ESC 1999

Exercice 3

Sujet : nombre moyen de tirages avant de vider une urne.

Moyen

Abordable en première année : ✓

Intérêt : ★★☆

Thèmes du programme abordés : probabilités discrètes

La numérotation des questions n'est pas celle du sujet original.

Modifications apportées au sujet d'origine : Une question trop calculatoire a été supprimée (loi de X_4) , une question a été ajoutée à la fin (équivalent de $E(X_n)$)

Soit n un entier naturel non nul.
Une urne U_n contient n boules numérotées de 1 à n. On effectue dans cette urne une succession de tirages d'une boule, en appliquant la règle suivante : si une boule tirée porte le numéro k, avant de procéder au tirage suivant, on enlève de l'urne toutes les boules dont le numéro est supérieur ou égal à k.
On note X_n la variable aléatoire égale au nombre de tirages nécessaires pour vider l'urne U_n de toutes ses boules.

Partie A : étude de cas particuliers.

1. Donner la loi de X_1, la loi de X_2 et leurs espérances.
2. Déterminer la loi de X_3 et calculer $E(X_3)$.

Partie B : calcul d'une probabilité.

On étudie désormais le cas général.

3. Calculer $P(X_n = 1)$ et $P(X_n = n)$.
4. Soit N_1 la variable aléatoire égale au numéro de la première boule tirée.
   a. Reconnaître la loi de N_1.
   b. Justifier que $\forall i \in [\![2, n]\!],\ \forall k \in [\![2, n]\!],\ P_{[N_1=i]}(X_n = k) = \begin{cases} 0 & \text{si } i \leqslant k-1 \\ P(X_{i-1} = k-1) & \text{si } i \geqslant k \end{cases}$
   c. Montrer que : $\forall k \in [\![2, n]\!],\ P(X_n = k) = \dfrac{1}{n} \displaystyle\sum_{i=k-1}^{n-1} P(X_i = k-1)$.
5. Calculer $P(X_n = 2)$.
6. Pour $n \geqslant 2$, on pose : $v_n = n!P(X_n = n-1)$.
   a. Établir que : $\forall n \geqslant 2,\ v_{n+1} = v_n + n$.
   b. En déduire $P(X_n = n-1)$.

Partie C : recherche d'un équivalent de $E(X_n)$.

7. En utilisant le résultat de la question 4.c, montrer que : $\forall n \geqslant 2,\ E(X_n) = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n-1} E(X_i) + 1$.
8. En déduire que : $\forall n \geqslant 2,\ E(X_n) = E(X_{n-1}) + \dfrac{1}{n}$.
9. Montrer enfin que : $\forall n \geqslant 1,\ E(X_n) = \displaystyle\sum_{k=1}^{n} \frac{1}{k}$.
10. a. Prouver que pour tout entier $k \geqslant 1$, on a $\dfrac{1}{k+1} \leqslant \displaystyle\int_k^{k+1} \frac{dt}{t} \leqslant \frac{1}{k}$.
    b. En sommant les inégalités obtenues à la question précédente, montrer que
    $$\int_1^{n+1} \frac{dt}{t} \leqslant \sum_{k=1}^{n} \frac{1}{k} \leqslant 1 + \int_1^{n} \frac{dt}{t}.$$
    c. Montrer que pour $n \geqslant 1$, $\ln(n) \leqslant E(X_n) \leqslant 1 + \ln(n)$, puis en déduire un équivalent simple de $E(X_n)$.

ESC 1999 : Corrigé

Exercice 3

Dans toute la suite, on note $A_{i,k}$ l'événement : «on tire la boule k au i-ème tirage».

Partie A

1. Si l'urne ne contient que la boule numérotée 1, alors nécessairement celle-ci est obtenue au premier tirage, puis retirée de l'urne. Et donc l'urne est vide après un seul tirage : X_1 est la variable certaine égale à 1.
Et donc $E(X_1) = 1$.

