Devoir maison n°4
Corrigé

Exercice 1

On désigne par n un entier naturel non nul.
On lance n fois une pièce de monnaie donnant « Pile » avec la probabilité p (avec $0 < p < 1$) et « Face » avec la probabilité $q = 1 - p$. On appelle k-chaîne de « Pile » une suite de k lancers consécutifs ayant tous donnés « Pile » , cette suite devant être précédée d'un « Face » ou débuter le tirage, et suivie d'un « Face » ou être la dernière suite du tirage.
Pour tout k de $[\![1, n]\!]$ on note Y_k la variable aléatoire égale au nombre total de k-chaînes de « Pile » obtenues au cours des n lancers.
Pour tout k de $[\![1, n]\!]$, on pourra noter P_k l'événement « on obtient « Pile » au k-ème lancer ».
Par exemple, avec $n = 11$, si l'on a obtenu les résultats $P_1P_2F_3F_4P_5P_6P_7F_8P_9F_{10}P_{11}$ alors $Y_1 = 2$, $Y_2 = 1$, $Y_3 = 1$ et pour tout $k \geqslant 4$, $Y_k = 0$.
Le but de cet exercice est de déterminer, pour tout k de $[\![1, n]\!]$ l'espérance de Y_k, notée $E(Y_k)$.

1. **Déterminer la loi de Y_n et donner $E(Y_n)$.**

   Y_n compte le nombre de n-chaînes de Pile sur n lancers : elle vaut donc 1 si tous les lancers ont donné Pile, et 0 sinon.
   $P(Y_n = 1) = P(P_1 \cap \ldots \cap P_n) = p^n$ par indépendance des lancers ; et comme $Y_n(\Omega) = \{0, 1\}$ on conclut $Y_n \hookrightarrow \mathscr{B}(p^n)$ et donc $E(Y_n) = p^n$.

2. **Montrer que $P(Y_{n-1} = 1) = 2qp^{n-1}$. et donner $E(Y_{n-1})$.**

   Y_{n-1} vaut 1 si on a une succession de $n-1$ Pile. En termes d'événements :

   $$(Y_{n-1} = 1) = (P_1 \cap \ldots \cap P_{n-1}F_n) \cup (F_1 \cap P_2 \ldots \cap P_n)$$

   Par indépendance des lancers et incompatibilité, $P(Y_{n-1} = 1) = p^{n-1}q + qp^{n-1} = 2qp^{n-1}$.

   On s'intéresse maintenant à $Y_{n-1}(\Omega)$. On voit qu'il est impossible d'avoir plus d'une $(n-1)$ chaîne (y compris dans le cas $n = 2$, car si on a «deux 1-chaîne», on a P_1P_2 donc en fait c'est une 2-chaîne !).
   Donc ici encore $Y_{n-1}(\Omega) = \{0, 1\}$ et le calcul précédent montre que $Y_{n-1} \hookrightarrow \mathscr{B}(2qp^{n-1})$ et donc $E(Y_{n-1}) = 2qp^{n-1}$.

3. **Dans cette question, k désigne un entier de $[\![1, n-2]\!]$.**
   Pour tout i de $[\![1, n]\!]$ on note $X_{i,k}$ la variable aléatoire qui vaut 1 si une k-chaîne de « Pile » commence au i-ème lancer et qui vaut 0 sinon.

   (a) **Calculer $P(X_{1,k} = 1)$.**

   $X[_{1,k}= 1]$ ssi une k-chaîne commence au 1er lancer, donc ssi les $k+1$ premiers lancers donnent $P_1 \ldots P_k F_{k+1}$ (attention il faut terminer la chaîne !!).
   Donc
   $$P(X_{1,k} = 1) = P\big(P_1 \cap \cdots \cap P_k \cap F_{k+1}\big) = p^k q$$
   par indépendance des lancers.

