Corrigé du devoir n^o 6

Partie I

1. (a) $J_1 = \displaystyle\int_0^{+\infty} x^{-\frac{1}{2}} e^{-\frac{x}{2}}\ \mathrm{d}x$

   J_1 est une intégrale doublement impropre, en posant $u = \sqrt{x}$, on obtient : $\mathrm{d}u = \dfrac{\mathrm{d}x}{2\sqrt{x}}$, $\quad e^{-\frac{x}{2}} = e^{-\frac{u^2}{2}}$, puis $J_1 = \displaystyle\int_0^{+\infty} 2\,e^{-\frac{u^2}{2}}\ \mathrm{d}u$. On reconnaît une intégrale de Gauss sur $\mathbb{R}^+$, donc convergente, et par parité de l'intégrande, $J_1 = \dfrac{1}{2}\displaystyle\int_{-\infty}^{+\infty} 2\,e^{-\frac{u^2}{2}}\ \mathrm{d}u = \sqrt{2\,\pi}$

   (b) X^2 est à valeurs dans $\mathbb{R}^+$ et $\forall x \geqslant 0$, $\left[X^2 \leqslant x\right] = \left[-\sqrt{x} \leqslant X \leqslant \sqrt{x}\right]$, donc
   $P\left(X^2 \leqslant x\right) = P\left(-\sqrt{x} \leqslant X \leqslant \sqrt{x}\right) = P\left(-\sqrt{x} < X \leqslant \sqrt{x}\right) = \Phi\left(\sqrt{x}\right) - \Phi\left(-\sqrt{x}\right)$

   Comme $\Phi\left(-\sqrt{x}\right) = 1 - \Phi\left(\sqrt{x}\right)$, on en déduit $P\left(X^2 \leqslant x\right) = 2\,\Phi\left(\sqrt{x}\right) - 1 = \dfrac{2}{\sqrt{2\,\pi}}\displaystyle\int_{-\infty}^{\sqrt{x}} e^{-\frac{t^2}{2}}\,\mathrm{d}t\ - 1$

   Une densité de X^2 sur $\mathbb{R}^+$ est donc donnée par $f_{X^2}(t) = \dfrac{1}{\sqrt{2\,\pi\,t}}\,e^{-\frac{t}{2}}$; sur $\mathbb{R}^-$ une densité nulle convient.

   Ainsi, d'après la question précédente et le changement de variable effectué dans J_1, on reconnaît une loi du χ^2 à un degré de liberté.

2. (a) $J_2 = \displaystyle\int_0^{+\infty} e^{-\frac{x}{2}}\ \mathrm{d}x = 2$; on reconnaît le double de la densité d'une loi exponentielle de paramètre $\frac{1}{2}$.

   Ainsi la loi du χ^2 à deux degrés de liberté est une loi exponentielle de paramètre $\frac{1}{2}$.

   (b) $X^2 + Y^2$ est à valeurs dans $\mathbb{R}^+$, donc sur $\mathbb{R}^-$ on peut prendre 0 comme densité, et pour $s \geqslant 0$, une densité f_S est donnée par $f_S(s) = \displaystyle\int_{-\infty}^{+\infty} f_{X^2}(t)\,f_{Y^2}(s-t)\mathrm{d}t = \int_0^{+\infty} f_{X^2}(t)\,f_{Y^2}(s-t)\mathrm{d}t$ puisque f_{X^2} est nulle sur $\mathbb{R}^-$.

   D'autre part, $f_{Y^2}(s-t) = 0$ si $s-t < 0$, donc $f_S(s) = \displaystyle\int_0^s f_{X^2}(t)\,f_{Y^2}(s-t)\mathrm{d}t = \dfrac{1}{\left(\sqrt{2\,\pi}\right)^2}\int_0^s \dfrac{e^{-\frac{t}{2}}}{\sqrt{t}}\,\dfrac{e^{-\frac{s-t}{2}}}{\sqrt{s-t}}\,\mathrm{d}t$

