## Exercice 1 : test du Khi-deux - sujet A

1. On obtient à la calculatrice : $\overline{x_e} \approx 4,528$ à $10^{-3}$ près par excès
   et $v_e \approx 4,505$ à $10^{-3}$ près par excès.

2. On constate que $\overline{x_e}$ et $v_e$ sont proches. Pour une variable aléatoire $X$ suivant une loi de Poisson de paramètre $\lambda$, $E(X) = V(X) = \lambda$.

   $\lambda$ est donc la valeur moyenne de $X$, paramètre inconnu que l'on estime ponctuellement par $\overline{x_e} \approx 4,5$.

3. On va tester l'hypothèse nulle $H_0$ contre l'hypothèse alternative $H_1$ :

$$\begin{cases} H_0 : & \text{la variable aléatoire } X \text{ suit la loi de Poisson } \mathscr{P}(\lambda), \\ & \text{les observations suivent la distribution théorique spécifiée,} \\ H_1 : & X \text{ ne suit pas la loi } \mathscr{P}(\lambda), \\ & \text{les observations ne suivent pas la distribution théorique spécifiée.} \end{cases}$$

4. (a) Dans le test du Khi-deux, chaque effectif théorique $c_i$ doit être supérieur ou égal à 5. Si cette condition n'est pas satisfaite, il y a lieu de regrouper deux ou plusieurs classes adjacentes. Avant regroupement : $c_0 = 2.00$ et $c_9 = 4.17$, $c_{10} = 1.88$ , $c_{11} = 0.77$ , $c_{12} = 0.29$ , $c_{13+} = 0.14$

   Le regroupement s'est effectué sur les deux premières classes de la distribution et sur les 5 dernières classes (car $c_{13} < c_{12} < c_{11} < c_{10} < c_9 < 5$)
   $k$ représente donc le nombre de classes après regroupement : $k = 9$.

   On dispose automatiquement de la relation $\displaystyle\sum_{i=1}^{9} c_i = 180$

   (b) $c_6 = n\,p_6$ avec $n = 180$ et $p_6 = P([X = x_6]) = P(X = 6) = \mathrm{e}^{-\lambda}\,\dfrac{\lambda^6}{6!}$

$$\text{D'où} \quad c_6 = 180 \times \mathrm{e}^{-4,5} \times \frac{4,5^6}{6!} \approx 23,06$$

5. (a) On prend pour variable aléatoire de test : $\displaystyle Z = \sum_{i=1}^{9} \frac{(N_i - c_i)^2}{c_i}$

   où $N_i$ est la variable aléatoire qui compte le nombre d'intervalles de 1 minute (parmi 180 intervalles), qui prennent $i$ clients.

   $Z$ suit approximativement la loi du Khi-deux à $k - 1 - r$ degrés de liberté où $k$ est le nombre de valeurs observées après regroupement et $r$ le nombre de paramètres de la loi de $X$ qu'il a fallu estimer. Ici $r = 1$.

   Donc $Z$ suit approximativement la loi $\chi^2_7$.

   (b)

$$\chi^2_{obs} = \frac{(n_1 - c_1)^2}{c_1} + \frac{(n_2 - c_2)^2}{c_2} + \cdots + \frac{(n_9 - c_9)^2}{c_9} \approx 1,05$$

   (c) On cherche le réel positif $z_\alpha$ tel que $P(Z \leqslant z_\alpha) = 95\%$.
   On obtient à la calculatrice : $z_\alpha \approx 14,067$.
   Pour l'échantillon observé, $\chi^2_{obs} < z_\alpha$. L'écart observé n'est pas significatif au seuil $\alpha$. Cet écart est imputable aux fluctuations d'échantillonnage.
   On ne peut pas rejeter l'hypothèse $H_0$ selon laquelle $X$ suit une loi de Poisson.

