Devoir surveillé n°9
CORRECTION

Temps d'attente et 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 $\displaystyle\sum_{k=1}^{n} kx^{k-1}$.

f_n est une application polynomiale qui est également égale sur $[0,1[$ à $x \mapsto \dfrac{1-x^{n+1}}{1-x}$.
Elle est dérivable sur $[0,1[$ et on a donc

$$\forall\, x \in [0,1[, \qquad f_n'(x) = \sum_{k=0}^{n} kx^{k-1} = \sum_{k=1}^{n} kx^{k-1} = \frac{-(n+1)x^n(1-x)+1(1-x^{n+1}}{(1-x)^2}$$

$$\boxed{\forall\, x \in [0,1[\quad \sum_{k=1}^{n} kx^{k-1} = \frac{1-(n+1)x^n+nx^{n+1}}{(1-x)^2}}$$

On notera une autre façon de démontrer ce résultat. Pour tout $x \neq 1$:

$$\begin{aligned}\sum_{k=1}^{n} kx^{k-1} &= \sum_{k=1}^{n}\sum_{i=1}^{k} x^{k-1} = \sum_{1\leqslant i\leqslant k\leqslant n} x^{k-1} = \sum_{i=1}^{n}\sum_{k=i}^{n} x^{k-1} = \sum_{i=1}^{n} \frac{1-x^{n-i+1}}{1-x}x^{i-1}\\ &= \frac{1}{1-x}\sum_{i=1}^{n}\left(x^{i-1}-x^n\right) = \frac{1-x^n}{(1-x)^2} - n\frac{x^n}{1-x} = \frac{1-(n+1)x^n+nx^{n+1}}{(1-x)^2}\end{aligned}$$

   (b) Montrer que $\displaystyle\sum_{k=1}^{+\infty} kx^{k-1} = \frac{1}{(1-x)^2}$.

x est fixé dans $[0,1[$.
Notons $S_n = \displaystyle\sum_{k=1}^{n} kx^{k-1}$. C'est une suite, vérifiant pour $n \in \mathbb{N}^*$: $S_n = \dfrac{1}{(1-x)^2}(1-v_n+w_n)$,
où $v_n = (n+1)x^n$ et $w_n = nx^{n+1}$.
Or $x \in [0,1[$, donc par croissance comparée : $(v_n) \to 0$ et $(w_n) \to 0$.
Par addition de suites convergentes :

$$\boxed{(S_n) \text{ converge et } \sum_{k=1}^{+\infty} kx^{k-1} = \lim(S_n) = \frac{1}{(1-x)^2}.}$$

2. On considère une variable aléatoire U à valeurs dans $[\![0,n]\!]$. Montrer que $\mathbf{E}(U) = \displaystyle\sum_{k=0}^{n} \mathbf{P}(U>k)$

Pour tout $k \in [\![0,n]\!]$, $[U>k] = \biguplus\limits_{i=k+1}^{n} [U=i]$ (réunion disjointe).

donc $\mathbf{P}(U>k) = \displaystyle\sum_{i=k+1}^{n} \mathbf{P}(U=i)$ (par σ-additivité).
Ainsi (puisque $\mathbf{P}(U>n)=0$) :

$$\begin{aligned}\sum_{k=0}^{n} \mathbf{P}(U>k) &= \sum_{k=0}^{n-1} \mathbf{P}(U>k) = \sum_{k=0}^{n-1}\sum_{i=k+1}^{n} \mathbf{P}(U=i) = \sum_{0\leqslant k<i\leqslant n} \mathbf{P}(U=i)\\ &= \sum_{i=1}^{n}\sum_{k=0}^{i-1} \mathbf{P}(U=i) = \sum_{i=1}^{n} i\mathbf{P}(U=i) = \mathbf{E}(U)\end{aligned}$$

$$\boxed{\mathbf{E}(U)=\sum_{k=0}^{n}\mathbf{P}(U>k)}$$

---

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 $k<m$.

Il n'y a qu'une application de E_n sur E_1, l'application $f:x\mapsto 1$.
Si f est surjective de F sur G, alors nécessairement $\mathrm{card}(G)\leqslant \mathrm{card}F$,
donc $\sigma_{k,m}=0$ si $k<m$.
On en déduit $\sigma_{1,m}=0$ si $m>1$ et $\sigma_{1,1}=1$.
Enfin, une application surjective entre deux ensembles de même cardinaux est nécessairement bijective (et réciproquement).
Donc $\sigma_{n,n}=n!$: le nombre de bijections de E_n sur E_n.

$$\boxed{\sigma_{n,1}=1 \qquad \sigma_{1,m}=0 \text{ si } m>1, \qquad \sigma_{n,n}=n! \qquad \sigma_{k,m}=0 \text{ dès que } k<m.}$$

(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

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

Soit φ une surjection de E_n sur E_m.
— Notons $b=\varphi(a_n)$. Il peut être n'importe lequel des éléments de E_m. Il y a m possibilités.
Puis de deux choses l'une :
  — ou bien b n'admet qu'un antécédent par φ (i.e. a_n),
  et dans ce cas là $\varphi_{|E_n\setminus\{a_n\}}$ est une surjection de $E_n\setminus\{a_n\}$ sur $E_m\setminus\{b\}$.
  — ou bien b admet plusieurs antécédents par φ,
  et dans ce cas là $\varphi_{|E_n\setminus\{a_n\}}$ est une surjection de $E_n\setminus\{a_n\}$ sur E_m.

Réciproquement,
— à toute $b\in E_m$ et toute application surjective de $E_n\setminus\{a_n\}$ sur $E_m\setminus\{b\}$,
on peut associer une surjection de E_n sur E_m.
On peut faire ceci, m fois (les m éléments b différents).
— à toute application surjective de $E_n\setminus\{a_n\}$ sur un ensemble à E_m éléments,
on peut associer m surjections de E_n sur E_m.

Donc

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

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

---

m^n est le nombre d'applications de E_n vers E_m. Essayons de les décrire avec les surjections.
Une application ψ de E_n sur E_m est parfaitement décrite par :
— La donnée de $\psi(E_n)$, qui est un sous-ensemble de E_m.
— La surjection exacte de E_n sur $\psi(E_n)$.