   $f_S(s) = \dfrac{e^{-\frac{s}{2}}}{2\,\pi}\displaystyle\int_0^s \dfrac{1}{\sqrt{t\,(s-t)}}\,\mathrm{d}t$ avec $t = s\,\cos^2(u), \mathrm{d}t = -2\,s\,\sin u\,\cos u\,\mathrm{d}u$ et par exemple, $u = \dfrac{\pi}{2}$ lorsque $t = 0$ et $u = 0$ pour $t = s$. On obtient alors $t\,(s-t) = s^2\cos^2(u)\,\sin^2(u)$ et comme on a choisi $u \in \left[0\,,\,\dfrac{\pi}{2}\right]$, on a $\sqrt{t\,(s-t)} = s\,\cos u\,\sin u$ et $f_S(s) = \dfrac{e^{-\frac{s}{2}}}{2\,\pi}\displaystyle\int_{\frac{\pi}{2}}^0 -2\,\mathrm{d}u = \dfrac{e^{-\frac{s}{2}}}{2}$

   On reconnaît une loi du χ^2 à deux degrés de liberté

Partie II

1. On calcule $\|(1,1,1,1)\| = 2$, donc on peut choisir $\overrightarrow{\varepsilon_4} = \frac{1}{2}\,(1,1,1,1)$.

   Soit $\overrightarrow{u} = (x,y,z,t) \in \mathbb{R}^4$, on a $p(\overrightarrow{u}) = (\overrightarrow{u}\cdot\overrightarrow{\varepsilon_4})\,\overrightarrow{\varepsilon_4} = \left(\dfrac{x+y+z+t}{2}\right)\overrightarrow{\varepsilon_4}$

2. On obtient alors $N = \dfrac{1}{4}\begin{pmatrix}1&1&1&1\\1&1&1&1\\1&1&1&1\\1&1&1&1\end{pmatrix}$

3. $\overrightarrow{u}(x,y,z,t) \in \operatorname{Ker}(p) \iff \overrightarrow{u}\cdot\overrightarrow{\varepsilon_4} = 0$ soit $x+y+z+t = 0$.

4. (a) Pour montrer que $(\overrightarrow{\varepsilon_1}, \overrightarrow{\varepsilon_2}, \overrightarrow{\varepsilon_3})$ est libre, on peut :
   - ⋆ Soit remarquer que leurs coordonnées dans la base canonique sont échelonnées ;
   - ⋆ Soit remarquer que c'est une famille orthogonale, et même orthonormale.

   Ensuite, on constate que $\overrightarrow{\varepsilon_1}, \overrightarrow{\varepsilon_2}, \overrightarrow{\varepsilon_3}$ sont dans $\operatorname{Ker}(p)$, donc orthogonaux à $\overrightarrow{\varepsilon_4}$, ainsi $\overrightarrow{\varepsilon_4} \notin \operatorname{Vect}\langle\overrightarrow{\varepsilon_1}, \overrightarrow{\varepsilon_2}, \overrightarrow{\varepsilon_3}\rangle$ et la famille des 4 vecteurs est libre ; on peut également remarquer qu'elle est orthonormale, ce qui suffit à prouver que c'est une base.

(b) Comme $p(\overrightarrow{\varepsilon_1}) = p(\overrightarrow{\varepsilon_2}) = p(\overrightarrow{\varepsilon_3}) = \vec{0}$ et que $p(\overrightarrow{\varepsilon_4}) = \varepsilon_4$, on en déduit $D = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, enfin Q est la matrice de passage de la base canonique $\mathscr{B}$ à la base $\mathscr{C}$; comme on a remarqué que $\mathscr{C}$ est orthonormale, il en résulte que Q est orthogonale donc $Q^{-1} = {}^tQ$ et $Q = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/2\sqrt{3} & 1/2 \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/2\sqrt{3} & 1/2 \\ 0 & -2/\sqrt{6} & 1/2\sqrt{3} & 1/2 \\ 0 & 0 & -\sqrt{3}/2 & 1/2 \end{pmatrix}$

