthonormale de F .

a) Soit $\omega \in \Omega$. Justifier l'existence de $Q_F(\omega) = \inf\{\|Y(\omega) - \mathcal{E}(Y) - x\|^2; x \in F\}$ et montrer que :

$$\|Y(\omega) - \mathcal{E}(Y)\|^2 = Q_F(\omega) + \sum_{j=1}^q \langle Y(\omega) - \mathcal{E}(Y), f^{(j)} \rangle^2.$$

b) À l'aide de la question 8, établir l'égalité : $E(Q_F) = \sum_{i=1}^k \lambda_i - \sum_{j=1}^q {}^t f^{(j)} \mathcal{V}(Y) f^{(j)}$.

c) Que devient l'égalité précédente lorsque $F = \mathcal{M}_{k,1}(\mathbb{R})$?

10.a) Montrer que pour toute matrice-colonne f de $\mathcal{M}_{k,1}(\mathbb{R})$ vérifiant $\|f\| = 1$, on a : ${}^t f \mathcal{V}(Y) f \leq \lambda_1$.

b) En déduire la borne inférieure de $E(Q_F)$ lorsque F décrit l'ensemble des droites vectorielles de $\mathcal{M}_{k,1}(\mathbb{R})$.

c) Dans cette question, on suppose que $(Y_1, Y_2, \dots, Y_k)$ suit la loi $\mathcal{B}_k(p)$, où pour tout $i \in \llbracket 1, k \rrbracket$, on a $p_i = \frac{1}{k}$.

Calculer les valeurs propres de $\mathcal{V}(Y)$ et la borne inférieure de $E(Q_F)$ pour l'ensemble des droites vectorielles F de $\mathcal{M}_{k,1}(\mathbb{R})$, puis préciser pour quelle(s) droite(s) cette borne est atteinte.

11. On suppose que le rang r de $\mathcal{V}(Y)$ est non nul. On note F_0 la somme des sous-espaces propres associés aux valeurs propres non nulles de $\mathcal{V}(Y)$ et F un sous-espace vectoriel de $\mathcal{M}_{k,1}(\mathbb{R})$ tel que $F \subset F_0$ et $F \neq F_0$.

a) Calculer $E(Q_{F_0})$ et en déduire que F_0 est un support vectoriel de Y .

b) Justifier l'existence d'un vecteur $f^{(r)}$ de F_0 , orthogonal à F et de norme 1.

c) Montrer que ${}^t f^{(r)} \mathcal{V}(Y) f^{(r)} > 0$ et en déduire que $E(Q_F) \neq 0$.

d) Montrer que le rang stochastique $R_s(Y)$ de Y est égal à r .

12. Dans cette question, on reprend les définitions et notations de la question 6.

a) À l'aide de la question 6.d), montrer que le rang stochastique $R_s(T)$ de T est égal à k .

b) Montrer que pour tout $i \in \mathbb{N}^*$, on a : $E(T_1 T_2 | [T_1 = i] \cap [T_2 > i]) = i \left(i + \frac{1}{p_2} \right)$.

c) Établir la relation : $E(T_1 T_2) = \frac{1}{p_1 p_2} - \frac{1}{p_1 + p_2}$.

d) On note $\Pi = (\pi_{i,j})_{1 \leq i,j \leq k}$ la matrice de $\mathcal{M}_k(\mathbb{R})$ définie par : $\pi_{i,j} = \begin{cases} \frac{1-p_i}{p_i^2} & \text{si } i = j \\ -\frac{1}{p_i + p_j} & \text{si } i \neq j \end{cases}$.

Montrer que la matrice Π est inversible.

Conception : HEC Paris – ESCP Europe

MATHÉMATIQUES II

OPTION SCIENTIFIQUE

Mercredi 6 mai 2015, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
Les candidats sont invités à **encadrer** dans la mesure du possible les résultats de leurs calculs.
Ils ne doivent faire usage d'aucun document : **l'utilisation de toute calculatrice et de tout matériel électronique est interdite**. Seule l'utilisation d'une règle graduée est autorisée.
Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre

Dans tout le problème :

- On note $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X une variable aléatoire définie sur $(\Omega, \mathcal{A})$, à valeurs dans $\mathbb{R}_+$.
- Toutes les variables aléatoires intervenant dans le problème sont définies sur le même espace $(\Omega, \mathcal{A})$ qui est, sauf mention contraire, muni de la probabilité P .
- On note S_X la fonction définie sur $\mathbb{R}$ à valeurs réelles, telle que : $\forall x \in \mathbb{R}, S_X(x) = P([X > x])$.

Dans le cadre de l'évaluation des risques encourus par des établissements financiers, il est nécessaire de retrancher à la valeur moyenne attendue des investissements (espérance mathématique "pure") un terme correctif d'autant plus important que le risque est plus grand.

L'objet du problème est de déterminer grâce à une "fonction de distorsion", une "espérance corrigée" qui prend en compte cette notion de risque et qui possède les propriétés requises pour une évaluation cohérente de risques financiers, en particulier, une propriété de sous-additivité nécessaire pour valoriser équitablement les avantages éventuels de la diversification.

Partie I. Probabilité de surpassement et espérance.

- On suppose uniquement dans cette question que X suit la loi exponentielle $\mathcal{E}(\lambda)$ (avec $\lambda > 0$).
 - Vérifier l'égalité : $E(X) = \int_0^{+\infty} S_X(x) dx$.
 - Donner l'allure de la courbe représentative de la fonction de répartition F de X et interpréter $E(X)$ en terme d'aire grâce à la formule précédente.
- Soit h la fonction définie sur $\mathbb{R}_+$ par : $h(x) = \frac{1}{(x+1)(x+2)}$.
 - Justifier la convergence de l'intégrale $\int_0^{+\infty} h(x) dx$.
 - Déterminer deux réels c et d vérifiant pour tout réel $x \geq 0$, la relation : $h(x) = \frac{c}{x+1} + \frac{d}{x+2}$.
En déduire une primitive de h sur $\mathbb{R}_+$.
 - Montrer que la fonction $f_0 : x \mapsto \begin{cases} \frac{1}{(x+1)(x+2) \ln 2} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$ est une densité de probabilité.

3. On suppose dans cette question que X admet pour densité la fonction f_0 définie à la question 2.c).

a) La variable aléatoire X admet-elle une espérance ?

b) Pour tout x réel, calculer $S_X(x)$ et en trouver un équivalent lorsque x tend vers $+\infty$.

c) Étudier la convergence de l'intégrale $\int_0^{+\infty} S_X(x) dx$.

4.a) Justifier la monotonie de la fonction S_X et trouver la limite de $S_X(x)$ lorsque x tend vers $+\infty$.

b) Montrer que la fonction S_X est continue à droite. À quelle condition est-elle continue en 0 ?

5. Dans cette question, on suppose que X admet une densité f nulle sur $]-\infty, 0]$, continue sur $]0, +\infty[$ mais non nécessairement en 0.

a) Montrer que la fonction S_X est continue sur $\mathbb{R}$ et de classe C^1 sur $]0, +\infty[$.

b) Justifier la convergence de l'intégrale $\int_0^1 x f(x) dx$.

c) Établir pour tout réel $A \geq 0$, l'égalité : $\int_0^A S_X(x) dx = AS_X(A) + \int_0^A x f(x) dx$.

d) En déduire que si l'intégrale $\int_0^{+\infty} S_X(x) dx$ est convergente, alors X admet une espérance.

e) Montrer que si X admet une espérance, alors on a pour tout réel $A \geq 0$: $\int_A^{+\infty} x f(x) dx \geq AS_X(A)$.

f) Déduire des résultats précédents que X admet une espérance si et seulement si l'intégrale $\int_0^{+\infty} S_X(x) dx$ est convergente, et que dans ce cas, on a : $E(X) = \int_0^{+\infty} S_X(x) dx$ [1].

6. Dans cette question, on suppose que X est discrète et à valeurs dans $\mathbb{N}$.

a) Établir pour tout entier naturel n , l'égalité : $\sum_{k=0}^n S_X(k) = (n+1)P([X \geq n+1]) + \sum_{k=0}^n k P([X = k])$.

b) En déduire que si la série de terme général $S_X(n)$ est convergente, alors X admet une espérance.

c) Montrer que X admet une espérance si et seulement si la série de terme général $S_X(n)$ est convergente, et que dans ce cas, on a : $E(X) = \sum_{n=0}^{+\infty} S_X(n)$.

d) On suppose que X admet une espérance.

(i) Exprimer pour tout $N \in \mathbb{N}$, l'intégrale $\int_0^N S_X(x) dx$ à l'aide d'une somme partielle de la série de terme général $S_X(n)$.

(ii) En déduire que $E(X) = \lim_{A \rightarrow +\infty} \int_0^A S_X(x) dx$.

Ainsi, la relation [1] reste applicable dans le cas des variables aléatoires à valeurs dans $\mathbb{N}$; on admet qu'elle reste applicable à toute variable aléatoire pour laquelle l'intégrale $\int_0^{+\infty} S_X(x) dx$ est convergente.

Partie II. Fonctions de distorsion et espérances corrigées : un exemple.

On appelle *fonction de distorsion* toute fonction g définie, continue et croissante sur l'intervalle $[0, 1]$ qui vérifie les trois propriétés supplémentaires suivantes : $g(0) = 0$, $g(1) = 1$ et g est concave sur $]0, 1[$.

Pour toute fonction de distorsion g , on dit que X admet une *espérance corrigée par g* , si la fonction composée $g \circ S_X$ admet une intégrale convergente sur $[0, +\infty[$.

Cette intégrale, notée $E_g(X)$, est appelée espérance de X corrigée par g . Ainsi : $E_g(X) = \int_0^{+\infty} g(S_X(x)) dx$.

7. Exemple . Soit Φ la fonction de répartition de la loi normale centrée réduite dont la dérivée, notée φ , est telle que : $\forall t \in \mathbb{R}, \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$.

a) Justifier que Φ est une bijection de $\mathbb{R}$ sur $]0, 1[$. On note Ψ la bijection réciproque de Φ .

Soit $\alpha \in]0, \frac{1}{2}[$ et w_α la fonction définie sur $[0, 1]$ telle que : $w_\alpha(x) = \begin{cases} 0 & \text{si } x = 0 \\ \Phi(\Psi(x) - \Psi(\alpha)) & \text{si } 0 < x < 1 \\ 1 & \text{si } x = 1 \end{cases}$.

b) Montrer que la fonction w_α est continue sur $[0, 1]$ et dérivable sur $]0, 1[$.

c) On note w'_α la dérivée de w_α . Établir pour tout $x \in]0, 1[$, la relation : $w'_\alpha(x) = \exp\left(\Psi(\alpha)\Psi(x) - \frac{(\Psi(\alpha))^2}{2}\right)$.

d) Déterminer l'équation de la tangente à la courbe représentative de la fonction w_α au point d'abscisse α .

e) Vérifier que w_α est une fonction de distorsion.

8. On considère à nouveau la fonction de distorsion w_α définie dans la question 7.

Soit Y une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$ qui suit une loi normale d'écart-type égal à 1, dont l'espérance est notée μ , et soit X une variable aléatoire qui suit la même loi que $\exp(Y)$.

a) Justifier l'existence de l'espérance de X et la calculer.

b) Montrer que pour tout réel $x > 0$, on a : $w_\alpha(S_X(x)) = P([X e^{-\Psi(\alpha)} > x])$.

c) En déduire l'existence et la valeur de $E_{w_\alpha}(X)$.

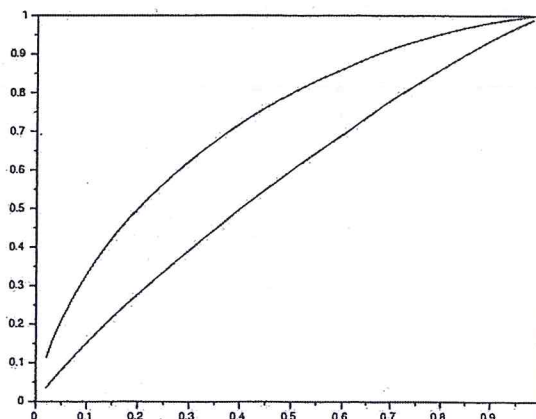
9. Pour faire tracer par *Scilab* la courbe représentative de w_α , on utilise la fonction *cdfnor* qui permet de calculer les valeurs de la fonction de répartition de variables aléatoires de loi normale.

Si une variable aléatoire Z suit la loi normale centrée réduite et si x et p sont deux réels reliés par l'égalité $P([Z \leq x]) = p$, alors :

- p est calculable en *Scilab* par *cdfnor*("PQ", x, 0, 1) ;
- x est calculable en *Scilab* par *cdfnor*("X", 0, 1, p, 1-p).

Le graphique ci-dessous a été obtenu en affectant successivement à la variable α les valeurs 0.2 et 0.4, et en exécutant les cinq instructions codées comme suit, la quatrième étant incomplète.

```
(1) qa=cdfnor("X", 0, 1, alpha, 1-alpha)
(2) p=[0.02 : 0.01 : 0.98]
(3) q=cdfnor("X", zeros(p), ones(p), p, 1-p)-qa*ones(p)
(4) wa=cdfnor("PQ", ?, ?, ?)
(5) plot(p, wa)
```



a) Quelles sont les valeurs affectées aux variables p et q par les instructions (2) et (3) (on en précisera le format matriciel) ?

b) Compléter la quatrième ligne de code.

c) À laquelle des deux courbes correspond la valeur $\alpha = 0.2$ (on justifiera mathématiquement la réponse) ?

d) Comment trouver les tangentes aux deux courbes en $(0,0)$ et $(1,1)$?

e) Que deviendrait la courbe représentative de w_α si on faisait tendre α vers 0 ?

Partie III. Sous-additivité des espérances corrigées.

Les notations et le contexte de cette partie sont identiques à ceux des parties I et II.

Dans cette partie, on note g une fonction de distorsion arbitraire et on suppose l'existence de $E_g(X)$.

Soit B un réel positif et soit Y une variable aléatoire à valeurs dans $\mathbb{R}_+$ telle que $P([Y \in [0, B]]) = 1$.

L'objectif de cette partie consiste à établir l'inégalité : $E_g(X + Y) \leq E_g(X) + E_g(Y)$ [2].

10. a) Soit x un réel fixé de $[0, 1]$. Justifier pour tout entier $n \geq 1$, l'inégalité :

$$g\left(x\left(1 - \frac{1}{n}\right) + (1-x)\frac{1}{n}\right) \geq xg\left(1 - \frac{1}{n}\right) + (1-x)g\left(\frac{1}{n}\right).$$

b) En déduire que $g(x) \geq x$.

c) Soit a, b et ε trois réels tels que : $0 < a < b < b + \varepsilon < 1$.

Justifier l'existence d'un réel $\lambda \in [0, 1]$ vérifiant les deux égalités :
$$\begin{cases} \lambda a + (1-\lambda)(b + \varepsilon) = b \\ (1-\lambda)a + \lambda(b + \varepsilon) = a + \varepsilon \end{cases}$$

En déduire l'inégalité : $g(b + \varepsilon) - g(a + \varepsilon) \leq g(b) - g(a)$.

11. a) Montrer que $E(X)$ existe et que $E(X) \leq E_g(X)$.

b) Montrer que si X est une variable aléatoire certaine, on a $E_g(X) = E(X)$.

c) Soit $r > 0$ et $s \geq 0$. Montrer que $E_g(rX + s)$ existe et que : $E_g(rX + s) = rE_g(X) + s$.

d) Soit T et W deux variables aléatoires telles que $P([0 \leq T \leq W]) = 1$.

Sous réserve d'existence, comparer $E_g(T)$ et $E_g(W)$.

12. Justifier l'existence de $E_g(Y)$ et de $E_g(X + Y)$; établir les inégalités : $E_g(Y) \leq B$ et $E_g(X + Y) \leq E_g(X) + B$.

13. On se propose de montrer par récurrence sur l'entier n que l'inégalité [2] est vraie pour toute variable aléatoire U telle que $P([U \in [0, n]]) = 1$.

Soit n un entier naturel donné. On suppose que quelle que soit la probabilité P sur $(\Omega, \mathcal{A})$, l'inégalité [2] est vraie pour toute fonction de distorsion g , pour toute variable aléatoire X possédant une espérance corrigée par g et pour toute variable aléatoire U telle que $P([U \in [0, n]]) = 1$.

a) Déduire de la question 12 que la propriété ci-dessus est vérifiée pour $n = 0$.

b) On suppose la propriété ci-dessus vérifiée pour un entier naturel n donné.

Soit Z une variable aléatoire telle que $P([Z \in [0, n + 1]]) = 1$ et $p = P([Z > 0]) > 0$.

On pose $P^* = P_{[Z > 0]}$ (probabilité conditionnelle sachant $[Z > 0]$). Pour tout réel $x > 0$, on pose :

$$S_X^*(x) = P^*([X > x]), \quad S_Z^*(x) = P^*([Z > x]) \quad \text{et} \quad S_{X+Z}^*(x) = P^*([X + Z > x]).$$

(i) Établir l'égalité : $S_{X+Z}(x) = (1-p)P_{[Z=0]}([X > x]) + pS_{X+Z}^*(x)$.

(ii) Exprimer $S_X(x)$ et $S_Z(x)$ en fonction de p , $P_{[Z=0]}([X > x])$, $S_X^*(x)$ et $S_Z^*(x)$.

(iii) En utilisant le résultat de la question 10.c), déduire des relations précédentes l'inégalité :

$$g(S_{X+Z}(x)) - g(S_X(x)) - g(S_Z(x)) \leq g(pS_{X+Z}^*(x)) - g(pS_X^*(x)) - g(pS_Z^*(x)).$$

c) Justifier que la fonction $h : x \mapsto \frac{g(px)}{g(p)}$ est une fonction de distorsion et établir l'inégalité :

$$\int_0^{+\infty} h(S_{X+Z}^*(x)) dx \leq \int_0^{+\infty} h(S_X^*(x)) dx + \int_0^{+\infty} h(S_Z^*(x)) dx.$$

d) En déduire l'inégalité : $E_g(X + Z) \leq E_g(X) + E_g(Z)$. Conclure.

14. Pour tout entier $n \geq 1$ et tout $\omega \in \Omega$, on pose : $Y_n(\omega) = \frac{1}{n} [nY(\omega)]$, où $[u]$ désigne la partie entière de u .

a) Justifier l'existence de $E_g(Y_n)$ et $E_g(X + Y_n)$; établir l'inégalité : $E_g(X + Y_n) \leq E_g(X) + E_g(Y_n)$.

b) Pour $x > 0$, comparer les événements $[Y_n > x]$ et $[Y > x]$, et montrer que $E_g(Y_n) \leq E_g(Y)$.

c) Montrer que pour tout entier $n \geq 1$, on a :

$$\int_{\frac{1}{n}}^{+\infty} g(S_{X+Y}(x)) dx = \int_0^{+\infty} g\left(S_{X+Y}\left(x + \frac{1}{n}\right)\right) dx \leq E_g(X + Y_n).$$

d) En déduire l'inégalité [2].

Conception : HEC Paris – ESCP Europe

OPTION SCIENTIFIQUE

MATHÉMATIQUES II

Mercredi 4 mai 2016, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à **encadrer** dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document. **L'utilisation de toute calculatrice et de tout matériel électronique est interdite.** Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

La simulation de vecteurs aléatoires dont les composantes ne sont pas indépendantes intervient dans l'évaluation de risques cumulés dans des domaines tels que l'assurance, la finance, la médecine ou l'écologie.

On résume les liaisons entre les composantes à l'aide de fonctions de plusieurs variables appelées copules.

L'objet du problème consiste à présenter cette notion de copule dans le cadre de la simulation d'un vecteur aléatoire à deux composantes.

On suppose que toutes les variables aléatoires et tous les vecteurs aléatoires qui interviennent dans ce problème sont définis sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$.

On rappelle que la loi d'un vecteur aléatoire (X, Y) à valeurs dans $\mathbf{R}^2$ est caractérisée par la fonction $F_{(X,Y)}$ définie sur $\mathbf{R}^2$ par : $\forall (x, y) \in \mathbf{R}^2, F_{(X,Y)}(x, y) = P([X \leq x] \cap [Y \leq y])$. On dit que $F_{(X,Y)}$ est la *fonction de répartition conjointe* de X et Y .

Partie I. Simulation d'une variable aléatoire à densité.

1.a) Justifier la convergence de l'intégrale $\int_0^{+\infty} \frac{dx}{\sqrt{x}(1+x)}$.

b) Soit V une variable aléatoire telle que $V(\Omega) = [0, \pi/2[$ suivant la loi uniforme sur $[0, \pi/2[$.

On pose : $X = \tan^2(V)$. Montrer que X est une variable aléatoire à densité.

c) En déduire que la fonction $f : x \mapsto \begin{cases} \frac{1}{\pi \sqrt{x}(1+x)} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$ est une densité de probabilité.