Donc $\{\psi:E_n\to E_m\}=\displaystyle\biguplus_{A\subset E_m}\{\psi:E_n\to A,\ \text{surjective}\}$. La réunion étant disjointe :

$$m^n=\mathrm{card}\{\psi:E_n\to E_m\}=\sum_{A\subset E_m}\mathrm{card}\{\psi:E_n\to A,\ \text{surjective}\}$$

Or $\mathrm{card}\{\psi:E_n\to A,\ \text{surjective}\}$ ne dépend que de $\mathrm{card}(A)$ de 1 à m, on a donc

$$\begin{aligned}m^n&=\mathrm{card}\{\psi:E_n\to E_m\}=\sum_{k=1}^{m}\sum_{A\subset E_m,\mathrm{card}(A)=k}\mathrm{card}\{\psi:E_n\to A,\ \text{surjective}\}\\&=\sum_{k=1}^{m}\sum_{A\subset E_m\ \mid\ \mathrm{card}A=k}\sigma_{n,k}=\sum_{k=1}^{m}\sigma_{n,k}\sum_{A\subset E_m\ \mid\ \mathrm{card}A=k}1=\sum_{k=1}^{m}\sigma_{n,k}\binom{m}{k}\end{aligned}$$

Puisque $\sigma_{n,0} = 0$, on peut commencer la somme avec $k = 0$.

$$\boxed{m^n = \sum_{k=0}^{m} \binom{m}{k} \sigma_{n,k}}$$

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

On va exploiter le résultat de la question précédente :

$$\begin{aligned}
\sum_{j=0}^{m} (-1)^{m-j} \binom{m}{j} j^n &= \sum_{j=0}^{m} \left((-1)^{m-j} \binom{m}{j} \sum_{k=0}^{j} \binom{j}{k} \sigma_{n,k} \right) = \sum_{0 \leqslant k \leqslant j \leqslant m} (-1)^{m-j} \binom{m}{j} \binom{j}{k} \sigma_{n,k} \\
&= \sum_{k=0}^{m} \sigma_{n,k} \left(\sum_{j=k}^{m} (-1)^{m-j} \frac{m!j!}{j!(m-j)!k!(j-k)!} \right) = \sum_{k=0}^{m} \sigma_{n,k} \binom{m}{k} \left(\sum_{j=k}^{m} (-1)^{m-j} \binom{m-k}{j-k} \right) \\
&= \sum_{k=0}^{m} \sigma_{n,k} \binom{m}{k} \underbrace{\left(\sum_{i=0}^{m-k} (-1)^{m+k-i} \binom{m-k}{i} \right)}_{i=j-k}
\end{aligned}$$

Puis, comme $m + k - i$ et $m - k - i$ ont même parité ($(m + k - i) - (m - k - i) = 2k$ pair), on a donc $(-1)^{m+k-i} = (-1)^{m-k-i}$,

Et ensuite d'après la formule de Newton, pour $m - k \neq 0$:

$$\sum_{i=0}^{m-k} (-1)^{m-k-i} \binom{m-k}{i} = ((-1) + 1)^{m-k} = 0$$

alors que pour $m - k = 0$: $\displaystyle\sum_{i=0}^{0} (-1)^{-i} \binom{0}{i} = 1$

Donc

$$\sum_{j=0}^{m} (-1)^{m-j} \binom{m}{j} j^n = \sum_{k=0}^{m} \sigma_{n,k} \binom{m}{k} \delta_{m-k,0} = \sigma_{n,m} \binom{m}{m}$$

$$\boxed{\sigma_{n,m} = \sum_{j=0}^{m} (-1)^{m-j} \binom{m}{j} j^n}$$

Remarque :

On a fait une inversion de Pascal qui permet de passer d'une relation à une autre :

$$a_m = \sum_{k=0}^{m} \binom{m}{k} b_k \Longleftrightarrow b_m = \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} a_k$$

On aurait également raisonner :

— avec un produit matriciel : la matrice du triangle de Pascal qui s'inverse en celle avec des $(-1)^{i+j} \binom{i}{j}$.
(On peut démontrer ce résultat avec l'isomorphisme de $\mathbb{R}_n[X]$, $P \mapsto P(X + 1)$ dont la matrice dans la base canonique est exactement la première matrice, et celle de sa réciproque : $P \mapsto P(X - 1)$ est la seconde matrice).

— avec un raisonnement sur les ensembles et le crible de Poincaré.
Il s'agit alors de faire un raisonnement comparable à celui de la question précédente, mais dans l'autre sens. . .
Notons, pour tout $b \in E_m$, A_b, l'ensemble des applications de E_n dans $E_m \setminus \{b\}$.
Alors une surjection est une application de $E_n \to E_m$ qui n'appartient pas à $\bigcup_{b \in E_m} A_b$
La réunion étant non (!) disjointe, on applique la formule du crible de Poincaré :

$$\text{Card} \bigcup_{b \in E_m} A_b = \sum_{k=1}^{m} (-1)^{k-1} \left(\sum_{I=\{b_{i_1},\ldots b_{i_k}\} \subset E_m} \text{card} A_{b_{i_1}} \cap \cdots \cap A_{b_{i_k}} \right)$$

Or card $\left(A_{b_{i_1}} \cap \cdots \cap A_{b_{i_k}} \right) = (m-k)^n$: nombre d'applications de E_n dans $E_m \setminus \{b_{i_1}, \ldots b_{i_k}\}$.

$$\text{Card} \bigcup_{b \in E_m} A_b = \sum_{k=1}^{m} (-1)^{k-1} \left(\sum_{I=\{b_{i_1},\ldots b_{i_k}\} \subset E_m} (m-k)^n \right) = \sum_{k=1}^{m} (-1)^{k-1} \left(\binom{m}{k} (m-k)^n \right)$$

On cherche alors

$$\sigma_{n,m} = m^n - \text{card}\left(\bigcup_{b\in E_m} A_b\right) = m^n - \sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k}(m-k)^n = +\sum_{k=0}^{m}(-1)^k\binom{m}{k}(m-k)^n$$

Le changement d'indice $h = m - k$ donne la formule recherchée.

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

D'après la question précédente, $\sigma_{n,m} = m^n + \sum\limits_{j=0}^{m-1}(-1)^{m-j}\binom{m}{j}j^n$. Donc

$$\frac{\sigma_{n,m}}{m^n} = 1 + \sum_{j=0}^{m-1}(-1)^{m-j}\binom{m}{j}\left(\frac{j}{m}\right)^n$$

Or pour tout $j \in [\![0, m-1]\!]$, $\left(\frac{j}{n}\right)^m \underset{n\to+\infty}{\longrightarrow} 0$ car $0 \leqslant \frac{j}{n} < 1$.
Ainsi, par addition de m suites convergentes vers 0 :

$$\boxed{\lim_{n\to+\infty}\frac{\sigma_{n,m}}{m^n}1}$$

m^n est le nombre d'application de E_n sur E_m.
Ainsi, lorsque n grandit, le nombre de fonctions surjectives tend vers le nombre d'applications. C'est cohérent : le nombre d'éléments de l'ensemble de départ tend vers l'infini, alors que celui du but reste constant ; les fonctions sont de plus en plus probablement surjective...

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 de lancers consécutifs qui donnent le même côté.
On dit que la première série 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 (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 $[L_1 > k]$.