   (b) **Soit $i \in [\![2, n-k]\!]$. Montrer que $P(X_{i,k} = 1) = q^2 p^k$.**

   Cette fois la chaîne commence en i, donc concerne les lancers i, $i+1$, ..., $i+k-1$.
   $i \geqslant 2$ donc pour que la chaîne *commence* en i, il faut un Face au lancer $i-1$; et avec $i \leqslant n-k$ par hypothèse, on a $i+k-1 \leqslant n-k+k-1 = n-1$ donc le lancer $i+k-1$ n'est pas le dernier : il faut un Face en $i+k$ pour arrêter la chaîne.
   Ainsi :
   $$P(X_{i,k} = 1) = P\big(F_{i-1} \cap P_i \cdots \cap P_{i+k-1} \cap F_{i+k}\big) = qp^k q = q^2 p^k$$

(c) **Montrer que** $P(X_{n-k+1,k}=1)=qp^k$**.**

Une k-chaîne de Pile qui commence au lancer $n-k+1$ va jusqu'au lancer $(n-k+1)+k-1=n$: elle se termine avec la série de lancers ! Donc le «Face» qui clôt la chaîne dans la question précédente n'a pas lieu ici.
Par contre, avec $k \leqslant n-2$, $n-k+1 \geqslant n-(n-2)+1 \geqslant 3$ donc le lancer $n-k+1$ n'est pas le premier : il faut un Face au lancer $n-k$ avant le début de la k-chaîne.
Pour résumer

$$[X_{n-k+1,k}=1]=F_{n-k}\cap P_{n-k+1}\cap P_{n-k+2}\cdots\cap P_n$$

et donc

$$P(X_{n-k+1,k}=1)=P(F_{n-k}\cap P_{n-k+1}\cap P_{n-k+2}\cdots\cap P_n)=qp^k$$

(d) **Exprimer** Y_k **en fonction des variables** $X_{i,k}$**, puis déterminer** $E(Y_k)$**.**

Argument classique derrière lequel se cachent des indicatrices : pour avoir le nombre de k-chaînes, on somme sur i les $X_{i,k}$ car $X_{i,k}$ donne le nombre de k-chaînes qui commencent au lancer i (il y en a 0 ou 1 !).
i varie ici de 1 à $n-k+1$ car il ne peut pas y avoir de k-chaîne commençant plus loin (elle ne rentrera pas : dans le cas de 5 lancers, on ne peut pas avoir une 3-chaîne qui commence au 4ème lancer par exemple !!).
Autrement dit $Y_k=\sum_{i=1}^{n-k+1}X_{i,k}$. Alors par linéarité de l'espérance : $E(Y_k)=\sum_{i=1}^{n-k+1}E(X_{i,k})$.
Les $X_{i,k}$ étant des variables de Bernoulli, on a toujours $E(X_{i,k})=P(X_{i,k}=1)$. On déduit, en découpant la somme suivant les cas particuliers traités aux questions précédentes :

$$\begin{aligned}E(Y_k)=\sum_{i=1}^{n-k+1}E(X_{i,k})&=E(X_{1,k})+\sum_{i=2}^{n-k}E(X_{i,k})+E(X_{n-k+1,1})\\&=p^kq+\sum_{i=2}^{n-k}p^kq^2+p^kq\\&=2p^kq+\big((n-k)-2+1\big)p^kq^2\\E(Y_k)&=2p^kq+(n-k-1)p^kq^2\end{aligned}$$

Exercice 2 : Problème du collectionneur

Une marque de céréales propose de collectionner des vignettes : chaque paquet contient une vignette, et il y a en tout N vignettes différentes.
On suppose que les vignettes insérées dans les paquets successifs sont indépendantes les unes des autres, et que chaque vignette est choisie au hasard uniformément. Une personne achète un paquet de céréales par jour : on étudie l'évolution de sa collection au cours du temps.