2.a) Compléter le code *Scilab* de la fonction `simulX` suivante de sorte que son application à l'entier N ($N \geq 2$) fournisse une matrice colonne contenant N simulations indépendantes de la variable aléatoire X .

```
function x=simulX(N)
    u=rand(...,...);
    x=ones(u); // matrice de même format que u.
    for i=1:N
        x(i,1)= .....
    end;
endfunction
```

Tournez la page S.V.P.

- b) Après avoir affecté une valeur entière supérieure ou égale à 2 à la variable N , on exécute les commandes suivantes :

```
x=simulX(N);
y=0;
for i=1:N if x(i,1)>1 then y=y+1; end; end;
q=y/N;
```

Trouver la loi d'une variable aléatoire dont la valeur de y est, en fin de boucle, une simulation.

De quel nombre peut-on s'attendre que q soit proche lorsque la valeur affectée à N est grande et pourquoi ?

3. Soit X une variable aléatoire à densité dont la fonction de répartition est notée F_X .

Soit p un réel de $]0, 1[$. On pose : $K_p = \{x \in \mathbf{R}; F_X(x) = p\}$ et $J_p = \{x \in \mathbf{R}; F_X(x) < p\}$.

- a) Justifier que les deux ensembles J_p et K_p ne sont pas vides et montrer que si $a \in J_p$ et $b \in K_p$, on a nécessairement : $a < b$.

- b) On pose : $G_X(p) = \inf(K_p)$. Justifier l'existence de $G_X(p)$ et établir l'égalité : $F_X(G_X(p)) = p$.

- c) En déduire pour tout x réel, l'équivalence : $x < G_X(p) \iff F_X(x) < p$.

- d) Soit U une variable aléatoire telle que $U(\Omega) =]0, 1[$ et qui suit la loi uniforme sur $]0, 1[$.

Montrer que la variable aléatoire $G_X(U)$ suit la même loi que X .

- e) On suppose que X admet pour densité la fonction f définie dans la question 1.c) et que p n'est plus fixé. Déterminer la fonction $G_X : p \mapsto G_X(p)$ définie sur $]0, 1[$.

Partie II. Fonction de répartition conjointe de deux variables aléatoires de lois uniformes.

4. Soit (X, Y) un couple de variables aléatoires et F la fonction de répartition conjointe de X et Y .

Soit $x \in \mathbf{R}$.

- a) Montrer que la fonction $y \mapsto F(x, y)$ est croissante sur $\mathbf{R}$.

- b) Établir l'égalité : $[X \leq x] = \bigcup_{n \in \mathbf{N}^*} ([X \leq x] \cap [Y \leq n])$.

- c) Montrer que $F(x, y)$ tend vers $P([X \leq x])$ lorsque y tend vers $+\infty$.

- d) Quelle est la limite de $F(x, y)$ lorsque y tend vers $-\infty$?

- e) Soit a, a', b, b' des réels vérifiant : $a \leq a'$ et $b \leq b'$. On pose : $A = [a < X \leq a']$ et $B = [b < Y \leq b']$.

(i) Exprimer la probabilité $P(A \cap B)$ en fonction de $P([X \leq a] \cap B)$ et $P([X \leq a'] \cap B)$.

(ii) Établir l'égalité : $P(A \cap B) = F(a', b') - F(a', b) - F(a, b') + F(a, b)$.

Dans les questions 5 et 6, on note U et V deux variables aléatoires suivant chacune la loi uniforme sur $[0, 1]$ et $F_{(U,V)}$ leur fonction de répartition conjointe.

On note C la restriction de $F_{(U,V)}$ à $[0, 1]^2$: $\forall (u, v) \in [0, 1]^2, C(u, v) = P([U \leq u] \cap [V \leq v])$.

Pour tout couple $(u, v) \in [0, 1]^2$, on pose : $C_+(u, v) = \min\{u, v\}$ et $C_-(u, v) = \max\{u + v - 1, 0\}$.

Pour $(u, v) \in [0, 1]^2$, on note $\overline{[U > u] \cup [V > v]}$ l'événement contraire de l'événement $[U > u] \cup [V > v]$.

On rappelle que si deux vecteurs aléatoires (X_1, Y_1) et (X_2, Y_2) ont même loi et si g est une fonction continue sur $\mathbf{R}^2$ à valeurs dans $\mathbf{R}$, alors les variables aléatoires $g(X_1, Y_1)$ et $g(X_2, Y_2)$ ont même loi.

- 5.a) Comparer pour tout $(u, v) \in [0, 1]^2$, les trois événements : $\overline{[U > u] \cup [V > v]}$, $[U \leq u] \cap [V \leq v]$ et $[U \leq u]$.

- b) Justifier pour tout $(u, v) \in [0, 1]^2$, la double inégalité : $u + v - 1 \leq C(u, v) \leq u$.

- c) En déduire l'encadrement suivant : $\forall (u, v) \in [0, 1]^2, C_-(u, v) \leq C(u, v) \leq C_+(u, v)$.

- 6.a) Calculer $F_{(U,U)}(x, y)$ selon les valeurs du couple (x, y) de $\mathbf{R}^2$.
- b) Représenter dans le plan rapporté à un repère orthonormé une ligne de niveau pour la fonction de deux variables $F_{(U,U)}$, correspondant à une valeur de la fonction strictement comprise entre 0 et 1.
Hachurer sur la même figure, l'ensemble des couples $(x, y) \in \mathbf{R}^2$ pour lesquels $F_{(U,U)}(x, y) = x$.
- c) Montrer que C est égale à C_+ si et seulement si les variables aléatoires U et V sont égales presque sûrement.
- d) Calculer la fonction de répartition conjointe $F_{(U,1-U)}$ et donner une condition nécessaire et suffisante portant sur U et V pour que C soit égale à C_- .

Partie III. Copules.

On appelle *copule* toute fonction Φ définie sur $[0, 1]^2$, à valeurs réelles, vérifiant les trois propriétés suivantes :

- $\forall u \in [0, 1], \Phi(u, 0) = \Phi(0, u) = 0$;
- $\forall u \in [0, 1], \Phi(u, 1) = \Phi(1, u) = u$;
- $\forall (u, u', v, v') \in [0, 1]^4, u \leq u' \text{ et } v \leq v' \implies \Phi(u', v') - \Phi(u', v) - \Phi(u, v') + \Phi(u, v) \geq 0$.

On appelle *copule à densité* toute copule Φ dont la restriction à l'ouvert $]0, 1[^2$ est de classe $\mathcal{C}^2$ sur $]0, 1[^2$.

7. *Exemples.* On reprend le contexte et les notations du préambule des questions 5 et 6.

- a) Vérifier que C est une copule. Dans la suite (Partie IV), on l'appelle la *copule associée au couple* (U, V) .
- b) En déduire que C_+ , C_- ainsi que la fonction Π définie sur $[0, 1]^2$ par $\Pi(u, v) = uv$ sont des copules.
8. Soit Φ une copule à densité et $(a, b) \in]0, 1[^2$.

Pour tout couple (h, k) de réels non nuls tels que $(a + h, b + k) \in]0, 1[^2$, on pose :

$$G(h, k) = \frac{1}{hk} (\Phi(a + h, b + k) - \Phi(a + h, b) - \Phi(a, b + k) + \Phi(a, b)).$$

- a) Soit h un réel non nul tel que $a + h \in]0, 1[$.

Justifier que $G(h, k)$ admet une limite $H(h)$ lorsque k tend vers 0 et exprimer $H(h)$ à l'aide de la dérivée partielle $\partial_2(\Phi)$ de Φ par rapport à sa seconde variable.

- b) On note $\partial_{1,2}^2(\Phi)$ la dérivée partielle seconde croisée de Φ sur $]0, 1[^2$ et on rappelle que $\partial_{1,2}^2(\Phi) = \partial_{2,1}^2(\Phi)$.

Trouver la limite de $H(h)$ lorsque h tend vers 0 et en déduire que $\partial_{1,2}^2(\Phi)(a, b) \geq 0$.

9. Soit φ une fonction définie sur $[0, 1]^2$, à valeurs réelles, continue sur $[0, 1]^2$ et de classe $\mathcal{C}^2$ sur $]0, 1[^2$.

Pour tout $(u, u', v, v') \in [0, 1]^4$, on pose : $\Psi(u, u', v, v') = \varphi(u', v') - \varphi(u', v) - \varphi(u, v') + \varphi(u, v)$.

- a) Pour tout $(u, u', v, v') \in]0, 1[^4$, justifier l'égalité : $\Psi(u, u', v, v') = \int_v^{v'} \left(\int_u^{u'} \partial_{1,2}^2(\varphi)(x, y) dx \right) dy$.

- b) Soit u et u' des réels tels que : $0 \leq u \leq u' \leq 1$. Pour tout $n \in \mathbf{N}^*$, on pose :
$$\begin{cases} u_n = \frac{1}{3n} + \left(1 - \frac{1}{n}\right)u \\ u'_n = \frac{2}{3n} + \left(1 - \frac{1}{n}\right)u' \end{cases}$$

Vérifier pour tout $n \in \mathbf{N}^*$, les inégalités strictes suivantes : $0 < u_n < u'_n < 1$.

- c) On pose :
$$\begin{cases} T_0 = \{(u, u', v, v') \in \mathbf{R}^4 ; 0 < u < u' < 1, 0 < v < v' < 1\} \\ T = \{(u, u', v, v') \in \mathbf{R}^4 ; 0 \leq u \leq u' \leq 1, 0 \leq v \leq v' \leq 1\} \end{cases}$$

On suppose que la fonction Ψ est positive ou nulle sur T_0 . Montrer que Ψ est positive ou nulle sur T .

d) En déduire que, pour que la fonction φ soit une copule, il suffit qu'elle vérifie les trois propriétés suivantes :

- $\forall u \in [0, 1], \varphi(u, 0) = \varphi(0, u) = 0$;
- $\forall u \in [0, 1], \varphi(u, 1) = \varphi(1, u) = u$;
- $\forall (x, y) \in]0, 1[^2, \partial_{1,2}^2(\varphi)(x, y) \geq 0$.

Partie IV. Familles de copules et simulation.

10. Soit M la fonction définie sur $[0, 1]^2$ par : $\forall (u, v) \in [0, 1]^2, M(u, v) = uv(u + v - uv)$.

a) Montrer que la fonction $S : (u, v) \mapsto u + v - 2uv$ admet sur $[0, 1]^2$ un minimum global c et un maximum global d et les calculer.

b) Montrer que M est une copule à densité.

c) Soit $\theta \in \mathbf{R}$ et M_θ la fonction définie sur $[0, 1]^2$ par : $\forall (u, v) \in [0, 1]^2, M_\theta(u, v) = (1 - \theta)uv + \theta M(u, v)$.
Pour quelles valeurs de θ la fonction M_θ est-elle une copule ?

11. Soit N une variable aléatoire suivant la loi de Bernoulli de paramètre $p \in]0, 1[$ et U_0, V_0, U_1, V_1 quatre variables aléatoires suivant chacune la loi uniforme sur $[0, 1]$.

On suppose que $(U_0, V_0), (U_1, V_1)$ et N sont mutuellement indépendants, autrement dit, on suppose que pour tout $(u_0, v_0, u_1, v_1, x) \in \mathbf{R}^5$, on a :

$$P([U_0 \leq u_0] \cap [V_0 \leq v_0] \cap [U_1 \leq u_1] \cap [V_1 \leq v_1] \cap [N \leq x]) = F_{(U_0, V_0)}(u_0, v_0) F_{(U_1, V_1)}(u_1, v_1) P([N \leq x]).$$

a) Pour tout $\omega \in \Omega$, on pose : $U_N(\omega) = \begin{cases} U_0(\omega) & \text{si } N(\omega) = 0 \\ U_1(\omega) & \text{si } N(\omega) = 1 \end{cases}$ et $V_N(\omega) = \begin{cases} V_0(\omega) & \text{si } N(\omega) = 0 \\ V_1(\omega) & \text{si } N(\omega) = 1 \end{cases}$.

Montrer que U_N et V_N sont des variables aléatoires et suivent chacune la loi uniforme sur $[0, 1]$.

b) Exprimer la copule associée au couple (U_N, V_N) à l'aide des copules associées aux deux couples (U_0, V_0) et (U_1, V_1) .

12. Soit $p \in]0, 1[$ et C_p la fonction définie sur $[0, 1]^2$ par : $\forall (u, v) \in [0, 1]^2, C_p(u, v) = puv + (1 - p)\min\{u, v\}$.

a) Montrer que C_p est une copule.

b) Proposer une méthode de simulation d'un couple aléatoire (U, V) auquel est associée la copule C_p et donner le code *Scilab* d'une fonction `simulc` réalisant cette simulation pour toute valeur donnée de p .

Cette fonction aura pour seul argument le paramètre p et retournera le couple (u, v) .

13. Soit X et Y deux variables aléatoires à densité, de fonctions de répartition respectives F_X et F_Y .

Pour tout $p \in]0, 1[$, on pose :
$$\begin{cases} G_X(p) = \inf(\{x \in \mathbf{R}; F_X(x) = p\}) \\ G_Y(p) = \inf(\{x \in \mathbf{R}; F_Y(x) = p\}) \end{cases}.$$

Soit C la fonction définie sur $[0, 1]^2$ par :
$$C(u, v) = \begin{cases} F_{(X, Y)}(G_X(u), G_Y(v)) & \text{si } (u, v) \in]0, 1[^2 \\ 0 & \text{si } uv = 0 \\ u & \text{si } v = 1 \\ v & \text{si } u = 1 \end{cases}.$$

Montrer que C est une copule. En déduire un procédé de simulation du couple (X, Y) à partir de la simulation (u, v) d'un couple (U, V) auquel la copule C est associée.

Conception : HEC Paris – ESCP Europe

OPTION SCIENTIFIQUE

MATHÉMATIQUES II

Mercredi 3 mai 2017, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à **encadrer** dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document. **L'utilisation de toute calculatrice et de tout matériel électronique est interdite.** Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

Dans tout le problème :

- toutes les variables aléatoires introduites sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$;
- on note θ un paramètre réel.

Partie I. Une démonstration probabiliste de la formule de Stirling

Pour tout $n \in \mathbb{N}^*$, soit h_n la fonction définie par : $\forall x \in [0, 1], h_n(x) = ((1-x)e^x)^n$.

Pour tout $n \in \mathbb{N}^*$, on pose : $I_n = \int_0^1 h_n(x) dx$.

1.a) À l'aide du changement de variable $u = n(1-x)$, montrer que : $\forall n \in \mathbb{N}^*, I_n = \frac{e^n}{n^{n+1}} \int_0^n u^n e^{-u} du$.

b) Montrer que pour tout $x \in [0, 1[$, on a : $x + \ln(1-x) \leq -\frac{x^2}{2}$.

c) En se référant à une densité de la loi normale centrée réduite, en déduire que : $\forall n \in \mathbb{N}^*, 0 \leq I_n \leq \sqrt{\frac{\pi}{2n}}$.

2. On note h_n^* la restriction à l'intervalle $]0, 1[$ de la fonction h_n .

On pose pour tout $x \in]0, 1[$: $h_n^*(x) = \exp\left(-\frac{nx^2}{2} H(x)\right)$ et $g(x) = (1-x)\ln(1-x) + x - \frac{x^2}{2}$.

a) Montrer que H est prolongeable par continuité en 0. On note encore H la fonction ainsi prolongée.

b) Montrer que la fonction g est convexe et strictement positive sur $]0, 1[$.

c) En déduire que la fonction H réalise une bijection strictement croissante de $[0, 1]$ sur $[1, +\infty[$.

3. Soit $(u_n)_{n \in \mathbb{N}^*}$ une suite convergente de limite nulle telle que : $\lim_{n \rightarrow +\infty} u_n \sqrt{n} = +\infty$ et $\forall n \in \mathbb{N}^*, 0 < u_n < 1$.

a) Donner un exemple d'une telle suite $(u_n)_{n \in \mathbb{N}^*}$.

b) Soit $(v_n)_{n \in \mathbb{N}^*}$ la suite définie par : $\forall n \in \mathbb{N}^*, v_n = H(u_n)$. Montrer que la suite $(v_n)_{n \in \mathbb{N}^*}$ est convergente et préciser sa limite.

- c) Établir pour tout $n \in \mathbb{N}^*$, l'encadrement : $I_n \geq \int_0^{u_n} h_n(x) dx \geq \frac{1}{\sqrt{nv_n}} \int_0^{u_n \sqrt{nv_n}} \exp\left(-\frac{y^2}{2}\right) dy$.
- d) Dédurre des questions 1.c) et 3.c), un équivalent de I_n lorsque n tend vers $+\infty$.
4. Soit $(T_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$, mutuellement indépendantes et de même loi exponentielle de paramètre 1. Pour tout $n \in \mathbb{N}^*$, on pose : $S_n = T_1 + T_2 + \dots + T_n$.
- a) Rappeler la loi suivie par la variable aléatoire S_n et montrer que $\lim_{n \rightarrow +\infty} P(|S_n| \leq n) = \frac{1}{2}$.
- b) Pour tout $n \in \mathbb{N}^*$, on pose : $U_n = \frac{T_{n+1}}{\sqrt{n}}$. Montrer que la suite $(U_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers la constante 0.
- c) En déduire que $\lim_{n \rightarrow +\infty} P(|S_{n+1}| \leq n) = \frac{1}{2}$.
5. Montrer que $n! \sim n^n e^{-n} \sqrt{2\pi n}$ (formule de Stirling).

Partie II. Quelques propriétés de la loi de Cauchy

6. On rappelle que la fonction Arctan est la fonction réciproque de la restriction à l'intervalle ouvert $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$ de la fonction tan, qu'elle est de classe C^∞ sur $\mathbb{R}$ admettant pour dérivée la fonction $x \mapsto \frac{1}{1+x^2}$ pour tout $x \in \mathbb{R}$ et qu'elle réalise une bijection de $\mathbb{R}$ sur $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$.
- a) Montrer que la fonction Arctan est impaire.
- b) Justifier l'existence d'un développement limité à l'ordre 3 de la fonction Arctan en 0 et le déterminer.
- c) Établir pour tout $x \in \mathbb{R}_+$, l'encadrement : $0 \leq \text{Arctan}(x) \leq x$.
- d) Montrer que pour tout $x \in \mathbb{R}_+$, on a : $\text{Arctan}(x) + \text{Arctan}\left(\frac{1}{x}\right) = \frac{\pi}{2}$.
- 7.a) Montrer que la fonction $x \mapsto \frac{1}{\pi} \times \frac{1}{1+(x-\theta)^2}$ est une densité de probabilité sur $\mathbb{R}$.

Dans toute la suite du problème, on note X une variable aléatoire à valeurs réelles, de densité f_X telle que :

$$\forall x \in \mathbb{R}, f_X(x) = \frac{1}{\pi} \times \frac{1}{1+(x-\theta)^2}.$$

On dit que X suit une loi de Cauchy de paramètre θ et on note : $X \hookrightarrow \mathcal{C}_\theta$.

- b) La variable aléatoire X admet-elle une espérance ?
- c) Pour $\theta = 0$, tracer la courbe représentative de f_X dans le plan rapporté à un repère orthogonal.
- 8.a) On note F_X la fonction de répartition de X . Pour tout $x \in \mathbb{R}$, calculer $F_X(x)$.
- b) Montrer que l'équation $F_X(x) = \frac{1}{2}$ d'inconnue x , admet une unique solution que l'on déterminera. Cette solution est la médiane théorique de X .

Partie III. La loi de la moyenne empirique

9. Pour tout $n \in \mathbb{N}^*$ et pour tout $x \in \mathbb{R}$, soit $\varphi_{n,x}$ la fonction définie sur $\mathbb{R}$ par :

$$\forall t \in \mathbb{R}, \varphi_{n,x}(t) = \frac{1}{(1+t^2)(1+(x-nt)^2)}.$$

On admet l'existence d'un unique quadruplet $(\alpha, \beta, \gamma, \delta)$ de réels indépendants de t pour lesquels on a :

$$\forall t \in \mathbf{R}, \varphi_{n,x}(t) = \frac{\alpha t + \beta}{1+t^2} + \frac{\gamma t + \delta}{1+(x-nt)^2}.$$

Pour tout $n \in \mathbf{N}^*$ et pour tout $x \in \mathbf{R}$, on pose : $\sigma_{n,x} = (x^2 + (n+1)^2)(x^2 + (n-1)^2)$.