L_1 peut pendre toutes la valeurs strictement entières et ne peut dépasser N (nombre maximum de lancers).
Pour tout $k \in [\![0, N-1]\!]$, $L_1 > k$ est l'événement : « les $k+1$ premiers lancers donnent que des piles ou ils ne donnent que des faces ».

$$[L_1 > k] = \left(\bigcap_{i=1}^{k+1} P_i\right) \biguplus \left(\bigcap_{i=1}^{k+1} \overline{P}_i\right)$$

Ces événements sont incompatibles, puis indépendants :

$$\mathbf{P}(L_1 > k) = \mathbf{P}\left(\bigcap_{i=1}^{k+1} P_i\right) + \mathbf{P}\left(\bigcap_{i=1}^{k+1} \overline{P}_i\right) = p^{k+1} + q^{k+1}$$

On a aussi, pour tout $k \in [\![1, N-1]\!]$, $[L_1 > k-1] = [L_1 = k] \uplus [L_1 > k]$. Donc

$$\mathbf{P}(L_1 = k) = \mathbf{P}(L_1 > k-1) - \mathbf{P}(L_1 > k) = p^k + q^k - p^{k+1} - q^{k+1} = p^k(1-p) + q^k(1-q)$$

$$\boxed{\forall\, k \in [\![1, N-1]\!], \mathbf{P}(L_1 = k) = pq(p^{k-1} + q^{k-1}) \text{ et } \mathbf{P}(L_1 = N) = \mathbf{P}(L_1 > N-1) = p^N + q^N}$$

Remarque :
On aurait pu évaluer directement la probabilité de l'événement

$$[L_1 = k] = \left(\bigcap_{i=1}^{k} P_i \cap \overline{P}_{k+1}\right) \biguplus \left(\bigcap_{i=1}^{k} \overline{P}_i \cap P_{k+1}\right)$$

(b) Vérifier que $\displaystyle\sum_{k=1}^{N} \mathbf{P}(L_1 = k) = 1$.

On peut vérifier avec le calcul

$$\boxed{\sum_{k=1}^{N} \mathbf{P}(L_1 = k) = pq\frac{1-p^{N-1}}{1-p} + pq\frac{1-q^{N-1}}{1-q} + p^N + q^N = p + q = 1}$$

(c) Montrer que

$$\mathbf{E}(L_1) = \frac{p^2+q^2}{pq} - \left(\frac{p^{N+1}}{q} + \frac{q^{N+1}}{p}\right)$$

On peut exploiter la formule trouvée dans le préliminaire (en ajoutant le cas $L_1 > 0$ de probabilité égale à $1 = p^1 + q^1$) :

$$\mathbf{E}(L_1) = \sum_{k=0}^{N} \mathbf{P}(L_1 > k) = \sum_{k=0}^{N-1} p^{k+1} + q^{k+1} + 0 = p\frac{1-p^N}{1-p} + q\frac{1-q^N}{1-q} = \frac{p^2+q^2-p^{N+2}-q^{N+2}}{pq}$$

$$\boxed{\mathbf{E}(L_1) = \frac{p^2+q^2}{pq} - \left(\frac{p^{N+1}}{q} + \frac{q^{N+1}}{p}\right)}$$

Remarque :

On peut aussi exploiter la formule plus classique (et la première formule des préliminaires) :

$$\begin{aligned}
\mathbf{E}(L_1) &= \sum_{k=1}^{N} k\mathbf{P}(L_1 = k) = pq\sum_{k=1}^{N-1} kp^{k-1} + kq^{k-1} + N(p^N + q^N) \\
&= pq\left(\frac{1 - Np^{N-1} + (N-1)p^N}{(1-p)^2} + \frac{1 - Np^{N-1} + (N-1)p^N}{(1-p)^2}\right) + Np^N + Nq^N \\
&= \frac{p - Np^N + (N-1)p^{N+1}}{q} + \frac{q - Nq^N + (N-1)q^{N+1}}{p} + \frac{Nqp^{N+1} + Npq^{N+1}}{pq} \\
&= \frac{p^2 + q^2 + N(qp^{N+1} + pq^{N+1} - p^{N+1} - q^{N+1}) + (N-1)(p^{N+2} + q^{N+2})}{pq} \\
&= \frac{p^2 + q^2 + N(p^{N+1}(q-1) + q^{N+1}(p-1)) + (N-1)(p^{N+2} + q^{N+2})}{pq} \\
&= \frac{p^2 + q^2 + N(-p^{N+2} - q^{N+2}) + (N-1)(p^{N+2} + q^{N+2})}{pq} = \frac{p^2 + q^2 - (p^{N+2} + q^{N+2})}{pq}
\end{aligned}$$

---

5. Etude de L_2

(a) Déterminer la loi conjointe du couple (L_1, L_2).

$L_2(\Omega) = [\![1, N-1]\!]$ Et pour tout $(h,k) \in [\![1, N]\!] \times [\![1, N-1]\!]$, on a l'égalité d'événements :

$$[L_1 = h, L_2 = k] = \begin{cases} \emptyset & \text{si } h+k > N \\ \left(\bigcap_{i=1}^{h} P_i \cap \bigcap_{i=h+1}^{h+k} \overline{P}_i \cap P_{h+k+1}\right) \biguplus \left(\bigcap_{i=1}^{h} \overline{P}_i \cap \bigcap_{i=h+1}^{h+k} P_i \cap \overline{P}_{h+k+1}\right) & \text{si } h+k < N \\ \left(\bigcap_{i=1}^{h} P_i \cap \bigcap_{i=h+1}^{h+k} \overline{P}_i\right) \biguplus \left(\bigcap_{i=1}^{h} \overline{P}_i \cap \bigcap_{i=h+1}^{h+k} P_i\right) & \text{si } h+k = N \end{cases}$$

Par incompatibilité puis indépendance :

$$\boxed{\mathbf{P}[L_1 = h, L_2 = k] = \begin{cases} 0 & \text{si } h+k > N \\ p^h q^k p + q^h p^k q = p^{h+1}q^k + q^{h+1}p^k & \text{si } h+k < N \\ p^h q^k + q^h p^k & \text{si } h+k = N \end{cases}}$$

(b) Déterminer la loi de L_2.

Les événements $([L_1 = h])_{h\in[\![1,N]\!]}$ forment un système complet d'événements, donc (F.P.T.) ; Pour tout $k \in [\![1, N-1]\!]$,

$$\begin{aligned} \mathbf{P}(L_2 = k) &= \sum_{h=1}^{N-1} \mathbf{P}(L_1 = h, L_2 = k) = \sum_{h=1}^{N-k-1} p^{h+1}q^k + q^{h+1}p^k + p^{N-k}q^k + q^{N-k}p^k \\ &= q^k \frac{1-p^{N-k-2}}{1-p} p^2 + p^k \frac{1-q^{N-k-2}}{1-q} q^2 + p^{N-k}q^k + q^{N-k}p^k \\ &= q^{k-1}(1-p^{N-k-2})p^2 + p^{k-1}(1-q^{N-k-2})q^2 + p^{N-k}q^k + q^{N-k}p^k \end{aligned}$$

$$\boxed{\mathbf{P}(L_2 = k) = q^{k-1}p^2 + p^{k-1}q^2 + q^{k-1}p^{N-k}(1+q) + p^{k-1}q^{N-k}(1+p)}$$

(c) L_1 et L_2 sont elles indépendantes ?