1. **Pour** $k\in\mathbb{N}^*$**, on pose** X_k **la variable aléatoire comptant le nombre de vignettes distinctes obtenues au bout de** k **paquets achetés.**
Préciser $X_k(\Omega)$ **(on distinguera** $k\leqslant N$ **et** $k>N$**). Donner les lois de** X_1**,** X_2 **et** X_3**.**

   On suppose $1\leqslant k\leqslant N$.
   En k jours on peut obtenir au minimum une vignette (on a tous les jours la même) et au maximum k vignette (une nouvelle chaque jour). Donc $X_k(\Omega)=[\![1,k]\!]$.
   Toutefois ce raisonnement ne tient pas si $k>N$, car il n'y a que N sortes de vignettes : si $k>N$, on a $X_k(\Omega)=[\![1,N]\!]$.

   X_1 est le nombre de vignettes distinctes obtenues à l'achat du premier paquet : on a donc $X_1=1$ (variable constante).

   X_2 est le nombre de vignettes distinctes obtenues à l'achat des deux premiers paquets : elle peut donc prendre les valeurs 1 (on a deux fois la même vignette) ou 2 (on a deux vignettes différentes).
   On peut regarder $P(X_2=1)$: c'est la probabilité que la vignette du second paquet soit égale à celle du premier. Par indépendance des vignettes successives, cette proba vaut $\frac{1}{N}$.
   Ainsi $P(X_2=1)=\frac{1}{N}$ et donc $P(X_2=2)=1-\frac{1}{N}$.

Enfin par des raisonnements similaires, $P(X_3 = 1) = (\frac{1}{N})^2$; et $P(X_3 = 3) = \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)$ (car la seconde doit être différente de la première, et la troisième doit être différente des deux premières).
On en déduit $P(X_3 = 2) = 1 - P(X_3 = 1) - P(X_3 = 3) = 1 - \frac{1}{N^2} - \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) = \frac{3}{N} - \frac{1}{N^2}$.

Pour $k \in [\![1, N]\!]$, on note S_k le nombre de jours nécessaires que doit attendre le collectionneur pour que le nombre de vignettes différentes qu'il possède passe de $k-1$ à k.
On note $S = S_1 + \cdots + S_N$ le nombre de jours nécessaires pour obtenir la collection complète des N vignettes.

2. **Que vaut $S_k(\Omega)$ pour $k \in [\![1, N]\!]$?**

   S_1 est le temps nécessaire pour passer de 0 vignette à 1 vignette : ceci arrivera au premier jour. Ainsi
   $$S_1(\Omega) = \{1\}$$
   Pour $k \geqslant 2$, on peut attendre arbitrairement longtemps pour obtenir sa k-ème vignette distincte (en n'achetant que des paquets qui contiennent des vignettes qu'on a déjà !) :
   $$\forall\, k \geqslant 2,\ S_k(\Omega) = \mathbb{N}^*$$

3. **Justifier que S_k suit une loi géométrique dont on donnera le paramètre.**
   S_k mesure le nombre d'achats nécessaires pour qu'on obtienne une vignette parmi les $N-(k-1)$ qu'on n'a pas encore : c'est le temps d'attente d'un succès survenant avec probabilité $\frac{N-(k-1)}{N}$. Les différentes vignette sont indépendantes donc on est bien dans le cadre d'une loi géométrique :
   $$\forall\, k \geqslant 1,\ S_k \hookrightarrow \mathcal{G}\left(\frac{N-(k-1)}{N}\right)$$
   (NB : ça marche aussi pour $k = 1$: la loi $\mathcal{G}(1)$ (temps d'attente d'un succès arrivant avec proba 1) est bien celle d'une variable constante égale à 1.)