On admet sans démonstration que : $\alpha = \frac{2nx}{\sigma_{n,x}}, \quad \beta = \frac{1+x^2-n^2}{\sigma_{n,x}}, \quad \gamma = -\frac{2n^3x}{\sigma_{n,x}}, \quad \delta = \frac{n^2(3x^2+n^2-1)}{\sigma_{n,x}}.$

- Établir la convergence de l'intégrale $\int_{-\infty}^{+\infty} \varphi_{n,x}(t) dt$.
 - À l'aide d'une primitive de la fonction $\psi_{n,x} : t \mapsto \frac{2t}{1+t^2} - \frac{2n(nt-x)}{1+(x-nt)^2}$, montrer que $\int_{-\infty}^{+\infty} \psi_{n,x}(t) dt = 0$.
 - Établir la relation : $\forall n \in \mathbf{N}^*, \forall x \in \mathbf{R}, \int_{-\infty}^{+\infty} \varphi_{n,x}(t) dt = \frac{(n+1)\pi}{x^2 + (n+1)^2}.$
10. On pose : $Y = X - \theta$. Pour n entier de $\mathbf{N}^*$, soit $(Y_1, Y_2, \dots, Y_n)$ un n -échantillon de variables aléatoires indépendantes et de même loi que Y .
- Pour tout $n \in \mathbf{N}^*$, on pose : $S_n = \sum_{i=1}^n Y_i$ et $\bar{Y}_n = \frac{S_n}{n}$ (moyenne empirique de l'échantillon $(Y_1, Y_2, \dots, Y_n)$).
- Déterminer la fonction de répartition F_Y de Y . Quelle est la loi de Y ?
 - Quelle est la fonction de répartition de S_2 ? En déduire la loi de $\bar{Y}_2$.
 - Déterminer pour tout $n \in \mathbf{N}^*$, la loi de la variable aléatoire $\bar{Y}_n$.
 - La loi faible des grands nombres s'applique-t-elle à la suite $(\bar{Y}_n)_{n \in \mathbf{N}^*}$? Pourquoi ?
11. Soit $(N, n) \in \mathbf{N}^{*2}$. On veut simuler N réalisations de la moyenne empirique $\bar{Y}_n$.

On suppose que l'on connaît une fonction *Scilab* cauchy telle que la commande `A=cauchy(N,n)` retourne une matrice $A \in \mathcal{M}_{N,n}(\mathbf{R})$, réalisation d'une famille $(Y_{i,j})_{\substack{1 \leq i \leq N \\ 1 \leq j \leq n}}$ de variables aléatoires indépendantes de loi C_0 .

Soit M une matrice de $\mathcal{M}_{N,n}(\mathbf{R})$ avec $(N, n) \in \mathbf{N}^{*2}$. On rappelle que dans le langage *Scilab* :

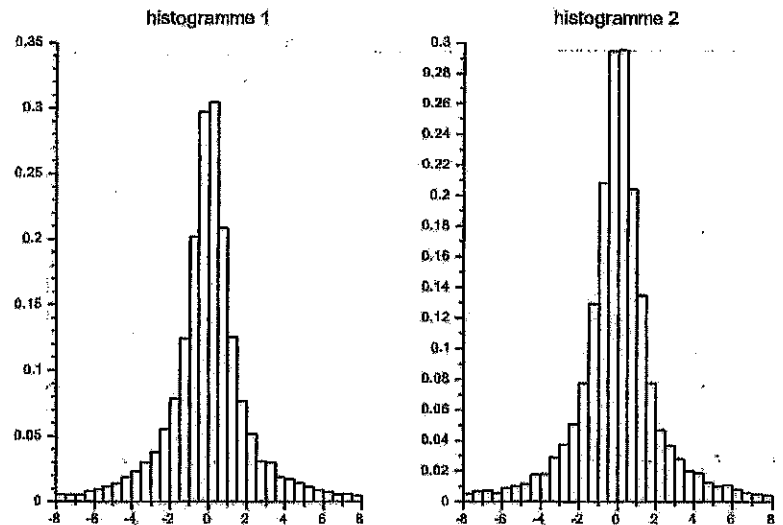
- la commande `sum(M)` retourne une matrice de $\mathcal{M}_{1,1}(\mathbf{R})$ contenant la somme de tous les éléments de M ;
- la commande `sum(M, 'r')` retourne un vecteur ligne de $\mathcal{M}_{1,n}(\mathbf{R})$ contenant les sommes des éléments de M calculées colonne par colonne ;
- la commande `sum(M, 'c')` retourne un vecteur colonne de $\mathcal{M}_{N,1}(\mathbf{R})$ contenant les sommes des éléments de M calculées ligne par ligne ;
- la commande `linspace(a,b,m)` retourne un vecteur ligne de m valeurs régulièrement espacées entre a et b et l'on obtient le même vecteur avec la commande `(a : l : b)` en prenant $l = \frac{b-a}{m-1}$;
- la commande `histplot(y,data)` permet de représenter les éléments du vecteur `data` sous la forme d'un histogramme ; les classes de l'histogramme sont définies par le vecteur strictement croissant `y` : si ce vecteur contient m éléments `y(1), y(2), ..., y(m)` tels que `y(1) < y(2) < ... < y(m)`, alors la première classe de l'histogramme est l'intervalle `[y(1), y(2)]` et les autres classes sont les intervalles `[y(i), y(i+1)]` pour $2 \leq i \leq m$.

- Compléter le programme suivant afin que la matrice `MoyEmp` contienne 12000 réalisations de la moyenne empirique $\bar{Y}_{200}$.

```

N=12000;n=200;
A=cauchy(N,n)
MoyEmp= .....
x=(-8; 0.5; 8)
histplot(x,MoyEmp) \\ histogramme 1
histplot(x,A(:,1)) \\ histogramme 2

```



- b) Les histogrammes 1 et 2 ont été obtenus à l'aide de ce programme. Expliquer en quoi ce couple d'histogrammes illustre le résultat de la question 10.c).

Partie IV. La loi de la médiane empirique

Dans les questions 13, 14 et 15, on suppose que le paramètre θ est inconnu.

On rappelle que $X \hookrightarrow \mathcal{C}_\theta$. Pour n entier de $\mathbb{N}^*$, on note $(X_1, X_2, \dots, X_{2n+1})$ un $(2n+1)$ -échantillon de variables aléatoires indépendantes et de même loi que X .

On admet l'existence de $(2n+1)$ fonctions $g_1, g_2, \dots, g_{2n+1}$ continues sur $\mathbb{R}^{2n+1}$ à valeurs réelles, telles que les variables aléatoires réelles $\hat{X}_1, \hat{X}_2, \dots, \hat{X}_{2n+1}$ définies par : $\forall k \in \llbracket 1, 2n+1 \rrbracket, \hat{X}_k = g_k(X_1, X_2, \dots, X_{2n+1})$ soient des variables aléatoires à densité et que pour tout $\omega \in \Omega$, les réels $\hat{X}_1(\omega), \hat{X}_2(\omega), \dots, \hat{X}_{2n+1}(\omega)$ soient un réarrangement par ordre croissant de $X_1(\omega), X_2(\omega), \dots, X_{2n+1}(\omega)$: $\forall \omega \in \Omega, \hat{X}_1(\omega) \leq \hat{X}_2(\omega) \leq \dots \leq \hat{X}_{2n+1}(\omega)$. En particulier, la variable aléatoire $\hat{X}_{n+1}$ est la médiane empirique de l'échantillon $(X_1, X_2, \dots, X_{2n+1})$.

12. Pour tout $h \in \mathbb{R}_+^*$ et pour tout $x \in \mathbb{R}$, on note Z la variable aléatoire discrète définie par :

$$\forall \omega \in \Omega, Z(\omega) = \text{Card} \{i \in \llbracket 1, 2n+1 \rrbracket; x < X_i(\omega) < x+h\}.$$

On note A et B les deux événements suivants : $A = [x < \hat{X}_{n+1} \leq x+h]$ et $B = A \cap [Z=1]$.

a) Établir la relation : $P(B) = (n+1) \binom{2n+1}{n} (F_X(x))^n (F_X(x+h) - F_X(x)) (1 - F_X(x+h))^n$.

b) On suppose que le réel x est fixé. Montrer qu'il existe un réel K indépendant de h pour lequel on a :

$$0 \leq P(A) - P(B) \leq K (F_X(x+h) - F_X(x))^2.$$

c) Montrer que $\hat{X}_{n+1}$ admet une densité $f_{\hat{X}_{n+1}}$ donnée par :

$$\forall x \in \mathbb{R}, f_{\hat{X}_{n+1}}(x) = (n+1) \binom{2n+1}{n} (F_X(x))^n (1 - F_X(x))^n f_X(x).$$

13.a) Établir l'équivalence suivante : $x f_{\hat{X}_{n+1}}(x) \underset{x \rightarrow +\infty}{\sim} (n+1) \binom{2n+1}{n} \frac{1}{x^{n+1}} \times \frac{1}{x^{n+1}}$.

b) En déduire l'existence de l'espérance $E(\hat{X}_{n+1})$ de la variable aléatoire $\hat{X}_{n+1}$.

c) Justifier que $\hat{X}_{n+1}$ est un estimateur du paramètre θ . Calculer $E(\hat{X}_{n+1} - \theta)$. Conclure.

d) À quelle condition nécessaire et suffisante portant sur n , la variable aléatoire $\hat{X}_{n+1}$ admet-elle une variance ?

14. On note $F_{\hat{X}_{n+1}}$ la fonction de répartition de $\hat{X}_{n+1}$. Soit ε un réel strictement positif.

a) Établir la relation : $\forall t \in \mathbf{R}, f_{\hat{X}_{n+1}}(2\theta - t) = f_{\hat{X}_{n+1}}(t)$. En déduire que $P(|\hat{X}_{n+1} - \theta| \geq \varepsilon) = 2 F_{\hat{X}_{n+1}}(\theta - \varepsilon)$.

b) Montrer que la suite d'estimateurs $(\hat{X}_{n+1})_{n \in \mathbf{N}^*}$ converge en probabilité vers θ .

c) La suite $(\hat{X}_{n+1})_{n \in \mathbf{N}^*}$ converge-t-elle en loi vers la variable certaine θ ?

15. Pour tout entier $n \geq 2$, on pose : $W_{n+1} = \frac{2\sqrt{2n+1}}{\pi}(\hat{X}_{n+1} - \theta)$.

a) On note $f_{W_{n+1}}$ la densité continue sur $\mathbf{R}$ de W_{n+1} . Montrer que :

$$\forall x \in \mathbf{R}, f_{W_{n+1}}(x) = \frac{n+1}{2\sqrt{2n+1}} \binom{2n+1}{n} \left[\frac{1}{4} - \frac{1}{\pi^2} \left(\operatorname{Arctan} \left(\frac{\pi x}{2\sqrt{2n+1}} \right) \right)^2 \right]^n \times \left(1 + \frac{\pi^2 x^2}{4(2n+1)} \right)^{-1}.$$

b) Montrer que pour tout $x \in \mathbf{R}$, on a : $\lim_{n \rightarrow +\infty} f_{W_{n+1}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$.

On admet que ce résultat implique la convergence en loi de la suite de variables aléatoires $(W_{n+1})_{n \geq 2}$ vers une variable aléatoire T qui suit la loi normale centrée réduite.

c) On note Φ la fonction de répartition de T . Soit α un réel vérifiant $0 < \alpha < 1$; on pose : $t_\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

Déterminer un intervalle de confiance asymptotique pour θ , centré sur $\hat{X}_{n+1}$, au niveau de confiance $1 - \alpha$.

16. Dans le langage *Scilab*, la fonction `gsort` permet de trier les éléments d'une matrice réelle A :

- la commande `gsort(A, 'r')` renvoie une copie de A triée colonne par colonne, par ordre décroissant (chaque colonne est triée indépendamment des autres) ;
- la commande `gsort(A, 'c')` renvoie une copie de A triée ligne par ligne, par ordre décroissant (chaque ligne est triée indépendamment des autres).

On suppose que $\theta = 0$ et on considère p réalisations ($p \geq 10^4$) du $(2n+1)$ -échantillon $(X_1, X_2, \dots, X_{2n+1})$. Recopier et compléter le code suivant afin que son exécution retourne un vecteur `MedianeEmp` de p réalisations de la médiane empirique $\hat{X}_{n+1}$, puis un vecteur W de p réalisations de W_{n+1} .

```
A=cauchy(p,2*n+1)
S=gsort .....
MedianeEmp= .....
W= .....
```

Conception : HEC Paris – ESCP Europe

OPTION SCIENTIFIQUE

MATHÉMATIQUES II

Lundi 7 mai 2018, de 14 h. à 18 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à **encadrer** dans la mesure du possible les résultats de leurs calculs.

Aucun document n'est autorisé. **L'utilisation de toute calculatrice et de tout matériel électronique est interdite.** Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

- On rappelle que la série $\sum_{n \geq 1} \frac{1}{n^x}$ converge si et seulement si le réel x est strictement supérieur à 1.
- On note ζ la fonction définie sur $]1, +\infty[$ par : $\forall x > 1, \zeta(x) = \sum_{n=1}^{+\infty} \frac{1}{n^x}$; on admet que $\zeta(2) = \frac{\pi^2}{6}$.
- Toutes les variables aléatoires introduites dans le problème sont supposées définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$.
- Si R est un élément de la tribu $\mathcal{A}$, on note $\bar{R}$ l'événement contraire de R .

L'objet du problème est l'étude de la convergence de séries dont les termes sont des variables aléatoires.

La convergence de telles séries, en loi ou en probabilité, est celle de la suite des sommes partielles associées.

Autrement dit, pour toute suite de variables aléatoires $(U_n)_{n \in \mathbb{N}^*}$, on dit que la série $\sum_{n \geq 1} U_n$ converge (en loi ou

en probabilité) lorsque la suite de variables aléatoires $\left(\sum_{k=1}^n U_k\right)_{n \in \mathbb{N}^*}$ converge (en loi ou en probabilité).

Partie I. Séries télescopiques

Dans cette partie, on considère une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires indépendantes, de même loi qu'une variable aléatoire X de référence, et on étudie la convergence de la série aléatoire $\sum_{n \geq 1} \left(\frac{X_n}{n} - \frac{X_{n+1}}{n+1}\right)$.

1.a) Justifier la convergence de la série $\sum_{n \geq 1} \frac{1}{n(n+1)}$. Calculer la somme de cette série.

b) Dans cet exemple, quelle est la loi de la variable aléatoire de référence X ?

2. Pour tout $n \in \mathbf{N}^*$, soit Y_n une variable aléatoire admettant pour densité la fonction f_n définie par :

$$f_n(x) = \begin{cases} 0 & \text{si } x < -\frac{1}{n+1} \quad \text{ou} \quad \text{si } x > 1 \\ 1 + (n+1)x & \text{si } -\frac{1}{n+1} \leq x \leq 0 \\ c_n & \text{si } 0 < x < \frac{n}{n+1} \\ (n+1)(1-x) & \text{si } \frac{n}{n+1} \leq x \leq 1 \end{cases}$$

où c_n est une constante strictement positive.

- Calculer la valeur de c_n et représenter graphiquement f_3 dans le plan rapporté à un repère orthonormé.
 - Déterminer la fonction de répartition F_n de Y_n . La fonction F_n est-elle de classe C^1 sur $\mathbf{R}$?
 - Montrer que la suite $(Y_n)_{n \in \mathbf{N}^*}$ converge en loi vers une variable aléatoire Y dont on précisera la loi.
3. Dans cette question, on suppose que la variable aléatoire de référence X possède une densité f bornée.

Pour tout $n \in \mathbf{N}^*$, on pose : $D_n = X_1 - \frac{X_{n+1}}{n+1}$.

- Montrer que la suite de variables aléatoires $\left(\frac{X_{n+1}}{n+1}\right)_{n \in \mathbf{N}^*}$ converge en probabilité vers 0.
- En déduire que la suite de variables aléatoires $(D_n)_{n \in \mathbf{N}^*}$ converge en loi vers X .
- Justifier que la variable aléatoire D_n admet pour densité la fonction f_{D_n} telle que :

$$\forall x \in \mathbf{R}, f_{D_n}(x) = (n+1) \int_{-\infty}^{+\infty} f(t) f((n+1)(t-x)) dt.$$

- En déduire une nouvelle démonstration du résultat obtenu dans la question 2.c).
4. Dans cette question, on suppose que la variable aléatoire de référence X suit la loi normale centrée réduite.

Pour tout $n \in \mathbf{N}^*$, on pose : $U_n = \frac{X_n}{n} - \frac{X_{n+1}}{n+1}$ et $T_n = \sum_{k=1}^n U_k$.

- Pour tout $n \in \mathbf{N}^*$, déterminer la loi de la variable aléatoire U_n .
- Justifier la convergence en loi de la série $\sum_{n \geq 1} U_n$.
- Soit $(U'_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires indépendantes vérifiant : $\forall n \in \mathbf{N}^*$, U'_n et U_n ont même loi.

Pour tout $n \in \mathbf{N}^*$, on pose : $T'_n = \sum_{k=1}^n U'_k$.

(i) Justifier que la suite de variables aléatoires $(T'_n)_{n \in \mathbf{N}^*}$ converge en loi vers une variable aléatoire qui suit la loi normale centrée de variance $\frac{\pi^2}{3} - 1$.

(ii) Pourquoi ce résultat ne contredit-il pas ceux obtenus dans les questions 3.b) et 4.b) ?

Partie II. Séries harmoniques « lacunaires »

Dans cette partie, on étudie des séries numériques obtenues à partir de la série harmonique divergente $\sum_{n \geq 1} \frac{1}{n}$ par effacement de certains de ses termes.

Pour toute partie $\mathcal{G}$ de $\mathbf{N}^*$, on note $\mathbf{1}_{\mathcal{G}}$ la fonction indicatrice de $\mathcal{G}$, c'est-à-dire la fonction définie sur $\mathbf{N}^*$ à valeurs dans $\{0, 1\}$ telle que : $\forall k \in \mathbf{N}^*, \mathbf{1}_{\mathcal{G}}(k) = \begin{cases} 1 & \text{si } k \in \mathcal{G} \\ 0 & \text{sinon} \end{cases}$.

Pour tout entier n supérieur ou égal à 1, on pose : $\forall \mathcal{G} \subset \mathbf{N}^*, h_n(\mathcal{G}) = \sum_{k=1}^n \frac{\mathbf{1}_{\mathcal{G}}(k)}{k}$.

Dans la question 5, on étudie deux cas de convergence et la question 6 est consacrée à un cas de divergence.

5. On pose : $\mathcal{D} = \{n^2; n \in \mathbb{N}^*\}$ et $\mathcal{T} = \{n^3; n \in \mathbb{N}^*\}$.

- Exprimer $h_n(\mathcal{D})$ à l'aide d'une somme partielle de la série de Riemann $\sum_{n \geq 1} \frac{1}{n^2}$.
- En déduire la convergence de la suite $(h_n(\mathcal{D}))_{n \in \mathbb{N}^*}$ et calculer la somme $\sum_{n=1}^{+\infty} \frac{1_{\mathcal{D}}(n)}{n}$.
- Justifier que $\mathcal{D} \cap \mathcal{T}$ est l'ensemble des entiers m pour lesquels $m^{1/6} \in \mathbb{N}^*$.

Pour traiter cette question, on admet que la racine carrée d'un entier naturel qui n'appartient pas à $\mathcal{D}$ est un nombre irrationnel, c'est-à-dire, un nombre qui ne peut pas s'écrire comme le quotient de deux entiers.

- Montrer que la suite $(h_n(\mathcal{D} \cup \mathcal{T}))_{n \in \mathbb{N}^*}$ est convergente et exprimer la somme $\sum_{n=1}^{+\infty} \frac{1_{\mathcal{D} \cup \mathcal{T}}(n)}{n}$ à l'aide de certaines valeurs de la fonction ζ .

6. On note $\mathcal{I}$ l'ensemble des entiers naturels impairs : $\mathcal{I} = \{2n-1; n \in \mathbb{N}^*\}$.

Pour tout $n \in \mathbb{N}^*$, on pose : $u_n = \int_n^{n+1} \left(\frac{1}{2n-1} - \frac{1}{2t-1} \right) dt$.

- Pour tout $n \in \mathbb{N}^*$, établir l'encadrement : $0 \leq u_n \leq \frac{2}{(2n-1)^2}$.
- Montrer que pour tout $n \in \mathbb{N}^*$, on a : $h_n(\mathcal{I}) = \sum_{k=1}^{\lfloor \frac{n+1}{2} \rfloor} u_k + \int_1^{\lfloor \frac{n+3}{2} \rfloor} \frac{1}{2t-1} dt$.
- Pour tout $n \in \mathbb{N}^*$, justifier l'encadrement : $0 \leq \ln \left(\frac{1}{n} \left(2 \left\lfloor \frac{n+3}{2} \right\rfloor - 1 \right) \right) \leq \frac{2}{n}$.
- Montrer que la série de terme général u_n est convergente.