Comme $L_1 + L_2 \leqslant N$. Nécessairement, on a

$$\mathbf{P}(L_1 = N-1) \times \mathbf{P}(L_2 = N-1) \neq \mathbf{P}(L_1 = N-1 \cap L_2 = N-1) = 0$$

$$\boxed{\text{Donc } L_1 \text{ et } L_2 \text{ ne sont pas indépendantes.}}$$

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

Toutes séries suivantes sont à termes positifs. Le fait que la somme est finie (bornée) est équivalente à la convergence de la série.
Nous ferons donc d'abord les calculs dans $\overline{\mathbb{R}_+}$, selon le résultat (fini ou non), nous en conclue-rons la convergence (ou non) et la valeur de la somme.

6. Déterminer la loi de L_1 en n'oubliant pas $\mathbf{P}(L_1 = +\infty)$ puis $\mathbf{E}(L_1)$.

Cette fois-ci $L_1(\Omega) = \mathbb{N}^* \cup \{+\infty\}$. Mais on peut toujours exploiter la même méthode, ce qui donne toujours :

$$\mathbf{P}(L_1 = k) = pq(p^{k-1} + q^{k-1})$$

On vérifie alors $\displaystyle\sum_{k=1}^{+\infty} \mathbf{P}(L_1 = k) = pq\left(\frac{1}{1-p} + \frac{1}{1-q}\right) = p + q = 1$ et donc $\mathbf{P}(L_1 = +\infty) = 0$.

$$\boxed{L_1(\Omega) = \mathbb{N}^* \cup \{+\infty\},\ \mathbf{P}(L_1 = k) = pq(p^{k-1} + q^{k-1}) \text{ et } \mathbf{P}(L_1 = +\infty) = 0.}$$

Le calcul de l'espérance s'obtient comme par passage à la limite pour $N \to +\infty$

$$\boxed{\mathbf{E}(L_1) = \frac{p^2 + q^2}{pq}}$$

---

7. Montrer que pour tout $k \in \mathbb{N}^*$, $\mathbf{P}(L_2 = k) = p^2q^{k-1} + q^2p^{k-1}$.

Remarque : Il s'agit du passage à la limite $N \to +\infty$ de la réponse à la question 5.b).
Si on veut exploiter cela, il faut alors le justifier ! Proprement !

Sinon, on peut tout recalculer, mais c'est plus simple car il n'y a pas de problème du type

$h+k \geqslant N$.
On a en effet, pour tout $h,k \in \mathbb{N}^*$

$$[L_1 = h, L_2 = k] = \left(\bigcap_{i=1}^{h} P_i \cap \bigcap_{i=h+1}^{h+k} \overline{P}_i \cap P_{h+k+1}\right) \biguplus \left(\bigcap_{i=1}^{h} \overline{P}_i \cap \bigcap_{i=h+1}^{h+k} P_i \cap \overline{P}_{h+k+1}\right)$$

Par incompatibilité puis indépendance :

$$\mathbf{P}[L_1 = h, L_2 = k] = p^h q^k p + q^h p^k q = p^{h+1} q^k + q^{h+1} p^k$$

Puis par la formule des probabilités totales (série convergente) :

$$\mathbf{P}(L_2 = k) = \sum_{h=1}^{+\infty} \mathbf{P}(L_1 = h, L_2 = k) = \sum_{h=1}^{+\infty} p^{h+1} q^k + q^{h+1} p^k = q^k p^2 \frac{1}{1-p} + p^k q^2 \frac{1}{1-q}$$

$$\boxed{\text{Pour tout } k \in \mathbb{N}^*,\ \mathbf{P}(L_2 = k) = p^2 q^{k-1} + q^2 p^{k-1}.}$$

---

8. Calculer $\mathbf{E}(L_2)$ et la comparer à $\mathbf{E}(L_1)$.

La série suivante est convergente (somme de séries géométriques dérivées), on peut directement passer à la limite (en exploitant les résultats des préliminaires) :

$$\sum_{k=1}^{+\infty} k\mathbf{P}(L_2 = k) = p^2 \sum_{k=1}^{+\infty} q^{k-1} + q^2 \sum_{k=1}^{+\infty} p^{k-1} = p^2 \frac{1}{(1-p)^2} + q^2 \frac{1}{(1-q)^2} = 2$$

$$\boxed{\mathbf{E}(L_2) = 2 \neq \mathbf{E}(L_1) \text{ (sauf si } p = q = \frac{1}{2})}$$

---

9. Donner une C.N.S simple pour que L_1 et L_2 soient indépendantes.
On rappelle que $p \in]0,1[$.

On a calculé plus haut : $\mathbf{P}(L_1 = h) = pq(p^{h-1} + q^{h-1})$ et $\mathbf{P}(L_2 = k) = p^2 q^{k-1} + q^2 p^{k-1}$,
mais aussi $\mathbf{P}(L_1 = h, L_2 = k) = p^{h+1} q^k + q^{h+1} p^k$.
Si $p = q = \frac{1}{2}$,
Alors $\mathbf{P}(L_1 = h) = \frac{1}{2^h}$ et $\mathbf{P}(L_2 = k) = \frac{1}{2^k}$ et $\mathbf{P}(L_1 = h, L_2 = k) = \frac{1}{2^{h+k}}$
Et donc pour tout $h,k \in \mathbb{N}^*$, $\mathbf{P}(L_1 = h, L_2 = k) = \mathbf{P}(L_1 = h) \times \mathbf{P}(L_2 = k)$.
Ainsi L_1 et L_2 sont indépendantes (pour $\mathbf{P}$).
Réciproquement, si L_1 et L_2 sont indépendantes,
alors pour tout $h,k \in \mathbb{N}^*$, $\mathbf{P}(L_1 = h, L_2 = k) = \mathbf{P}(L_1 = h) \times \mathbf{P}(L_2 = k)$.
En particulier pour $h = 1, k = 1$: $\mathbf{P}(L_1 = 1) \times \mathbf{P}(L_2 = 1) = 2pq(p^2 + q^2)$,
alors que $\mathbf{P}(L_1 = 1, L_2 = 1) = p^2 q + q^2 p$.
On a donc nécessairement : $2pq(p^2+q^2) = p^2 q + q^2 p$ i.e. $2(p^2+q^2) = p+q = 1$ en simplifiant par $pq \neq 0$.
Donc $p^2+q^2 = \frac{1}{2}$ donc $\frac{1}{2} = p^2+(1-p)^2 = p^2+1-2p+p^2$, donc $0 = 2p^2-2p+\frac{1}{2} = (2p-1)(p-\frac{1}{2})$.
Ainsi, nécessairement $p = \frac{1}{2}$ et $q = 1-p = \frac{1}{2}$.

$$\boxed{L_1 \text{ et } L_2 \text{ soient indépendantes si et seulement si } p = q = \tfrac{1}{2}.}$$

---

10. Donner de façon générale la loi de L_i.

Cette question n'est pas triviale, on voit que dans le cas général, les lois de L_1 et de L_2 sont assez différentes, en particulier en ce qui concerne leur espérance.