Partie III

A. Étude des variables X_i

1. On a affaire à un schéma de Bernoulli, avec probabilité de succès égale à p_i, donc $\forall i \in [\![1, s]\!],\ X_i \hookrightarrow \mathscr{B}(n, p_i)$
Ainsi, d'après le cours, $E(X_i) = n\,p_i$ et $V(X_i) = n\,p_i\,(1-p_i)$
2. $X_i + X_j$ suit également une loi binomiale, mais de paramètres n et $p_i + p_j$, ainsi d'après les règles de calcul sur la variance :
$V(X_i + X_j) = V(X_i) + V(X_j) + 2\operatorname{cov}(X_i, X_j)$
$n\,(p_i + p_j)\,(1 - p_i - p_j) = n\,p_i\,(1-p_i) + n\,p_j\,(1-p_j) + 2\operatorname{cov}(X_i, X_j)$
d'où $\boxed{\operatorname{cov}(X_i, X_j) = -n\,p_i\,p_j}$

B. On suppose dans cette partie que s = 2

1. On remarque que $X_1 + X_2 = n$ et $p_1 + p_2 = 1$, donc
$U_n = \dfrac{(X_1 - n\,p_1)^2}{n\,p_1} + \dfrac{(X_2 - n\,p_2)^2}{n\,p_2} = \dfrac{(X_1 - n\,p_1)^2}{n\,p_1} + \dfrac{(n - X_1 - n\,(1-p_1))^2}{n\,p_2} = (X_1 - n\,p_1)^2 \left(\dfrac{1}{n\,p_1} + \dfrac{1}{n\,p_2}\right)$
Finalement
$$\boxed{U_n = \frac{(X_1 - n\,p_1)^2}{n\,p_1\,p_2} = Z_1^2}$$
2. Soit B_i la variable égale à 1 si le i^{e} tirage donne une boule de couleur C_1 et 0 sinon ; on constate que les variables B_i, $1 \leqslant i \leqslant n$ sont indépendantes, de même loi de Bernoulli de paramètre p_1.
Ainsi par le théorème central limite, la variable centrée réduite associée à X_1, c'est à dire Z_1, converge en loi vers une variable de loi normale centrée réduite.

C. On suppose dans cette partie que $s = 3$, $p_1 = p_2 = \dfrac{1}{4}$ et $p_3 = \dfrac{1}{2}$

On pose $Z_1 = \dfrac{2}{\sqrt{n}}\left(X_3 - \dfrac{n}{2}\right)$ et $Z_2 = \sqrt{\dfrac{2}{n}}\,(X_1 - X_2)$

1. $Z_1^2 = \dfrac{4}{n}\left(X_3 - \dfrac{n}{2}\right)^2 = \dfrac{4}{n}\left(n - X_1 - X_2 - \dfrac{n}{2}\right)^2 = \dfrac{4}{n}\left[\left(\dfrac{n}{4} - X_1\right) + \left(\dfrac{n}{4} - X_2\right)\right]^2$
Or $\left[\left(\dfrac{n}{4} - X_1\right) + \left(\dfrac{n}{4} - X_2\right)\right]^2 = \left(\dfrac{n}{4} - X_1\right)^2 + \left(\dfrac{n}{4} - X_2\right)^2 + 2\left(\dfrac{n}{4} - X_1\right)\left(\dfrac{n}{4} - X_2\right)$
De même $Z_2^2 = \dfrac{2}{n}\left(X_1 - \dfrac{n}{4} + \dfrac{n}{4} - X_2\right)^2 = \dfrac{2}{n}\left[\left(X_1 - \dfrac{n}{4}\right)^2 + \left(\dfrac{n}{4} - X_2\right)^2 - 2\left(X_1 - \dfrac{n}{4}\right)\left(X_2 - \dfrac{n}{4}\right)\right]$
$$\begin{aligned} \text{Donc } Z_1^2 + Z_2^2 &= \frac{4}{n}\left(\frac{n}{4} - X_1\right)^2 + \frac{4}{n}\left(\frac{n}{4} - X_2\right)^2 + \frac{8}{n}\left(\frac{n}{4} - X_1\right)\left(\frac{n}{4} - X_2\right) \\ &+ \frac{2}{n}\left(X_1 - \frac{n}{4}\right)^2 + \frac{2}{n}\left(\frac{n}{4} - X_2\right)^2 - \frac{4}{n}\left(X_1 - \frac{n}{4}\right)\left(X_2 - \frac{n}{4}\right) \\ &= \frac{4}{n}\left(X_1 - \frac{n}{4}\right)^2 + \frac{4}{n}\left(\frac{n}{4} - X_2\right)^2 + \underbrace{\frac{2}{n}\left(X_1 - \frac{n}{4}\right)^2 + \frac{2}{n}\left(X_2 - \frac{n}{4}\right)^2 + \frac{4}{n}\left(X_1 - \frac{n}{4}\right)\left(X_2 - \frac{n}{4}\right)}_{\frac{2}{n}(X_1 + X_2 - \frac{n}{2})^2} \\ &= \frac{4}{n}\left(X_1 - \frac{n}{4}\right)^2 + \frac{4}{n}\left(X_2 - \frac{n}{4}\right)^2 + \frac{2}{n}\left(X_3 - \frac{n}{2}\right)^2 = U_n \text{ (car } X_3 = n - X_1 - X_2) \end{aligned}$$
2. $E(Z_1) = \dfrac{2}{\sqrt{n}}E\left(X_3 - \dfrac{n}{2}\right) = \dfrac{2}{\sqrt{n}}\left(\dfrac{n}{2} - \dfrac{n}{2}\right) = 0$ et $E(Z_2) = \sqrt{\dfrac{2}{n}}\,E\left[(X_1) - E\left(X_2\right)\right] = 0$

