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

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 :

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

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 : $\qquad e^{\lambda} = \sum_{k=0}^{n-1} \frac{\lambda^k}{k!} + \int_0^{\lambda} e^{\lambda - t} \frac{t^{n-1}}{(n-1)!} dt$.

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, ..., X_k des variables indépendantes gaussiennes centrées réduites.

a) Déterminer la loi de X_1^2 .

b) En déduire la loi de $X_1^2+..+X_k^2$ (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 , ..,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 $\{1,..,s\}$, 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 $(i,j) \in \{1,..,s\}^2$ tel que $i \neq j$. Déterminer la loi de $X_i + X_j$ et sa variance .
En déduire que $cov(X_i, X_j) = -np_i p_j$.

B. On suppose dans cette partie que $s = 2$.

1) Montrer que $U_n = Z_1^2$ où $Z_1 = \frac{X_1 - np_1}{\sqrt{np_1p_2}}$.

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 $cov(Z_1, Z_2)$.

3) Par quelle loi peut-on approcher la loi de Z_1 lorsque n est grand?

4) Pour i élément de $\{1,..,n\}$, 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 $Y_i = \dfrac{X_i - np_i}{\sqrt{np_i}}$. On note M la matrice de covariance des variables Y_1 ,..,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 ${}^tQ\,Q = I$ et telle que $M = Q\,J_{s-1}\,{}^tQ$ 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,.., Z_s par : $\begin{pmatrix} Z_1 \\ .. \\ .. \\ Z_s \end{pmatrix} = {}^tQ \begin{pmatrix} Y_1 \\ .. \\ .. \\ Y_s \end{pmatrix}$.

On note a_{ij} l'élément de la ligne i et de la colonne j de ${}^tQ$. Exprimer chaque Z_i en fonction de Y_1 , .., 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 , .., 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 ,.., 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,
- $8000\displaystyle\sum_{i=1}^{4}\frac{(f_i - p_i)^2}{p_i} = 12{,}03$.

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 $P(X_\lambda \le x)$.

x \ λ	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