En fait cela dépend de la parité de i. D'abord notons que $L_i(\Omega)=\mathbb{N}^*$.

$$\begin{aligned}
\mathbf{P}(L_i=h) &= \sum_{k_1=1}^{+\infty}\sum_{k_2=1}^{+\infty}\cdots\sum_{k_{i-1}=1}^{+\infty}\mathbf{P}(L_1=k_1,L_2=k_2,\dots L_{i-1}=k_{i_1},L_i=h)\\
&= \begin{cases}\displaystyle\sum_{k_1=1}^{+\infty}\sum_{k_2=1}^{+\infty}\cdots\sum_{k_{i-1}=1}^{+\infty} p^{k_1}q^{k_2}\cdots p^{k_{i-1}}q^hp+q^{k_1}p^{k_2}\cdots q^{k_{i-1}}p^hq+ & \text{si } i \text{ pair}\\ \displaystyle\sum_{k_1=1}^{+\infty}\sum_{k_2=1}^{+\infty}\cdots\sum_{k_{i-1}=1}^{+\infty} p^{k_1}q^{k_2}\cdots q^{k_{i-1}}p^hq+q^{k_1}p^{k_2}\cdots p^{k_{i-1}}q^hp+ & \text{si } i \text{ impair}\end{cases}\\
&= \begin{cases}\frac{p}{1-p}\frac{q}{1-q}\cdots\frac{p}{1-p}q^hp+\frac{q}{1-q}\frac{p}{1-p}\cdots\frac{q}{1-q}p^hq & \text{si } i \text{ pair}\\ \frac{p}{1-p}\frac{q}{1-q}\cdots\frac{q}{1-q}p^hq+\frac{q}{1-q}\frac{p}{1-p}\cdots\frac{p}{1-p}q^hp & \text{si } i \text{ impair}\end{cases}\\
&= \begin{cases}q^{h-1}p^2+p^{h-1}q^2 & \text{si } i \text{ pair}\\ p^hq+q^hp=pq(p^{h-1}+q^{h-1}) & \text{si } i \text{ impair}\end{cases}
\end{aligned}$$

$$\boxed{L_i=\begin{cases}L_1 & \text{si } i \text{ impair}\\ L_2 & \text{si } i \text{ pair}\end{cases}}$$

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 observées lors des n premiers lancers.

11. Donner les lois de N_1, N_2 et leur espérance.

N_1 est une variable aléatoire certaine, égale à 1.
$N_2(\Omega)=\{1,2\}$. On a $[N_2=2]=P_1\cap\overline{P_2}\uplus\overline{P_1}\cap P_2$.
Donc $\mathbf{P}(N_2=2)=p(1-p)+(1-p)p=2\frac{1}{4}=\frac{1}{2}$.
Puis $\mathbf{P}(N_2=1)=1-\mathbf{P}(N_2=2)=\frac{1}{2}$ et enfin $\mathbf{E}(N_2)=\frac{1}{2}\times1+\frac{1}{2}\times2=\frac{3}{2}$

$$\boxed{N_1(\Omega)=\{1\},\quad \mathbf{E}(N_1)=1 \qquad\text{et}\qquad N_2\hookrightarrow U([\![1,2]\!]),\quad \mathbf{E}(N_2)=\frac{3}{2}}$$

---

12. Montrer de manière rigoureuse que pour $n\geqslant 2$ et $k\in[\![1,n]\!]$,

$$\mathbf{P}(N_n=k)=\frac{1}{2}\mathbf{P}(N_{n-1}=k)+\frac{1}{2}\mathbf{P}(N_{n-1}=k-1)$$

On raisonne selon que le dernier lancer (le n-ième) débute une nouvelle série ou non.
Notons D_n, l'événement : « le lancer n est différent du lancer $n-1$ ».
On a donc

$$[N_n=k]=[(N_n=k)\cap D_n]\uplus[(N_n=k)\cap\overline{D_n}]=[(N_{n-1}=k-1)\cap D_n]\uplus[(N_{n-1}=k)\cap\overline{D_n}]$$

Donc (événements incompatibles) :

$$\mathbf{P}(N_n=k)=\mathbf{P}[(N_{n-1}=k-1)\cap D_n]+\mathbf{P}[(N_{n-1}=k)\cap\overline{D_n}]$$

$$\mathbf{P}(N_n=k)=\mathbf{P}(N_{n-1}=k-1)\times\mathbf{P}_{N_{n-1}=k-1}(D_n)+\mathbf{P}(N_{n-1}=k)\times\mathbf{P}_{D_{n-1}=k}(\overline{D_n})$$

Enfin, la probabilité de réalisation de l'événement D_n vaut $\frac{1}{2}$, quels que soient les tirages précédents. De même pour $\overline{D_n}$.

$$\boxed{\mathbf{P}(N_n=k)=\frac{1}{2}\mathbf{P}(N_{n-1}=k)+\frac{1}{2}\mathbf{P}(N_{n-1}=k-1)}$$

On notera que la formule est vraie même pour $k=1$, où $\mathbf{P}(N_{n-1}=0)=0$ et pour $k=n$ où $\mathbf{P}(N_{n-1}=n)=0$.

---

13. On pose $G_n(s)=\displaystyle\sum_{k=1}^{n}\mathbf{P}(N_n=k)s^k$.

(a) Montrer que pour $n \geqslant 2$, $G_n(s) = \frac{1+s}{2} G_{n-1}(s)$.

Soit $n \geqslant 2$. D'après la formule précédente, pour $s \in \mathbb{R}$:

$$\begin{aligned} G_n(s) &= \sum_{k=1}^{n} \mathbf{P}(N_n = k)s^k = \frac{1}{2}\left(\sum_{k=1}^{n} \mathbf{P}(N_{n-1} = k)s^k + \sum_{k=1}^{n} \mathbf{P}(N_{n-1} = k-1)s^k\right) \\ &= \frac{1}{2}G_{n-1}(s) + \frac{s}{2}\sum_{k=1}^{n} \mathbf{P}(N_{n-1} = k-1)s^{k-1} = \frac{1}{2}G_{n-1}(s) + \frac{s}{2}\sum_{k=1}^{n-1} \mathbf{P}(N_{n-1} = h)s^h \\ &= \frac{1}{2}(1+s)G_{n-1}(s) \end{aligned}$$

où l'on a exploité que $\mathbf{P}(N_{n-1} = n) = \mathbf{P}(N_{n-1} = 0) = 0$.

$$\boxed{\forall\, n \geqslant 2, \forall\, s \in \mathbb{R}, \qquad G_n(s) = \frac{1+s}{2} G_{n-1}(s)}$$

(b) En déduire $G_n(s)$ puis le nombre moyen de séries lors des n premiers lancers.

