

Corrigé

Total sur 16,5 points - dont rédaction/présentation/clarté : 1 point

Dans tout le sujet, concernant les codes Python, on supposera l'importation suivante faite :

```
import numpy.random as rd
```

Questions indépendantes

8 points

1. Enoncer les inégalités de Markov et de Bienaymé-Tchebychev pour une variable aléatoire X 1 point

Voir cours.

2. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires, telle que, pour tout $n \in \mathbb{N}^*$, $X_n \hookrightarrow \mathcal{E}\left(1 + \frac{1}{n}\right)$ 2 points

Montrer que $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire Z à déterminer.

Avec F_n la fonction de répartition de la loi exponentielle de paramètre $1 + \frac{1}{n}$

et pour $x \in \mathbb{R}_+$, $F_n(x) = 1 - \exp\left(-\left(1 + \frac{1}{n}\right)x\right)$

or $1 + \frac{1}{n} \rightarrow 0$ donc par opérations et continuité de l'exponentielle $\lim_{n \rightarrow +\infty} F_n(x) = 1 - e^{-x}$

par ailleurs $\forall x < 0, F_n(x) = 0$ donc $\lim_{n \rightarrow +\infty} F_n(x) = 0$

finalement $\lim_{n \rightarrow +\infty} F_n(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-x} & \text{si } x \geq 0 \end{cases}$, autrement dit $\forall x \in \mathbb{R}, \lim_{n \rightarrow +\infty} P(X_n \leq x) = F(x)$ où F est

la fonction de répartition de la loi exponentielle de paramètre 1 donc

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X \hookrightarrow \mathcal{E}(1)$$

3. Soit S_n une variable aléatoire telle que $S_n \hookrightarrow \mathcal{B}(n, p)$ 3 points

Voir cours, approximation d'une loi binomiale par une loi normale, on doit trouver $\mathcal{N}(np, npq)$

- a. Justifier que S_n peut s'écrire $S_n = \sum_{i=1}^n X_i$ où $\forall i \in \llbracket 1, n \rrbracket, X_i \hookrightarrow \mathcal{B}(p)$ 0,5 point

Précision : pour bien faire, il faut le démontrer par récurrence, ici on attendait simplement de dire qu'on peut « décomposer » la loi binomiale en ses n épreuves de Bernoulli identiques et indépendantes, S_n compte le nombre de succès, soit la somme des « 1 » obtenus.

- b. Déterminer S_n^* et justifier que S_n^* peut être approchée par la loi normale centrée réduite. 1,5 points

Précision : on peut donner directement $E(S_n) = np$ et $V(S_n) = npq$ car $S_n \hookrightarrow \mathcal{B}(n, p)$

- c. Déterminer une loi normale permettant d'approcher S_n 1 point

4. On exécute le programme ci-contre. Expliquer son fonctionnement et le résultat que l'on peut attendre, on justifiera. 2 points

La fonction effectue N (petite erreur dans l'énoncé initial, c'était $N - 1$) épreuves de Bernoulli de paramètre p et compte le nombre de succès avec le paramètre c , puis elle renvoie la fréquence de succès, soit la moyenne empirique de l'épreuve aléatoire.

```
def simulation(N,p) :
    c=0
    for i in range(1,N+1):
        c=rd.binomial(1,p)
    return c/N

simulation(1000, 1/3)
```

Enfin `simulation(1000, 1/3)` effectue donc la simulation pour 1 000 expériences de paramètre de succès $\frac{1}{3}$ et d'après la loi faible des grands nombres, on s'attend à ce que le résultat soit proche de $E(X_i) = p = \frac{1}{3}$ (où X_i est la variable aléatoire donnant le résultat de l'épreuve i pour $i \in \llbracket 1, 1000 \rrbracket$ ici).

Exercice

7,5 points

Un institut de sondage a été missionné pour estimer la proportion p de végétariens en France. Il interroge pour cela n français(es).

