



PRÉPASIMPLE

prepasimple.fr

HEC — 1999

Corrigé de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Corrigé

Partie 1

La loi du χ^2 est utilisée uniquement avec la définition donnée dans le sujet. On note F_r sa fonction de répartition, Φ la fonction de répartition de la loi normale centrée réduite, et $\Gamma(a) = \int_0^{+\infty} t^{a-1} e^{-t} dt$ pour $a > 0$.

Question adaptée au programme actuel — Le préambule précise $r \in \mathbb{N}^*$ et identifie la loi définie à une transformation affine d'une loi gamma de taux 1. Aucune propriété de la loi du χ^2 n'est supposée connue avant les questions qui l'établissent.

1.

Posons $a = r/2 > 0$.

Méthode 1 — Transformation d'une loi gamma

Par la définition du sujet, $T = X/2$ suit la loi gamma de forme a et de taux 1. Ses moments usuels existent et valent $E(T) = a$ et $V(T) = a$.

Comme $X = 2T$, les règles de transformation affine donnent

$$E(X) = 2a = r, \quad V(X) = 4a = 2r.$$

Méthode 2 — Calcul des deux premiers moments par transfert

Pour $p = 1$ ou $p = 2$, la variable X est positive presque sûrement. Le changement de variable affine $x = 2t$, d'abord sur un segment puis par passage aux limites, donne

$$\int_0^{+\infty} x^p f_r(x) dx = \frac{2^p}{\Gamma(a)} \int_0^{+\infty} t^{a+p-1} e^{-t} dt = \frac{2^p \Gamma(a+p)}{\Gamma(a)}.$$

L'intégrale est finie puisque $a + p > 0$. Elle justifie donc l'existence des moments avant leur utilisation. Avec $\Gamma(a + 1) = a\Gamma(a)$ et $\Gamma(a + 2) = (a + 1)a\Gamma(a)$,

$$E(X) = 2a = r, \quad E(X^2) = 4a(a + 1) = r^2 + 2r.$$

La formule de Huygens donne alors $V(X) = r^2 + 2r - r^2 = 2r$.

La première méthode exploite directement la loi gamma ; la seconde établit les résultats à partir de la densité.

2.a)

Méthode 1 — Formule de Taylor avec reste intégral

La fonction exponentielle est de classe $\mathcal{C}^n$ sur le segment $[0, \lambda]$. La formule de Taylor en 0, avec reste intégral d'ordre n , donne

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

Dans l'intégrale, le changement de variable affine $t = \lambda - u$ transforme les bornes $u = 0, \lambda$ en $t = \lambda, 0$. Il vient

$$\int_0^\lambda \frac{(\lambda - u)^{n-1}}{(n-1)!} e^u du = \int_0^\lambda e^{\lambda-t} \frac{t^{n-1}}{(n-1)!} dt,$$

ce qui est la formule demandée.

Méthode 2 — Intégrations par parties et récurrence

Notons $R_n(\lambda) = \int_0^\lambda e^{\lambda-t} t^{n-1} / (n-1)! dt$.

Pour $n = 1$,

$$R_1(\lambda) = [-e^{\lambda-t}]_0^\lambda = e^\lambda - 1.$$

Pour $n \geq 1$, une intégration par parties sur le segment $[0, \lambda]$, avec la primitive $-e^{\lambda-t}$, donne

$$R_{n+1}(\lambda) = -\frac{\lambda^n}{n!} + R_n(\lambda).$$

Le terme en 0 est nul parce que $n \geq 1$. Si

$$R_n(\lambda) = e^\lambda - \sum_{k=0}^{n-1} \lambda^k / k!,$$

la relation précédente donne

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

La formule est donc établie par récurrence pour tout $n \geq 1$. La première méthode reconnaît une formule du cours ; la seconde explique comment se construit le reste.

2.b)

La variable X_{2n} possède une densité et $\Gamma(n) = (n-1)!$. Sa fonction de répartition en 2λ vaut

$$F_{2n}(2\lambda) = \frac{1}{(n-1)!2^n} \int_0^{2\lambda} x^{n-1} e^{-x/2} dx.$$

Avec $x = 2t$, changement de variable affine, on obtient

$$F_{2n}(2\lambda) = \int_0^\lambda \frac{t^{n-1}}{(n-1)!} e^{-t} dt.$$

En multipliant l'identité de 2.a) par $e^{-\lambda}$, il vient

$$1 = e^{-\lambda} \sum_{k=0}^{n-1} \frac{\lambda^k}{k!} + F_{2n}(2\lambda).$$

Ainsi

$$P(X_{2n} > 2\lambda) = 1 - F_{2n}(2\lambda) = \sum_{k=0}^{n-1} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Or $Y \sim \mathcal{P}(\lambda)$ et $\{Y < n\} = \{Y \in \llbracket 0, n-1 \rrbracket\}$. Le dernier membre vaut $P(Y < n)$, d'où l'égalité demandée. Aucune indépendance entre X_{2n} et Y n'est nécessaire : il s'agit d'une égalité de probabilités marginales.

2.c)

Code adapté en Python. Pour $x > 0$, on applique 2.b) avec $\lambda = x/2$. Pour $x \leq 0$, la positivité presque sûre de X_{2n} donne une probabilité égale à 1.

Python

```
import numpy as np
```

```
def proba_superieure(n, x):
```

```

if x <= 0:
    return 1.0
lam = x / 2
terme = np.exp(-lam)
total = terme
for k in range(1, n):
    terme = terme * lam / k
    total = total + terme
return total

```

À l'entrée de l'itération d'indice k , la mise à jour produit le terme $e^{-\lambda} \lambda^k / k!$ à partir du précédent. La somme finale porte exactement sur les indices 0 à $n - 1$.

