



PRÉPASIMPLE

prepasimple.fr

HEC — 1999

Sujet de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Sujet

Partie 1

Sous réserve d'existence, on note $\text{Cov}(X, Y)$ la covariance de deux variables aléatoires X et Y .

Soit $r \in \mathbb{N}^*$. On définit la loi du χ^2 à r degrés de liberté comme la loi admettant pour densité

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

On rappelle que, pour $a > 0$,

$$\Gamma(a) = \int_0^{+\infty} t^{a-1} e^{-t} dt.$$

La loi gamma de forme a et de taux 1 a pour densité $t \mapsto t^{a-1} e^{-t} / \Gamma(a)$ sur $]0, +\infty[$. Ainsi, dire que X suit la loi du χ^2 à r degrés de liberté revient à dire que $X/2$ suit cette loi gamma de forme $r/2$.

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 $\lambda > 0$ et tout entier $n \geq 1$,

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

2.b) Soit Y une variable aléatoire suivant la loi de Poisson de paramètre $\lambda > 0$ et X_{2n} une variable aléatoire suivant la loi du χ^2 à $2n$ degrés de liberté. Montrer que

$$P(X_{2n} > 2\lambda) = P(Y < n).$$

2.c) Écrire une fonction Python de paramètres n , entier strictement positif, et x , réel, qui retourne la valeur de $P(X_{2n} > x)$.

2.d) À 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 $F_6(0)$, $F_6(4)$ et $F_6(8)$.

3. Soit $k \in \mathbb{N}^*$ et $X_1, \dots, X_k$ des variables indépendantes gaussiennes centrées réduites.

3.a) Déterminer la loi de X_1^2 .

Pour identifier la constante de normalisation, on pourra établir la valeur de $\Gamma(1/2)$ à partir de l'intégrale d'une densité normale centrée réduite, en utilisant le changement de variable $t = u^2$ sur $]0, +\infty[$.

3.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.

3.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 $r, r' \in \mathbb{N}^*$ et $r < r'$.

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 $p_i > 0$ pour tout i .

On effectue n tirages successifs d'une boule avec remise ; les tirages sont indépendants.

Pour $i \in \llbracket 1, s \rrbracket$, on note X_i le nombre de boules de couleur C_i obtenues à l'issue des n

tirages. La variable X_i dépend de n , et

$$\sum_{i=1}^s X_i = n.$$

On définit

$$U_n = \sum_{i=1}^s \frac{(X_i - np_i)^2}{np_i}.$$

A. Étude des variables X_i

1. Déterminer la loi de X_i , son espérance et sa variance.
2. Soit $(i, j) \in \llbracket 1, s \rrbracket^2$ avec $i \neq j$. 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 ?

2.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 $s = 3$, $p_1 = p_2 = 1/4$ et $p_3 = 1/2$

On pose

$$Z_1 = \frac{2}{\sqrt{n}} \left(X_3 - \frac{n}{2} \right), \quad Z_2 = \sqrt{\frac{2}{n}} (X_1 - X_2).$$

1. Montrer que $U_n = Z_1^2 + Z_2^2$.
2. Déterminer les espérances et les variances de Z_1 et Z_2 , ainsi que $\text{Cov}(Z_1, Z_2)$.
3. Par quelle loi peut-on approcher la loi de Z_1 lorsque n est grand ?
4. Pour $i \in \llbracket 1, n \rrbracket$, on définit T_i par : $T_i = 1$ si, au i -ième tirage, on a obtenu une boule de couleur C_1 ; $T_i = -1$ si on a obtenu une boule de couleur C_2 ; $T_i = 0$ si on a obtenu une boule de couleur C_3 .
 - 4.a) Exprimer $X_1 - X_2$ à l'aide des variables T_i .
 - 4.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 admet le résultat particulier suivant : si G_1 et G_2 sont deux variables indépendantes de loi normale centrée réduite, alors, pour tout réel x ,

$$\lim_{n \rightarrow +\infty} P(Z_1^2 + Z_2^2 \leq x) = P(G_1^2 + G_2^2 \leq x).$$

En déduire la loi limite de U_n et la loi par laquelle on peut approcher sa loi lorsque n est grand. Préciser pourquoi il ne s'agit pas d'une loi exacte pour un entier n fixé.

