Devoir surveillé n°9

Sujet donné le 26 mai 2023, 4h.
L'usage de la calculatrice n'est pas autorisé.
La notation tiendra particulièrement compte de **la qualité de la rédaction, la précision des raisonnements et l'énoncé des formules utilisées. Les réponses aux questions seront numérotées et séparées par un trait horizontal. Les résultats essentiels devront être encadrés ou soulignés.**

BON TRAVAIL

Temps d'attente et problème du collectionneur

Notations et remarques préliminaires

Dans tout ce sujet, on considère des variables aléatoires à valeurs dans $\mathbb{N}$ ou une partie de $\mathbb{N}$.

- Si X est une variable aléatoire finie c'est-à-dire ici que $X(\Omega)$ est une partie finie de $\mathbb{N}$,
  on note $\mathbf{E}(X) = \displaystyle\sum_{k\in X(\Omega)} k\mathbf{P}(X = k)$, l'espérance de X (comme dans le cours).
  on note également $\mathbf{V}(X) = \mathbf{E}([X - \mathbf{E}(X)]^2)$, la variance de X (comme dans le cours).
- Si X est une variable aléatoire dénombrable, c'est-à-dire ici que $X(\Omega) = \mathbb{N} \cup \{+\infty\}$,
  on dira que X admet une espérance si $\mathbf{P}(X = +\infty) = 0$ et si la série $\displaystyle\sum_{k\geqslant 0} k\mathbf{P}(X = k)$ converge.

  Dans ce cas, on notera $\mathbf{E}(X)$, la valeur de $\displaystyle\sum_{k=0}^{+\infty} k\mathbf{P}(X = k)$.

Objectifs

Le but de ce problème est d'étudier des temps d'attente pour la réalisation de différentes situations, et en particulier le fameux « problème du collectionneur ».

Le préliminaires donnent 3 résultats utiles pour la suite (à différents endroits).
La première partie se consacre à l'obtention de séquences consécutives, pour le jeu du Pile ou Face.
La seconde partie élargit un peu la problématique en se consacrant aux nombres de séries obtenues en limitant le nombre de lancers.
Dans la dernière partie nous nous concentrons tout particulièrement sur différentes questions du problème du collectionneur.

Préliminaires

1. Pour $x \in [0, 1[$, on pose $f_n(x) = \sum\limits_{k=0}^{n} x^k$.
   (a) En exploitant une autre expression de $f_n(x)$, calculer $\displaystyle\sum_{k=1}^{n} kx^{k-1}$.
   (b) Montrer que $\displaystyle\sum_{k=1}^{+\infty} kx^{k-1} = \frac{1}{(1-x)^2}$.
2. On considère une variable aléatoire U à valeurs dans $[\![0, n]\!]$.
   Montrer que $\mathbf{E}(U) = \displaystyle\sum_{k=0}^{n} \mathbf{P}(U > k)$
3. Nombre de surjections de E_n sur E_m.
   Soient $n, m \in \mathbb{N}^*$. On considère deux ensembles E_n et E_m à respectivement n et m éléments.
   On note $\sigma_{n,m}$ le nombre de surjections de E_n sur E_m.
   (a) Sans justification particulière, donner les valeurs de $\sigma_{n,1}$, $\sigma_{1,m}$, $\sigma_{n,n}$ et $\sigma_{k,m}$ avec $k < m$.

(b) En associant, à toute surjection φ de $E_n = \{a_1, \dots a_n\}$ sur E_m, la fonction $\varphi_{|E_n \setminus \{a_n\}}$ à valeurs dans E_m ou dans $E_m \setminus \{\varphi(a_n)\}$, montrer que

$$\sigma_{n,m} = m\big(\sigma_{n-1,m-1} + \sigma_{n-1,m}\big)$$

(c) Montrer, <u>par ailleurs</u>, que $m^n = \displaystyle\sum_{k=0}^{m} \binom{m}{k} \sigma_{n,k}$.

(d) Montrer, alors que $\sigma_{n,m} = \displaystyle\sum_{j=0}^{m} (-1)^{m-j} \binom{m}{j} j^n$.

(e) En déduire la valeur de $\displaystyle\lim_{n\to+\infty} \frac{\sigma_{n,m}}{m^n}$. Interpréter ce résultat.

I. Séquences dans le jeu du Pile ou Face infini

On effectue une succession de lancers indépendants d'une pièce qui donne Pile avec la probabilité $p \in]0, 1[$, Face avec la probabilité $q = 1 - p$. On s'intéresse, dans cette partie, aux séquences (ou série) de lancers consécutifs qui donnent le même côté.
On dit que la première série (ou séquence) est de longueur n si les n premiers lancers ont amené le même côté de la pièce et le $n + 1$-ième l'autre côté. La deuxième série commence au lancer suivant la fin de la première série et se termine (si elle se termine) au lancer précédent un changement de côté. On définit de même les séries suivantes.