## Exercice 2 : 6 questions indépendantes - sujet B

1. Pour avoir des tirages équiprobables avec des jetons 2 à 2 distincts, on décide de numéroter artificiellement les jetons. On choisit comme univers $\Omega$ l'ensemble des combinaisons de 3 jetons pris dans l'ensemble $\{B1, B2, C1, C2, C3, R1, R2, R3, R4\}$ des neuf jetons et on fait l'hypothèse d'équiprobabilité sur $\Omega$. Le nombre total de tirages possibles est
   $\text{Card}(\Omega) = \dbinom{9}{3} = \dfrac{9 \times 8 \times 7}{3!} = 3 \times 4 \times 7 = 84$

   Soit $A$ l'événement : «tirer 3 jetons de 3 couleurs différentes». Alors

$$P(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)} = \frac{2 \times 3 \times 4}{3 \times 4 \times 7} = \frac{2}{7}$$

2. L'expérience aléatoire consiste à répéter 315 fois la même épreuve de Bernoulli dans des conditions identiques et indépendantes dont les issues contraires sont : un étudiant appelé sur APB se présente le jour de la rentrée à l'UTBM («succès» de probabilité $p = 0,8$) ou ne se présente pas le jour de la rentrée («échec» de probabilité $1 - p$).

   On sait alors que la variable aléatoire $X$ qui compte le nombre d'étudiants présents le jour de la rentrée sur les 315 appelés (c'est-à-dire le nombre de «succès») suit la loi binomiale de paramètres $n = 315$ et $p = 0,8$.

   «L'UTBM se retrouve avec plus de 250 étudiants» est l'événement $[X \geqslant 251]$.

$$P(X \geqslant 251) = 1 - P(X \leqslant 250) \approx 0,589 \quad \text{obtenu à la calculatrice.}$$

Comme $n > 30$, $np = 252 > 15$ et $np(1-p) = 50,4 > 5$, on aurait pu remplacer la loi de $X$ par la loi normale $\mathcal{N}\left(np\,,\,\sqrt{np(1-p)}\right)$.

Soit $Y$ une variable aléatoire suivant la loi $\mathcal{N}\left(252\,,\,\sqrt{50,4}\right)$.

En notant $Y^* = \dfrac{Y - np}{\sqrt{np(1-p)}} = \dfrac{Y-252}{\sqrt{50,4}}$ la variable centrée réduite associée à $Y$, on sait que $Y^* \hookrightarrow \mathcal{N}(0,1)$. Avec la *correction de continuité*, on obtient :

$$\begin{array}{rcl} P(X \geqslant 251) & \approx & P(Y > 250,5) \\ & = & 1 - P(Y \leqslant 250,5) \\ & = & P\left(Y^* \leqslant \dfrac{250,5-252}{\sqrt{50,4}}\right) \\ & = & 1 - \Phi\left(-\dfrac{1,5}{\sqrt{50,4}}\right) \\ & = & \Phi\left(\dfrac{1,5}{\sqrt{50,4}}\right) \\ & \approx & 0,584 \end{array}$$

3. On reconnaît la loi des tirages simultanés sans remise : $X$ suit la loi hypergéométrique $\mathcal{H}(N,n,p)$ avec $N = 32$, $n = 4$ et $p = \dfrac{4}{32} = \dfrac{1}{8}$.

   Alors $X(\Omega) = [\![0\,,\,4]\!]$ et pour tout entier $k$ tel que $0 \leqslant k \leqslant 4$,

$$P([X=k]) = \frac{\binom{4}{k}\binom{32-4}{4-k}}{\binom{32}{4}} = \frac{\binom{4}{k}\binom{28}{4-k}}{35960}$$

   Avec un logiciel de calcul formel, on en déduit le tableau :

| $k$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P(X=k)$ | $\dfrac{4095}{7192}$ | $\dfrac{1638}{4495}$ | $\dfrac{567}{8990}$ | $\dfrac{14}{4495}$ | $\dfrac{1}{35960}$ |

4. $Y \hookrightarrow \mathcal{N}(\mu\,,\,\sigma)$ avec $\mu = 1000$ et $\sigma$ à déterminer.
   On pose $U = \dfrac{Y-\mu}{\sigma}$ alors $U \hookrightarrow \mathcal{N}(0\,,\,1)$.
   On sait que 10% des paquets de sucre pèsent plus de 1020 grammes, ce que l'on traduit en termes de probabilités par :