Si $n = 2$, l'urne contient initialement deux boules numérotées 1 et 2.
Puisqu'on retire au moins une boule de l'urne après chaque tirage, $X_2(\Omega) = \{1, 2\}$.
De plus, on a $[X_2 = 1] = A_{1,1}$ et donc $P(X_2 = 1) = P(A_{1,1}) = \frac{1}{2}$.
Et puisque $\{[X_2 = 1], [X_2 = 2]\}$ est un système complet d'événements, on a

$$P(X_2 = 1) + P(X_2 = 2) = 1 \Leftrightarrow P(X_2 = 2) = 1 - P(X_2 = 1) = \frac{1}{2}.$$

Et donc $E(X_2) = \frac{1}{2} + 2\frac{1}{2} = \boxed{\frac{3}{2}.}$

2. Si $n = 3$, l'urne contient initialement les boules numérotées de 1 à 3, et comme précédemment, $X_3(\Omega) = [\![1, 3]\!]$.
On a $[X_3 = 1] = A_{1,1}$ donc $P(X_3 = 1) = P(A_{1,1}) = \frac{1}{3}$.
De même, le seul moyen de nécessiter trois tirages pour vider l'urne est de tirer successivement les boules 3, 2 et 1. Et donc on a $[X_3 = 3] = A_{1,3} \cap A_{2,2} \cap A_{3,1}$, de sorte que $P(X_3 = 3) = P(A_{1,3} \cap A_{2,2} \cap A_{3,1})$.
Par la formule des probabilités composées, on a donc

$$P(X_3 = 3) = P(A_{1,3})P_{A_{1,3}}(A_{2,2})P_{A_{1,3}\cap A_{2,2}}(A_{3,1}) = \frac{1}{3} \times \frac{1}{2} \times 1 = \frac{1}{6}.$$

> **Détails**
> Connaissant les résultats des tirages précédents, on connait la composition de l'urne avant un tirage donné.

Enfin, comme à la question 1,

$$P(X_3 = 2) = 1 - P(X_3 = 1) - P(X_3 = 3) = \frac{1}{2}.$$

On en déduit que $E(X_3) = \frac{1}{3} + 2\frac{1}{2} + 3\frac{1}{6} = \boxed{\frac{11}{6}.}$

Partie B

3. Comme dans les cas particuliers précédents, on a $[X_n = 1] = A_{1,1}$ et donc

$$\boxed{P(X_n = 1) = P(A_{1,1}) = \frac{1}{n}.}$$

De même, le seul moyen de nécessiter n tirages pour vider l'urne est de tirer successivement les boules $n, n-1, \dots, 2, 1$ et donc

$$[X_n = n] = A_{1,n} \cap A_{2,n-1} \cap \dots \cap A_{n,1}.$$

Et donc $P(X_n = n) = P(A_{1,n} \cap A_{2,n-1} \cap \dots \cap A_{n,1})$.
Par la formule des probabilités composées, on a donc

$$\begin{aligned} P(X_n = n) &= P(A_{1,n})P_{A_{1,n}}(A_{2,n-1}) \cdots P_{A_{1,n}\cap\cdots\cap A_{n-1,2}}(A_{n,1}) \\ &= \frac{1}{n}\frac{1}{n-1} \cdots 1 = \boxed{\frac{1}{n!}.} \end{aligned}$$

4.a. La première boule tirée est choisie au hasard parmi les boules de 1 à n, qui sont toutes équiprobables, de sorte que $N_1 \hookrightarrow \mathcal{U}([\![1,n]\!])$.

4.b. Sachant que $[N_1 = i]$, à l'issue du premier tirage, il ne reste que les boules 1 à $i-1$ dans l'urne.
Si $i \leqslant k-1$, alors, après le premier tirage il faudra au plus $i-1 \leqslant k-2$ autres tirages pour vider l'urne.
Et donc X_n ne peut prendre que des valeurs inférieures ou égales à $k-1$, de sorte que $P_{[N_1=i]}(X_n = k) = 0$.
En revanche, si $i \geqslant k$, alors le second tirage a lieu dans une urne contenant les boules de 1 à $i-1$.
Et donc le nombre de tirages nécessaires pour virer l'urne **après le premier tirage** est le même que le nombre de tirages nécessaires pour vider l'urne U_{i-1}.
En particulier, $P_{[N_1=i]}(X_n = k) = P(X_{i-1} = k-1)$.

Remarque
Ce raisonnement est encore valable si $i \leqslant k-1$, mais alors $P(X_{i-1} = k-1) = 0$.

4.c. D'après la formule des probabilités totales appliquée au système complet d'événements $\{[N_1 = i],\ i \in [\![1,n]\!]\}$, on a, pour $k \geqslant 2$,

$$\begin{aligned}P(X_n = k) &= \sum_{i=1}^{n} P(N_1 = i)P(X_n = k \cap N_1 = i) = \sum_{i=1}^{n} \frac{1}{i} P_{[N_1=i]}(X_n = k)\\ &= \frac{1}{n}\sum_{i=k}^{n} P(X_{i-1} = k-1)\\ &= \frac{1}{n}\sum_{j=k-1}^{n-1} P(X_j = k-1).\end{aligned}$$

Chgt d'indice
$j = i-1$.

