Exercice 16

1. Soit X une variable aléatoire réelle finie et telle que $X(\Omega) \subset \mathbb{R}^+$, et soit $a > 0$. Démontrer l'inégalité de Markov :
$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$
2. Soit X une variable aléatoire réelle finie quelconque, et soit $a > 0$. En utilisant l'inégalité de Markov démontrer l'inégalité de Bienaymé-Tchebychev :
$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{V(X)}{a^2}.$$
On pourra remarquer que $|x - m| \geq a \iff (x - m)^2 \geq a^2$.
3. Soit $X \hookrightarrow \mathcal{B}(n, p)$ et soit $\varepsilon > 0$. Démontrer que
$$\mathbb{P}(|\frac{X}{n} - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}$$
4. Si on lance un dé équilibré n fois, la fréquence d'apparition du chiffre 3 (i.e. le nombre de 3 obtenus divisé par le nombre total de lancers n) est proche de $\frac{1}{6}$. Donner une condition sur n pour pouvoir affirmer avec une fiabilité de 95% que cette fréquence est comprise entre $\frac{1}{6} - \frac{1}{100}$ et $\frac{1}{6} + \frac{1}{100}$.

Solution :

1. Écrivons $X(\Omega) = \{x_1, x_2, \dots, x_n\}$ avec $x_1 > x_2 > \cdots > x_k \geq a > x_{k+1} > \cdots > x_n \geq 0$ pour un certain $n \in \mathbb{N}^*$ et un certain $k \in [\![0, n]\!]$ (avec par convention $k = 0$ si $a > x_1$ et $k = n$ si $x_n > a$ et les sommes vides correspondantes).

   Alors $\mathbb{P}(X \geq a) = \mathbb{P}((X = x_1) \cup (X = x_2) \cup \cdots \cup (X = x_k)) = \sum_{i=1}^{k} \mathbb{P}(X = x_i)$. Dès lors, si on "coupe" le calcul de $\mathbb{E}(X)$ en deux parties :
$$\begin{aligned}\mathbb{E}(X) = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i) &= \sum_{i=1}^{k} x_i \mathbb{P}(X = x_i) + \sum_{i=k+1}^{n} x_i \mathbb{P}(X = x_i) \\ &\geq \sum_{i=1}^{k} x_i \mathbb{P}(X = x_i) \quad (\text{car pour tout } i \in [\![k+1, n]\!],\ x_i \geq 0) \\ &\geq \sum_{i=1}^{k} a \mathbb{P}(X = x_i) \quad (\text{car pour tout } i \in [\![1, k]\!],\ x_i \geq a) \\ &\geq a\mathbb{P}(X \geq a).\end{aligned}$$
   En divisant par $a > 0$ on a finalement $\frac{\mathbb{E}(X)}{a} \geq \mathbb{P}(X \geq a)$.
2. Soit $Y = (X - \mathbb{E}(X))^2$ et soit $b = a^2$. Comme $Y \geq 0$ on a d'après la première question
$$\mathbb{P}(Y \geq b) \leq \frac{\mathbb{E}(Y)}{b}$$

c'est-à-dire
$$\mathbb{P}((X-\mathbb{E}(X))^2 \geq a^2) \leq \frac{\mathbb{E}((X-\mathbb{E}(X))^2)}{a^2} = \frac{V(X)}{a^2}$$
Or comme l'indique la remarque
$$((X-\mathbb{E}(X))^2 \geq a^2) \ = \ (|X-\mathbb{E}(X)| \geq a)$$
donc finalement $\mathbb{P}(|X-\mathbb{E}(X)| \geq a) \leq \dfrac{V(X)}{a^2}$.

3. Appliquons le résultat de la question précédente à $X \hookrightarrow \mathcal{B}(n,p)$ et $a = n\,\varepsilon$. Comme $\mathbb{E}(X) = np$ et $V(X) = np(1-p)$, on a
$$\mathbb{P}(|X-np| \geq n\varepsilon) \leq \frac{np(1-p)}{(n\varepsilon)^2} = \frac{p(1-p)}{n\varepsilon^2}.$$
Enfin, comme pour tout $x \in \mathbb{R}$ on a $|x-np| \geq n\varepsilon \iff |\dfrac{x}{n}-p| \geq \varepsilon$, on a
$$(|X-np| \geq n\varepsilon) \ = \ (|\frac{X}{n}-p| \geq \varepsilon).$$
Finalement, $\mathbb{P}(|\dfrac{X}{n}-p| \geq \varepsilon) \leq \dfrac{p(1-p)}{n\varepsilon^2}$.

4. Si on lance un dé équilibré n fois, le nombre X de chiffres 3 apparus suit une loi binomiale de paramètres n et $1/6$: $X \hookrightarrow \mathcal{B}(n,1/6)$. La fréquence d'apparition du chiffre 3 est donc $\dfrac{X}{n}$, et se doit d'être proche de $p = 1/6$ pour n grand. En utilisant la question précédente avec $\varepsilon = 1/100$ on a :
$$\mathbb{P}(|\frac{X}{n}-\frac{1}{6}| \geq \frac{1}{100}) \leq \frac{\frac{1}{6}(1-\frac{1}{6})}{n(\frac{1}{100})^2} = \frac{50000}{36\,n}.$$
En passant à l'évènement complémentaire on a donc
$$\mathbb{P}(|\frac{X}{n}-\frac{1}{6}| < \frac{1}{100}) = 1-\mathbb{P}(|\frac{X}{n}-\frac{1}{6}| \geq \frac{1}{100}) \geq 1-\frac{50000}{36\,n}.$$
Or
$$\begin{aligned}\left(\left|\frac{X}{n}-\frac{1}{6}\right| < \frac{1}{100}\right) &= \left(-\frac{1}{100} < \frac{X}{n}-\frac{1}{6} < \frac{1}{100}\right)\\ &= \left(\frac{1}{6}-\frac{1}{100} < \frac{X}{n} < \frac{1}{6}+\frac{1}{100}\right)\\ &= \left(\frac{X}{n} \in \left]\frac{1}{6}-\frac{1}{100}, \frac{1}{6}+\frac{1}{100}\right[\right).\end{aligned}$$
Ainsi $\mathbb{P}\left(\dfrac{X}{n} \in \left]\dfrac{1}{6}-\dfrac{1}{100}, \dfrac{1}{6}+\dfrac{1}{100}\right[\right) \geq 1-\dfrac{50000}{36\,n}$.

On peut donc affirmer que $\dfrac{X}{n} \in \left]\dfrac{1}{6}-\dfrac{1}{100}, \dfrac{1}{6}+\dfrac{1}{100}\right[$ avec une fiabilité de 95% dès que $1-\dfrac{50000}{36\,n} \geq 0,95$ soit, après résolution, dès que $n \geq \dfrac{10^6}{36} \simeq 27\,778$.