On constate que les variables Z_1 et Z_2 sont centrées.

$\operatorname{cov}(Z_1, Z_2) = \dfrac{4}{\sqrt{n}}\sqrt{\dfrac{2}{n}}\Big(\operatorname{cov}(X_3, X_1) - \operatorname{cov}(X_3, X_2)\Big) = 0$

$V(Z_1) = \dfrac{4}{n}V\,(X_3) = 1$ et $V(Z_2) = \dfrac{2}{n}\,(V\,(X_1) + V\,(X_2) - 2\operatorname{cov}(X_1, X_2)) = \dfrac{2}{n}\left(\left(\dfrac{3\,n}{16}\right) + \left(\dfrac{3\,n}{16}\right) + 2\,\dfrac{n}{16}\right) = 1$

Z_1 et Z_2 sont centrées réduites et non corrélées.

3. On peut à nouveau appliquer le théorème central limite à la variable centrée réduite associée à X_3, à savoir Z_1, et approcher sa loi par une loi normale centrée réduite.

4. (a) i. $X_1 - X_2 = \displaystyle\sum_{i=1}^{n} T_i$ *(on compte positivement chaque boule de couleur 1, négativement chaque boule de couleur 2, et 0 pour les autres).*

   ii. Les variables T_i sont indépendantes, de même loi, donc on peut ici encore appliquer le théorème central limite à la variable centrée réduite associée, soit Z_2.

5. (a)

```python
from random import *
def TIRAGE(n):
    C=[]
    for k in range(n):
        t=random()
        if t<0.25:
            C.append(1)
        elif t<0.5:
            C.append(2)
        else:
            C.append(3)
    return C
```

(b)

```python
def DIFFERENCE(C):
    n=len(C)
    Un=0
    Deux=0
    for k in range(n):
        if C[k]==1:
            Un+=1
        elif C[k]==2:
            Deux+=1
    return n-Un-Deux,Un-Deux
```

(c)

```python
def SIMULE_Un(C):
    n=len(C)
    Trois=DIFFERENCE(C)[0]
    return ((4/n)*(Trois-(n/2))**2))+((2/n)*(DIFFERENCE(C))**2)
```

Les lois de Z_1 et de Z_2 sont approchées par des lois normales centrées réduites, de plus Z_1 et Z_2 sont non corrélées et même indépendantes, donc U_n suit approximativement une loi du χ^2 à deux degrés de liberté, c'est à dire aussi une loi exponentielle de paramètre $\frac{1}{2}$

