Correction de la feuille Exo_23 : Inégalités. Loi faible des grands nombres.

Ex 1 : 1) a. F_n est le nombre d'électeurs pour A divisé par le nombre d'électeurs sondé donc

$$\boxed{F_n = \frac{1}{n}\sum_{k=1}^{n} X_k}$$

b. La linéarité de l'espérance donne $E(F_n) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} E(X_k)$ donc $\boxed{E(F_n) = p}$

L'indépendance des X_k donne $V(F_n) = \dfrac{1}{n^2}\displaystyle\sum_{k=1}^{n} V(X_k)$ donc $\boxed{V(F_n) = \dfrac{p(1-p)}{n}}$

L'étude de la fonction $x \longmapsto x(1-x)$ donne l'inégalité usuelle $p(1-p) \leqslant \dfrac{1}{4}$ on en déduit que :

$$\boxed{V(F_n)\ \frac{1}{4n}}$$

2) a. F_n admet une espérance et une variance donc on peut appliquer l'inégalité de Bienaymé-Tchebychev :

$$\text{pour tout } t > 0,\ \mathbb{P}\big(|F_n - E(F_n)| \geqslant t\big) \leqslant \frac{V(F_n)}{t^2}$$

or on a $V(F_n) \leqslant \dfrac{1}{4n}$ donc

$$\boxed{\text{pour tout } t > 0,\ \mathbb{P}\big(|F_n - p| \geqslant t\big) \leqslant \frac{1}{4nt^2}}$$

b. On déduit de la question précédente que pour tout $t > 0$, $\mathbb{P}\big(|F_n - p| < t\big) \geqslant 1 - \dfrac{1}{4nt^2}$

donc en prenant $t = \sqrt{\dfrac{5}{n}}$, on obtient : $\mathbb{P}\left(|F_n - p| < \sqrt{\dfrac{5}{n}}\right) \geqslant \dfrac{95}{100}$

$$\boxed{\left[F_n - \sqrt{\frac{5}{n}}; F_n + \sqrt{\frac{5}{n}}\right] \text{ est un intervalle de confiance au niveau } 95\% \text{ de } p}$$

3) a. On est dans les conditions d'application du théorème central limite donc F_n^* converge en loi vers une variable aléatoire suivant la loi $\mathcal{N}(0,1)$ donc $\lim\limits_{n\to+\infty} \mathbb{P}\left(|F_n^*| \leqslant 2\right) = 2\Phi(2) - 1$

Or d'une part comme $F_n^* = \dfrac{F_n - p}{\sqrt{\frac{p(1-p)}{n}}}$ on a : $\mathbb{P}\left(|F_n^*| \leqslant 2\right) = \mathbb{P}\left(|F_n - p| \leqslant 2\sqrt{\frac{p(1-p)}{n}}\right)$

et d'autre part $2\Phi(2) - 1 \approx 0{,}9545 \geqslant 95\%$

on obtient bien :

$$\boxed{\lim_{n\to+\infty} \mathbb{P}\left(|F_n - p| \leqslant 2\sqrt{\tfrac{p(1-p)}{n}}\right) \geqslant 0{,}95}$$

b. comme $p(1-p) \leqslant \dfrac{1}{4}$ alors $\lim\limits_{n\to+\infty} \mathbb{P}\left(|F_n - p| \leqslant \dfrac{1}{\sqrt{n}}\right) \geqslant 0{,}95$

$$\boxed{\text{pour } n \text{ assez grand, } \left[F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}}\right] \text{ est un intervalle de confiance à } 95\% \text{ de } p}$$

4) On utilise les deux intervalles obtenus aux questions 2) et 3).

a. Pour $n = 100$, on a $f_{100} = 0{,}52$.

— Avec la question 2), on obtient :

$$\left[f_{100} - \sqrt{\frac{5}{100}}\ ;\ f_{100} + \sqrt{\frac{5}{100}}\right] = \left[0{,}52 - \sqrt{0{,}05}\ ;\ 0{,}52 + \sqrt{0{,}05}\right] = [0{,}296\ ;\ 0{,}744]$$

— Avec la question 3), on obtient :
$$\left[f_{100} - \frac{1}{\sqrt{100}}\ ;\ f_{100} + \frac{1}{\sqrt{100}}\right] = [0{,}52 - 0{,}1\ ;\ 0{,}52 + 0{,}1] = [0{,}42\ ;\ 0{,}62]$$

Le second intervalle est nettement plus précis que le premier.

b. Pour $n = 10000$, on a $f_{10000} = 0{,}52$.

— Avec la question 2), on obtient :
$$\left[f_{10000} - \sqrt{\frac{5}{10000}}\ ;\ f_{10000} + \sqrt{\frac{5}{10000}}\right] = [0{,}4976\ ;\ 0{,}5424]$$

— Avec la question 3), on obtient :
$$\left[f_{10000} - \frac{1}{\sqrt{10000}}\ ;\ f_{10000} + \frac{1}{\sqrt{10000}}\right] = [0{,}51\ ;\ 0{,}53]$$

Là encore, le second intervalle est plus précis.