On notera P_k l'événement « le k-ième lancer a donne Pile ».
On note L_k la variable aléatoire donnant longueur de la k-ième série (et on convient que $L_k = 0$ s'il n'y a pas de k-ième série et $L_k = +\infty$ si la série est infinie).

Par exemple, si le début d'une partie conduit à : ω =Pile,Pile,Pile,Face,Face,Face,Face,Pile,Face. . ., on peut voir ici trois séries et le début d'une quatrième. On a $L_1(\omega) = 3$, $L_2(\omega) = 4$ et $L_3(\omega) = 1$.

On suppose dans les questions 4 et 5 que l'on effectue un nombre $N \geqslant 1$ fixé de lancers.

4. Etude de L_1
   (a) Déterminer la loi de L_1.
   On pourra, par exemple, commencer par s'intéresser à l'événement $[L_1 > k]$.
   (b) Vérifier que $\displaystyle\sum_{k=1}^{N} \mathbf{P}(L_1 = k) = 1$.
   (c) Montrer que
   $$\mathbf{E}(L_1) = \frac{p^2 + q^2}{pq} - \left(\frac{p^{N+1}}{q} + \frac{q^{N+1}}{p}\right)$$
5. Etude de L_2
   (a) Déterminer la loi conjointe du couple (L_1, L_2).
   (b) Déterminer la loi de L_2.
   (c) L_1 et L_2 sont elles indépendantes ?

*On suppose désormais que l'on effectue un **nombre infini de lancers**.*

6. Déterminer la loi de L_1 en n'oubliant pas $\mathbf{P}(L_1 = +\infty)$ puis $\mathbf{E}(L_1)$.
7. Montrer que pour tout $k \in \mathbb{N}^*$, $\mathbf{P}(L_2 = k) = p^2 q^{k-1} + q^2 p^{k-1}$.
8. Calculer $\mathbf{E}(L_2)$ et la comparer à $\mathbf{E}(L_1)$.
9. Donner une C.N.S simple pour que L_1 et L_2 soient indépendantes.
   On rappelle que $p \in]0, 1[$.
10. Donner de façon générale la loi de L_i.

II - Nombre de séquences dans le jeu fini

On suppose dans cette partie que $p = q = \frac{1}{2}$. On s'intéresse à la variable aléatoire N_n donnant le nombre de séries ou séquences observées lors des n premiers lancers.

11. Donner les lois de N_1, N_2 et leur espérance.
12. Montrer de manière rigoureuse que pour $n \geqslant 2$ et $k \in [\![1, n]\!]$,
$$P(N_n = k) = \frac{1}{2} P(N_{n-1} = k) + \frac{1}{2} P(N_{n-1} = k - 1)$$

13. On pose $G_n(s) = \sum_{k=1}^{n} \mathbf{P}(N_n = k)s^k$.
   (a) Montrer que pour $n \geqslant 2$, $G_n(s) = \frac{1+s}{2} G_{n-1}(s)$.
   (b) En déduire $G_n(s)$ puis le nombre moyen de séries lors des n premiers lancers.
   (c) Calculer $G_n''(s)$, en déduire $\mathbf{V}(N_n)$.
   On pourra commencer par montrer que $\mathbf{V}(N_n) = G_n''(1) + G_n'(1) - [G_n'(1)]^2$.
   (d) Déduire des questions précédentes que $\mathbf{P}(\frac{n}{2} - \sqrt{n} \leqslant N_n \leqslant \frac{n}{2} + 1 + \sqrt{n})$ est minoré par $\dfrac{3}{4}$.

III - Problème du collectionneur

On fixe pour cette partie un entier $m > 2$.
Le problème du collectionneur d'images ou de vignettes (« coupon collector's problem ») est un problème de probabilités classique. Un fabricant de tablettes de chocolat propose à ses acheteurs de collectionner des vignettes. Chaque tablette vendue contient une vignette de la collection, que l'on découvre à l'ouverture de la tablette. Il y a m vignettes différentes à collectionner.
On note $E_m = \{x_1, x_2, \ldots x_m\}$, l'ensemble des m vignettes, et pour tout $i \in \mathbb{N}$, V_i est la variable aléatoire qui indique le numéro j de la vignette x_j obtenue lors de l'achat i.
On suppose que les variables $(V_1, V_2, \ldots V_i \ldots)$ sont indépendantes. Enfin, on note pour tout $i \in \mathbb{N}^*$, $j \in [\![1, m]\!]$, $p_j = \mathbf{P}(V_i = j) \in]0, 1[$.
On se propose de déterminer le nombre moyen d'achats nécessaires pour constituer la collection complètes des m images.