On pose : $\delta = \sum_{n=1}^{+\infty} u_n$. Établir l'égalité : $\lim_{n \rightarrow +\infty} (h_n(\mathcal{I}) - \ln(\sqrt{n})) = \delta$.

- Pour tout $n \in \mathbb{N}^*$, montrer que l'on a : $\sum_{k=n+1}^{+\infty} u_k \leq \frac{1}{2n-1}$.

- Justifier pour tout entier $n \geq 3$, l'encadrement : $-\frac{1}{n} \leq \delta - (h_n(\mathcal{I}) - \ln(\sqrt{n})) \leq \frac{1}{n-2}$.

g) La fonction *Scilab* suivante, dont le script est incomplet (ligne (6)), permet de donner une valeur approchée de δ en calculant successivement des valeurs de $h_n(\mathcal{I}) - \ln(\sqrt{n})$ jusqu'à atteindre une précision donnée.

```
(1)  function s=delta(eps)
(2)  n=3;
(3)  s=1+1/3-(log(3)/2);
(4)  while 1/(n-2)>eps
(5)  n=n+2;
(6)  s=s+1/n+ .....;
(7)  end;
(8)  endfunction
```

- Quelles sont les valeurs de $h_n(\mathcal{I}) - \ln(\sqrt{n})$ affectées successivement à la variable **s** lorsqu'on applique cette fonction à **eps=0.2**?
- Compléter la ligne (6).
- Pour quelles raisons l'algorithme utilisé peut-il assurer une précision arbitraire au calcul de la valeur approchée de δ ?

Partie III. Séries de Riemann alternées

Dans cette partie, on note $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires indépendantes, de même loi uniforme sur la paire $\{-1, +1\}$, c'est-à-dire : pour tout $n \in \mathbf{N}^*$, $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$.

Pour tout $n \in \mathbf{N}^*$, on pose : $S_n = \sum_{k=1}^n X_k$.

7. Soit $(x_n)_{n \in \mathbf{N}^*}$ une suite réelle. Pour tout $n \in \mathbf{N}^*$, on pose : $s_n = \sum_{k=1}^n x_k$.

On suppose l'existence d'un réel $\alpha \geq 0$ et d'un réel $M > 0$ tels que : $\forall n \in \mathbf{N}^*, |s_n| \leq M n^\alpha$.

a) Soit β un réel tel que $\beta > \alpha$.

(i) Montrer pour tout entier $n \geq 2$, l'égalité : $\sum_{k=1}^n \frac{x_k}{k^\beta} = \frac{s_n}{n^\beta} + \sum_{k=1}^{n-1} s_k \left(\frac{1}{k^\beta} - \frac{1}{(k+1)^\beta} \right)$.

(ii) En déduire que la série $\sum_{n \geq 1} \frac{x_n}{n^\beta}$ est convergente.

b) Justifier pour tout réel $x > 0$, la convergence de la série $\sum_{n \geq 1} \frac{(-1)^n}{n^x}$.

8. Soit s et t des réels strictement positifs et n un entier supérieur ou égal à 1.

a) Calculer l'espérance de la variable aléatoire e^{tS_n} .

b) En utilisant l'écriture de e^u ($u \in \mathbf{R}$) sous forme de somme d'une série, établir l'inégalité :

$$\frac{1}{2} (e^t + e^{-t}) \leq e^{\frac{t^2}{2}}.$$

c) À l'aide de l'inégalité de Markov, montrer que : $P(|S_n| > s) \leq \exp\left(\frac{n t^2}{2} - ts\right)$.

d) Justifier l'inégalité : $P(|S_n| > s) \leq 2 \exp\left(-\frac{s^2}{2n}\right)$.

9. Pour tout réel $\alpha \geq 0$, on pose : $\mathcal{C}_\alpha = \bigcap_{n=1}^{+\infty} \left(\bigcup_{k=n}^{+\infty} [|S_k| > k^\alpha] \right)$.

a) Justifier que $\mathcal{C}_\alpha$ est un élément de la tribu $\mathcal{A}$.

b) Montrer que si $\alpha > \frac{1}{2}$, alors la série $\sum_{n \geq 1} P(|S_n| > n^\alpha)$ est convergente.

c) En déduire que pour tout réel $\alpha > \frac{1}{2}$, on a $P(\mathcal{C}_\alpha) = 0$.

10. Dans cette question, on s'intéresse à la série aléatoire $\sum_{n \geq 1} \frac{X_n}{n}$.

On note $\mathcal{C}$ l'ensemble des $\omega \in \Omega$ pour lesquels la série numérique $\sum_{n \geq 1} \frac{X_n(\omega)}{n}$ converge et pour tout $n \in \mathbf{N}^*$,

on pose : $K_n = \sum_{k=1}^n \frac{X_k}{k}$.

Soit K l'application définie sur Ω par : $K(\omega) = \begin{cases} \lim_{n \rightarrow +\infty} K_n(\omega) & \text{si } \omega \in \mathcal{C} \\ 0 & \text{sinon} \end{cases}$.

On admet sans démonstration que $\mathcal{C}$ est un élément de la tribu $\mathcal{A}$ et que K est une variable aléatoire.

a) En utilisant le résultat de la question 7.a), montrer que si α vérifie $0 \leq \alpha < 1$, alors on a : $\overline{\mathcal{C}_\alpha} \subset \mathcal{C}$.

b) À l'aide des résultats de la question 9, montrer que $P(\mathcal{C}) = 1$.

c) Pour tout réel $\varepsilon > 0$, on considère l'événement $E(\varepsilon)$ défini par : $E(\varepsilon) = \bigcap_{N=1}^{+\infty} \left(\bigcup_{n=N}^{+\infty} [|K - K_n| > \varepsilon] \right)$.

Montrer que $P(E(\varepsilon)) = 0$ et en déduire que la suite de variables aléatoires $(K_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers la variable aléatoire K .

On admet sans démonstration que la suite $(K_n)_{n \in \mathbb{N}^*}$ converge alors en loi vers K .

Dans les questions 11 et 12, on considère une suite $(B_n)_{n \in \mathbb{N}^*}$ de variables aléatoires indépendantes suivant chacune la loi de Bernoulli de paramètre $\frac{1}{2}$. Pour tout $n \in \mathbb{N}^*$, on pose : $H_n = \sum_{k=1}^n \frac{B_k}{k}$.

11.a) Pour tout $n \in \mathbb{N}^*$, déterminer l'espérance et la variance de H_n et trouver leurs limites respectives lorsque n tend vers $+\infty$.

b) Montrer que, quel que soit le réel $r > 0$, on a : $\lim_{n \rightarrow +\infty} P([H_n \leq r]) = 0$.

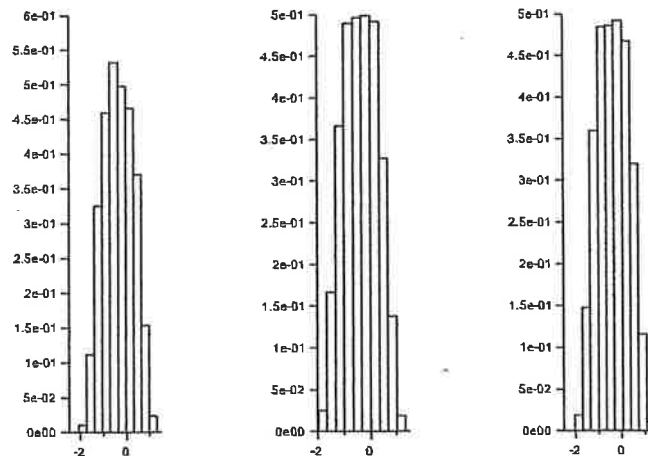
c) La fonction Scilab suivante, dont le script est incomplet (ligne(5)), permet d'effectuer p simulations de la variable aléatoire $H_n - h_n(\mathcal{I})$, où $h_n(\mathcal{I})$ a été définie dans la partie II (préambule et question 6).

```
(1)  function y=simul(n,p)
(2)  y=zeros(p,1);
(3)  for i=1:p
(4)  for k=1:n
(5)  y(i,1)=y(i,1)+(grand(1,1,'bin',1,0.5)+.....)/k;
(6)  end;
(7)  end;
(8)  endfunction
```

(i) Compléter la ligne (5).

(ii) Les trois histogrammes suivants représentent la distribution simulée de la variable aléatoire $H_n - h_n(\mathcal{I})$ pour $n = 50$, $n = 100$ et $n = 200$. Par quelles instructions ont-ils pu être obtenus ?

(iii) Pourquoi ces histogrammes suggèrent-ils une convergence en loi de la suite $(H_n - h_n(\mathcal{I}))_{n \in \mathbb{N}^*}$?



12. Pour tout $n \in \mathbb{N}^*$, on pose : $B'_n = \frac{1 + X_n}{2}$.

a) Justifier pour tout $n \in \mathbb{N}^*$, la relation : $\sum_{k=1}^n \frac{B'_k}{k} - h_n(\mathcal{I}) = \frac{K_n}{2} + \sum_{k=1}^n \frac{(-1)^k}{2k}$.

b) En déduire que la suite de variables aléatoires $(H_n - h_n(\mathcal{I}))_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire de la forme $\lambda K + \mu$, où λ et μ sont des réels dont on précisera la valeur.

FIN

Conception : HEC Paris – ESCP Europe

OPTION SCIENTIFIQUE

MATHÉMATIQUES II

Jeudi 2 mai 2019, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à **encadrer** dans la mesure du possible les résultats de leurs calculs.

Aucun document n'est autorisé. **L'utilisation de toute calculatrice et de tout matériel électronique est interdite.** Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

La régression logistique permet de modéliser l'influence qu'exercent des facteurs exogènes sur une variable binaire, c'est-à-dire une variable ne pouvant prendre que deux valeurs.

Outre son domaine d'application privilégié qui est l'apprentissage automatique (machine learning), la régression logistique est couramment utilisée aussi bien en médecine qu'en actuariat et en économétrie.

Partie I. Fonction logistique et lois logistiques

On appelle *fonction logistique* la fonction Λ définie sur $\mathbf{R}$ par : $\forall x \in \mathbf{R}, \Lambda(x) = \frac{1}{1 + e^{-x}}$.

1.a) Montrer que Λ est une bijection de $\mathbf{R}$ sur $]0, 1[$, dont la bijection réciproque est la fonction L définie par :

$$\forall x \in]0, 1[, L(x) = \ln\left(\frac{x}{1-x}\right).$$

b) Calculer la dérivée de la fonction Λ .

c) Justifier l'existence d'un unique réel x_0 tel que : $\Lambda(x_0) = x_0$.

d) Établir pour tout $x \in \mathbf{R}$, l'inégalité : $|\Lambda(x) - x| \leq |x - x_0|$.

2. Le script *Scilab* suivant, dont la ligne (1) définit la fonction Λ , permet de calculer une valeur approchée de x_0 par la méthode de dichotomie.

```
(1) def f('y=Lambda(x)', 'y=1/(1+exp(-x))');
(2) a=0;
(3) b=1;
(4) eps= .....;
(5) while b-a>eps;
(6)   c=(a+b)/2;
(7)   if Lambda(c)>c then .....; else b= .....; end;
(8) end;
(9) x0=(a+b)/2
```

- a) Compléter la ligne (7) et justifier le choix des valeurs affectées en lignes (2) et (3) aux variables **a** et **b**.
- b) Quelle valeur maximale peut-on affecter en ligne (4) à la variable **eps** pour être assuré que l'erreur d'approximation commise ne dépasse pas 10^{-4} ?
- c) Que peut-on dire de la valeur numérique obtenue par l'instruction (10) suivante ?

(10) `Lambda(x0)-x0`

3. On note λ la dérivée de la fonction Λ .

- a) Vérifier que λ est une densité de probabilité.
- b) Préciser la parité de la fonction λ ; donner l'allure de sa courbe représentative dans le plan rapporté à un repère orthogonal et en déterminer les points d'inflexion.

On dit qu'une variable aléatoire Z suit la *loi logistique standard* si elle admet la fonction λ pour densité.

Pour tout couple $(r, s) \in \mathbf{R} \times \mathbf{R}_+^*$, on dit qu'une variable aléatoire Y suit la loi logistique $\mathcal{L}(r, s)$ si la variable aléatoire Z définie par $Z = \frac{Y-r}{s}$ suit la loi logistique standard.

- 4.a) Justifier qu'une variable aléatoire qui suit une loi logistique $\mathcal{L}(r, s)$ admet des moments de n'importe quel ordre et en indiquer l'espérance.
 - b) En utilisant la méthode d'inversion, écrire le script d'une fonction *Scilab*, `function S=grandlogis(n,p,r,s)`, fournissant pour tout couple (n,p) d'entiers strictement positifs, une matrice **S** à **n** lignes et **p** colonnes dont les coefficients sont des simulations de variables aléatoires indépendantes suivant la loi logistique $\mathcal{L}(r, s)$.
 - c) Décrire un procédé permettant de calculer une valeur approchée de la variance de la loi logistique standard à l'aide de la fonction `grandlogis`.
5. Soit U_1 et U_2 deux variables aléatoires indépendantes suivant chacune la loi exponentielle de paramètre 1.
- a) Montrer que la variable aléatoire $Z = \ln\left(\frac{U_1}{U_2}\right)$ suit la loi logistique standard (on pourra utiliser un changement de variable exponentiel, c'est-à-dire de la forme $t = e^x$).
 - b) En déduire un nouveau script *Scilab* permettant de simuler une variable aléatoire suivant la loi logistique standard à l'aide de la fonction `grand`.

Partie II. Variance de la loi logistique standard

- Pour tout couple $(a, b) \in \mathbf{R}^2$, on note $\mathcal{I}m(z)$ la partie imaginaire b du nombre complexe $z = a + ib$.
- Pour tout polynôme $P = \sum_{k=0}^d a_k X^k \in \mathbf{R}[X]$ de degré $d \in \mathbf{N}$, les termes non nuls $a_k X^k$ sont appelés les monômes de P et les a_k leurs coefficients.
- Dans la factorisation $P = a_d \prod_{k=1}^d (X - z_k)$ de P dans $\mathbf{C}[X]$ (lorsque $d \neq 0$), la somme $\sum_{k=1}^d z_k$ est appelée la somme des racines complexes de P , que les nombres complexes $z_1, z_2, \dots, z_d$ soient distincts ou non.

Pour tout $n \in \mathbf{N}$, on pose : $P_n = \sum_{k=0}^n (-1)^k \binom{2n+1}{2k+1} (X-1)^{n-k}$.

- 6.a) Expliciter les polynômes P_0 et P_1 .
- b) Pour tout $n \in \mathbf{N}^*$, préciser le degré du polynôme P_n et donner les coefficients de ses deux monômes de plus hauts degrés.
- c) Utiliser le résultat précédent pour montrer que pour tout $n \in \mathbf{N}^*$, la somme des racines complexes de P_n est égale à $\frac{2n(n+1)}{3}$.

7. Soit $x \in \mathbf{R}$ et $n \in \mathbf{N}$.

a) Justifier les égalités suivantes :

$$\sin((2n+1)x) = \operatorname{Im} \left((\cos(x) + i \sin(x))^{2n+1} \right) = \sum_{k=0}^n (-1)^k \binom{2n+1}{2k+1} \cos^{2(n-k)}(x) \times \sin^{2k+1}(x).$$

b) En déduire, pour tout $x \in]0, \pi[$, la relation : $\frac{\sin((2n+1)x)}{\sin^{2n+1}(x)} = P_n\left(\frac{1}{\sin^2(x)}\right)$.

c) À l'aide du résultat de la question 6.c), montrer que pour tout $n \in \mathbf{N}^*$, on a :

$$\sum_{k=1}^n \frac{1}{\sin^2\left(\frac{k\pi}{2n+1}\right)} = \frac{2n(n+1)}{3}.$$

8. Soit $x \in]0, \frac{\pi}{2}[$.

a) Justifier les inégalités suivantes : $\sin(x) \leq x \leq \tan(x)$ et $\frac{1}{\sin^2(x)} - 1 \leq \frac{1}{x^2} \leq \frac{1}{\sin^2(x)}$.

b) En utilisant le résultat de la question 7.c), en déduire, pour tout $n \in \mathbf{N}^*$, l'encadrement :

$$\frac{n(2n-1)}{3} \leq \frac{(2n+1)^2}{\pi^2} \sum_{k=1}^n \frac{1}{k^2} \leq \frac{2n(n+1)}{3}.$$

c) Établir l'égalité : $\sum_{k=1}^{+\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$.

9. Soit Z une variable aléatoire suivant la loi logistique standard.

a) À l'aide d'une intégration par parties, justifier que la variance de Z , notée $V(Z)$, vérifie l'égalité :

$$V(Z) = 4 \int_0^{+\infty} \frac{x e^{-x}}{1 + e^{-x}} dx.$$

b) Établir pour tout $n \in \mathbf{N}$, l'égalité :

$$\int_0^{+\infty} \frac{x e^{-x}}{1 + e^{-x}} dx = \sum_{k=0}^n (-1)^k \int_0^{+\infty} x e^{-(k+1)x} dx + I_n, \quad \text{où } I_n = (-1)^{n+1} \int_0^{+\infty} \frac{x e^{-(n+2)x}}{1 + e^{-x}} dx.$$

c) Montrer que l'intégrale I_n tend vers 0 lorsque n tend vers $+\infty$ et en déduire l'égalité :

$$\int_0^{+\infty} \frac{x e^{-x}}{1 + e^{-x}} dx = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)^2}.$$

d) En utilisant la formule établie en 8.c), déduire de l'égalité précédente que la variance de Z est égale à $\frac{\pi^2}{3}$.

10.a) Établir la convergence des deux intégrales $\int_0^{+\infty} \ln(x) e^{-x} dx$ et $\int_0^{+\infty} (\ln(x))^2 e^{-x} dx$.

b) On pose $I = \int_0^{+\infty} \ln(x) e^{-x} dx$ et $J = \int_0^{+\infty} (\ln(x))^2 e^{-x} dx$.

En utilisant le résultat de la question 5.a), calculer $J - I^2$.

Partie III. Estimation à partir de données binaires

Dans cette partie, θ est un paramètre réel inconnu et F désigne la fonction de répartition d'une variable aléatoire à densité dont une densité f est continue et strictement positive sur $\mathbf{R}$.

Soit $(Y_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires indépendantes définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P}_\theta)$ suivant chacune la loi de Bernoulli de paramètre $F(\theta)$.

11. Justifier que F est une bijection de $\mathbf{R}$ sur $]0, 1[$. On note F^{-1} sa bijection réciproque.

12. Pour tout $n \in \mathbf{N}^*$, on pose : $\bar{Y}_n = \frac{1}{n} \sum_{j=1}^n Y_j$.

Montrer que la suite $(\sqrt{n}(\bar{Y}_n - F(\theta)))_{n \in \mathbf{N}^*}$ converge en loi vers une variable aléatoire suivant une loi normale centrée dont on précisera la variance.

13. Pour tout $n \in \mathbf{N}^*$ et tout $\omega \in \Omega$, on pose : $T_n(\omega) = \begin{cases} F^{-1}(\bar{Y}_n(\omega)) & \text{si } 0 < \bar{Y}_n(\omega) < 1 \\ 0 & \text{sinon} \end{cases}$.

De plus, pour tout $n \in \mathbf{N}^*$, on note E_n l'événement $[0 < \bar{Y}_n < 1]$.

a) Calculer $\mathbf{P}_\theta(E_n)$ et trouver la limite de cette probabilité lorsque n tend vers $+\infty$.

b) Soit $x \in \mathbf{R}$ et $n \in \mathbf{N}^*$.

(i) Établir l'égalité ensembliste $\{\omega \in E_n / T_n(\omega) \leq x\} = [\bar{Y}_n \leq F(x)] \cap E_n$ et montrer que $[T_n \leq x]$ est un élément de la tribu $\mathcal{A}$.

(ii) Justifier l'encadrement :

$$\mathbf{P}_\theta([\bar{Y}_n \leq F(x)] \cap E_n) \leq \mathbf{P}_\theta([T_n \leq x]) \leq \mathbf{P}_\theta([\bar{Y}_n \leq F(x)] \cap E_n) + 1 - \mathbf{P}_\theta(E_n).$$

c) Montrer que pour tout $x \neq \theta$, on a : $\lim_{n \rightarrow +\infty} \mathbf{P}_\theta([T_n \leq x]) = \begin{cases} 0 & \text{si } x < \theta \\ 1 & \text{si } x > \theta \end{cases}$.

d) En déduire que $(T_n)_{n \in \mathbf{N}^*}$ est une suite convergente d'estimateurs du paramètre θ .

