Feuille Exo_24 : Inégalités. Loi faible des grands nombres.

Ex 1 : Soit $X \hookrightarrow \mathscr{B}\left(10\,;\,\frac{1}{2}\right)$

1) Donner l'espérance et la variance de X.
2) Quel majorant de $\mathbb{P}(\left[|X-5| \geqslant 4\right])$ obtient-on en appliquant l'inégalité de Bienaymé-Tchebychev ?
3) Calculer $P(\left[|X-5| \geqslant 4\right])$.

Ex 2 : Dans une population, la moitié possède un ordinateur. On fait un sondage auprès de n personnes de cette population. On suppose que la population est suffisamment importante pour assimiler ce sondage à un tirage avec remise. On note Y_n la proportion de sondés ayant un ordinateur. Déterminer à l'aide de l'inégalité de Bienaymé-Tchebychev, une valeur de n à partir de laquelle cette proportion se trouve dans l'intervalle $]0{,}48\,,\,0{,}52[$, avec une probabilité supérieure à 0,5

Ex 3 : En appliquant l'inégalité de Bienaymé-Tchebychev à $X \hookrightarrow \mathcal{N}(0,1)$, montrer que, pour tout $x > 0$,

$$\int_x^{+\infty} e^{-\frac{t^2}{2}}\,\mathrm{d}t \leqslant \frac{1}{x^2}\sqrt{\frac{\pi}{2}}$$

Ex 4 : Soit (X_n) une suite de variables aléatoires mutuellement indépendantes et suivant toutes la même loi uniforme sur $[0,1]$. On note pour $n \in \mathbb{N}*$, $Y_n = \max(X_1, ..., X_n)$.

Montrer que :

$$\forall \varepsilon \in \mathbb{R}_+^*, \quad \lim_{n\to+\infty} P(|Y_n - 1| \geqslant \varepsilon) = 0$$

Ex 5 : Une urne contient 6 boules vertes et 2 boules rouges.
On tire successivement et sans remise une boule de cette urne et on s'arrête dès que l'on voit une boule rouge.
Estimer le nombre moyen de tirages à l'aide d'un programme Python.
(A vérifier par le calcul.)

Ex 6 : On lance n fois un dé normal. Si on sort au moins un as (la face 1), on ne gagne rien et si on n'a jamais eu d'as, on gagne la somme des nombres obtenus. On note X le gain obtenu à la fin de ce jeu.
Pour tout $i \in [\![1;n]\!]$, on note X_i la variable aléatoire définie par X_i prend la valeur 0 si le i ème lancer a donné un as et le numéro sorti au $i$ème lancer dans le cas contraire.

1) Ecrire une fonction Python prenant en entrée un entier n et permettant de simuler la variable aléatoire X.
2) Ecrire une fonction Python prenant en entrée un entier n et retournant une estimation de $\mathbb{E}(X)$.
3) Ecrire un programme Python permettant de tracer sur le même graphique pour n compris entre 1 et 20 :
   - l'expression de l'espérance $\mathbb{E}(X) = 4n \times \dfrac{5^n}{6^n}$
   - une estimation de $\mathbb{E}(X)$ grâce à la fonction de la question b).

Ex 7 : On considère n un entier naturel non nul quelconque.
Une urne contient n boules numérotées de 1 à n. On effectue des tirages (avec remise) tant que le numéro obtenu est strictement inférieur au précédent. On note X le nombre de tirages effectués.

1) Estimer la loi de X.
2) Représenter graphiquement l'estimation précédente.
3) Estimer l'espérance de X par des séries de simulations.
4) On note $u_n = E(X)$,
   Représenter graphiquement u_n en fonction de n.

(Pour chaque n, u_n sera estimé par des séries de simulations).

Ex 8 : On souhaite estimer un paramètre $p \in]0;1[$. On note : $q = 1 - p$. Soit un entier $n \geqslant 1$ fixé.
On considère $X_1, \dots, X_n$ des variables aléatoires, indépendantes, suivant une même loi de Bernoulli de paramètre p et définies sur un même espace probabilités.

On note : $\overline{X_n} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$.

1) a. Justifier que $p(1-p) \leqslant \dfrac{1}{4}$.

   b. En utilisant l'inégalité de Bienaymé-Tchebychev, montrer que l'intervalle $\left[\overline{X_n} - \sqrt{\dfrac{5}{n}}\,;\, \overline{X_n} + \sqrt{\dfrac{5}{n}}\right]$ est un intervalle de confiance de p au niveau de confiance $0,95$.