D. On suppose désormais que $s = 4$ **et** $p_1 = p_2 = p_3 = p_4 = \dfrac{1}{4}$

1. Comme tous les p_i sont égaux, on a pour $i \neq j$, $\operatorname{cov}(Y_i, Y_j) = \dfrac{1}{n\,p}\operatorname{cov}(X_i, X_j) = -\dfrac{n\,p^2}{n\,p} = -\dfrac{1}{4}$

   Et pour $i = j$, $\operatorname{cov}(Y_i, Y_i) = V(Y_i) = \dfrac{1}{n\,p}V(X_i) = \dfrac{n\,p\,(1-p)}{n\,p} = \dfrac{3}{4}$

   On observe que $M = I_4 - N$

2. (a)

$$\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \\ 1/2\sqrt{3} & 1/2\sqrt{3} & 1/2\sqrt{3} & -\sqrt{3}/2 \\ 1/2 & 1/2 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} \quad \text{donc}$$

$$\begin{cases} Z_1 &= \dfrac{Y_1}{\sqrt{2}} - \dfrac{Y_2}{\sqrt{2}} \\[2ex] Z_2 &= \dfrac{Y_1}{\sqrt{6}} + \dfrac{Y_2}{\sqrt{6}} - \dfrac{2\,Y_3}{\sqrt{6}} \\[2ex] Z_3 &= \dfrac{Y_1}{\sqrt{2}} + \dfrac{Y_2}{\sqrt{2}} + \dfrac{Y_3}{\sqrt{2}} - \dfrac{\sqrt{3}\,Y_4}{2} \\[2ex] Z_4 &= \dfrac{Y_1}{2} + \dfrac{Y_2}{2} + \dfrac{Y_3}{2} + \dfrac{Y_4}{2} = 0 \quad \text{car} \quad X_1 + X_2 + X_3 + X_4 = n \end{cases}$$

   (b) Les variables X_i ont toutes la même espérance égale à $\frac{n}{4}$ donc toutes les variables Y_i sont centrées, et par combinaisons linéaires des Y_i, toutes les variables Z_i le sont également.

   (c) Par définition, $U_n = Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2$, or $\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} = Q \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix}$ donc

$$Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2 = \begin{pmatrix} Y_1 & Y_2 & Y_3 & Y_4 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} = \begin{pmatrix} Z_1 & Z_2 & Z_3 & Z_4 \end{pmatrix} \underbrace{{}^tQ\,Q}_{I_4} \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} = Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2.$$

   Comme $Z_4 = 0$, on obtient

$$\boxed{U_n = Z_1^2 + Z_2^2 + Z_3^2}$$

Partie IV

Pour tout $i \in [\![1, 4]\!]$, soit X_i la variable aléatoire égale au nombre d'occurrences du nombre i ; $X_i \hookrightarrow \mathscr{B}\left(10000, \frac{1}{4}\right)$.
$U_n = Z_1^2 + Z_2^2 + Z_3^2$ suit approximativement une loi du χ^2 à 3 degrés de liberté, donc d'après les valeurs fournies, $P\left(U_n \leqslant 7,81\right) = 0,95$. Or les résultats des 10000 tirages donnent : $Y_1^2 = \dfrac{(2602-2500)^2}{2500} = \dfrac{102^2}{2500}$,
$Y_2^2 = \dfrac{(2534-2500)^2}{2500} = \dfrac{34^2}{2500}$, $Y_3^2 = \dfrac{(2422-2500)^2}{2500} = \dfrac{78^2}{2500}$, $Y_4^2 = \dfrac{(2442-2500)^2}{2500} = \dfrac{58^2}{2500}$
Et finalement $U_n \simeq 8,4032$; donc on peut rejeter l'hypothèse de répartition uniforme avec une probabilité supérieure à $0,95$.
Pour une valeur plus précise il faudrait disposer d'une table de lois du χ^2, assez semblable à celle de la loi normale. On verrait que $F_3(9,35) = 0,975$ et on pourrait faire une interpolation linéaire pour évaluer plus précisément la probabilité de rejet, en gardant à l'esprit toutefois que la loi du χ^2 est déjà une approximation de celle de U_n.