III.A. Nombre de vignettes différentes après n achats

On commence par s'intéresser au nombre moyen de vignettes que l'on a après un nombre n (fixé) d'achats de tablettes.

14. Soit $j \in [\![1, n]\!]$. On note U_n^j, la variable aléatoire qui indique si la vignette j a été obtenue lors de n premiers achats.
   (a) Quelle est la loi de U_n^j ?
   (b) Les variables $(U_n^j)_{n \in \mathbb{N}}$ sont-elles indépendantes ?
15. On note T_n, la variable aléatoire qui indique le nombre de vignettes obtenues après les n premiers achats (certaines peuvent être en multiple exemplaire).
   (a) Exprimer T_n en fonction des (U_n^j).
   (b) En déduire $\mathbf{E}(T_n) = \sum_{j=1}^{m} (1 - (1 - p_j)^n)$.
   Quelle est la limite de $(\mathbf{E}(T_n))_{n \in \mathbb{N}}$? Qu'en pensez-vous ?
16. Comparaison au cas uniforme.
   (a) Que vaut $\mathbf{E}(T_n)$ dans le cas où, pour tout $i \in \mathbb{N}$, V_i suit une loi uniforme ?
   (b) Montrer que pour tout $n \in \mathbb{N}^*$, et tout $u \in]-1, 1]$, $(1 + u)^n \geqslant 1 + nu$.
   (c) Posons pour tout $j \in [\![1, m]\!]$, $\epsilon_j = p_j - \frac{1}{m}$. Que vaut $\sum_{j=1}^{m} \epsilon_j$?
   (d) Montrer que
   $$(1 - p_j)^n \geqslant \left(1 - \frac{1}{m}\right)^n - n\epsilon_j \left(1 - \frac{1}{m}\right)^{n-1}$$
   Puis en déduire que
   $$\mathbf{E}(T_n) \leqslant m \left(1 - \left(\frac{m-1}{m}\right)^n\right)$$

III.B. Nombre de tablettes à acheter pour avoir la collection complète

On s'intéresse maintenant au nombre n (variable) d'achats à effectuer pour avoir toute la collection. Nous ne considérons que des distributions équiprobables des vignettes, c'est-à-dire : pour tout $j \in [\![1, m]\!]$, $p_j = \dfrac{1}{m}$.
Pour $j \in [\![1, m]\!]$, on note X_j, la variable aléatoire donnant le nombre d'achats effectués au moment où la collection compte pour la première fois j vignettes. On définit alors $Z_j = X_{j+1} - X_j$.

17. Préciser la loi de la variable aléatoire X_1 et donner $\mathbf{E}(X_1)$ son espérance.
18. Pour $j \in [\![1, m-1]\!]$, montrer que $Z_j(\Omega) = \mathbb{N}^* \cup \{+\infty\}$ et $\mathbf{P}(Z_j = k) = \left(\dfrac{j}{m}\right)^{k-1} \times \dfrac{m-j}{m}$.
    En déduire $\mathbf{E}(Z_j) = \dfrac{m}{m-j}$.
19. En exploitant un télescopage, donner une expression de $\mathbf{E}(X_m)$ sous la forme d'une somme.
20. On note, pour tout $n \in \mathbb{N}^*$, $u_n = \displaystyle\sum_{i=1}^{n} \frac{1}{i} - \ln(n)$ et pour $n \geqslant 2$, $v_n = u_n - u_{n-1}$.
    (a) Donner un équivalent de v_n.
    (b) En déduire que la suite (u_n) est convergente.

On note $\gamma = \lim(u_n)$.

21. Donner un développement asymptotique à deux termes de $\mathbf{E}(X_m)$ pour $m \to +\infty$. Interpréter ce résultat.

On fixe de nouveau m et on essaye maintenant de préciser la valeur de $\mathbf{P}(X_m = n)$ pour différentes valeurs de n.

22. Quelle est la valeur de $\mathbf{P}(X_m = n)$ pour n compris entre 1 et $m - 1$?
23. On rappelle que $\sigma_{n,m}$ est le nombre de surjections d'un ensemble de n éléments vers un ensemble de m éléments.
    Montrer que pour tout $h \in \mathbb{N}$, $\mathbf{P}(X_m = h) = \dfrac{m\sigma_{h-1,m-1}}{m^h}$.
24. En déduire la limite de $\mathbf{P}(X_m = h)$ lorsque $h \to +\infty$. Interpréter ce résultat.