Le cas $n = 1$ est traité correctement : la boucle est vide et la fonction renvoie $e^{-x/2}$ pour $x > 0$. Le code suppose $n \in \mathbb{N}^*$, conformément au sujet.

Ce code a été exécuté. Pour $n = 3$, il donne 1 en $x = 0$, environ 0,6766764 en $x = 4$ et 0,2381033 en $x = 8$, en accord avec l'annexe. Pour $n = 1$, les résultats ont également été comparés à $e^{-x/2}$.

2.d)

La densité étant nulle sur $] -\infty, 0]$, $F_6(0) = 0$.

Pour $x = 4$, la question 2.b) s'applique avec $n = 3$ et $\lambda = 2$. La table donne

$$F_6(4) = 1 - P(\mathcal{P}(2) \leq 2) \simeq 1 - 0,6767 = 0,3233.$$

Pour $x = 8$, on prend $\lambda = 4$, d'où

$$F_6(8) = 1 - P(\mathcal{P}(4) \leq 2) \simeq 1 - 0,2381 = 0,7619.$$

Pour l'allure complète, 2.b) donne l'expression exacte

$$\forall x \geq 0, \quad F_6(x) = 1 - e^{-x/2} \left(1 + \frac{x}{2} + \frac{x^2}{8} \right).$$

Sur $]0, +\infty[$, le lien entre une densité continue et sa fonction de répartition donne

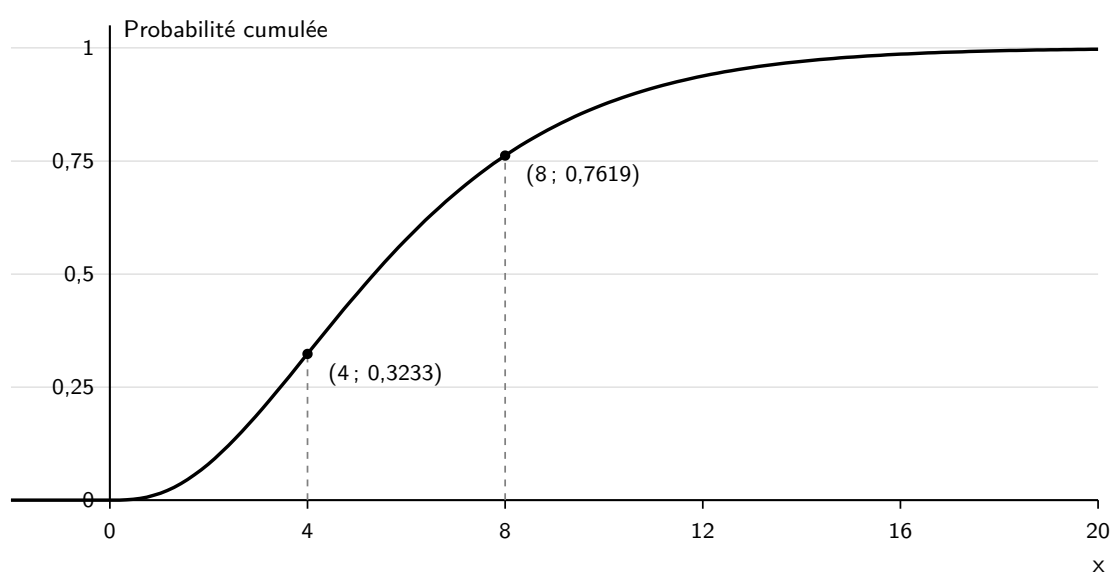
$$F'_6(x) = f_6(x) = \frac{x^2}{16} e^{-x/2} > 0$$

et

$$F''_6(x) = \frac{x(4-x)}{32} e^{-x/2}.$$

La fonction est nulle pour $x \leq 0$, strictement croissante pour $x > 0$, convexe sur $]0, 4[$ et concave sur $]4, +\infty[$. Elle possède un point d'inflexion en $x = 4$, une tangente horizontale en 0 et tend vers 1 en $+\infty$, par croissance comparée de l'exponentielle et des polynômes.

Le graphique ci-dessous est calculé à partir de l'expression exacte ; les deux ordonnées indiquées sont arrondies comme dans la table.



3.a)

Posons $V = X_1^2$. Comme X_1 est normale centrée réduite, elle admet la densité $u \mapsto e^{-u^2/2}/\sqrt{2\pi}$ et n'a pas d'atome.

Pour $x < 0$, $P(V \leq x) = 0$. Pour $x \geq 0$,

$$P(V \leq x) = P(-\sqrt{x} \leq X_1 \leq \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = 2\Phi(\sqrt{x}) - 1.$$

Cette fonction est continue en 0, puisque $\Phi(0) = 1/2$. Pour $x > 0$, sa dérivée vaut

$$\frac{e^{-x/2}}{\sqrt{2\pi x}}.$$

Elle fournit une densité de V sur $]0, +\infty[$, que l'on complète par 0 ailleurs.