14. Pour tout $n \in \mathbf{N}^*$ et tout $\omega \in \Omega$, on pose : $U_n(\omega) = \begin{cases} \frac{T_n(\omega) - \theta}{\bar{Y}_n(\omega) - F(\theta)} & \text{si } \bar{Y}_n(\omega) \neq F(\theta) \\ \frac{1}{f(\theta)} & \text{si } \bar{Y}_n(\omega) = F(\theta) \end{cases}$.

On admet sans démonstration que pour tout $n \in \mathbf{N}^*$, U_n est une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbf{P}_\theta)$.

a) Soit $\varepsilon > 0$.

Pour tout $n \in \mathbf{N}^*$, on note $B_n(\varepsilon)$ l'événement $\left[\left| U_n - \frac{1}{f(\theta)} \right| \leq \varepsilon \right]$.

(i) Établir l'existence d'un réel $\alpha > 0$ tel que : $\forall x \in [\theta - \alpha, \theta + \alpha], \left| \frac{1}{f(x)} - \frac{1}{f(\theta)} \right| \leq \varepsilon$.

(ii) Pour un tel α , justifier l'inclusion : $[|T_n - \theta| \leq \alpha] \cap E_n \subset B_n(\varepsilon)$, où E_n a été défini dans la question 13.

b) Montrer que la suite $(U_n)_{n \in \mathbf{N}^*}$ converge en probabilité vers $\frac{1}{f(\theta)}$.

c) En déduire que la suite $(\sqrt{n}(T_n - \theta))_{n \in \mathbf{N}^*}$ converge en loi vers une variable aléatoire suivant une loi normale centrée dont on précisera la variance.

Partie IV. Régression logistique

- Dans toute cette partie, p désigne un entier supérieur ou égal à 2.
- Pour tout couple $(n, m) \in (\mathbf{N}^*)^2$, on note $\mathcal{M}_{n,m}(\mathbf{R})$ l'ensemble des matrices à n lignes et m colonnes à coefficients réels et tM la transposée de toute matrice $M \in \mathcal{M}_{n,m}(\mathbf{R})$.
- Pour tout $m \in \mathbf{N}^*$, le produit scalaire usuel de deux vecteurs u et v de $\mathbf{R}^m$ est noté $\langle u, v \rangle$. Si U et V sont les matrices colonnes représentant u et v dans la base canonique, le produit scalaire $\langle u, v \rangle$ est donc l'unique coefficient de la matrice tUV .
- On rappelle que les fonctions Λ et L ont été définies dans la partie I.

Dans cette partie, on note Y une variable aléatoire de Bernoulli, dite *variable endogène*, dont la loi dépend du niveau de p *facteurs exogènes*.

L'influence de ces facteurs sur la loi de Y est résumée par la fonction b qui associe à un vecteur $x \in \mathbf{R}^p$, la probabilité $b(x)$ que Y soit égale à 1 lorsque les niveaux des facteurs sont donnés par les composantes du vecteur x .

Dans le *modèle de régression logistique* envisagé dans cette partie, la fonction b est supposée de la forme :

$$b : x \mapsto \Lambda(\langle \alpha, x \rangle)$$

où $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)$ est un vecteur de $\mathbf{R}^p$ dont les composantes $\alpha_1, \alpha_2, \dots, \alpha_p$ sont des paramètres inconnus qui représentent les degrés d'influence des divers facteurs exogènes sur la variable endogène Y .

Pour estimer les paramètres du modèle, on dispose de k vecteurs $x^{(1)}, x^{(2)}, \dots, x^{(k)}$ de $\mathbf{R}^p$ ($k \in \mathbf{N}^*$) et pour tout $i \in \llbracket 1, k \rrbracket$, d'une suite $(Y_{i,n})_{n \in \mathbf{N}^*}$ de variables aléatoires indépendantes suivant chacune la loi de Bernoulli de paramètre $b(x^{(i)}) = \Lambda(\langle \alpha, x^{(i)} \rangle)$.

Pour chaque indice fixé i et pour tout $n \in \mathbf{N}^*$, les variables aléatoires $Y_{i,1}, Y_{i,2}, \dots, Y_{i,n}$ définissent donc un n -échantillon associé à la loi de la variable endogène lorsque les niveaux des facteurs exogènes sont les composantes $x_1^{(i)}, x_2^{(i)}, \dots, x_p^{(i)}$ du vecteur $x^{(i)}$ dans la base canonique de $\mathbf{R}^p$.

15. On note respectivement A et M la matrice du vecteur α et la matrice de la famille $(x^{(1)}, x^{(2)}, \dots, x^{(k)})$ dans la base canonique de $\mathbf{R}^p$:

$$A = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} \in \mathcal{M}_{p,1}(\mathbf{R}) \quad \text{et} \quad M = \begin{pmatrix} x_1^{(1)} & \dots & x_1^{(k)} \\ \vdots & & \vdots \\ x_p^{(1)} & \dots & x_p^{(k)} \end{pmatrix} \in \mathcal{M}_{p,k}(\mathbf{R}).$$

On suppose que le rang de la matrice M est égal à p .

a) Montrer que la matrice $M^t M$ est inversible.

b) Montrer que pour toute matrice $H \in \mathcal{M}_{k,1}(\mathbf{R})$, la matrice $U \in \mathcal{M}_{p,1}(\mathbf{R})$ pour laquelle l'unique coefficient de la matrice ${}^t(MU - H)(MU - H)$ est le plus petit possible, est la matrice $(M^t M)^{-1} M H$.

c) Expliquer pourquoi les lois des variables aléatoires $Y_{i,n}$ ne suffiraient pas à définir le vecteur α si le rang de M n'était pas égal à p .

16. Pour tout $n \in \mathbf{N}^*$ et tout $i \in \llbracket 1, k \rrbracket$, on pose $\bar{Y}_{i,n} = \frac{1}{n} \sum_{j=1}^n Y_{i,j}$ et pour tout $\omega \in \Omega$:

$$T_{i,n} = \begin{cases} L(\bar{Y}_{i,n}(\omega)) & \text{si } 0 < \bar{Y}_{i,n}(\omega) < 1 \\ 0 & \text{sinon} \end{cases}.$$

a) Soit $(c_1, c_2, \dots, c_k) \in \mathbf{R}^k$. En utilisant les résultats de la partie III, montrer que $\left(\sum_{i=1}^k c_i T_{i,n} \right)_{n \in \mathbf{N}^*}$ est une

suite convergente d'estimateurs du paramètre $\sum_{i=1}^k c_i \langle \alpha, x^{(i)} \rangle$.

b) Pour tout $n \in \mathbf{N}^*$ et tout $\omega \in \Omega$, on pose, $H_n(\omega) = \begin{pmatrix} T_{1,n}(\omega) \\ T_{2,n}(\omega) \\ \vdots \\ T_{k,n}(\omega) \end{pmatrix}$ et $\begin{pmatrix} A_{1,n}(\omega) \\ A_{2,n}(\omega) \\ \vdots \\ A_{p,n}(\omega) \end{pmatrix} = (M^t M)^{-1} M H_n(\omega)$.

Montrer que pour tout $j \in \llbracket 1, p \rrbracket$, la suite $(A_{j,n})_{n \in \mathbf{N}^*}$ est une suite convergente d'estimateurs de α_j .

FIN

Conception : HEC Paris – ESCP BS

OPTION SCIENTIFIQUE

MATHÉMATIQUES II

Mercredi 29 avril 2020, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à **encadrer** dans la mesure du possible les résultats de leurs calculs.

Aucun document n'est autorisé. **L'utilisation de toute calculatrice et de tout matériel électronique est interdite.** Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

Lorsque l'on cherche à estimer un paramètre inconnu à partir d'un échantillon de données, on appelle statistique exhaustive toute fonction de ces données qui résume à elle seule l'information que ces données fournissent sur le paramètre.

On donne ici une définition précise de cette notion d'exhaustivité dans le cas des échantillons de variables aléatoires discrètes, illustrée de plusieurs exemples qui en montrent l'intérêt.

On s'intéressera dans ce problème à l'estimation d'un paramètre réel inconnu θ appartenant à un intervalle Θ .

On dispose pour cela de plusieurs observations $x_1, \dots, x_n$ considérées comme les réalisations de variables aléatoires discrètes $X_1, \dots, X_n$ définies sur le même espace probabilisable $(\Omega, \mathcal{A})$, à valeurs dans une partie B de $\mathbb{N}$.

L'espace probabilisable $(\Omega, \mathcal{A})$ est muni d'une famille $(P^\theta)_{\theta \in \Theta}$ de probabilités indexées par le paramètre θ .

On fait, pour toutes les valeurs du paramètre θ , les trois hypothèses suivantes.

- Les variables aléatoires $X_1, \dots, X_n$ sont mutuellement indépendantes, c'est-à-dire :

$$\forall (x_1, \dots, x_n) \in B^n, \quad P^\theta \left(\bigcap_{i=1}^n [X_i = x_i] \right) = \prod_{i=1}^n P^\theta ([X_i = x_i]) \quad (1)$$

- Les variables aléatoires $X_1, \dots, X_n$ suivent toutes la même loi qu'une variable aléatoire de référence, notée X , à valeurs dans B , c'est-à-dire :

$$\forall i \in [1, n], \quad \forall x \in B, \quad P^\theta ([X_i = x]) = P^\theta ([X = x]) \quad (2)$$

- Tous les éléments de B sont des valeurs effectivement possibles de X , c'est-à-dire :

$$\forall x \in B, \quad P^\theta([X = x]) > 0 \quad (3)$$

On appelle *statistique* toute variable aléatoire S de la forme $\omega \mapsto s(X_1(\omega), \dots, X_n(\omega))$, où s désigne une application définie sur B^n et à valeurs réelles. On note alors $S = s(X_1, \dots, X_n)$.

Pour tout $\theta \in \Theta$, on note $E^\theta(S)$ l'espérance de S lorsque $(\Omega, \mathcal{A})$ est muni de la probabilité P^θ (si cette espérance existe). On note de même $V^\theta(S)$ la variance de S (si elle existe).

Partie 1 : développements en série

1. Dans cette question, x désigne un nombre réel strictement compris entre 0 et 1.

a) Justifier la convergence de la série $\sum_{k \geq 1} \frac{x^k}{k}$.

b) Vérifier, pour tout $m \in \mathbb{N}^*$ et tout $t \in]0, 1[$, l'égalité :

$$\frac{1}{1-t} = \frac{t^m}{1-t} + \sum_{k=0}^{m-1} t^k.$$

c) Démontrer que l'intégrale $\int_0^x \frac{t^m}{1-t} dt$ tend vers 0 quand l'entier m tend vers l'infini.

d) En déduire la somme de la série $\sum_{k \geq 1} \frac{x^k}{k}$.

2. Dans cette question, indépendante de la précédente, $(a_k)_{k \in \mathbb{N}}$ désigne une suite de nombres réels telle que la série $\sum_{k \geq 0} a_k c^k$ est absolument convergente pour un réel strictement positif c .

a) Justifier que la fonction $f : x \mapsto a_0 + \sum_{k=1}^{+\infty} a_k x^k$ est bien définie sur le segment $[-c, +c]$.

b) Pour un entier naturel m , on pose : $M_m = \sum_{k=m+1}^{+\infty} |a_k| c^{k-m-1}$.

Justifier, pour tout $x \in [-c, +c]$, l'inégalité :

$$\left| \sum_{k=m+1}^{+\infty} a_k x^k \right| \leq M_m |x|^{m+1}.$$

c) Justifier, pour tout $m \in \mathbb{N}^*$, le développement limité au voisinage de 0 :

$$f(x) = a_0 + \sum_{k=1}^m a_k x^k + o(x^m).$$

d) Démontrer que si la fonction f est nulle sur l'intervalle $]0, +c]$, alors $(a_k)_{k \in \mathbb{N}}$ est la suite nulle.

Dans toute la suite du problème, pour tout $\theta \in \Theta$ et tout $(x_1, \dots, x_n) \in B^n$, on note :

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n P^\theta([X_i = x_i]) \quad (4)$$

Cette quantité, qui s'écrit aussi $\prod_{i=1}^n P^\theta([X = x_i])$ d'après (2), est appelée la *vraisemblance* de la valeur θ du paramètre au vu des observations $x_1, \dots, x_n$.

Partie II : estimateur du maximum de vraisemblance, un exemple

Dans cette partie, Θ est l'intervalle ouvert $]0, 1[$, B est égal à $\mathbb{N}^*$ et on a :

$$\forall x \in B, \quad P^\theta([X = x]) = (1 - \theta)^{x-1} \theta.$$

On note $\bar{X}$ la variable aléatoire $\frac{1}{n} \sum_{i=1}^n X_i$.

3. Soit $\theta \in \Theta$.

- Reconnaître la loi de X lorsque $(\Omega, \mathcal{A})$ est muni de la probabilité P^θ .
- En déduire que $\bar{X}$ est un estimateur sans biais du paramètre $1/\theta$.
- Quel est le risque quadratique de cet estimateur ?

4. On note T la variable aléatoire $\frac{1}{n} \sum_{i=1}^n \frac{1}{X_i}$.

- En utilisant le résultat de la question 1.d, justifier que :

$$\forall \theta \in \Theta, \quad E^\theta(T) = \frac{\theta \ln(\theta)}{\theta - 1}.$$

- En déduire que T est un estimateur de θ dont le biais $b_\theta(T)$ est strictement positif.

5. Soit $(x_1, \dots, x_n) \in B^n$.

- Justifier, pour tout $\theta \in \Theta$, l'égalité :

$$\ln(L(x_1, \dots, x_n, \theta)) = n \ln(\theta) - \left(n - \sum_{i=1}^n x_i\right) \ln(1 - \theta).$$

- En déduire que, lorsque les x_i ne sont pas tous égaux à 1, le nombre $\frac{n}{\sum_{i=1}^n x_i}$ est l'unique

valeur de θ qui maximise la vraisemblance $L(x_1, \dots, x_n, \theta)$.

6. On note U la variable aléatoire $\frac{n}{\sum_{i=1}^n X_i}$.

- Établir, pour tout $\theta \in \Theta$ et tout entier $k \geq n$, l'égalité :

$$\frac{n}{k} = \theta - \theta^2 \left(\frac{k}{n} - \frac{1}{\theta}\right) + \int_{1/\theta}^{k/n} \left(\frac{k}{n} - t\right) \frac{2}{t^3} dt.$$

- En déduire que U est un estimateur de θ dont le biais $b_\theta(U)$ est donné par :

$$\forall \theta \in \Theta, \quad b_\theta(U) = \sum_{k=n}^{+\infty} P\left(\sum_{i=1}^n X_i = k\right) \int_{1/\theta}^{k/n} \left(\frac{k}{n} - t\right) \frac{2}{t^3} dt.$$

- Justifier que $b_\theta(U)$ est strictement positif, quelle que soit la valeur du paramètre θ .

7. Dans cette question, on suppose que le nombre des observations est illimité. On dispose donc, pour estimer le paramètre θ , d'une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires mutuellement indépendantes et de même loi.

Pour tout entier $n \in \mathbb{N}^*$, on note $T_n = \frac{1}{n} \sum_{i=1}^n \frac{1}{X_i}$ et $U_n = \frac{n}{\sum_{i=1}^n X_i}$.

Étudier la convergence des deux suites d'estimateurs $(T_n)_{n \in \mathbb{N}^*}$ et $(U_n)_{n \in \mathbb{N}^*}$ du paramètre θ .

Dans toute la suite du problème, on dit qu'une statistique $S = s(X_1, \dots, X_n)$ est *exhaustive* s'il existe une application g de $s(B^n) \times \Theta$ dans $\mathbb{R}_+$ et une application h de B^n dans $\mathbb{R}_+$ telles que :

$$\forall \theta \in \Theta, \forall (x_1, \dots, x_n) \in B^n, L(x_1, \dots, x_n, \theta) = g(s(x_1, \dots, x_n), \theta) h(x_1, \dots, x_n) \quad (5)$$

Partie III : statistique exhaustive, un exemple

Dans cette partie, on suppose que $B = \{0, 1\}$, $\Theta =]0, 1[$ et que, quel que soit $\theta \in \Theta$, les variables aléatoires $X_1, \dots, X_n$ suivent la loi de Bernoulli de paramètre θ , lorsque l'espace probabilisable $(\Omega, \mathcal{A})$ est muni de la probabilité P^θ .

On pose : $S = \sum_{i=1}^n X_i$.

8. a) Démontrer que la vraisemblance de n'importe quelle valeur $\theta \in \Theta$ du paramètre est donnée par :

$$\forall (x_1, \dots, x_n) \in \{0, 1\}^n, L(x_1, \dots, x_n, \theta) = \theta^{\left(\sum_{i=1}^n x_i\right)} \times (1 - \theta)^{\left(\sum_{i=1}^n (1 - x_i)\right)}.$$

b) En déduire que la statistique S est exhaustive.

9. Soit $k \in \llbracket 0, n \rrbracket$ et $(x_1, \dots, x_n) \in \{0, 1\}^n$.

a) Calculer la probabilité conditionnelle $P_{[S=k]}^\theta([X_1 = x_1] \cap \dots \cap [X_n = x_n])$ et vérifier que la loi conditionnelle du vecteur aléatoire $(X_1, \dots, X_n)$ sachant l'événement $[S = k]$ ne dépend pas du paramètre θ .

b) Établir, pour tout $\theta \in \Theta$, l'égalité : $P_{[S=k]}^\theta([X_1 = 1]) = \frac{k}{n}$.

10. Le script *Scilab* suivant permet d'effectuer des simulations, qu'il place dans une matrice Y, dont il évalue ensuite la moyenne de chaque colonne.

```
--> theta=0.3;
--> N=100000;
--> n=10;
--> k=4;

--> U=grand(n,N,'bin',1,theta);
--> S=sum(U,'r'); // somme des lignes de U, colonne par colonne
--> K=find(S==k); // recherche des coefficients de S égaux à k
```

```
--> Y=U(1:n,K);

--> M=mean(Y,'c') // moyenne des colonnes de Y, ligne par ligne
ans =
    0.4019917
    0.4042436
    0.4008908
    0.3962868
    0.4054947
    0.3953861
    0.3990892
    0.4002402
    0.3941851
    0.4021919
```

- a) Décrire avec précision ce que représente une colonne de la matrice U.
- b) Expliquer pourquoi les coefficients de Y fournissent une simulation d'une loi conditionnelle du vecteur $(X_1, \dots, X_n)$.
- c) Commenter les résultats trouvés pour les coefficients de M.

11. À la suite du script précédent, on exécute l'instruction suivante :

```
--> C=Y*Y'/length(K);
```

- a) Donner le format de la matrice C et indiquer la valeur de son coefficient C(1,1).
- b) À quelle valeur approchée peut-on s'attendre pour C(1,2) et pour les autres coefficients non diagonaux de la matrice C ?
- c) Quelle est la somme totale des coefficients de la matrice C ?

Partie IV : inégalité de Rao-Blackwell

Dans cette partie, on reprend les hypothèses générales du préambule et on considère une statistique exhaustive $S = s(X_1, \dots, X_n)$, au sens donné par (5).

On admet que, pour tout élément u de $s(B^n)$ et tout élément $(x_1, \dots, x_n)$ de B^n , la probabilité conditionnelle $P_{[S=u]}^\theta([X_1 = x_1] \cap \dots \cap [X_n = x_n])$ ne dépend pas de θ .

12. Soit T un estimateur sans biais du paramètre θ .

- a) Démontrer que, pour tout $u \in s(B^n)$, l'espérance conditionnelle $E_{[S=u]}^\theta(T)$ existe et que sa valeur ne dépend pas de θ .
- b) Justifier que $([S = u])_{u \in s(B^n)}$ est un système complet d'événements.

13. Comme l'espérance conditionnelle $E_{[S=u]}^\theta(T)$ ne dépend pas de la valeur de θ , on peut la noter $E_{[S=u]}(T)$ et définir une application r de B^n dans $\mathbb{R}$ par :

$$\forall (x_1, \dots, x_n) \in B^n, \quad r(x_1, \dots, x_n) = E_{[S=s(x_1, \dots, x_n)]}(T) .$$

- a) En utilisant la formule de l'espérance totale, démontrer que $R = r(X_1, \dots, X_n)$ est un estimateur sans biais de θ .

b) On suppose que T admet une variance, quelle que soit la valeur du paramètre θ . Justifier qu'il en est de même pour R et en utilisant les inégalités

$$(E_{[S=u]}(T - \theta))^2 \leq E_{[S=u]}((T - \theta)^2)$$

établir, pour tout $\theta \in \Theta$, l'inégalité (appelée inégalité de Rao-Blackwell) :

$$V^\theta(R) \leq V^\theta(T) .$$

14. Un exemple d'estimateur sans biais optimal

Dans cette question uniquement, on suppose que $B = \mathbb{N}$, $\Theta =]0, +\infty[$ et que, pour tout $\theta \in \Theta$, la loi commune des variables aléatoires $X_1, \dots, X_n$ sur l'espace probabilisé $(\Omega, \mathcal{A}, P^\theta)$ est la loi de Poisson de paramètre θ .

a) Justifier que la statistique $S = \sum_{i=1}^n X_i$ est exhaustive.

b) Soit $u \in \mathbb{N}$ et $(x_1, \dots, x_n) \in \mathbb{N}^n$.