Puisque le choix des sondé(e)s s'effectue sur une population très grande, on admet que l'expérience peut s'apparenter à une suite de n tirages indépendants avec remise. On note X_n la variable aléatoire égale au nombre de végétariens sondés et on souhaite quantifier à quel point la fréquence $F_n = \frac{X_n}{n}$ approche la proportion p

1. Déterminer la loi de X_n

1 point

X_n compte le nombre de « succès » (une personne sondée est végétarienne) lors de n itérations d'une même épreuve de Bernoulli de paramètre p (chaque personne interrogée a une probabilité p d'être végétarienne et on « compte » 1 si elle l'est). Les épreuves étant de plus indépendantes par hypothèse, alors

$$X_n \hookrightarrow \mathcal{B}(n, p)$$

2. En considérant la fonction $f : x \mapsto x(1-x)$ montrer que $p(1-p) \leq \frac{1}{4}$

1,5 points

f est définie et dérivable sur $\mathbb{R}$ (c'est une fonction polynomiale) et $\forall x \in \mathbb{R}, f(x) = x - x^2$ donc $f'(x) = 1 - 2x$ alors $f'(x) \leq 0 \Leftrightarrow 1 - 2x \leq 0 \Leftrightarrow 2x \geq 1 \Leftrightarrow x \geq \frac{1}{2}$ et de même $f'(x) \geq 0 \Leftrightarrow x \leq \frac{1}{2}$

donc f est croissante sur $]-\infty, \frac{1}{2}]$ et décroissante sur $[\frac{1}{2}, +\infty[$ donc f admet un maximum en $\frac{1}{2}$

or $f\left(\frac{1}{2}\right) = \frac{1}{2} \left(1 - \frac{1}{2}\right) = \frac{1}{4}$ donc $\forall x \in \mathbb{R}, f(x) = x(1-x) \leq \frac{1}{4}$ et a fortiori ($p \in [0, 1]$),

$$p(1-p) \leq \frac{1}{4}$$

Remarque : on peut aussi démontrer cette inégalité avec les propriétés du polynôme ou avec une identité remarquable $p^2 - p + \frac{1}{4} = \left(p - \frac{1}{2}\right)^2 \geq 0$

3. Montrer que, pour tout $\varepsilon > 0$, $P(|F_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$

2 points

D'après l'inégalité de Bienaymé-Tchebychev (X_n admet un moment d'ordre 2 car $X_n(\Omega)$ est fini et donc F_n aussi) : $\forall \varepsilon > 0, P(|F_n - E(F_n)| \geq \varepsilon) \leq \frac{V(F_n)}{\varepsilon^2}$

or $E(F_n) = E\left(\frac{X_n}{n}\right) = \frac{1}{n}E(X_n)$ par linéarité de l'espérance donc $E(F_n) = \frac{1}{n} \times np = p$ par propriété de la loi binomiale ($E(X_n) = np$ et $V(X_n) = np(1-p)$)

par propriété de la variance $V(F_n) = V\left(\frac{X_n}{n}\right) = \frac{1}{n^2}V(X_n) = \frac{1}{n^2} \times np(1-p) = \frac{p(1-p)}{n}$

d'où en injectant dans l'inégalité $P(|F_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}$

or $p(1-p) \leq \frac{1}{4} \Rightarrow \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$ et donc

$$P(|F_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$$

4. En déduire une condition sur n pour que F_n soit une approximation de p à 10^{-2} près avec une probabilité supérieure ou égale à 95%

3 points

On veut donc que l'écart entre F_n et p soit inférieur à 10^{-2} autrement dit on cherche $|F_n - p| \leq 10^{-2}$ et on va donc appliquer l'inégalité avec $\varepsilon = 10^{-2}$; par ailleurs, on cherche une probabilité supérieure à 0,95

or s'intéresse donc à $P(|F_n - p| \leq 10^{-2})$ or $P(|F_n - p| \geq 10^{-2}) = 1 - P(|F_n - p| \leq 10^{-2})$ et d'après la question précédente $P(|F_n - p| \geq 10^{-2}) \leq \frac{1}{4n(10^{-2})^2}$ i.e. $1 - P(|F_n - p| \leq 10^{-2}) \leq \frac{1}{4n10^{-4}}$

soit $1 - \frac{10^4}{4n} \leq P(|F_n - p| \leq 10^{-2})$ et on souhaite que $P(|F_n - p| \leq 10^{-2})$ soit supérieur à 95% ce qui sera

réalisé dès que $1 - \frac{10^4}{4n} \geq 0,95 \Leftrightarrow \frac{10^4}{4n} \leq 0,05 \Leftrightarrow \frac{10^4}{n} \leq 2 \times 10^{-1} \Leftrightarrow n \geq \frac{10^5}{2}$ i.e.

$$n \geq 50\,000$$

Remarque : cela est en fait beaucoup (en France, généralement un échantillon de 1 000 personnes est suffisant pour un sondage). On peut supposer que la loi de Bienaymé-Tchebychev n'est pas la meilleure approximation, on pourrait tester l'approximation par la loi normale $\mathcal{N}(np, npq)$