On rappelle la linéarité de l'espérance : si** $X_1, \ldots, X_n$ **sont des variables aléatoires admettant une espérance, et** $\lambda_1, \ldots, \lambda_n$ **sont des réels, alors** $X = \sum_{k=1}^{n} \lambda_i X_i$ **admet une espérance donnée par :
$$E\left(\sum_{k=1}^{n} \lambda_i X_i\right) = \sum_{k=1}^{n} \lambda_i E(X_i)$$

4. **Montrer que $E(S) = N \sum_{k=1}^{N} \frac{1}{k}$.**

   $S = \sum_{k=1}^{N} S_k$ donc par linéarité $E(S) = \sum_{k=1}^{N} E(S_k)$.
   On sait que si $X \hookrightarrow \mathcal{G}(p)$, $E(X) = \frac{1}{p}$: on en déduit
   $$\begin{aligned} E(S) &= \sum_{k=1}^{N} E(S_k) \\ &= \sum_{k=1}^{N} \frac{1}{\frac{N-k+1}{N}} \\ &= \sum_{k=1}^{N} \frac{N}{N-k+1} \qquad \text{changement d'indice } \ell = N-k+1 \\ E(S) &= N \sum_{\ell=1}^{N} \frac{1}{\ell} \end{aligned}$$

   (NB : avec ce changement les bornes se comportent comme suit : si $k = 1$, $\ell = N - 1 + 1 = N$, et si $k = N$, $\ell = N - N + 1 = 1$; et $\sum_{\ell=N}^{1} = \sum_{\ell=1}^{N}$... car l'addition est commutative !)

Voir le TD3 : on montre alors que $\mathrm{E(S)} \underset{\mathrm{N}\to+\infty}{\sim} \mathrm{N}\ln(\mathrm{N})$**.**

On donne maintenant des majorations sur les déviations à cette espérance : le temps moyen étant celui calculé ci-dessus, quelle est la probabilité qu'on doive attendre plus longtemps ?

Pour $m \in \mathbb{N}^*$**, et** $i \in [\![1,\mathrm{N}]\!]$**, on note** $\mathrm{B}_{i,m}$ **l'événement : «au jour** m**, on n'a toujours pas obtenu la vignette** i**».**

5. **Calculer** $\mathrm{P}(\mathrm{B}_{i,m})$ **pour tout** $m \in \mathbb{N}^*$ **et** $i \in [\![1,\mathrm{N}]\!]$**.**

   La probabilité qu'un paquet contienne une vignette autre que la i est égale à $\dfrac{\mathrm{N}-1}{\mathrm{N}}$ (toutes les vignettes apparaissant de manière équiprobable).
   $\mathrm{B}_{i,m}$ est l'événement : «les m premiers paquets achetés contiennent une vignette autre que i». Les vignettes des paquets successifs sont indépendantes : on en déduit

   $$\forall\, i \in [\![1,\mathrm{N}]\!],\ \forall\, m \in \mathbb{N}^*,\ \mathrm{P}(\mathrm{B}_{i,m}) = \left(\frac{\mathrm{N}-1}{\mathrm{N}}\right)^m$$

6. **Exprimer l'événement** $(\mathrm{S} > m)$ **en fonction des événements** $\mathrm{B}_{i,m}$ **(avec explications !!).**

   $(\mathrm{S} > m)$ est l'événement : «il faut un temps $> m$ pour compléter sa collection» ; ou encore : «au bout du temps m, on n'a pas eu toutes les vignettes» ; ou encore : «il existe (au moins) une vignette qu'on n'a toujours pas obtenue au jour m».