Vérifier que la probabilité conditionnelle $P_{[S=u]}^\theta([X_1 = x_1] \cap \dots \cap [X_n = x_n])$ ne dépend pas de θ .

c) Soit $u \in \mathbb{N}$.

Démontrer que chacune des variables aléatoires $X_1, \dots, X_n$ suit une loi binomiale lorsque l'espace probabilisable $(\Omega, \mathcal{A})$ est muni de la probabilité $P_{[S=u]}^\theta$. Sont-elles indépendantes pour cette probabilité?

d) Trouver une suite réelle $(\varphi_k)_{k \in \mathbb{N}}$ telle que

$$\forall \theta > 0, \quad \sum_{k=0}^{+\infty} \varphi_k \frac{(n\theta)^k}{k!} = \theta e^{n\theta}$$

et en prouver l'unicité à l'aide du résultat de la question 2.

e) En exploitant le résultat de la question 13, démontrer que, parmi les estimateurs sans biais de θ , l'estimateur $\frac{1}{n} \sum_{i=1}^n X_i$ est optimal, c'est-à-dire que son risque quadratique est inférieur ou égal à celui de tout autre estimateur sans biais de θ .

Conception : HEC Paris – ESCP BS

OPTION SCIENTIFIQUE

MATHÉMATIQUES II

Jeudi 29 avril 2021, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à **encadrer** dans la mesure du possible les résultats de leurs calculs.

Aucun document n'est autorisé. **L'utilisation de toute calculatrice et de tout matériel électronique est interdite.** Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

Partie 1 – Polynômes factoriels

On note F , l'espace vectoriel des fonctions polynomiales à coefficients réels et, pour tout entier naturel r , on note F_r , le sous-espace vectoriel des fonctions polynomiales à coefficients réels de degré inférieur ou égal à r .

On note U_k la fonction $x \mapsto x^k$ avec la convention habituelle $x^0 = 1$, de telle sorte que la base canonique de F_r est notée $(U_0, U_1, \dots, U_r)$.

1) Soit r , un entier naturel. On considère une famille $(Q_0, Q_1, \dots, Q_r)$ de fonctions polynomiales de degrés respectifs $d_0, d_1, \dots, d_r$ avec $d_0 < d_1 < \dots < d_r$.

1.a) On suppose qu'il existe des réels $\lambda_0, \lambda_1, \dots, \lambda_r$ non tous nuls tels que

$$\lambda_0 Q_0 + \lambda_1 Q_1 + \dots + \lambda_r Q_r = 0.$$

En considérant $m = \max\{k \in \llbracket 0, r \rrbracket : \lambda_k \neq 0\}$, démontrer que l'hypothèse précédente est absurde.

Qu'a-t-on ainsi démontré ?

1.b) À quelle condition la famille $(Q_0, Q_1, \dots, Q_r)$ est-elle une base de F_r ? (On précisera s'il s'agit d'une condition *nécessaire*, d'une condition *suffisante* ou d'une condition *nécessaire et suffisante*.)

2) Pour tout entier naturel r , le réel " x puissance r descendante" est noté $x^{\downarrow r}$ et défini par

$$\forall x \in \mathbb{R}, \quad x^{\downarrow r} = x(x-1) \times \dots \times (x-r+1) = \prod_{k=0}^{r-1} (x-k)$$

avec la convention $x^{\downarrow 0} = 1$. On pose alors

$$\forall r \in \mathbb{N}, \quad V_r : x \mapsto x^{\downarrow r}.$$

Il est clair que V_r appartient à F_r .

2.a) Quelles sont les racines de V_r ?

2.b) Démontrer que la famille $(V_0, V_1, \dots, V_r)$ est une base de F_r .

2.c) Démontrer que, pour tout entier $r \geq 2$,

$$x^r \underset{x \rightarrow +\infty}{=} x^r - \frac{r(r-1)}{2} x^{r-1} + o(x^{r-1}).$$

(On pourra raisonner par récurrence sur r .)

3) Ici, l'entier $r \geq 1$ est fixé et on compare la famille $(V_k)_{0 \leq k \leq r}$ à la base canonique $(U_k)_{0 \leq k \leq r}$ de l'espace F_r des fonctions polynomiales de degré inférieur ou égal à r .

3.a) Démontrer qu'il existe une unique famille $(\sigma(r, k))_{0 \leq k \leq r}$ de nombres réels tels que

$$U_r = \sum_{k=0}^r \sigma(r, k) V_k.$$

3.b) Établir les relations suivantes :

$$\forall r \in \mathbb{N}^*, \quad \sigma(r, 0) = 0 \quad (1)$$

$$\sigma(r, 1) = \sigma(r, r) = 1 \quad (2)$$

$$\sigma(r, r-1) = \frac{r(r-1)}{2} \quad (3)$$

$$\forall r \in \llbracket 2, +\infty \rrbracket, \quad \sigma(r, 2) = 2^{r-1} - 1 \quad (4)$$

3.c) Démontrer que, pour tout $r \in \mathbb{N}^*$ et pour tout $k \in \llbracket 1, r \rrbracket$, on a

$$\sigma(r+1, k) = \sigma(r, k-1) + k \sigma(r, k).$$

3.d) En déduire que $\sigma(r, k)$ est un entier naturel non nul pour tout $r \in \mathbb{N}^*$ et pour tout $k \in \llbracket 1, r \rrbracket$.

3.e) Écrire un code scilab qui affiche (au moyen de la commande `disp`) les listes $(\sigma(r, k))_{1 \leq k \leq r}$ pour r variant de 2 à 5.

On pourra utiliser la commande `ones(n,p)` qui retourne la matrice de $\mathfrak{M}_{n,p}(\mathbb{R})$ dont tous les coefficients sont égaux à 1.

4.a) Démontrer que, pour tout entier naturel r , il existe une unique famille $(s(r, k))_{0 \leq k \leq r}$ de nombres réels tels que

$$\forall x \in \mathbb{R}, \quad x^r = \sum_{k=0}^r s(r, k) x^k.$$

4.b) Soit $r \in \mathbb{N}^*$. Démontrer que

$$\forall k \in \llbracket 1, r \rrbracket, \quad s(r+1, k) = s(r, k-1) - r s(r, k).$$

En déduire la valeur de $s(r, 1)$.

4.c) Déduire de 4.b) que, pour tout $r \in \mathbb{N}^*$ et pour tout $k \in \llbracket 1, r \rrbracket$, le signe de $s(r, k)$ est celui de $(-1)^{r+k}$.

4.d) Démontrer que $\sigma(r, r) s(r, r) = 1$ et que

$$\forall \ell \in \llbracket 0, r-1 \rrbracket, \quad \sum_{k=\ell}^r \sigma(r, k) s(k, \ell) = 0.$$

4.e) Calculer $s(r, r)$ pour tout $r \in \mathbb{N}^*$ et $s(r, r-1)$ pour tout entier $r \geq 2$.

Partie 2 – Quelques propriétés de la loi de Poisson

Sous réserve d'existence, on note respectivement $\mathbf{E}(A)$ et $\mathbf{V}(A)$, l'espérance et la variance d'une variable aléatoire A et $\mathbf{Cov}(A, B)$, la covariance de deux variables aléatoires discrètes A et B .

Dans cette partie, on note X , une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ qui suit la loi de Poisson de paramètre $\theta > 0$.

Pour tout entier $r \geq 1$, on pose $X^r = X(X-1) \times \cdots \times (X-r+1)$ avec la convention $X^0 = X^0 = 1$.

Avec la suite $(\sigma(r, k))_{0 \leq k \leq r}$ définie au 3.a) et la suite $(s(r, k))_{0 \leq k \leq r}$ définie au 4.a), on a

$$\forall r \in \llbracket 1, +\infty \rrbracket, \quad X^r = \sum_{k=0}^r \sigma(r, k) X^k \quad \text{et} \quad X^r = \sum_{k=0}^r s(r, k) X^k.$$

On admet ces deux résultats sans démonstration.

- 5.a) Rappeler sans démonstration les valeurs respectives de $\mathbf{E}(X)$, $\mathbf{V}(X)$ et $\mathbf{E}(X^2)$.
- 5.b) Exprimer X , X^2 , X^3 et X^4 en fonction des variables aléatoires X^1 , X^2 , X^3 et X^4 .
- 5.c) Démontrer que la variable aléatoire X admet des moments de tous ordres.
- 6.a) Justifier que, pour tout entier $r \geq 1$, la variable aléatoire X^r admet des moments de tous ordres.
- 6.b) Pour tout entier $r \geq 1$, exprimer $\mathbf{E}(X^r)$ en fonction de r et de θ .
- 6.c) Calculer $\mathbf{E}(X^3)$ et $\mathbf{E}(X^4)$ en fonction de θ .
- 7) Pour tout entier naturel k et pour tout réel $\theta > 0$, on pose

$$f(\theta, k) = \frac{\theta^k}{k!} e^{-\theta} \quad \text{et} \quad g(\theta, k) = \ln(f(\theta, k)).$$

- 7.a) Pour tout entier $k \geq 0$, calculer l'expression de la dérivée partielle $\partial_1(g)(\theta, k)$ et exprimer la variable aléatoire $\partial_1(g)(\theta, X)$ en fonction de X et de θ .
- 7.b) Vérifier que $XX^r = X^{r+1} + rX^r$ pour tout entier $r \geq 1$. En déduire que $\mathbf{Cov}(X, X^r) = r\theta^r$.
- 7.c) Calculer $\mathbf{Cov}(\partial_1(g)(\theta, X), X^r)$ et en déduire l'inégalité

$$\forall \theta > 0, \forall r \in \mathbf{N}^*, \quad \mathbf{V}(X^r) \geq r^2 \theta^{2r-1}.$$

Partie 3 – Estimation ponctuelle de fonctions du paramètre θ

Le contexte et les notations sont ceux de la partie 2.

On suppose que le paramètre $\theta \in \mathbf{R}_+^*$ est inconnu et on cherche ici à estimer $\varphi(\theta)$, où φ est une fonction dérivable sur $\mathbf{R}_+^*$.

Pour n entier de $\mathbf{N}^*$, on considère dans toute cette partie un n -échantillon $(X_1, \dots, X_n)$ de variables aléatoires mutuellement indépendantes qui, comme X , suivent toutes la loi de Poisson de paramètre θ .

Pour tout $\theta > 0$ et pour tout $(k_1, \dots, k_n) \in \mathbf{N}^n$, on pose

$$F(\theta, k_1, \dots, k_n) = \mathbf{P}\left(\bigcap_{i=1}^n [X_i = k_i]\right) = \prod_{i=1}^n f(\theta, k_i)$$

et

$$G(\theta, k_1, \dots, k_n) = \ln(F(\theta, k_1, \dots, k_n)).$$

- 8.a) Démontrer que les fonctions $\theta \mapsto F(\theta, k_1, \dots, k_n)$ et $\theta \mapsto G(\theta, k_1, \dots, k_n)$ sont dérivables sur $\mathbf{R}_+^*$ et calculer les dérivées partielles $\partial_1(F)(\theta, k_1, \dots, k_n)$ et $\partial_1(G)(\theta, k_1, \dots, k_n)$.
- 8.b) Pour tout $\theta > 0$, on pose $Z_\theta = \partial_1(G)(\theta, X_1, \dots, X_n)$. Démontrer que la variable aléatoire Z_θ est centrée et admet une variance strictement positive, notée $I(\theta)$, que l'on calculera.

On rappelle que : s'il existe n séries absolument convergentes $\sum_{k_1 \in \mathbf{N}} v_{1,k_1}, \dots, \sum_{k_n \in \mathbf{N}} v_{n,k_n}$ telles que

$$\forall (k_1, \dots, k_n) \in \mathbf{N}^n, \quad |u_{k_1, \dots, k_n}| \leq |v_{1,k_1}| \times \dots \times |v_{n,k_n}|$$

alors la série $\sum u_{k_1, \dots, k_n}$ est dite absolument convergente. On admet que, dans ce cas, la somme

$$\sum_{(k_1, \dots, k_n) \in \mathbf{N}^n} u_{k_1, \dots, k_n}$$

est bien définie.

On rappelle l'énoncé de la Formule de transfert : si la série $\sum u_\theta(k_1, \dots, k_n) F(\theta, k_1, \dots, k_n)$ est absolument convergente (au sens qui vient d'être rappelé), alors la variable aléatoire discrète $U_\theta = u_\theta(X_1, \dots, X_n)$ est d'espérance finie et

$$\begin{aligned} \mathbf{E}(U_\theta) &= \sum_{(k_1, \dots, k_n) \in \mathbf{N}^n} u_\theta(k_1, \dots, k_n) F(\theta, k_1, \dots, k_n) \\ &= \sum_{(k_1, \dots, k_n) \in \mathbf{N}^n} u_\theta(k_1, \dots, k_n) \mathbf{P}([X_1 = k_1] \cap \dots \cap [X_n = k_n]). \end{aligned}$$

Soit $t : \mathbf{N}^n \rightarrow \mathbf{R}$, une application indépendante de θ . On peut alors considérer la variable aléatoire discrète

$$T = t(X_1, \dots, X_n)$$

comme un estimateur de $\varphi(\theta)$.

On dira que la variable aléatoire T est un estimateur régulier de $\varphi(\theta)$ lorsque les trois conditions suivantes (R_1) , (R_2) et (R_3) sont satisfaites.

$$\mathbf{E}(T) = \varphi(\theta) \quad (R_1)$$

$$\mathbf{V}(T) \text{ existe} \quad (R_2)$$

$$\forall \theta > 0, \quad \varphi'(\theta) = \sum_{(k_1, \dots, k_n) \in \mathbf{N}^n} t(k_1, \dots, k_n) \partial_1(F)(\theta, k_1, \dots, k_n) \quad (R_3)$$

On notera que la condition (R_3) sous-entend que le second membre est la somme d'une série absolument convergente (au sens rappelé plus haut).

9) Dans cette question, on suppose que T est un estimateur régulier de $\varphi(\theta)$.

9.a) La variable aléatoire T est-elle un estimateur sans biais de $\varphi(\theta)$?

9.b) Pourquoi la condition (R_3) n'est-elle pas une conséquence directe de la condition (R_1) ?

10) Soit T , un estimateur régulier du paramètre $\varphi(\theta)$.

10.a) Établir les égalités suivantes :

$$\forall \theta > 0, \quad \varphi'(\theta) = \mathbf{E}(T \times Z_\theta) = \mathbf{Cov}(T, Z_\theta).$$

10.b) En déduire l'inégalité

$$\forall \theta > 0, \quad \mathbf{V}(T) \geq \frac{(\varphi'(\theta))^2}{I(\theta)}$$

où $I(\theta)$ a été défini au 8.b).

11) On cherche à simuler un échantillon de N réalisations de Z_θ pour différents couples (n, θ) .

11.a) Compléter le code scilab suivant en justifiant votre réponse.

```
function ech=Z_th(n, theta)
    X=grand(n,N,'poi',theta);
    ech = (sum(X,      ) - n*theta)/theta
endfunction
```

On rappelle l'usage de la commande `sum` : pour un tableau $M = (M_{i,j})_{1 \leq i \leq n, 1 \leq j \leq p}$, les deux instructions `sum(M, 'r')` et `sum(M, 'c')` retournent respectivement les tableaux

$$\left(\sum_{i=1}^n M_{i,j} \right)_{1 \leq j \leq p} \quad \text{et} \quad \left(\sum_{j=1}^p M_{i,j} \right)_{1 \leq i \leq n}$$

de tailles (*size*) respectives $(1, p)$ et $(n, 1)$.

11.b) À l'aide de la commande `histplot`, on a tracé les histogrammes des échantillons obtenus pour les couples $(n, \theta) = (10, 4), (20, 4), (40, 4)$ et $(50, 5)$.

À quels couples correspondent les figures suivantes ? (On pourra admettre que $I(\theta) = n/\theta$.)

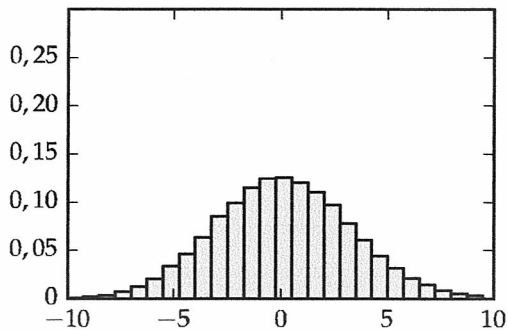


Figure A

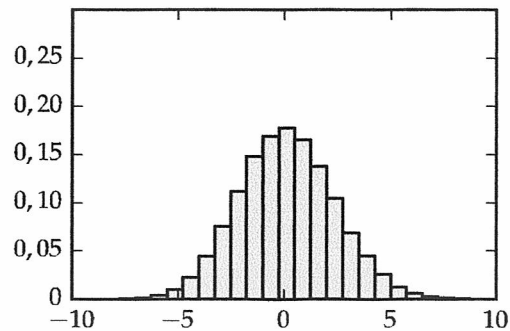


Figure B

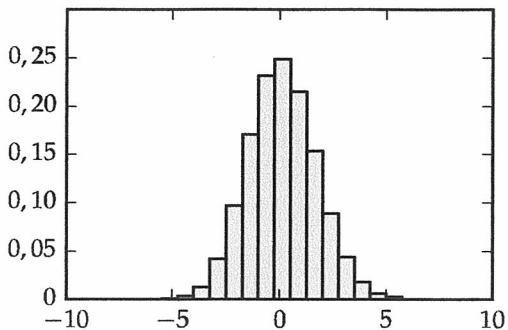


Figure C

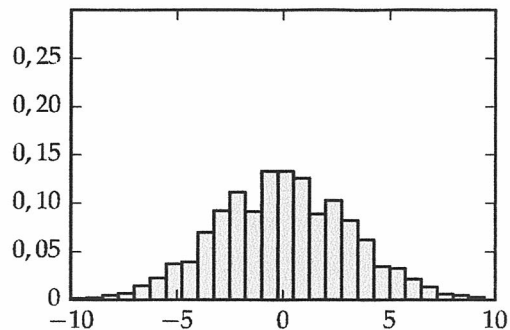


Figure D

12) Soit un entier $r \geq 1$. On suppose ici que $\varphi(\theta) = \theta^r$ et pour tout entier n supérieur ou égal à 1, on pose

$$S_n = X_1 + \dots + X_n \quad \text{et} \quad M_{r,n} = \frac{S_n(S_n - 1) \times \dots \times (S_n - r + 1)}{n^r}.$$

12.a) Rappeler (sans démonstration) la loi de la variable aléatoire S_n ainsi que son espérance et sa variance.

12.b) Démontrer que $M_{r,n}$ est un estimateur régulier de θ^r .

NB : Pour établir la propriété (R_3) , on *admettra* que la série est absolument convergente.

En déduire que

$$\forall \theta > 0, \quad \mathbf{V}(M_{r,n}) \geq \frac{r^2 \theta^{2r-1}}{n}.$$

12.c) Dans cette question, on suppose que $r = 2$. Calculer la variance de $M_{2,n}$ et démontrer que la suite d'estimateurs de θ^2

$$(M_{2,n})_{n \geq 1}$$

est convergente.

12.d) Pour un entier $r \geq 1$ quelconque, la suite

$$(M_{r,n})_{n \geq 1}$$

d'estimateurs de θ^r est-elle convergente ? (On pourra commencer par calculer, en fonction de l'entier $k \in \mathbb{N}^*$ un équivalent de $\mathbf{E}(S_n^k)$ lorsque n tend vers $+\infty$.)

Partie 4 – Le cas $\varphi(\theta) = \theta$

Le contexte et les notations sont ceux des parties 2 et 3.

Dans cette partie, on compare deux estimateurs du paramètre inconnu θ .

13) Pour tout entier n supérieur ou égal à 1, on pose

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

13.a) Démontrer que $\bar{X}_n$ est un estimateur régulier du paramètre θ .

13.b) Que devient l'inégalité du 10.b) ?

On dit qu'un estimateur régulier de θ est efficace lorsque sa variance est minimale parmi les estimateurs réguliers de θ .

14) Soit Y , un estimateur régulier de θ . Pour tout réel α , on pose

$$\psi(\alpha) = \bar{X}_n + \alpha(Y - \bar{X}_n).$$

14.a) Vérifier que $\psi(\alpha)$ est un estimateur régulier de θ pour tout $\alpha \in \mathbb{R}$.