   $P(Y > 1020) = 0,1$ d'où $1 - P(Y \leqslant 1020) = 0,1$ d'où $P(Y \leqslant 1020) = 0,9$
   d'où $P\left(U \leqslant \dfrac{20}{\sigma}\right) = 0,9$ puis $\dfrac{20}{\sigma} = \Phi^{-1}(0,9)$

   et enfin $\boxed{\sigma = \dfrac{20}{\Phi^{-1}(0,9)} \approx 15,61}$

5. Soit $t \in \mathbb{R}^+$. Puisque les composants sont montés en série,

$$[Z > t] = [T_1 > t] \cap [T_2 > t]$$

   Notons $F_1$ et $F_2$ les fonctions de répartition respectives des variables aléatoires $T_1$ et $T_2$. On sait que $F_1(t) = P(T_1 \leqslant t) = \displaystyle\int_0^t \lambda\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x = 1 - \mathrm{e}^{-\lambda\,t}$.
   De même $F_2(t) = 1 - \mathrm{e}^{-\mu t}$. Donc

$$\begin{array}{rcll} P\big([Z>t]\big) & = & P\big([T_1 > t] \cap [T_2 > t]\big) & \\ & = & P\big([T_1>t]\big)\,P\big([T_2>t]\big) & \text{par indépendance de } T_1 \text{ et } T_2 \\ & = & (1-F_1(t))\,(1-F_2(t)) & \\ & = & \mathrm{e}^{-\lambda t}\,\mathrm{e}^{-\mu t} & \\ & = & \mathrm{e}^{-(\lambda+\mu)t} & \end{array}$$

6. $X(\Omega) = Y(\Omega) = \mathbb{N}^*$. Donc $[X=Y] = \displaystyle\bigcup_{n=1}^{+\infty}([X=n]\cap[Y=n])$.

   Comme les événements $[X=n]\cap[Y=n]$ sont deux à deux incompatibles, nous avons
   $P\big([X=Y]\big) = P\left(\displaystyle\bigcup_{n=1}^{+\infty}([X=n]\cap[Y=n])\right) = \displaystyle\sum_{n=1}^{+\infty} P\big([X=n]\cap[Y=n]\big)$
   Or les variables aléatoires $X$ et $Y$ sont indépendantes.
   Donc $\forall\, n \in \mathbb{N}^*$, $P\big([X=n]\cap[Y=n]\big) = P(X=n)\,P(Y=n)$

   D'où $P\big([X=Y]\big) = \displaystyle\sum_{n=1}^{+\infty} p\,(1-p)^{n-1} \times p\,(1-p)^{n-1} = \sum_{n=1}^{+\infty}\left(\frac{1}{2}\right)^{2n}$ car $p = 1-p = 1/2$.
   Ainsi $P\big([X=Y]\big) = \displaystyle\sum_{n=1}^{+\infty}\left(\frac{1}{4}\right)^n = \frac{1}{4}\sum_{n=0}^{+\infty}\left(\frac{1}{4}\right)^n = \frac{1}{4}\times\frac{1}{1-\frac{1}{4}} = \frac{1}{4}\times\frac{4}{3}$

$$\boxed{P\big([X=Y]\big) = \frac{1}{3}}$$

## Exercice 2 : 5 questions indépendantes - sujet A

1. $P(A) = \dfrac{2\times 4\times 6}{\binom{12}{3}} = \dfrac{2\times 4\times 6}{2\times 11\times 10} = \dfrac{12}{55}$ avec les combinaisons

   ou bien $P(A) = 6\left(\dfrac{2}{12}\times\dfrac{4}{11}\times\dfrac{6}{10}\right)$ avec un arbre pondéré et 6 chemins.

2. Le nombre de présents : $X \hookrightarrow \mathscr{B}(n,p)$ avec $n = 300$ et $p = 0,85$.

$$P(X \geqslant 251) = 1 - P(X \leqslant 250) \approx 0,7693$$

   Avec approximation normale si $Y \hookrightarrow \mathcal{N}\left(255\,,\,\sqrt{38,25}\right)$ :
   $P(X \geqslant 251) \approx P(Y > 250,5) \approx 0,7666$

3. $X \hookrightarrow \mathcal{H}(N,n,p)$ avec $N = 32$, $n = 4$ et $p = \dfrac{8}{32} = \dfrac{1}{4}$.

$$\forall\, k \in [\![0\,,\,4]\!],\; P([X=k]) = \frac{\binom{8}{k}\binom{32-8}{4-k}}{\binom{32}{4}} = \frac{\binom{8}{k}\binom{24}{4-k}}{35960}$$

4. $Y \hookrightarrow \mathcal{N}(\mu\,,\,\sigma)$ avec $\mu = 1000$. $\quad U = \dfrac{Y-\mu}{\sigma} \hookrightarrow \mathcal{N}(0\,,\,1)$

$$P(Y > 1010) = 0,2 \Rightarrow P(Y \leqslant 1010) = 0,8 \Rightarrow P\left(U \leqslant \frac{10}{\sigma}\right) = 0,8$$

$$\sigma = \frac{10}{\Phi^{-1}(0,8)} \approx 11,8818...$$

5. $P\big([X=Y]\big) = P\left(\bigcup_{n=1}^{+\infty}([X=n]\cap[Y=n])\right) = \sum_{n=1}^{+\infty} P\big([X=n]\cap[Y=n]\big)$

