PROBABILITES :

Soit un entier $N \geqslant 2$. Une urne contient N boules numérotées de 1 à N. On y effectue une suite indéfinie de tirages avec remise. L'expérience est modélisée dans un espace probabilisé $(\Omega, \mathcal{A}, P)$.

Pour tout $k \in [\![1, N]\!]$, on note X_k la variable aléatoire correspondant au numéro du premier tirage pour lequel k boules distinctes ont été obtenues au moins une fois.
En particulier, X_N représente le numéro du premier tirage pour lequel toutes les boules de l'urne sont sorties au moins une fois.

1. a) On note $T_1 = X_1$ et, pour tout $k \in [\![2, N]\!]$, $T_k = X_k - X_{k-1}$. Montrer que T_k suit une loi géométrique dont on précisera le paramètre p_k.

   b) En déduire l'espérance et la variance de T_k.

Dans la suite, on pourra admettre que les variables T_k sont mutuellement indépendantes.

2. Montrer que X_N admet une espérance et une variance, et déterminer $E(X_N)$ et $V(X_N)$.

3. Pour tout entier $n \geq 1$, on note $H_n = \sum\limits_{k=1}^{n} \frac{1}{k}$.

   a) Montrer que : $\forall n \in \mathbb{N}^*, \ln(n+1) \leqslant H_n \leqslant 1 + \ln n$.

   b) En déduire que $E(X_N) \underset{(N \to \infty)}{\sim} \ln(N)$.

4. a) Montrer que $V(X_N) \leqslant 2N^2$.

   b) Montrer que, pour tout réel $\alpha > 0$, $P\big(|X_N - NH_N| \geqslant \alpha N\big) \leqslant \dfrac{2}{\alpha^2}$.

CORRIGE :

1. a) La variable T_k correspond au nombre de tirages nécessaires pour obtenir une nouvelle boule à partir du moment où $(k-1)$ boules distinctes ont déjà été tirées.

Cela correspond au temps d'attente d'un premier succès dans un processus de Bernoulli dont la probabilité de succès vaut $\frac{N+1-k}{N}$, qui est la proportion de boules qui n'ont jamais été tirées.

Donc T_k suit une loi géométrique de paramètre $p_k = \frac{N+1-k}{N}$.

b) On en déduit que :

$$E(T_k) = \frac{1}{p_k} = \frac{N}{N+1-k} \text{ et } V(T_k) = \frac{(1-p_k)}{p_k^2} = \frac{(k-1)N}{(N+1-k)^2}.$$

2. Comme $X_n = \sum_{k=1}^{N} T_k$, par linéarité de l'espérance,

$$E(X_N) = \sum_{k=1}^{N} E(T_k) = \sum_{k=1}^{N} \frac{N}{(N+1-k)} = N \sum_{k=1}^{N} \frac{1}{k}$$

De plus, comme les variables T_k sont mutuellement indépendantes,

$$V(X_N) = \sum_{k=1}^{N} V(T_k) = \sum_{k=1}^{N} \frac{(k-1)N}{(N+1-k)^2} = N \sum_{k=1}^{N} \frac{N-k}{k^2}$$

3. Classiquement par décroissance de la fonction $t \mapsto \frac{1}{t}$ sur $\mathbb{R}_+^*$:

$$\frac{1}{k+1} \leqslant \ln\left(\frac{k+1}{k}\right) \leqslant \frac{1}{k} \text{ pour tout } k \text{ de } \mathbb{N}^*$$

et, par sommations télescopiques : $\sum_{k=1}^{n} \frac{1}{k} \leqslant 1 + \sum_{k=2}^{n} \frac{1}{k} \leqslant 1 + \ln n$; $\sum_{k=1}^{n} \frac{1}{k} \geqslant \ln(n+1)$.

b) Pour tout entier $n > 1$, $\frac{\ln(n+1)}{\ln n} \leqslant \frac{H_n}{\ln n} \leqslant 1 + \frac{1}{\ln n}$.

Par encadrement $\lim_{n\to\infty} \frac{H_n}{\ln n} = 1$, puis $E(X_N) \underset{(\infty)}{\sim} N \ln N$.

4. a) On a :

$$\begin{aligned} V(X_N) &= N \sum_{k=1}^{N} \frac{N-k}{k^2} \leqslant N^2 \sum_{k=1}^{N} \frac{1}{k^2} \\ &\leqslant N^2\Big(1 + \sum_{k=2}^{N} \frac{1}{k(k-1)}\Big) = N^2\Big(1 + \sum_{k=2}^{N} \frac{1}{k-1} - \sum_{k=2}^{N} \frac{1}{k}\Big) \end{aligned}$$

Donc $V(X_N) \leqslant N^2(1 + 1 - \frac{1}{N}) \leqslant 2N^2$.

b) Par l'inégalité de Bienaymé-Tchebychev,

$$P(|X_N - NH_N| \geqslant \alpha N) = P(|X_N - E(X_N)| \geqslant \alpha N) \leqslant \frac{V(X_N)}{\alpha^2 N^2} \leqslant \frac{2}{\alpha^2}.$$