$G_1(s) = \mathbf{P}(N_1 = 1)s^1 = 1s = s$.
On reconnait une suite géométrique. On a donc, pour tout $n \in \mathbb{N}^*$, $s \in \mathbb{R}$:

$$\boxed{G_n(s) = \left(\frac{1+s}{2}\right)^{n-1} G_1(s) = s\left(\frac{1+s}{2}\right)^{n-1}}$$

On a alors, par dérivation polynomiale (pour $n \geqslant 2$) :

$$G_n'(s) = \sum_{k=1}^{n} k\mathbf{P}(N_n = k)s^{k-1} = \left(\frac{1+s}{2}\right)^{n-1} + (n-1)\frac{s}{2}\left(\frac{1+s}{2}\right)^{n-2}$$

Et donc en prenant la valeur en $s = 1$

$$\boxed{\mathbf{E}(N_n) = \sum_{k=1}^{n} k\mathbf{P}(N_n = k) = G_n'(1) = 1 + \frac{n-1}{2} = \frac{n+1}{2}}$$

Pour $n = 1$, ou $n = 2$, cette expression reste juste (cf. calcul en 11.)

(c) On considère $n \geqslant 2$. Calculer $G_n''(s)$, en déduire $\mathbf{V}(N_n)$.
On pourra commencer par montrer que $\mathbf{V}(N_n) = G_n''(1) + G_n'(1) - [G_n'(1)]^2$.

G_n est dérivable deux fois (polynôme) et pour tout $s \in \mathbb{R}$

$$G_n''(s) = \sum_{k=1}^{n} k(k-1)\mathbf{P}(N_n = k)s^{k-2} = \frac{n-1}{2}\left(\frac{1+s}{2}\right)^{n-2} + (n-1)\frac{1}{2}\left(\frac{1+s}{2}\right)^{n-2} + (n-1)(n-2)\frac{s}{4}\left(\frac{1+s}{2}\right)^{n-3}$$

(vraie également en $n = 2$). On a donc

$$G_n''(1) = \sum_{k=1}^{n} k(k-1)\mathbf{P}(N_n = k) = \mathbf{E}(N_n(N_n - 1)) = \frac{n-1}{2} + \frac{n-1}{2} + \frac{(n-1)(n-2)}{4}$$

d'après la formule de transfert.
Donc (par linéarité de l'espérance) :

$$\mathbf{V}(N_n) = \mathbf{E}(N_n^2) - [\mathbf{E}(N_n)]^2 = \mathbf{E}(N_n(N_n - 1)) + \mathbf{E}(N_n) - [\mathbf{E}(N_n)]^2 = G_n''(1) + G_n'(1) - [G_n'(1)]^2$$

$$\boxed{\mathbf{V}(N_n) = \frac{(n-1)(n+2)}{4} + \frac{n+1}{2} - \frac{(n+1)^2}{4} = \frac{n^2 + n - 2 + 2n + 2 - n^2 - 2n - 1}{4} = \frac{n-1}{4}}$$

On vérifie que la formule marche également pour $n = 1$: $\mathbf{V}(N_1) = 0$
et pour $n = 2$: $\mathbf{V}(N_2) = \frac{1}{4}$ (loi uniforme avec $n = 2$)

(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 $\dfrac{3}{4}$.

On a les équivalences d'événements :

$$\frac{n}{2}+\sqrt{n} \leqslant N_n \leqslant \frac{n}{2}+1+\sqrt{n} \Longleftrightarrow -\frac{1}{2}-\sqrt{n} \leqslant N_n - \frac{n+1}{2} \leqslant \frac{1}{2}+\sqrt{n} \Longleftrightarrow \left|N_n - \mathbf{E}(N_n)\right| \leqslant \frac{1}{2}+\sqrt{n}$$

Or l'inégalité de Bienaymé-Tchebychev donne (pour tout $\epsilon > 0$) :

$$\mathbf{P}\big(|N_n - \mathbf{E}(N_n)| \geqslant \epsilon\big) \leqslant \frac{\mathbf{V}(N_n)}{\epsilon^2}$$

Donc avec $\epsilon \leftarrow \frac{1}{2} + \sqrt{n}$:

$$\mathbf{P}\left(\frac{n}{2} - \sqrt{n} \leqslant N_n \leqslant \frac{n}{2} + 1 + \sqrt{n}\right) = \mathbf{P}\left(\left|N_n - \mathbf{E}(N_n)\right| \leqslant \frac{1}{2} + \sqrt{n}\right) = 1 - \mathbf{P}\left(\left|N_n - \mathbf{E}(N_n)\right| \geqslant \frac{1}{2} + \sqrt{n}\right)$$

$$\mathbf{P}\left(\frac{n}{2} - \sqrt{n} \leqslant N_n \leqslant \frac{n}{2} + 1 + \sqrt{n}\right) \geqslant 1 - \frac{\mathbf{V}(N_n)}{(\frac{1}{2} + \sqrt{n})^2} = 1 - \frac{(n-1)}{4(n + \sqrt{n} + \frac{1}{4})}$$

Or $\sqrt{n} + \frac{1}{4} \geqslant -1$, donc $\dfrac{1}{n + \sqrt{n} + \frac{1}{4}} \leqslant \dfrac{1}{n-1}$, ainsi $-\dfrac{n-1}{4(n + \sqrt{n} + \frac{1}{4})} \geqslant -\dfrac{1}{4}$.

Et donc

$$\boxed{\mathbf{P}\left(\frac{n}{2} - \sqrt{n} \leqslant N_n \leqslant \frac{n}{2} + 1 + \sqrt{n}\right) \geqslant 1 - \frac{1}{4} = \frac{3}{4}}$$

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 $E_m = \{x_1, x_2, \dots x_m\}$, 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 $(V_1, V_2, \dots V_i \dots)$ sont indépendantes. Enfin, on note pour tout $i \in \mathbb{N}^*$, $j \in [\![1, m]\!]$, $p_j = \mathbf{P}(V_i = x_j) \in]0, 1[$.
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'achat 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 ?

U_n^j est une indicatrice, donc elle suit une loi de Bernoulli de paramètre $\mathbf{P}(U_n^j = 1)$.
Il est plus simple de raisonner avec l'événement complémentaire :

$$U_n^j = 0 \Longleftrightarrow (V_1 \neq j) \cap (V_2 \neq j) \cap \dots \cap (V_n \neq j)$$

Ce sont des événements indépendants par hypothèse :

$$\mathbf{P}(U_n^j = 0) = \prod_{h=1}^{n} \mathbf{P}(V_h \neq j) = \prod_{h=1}^{n} (1 - p_j) = (1 - p_j)^n$$

$$\boxed{\text{Donc } U_n^j \text{ suit la loi de Bernoulli de paramètre } 1 - (1 - p_j)^n.}$$

(b) Les variables $(U_n^j)_{n\in\mathbb{N}}$ sont-elles indépendantes ?