Il reste à identifier la constante avec celle de f_1 . La normalisation de la densité normale et sa parité donnent

$$\int_0^{+\infty} e^{-u^2/2} du = \sqrt{\frac{\pi}{2}}.$$

Le changement affine $u = \sqrt{2}v$ entraîne

$$\int_0^{+\infty} e^{-v^2} dv = \frac{\sqrt{\pi}}{2}.$$

Dans l'intégrale définissant $\Gamma(1/2)$, on utilise le changement de variable indiqué $t = v^2$, d'abord sur $[\varepsilon, R]$ avec $0 < \varepsilon < R$, puis par passage aux limites. Comme $dt = 2v dv$,

$$\Gamma(1/2) = 2 \int_0^{+\infty} e^{-v^2} dv = \sqrt{\pi}.$$

La densité obtenue est donc

$$\frac{x^{-1/2}e^{-x/2}}{\Gamma(1/2)2^{1/2}} = f_1(x) \quad (x > 0).$$

Ainsi X_1^2 suit la loi du χ^2 à 1 degré de liberté, au sens de la définition du sujet. On a établi cette loi par la fonction de répartition, sans formule de transformation par une densité jointe.

3.b)

D'après 3.a), pour chaque i , $X_i^2/2$ suit la loi gamma de forme $1/2$ et de taux 1 .

Les variables X_i^2 sont indépendantes, comme l'autorise l'énoncé; les variables $X_i^2/2$ le sont donc aussi. Elles ont le même taux, et la stabilité par somme des lois gamma indépendantes donne

$$\frac{1}{2} \sum_{i=1}^k X_i^2 \sim \Gamma(k/2, \text{taux } 1).$$

Par la définition de la loi étudiée,

$$X_1^2 + \dots + X_k^2 \text{ suit la loi du } \chi^2 \text{ à } k \text{ degrés de liberté.}$$

On a ainsi démontré la propriété utile des sommes de carrés de normales indépendantes à partir de la loi gamma.

3.c)

Pour comparer les fonctions de répartition, on construit les deux variables à partir d'une même famille de normales indépendantes. Prenons $G_1, \dots, G_{r'}$ indépendantes et normales centrées réduites, et posons

$$A = \sum_{i=1}^r G_i^2, \quad B = \sum_{i=r+1}^{r'} G_i^2.$$

D'après 3.b), A a pour fonction de répartition F_r et $A + B$ a pour fonction de répartition $F_{r'}$. Puisque $B \geq 0$, on a $A + B \geq A$, donc

$$\forall x \in \mathbb{R}, \quad F_{r'}(x) = P(A + B \leq x) \leq P(A \leq x) = F_r(x).$$

Pour $x > 0$, cette inégalité est stricte. En effet, A et B sont indépendantes par le lemme des coalitions. Leurs densités sont strictement positives sur $]0, +\infty[$, car leurs nombres de degrés de liberté sont strictement positifs. Ainsi

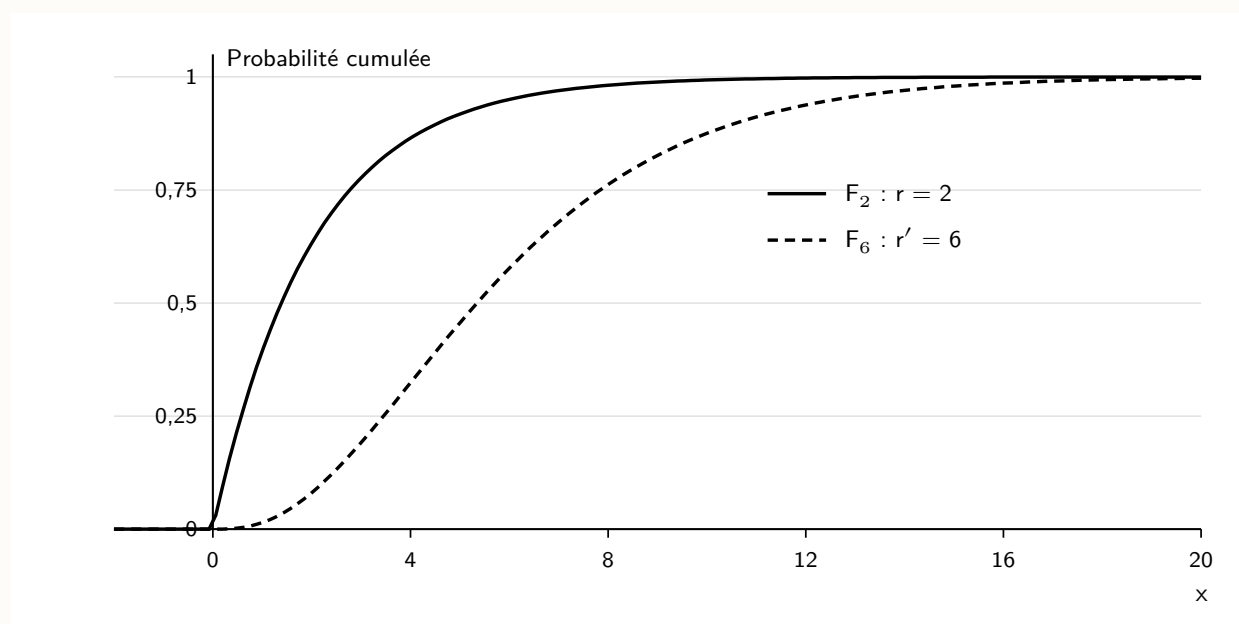
$$P(x/3 < A < 2x/3) > 0, \quad P(B > x) > 0.$$

Par indépendance, l'intersection de ces événements a une probabilité strictement positive. Sur cette intersection, $A \leq x$ et $A + B > x$. Il en résulte $F_r(x) - F_{r'}(x) > 0$.