On cherche par la suite un intervalle de confiance de p au niveau de confiance $0,95$ d'amplitude plus petite.

2) On fixe un réel strictement positif t quelconque et ε un réel strictement positif quelconque.

   a. Établir l'égalité : $P\left(\overline{X_n} - p \geqslant \varepsilon\right) = P\left(e^{nt\overline{X_n}} \geqslant e^{nt(p+\varepsilon)}\right)$.

   b. En utilisant l'inégalité de Markov, établir l'inégalité suivante : $P\left(\overline{X_n} - p \geqslant \varepsilon\right) \leqslant e^{n\left(\ln(pe^t+q) - t(p+\varepsilon)\right)}$.

   c. On admet l'inégalité : $\ln(pe^t + q) - tp \leqslant \dfrac{t^2}{8}$.

   Ainsi, on a l'inégalité suivante : $\forall t \in \mathbb{R}_+, \quad P\left(\overline{X_n} - p \geqslant \varepsilon\right) \leqslant e^{n\left(\frac{t^2}{8} - t\varepsilon\right)}$.

   En déduire l'inégalité : $P\left(\overline{X_n} - p \geqslant \varepsilon\right) \leqslant e^{-2n\varepsilon^2}$.

3) On pose $\overline{Y_n} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(1 - X_k)$. Établir l'inégalité : $P\left(\overline{Y_n} - q \geqslant \varepsilon\right) \leqslant e^{-2n\varepsilon^2}$.

4) Déduire des questions 2)c) et 3. l'inégalité :$\forall \varepsilon \in \mathbb{R}_+^*, \quad P\left(\left|\overline{X_n} - p\right| \geqslant \varepsilon\right) \leqslant 2e^{-2n\varepsilon^2}$.

5) Comment choisir ε pour obtenir un intervalle de confiance de p au niveau de confiance 0,95 ?
L'amplitude de l'intervalle de confiance est-elle plus réduite que celle obtenue à la question 1) b. ?

Ex 9 : On considère deux variables aléatoires indépendantes : U et V suivant, chacune, la loi uniforme sur $[0;1]$.

1) Justifier son existence, puis déterminer une densité f des variables aléatoires U^2 et V^2.

2) On considère la variable aléatoire $Z = U^2 + V^2$. Justifier que Z admet une densité de probabilité, notée h.

3) Écrire un programme permettant de simuler la variable aléatoire Z et d'estimer $P(Z \leq 1)$.

4) a. Montrer que, pour $0 < x \leq 1$, on a : $h(x) = \dfrac{1}{4}\displaystyle\int_0^x \dfrac{1}{\sqrt{x-t}}\dfrac{1}{\sqrt{t}}dt$.

   b. Montrer que, pour $0 < x \leq 1$, on a : $h(x) = \dfrac{1}{4}\displaystyle\int_0^1 \dfrac{1}{\sqrt{1-y}}\dfrac{1}{\sqrt{y}}dy$.

   c. Montrer que, sur $]0;1]$, on a : $h(x) = \dfrac{\pi}{4}$. *(On pourra utiliser le changement de variable $y = sin^2(u)$).*

   d. Interpréter graphiquement le résultat en terme d'aire.

5) On considère une suite de variables aléatoires de Bernoulli $(Y_n)_{n\geq 1}$ mutuellement indépendantes et de même paramètre $\dfrac{\pi}{4}$, et on note $S_n = \dfrac{Y_1 + \cdots + Y_n}{n}$ pour tout entier $n \geq 1$.

   a. Soit $\varepsilon > 0$, déterminer, en fonction de n et ε, une majoration de $P\left(\left|S_n - \dfrac{\pi}{4}\right| \geq \varepsilon\right)$.

   b. En déduire, à partir de quelle valeur de n, il est possible de définir un intervalle de confiance de niveau de confiance $0,95$ de $\dfrac{\pi}{4}$ et d'amplitude 2×10^{-2}.

   c. Á l'aide de la simulation précédente, déterminer un intervalle de confiance de niveau de confiance $0,95$ de $\dfrac{\pi}{4}$ et d'amplitude 2×10^{-2}.