



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 LETTRES & SCIENCES-HUMAINES

MATHEMATIQUES

PROGRAMME ENS (B/L)

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.*

On désire tester si la répartition des voyelles dans un texte ancien correspond à celle de la langue française actuelle.
On étudie quelques aspects probabilistes permettant de répondre à cette question.
Les parties 1 et 2 sont indépendantes. La partie 3 utilise les notations et les résultats des parties précédentes.

Notation

Si X et Y sont deux variables aléatoires discrètes, on note $\text{cov}(X, Y)$ leur covariance, si celle-ci existe.

Partie 1

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). 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 $Z_1 = \frac{X_1 - np_1}{\sqrt{np_1 p_2}}$. (On utilisera la relation : $X_1 + X_2 = n$).
- 2) Par quelle loi peut-on approcher la loi de Z_1 lorsque n est grand ?

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 $U_n = Z_1^2 + Z_2^2$. (On utilisera la relation : $X_1 + X_2 + X_3 = n$).
- 2) Déterminer les espérances et variances de Z_1 et Z_2 et $cov(Z_1, Z_2)$.
- 3) Par quelle loi peut-on approcher celle 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.

D. On suppose $s = 4$ et $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$.

- 1) Pour i élément de $\{1, 2, 3, 4\}$, on pose $Y_i = \frac{X_i - np_i}{\sqrt{np_i}}$.

On note M la matrice carrée d'ordre 4 dont le coefficient en ligne i et colonne j vaut $cov(Y_i, Y_j)$. On définit N la matrice carrée d'ordre 4 dont tous les coefficients sont égaux à $\frac{1}{4}$.

Exprimer M en fonction de N et I , où I désigne la matrice unité.

- 2) a) On définit 4 vecteurs de $\mathbb{R}^4$: $e_1 = \frac{1}{\sqrt{2}}(1, -1, 0, 0)$, $e_2 = \frac{1}{\sqrt{6}}(1, 1, -2, 0)$, $e_3 = \frac{1}{2\sqrt{3}}(1, 1, 1, -3)$, $e_4 = \frac{1}{2}(1, 1, 1, 1)$.

Montrer que ces 4 vecteurs sont des vecteurs propres de N et qu'ils forment une base de $\mathbb{R}^4$.

- b) Soit Q la matrice de passage de la base canonique de $\mathbb{R}^4$ à la base (e_1, e_2, e_3, e_4) .

Montrer que ${}^t Q Q = I$. (${}^t Q$ désigne la transposée de Q).

Expliciter la matrice ${}^t Q M Q$.

- 3) On définit les variables Z_1, Z_2, Z_3 et Z_4 par :
$$\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} = {}^t Q \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}.$$

- a) Exprimer chaque Z_i en fonction de Y_1, Y_2, Y_3 et Y_4 et montrer que $Z_4 = 0$.

- b) Montrer que, pour $i = 1, 2, 3$, les variables Z_i sont centrées.

- c) Montrer que $U_n = Z_1^2 + Z_2^2 + Z_3^2$. (On pourra calculer $(Y_1, Y_2, Y_3, Y_4) \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}$).

Partie 2

1) Soit r un entier naturel non nul.

a) Montrer la convergence de l'intégrale $\int_0^{+\infty} x^{\frac{r}{2}-1} e^{-\frac{x}{2}} dx$. On note J_r sa valeur.

b) Montrer que pour tout entier r non nul, $J_{r+2} = r J_r$.

2) a) Soit Y une variable aléatoire gaussienne centrée réduite.

Déterminer la fonction de répartition de la variable aléatoire Y^2 et montrer que Y^2 admet une densité que l'on déterminera.

b) En déduire la valeur de J_1 .

3) Soit r un entier naturel non nul. On définit la fonction f_r par :

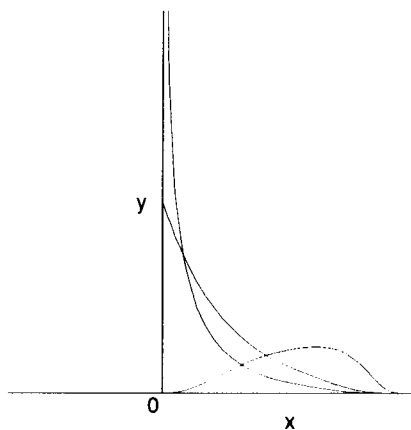
$$\forall x > 0, \quad f_r(x) = \frac{1}{J_r} x^{\frac{r}{2}-1} e^{-\frac{x}{2}} \quad \forall x \leq 0, \quad f_r(x) = 0.$$

a) Montrer que f_r est une densité de probabilité.

On dit qu'une variable aléatoire X suit la loi du χ^2 (lire khi-deux) à r degrés de liberté si et seulement si X admet f_r pour densité.

b) Quelle loi reconnaît-on pour $r = 2$?

c) Recopier le graphique ci-dessous en identifiant les courbes représentatives des trois fonctions f_1, f_2, f_3 , en justifiant avec précision la réponse.



d) Soit X une variable aléatoire suivant la loi du χ^2 à r degrés de liberté. Montrer que X admet une espérance et une variance et donner leur valeur.

Partie 3

On reprend les notations de la partie 1, s étant un entier quelconque supérieur ou égal à 2.

On admet que, n étant grand, la variable aléatoire U_n suit la loi du χ^2 à $s-1$ degrés de liberté.

Les linguistes ont déterminé la répartition, dans la langue française actuelle, des voyelles a, e, i, o, u, parmi l'ensemble de ces voyelles :

$$p_1 = 0,1743 \text{ de a}$$

$$p_2 = 0,4031 \text{ de e}$$

$$p_3 = 0,1641 \text{ de i}$$

$$p_4 = 0,1212 \text{ de o}$$

$$p_5 = 0,1373 \text{ de u.}$$

Dans un extrait (comportant 1145 voyelles a, e, i, o, u) d'un texte de Pascal, « Usage du triangle arithmétique pour les combinaisons » (confer « Traité du triangle arithmétique » 1654), on observe les fréquences :

$$q_1 = 0,1476 \text{ de a}$$

$$q_2 = 0,4105 \text{ de e}$$

$$q_3 = 0,1607 \text{ de i}$$

$$q_4 = 0,1476 \text{ de o}$$

$$q_5 = 0,1336 \text{ de u.}$$

On cherche, à l'aide de cet extrait, à mettre en évidence l'évolution des fréquences lexicales de la langue française. Pour cela on fait l'hypothèse, a contrario, que les fréquences des voyelles dans un texte, parmi l'ensemble des voyelles, sont les mêmes au 17^{ième} siècle et au 20^{ième} siècle.

On donne les valeurs numériques suivantes :

- $0,975 = F_4(11, 14)$, où F_4 désigne la fonction de répartition de la loi du χ^2 à 4 degrés de liberté,

- $1145 \sum_{i=1}^5 \frac{(q_i - p_i)^2}{p_i} = 11,62$.

En introduisant des variables aléatoires convenables X_1, X_2, X_3, X_4, X_5 et la variable U_n associée, justifier le rejet de l'hypothèse de la stabilité de la fréquence des voyelles.