Les deux fonctions sont nulles sur $] -\infty, 0]$, continues, strictement croissantes sur $]0, +\infty[$ et tendent vers 1 en $+\infty$. La courbe associée au plus grand nombre de degrés de liberté se situe sous l'autre pour tout $x > 0$.

Pour illustrer cette disposition sans fixer les entiers du sujet, le dessin suivant prend les valeurs représentatives $r = 2$ et $r' = 6$. Les courbes sont calculées à partir des formules exactes

$$F_2(x) = 1 - e^{-x/2}, \quad F_6(x) = 1 - e^{-x/2}(1 + x/2 + x^2/8) \quad (x \geq 0).$$



Partie 2

Dans cette partie, les variables de comptage X_i sont celles des tirages dans l'urne ; elles ne désignent plus les variables normales de la partie 1.

A. Étude des variables X_i

1.

Pour chaque tirage, l'événement « obtenir la couleur C_i » a pour probabilité p_i . Les tirages sont indépendants et effectués avec remise, donc ces événements donnent n épreuves de Bernoulli indépendantes de même paramètre p_i .

La variable X_i en compte les succès :

$$X_i \sim \mathcal{B}(n, p_i).$$

Par les moments usuels de la loi binomiale,

$$E(X_i) = np_i, \quad V(X_i) = np_i(1 - p_i).$$

Les variables étant bornées entre 0 et n , tous leurs moments existent.

2.

Pour $i \neq j$, les deux couleurs sont incompatibles au cours d'un même tirage. La probabilité d'obtenir l'une ou l'autre vaut $p_i + p_j$. La variable $X_i + X_j$ compte ces succès dans les n tirages indépendants, donc

$$X_i + X_j \sim \mathcal{B}(n, p_i + p_j),$$

avec

$$V(X_i + X_j) = n(p_i + p_j)(1 - p_i - p_j).$$

Si $s = 2$, on a $p_i + p_j = 1$: cette loi est alors concentrée en n , et sa variance vaut 0, ce qui reste conforme à la formule.

Méthode 1 — Variance de la somme

Les variables admettent des moments d'ordre 2, donc

$$2\text{Cov}(X_i, X_j) = V(X_i + X_j) - V(X_i) - V(X_j).$$

En remplaçant les variances,

$$2\text{Cov}(X_i, X_j) = n((p_i + p_j) - (p_i + p_j)^2 - p_i + p_i^2 - p_j + p_j^2) = -2np_i p_j.$$

Ainsi $\text{Cov}(X_i, X_j) = -np_i p_j$.

Méthode 2 — Indicatrices des tirages

Pour $\ell \in \llbracket 1, n \rrbracket$, notons $B_{\ell,i}$ l'indicatrice de l'obtention de C_i au tirage ℓ . On a

$$X_i = \sum_{\ell=1}^n B_{\ell,i}.$$

Si $\ell \neq m$, les indicatrices $B_{\ell,i}$ et $B_{m,j}$ concernent des tirages indépendants, donc leur covariance est nulle. Pour le même tirage, $B_{\ell,i}B_{\ell,j} = 0$, tandis que leurs espérances valent p_i et p_j . Ainsi

$$\text{Cov}(B_{\ell,i}, B_{\ell,j}) = 0 - p_i p_j = -p_i p_j.$$

Par bilinéarité et sommation finie,

$$\text{Cov}(X_i, X_j) = \sum_{\ell=1}^n \sum_{m=1}^n \text{Cov}(B_{\ell,i}, B_{m,j}) = -n p_i p_j.$$

La première méthode exploite la loi de la somme demandée ; la seconde explique la covariance négative par l'incompatibilité des couleurs à chaque tirage.

B. Cas de deux couleurs

1.

Comme $X_1 + X_2 = n$ et $p_1 + p_2 = 1$,

$$X_2 - n p_2 = n - X_1 - n(1 - p_1) = -(X_1 - n p_1).$$

Il vient

$$U_n = (X_1 - np_1)^2 \left(\frac{1}{np_1} + \frac{1}{np_2} \right) = \frac{(X_1 - np_1)^2}{np_1 p_2} = Z_1^2.$$

Les dénominateurs sont strictement positifs puisque $n \geq 2$ et $p_1, p_2 > 0$.

2.a)

La variable X_1 est une somme de n variables de Bernoulli indépendantes, de même paramètre p_1 . Comme $0 < p_1 < 1$, leur espérance est p_1 et leur variance $p_1(1 - p_1) = p_1 p_2$ est strictement positive.

Le théorème limite central s'applique et donne

$$Z_1 = \frac{X_1 - np_1}{\sqrt{np_1 p_2}} \longrightarrow G \quad \text{en loi,} \quad G \sim \mathcal{N}(0, 1).$$

Pour n grand, on peut donc approcher la loi de Z_1 par la loi normale centrée réduite. Cette approximation suppose que les proportions sont fixées lorsque n augmente.

2.b)

Méthode 1 — Composition par une fonction continue

D'après 2.a), Z_1 converge en loi vers $G \sim \mathcal{N}(0, 1)$. La fonction $h : x \mapsto x^2$ est continue sur $\mathbb{R}$: le résultat du programme sur la composition de la convergence en loi s'applique.

Ainsi $U_n = Z_1^2$ converge en loi vers G^2 . D'après la partie 1, question 3.a), G^2 suit la loi du χ^2 à 1 degré de liberté.

Toutes les hypothèses de cette composition sont vérifiées. Aucun résultat sur la somme de variables convergeant séparément en loi n'est utilisé.