14.b) En déduire que

$$\text{Cov}(\bar{X}_n, Y) = \frac{\theta}{n}.$$

14.c) Exprimer $\mathbf{V}(Y - \bar{X}_n)$ en fonction de $\mathbf{V}(Y)$ et de $\mathbf{V}(\bar{X}_n)$. En déduire qu'un estimateur efficace de θ est presque sûrement unique.

15) Pour tout entier $n \geq 2$, on pose

$$W_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

15.a) Exprimer W_n en fonction de $\sum_{i=1}^n X_i^2$ et de $\bar{X}_n^2$.

15.b) Démontrer que W_n est un estimateur sans biais de θ .

15.c) Démontrer que W_n admet une variance (qu'on ne cherchera pas à calculer).

15.d) Étudier la convergence des deux suites d'estimateurs $(\bar{X}_n)_{n \geq 1}$ et $(W_n)_{n \geq 2}$ du paramètre inconnu θ .

On pourra démontrer que : si une suite réelle $(a_n)_{n \in \mathbb{N}}$ converge vers $a \in \mathbb{R}$ et si deux suites $(Y_n)_{n \in \mathbb{N}}$ et $(Z_n)_{n \in \mathbb{N}}$ de variables aléatoires convergent en probabilité vers les réels y et z respectivement, alors la suite de variables aléatoires $(a_n(Y_n - Z_n))_{n \in \mathbb{N}}$ converge en probabilité vers le réel $a(y - z)$.

16) On simule des échantillons de N réalisations des estimateurs $\bar{X}_n, W_n$ et

$$W'_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

avec $n = 50$. En comparant les figures suivantes, relier chaque histogramme à l'estimateur qui lui correspond.

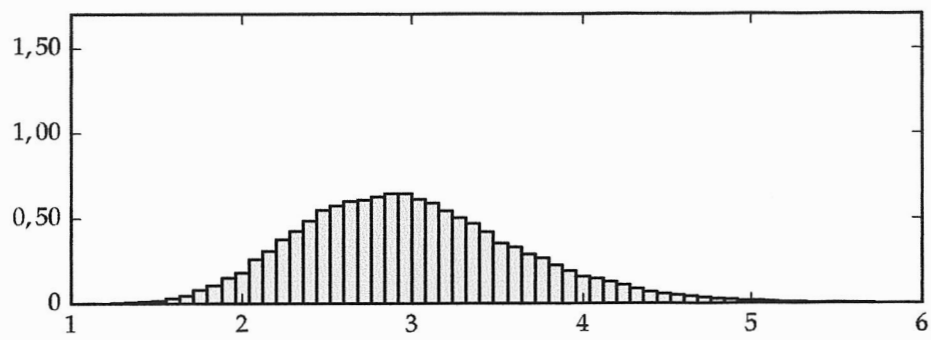


Figure E

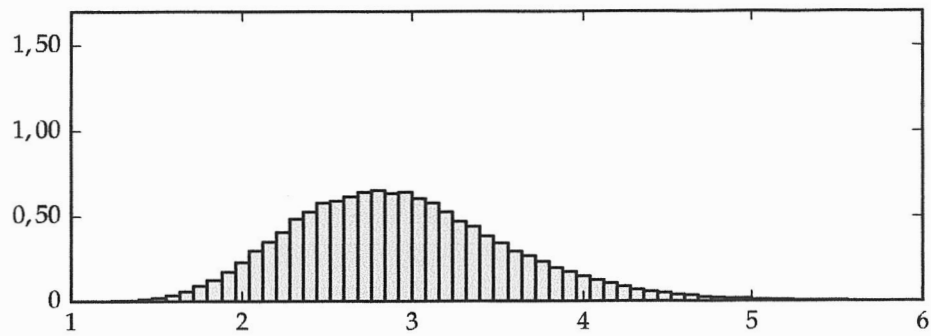


Figure F

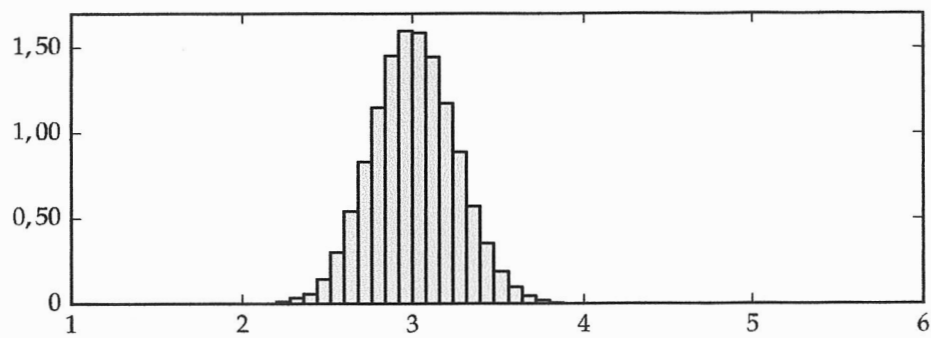


Figure G

Fin de l'énoncé

Conception : HEC Paris – ESCP BS

OPTION SCIENTIFIQUE

MATHÉMATIQUES II

Lundi 9 mai 2022, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

*Les candidats sont invités à **encadrer** dans la mesure du possible les résultats de leurs calculs.*

*Aucun document n'est autorisé. **L'utilisation de toute calculatrice et de tout matériel électronique est interdite.** Seule l'utilisation d'une règle graduée est autorisée.*

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

Dans ce problème, toutes les variables aléatoires sont supposées définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Si X est une variable aléatoire, on note respectivement $\mathbf{E}(X)$ et $\mathbf{Var}(X)$ son espérance et sa variance, sous réserve d'existence.

Le but de ce problème est de mettre en évidence quelques résultats asymptotiques liés au modèle du collectionneur de vignettes. Dans chaque paquet de céréales se trouve une vignette et il y a en tout des vignettes de n types différents, où n est un entier supérieur ou égal à 1. Chacun des n types de vignettes se retrouve avec la même fréquence dans les paquets de céréales. Une collection est alors complète lorsqu'elle comporte n vignettes de types différents.

On modélise le nombre total de paquets de céréales qu'il est nécessaire d'acheter pour obtenir la collection complète de n vignettes de types différents par la variable aléatoire notée C_n .

On pose par convention $C_0 = 0$ et pour tout $i \in \llbracket 1, n \rrbracket$, on note C_i le nombre d'achats de paquets de céréales nécessaires pour obtenir i vignettes de types différents.

De même, pour tout $i \in \llbracket 1, n \rrbracket$, on pose $X_i = C_i - C_{i-1}$, qui représente le nombre d'achats supplémentaires de paquets de céréales qu'il est nécessaire d'effectuer pour obtenir une nouvelle vignette d'un type différent des $(i - 1)$ vignettes de types différents déjà obtenues. Par convention, on pose $X_1 = C_1 = 1$.

On suppose que les variables aléatoires $X_1, \dots, X_n$ sont mutuellement indépendantes. Enfin, on pose :

$$V_n = \frac{C_n}{n} - \ln(n).$$

- Questions préliminaires -

1. (a) Montrer que :

$$C_n = X_1 + \cdots + X_n = \sum_{i=1}^n X_i.$$

- (b) Justifier que pour tout $i \in \llbracket 1, n \rrbracket$, la variable aléatoire X_i suit la loi géométrique $\mathcal{G}\left(\frac{n-i+1}{n}\right)$.

- Partie I -

Pour tout entier $n \geq 1$, on pose :

$$H_n = \sum_{k=1}^n \frac{1}{k} \quad , \quad S_n = \sum_{k=1}^n \frac{1}{k^2}.$$

Nous allons démontrer la convergence de la suite $(S_n)_{n \geq 1}$ et déterminer une valeur approchée de sa limite S .

2. (a) Montrer que pour tout entier $k \geq 2$, on a l'encadrement :

$$\frac{1}{k} - \frac{1}{k+1} \leq \frac{1}{k^2} \leq \frac{1}{k-1} - \frac{1}{k}.$$

- (b) Montrer que pour tout entier $n \geq 2$, on a :

$$\frac{3}{2} - \frac{1}{n+1} \leq S_n \leq 2 - \frac{1}{n}.$$

- (c) Montrer que la suite $(S_n)_{n \geq 1}$ est convergente et donner un encadrement de sa limite S .

- (d) Montrer que pour tout entier $n \geq 1$, on a l'encadrement :

$$\frac{1}{n+1} \leq S - S_n \leq \frac{1}{n}.$$

- (e) En déduire un programme Scilab qui permet d'obtenir une valeur approchée de S à 10^{-7} -près.

3. Pour tout entier $n \geq 1$, on pose $u_n = H_n - \ln(n)$.

- (a) Montrer que pour tout entier $n \geq 1$, on a l'encadrement :

$$\frac{1}{n+1} \leq \ln\left(\frac{n+1}{n}\right) \leq \frac{1}{n}.$$

- (b) Montrer que pour tout entier $n \geq 1$, on a : $0 \leq u_n \leq 1$.

- (c) Montrer que la suite $(u_n)_{n \geq 1}$ converge vers une certaine limite $\gamma \in [0, 1]$.

4. Justifier l'existence de l'espérance de C_n et montrer que $\mathbf{E}(C_n) = nH_n$.

5. Justifier l'existence de la variance de C_n et exprimer $\mathbf{Var}(C_n)$ en fonction de n, S_n et H_n .

6. (a) Montrer que pour tout réel $a > 0$, on a :

$$\mathbb{P}(|C_n - nH_n| \geq an) \leq \frac{S}{a^2}.$$

- (b) Montrer que pour tout réel $c > 1$, on a :

$$\mathbb{P}\left(\left|\frac{C_n}{n} - \ln(n)\right| \geq c\right) \leq \frac{S}{(c-1)^2}.$$

- (c) Pour cette question, on suppose que $n = 10^6$.
On donne l'approximation $\ln(n) \approx 13.816$.
Montrer que :

$$\mathbb{P}\left(\frac{C_n}{n} \in [7.81, 19.92]\right) \geq 0.92.$$

- Partie II -

7. Soient $T_1, \dots, T_n$, n variables aléatoires mutuellement indépendantes de même loi exponentielle $\mathcal{E}(1)$. On pose :

$$M_n = \max_{1 \leq i \leq n} T_i.$$

Montrer que M_n suit la loi de densité f_n donnée par :

$$\forall x \in \mathbb{R}, \quad f_n(x) = \begin{cases} ne^{-x}(1 - e^{-x})^{n-1} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

8. (a) Soit Z une variable aléatoire de loi exponentielle $\mathcal{E}(n+1)$ et indépendante de M_n . On note g la densité de la loi de Z qui est nulle sur $]-\infty, 0]$ et continue sur $]0, +\infty[$. Montrer que pour tout $x > 0$ et tout $t \in]0, x[$, on a :

$$f_n(t)g(x-t) = (n+1)e^{-(n+1)x}ne^t(e^t - 1)^{n-1}.$$

- (b) Soit $(Z_i)_{i \geq 1}$ une suite de variables aléatoires mutuellement indépendantes telle que pour tout entier $i \geq 1$, Z_i est de loi exponentielle $\mathcal{E}(i)$.

Montrer que pour tout entier $n \geq 1$, la variable aléatoire $\sum_{i=1}^n Z_i$ suit la loi de densité f_n .

9. On définit la fonction $f : \mathbb{R} \rightarrow \mathbb{R}_+$ par :

$$\forall x \in \mathbb{R}, \quad f(x) = e^{-x}e^{-e^{-x}}.$$

Montrer que f est une densité de probabilité sur $\mathbb{R}$. La loi de densité f est appelée loi de Gumbel.

10. (a) Soit $(Z_i)_{i \geq 1}$ la suite de variables aléatoires introduite précédemment. On pose $W_n = \sum_{i=1}^n Z_i - \ln(n)$.

Montrer que la fonction de répartition F_{W_n} de W_n est donnée par :

$$\forall x \in \mathbb{R}, \quad F_{W_n}(x) = \begin{cases} \left(1 - \frac{e^{-x}}{n}\right)^n & \text{si } x > -\ln(n), \\ 0 & \text{si } x \leq -\ln(n). \end{cases}$$

- (b) Montrer que la suite de variables aléatoires $(W_n)_{n \geq 1}$ converge en loi vers une variable aléatoire de loi de Gumbel.

11. (a) On rappelle qu'en Scilab l'instruction `grand(1,1,'geom',p)` permet la simulation d'une variable aléatoire suivant la loi géométrique de paramètre p .

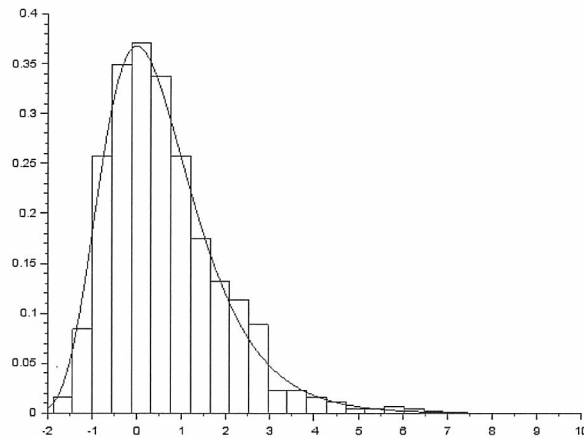
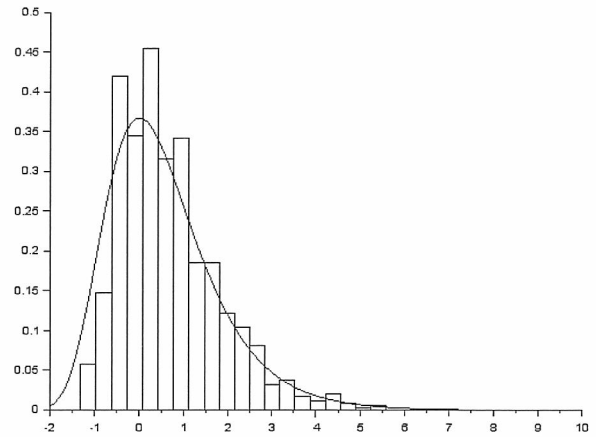
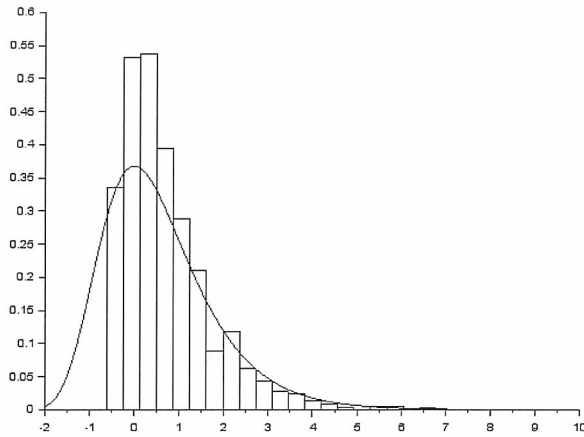
Écrire une fonction d'en-tête `function y=simulV(n)` qui pour un entier n fourni en entrée, renvoie une simulation de la variable aléatoire V_n définie en introduction de ce problème.

- (b) À la suite de la fonction `simulV`, écrire un programme Scilab qui construit un vecteur-ligne V contenant 1000 simulations indépendantes de la variable aléatoire V_n pour un certain entier n entré par l'utilisateur.

- (c) On complète ce programme par le code suivant :

```
histplot(20,V)
function y=f(x)
    y=exp(-x)*exp(-exp(-x))
endfunction
absc=linspace(-1,10,100)
fplot2d(absc,f)
```

On obtient les sorties graphiques suivantes en exécutant le programme pour $n = 5$, $n = 10$ puis $n = 50$:



Que peut-on observer sur ces figures ? Quelle conjecture peut-on en déduire pour la suite $(V_n)_{n \geq 1}$?

12. Soit $p \in]0, 1[$ et soit U une variable aléatoire de loi exponentielle $\mathcal{E}(p)$. On pose $\alpha = -\frac{p}{\ln(1-p)}$.

Montrer que la variable aléatoire $V = \lfloor \alpha U \rfloor + 1$ est de loi géométrique $\mathcal{G}(p)$.

On rappelle que $\lfloor t \rfloor$ désigne la partie entière du réel t .

13. (a) Montrer que la variable aléatoire $V - \alpha U$ est à valeurs dans $[0, 1]$. En déduire que $\mathbf{Var}(V - \alpha U) \leq 1$.
 (b) Soit X et Y deux variables aléatoires admettant un moment d'ordre 2. Justifier l'existence de la covariance $\mathbf{Cov}(X, Y)$ puis montrer que :

$$\mathbf{Var}(X + Y) \leq 2(\mathbf{Var}(X) + \mathbf{Var}(Y))$$

On pourra calculer $\mathbf{Var}(X + Y)$ et $\mathbf{Var}(X - Y)$.

- (c) En déduire que :

$$\mathbf{Var}(V - U) \leq 2 + 2 \frac{(1 - \alpha)^2}{p^2}.$$

14. Dans cette question, on suppose que $n \geq 2$.

Soient $Y_1, \dots, Y_n$, n variables aléatoires mutuellement indépendantes telles que pour tout $i \in \llbracket 1, n \rrbracket$, Y_i suit la loi exponentielle $\mathcal{E}\left(\frac{n-i+1}{n}\right)$.

On pose $\alpha_1 = 0$, $X'_1 = 1$ et pour tout $i \in \llbracket 2, n \rrbracket$, $\alpha_i = -\frac{n-i+1}{n} \times \frac{1}{\ln\left(\frac{i-1}{n}\right)}$ et $X'_i = \lfloor \alpha_i Y_i \rfloor + 1$.

De plus, on pose : $W'_n = \frac{Y_1 + \dots + Y_n}{n} - \ln(n)$ et $V'_n = \frac{X'_1 + \dots + X'_n}{n} - \ln(n)$.

(a) Montrer que :

$$\mathbf{Var}(V'_n - W'_n) \leq \frac{1}{n^2} + \frac{2}{n^2} \sum_{i=2}^n \left[\left(\frac{n}{n-i+1} \right)^2 (1 - \alpha_i)^2 + 1 \right].$$

(b) i. Montrer que la fonction $\phi :]0, 1[\rightarrow \mathbb{R}$ définie par :

$$\forall z \in]0, 1[, \phi(z) = \frac{1}{(1-z)^2} \left(1 + \frac{1-z}{\ln(z)} \right)^2$$

peut être prolongée en une fonction continue sur le segment $[0, 1]$.

ii. En déduire qu'il existe un réel $A > 0$ tel que :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=2}^n \left(\frac{n}{n-i+1} \right)^2 (1 - \alpha_i)^2 = A.$$

(c) Montrer que la suite de variables aléatoires $(V'_n - W'_n)_{n \geq 2}$ converge en probabilité vers 0.

(d) Montrer que la suite de variables aléatoires $(V'_n)_{n \geq 2}$ converge en loi vers une variable aléatoire de loi de Gumbel.

(e) On donne les valeurs numériques suivantes : $e^{-e^{-x}} \approx 0.96$ si $x = 3.20$ et $e^{-e^{-x}} \approx 0.04$ si $x = -1.17$.

Lorsque $n = 10^6$, montrer que :

$$\mathbb{P} \left(\frac{C_n}{n} \in [12.65, 17.02] \right) \approx 0.92.$$

Comparer avec le résultat de la première partie et commenter.

- Partie III -

Dans cette partie, on suppose que $n \geq 2$. On suppose de plus que les vignettes sont numérotées de 1 à n .

Pour tout $i \in \llbracket 1, n \rrbracket$ et tout entier $m \geq 1$, on note $A_{i,m}$ la variable aléatoire donnant le nombre de vignettes numérotées i obtenues dans les m premiers paquets de céréales achetés.

15. (a) Justifier que pour tout $i \in \llbracket 1, n \rrbracket$, la variable aléatoire $A_{i,m}$ suit la loi binomiale $\mathcal{B} \left(m, \frac{1}{n} \right)$.
- (b) Calculer la covariance $\text{Cov}(A_{1,m}, m - A_{1,m})$.
- (c) À l'aide d'un raisonnement par l'absurde, en déduire que les variables aléatoires $A_{1,m}, \dots, A_{n,m}$ ne sont pas mutuellement indépendantes.
16. (a) Montrer que si $(E_i)_{i \geq 1}$ est une famille d'événements, alors pour tout entier $r \geq 1$:

$$\mathbb{P} \left(\bigcup_{i=1}^r E_i \right) \leq \sum_{i=1}^r \mathbb{P}(E_i).$$

(b) Montrer que :

$$\mathbb{P}(C_n > m) \leq n \left(1 - \frac{1}{n} \right)^m \leq n e^{-\frac{m}{n}}.$$

(c) Soit c un réel strictement positif. Montrer qu'on a la majoration :

$$\mathbb{P}(C_n > c n \ln(n)) \leq n^{1-c}.$$

(d) Montrer que pour tout réel $x > -\ln(n)$, on a :

$$\mathbb{P}(V_n > x) \leq e^{-x}.$$

Dans la suite de cette partie, on introduit un modèle légèrement différent : le nombre N de paquets achetés est décrit par une variable aléatoire N de loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda > 0$. On cherche à calculer la probabilité de compléter, à partir des N vignettes obtenues, la collection de vignettes. On suppose toujours que les vignettes sont numérotées de 1 à n . On note $\tilde{A}_i$ la variable aléatoire donnant le nombre de vignettes numérotées i obtenues dans les N paquets de céréales achetés.