6. On représente en Python une suite de 100 tirages par un tableau `C` de taille `(100,)` contenant les entiers 1, 2 ou 3. L'élément `C[i]`, pour $0 \leq i \leq 99$, représente la couleur de la boule obtenue au $(i + 1)$ -ième tirage.

6.a) Écrire une fonction Python `Tirage()` qui simule les 100 tirages indépendants avec remise, avec les proportions $1/4$, $1/4$ et $1/2$, et renvoie le tableau `C`.

On utilisera `numpy.random.randint(0, 4)`, qui renvoie un entier aléatoire uniforme parmi 0, 1, 2, 3.

6.b) Écrire une fonction Python `Difference(C)` qui retourne la valeur de $X_1 - X_2$ associée au tableau `C`.

D. On suppose désormais que s est un 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 de $Y_1, \dots, Y_s$. Montrer que $M = I - N$, où I est la matrice identité d'ordre s et N la matrice dont le coefficient de ligne i et de 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 ${}^tQQ = I$ et

$$M = QJ_{s-1}{}^tQ,$$

où J_{s-1} est la matrice diagonale d'ordre s dont les $s - 1$ premiers coefficients diagonaux valent 1 et dont le dernier vaut 0.

On ne demande pas de calculer explicitement Q .

4. On définit $Z_1, \dots, Z_s$ par

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_s \end{pmatrix} = {}^t Q \begin{pmatrix} Y_1 \\ \vdots \\ Y_s \end{pmatrix}.$$

On note $a_{i,j}$ le coefficient de ligne i et de colonne j de ${}^t Q$. Exprimer chaque Z_i en fonction de $Y_1, \dots, Y_s$ et des $a_{i,j}$.

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 Z_s ?

5. On admet le résultat particulier suivant : si $G_1, \dots, G_{s-1}$ sont des variables mutuellement indépendantes de loi normale centrée réduite, alors, pour tout réel x ,

$$\lim_{n \rightarrow +\infty} P \left(\sum_{i=1}^{s-1} Z_i^2 \leq x \right) = P \left(\sum_{i=1}^{s-1} G_i^2 \leq x \right).$$

Montrer que U_n converge en loi vers la loi du χ^2 à $s - 1$ degrés de liberté. En déduire la loi approchée de U_n pour n grand.

Partie 3

On s'intéresse à la répartition le long de l'année des naissances en Suède dans les années 1930. Les quatre périodes et leurs proportions de journées dans l'année sont :

Période	Mois	Proportion p_i
T_1	Avril – juin	0,250
T_2	Juillet – août	0,169
T_3	Septembre – octobre	0,167

Période	Mois	Proportion p_i
T_4	Novembre – mars	0,414

On dispose d'un échantillon de 8 000 naissances, regroupées selon ces quatre périodes, avec les proportions observées suivantes :

Période	Proportion observée f_i
T_1	0,264
T_2	0,173
T_3	0,159
T_4	0,404

On fait l'hypothèse que, parmi la population observée, le jour de naissance d'un enfant dans l'année suit une loi uniforme. On considère les naissances de l'échantillon comme indépendantes et on confronte cette hypothèse aux données.

On donne les valeurs numériques suivantes :

$$F_3(11,3) \simeq 0,99,$$

où F_3 est la fonction de répartition de la loi du χ^2 à 3 degrés de liberté, et

$$8\,000 \sum_{i=1}^4 \frac{(f_i - p_i)^2}{p_i} \simeq 12,03.$$

On utilise la règle de décision suivante : on rejette l'hypothèse de répartition uniforme lorsque la valeur observée de U_n dépasse 11,3. Le seuil est choisi pour que, sous cette

hypothèse et avec l'approximation de la partie 2, la probabilité de rejet soit voisine de 1 %.

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 : pour chaque paramètre λ et chaque entier x indiqués, le tableau donne $P(Y \leq x)$, où $Y \sim \mathcal{P}(\lambda)$.

$x \backslash \lambda$	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