   Par indépendance de $X$ et $Y$

$$P\big([X=Y]\big) = \sum_{n=1}^{+\infty} p\,(1-p)^{n-1}\times p\,(1-p)^{n-1} = \sum_{n=1}^{+\infty}\left(\frac{1}{2}\right)^{2n}$$

$$= \sum_{n=1}^{+\infty}\left(\frac{1}{4}\right)^{n} = \frac{1}{4}\sum_{n=0}^{+\infty}\left(\frac{1}{4}\right)^{n} = \frac{1}{4}\times\frac{1}{1-\frac{1}{4}} = \frac{1}{4}\times\frac{4}{3} = \frac{1}{3}$$

## Exercice 3 : estimation

1. (a) Soit $k \in [\![1\,,n]\!]$. $X_k$ est une variable aléatoire discrète qui ne prend que deux valeurs : 0 et 1. Elle suit alors une loi de Bernoulli dont on note $\theta$ le paramètre. On a donc $\theta = P(X_k = 1)$.

   On note $A_k$ l'événement : «la $k-$ième personne a obtenu la face 1 lors du lancer du dé.»

   En appliquant la formule des probabilités totales avec le système complet d'événements $(A_k\,,\,\overline{A_k})$, on obtient :

$$\begin{aligned}\theta = P(X_k=1) &= P(A_k)\,P_{A_k}(X_k=1) + P(\overline{A_k})\,P_{\overline{A_k}}(X_k=1)\\ &= \frac{1}{6}\times p + \frac{5}{6}\times(1-p)\\ &= \frac{5}{6} - \frac{4}{6}p\end{aligned}$$

   Donc $\boxed{\theta = \dfrac{5-4p}{6}}$ puis $\boxed{p = \dfrac{5-6\theta}{4} = \dfrac{5}{4} - \dfrac{3}{2}\theta}$

   (b) $S_n = \sum_{k=1}^{n} X_k$ apparaît comme la somme de $n$ variables aléatoires mutuellement indépendantes, suivant toutes la même loi de Bernoulli de paramètre $\theta$. Donc par stabilité pour la somme, $S_n$ suit la loi binomiale de paramètres $n$ et $\theta$ : $\boxed{S_n \hookrightarrow \mathscr{B}(n,\theta)}$

   On en déduit que $E(S_n) = n\,\theta$ et que $V(S_n) = n\,\theta\,(1-\theta)$.

2. (a)
   - Par linéarité de l'espérance, $E(F_n) = E\left(\dfrac{S_n}{n}\right) = \dfrac{1}{n}E(S_n) = \theta$ ce qui signifie que $F_n$ est un estimateur sans biais de $\theta$.
   - Puisque que $F_n$ est un estimateur sans biais de $\theta$, pour montrer qu'il est convergent, **il suffit** de prouver que $\lim_{n\to+\infty} V(F_n) = 0$.

$$V(F_n) = V\left(\frac{S_n}{n}\right) = \frac{1}{n^2}V(S_n) = \frac{1}{n^2}n\,\theta\,(1-\theta) = \frac{\theta\,(1-\theta)}{n} \underset{(n\to+\infty)}{\longrightarrow} 0$$

   (b) On a exprimé précédemment $p$ en fonction de $\theta$ : $\quad p = \dfrac{5}{4} - \dfrac{3}{2}\theta$

   ce qui conduit à définir, à partir de $F_n$, un estimateur pour $p$ à savoir :

$$\boxed{T_n = \frac{5}{4} - \frac{3}{2}F_n} \qquad \text{et aussi } T_n = \frac{5}{4} - \frac{3}{2n}S_n$$

Par propriété de l'espérance : $E(T_n) = \dfrac{5}{4} - \dfrac{3}{2}E(F_n) = \dfrac{5}{4} - \dfrac{3}{2}\theta = p$
ce qui signifie que $T_n$ est un estimateur sans biais de $p$.