Méthode 2 — Fonctions de répartition et traitement des bornes

Notons H_n la fonction de répartition de Z_1 . Le théorème limite central donne $H_n(a) \rightarrow \Phi(a)$ pour tout réel a , puisque Φ est continue partout.

Pour une borne réelle a , il faut aussi traiter la probabilité stricte $P(Z_1 < a)$. Pour tout $\eta > 0$,

$$H_n(a - \eta) \leq P(Z_1 < a) \leq H_n(a).$$

En passant aux limites inférieure et supérieure puis en faisant tendre η vers 0, la continuité de Φ donne $P(Z_1 < a) \rightarrow \Phi(a)$.

Pour $x > 0$,

$$P(U_n \leq x) = P(-\sqrt{x} \leq Z_1 \leq \sqrt{x}) = H_n(\sqrt{x}) - P(Z_1 < -\sqrt{x}).$$

Par les convergences précédentes, cette probabilité tend vers $\Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = F_1(x)$.

Pour $x < 0$, elle vaut 0, comme $F_1(x)$. Pour $x = 0$ et tout $\eta > 0$,

$$0 \leq P(U_n \leq 0) = P(Z_1 = 0) \leq H_n(\eta) - H_n(-\eta).$$

Le majorant tend vers $\Phi(\eta) - \Phi(-\eta)$, puis vers 0 lorsque $\eta \rightarrow 0$. On obtient donc aussi la convergence vers $F_1(0) = 0$.

La fonction F_1 étant continue sur $\mathbb{R}$, ces limites prouvent la convergence en loi annoncée. Cette seconde méthode montre pourquoi on ne doit pas oublier les éventuels atomes de la variable discrète Z_1 aux bornes de l'intervalle.

C. Cas de trois couleurs avec proportions $1/4, 1/4, 1/2$

1.

Posons $A = X_1 - n/4$ et $B = X_2 - n/4$. Comme $X_1 + X_2 + X_3 = n$, on a $X_3 - n/2 = -(A + B)$. Par définition,

$$U_n = \frac{4}{n}(A^2 + B^2) + \frac{2}{n}(A + B)^2.$$

D'autre part,

$$Z_1^2 = \frac{4}{n}(A + B)^2, \quad Z_2^2 = \frac{2}{n}(A - B)^2.$$

En développant,

$$\frac{4}{n}(A + B)^2 + \frac{2}{n}(A - B)^2 = \frac{1}{n}(6A^2 + 4AB + 6B^2).$$

Le développement de U_n donne le même polynôme. Ainsi $U_n = Z_1^2 + Z_2^2$.

2.

Toutes les variables de comptage sont bornées, donc les moments et covariances utilisés existent. Les espérances de $X_3 - n/2$ et de $X_1 - X_2$ sont nulles. Par conséquent,

$$E(Z_1) = E(Z_2) = 0.$$

Comme $X_3 \sim \mathcal{B}(n, 1/2)$,

$$V(Z_1) = \frac{4}{n} V(X_3) = \frac{4}{n} \times \frac{n}{4} = 1.$$

Pour les deux autres comptages, la partie A donne

$$V(X_1) = V(X_2) = \frac{3n}{16}, \quad \text{Cov}(X_1, X_2) = -\frac{n}{16}.$$

Il en résulte

$$V(X_1 - X_2) = \frac{3n}{16} + \frac{3n}{16} - 2\left(-\frac{n}{16}\right) = \frac{n}{2},$$

puis $V(Z_2) = (2/n)(n/2) = 1$.

Enfin,

$$\text{Cov}(X_3, X_1) = \text{Cov}(X_3, X_2) = -\frac{n}{8}.$$

Par bilinéarité,

$$\text{Cov}(Z_1, Z_2) = \frac{2\sqrt{2}}{n} (\text{Cov}(X_3, X_1) - \text{Cov}(X_3, X_2)) = 0.$$

Les deux variables sont centrées, de variance 1 et non corrélées. Cette dernière propriété ne démontre pas leur indépendance.

3.

La variable X_3 est une somme de n Bernoulli indépendantes de paramètre $1/2$, dont la variance vaut $1/4 > 0$. Le théorème limite central donne

$$\frac{X_3 - n/2}{\sqrt{n/4}} = Z_1 \longrightarrow \mathcal{N}(0, 1) \quad \text{en loi.}$$

Pour n grand, la loi de Z_1 est donc approchée par la loi normale centrée réduite.

4.a)

À chaque tirage, T_i ajoute 1 au comptage de C_1 , retranche 1 au comptage de C_2 et ne contribue pas pour C_3 .

En sommant les contributions sur les n tirages,

$$X_1 - X_2 = \sum_{i=1}^n T_i.$$

4.b)

Les T_i sont des fonctions des tirages indépendants ; elles sont donc indépendantes et de même loi. Leur loi est

$$P(T_i = 1) = \frac{1}{4}, \quad P(T_i = -1) = \frac{1}{4}, \quad P(T_i = 0) = \frac{1}{2}.$$

Ces variables sont bornées. On calcule

$$E(T_i) = \frac{1}{4} - \frac{1}{4} = 0, \quad E(T_i^2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

donc $V(T_i) = 1/2 > 0$.

Le théorème limite central s'applique et donne

$$\frac{\sum_{i=1}^n T_i}{\sqrt{n/2}} \rightarrow \mathcal{N}(0, 1) \quad \text{en loi.}$$

Par 4.a), le membre de gauche vaut $\sqrt{2/n}(X_1 - X_2) = Z_2$. Pour n grand, la loi de Z_2 est donc approchée par la loi normale centrée réduite.