On a l'implication (ou la croissance) :

$$U_1^j = 1 \Longrightarrow V_1 = x_j \Longrightarrow U_2^j = 1$$

Donc $\mathbf{P}(U_1^j = 1 \cap U_2^j = 0) = 0 \neq \mathbf{P}(U_1^j = 1) \times \mathbf{P}(U_2^j = 0)$.

$$\boxed{\text{Les variables } (U_n^j)_{n\in\mathbb{N}} \text{ ne sont pas indépendantes.}}$$

---

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).

U_n^j indique si la vignette j a été récupérée.

Donc, pour tout $k \in \mathbb{N}_m$, $\sum_{j=1}^{m} U_n^j = k$ signifie qu'exactement « k vignettes ont été récupérée », c'est exactement $T_n = k$.

$$\boxed{T_n = \sum_{j=1}^{m} U_n^j}$$

(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 ?

L'espérance est linéaire (même si les variables ne sont pas indépendantes) :

$$\mathbf{E}(T_n) = \mathbf{E}\left(\sum_{j=1}^{m} U_n^j\right) = \sum_{j=1}^{m} \mathbf{E}(U_n^j)$$

Or pour une loi de Bernoulli de paramètre p, l'espérance vaut p, donc $\mathbf{E}(U_n^j) = 1 - (1 - p_j)^n$ (pour tout j).

$$\boxed{\mathbf{E}(T_n) = \sum_{j=1}^{m} (1 - (1 - p_j)^n)}$$

Pour n tendant vers $+\infty$, $(1 - p_j)^n \to 0$ car $p_j \neq 1$ et donc par addition de m suites :

$$\boxed{\mathbf{E}(T_n) \to \sum_{j=1}^{m} 1 = m}$$

En achetant une infinité de tablettes de chocolat, on obtient toute la collection de vignettes.
Remarque :
On peut aussi voir que T_n est un estimateur convergent vers m (mais avec biais).

---

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 ?

Dans le cas uniforme, $p_j = \frac{1}{m}$, on a donc

$$\boxed{\mathbf{E}(T_n) = \sum_{j=1}^{m} 1 - (1 - \frac{1}{m})^n = m\left(1 - \left(\frac{m-1}{m}\right)^n\right)}$$

(b) Montrer que pour tout $n \in \mathbb{N}^*$, et tout $u \in]-1,1]$, $(1+u)^n \geqslant 1+nu$.

L'application $\theta : u \mapsto (1+u)^n$ est polynomiale donc de classe $\mathcal{C}^2$.
Elle vérifie $\theta''(u) = n(n-1)(1+u)^{n-2}$ (formule vraie même pour $n=1$).
On a donc θ'' qui est positive sur $]-1,1]$ et donc θ est convexe sur $]-1,1]$.
Ainsi, elle se trouve AU-DESSUS de toutes ces tangentes.
Or la tangente en $u=0$ a pour équation : $y = \theta'(0)(u-0)+\theta(0) = n(1+0)^{n-1}(u-0)+(1+0)n = nu+1$.

$$\boxed{\text{pour tout } n \in \mathbb{N}^*, \text{ et tout } u \in]-1,1],\ \theta(u) \geqslant 0, \text{ donc } (1+u)^n \geqslant 1+nu}$$

(c) Posons pour tout $j \in [\![1,m]\!]$, $\epsilon_j = p_j - \frac{1}{m}$. Que vaut $\displaystyle\sum_{j=1}^{m} \epsilon_j$?

$$\boxed{\sum_{j=1}^{m} \epsilon_j = \sum_{j=1}^{m} p_j - \sum_{j=1}^{m} \frac{1}{m} = 1-1=0.}$$

(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

$$\mathbf{E}(T_n) \geqslant m\left(1-\left(\frac{m-1}{m}\right)^n\right)$$

On a $p_j = \frac{1}{m}+\epsilon_j$ et donc

$$(1-p_j)^n = \left(1-\frac{1}{m}-\epsilon_j\right)^n = \left(1-\frac{1}{m}\right)^n \times \left(1+\frac{-\epsilon_j}{1-\frac{1}{m}}\right)^n$$

On note $u_j = \dfrac{\epsilon_j}{1-\frac{1}{m}} = \dfrac{m\epsilon_j}{m-1}$.
Comme $\epsilon_j = p_j - \frac{1}{m}$, alors $-\frac{1}{m} < \epsilon_j < 1-\frac{1}{m}$, donc $\frac{1}{m-1} > -u_j > -1$.
On peut appliquer l'inégalité trouvée en (b) :

$$(1+(-u_j))^n \geqslant 1-nu_j = 1-n\frac{m}{m-1}\epsilon_j$$

Puis, on multiplie par $\left(1-\frac{1}{m}\right)^n$ positif : $(1-p_j)^n \geqslant (1-\frac{1}{m})^n\left(1-\dfrac{n\epsilon_j}{1-\frac{1}{m}}\right)$.

$$\boxed{(1-p_j)^n \geqslant \left(1-\frac{1}{m}\right)^n - n\epsilon_j\left(1-\frac{1}{m}\right)^{n-1}}$$

On trouve donc, en sommant pour j de 1 à m :

$$\mathbf{E}(T_n) = \sum_{j=1}^{m}(1-(1-p_j)^n) = m - \sum_{j=1}^{m}(1-p_j)^n \leqslant m - \sum_{j=1}^{m}\left(1-\frac{1}{m}\right)^n(1-n\frac{m}{m-1}\epsilon_j)$$

$$\mathbf{E}(T_n) \leqslant m - m(1-\frac{1}{m})^n + (1-\frac{1}{m})^n \times n \times \frac{m}{m-1}\sum_{j=1}^{m}\epsilon_j$$

Or $\displaystyle\sum_{j=1}^{m}\epsilon_j = 0$ et $m-m(1-\frac{1}{m})^n = m\left(1-\left(\frac{m-1}{m}\right)^n\right)$

$$\boxed{\mathbf{E}(T_n) \leqslant m\left(1-\left(\frac{m-1}{m}\right)^n\right) \text{ (cas équiprobable).}}$$

Remarque :
Le cas équiprobable est celui qui donne la plus grande des espérances du nombre de vignettes différentes obtenues après n achats.

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

On s'intéresse maintenant au nombre d'achat 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.

Avec un achat, on a assurément une image. Donc $[X_1 = 1]$ est un événement certain.

$$\boxed{X_1(\Omega) = \{1\} \text{ et } \mathbf{E}(X_1) = 1.}$$

---

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 $\mathbf{E}(Z_j) = \dfrac{m}{m-j}$.

On fixe $j \in [\![1, m-1]\!]$.
On note A_k l'événement : « le collectionneur, qui possède j vignettes différentes, a une nouvelle vignette lors du k-ième l'achat suivant ».