De plus, par propriété de la variance,

$$V(T_n) = V\left(\frac{5}{4} - \frac{3}{2}S_n\right) = V\left(-\frac{3}{2}F_n\right) = \frac{9}{4}V(F_n) \underset{(n\to+\infty)}{\longrightarrow} 0$$

Donc $T_n$ est un estimateur sans biais et convergent de $p$.

(c) $V(T_n) = \dfrac{9}{4}V(F_n) = \dfrac{9\,\theta\,(1-\theta)}{4n} = \dfrac{9}{4n}\left(\dfrac{5-4p}{6}\right)\left(1 - \dfrac{5-4p}{6}\right)$
$= \dfrac{9}{4n}\left(\dfrac{5-4p}{6}\right)\left(\dfrac{1+4p}{6}\right)$

D'où $\boxed{V(T_n) = \dfrac{(5-4p)(1+4p)}{16\,n}} = \dfrac{1}{n}\left(\dfrac{5}{16} + p - p^2\right)$

(d) L'énoncé dit que $n > 30$, $n\theta \geqslant 15$ et que $n\theta(1-\theta) > 5$.
D'après le théorème central limite, la variable aléatoire $S_n$ suit approximativement la loi normale $\mathcal{N}\left(n\theta\,,\ \sqrt{n\,\theta\,(1-\theta)}\right)$.
Or l'image par une fonction affine non constante d'une variable aléatoire suivant une loi normale, suit aussi une loi normale.

Donc $T_n = \dfrac{5}{4} - \dfrac{3}{2n}S_n$ suit approximativement la loi normale
$\mathcal{N}\left(p\,,\ \dfrac{1}{4}\sqrt{\dfrac{(5-4p)(1+4p)}{n}}\right)$

3. Sur $n = 1000$ personnes interrogées, le sondeur a recueilli $S_n(\omega) = 425$ avis favorables.

(a) Estimations ponctuelles : $\widehat{\theta} = F_n(\omega) = \dfrac{S_n(\omega)}{n} = \dfrac{425}{1000} = 0,425$.

D'où $\widehat{p} = T_n(\omega) = \dfrac{5}{4} - \dfrac{3}{2}F_n(\omega) = \dfrac{5}{4} - \dfrac{3}{2}\widehat{\theta} = 1,25 - 1,5 \times 0,425 = 0,6125$

(b) On prend $\alpha = 10\%$. Posons $Z_n = \dfrac{T_n - E(T_n)}{\sqrt{V(T_n)}} = \dfrac{T_n - p}{\frac{1}{4}\sqrt{\frac{(5-4p)(1+4p)}{n}}}$.

Alors la suite $(Z_n)$ converge en loi vers une variable aléatoire $U$ de loi normale centrée réduite.
On cherche le réel $t_\alpha$ tel que $P(-t_\alpha \leqslant U \leqslant t_\alpha) = 1 - \alpha$ c'est-à-dire
$\Phi(t_\alpha) - \Phi(-t_\alpha) = 1 - \alpha \iff 2\,\Phi(t_\alpha) - 1 = 1 - \alpha \iff \Phi(t_\alpha) = 1 - \dfrac{\alpha}{2}$

On obtient à la calculatrice $\boxed{t_\alpha = \Phi^{-1}(0,95) \approx 1,6449}$. On en déduit que

$$P\left(T_n - \frac{t_\alpha}{4}\sqrt{\frac{(5-4p)(1+4p)}{n}} \leqslant p \leqslant T_n + \frac{t_\alpha}{4}\sqrt{\frac{(5-4p)(1+4p)}{n}}\right) \underset{(n\to+\infty)}{\longrightarrow} 90\%$$

Ainsi un intervalle de confiance (asymptotique) observé pour $p$ au niveau de confiance 90% est :

$$\left[\widehat{p} - \frac{t_\alpha}{4}\sqrt{\frac{(5-4\widehat{p})(1+4\widehat{p})}{n}}\,,\ \widehat{p} + \frac{t_\alpha}{4}\sqrt{\frac{(5-4\widehat{p})(1+4\widehat{p})}{n}}\right] = [57.39\%, 65.11\%]$$

Remarque : à ne pas confondre avec un intervalle de confiance observé pour $\theta$ au niveau de confiance 90% qui est :

$$\left[\widehat{\theta} - t_\alpha\sqrt{\frac{\widehat{\theta}(1-\widehat{\theta})}{n}}\,,\ \widehat{\theta} + t_\alpha\sqrt{\frac{\widehat{\theta}(1-\widehat{\theta})}{n}}\right] = [39.92\%, 45.08\%]$$