5.

Question adaptée au programme actuel — La demande d'une loi exacte de U_n est remplacée par l'étude de sa loi limite, avec un résultat particulier admis. Les deux convergences marginales précédentes et la covariance nulle ne suffisent pas à établir cette limite à elles seules.

Résultat admis dans cette version adaptée : pour deux variables G_1, G_2 indépendantes de loi normale centrée réduite,

$$\forall x \in \mathbb{R}, \quad \lim_{n \rightarrow +\infty} P(Z_1^2 + Z_2^2 \leq x) = P(G_1^2 + G_2^2 \leq x).$$

La question C.1 donne $U_n = Z_1^2 + Z_2^2$. La partie 1, question 3.b), montre que $G_1^2 + G_2^2$ suit la loi du χ^2 à 2 degrés de liberté. Sa fonction de répartition est continue, puisqu'elle admet une densité. Le résultat admis donne donc, par définition,

$$U_n \longrightarrow \chi^2 \text{ à 2 degrés de liberté en loi.}$$

On peut approcher la loi de U_n par cette loi lorsque n est grand.

Pour un entier n fixé, les comptages X_1, X_2, X_3 ne prennent qu'un nombre fini de valeurs. La variable U_n a donc une loi discrète à support fini et ne peut pas avoir exactement la loi à densité définie en partie 1.

On peut également vérifier que Z_1 et Z_2 ne sont pas indépendantes à taille finie. L'événement $\{X_3 = n\}$ a pour probabilité $2^{-n} > 0$ et équivaut à $\{Z_1 = \sqrt{n}\}$. Sur cet événement, $X_1 = X_2 = 0$, donc $Z_2 = 0$.

Cependant, l'événement où tous les tirages donnent C_1 , de probabilité $4^{-n} > 0$, donne $Z_2 = \sqrt{2n} \neq 0$. Ainsi $P(Z_2 = 0) < 1$, alors que $P(Z_2 = 0 \mid Z_1 = \sqrt{n}) = 1$. Leur indépendance est impossible.

Le résultat admis concerne donc bien une limite de probabilités, et non une affirmation d'indépendance ou de normalité exacte pour un n fini.

6.a)

Code adapté en Python. On affecte les valeurs aléatoires 0 et 1 respectivement aux couleurs C_1 et C_2 , et les valeurs 2 ou 3 à la couleur C_3 .

Python

```
import numpy as np

def Tirage():
    C = np.zeros(100, dtype=int)
    for i in range(100):
        r = np.random.randint(0, 4)
        if r == 0:
            C[i] = 1
        elif r == 1:
            C[i] = 2
        else:
            C[i] = 3
    return C
```

La borne supérieure de `randint` est exclue : chacune des valeurs 0, 1, 2, 3 a pour probabilité $1/4$. Les probabilités des couleurs sont donc exactement $1/4$, $1/4$ et $1/2$.

La boucle parcourt les indices 0 à 99, soit exactement 100 tirages. Les appels successifs simulent les tirages indépendants requis. `dtype=int` demande un tableau d'entiers.

Ce code a été exécuté. La taille du tableau et l'appartenance de toutes ses valeurs à $\{1, 2, 3\}$ ont été contrôlées ; des répétitions ont donné des fréquences proches des proportions attendues. Ces fréquences constituent une vérification numérique, et non la preuve des probabilités, qui résulte de l'affectation des quatre valeurs équiprobables.

6.b)

Le code reprend directement la contribution définie pour T_i en C.4 :

Python

```
def Difference(C):
    d = 0
    for i in range(100):
        if C[i] == 1:
            d = d + 1
        elif C[i] == 2:
            d = d - 1
    return d
```

La couleur 3 n'a aucune contribution. À la fin de la boucle, d vaut donc le nombre de 1 moins le nombre de 2, c'est-à-dire $X_1 - X_2$.

La fonction prend en entrée un tableau de longueur 100, comme dans le sujet, et ne nécessite aucun import supplémentaire. Elle a été exécutée sur des tableaux simulés et sur des cas déterministes : un tableau ne contenant que des 1 donne 100, un tableau ne contenant que des 2 donne -100 , et un tableau ne contenant que des 3 donne 0.

D. Nombre quelconque de couleurs

1.

Les dénominateurs $\sqrt{np_i}$ sont strictement positifs. Les variables Y_i sont centrées, puisque $E(X_i) = np_i$, et elles admettent des moments d'ordre 2.

Pour $i = j$, la partie A donne

$$V(Y_i) = \frac{np_i(1 - p_i)}{np_i} = 1 - p_i.$$

Pour $i \neq j$,

$$\text{Cov}(Y_i, Y_j) = \frac{-np_i p_j}{\sqrt{np_i} \sqrt{np_j}} = -\sqrt{p_i p_j}.$$

La matrice $I - N$ a précisément pour coefficients diagonaux $1 - p_i$ et pour coefficients hors diagonale $-\sqrt{p_i p_j}$. Ainsi

$$M = I - N.$$

2.

Posons le vecteur colonne

$$v = \begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_s} \end{pmatrix}.$$

On a $N = v^t v$ et

$${}^t v v = \sum_{i=1}^s p_i = 1.$$

Par associativité du produit matriciel,

$$N^2 = v^t v v^t v = v({}^t v v)^t v = v^t v = N.$$

Pour tout vecteur colonne $u \in \mathbb{R}^s$,

$$Nu = v({}^t v u) \in \text{Vect}(v).$$

Ainsi $\text{Im } N \subset \text{Vect}(v)$. Comme $Nv = v$ et $v \neq 0$, l'image contient v et donc toute sa droite vectorielle. Finalement,

$$\text{Im } N = \text{Vect}(v), \quad \text{rg } N = 1.$$

On peut aussi lire N comme le projecteur orthogonal sur $\text{Vect}(v)$, car v est de norme 1 et $Nu = \langle v, u \rangle v$.