Par équiprobabilité du tirage des vignettes, pour tout $k \in \mathbb{N}^*$, $\mathbf{P}(\overline{A_k}) = \dfrac{j}{m}$ et $\mathbf{P}(A_k) = \dfrac{m-j}{m}$.
Il n'est pas impossible que Z_j prenne toutes les valeurs entières possibles, voire soit égale à $+\infty$.
Donc $Z_j(\Omega) = \mathbb{N}^* \cup \{+\infty\}$ Puis, pour $r \in \mathbb{N}^*$, on a l'égalité en événements :

$$[Z_j = r] \Longleftrightarrow \overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_{r-1}} \cap A_r$$

Ces événements étant indépendants :

$$\boxed{\mathbf{P}[Z_j = r] = \mathbf{P}(\bigcap_{i=1}^{r-1} \overline{A_i} \cap A_r) = \prod_{i=1}^{r^1} \mathbf{P}(\overline{A_i}) \times \mathbf{P}(A_r) = \left(\frac{j}{m}\right)^{r-1} \frac{m-j}{m}}$$

On a alors bien

$$\mathbf{P}(Z_j = +\infty) = 1 - \frac{m-j}{m} \sum_{r=1}^{+\infty} \left(\frac{j}{m}\right)^{r-1} = 1 - \frac{m-j}{m} \frac{1}{1 - \frac{j}{m}} = 0$$

Puis la série $\displaystyle\sum_{r \geqslant 1} r\mathbf{P}(Z = r)$ qui converge (géométrique de raison $\frac{j}{m} \in [0, 1[$), donc Z_j admet une espérance et d'après les questions du préliminaires :

$$\boxed{\mathbf{E}(Z_j) = \frac{m-j}{m} \times \frac{1}{(1 - \frac{j}{m})^2} = \frac{m}{m-j}}$$

---

19. En exploitant un télescopage, donner une expression de $\mathbf{E}(X_m)$ sous la forme d'une somme.

Donc : $X_m = \displaystyle\sum_{i=1}^{m-1} X_{i+1} - X_i + X_1 = X_1 + \sum_{i=1}^{m-1} Z_i$.
Par linéarité de l'espérance :

$$\mathbf{E}(X_m) = \mathbf{E}(X_1) + \sum_{i=1}^{m-1} \mathbf{E}(Z_i) = 1 + \sum_{i=1}^{m-1} \frac{m}{m-i}$$

Comme $1 = \frac{m}{m}$, en posant, $h = m - i$, on trouve

$$\boxed{\mathbf{E}(X_m) = m \sum_{h=1}^{m} \frac{1}{h}}$$

---

20. On note, pour tout $n \in \mathbb{N}^*$, $u_n = \sum_{i=1}^{n} \frac{1}{i} - \ln(n)$ et pour $n \geqslant 2$, $v_n = u_n - u_{n-1}$.

   (a) Donner un équivalent de v_n.

$$v_n = \sum_{i=1}^{n} \frac{1}{i} - \ln(n) - \sum_{i=1}^{n-1} \frac{1}{i} + \ln(n-1) = \frac{1}{n} + \ln(1 - \frac{1}{n})$$

$$v_n = \frac{1}{n} + \left(-\frac{1}{n} - \frac{1}{2n^2} + o(\frac{1}{n^2}) \right) = -\frac{1}{2n^2} + o(\frac{1}{n^2})$$

$$\boxed{v_n \sim \frac{-1}{2n^2}}$$

   (b) En déduire que la suite (u_n) est convergente.

La série de terme général v_n est donc une série à termes négatifs (à partir d'un certain rang). Son terme général est équivalent à $\frac{-1}{2n^2}$. Or d'après le critère de Riemann, la série $\sum \frac{1}{n^2}$ converge.
Donc la série de terme général v_n converge.
Donc la série de terme général $(u_{n+1} - u_n)$ converge, ce qui est équivalent à :

$$\boxed{\text{La suite } (u_n) \text{ converge.}}$$

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 a vu que $\mathbf{E}(X_m) = m \sum_{h=1}^{m} \frac{1}{h}$.
Or d'après la question précédente :

$$\sum_{i=1}^{n} \frac{1}{i} = \ln n + \gamma + o(1) \text{ pour } n \to +\infty$$

$$\boxed{\mathbf{E}(X_m) \underset{m \to +\infty}{=} m \ln m + \gamma + o(1)}$$

Plus le nombre de vignettes différentes est important, plus il sera nécessaire d'acheter des tablettes de chocolat, et ceci plus que linéairement !

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 $m - 1$?

Il s'agit de calculer la probabilité de l'événement : le nombre minimal d'achat pour obtenir les m vignettes différentes est n avec $n < m$.
Cet événement est impossible. Donc

$$\boxed{\text{Pour } n \text{ compris entre 1 et } m-1, \ \mathbf{P}(X_m = n) = 0.}$$

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 $h \in \mathbb{N}$, $\mathbf{P}(X_m = h) = \frac{m\sigma_{h-1,m-1}}{m^h}$.

Puisqu'il y a équiprobabilité entre les vignettes à chaque achat, on peut supposer qu'après h achats de tablette, toutes les distributions des vignettes possibles sont équiprobables.
Ainsi, à chaque situation (équiprobable) correspond une liste $(\varphi(1), \varphi(2), \dots \varphi(h))$ où pour tout $i \in [\![1, h]\!]$, $\varphi(i) \in E_m = \{x_1, x_2, \dots x_m\}$.
Il y a donc autant de situations que d'applications de $[\![1, h]\!]$ sur E_m.

Le nombre total de cas est donc m^h. Les situations favorables correspondent aux fonctions φ telle que $\varphi_{|[\![1,h-1]\!]}$ est une surjection de $[\![1,h-1]\!]$ sur $E_m \setminus \{\varphi(h)\}$. Il y a $m \times \sigma_{h-1,m-1}$ cas favorables (choix sur $\varphi(h)$ puis nombre de surjections).

$$\boxed{\text{Donc } \mathbf{P}(X_m = h) = \frac{m\sigma_{h-1,m-1}}{m^h} = \frac{\sigma_{h-1,m-1}}{m^{h-1}}.}$$

---

24. En déduire un équivalent et la limite de $\mathbf{P}(X_m = h)$ lorsque $h \to +\infty$. Interpréter ce résultat.

D'après la question 3.(e), $\sigma_{n,m} \underset{n\to+\infty}{\sim} m^n$.

Donc $\mathbf{P}(X_m = h) = \dfrac{\sigma_{h-1,m-1}}{m^{h-1}} \sim \dfrac{(m-1)^{h-1}}{m^{h-1}} = \left(1 - \frac{1}{m}\right)^{h-1}$. Ainsi :

$$\boxed{\mathbf{P}(X_m = h) \underset{h\to+\infty}{\sim} \left(1 - \frac{1}{m}\right)^{h-1} \quad \text{et} \quad \lim_{h\to+\infty} \mathbf{P}(X_m = h) = 0.}$$

La probabilité d'avoir besoin d'au moins h achats pour obtenir toute la collection de vignettes tend vers 0 pour h tendant vers $+\infty$ (quel que soit m).