5. D'après la question précédente[1], on a

$$P(X_n = 2) = \frac{1}{n}\sum_{i=1}^{n-1} P(X_i = 1) = \boxed{\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{i}.}$$

[1] Appliquée avec $k = 2$.

6.a. Par définition, on a $v_{n+1} = (n+1)!P(X_{n+1} = n)$. En utilisant la question 5.c, il vient alors

$$\begin{aligned}v_{n+1} &= (n+1)!\frac{1}{n+1}\sum_{i=n-1}^{n} P(X_i = n-1)\\ &= n!\big(P(X_{n-1} = n-1) + P(X_n = n-1)\big)\\ &= n!\frac{1}{(n-1)!} + n!P(X_n = n-1) = \boxed{n + v_n.}\end{aligned}$$

6.b. Nous savons déjà d'après la question 1 que $v_2 = 2! \times P(X_2 = 1) = 1$.
Et pour $n \geqslant 3$, en utilisant plusieurs fois le résultat de la question précédente,

$$\begin{aligned}v_n &= v_{n-1} + (n-1) = v_{n-2} + (n-2) + (n-1)\\ &= \cdots = v_2 + 2 + 3 + \cdots + n-1 = 1 + 2 + \cdots + n-1 = \frac{n(n-1)}{2}.\end{aligned}$$

Rappel
On a
$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

Et donc $P(X_n = n-1) = \dfrac{1}{n!}\dfrac{n(n-1)}{2} = \boxed{\dfrac{1}{2(n-2)!}.}$

Vérification
Il n'est pas inutile de vérifier que ce résultat est cohérent avec les lois de X_2 et de X_3 calculées dans la partie A afin de déceler d'éventuelles erreurs.

Partie C.

7. D'après la question 4.c, on a, pour tout $k \geqslant 2$,

$$kP(X_n = k) = \frac{1}{n}\sum_{i=k-1}^{n-1} \underbrace{k}_{=(k-1)+1} P(X_i = k-1) = \frac{1}{n}\sum_{i=k-1}^{n-1}(k-1)P(X_i = k-1) + \frac{1}{n}\sum_{i=k-1}^{n-1} P(X_i = k-1).$$

Et donc,

$$E(X_n) = \sum_{k=1}^{n} kP(X_n = k)$$

$$\begin{aligned}
&= P(X_n = 1) + \sum_{k=2}^{n}\left(\frac{1}{n}\sum_{i=k-1}^{n-1}(k-1)P(X_i = k-1) + \frac{1}{n}\sum_{i=k-1}^{n-1}P(X_i = k-1)\right)\\
&= \frac{1}{n} + \frac{1}{n}\sum_{k=2}^{n}\sum_{i=k-1}^{n-1}(k-1)P(X_i = k-1) + \frac{1}{n}\sum_{k=2}^{n}\sum_{i=k-1}^{n-1}P(X_i = k-1)\\
&= \frac{1}{n} + \frac{1}{n}\sum_{i=1}^{n-1}\sum_{k=2}^{i+1}(k-1)P(X_i = k-1) + \frac{1}{n}\sum_{i=1}^{n-1}\sum_{k=2}^{i+1}P(X_i = k-1)\\
&= \frac{1}{n} + \frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=1}^{i}jP(X_i = j) + \frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=1}^{i}P(X_i = j)\\
&= \frac{1}{n} + \frac{1}{n}\sum_{i=1}^{n-1}E(X_i) + \frac{1}{n}\sum_{i=1}^{n-1}1\\
&= \frac{1}{n}\sum_{i=1}^{n-1}E(X_i) + \frac{1}{n} + \frac{n-1}{n}\\
&= \boxed{\frac{1}{n}\sum_{i=1}^{n-1}E(X_i) + 1.}
\end{aligned}$$

On isole le terme correspondant à $k = 1$ pour lequel la relation de 4.c n'est pas valable.

Chgt d'indice
$j = k - 1$.