3.

Méthode 1 — Théorème spectral et espaces propres

La matrice $N = v^t v$ est symétrique, donc $M = I - N$ est symétrique réelle.

Le vecteur v vérifie $Nv = v$, d'où $Mv = 0$. Plus généralement, $Mu = 0$ équivaut à $u = Nu$, donc à $u \in \text{Vect}(v)$. L'espace propre de M pour la valeur 0 est ainsi de dimension 1.

D'autre part, $Mu = u$ équivaut à $Nu = 0$, donc à ${}^t v u = 0$, puisque $v \neq 0$. L'espace propre pour la valeur 1 est $v^\perp$, de dimension $s - 1$.

Ces deux espaces propres sont supplémentaires et de dimensions totales s . Le théorème spectral pour les matrices symétriques réelles fournit une base orthonormale de vecteurs propres. On la range avec les $s - 1$ vecteurs de valeur propre 1 en premier et le vecteur de valeur propre 0 en dernier.

Si Q a pour colonnes les vecteurs de cette base, alors ${}^t Q Q = I$ et

$$M = Q J_{s-1} {}^t Q.$$

Aucun déterminant ni polynôme caractéristique n'est nécessaire pour identifier les valeurs propres.

Méthode 2 — Construction d'une base orthonormale adaptée

Le vecteur v est de norme 1. Son orthogonal $v^\perp$ est un espace euclidien de dimension $s - 1$; il possède une base orthonormale $(q_1, \dots, q_{s-1})$.

La famille $(q_1, \dots, q_{s-1}, v)$ est une base orthonormale de $\mathbb{R}^s$. Prenons Q dont les colonnes sont ces vecteurs, avec $q_s = v$. Les produits scalaires des colonnes donnent ${}^tQQ = I$.

Pour $i \leq s - 1$, ${}^tvq_i = 0$, donc $Nq_i = 0$ et $Mq_i = q_i$. Pour la dernière colonne, $Mv = 0$. Ces relations s'écrivent

$$MQ = QJ_{s-1}.$$

La matrice Q est inversible, car ses colonnes forment une base. L'égalité ${}^tQQ = I$ implique $Q^{-1} = {}^tQ$. En multipliant $MQ = QJ_{s-1}$ à droite par tQ , on obtient la factorisation demandée.

La première méthode applique le théorème spectral après avoir déterminé les espaces propres ; la seconde construit directement une base adaptée à la droite $\text{Vect}(v)$ et à son orthogonal.

4.

Par définition du produit matrice-vecteur, pour chaque $i \in \llbracket 1, s \rrbracket$,

$$Z_i = \sum_{j=1}^s a_{i,j} Y_j.$$

Tous les Y_j sont centrés et admettent des moments d'ordre 2. Les sommes sont finies, donc les Z_i admettent aussi des moments d'ordre 2, et la linéarité donne

$$E(Z_i) = \sum_{j=1}^s a_{i,j} E(Y_j) = 0.$$

Notons C_Z la matrice de covariance de $Z_1, \dots, Z_s$. Par bilinéarité, pour tout couple (i, ℓ) ,

$$\text{Cov}(Z_i, Z_\ell) = \sum_{j=1}^s \sum_{k=1}^s a_{i,j} a_{\ell,k} \text{Cov}(Y_j, Y_k).$$

C'est le coefficient de ligne i et de colonne ℓ du produit tQMQ . Ainsi

$$C_Z = {}^tQMQ = {}^tQ(QJ_{s-1}{}^tQ)Q = J_{s-1},$$

car ${}^tQQ = I$.

On obtient donc $V(Z_i) = 1$ pour $i \leq s-1$, $V(Z_s) = 0$, et des covariances nulles entre variables distinctes. Ces covariances nulles ne suffisent pas à conclure à l'indépendance.

Comme Z_s est centrée et de variance nulle, $E(Z_s^2) = 0$. La variable positive Z_s^2 est donc nulle presque sûrement :

$$Z_s = 0 \quad \text{presque sûrement.}$$

Avec le choix de la deuxième méthode de D.3, on peut même lire cette égalité directement : la dernière colonne de Q étant v ,

$$Z_s = {}^tvY = \sum_{j=1}^s \sqrt{p_j} \frac{X_j - np_j}{\sqrt{np_j}} = \frac{\sum_{j=1}^s X_j - n \sum_{j=1}^s p_j}{\sqrt{n}} = 0.$$

5.

Question adaptée au programme actuel — L'approximation informelle de l'énoncé original est remplacée par un résultat particulier portant sur des probabilités de sommes de carrés. La conclusion demandée est une convergence en loi, et non une loi exacte à taille finie.