17. (a) Soit $p \in \mathbb{N}^*$. Justifier que pour tout $i \in \llbracket 1, n \rrbracket$, la loi conditionnelle du couple $(\tilde{A}_i, N)$ conditionnée par l'événement $[N = p]$ est la loi binomiale $\mathcal{B}\left(p, \frac{1}{n}\right)$ et en déduire que la variable aléatoire $\tilde{A}_i$ suit une loi de Poisson dont on déterminera le paramètre.

(b) Montrer que pour tout n -uplet $(k_1, \dots, k_n)$ avec $k_i \in \mathbb{N}$ et $k_1 + \dots + k_n = p$, on a :

$$\mathbb{P}_{[N=p]}([\tilde{A}_1 = k_1] \cap \dots \cap [\tilde{A}_n = k_n]) = \frac{p!}{k_1! \dots k_n!} \frac{1}{n^p}.$$

(c) Montrer que pour tout n -uplet $(k_1, \dots, k_n)$ avec $k_i \in \mathbb{N}$, on a :

$$\mathbb{P}([\tilde{A}_1 = k_1] \cap \dots \cap [\tilde{A}_n = k_n]) = \prod_{i=1}^n \frac{e^{-\frac{\lambda}{n}} \left(\frac{\lambda}{n}\right)^{k_i}}{k_i!}.$$

(d) Montrer que les variables aléatoires $\tilde{A}_1, \dots, \tilde{A}_n$ sont mutuellement indépendantes.
Commenter en comparant avec le résultat de la question 15.(c).

18. Soit D_n l'événement « à l'issue des N achats de paquets de céréales, la collection de vignettes est complète ». Montrer que :

$$\mathbb{P}(D_n) = \left(1 - e^{-\frac{\lambda}{n}}\right)^n.$$

On admet le résultat suivant : pour tout $p \in \mathbb{N}^*$, on a $\mathbb{P}_{[N=p]}(D_n) = \mathbb{P}(C_n \leq p)$.

19. (a) Montrer que :

$$\mathbb{P}(D_n) = \sum_{p=0}^{\infty} \mathbb{P}(C_n \leq p) \mathbb{P}(N = p).$$

(b) On suppose maintenant que $\lambda > 1$.

Soit $a \in]\sqrt{\lambda}, \lambda[$. On pose $k_1 = \lfloor \lambda - a \rfloor$ et $k_2 = \lfloor \lambda + a \rfloor + 1$.

Montrer que :

$$\sum_{p=k_1}^{k_2} \mathbb{P}(N = p) \geq 1 - \frac{\lambda}{a^2}.$$

(c) Montrer l'encadrement :

$$\mathbb{P}(C_n \leq k_1) \left(1 - \frac{\lambda}{a^2}\right) \leq \mathbb{P}(D_n) \leq \mathbb{P}(C_n \leq k_2) + \frac{\lambda}{a^2}.$$

20. (a) Soit $(c_n)_{n \geq 1}$ une suite convergente de nombres réels.

Montrer que pour tout entier n suffisamment grand, on a l'encadrement :

$$\mathbb{P}\left(V_n \leq c_n - \frac{1}{n^{1/3}} - \frac{1}{n}\right) \left(1 - \frac{\ln(n) + c_n}{n^{1/3}}\right) \leq \left(1 - \frac{e^{-c_n}}{n}\right)^n \leq \mathbb{P}\left(V_n \leq c_n + \frac{1}{n^{1/3}} + \frac{1}{n}\right) + \frac{\ln(n) + c_n}{n^{1/3}}.$$

On pourra appliquer la question précédente avec $\lambda = n \ln(n) + c_n n$ et $a = n^{2/3}$.

(b) Retrouver alors la convergence en loi de la suite $(V_n)_{n \geq 1}$ vers une variable aléatoire de loi de Gumbel.



Conception : ESCP BS – HEC Paris

MATHEMATIQUES 2 APPROFONDIES

FILIÈRE ÉCONOMIQUE ET COMMERCIALE

VOIE GENERALE

Mardi 2 mai 2023, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Aucun document n'est autorisé. L'utilisation de toute calculatrice et de tout matériel électronique est interdite. Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

Toutes les variables aléatoires considérées dans ce problème sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Questions préliminaires

Si $t \in \mathbb{R}$, on appelle partie entière de t que l'on note $[t]$, l'unique entier tel que $[t] \leq t < [t] + 1$.

1. Montrer que pour tout $x \in \mathbb{R}$, on a $\lim_{n \rightarrow +\infty} \frac{[nx]}{n} = x$.

2. Montrer que $\lim_{n \rightarrow +\infty} \left(n - \left\lfloor \frac{n}{2} \right\rfloor \right) = +\infty$.

Première Partie - La loi arcsinus

3. On rappelle que la restriction de la fonction $\sin$ à l'intervalle $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ est une bijection sur $[-1, 1]$.

On note $\arcsin : [-1, 1] \rightarrow \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ sa bijection réciproque.

(a) Montrer que pour tout $x \in [-1, 1]$, on a $\cos(\arcsin(x)) = \sqrt{1 - x^2}$.

(b) Montrer que la fonction arcsin est continue sur $[-1, 1]$ et dérivable sur $] -1, 1[$, de dérivée

$$x \mapsto \frac{1}{\sqrt{1-x^2}}.$$

4. (a) Soit $G : [0, 1] \rightarrow \mathbb{R}$ la fonction définie par $G(x) = 2 \arcsin(\sqrt{x})$.

Montrer que G est continue sur $[0, 1]$ et dérivable sur $]0, 1[$, de dérivée g donnée par

$$\forall x \in]0, 1[\quad g(x) = \frac{1}{\sqrt{x(1-x)}}.$$

(b) Etudier la fonction g et tracer son graphe.

5. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ la fonction définie par :

$$f(t) = \begin{cases} \frac{1}{\pi} g(t) & \text{si } t \in]0, 1[, \\ 0 & \text{sinon.} \end{cases}$$

(a) Montrer que la fonction f est une densité de probabilité.

(b) Soit X une variable aléatoire de densité f . On dit que X suit la loi arcsinus.

Montrer que X admet une espérance et calculer celle-ci.

(On pourra considérer le changement de variable $u = 1 - x$.)

6. (a) Soit U une variable aléatoire à densité suivant la loi uniforme sur $[0, 1]$.

Montrer que la variable aléatoire $V = \sin^2\left(\frac{\pi}{2} U\right)$ suit la loi arcsinus.

(b) À l'aide de la question précédente, compléter les deux lignes de la fonction Python suivante afin qu'elle renvoie une réalisation d'une variable aléatoire de loi arcsinus :

```
import numpy as np
import numpy.random as rd
```

```
def arcsinus() :
```

```
    U= ...
```

```
    V= ...
```

```
    return V
```

7. Soit $n \in \mathbb{N}$ tel que $n \geq 4$ et $1 \leq k \leq \frac{n}{2} - 1$. Montrer l'encadrement :

$$G\left(\frac{k+1}{n}\right) - G\left(\frac{k}{n}\right) \leq \frac{1}{\sqrt{k(n-k)}} \leq G\left(\frac{k}{n}\right) - G\left(\frac{k-1}{n}\right).$$

Montrer que les deux inégalités sont inversées si l'on suppose que $\frac{n}{2} + 1 \leq k \leq n - 1$.

8. Soient $x \in]0, 1/2[$ et $(a_n)_{n \in \mathbb{N}^*}$ une suite d'entiers telle que :

- Pour tout n suffisamment grand, on a $1 \leq a_n < \lfloor nx \rfloor$;

- $\lim_{n \rightarrow +\infty} \frac{a_n}{n} = 0$;

- $\lim_{n \rightarrow +\infty} a_n = +\infty$.

Donner un exemple d'une telle suite $(a_n)_{n \in \mathbb{N}^*}$ et montrer que :

$$\lim_{n \rightarrow +\infty} \sum_{k=a_n}^{\lfloor nx \rfloor} \frac{1}{\sqrt{k(n-k)}} = G(x).$$

On admet que pour tout $x \in]0, 1[$ et toute suite $(a_n)_{n \in \mathbb{N}^*}$ vérifiant les mêmes conditions qu'à la question précédente, on a :

$$\lim_{n \rightarrow +\infty} \sum_{k=a_n}^{\lfloor nx \rfloor} \frac{1}{\sqrt{k(n-k)}} = G(x).$$

Deuxième Partie - Marche aléatoire sur $\mathbb{Z}$

Soit $\theta \in]0, 1[$ et $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes et de même loi donnée par :

$$\forall i \in \mathbb{N}^*, \quad \mathbb{P}(X_i = 1) = \theta \quad \text{et} \quad \mathbb{P}(X_i = -1) = 1 - \theta.$$

On pose pour tout $n \in \mathbb{N}^*$, $S_n = X_1 + \dots + X_n$ et $S_0 = 0$.

La suite $(S_n)_{n \in \mathbb{N}}$ est appelée marche aléatoire sur $\mathbb{Z}$.

9. Pour tout $i \in \mathbb{N}^*$, déterminer la loi de $Y_i = \frac{1}{2} (X_i + 1)$.

En déduire la loi de $\frac{1}{2} (S_n + n)$ pour $n \in \mathbb{N}^*$.

10. Les variables aléatoires $(S_n)_{n \in \mathbb{N}^*}$ sont-elles mutuellement indépendantes ?

11. On pose pour tout $n \in \mathbb{N}^*$, $\overline{S}_n = \frac{1}{n} S_n$.

Soit $\alpha \in]0, 1[$. On note Φ la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0, 1)$ et t_α le réel positif tel que $t_\alpha = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right)$.

En majorant $\sqrt{\theta(1-\theta)}$, montrer que l'intervalle $\left[\frac{1}{2} \left(1 - \frac{t_\alpha}{\sqrt{n}} + \overline{S}_n \right), \frac{1}{2} \left(1 + \frac{t_\alpha}{\sqrt{n}} + \overline{S}_n \right) \right]$ est un intervalle de confiance asymptotique d'estimation du paramètre θ au niveau de confiance $1 - \alpha$.

12. (a) Justifier que pour tout $k \in \mathbb{N}$, on a $\mathbb{P}(S_{2k+1} = 0) = 0$.

(b) Montrer que pour tout $k \in \mathbb{N}$, on a :

$$\mathbb{P}(S_{2k} = 0) = \binom{2k}{k} \theta^k (1 - \theta)^k.$$

Soit $n \in \mathbb{N}^*$ fixé. On introduit la variable aléatoire L_n , à valeurs dans $\llbracket 0, n \rrbracket$, définie par :

$$L_n : \begin{array}{ll} \Omega & \rightarrow \mathbb{R} \\ \omega & \mapsto \max \{ k \in \llbracket 0, n \rrbracket, S_{2k}(\omega) = 0 \}. \end{array}$$

La variable aléatoire $2L_n$ décrit donc le dernier instant avant l'instant $2n$ où la marche aléatoire prend la valeur 0.

13. Déterminer $\mathbb{P}(L_n = n)$.

14. Soit $k \in \llbracket 0, n - 1 \rrbracket$.

(a) Montrer l'égalité d'événements suivante :

$$[L_n = k] = [S_{2k} = 0] \cap [S_{2k+1} \neq 0] \cap \dots \cap [S_{2n} \neq 0].$$

(b) Justifier que :

$$\mathbb{P}_{[S_{2k}=0]} ([S_{2k+1} \neq 0] \cap \dots \cap [S_{2n} \neq 0]) = \mathbb{P} ([S_1 \neq 0] \cap \dots \cap [S_{2(n-k)} \neq 0]) .$$

On notera p_{n-k} la valeur de cette probabilité. (*)

15. Montrer que pour tout $n \in \mathbb{N}^*$ et tout $k \in \llbracket 0, n \rrbracket$ on a :

$$\mathbb{P}(L_n = k) = \binom{2k}{k} \theta^k (1 - \theta)^k p_{n-k}$$

où par convention $p_0 = 1$.

A partir de maintenant et dans toute la suite du problème, $\theta = \frac{1}{2}$.
Le calcul explicite des $(p_k)_{k \in \mathbb{N}^*}$ est l'objet des questions suivantes.

Pour tout $n \in \mathbb{N}^*$, on considère les événements suivants :

$$\mathcal{D}_n = \bigcap_{k=1}^{2n} [S_k \neq 0] \quad , \quad \mathcal{D}_n^+ = \bigcap_{k=1}^{2n} [S_k > 0] \quad , \quad \mathcal{D}_n^- = \bigcap_{k=1}^{2n} [S_k < 0] .$$

On considère également, pour tout $n \in \mathbb{N}^*$ et tout $r \in \llbracket 1, n \rrbracket$, l'événement :

$$A_{n,r} = \left(\bigcap_{k=1}^{2n-1} [S_k > 0] \right) \cap [S_{2n} = 2r] .$$

Pour tout entier r , on pose $A_{n,r} = \emptyset$ lorsque $r > n$ ou $r \leq 0$ et on note, pour tout $n \in \mathbb{N}^*$ et tout $r \in \mathbb{Z}$, $a_{n,r} = \mathbb{P}(A_{n,r})$.

16. (a) Montrer que, pour tout $\omega \in \Omega$, l'ensemble $\{S_1(\omega), \dots, S_r(\omega)\}$ est un intervalle d'entiers.

(b) En déduire que :

$$\mathbb{P}(\mathcal{D}_n) = \mathbb{P}(\mathcal{D}_n^+) + \mathbb{P}(\mathcal{D}_n^-) .$$

17. Montrer que :

$$\mathbb{P}(\mathcal{D}_n^+) = \sum_{r=1}^n a_{n,r} .$$

18. Soit $n \in \mathbb{N}$ tel que $n \geq 2$.

(a) Soit $(r, q) \in (\mathbb{N}^*)^2$. Montrer que

$$\mathbb{P}(A_{n-1,q} \cap A_{n,r}) = \begin{cases} \frac{1}{4} a_{n-1,q} & \text{si } q = r-1 \text{ ou } q = r+1 \\ \frac{1}{2} a_{n-1,q} & \text{si } q = r \\ 0 & \text{sinon.} \end{cases}$$

(b) Montrer que pour tout $r \in \mathbb{N}^*$, on a :

$$a_{n,r} = \frac{1}{4} a_{n-1,r-1} + \frac{1}{2} a_{n-1,r} + \frac{1}{4} a_{n-1,r+1} .$$

19. Montrer par récurrence sur n que pour tout $(n, r) \in (\mathbb{N}^*)^2$ on a :

$$a_{n,r} = \frac{1}{4^n} \left(\binom{2n-1}{n+r-1} - \binom{2n-1}{n+r} \right),$$

où l'on pose $\binom{m}{k} = 0$ dès que $k \notin \llbracket 0, m \rrbracket$.

20. On rappelle que la suite $(p_n)_{n \in \mathbb{N}}$ a été définie précédemment (*).

Montrer que, pour tout $n \in \mathbb{N}$, on a : $p_n = \frac{1}{4^n} \binom{2n}{n}$.

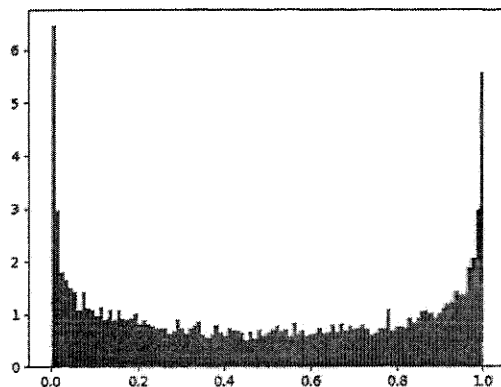
Troisième Partie - Comportement asymptotique

Dans cette troisième partie, on cherche à étudier la limite en loi de la suite de variables aléatoires $\left(\frac{L_n}{n} \right)_{n \in \mathbb{N}^*}$.

21. (a) Compléter par autant de lignes que nécessaire la fonction Python *DernierPassage* permettant de simuler la variable aléatoire L_n :

```
import numpy.random as rd
def DernierPassage(n):
    L=0
    S=0
    for i in range(1,2*n+1):
        ...
    return L
```

(b) A l'aide de cette fonction, on simule $N = 10000$ réalisations de la variable aléatoire $\frac{L_{100}}{100}$ puis on représente les valeurs obtenues sous la forme d'un histogramme à 100 classes. On obtient la figure suivante :



Quelle conjecture peut-on formuler à propos de la suite de variables aléatoires $\left(\frac{L_n}{n} \right)_{n \in \mathbb{N}^*}$?

22. Pour tout entier $n \in \mathbb{N}^*$, on pose $C_n = \frac{\sqrt{n}}{4^n} \binom{2n}{n}$ et $u_n = \ln(C_n)$.

(a) Montrer que la série de terme général $u_{n+1} - u_n$ converge.

(b) En déduire que la suite $(C_n)_{n \in \mathbb{N}^*}$ est convergente.

On admet que $\frac{1}{\sqrt{\pi}}$ est la limite de cette suite.

23. Soient $x \in]0, 1[$ et $(a_n)_{n \in \mathbb{N}^*}$ une suite d'entiers telle que :

- Pour tout n suffisamment grand, on a $1 \leq a_n < \lfloor nx \rfloor$;
- $\lim_{n \rightarrow +\infty} \frac{a_n}{n} = 0$;
- $\lim_{n \rightarrow +\infty} a_n = +\infty$.

(a) Montrer que :

$$\mathbb{P}\left(\frac{L_n}{n} \leq x\right) = \frac{1}{4^n} \sum_{k=0}^{\lfloor nx \rfloor} \binom{2k}{k} \binom{2(n-k)}{n-k}.$$

(b) Montrer, pour n suffisamment grand, l'encadrement :

$$m_n \sum_{k=a_n}^{\lfloor nx \rfloor} \frac{1}{\sqrt{k(n-k)}} \leq \frac{1}{4^n} \sum_{k=a_n}^{\lfloor nx \rfloor} \binom{2k}{k} \binom{2(n-k)}{n-k} \leq M_n \sum_{k=a_n}^{\lfloor nx \rfloor} \frac{1}{\sqrt{k(n-k)}},$$

où :

$$m_n = \min \{C_k C_{n-k}, k \in \llbracket a_n, \lfloor nx \rfloor \rrbracket\}, \quad M_n = \max \{C_k C_{n-k}, k \in \llbracket a_n, \lfloor nx \rfloor \rrbracket\}.$$

(c) Soient $(\alpha_n)_{n \in \mathbb{N}^*}$ et $(\beta_n)_{n \in \mathbb{N}^*}$ deux suites d'entiers telles que

$$1 \leq \alpha_n \leq \beta_n \text{ et } \lim_{n \rightarrow +\infty} \alpha_n = +\infty.$$

Montrer que les deux suites $\left(\min_{k \in \llbracket \alpha_n, \beta_n \rrbracket} C_k\right)_{n \in \mathbb{N}^*}$ et $\left(\max_{k \in \llbracket \alpha_n, \beta_n \rrbracket} C_k\right)_{n \in \mathbb{N}^*}$ sont convergentes, de même limite $\frac{1}{\sqrt{\pi}}$.

(d) En déduire que :

$$\lim_{n \rightarrow +\infty} \frac{1}{4^n} \sum_{k=a_n}^{\lfloor nx \rfloor} \binom{2k}{k} \binom{2(n-k)}{n-k} = \frac{G(x)}{\pi},$$

où la fonction G a été définie dans la première partie.

24. Soient $x \in]0, 1[$ et $(a_n)_{n \in \mathbb{N}^*}$ une suite d'entiers telle que :

- Pour tout n suffisamment grand, on a $1 \leq a_n < \lfloor nx \rfloor$;
- $\lim_{n \rightarrow +\infty} \frac{a_n}{\sqrt{n}} = 0$;
- $\lim_{n \rightarrow +\infty} a_n = +\infty$.

(a) Donner un exemple d'une telle suite $(a_n)_{n \in \mathbb{N}^*}$.

(b) En déduire que :

$$\lim_{n \rightarrow +\infty} \frac{1}{4^n} \sum_{k=0}^{a_n-1} \binom{2k}{k} \binom{2(n-k)}{n-k} = 0.$$

25. Montrer que la suite de variables aléatoires $\left(\frac{L_n}{n}\right)_{n \in \mathbb{N}^*}$ converge en loi.