8. Pour $n = 2$, on a $E(X_2) = \frac{3}{2} = E(X_1) + \frac{1}{2}$.
Et pour $n \geqslant 3$, alors

$$\begin{aligned}
E(X_n) &= \frac{1}{n}\sum_{i=1}^{n-1}E(X_i) + 1\\
&= \frac{1}{n}E(X_{n-1}) + \frac{1}{n}\sum_{i=1}^{n-2}E(X_i) + 1\\
&= \frac{1}{n}E(X_{n-1}) + \frac{n-1}{n}\,\frac{1}{n-1}\sum_{i=1}^{n-2}E(X_i) + 1\\
&= \frac{1}{n}E(X_{n-1}) + \frac{n-1}{n}\left(\sum_{i=1}^{n-2}E(X_i) + 1\right) + 1 - \frac{n-1}{n}\\
&= \frac{1}{n}E(X_{n-1}) + \frac{n-1}{n}E(X_{n-1}) + \frac{1}{n}\\
&= \boxed{E(X_{n-1}) + \frac{1}{n}.}
\end{aligned}$$

9. Nous savons déjà que $E(X_1) = 1 = \frac{1}{1}$.
Et pour $n \geqslant 2$,

$$E(X_n) = E(X_{n-1}) + \frac{1}{n} = E(X_{n-2}) + \frac{1}{n-1} + \frac{1}{n} = \cdots = E(X_1) + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1} + \frac{1}{n} = \sum_{k=1}^{n}\frac{1}{k}.$$

10.a. La fonction $t \mapsto \frac{1}{t}$ est décroissante sur $[k, k+1]$. En particulier,

$$\forall t \in [k, k+1],\ \frac{1}{k+1} \leqslant \frac{1}{t} \leqslant \frac{1}{k}.$$

Et donc, par croissance de l'intégrale,

$$\boxed{\frac{1}{k+1} = \int_k^{k+1}\frac{dt}{k+1} \leqslant \int_k^{k+1}\frac{dt}{t} \leqslant \int_k^{k+1}\frac{dt}{k} = \frac{1}{k}.}$$

Astuce
Ne perdons pas de temps à calculer une primitive de la constante $\frac{1}{k}$: l'intégrale d'une constante sur un segment est égale à cette constante fois la longueur du segment.

10.b. Sommons les inégalités précédemment obtenues pour k variant de 1 à $n - 1$. Alors

$$\sum_{k=1}^{n-1}\frac{1}{k+1} \leqslant \sum_{k=1}^{n-1}\int_k^{k+1}\frac{dt}{t} \leqslant \sum_{k=1}^{n-1}\frac{1}{k}.$$

Notons que par la relation de Chasles,

$$\sum_{k=1}^{n-1} \int_k^{k+1} \frac{dt}{t} = \int_1^2 \frac{dt}{t} + \int_2^3 \frac{dt}{t} + \cdots + \int_{n-1}^n \frac{dt}{t} = \int_1^n \frac{dt}{t}.$$

En particulier, on a $\displaystyle\sum_{k=1}^{n-1} \frac{1}{k+1} \leqslant \int_1^n \frac{dt}{t}$.

Mais $\displaystyle\sum_{k=1}^{n-1} \frac{1}{k+1} = \sum_{i=2}^{n} \frac{1}{i} = \sum_{k=1}^{n} \frac{1}{k} - 1$. Et donc

$$\sum_{k=1}^{n-1} \frac{1}{k+1} \leqslant \int_1^n \frac{dt}{t} \Longleftrightarrow \sum_{k=1}^{n} \frac{1}{k} \leqslant 1 + \int_1^n \frac{dt}{t}.$$

De même, toujours en sommant les inégalités de 10.a pour k variant de 1 à n, il vient

$$\int_1^{n+1} \frac{dt}{t} = \sum_{k=1}^{n} \int_k^{k+1} \frac{dt}{t} \leqslant \sum_{k=1}^{n} \frac{1}{k}.$$

10.c. Nous savons calculer les intégrales de la question précédente :

$$\int_1^n \frac{dt}{t} = [\ln t]_1^n = \ln n \text{ et de même } \int_1^{n+1} \frac{dt}{t} = \ln(n+1) \geqslant \ln(n).$$

Ainsi, on a bien

$$\boxed{\ln(n) \leqslant E(X_n) \leqslant \ln(n) + 1.}$$

En divisant par $\ln n > 0$, il vient

$$\frac{1}{\ln n} \leqslant \frac{E(X_n)}{\ln n} \leqslant 1 + \frac{1}{\ln n}.$$

Et donc par le théorème des gendarmes, $\displaystyle\lim_{n\to+\infty} \frac{E(X_n)}{\ln n} = 1$, de sorte que $\boxed{E(X_n) \underset{n\to+\infty}{\sim} \ln n.}$

Méthode

Une inégalité ne suffit pas à donner un équivalent, même si l'on sait que

$$1 + \ln(n) \underset{n\to+\infty}{\sim} \ln n.$$

Pour prouver un équivalent, il faudra toujours montrer que la limite du quotient tend vers 1 (et donc, si l'on souhaite, comme c'est le cas ici, utiliser une double inégalité, celle-ci se traduira en une limite grâce au théorème des gendarmes).