   Avec la dernière formulation, on voit que

   $$(\mathrm{S} > m) = \bigcup_{i=1}^{\mathrm{N}} \mathrm{B}_{i,m}$$

7. **On rappelle la *formule de Boole* : si** $(\mathrm{A}_1, \dots, \mathrm{A}_n)$ **sont des événements, on a :**

   $$\mathrm{P}\left(\bigcup_{i=1}^{n} \mathrm{A}_i\right) \leqslant \sum_{i=1}^{n} \mathrm{P}(\mathrm{A}_i)$$

   Montrer que $\mathrm{P}(\mathrm{S} > m) \leqslant \mathrm{N}\left(1 - \dfrac{1}{\mathrm{N}}\right)^m$**.** Directement avec la formule de Boole et les deux questions précédentes :

   $$\mathrm{P}(\mathrm{S} > m) = \mathrm{P}\left(\bigcup_{i=1}^{\mathrm{N}} \mathrm{B}_{i,m}\right) \leqslant \sum_{i=1}^{\mathrm{N}} \mathrm{P}(\mathrm{B}_{i,m}) = \sum_{i=1}^{\mathrm{N}} \mathrm{P}(\mathrm{B}_{i,m}) = \sum_{i=1}^{\mathrm{N}} \left(\frac{\mathrm{N}-1}{\mathrm{N}}\right)^m = \mathrm{N}\left(1 - \frac{1}{\mathrm{N}}\right)^m$$

8. **Montrer :** $\forall\, x > -1, \ln(1+x) \leqslant x$**. En déduire :**

   $$\forall\, m \in \mathbb{N}^*, \mathrm{P}(\mathrm{S} > m) \leqslant \mathrm{N}e^{-m/\mathrm{N}}$$

   On étudie la fonction g tq $\forall\, x > -1,\ g(x) = \ln(1+x) - x$. Avec dérivée + tableau de signes, on constate que sa valeur maximale est atteinte en $x = 0$, et vaut $g(0) = 0$. Donc : $\forall\, x > -1,\ g(x) \leqslant 0$ ce qui est ce qu'il fallait démontrer.
   Ou alors on remarque que f tq $f(x) = \ln(1+x)$ vérifie $f''(x) = -\dfrac{1}{(1+x)^2} > 0$, donc est concave, donc est au-dessous de sa tangente en 0 ; tangente qui a justement pour équation $y = x$.

   Ceci étant acquis : $\mathrm{N} \geqslant 2$ (non précisé mais sinon le problème est idiot), donc $-\dfrac{1}{\mathrm{N}} \geqslant -\dfrac{1}{2} > -1$ et on a donc $\ln\left(1 - \dfrac{1}{\mathrm{N}}\right) \leqslant -\dfrac{1}{\mathrm{N}}$.

   On reprend $\mathrm{P}(\mathrm{S} > m) \leqslant \mathrm{N}\left(1 - \dfrac{1}{\mathrm{N}}\right)^m = \mathrm{N}\exp\left(m \ln\left(1 - \dfrac{1}{\mathrm{N}}\right)\right)$.

$$\begin{aligned}
& \ln\left(1-\frac{1}{N}\right) \leqslant -\frac{1}{N} \\
\Rightarrow\ & m\ln\left(1-\frac{1}{N}\right) \leqslant -\frac{m}{N} \\
\Rightarrow\ & \exp\left(m\ln\left(1-\frac{1}{N}\right)\right) \leqslant e^{-m/N} \qquad \text{croissance de exp} \\
\Rightarrow\ & N\exp\left(m\ln\left(1-\frac{1}{N}\right)\right) \leqslant Ne^{-m/N}
\end{aligned}$$

d'où

$$P(S > m) \leqslant N\exp\left(m\ln\left(1-\frac{1}{N}\right)\right) \leqslant Ne^{-m/N}$$

Donner une majoration de la probabilité qu'il faille attendre au moins $N\ln(N)+c$ **jours pour obtenir la collection complète (avec** $c \in \mathbb{N}$**).**

On parle donc de majorer la probabilité $P(S > N\ln(N)+c)$. On utilise ce qu'on vient de démontrer :

$$P(S > N\ln(N)+c) \leqslant Ne^{-(N\ln(N)+c)/N} = Ne^{-\ln(N)-c/N} = N\underbrace{e^{-\ln(N)}}_{=1/N}e^{-c/N} = e^{-c/N}$$

Le majorant recherché est donc $e^{-c/N}$.