Les colonnes de Q forment une base orthonormale. En particulier, $Q^{-1} = {}^tQ$ et $Q^tQ = I$. Comme $Z = {}^tQY$, on a $Y = QZ$. Pour chaque réalisation,

$$\sum_{i=1}^s Y_i^2 = {}^tYY = {}^tZ^tQQZ = {}^tZZ = \sum_{i=1}^s Z_i^2.$$

Par définition de U_n et par D.4),

$$U_n = \sum_{i=1}^s Y_i^2 = \sum_{i=1}^{s-1} Z_i^2 \quad \text{presque sûrement.}$$

Résultat admis dans cette version adaptée : pour des variables $G_1, \dots, G_{s-1}$ mutuellement indépendantes et normales centrées réduites,

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

La partie 1, question 3.b), établit que

$$\sum_{i=1}^{s-1} G_i^2$$

suit la loi du χ^2 à $s - 1$ degrés de liberté. Sa fonction de répartition est continue. L'égalité presque sûre précédente et le résultat admis donnent donc

$$U_n \longrightarrow \chi^2 \text{ à } s - 1 \text{ degrés de liberté} \quad \text{en loi.}$$

Pour n grand, cette loi à densité constitue la loi approchée de U_n . Pour un n fixé, U_n reste une variable à valeurs dans un ensemble fini : cette approximation ne devient pas une identité de lois.

Le nombre $s - 1$ vient de la dimension de l'orthogonal à v , ou, de façon équivalente, de la contrainte exacte

$$\sum X_i = n,$$

qui donne une coordonnée Z_s nulle.

Partie 3

Question adaptée au programme actuel — La règle de décision est explicitée dans le sujet, ainsi que le modèle de naissances indépendantes. On applique cette règle à la loi approchée établie en partie 2, sans supposer connue une théorie générale des tests statistiques. Les égalités numériques de l'original sont écrites comme des approximations.

Notons $n = 8\,000$ et, pour chaque $i \in \llbracket 1, 4 \rrbracket$, X_i le nombre de naissances observées dans la période T_i .

Sous l'hypothèse de répartition uniforme des jours de naissance, la probabilité d'une naissance dans T_i est la proportion de journées p_i donnée par le sujet. Avec le modèle de naissances indépendantes, les comptages relèvent du modèle de la partie 2 avec $s = 4$.

On définit donc

$$U_n = \sum_{i=1}^4 \frac{(X_i - np_i)^2}{np_i}.$$

Les effectifs observés sont $x_i = nf_i$. Ils sont cohérents avec le total de l'échantillon :

Période	Effectif observé x_i	Effectif attendu np_i
T_1	2 112	2 000
T_2	1 384	1 352

Période	Effectif observé x_i	Effectif attendu np_i
T_3	1 272	1 336
T_4	3 232	3 312

Les deux colonnes totalisent 8 000. Pour cette réalisation, la valeur de la statistique est

$$u_{\text{obs}} = \sum_{i=1}^4 \frac{(nf_i - np_i)^2}{np_i} = n \sum_{i=1}^4 \frac{(f_i - p_i)^2}{p_i} \simeq 12,03.$$

Sous l'hypothèse étudiée, la partie 2, question D.5, permet d'approcher la loi de U_n par la loi du χ^2 à $4 - 1 = 3$ degrés de liberté. Le seuil donné vérifie

$$P(U_n > 11,3) \simeq 1 - F_3(11,3) \simeq 0,01.$$

Ainsi, sous l'hypothèse uniforme, un dépassement de ce seuil est un événement de probabilité voisine de 1 % dans l'approximation retenue.

Or $u_{\text{obs}} \simeq 12,03 > 11,3$. La réalisation appartient donc à la région de rejet définie dans le sujet. On rejette l'hypothèse de répartition uniforme selon cette règle, au seuil nominal de 1 % avec l'approximation utilisée.

Cela signifie que l'échantillon est peu compatible avec le modèle uniforme et indépendant retenu. Ce calcul ne donne pas la probabilité que l'hypothèse elle-même soit vraie ou fausse, et n'établit pas une impossibilité absolue du modèle.

Relevé des adaptations

– Partie 1, préambule : les degrés de liberté sont précisés comme entiers strictement positifs, et la loi définie est reliée explicitement à une loi gamma de taux 1. Ses propriétés sont établies dans le corrigé, sans les supposer connues comme un cours sur la loi du χ^2 .

- Partie 1, 2.a) et 2.b) : précision $n \geq 1$, nécessaire pour les factorielles, les bornes de somme et la densité utilisée.
- Partie 1, 2.c) : code adapté en Python, avec prise en compte des valeurs réelles $x \leq 0$.
- Partie 1, 3.a) : indication du changement de variable non affine $t = u^2$, utilisé pour établir $\Gamma(1/2) = \sqrt{\pi}$ et identifier la densité.
- Partie 2, préambule : indépendance des tirages avec remise rendue explicite pour les lois binomiales et les applications du théorème limite central.
- Partie 2, C.5) : rectification de la demande d'une loi exacte pour U_n . À taille finie, U_n est discrète ; la conclusion devient une loi limite et une approximation. Une convergence particulière des probabilités de sommes de carrés est admise, car les convergences marginales et une covariance nulle ne permettent pas de déduire la limite demandée.
- Partie 2, C.6.a) et C.6.b) : code adapté en Python ; le tableau est indexé de 0 à 99, et la fonction de tirage renvoie le tableau simulé. Les probabilités et le calcul de $X_1 - X_2$ sont conservés.
- Partie 2, D.5) : l'approximation gaussienne informelle de l'original est remplacée par une convergence particulière admise pour la somme des carrés. La loi du χ^2 est la loi limite et la loi approchée, sans affirmation de normalité ou d'indépendance exactes des coordonnées à taille finie.
- Partie 3 : explicitation de la règle de rejet et de l'indépendance dans le modèle, afin de traiter l'exemple avec les probabilités et la convergence établies dans le problème. Les valeurs numériques arrondies sont notées avec $\simeq$.