Commentaire général :
Les deux intervalles de confiance sont valables au niveau 95%, mais ils ne reposent pas sur les mêmes outils.

— L'intervalle obtenu avec Bienaymé-Tchebychev est **toujours valable**, sans hypothèse supplémentaire sur la loi, mais il est **très large** : il donne une estimation peu précise de p.

— L'intervalle issu du théorème central limite est **beaucoup plus précis** (plus resserré), mais il repose sur une **approximation asymptotique** : il est pertinent seulement pour n suffisamment grand.

— Lorsque n augmente, les deux intervalles se **resserrent** (en $1/\sqrt{n}$) : l'estimation devient plus fiable.

— En pratique, on privilégie l'intervalle issu du TCL, car il est **nettement plus informatif**, tandis que l?inégalité de Tchebychev sert surtout à garantir un résultat sans hypothèse.

Ex 2 : 1) C'est une question de cours :

$$\boxed{E(X) = 5 \text{ et } V(X) = \frac{5}{2}}$$

2) X admet une variance donc on peut appliquer l'inégalité de Bienaymé-Tchebychev :

$$\forall t > 0, \qquad P\Big(\,|X - E(X)| \geqslant t\,\Big) \leqslant \frac{\sigma^2}{t^2}$$

ce qui donne :

$$\boxed{P\Big(\,|X - 5| \geqslant 4\,\Big) \leqslant \frac{5}{32} \quad (\approx 0{,}16)}$$

3) On sait que : $\Big[|X - 5| \geqslant 4\Big] = \Big[X \leqslant 1\Big]$ ou $\Big[X \geqslant 9\Big]$ et $X(\Omega) = [\![0; 10]\!]$.

donc $P(|X - 5| \geqslant 4) = P(X = 0) + P(X = 1) + P(X = 9) + P(X = 10) = 2 \times (1 + 10)\dfrac{1}{2^{10}}$

$$\boxed{P(|X - 5| \geqslant 4) = \frac{11}{512} \approx 0{,}021}$$

Ex 3 : Cette expérience est constituée de n épreuves de Bernoulli identiques et indépendantes, le succès : "l'individu a un ordinateur" a pour probabilité $\dfrac{1}{2}$

En notant X_n le nombre de sondés ayant un ordinateur, on a $X_n \hookrightarrow \mathscr{B}\left(n, \dfrac{1}{2}\right)$

la proportion de sondés ayant un ordinateur est : $Y_n = \dfrac{X_n}{n}$ et ainsi :

$$E(Y_n) = \frac{1}{n}E(X_n) = \frac{1}{2} \qquad \text{et} \qquad V(Y_n) = \frac{1}{n^2}V(X_n) = \frac{1}{4n}$$

De plus : $P\left(\Big[Y_n \in]0{,}48\,,\,0{,}52[\Big]\right) \;=\; 1 - P(|Y_n - 0{,}5| \geqslant 0{,}02)$

et l'inégalité de Bienaymé-Tchebychev donne :

$$P(|Y_n - E(Y_n)| \geqslant 0{,}02) \leqslant \frac{V(Y_n)}{0{,}02^2}$$

On obtient donc

$$P\left(\left[Y_n \in]0{,}48\,,\,0{,}52[\right]\right) \;\geqslant\; 1 - \frac{1}{4n \times 0{,}02^2}$$

Il nous reste à trouver le premier n pour lequel $\dfrac{1}{4n \times 0{,}02^2} \leqslant \dfrac{1}{2}$

$\boxed{Y_n \text{ se trouve dans l'intervalle }]0{,}48\,,\,0{,}52[\text{, avec une probabilité supérieure à } 0{,}5 \text{ pour } n \geqslant 1250}$

Ex 4 : X suit loi $\mathcal{N}(0,1)$ donc X admet une espérance $E(X) = 0$ et une variance $V(X) = 1$, en appliquant l'inégalité de Bienaymé-Tchebychev on obtient :

$$\forall x \geqslant 0, \quad P(|X| \geqslant x) \leqslant \frac{1}{x^2} \qquad (*)$$

or pour $x > 0$,

$$\begin{aligned}
P(|X| \geqslant x) &= 1 - P(|X| < x) \\
&= 1 - P(|X| \leqslant x) \qquad \textit{(car X est à densité)} \\
&= 1 - P(-x \leqslant X \leqslant x) \\
&= 1 - (\Phi(x) - \Phi(-x)) \qquad \text{en notant } \Phi \text{ la fonction de répartition de la loi } \mathcal{N}(0,1) \\
&= 2\,(1 - \Phi(x)) \\
&= 2\int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,dt
\end{aligned}$$

En reprenant l'inégalité $(*)$ il vient bien :

$$\boxed{\text{Pour tout réel } x > 0, \qquad \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,dt \leqslant \frac{1}{x^2}\sqrt{\frac{\pi}{2}}}$$