Notion de probabilité. Espaces probabilisés.

1. Deux archers tirent chacun leur tour sur une cible. Le premier qui touche a gagné. Le joueur qui commence a la probabilité p_1 de toucher à chaque tour et le second la probabilité p_2 (avec $p_1,\, p_2 \in]0,1[$).

a. Quelle est la probabilité que le premier joueur gagne ?

b. Montrer qu'il est presque sûr que le jeu se termine.

- -

a. Pour $n \in \mathbb{N}^*$, notons S_n l'événement: "le n-ème tir atteint la cible". Alors $P(S_{2k+1}) = p_1$ pour $k \in \mathbb{N}$, et $P(S_{2k}) = p_2$ pour $k \in \mathbb{N}^*$.

Notons ensuite A_n l'événement: "la cible est atteinte, pour la première fois, au n-ème tir". On a alors $A_n = \overline{S_1} \cap \overline{S_2} \cap \cdots \cap \overline{S_{n-1}} \cap S_n$. Les événements S_k étant mutuellement indépendants, on a

$$P(A_{2k+1}) = (1-p_1)^k(1-p_2)^k p_1 \quad (k \in \mathbb{N}) \quad \text{et} \quad P(A_{2k}) = (1-p_1)^k(1-p_2)^{k-1}p_2 \quad (k \in \mathbb{N}^*)\ .$$

Soit enfin G_1 l'événement: "le premier joueur gagne la partie". Alors $G_1 = \bigsqcup_{k=0}^{+\infty} A_{2k+1}$ (réunion disjointe). Donc

$$P(G_1) = \sum_{k=0}^{+\infty} P(A_{2k+1}) = p_1 \sum_{k=0}^{+\infty} \big((1-p_1)(1-p_2)\big)^k = \frac{p_1}{1-(1-p_1)(1-p_2)} = \frac{p_1}{p_1+p_2-p_1p_2}\ .$$

b. De la même façon, si G_2 est l'événement: "le deuxième joueur gagne la partie", on a

$$P(G_2) = \sum_{k=1}^{+\infty} P(A_{2k}) = p_2\,(1-p_1) \sum_{k=0}^{+\infty} \big((1-p_1)(1-p_2)\big)^k = \frac{p_2-p_1p_2}{p_1+p_2-p_1p_2}\ .$$

On observe que $P(G_1)+P(G_2)=1$ et, les événements G_1 et G_2 étant incompatibles, cela signifie que $P(G_1 \cup G_2) = 1$. L'événement $G_1 \cup G_2$, c'est-à-dire "le jeu se termine", est donc presque sûr.

---

2. Une urne contient une boule blanche et une boule rouge. On tire dans cette urne une boule, on note sa couleur et on la remet dans l'urne accompagnée de deux autres boules de la même couleur, puis on répète l'opération.

a. Quelle est la probabilité que les n premières boules tirées soient rouges ?

b. Quelle est la probabilité de tirer indéfiniment des boules rouges ?

c. Le résultat précédent reste-t-il vrai si on remet la boule accompagnée de trois autres boules de la même couleur ?

- -

a. Pour $n \in \mathbb{N}^*$, notons R_n l'événement: "le n-ème tirage amène une boule rouge". On a $P(R_1) = \dfrac{1}{2}$, puis $P(R_2|R_1) = \dfrac{3}{4}$ (puisque, si R_1 est réalisé, l'urne contient ensuite une boule blanche et trois rouges), et plus généralement $P(R_k \mid R_1 \cap \cdots \cap R_{k-1}) = \dfrac{2k-1}{2k}$ pour tout $k \geq 2$. Par la formule des probabilités composées, on obtient alors

$$P(R_1 \cap \cdots \cap R_n) = P(R_1)\cdot P(R_2|R_1)\cdot \cdots \cdot P(R_n \mid R_1 \cap \cdots \cap R_{n-1}) = \frac{1}{2}\times\frac{3}{4}\times\cdots\times\frac{2n-1}{2n}$$

Donc $P(R_1 \cap \cdots \cap R_n) = \dfrac{(2n)!}{(2^n\, n!)^2}$.

b. Notons R l'événement: "on tire indéfiniment des boules rouges", alors $R = \bigcap_{n=1}^{+\infty} R_n$, donc $0 \leq P(R) \leq P(R_1 \cap \cdots \cap R_n)$ pour $n \in \mathbb{N}^*$. Mais on a, par la formule de Stirling, $P(R_1 \cap \cdots \cap R_n) \underset{n\to+\infty}{\sim} \dfrac{1}{\sqrt{\pi n}}$, donc $\lim\limits_{n\to+\infty} P(R_1 \cap \cdots \cap R_n) = 0$ et, par encadrement, $P(R) = 0$.

c. Dans ce nouveau modèle, on a $P(R_k \mid R_1 \cap \cdots \cap R_{k-1}) = \dfrac{3k-2}{3k-1}$ pour tout $k \geq 2$, donc $P(R_1 \cap \cdots \cap R_n) = \prod\limits_{k=1}^{n} \left(\dfrac{3k-2}{3k-1}\right) = \prod\limits_{k=1}^{n} \left(1 - \dfrac{1}{3k-1}\right)$. Donc

$$\ln\big(P(R_1 \cap \cdots \cap R_n)\big) = \sum_{k=1}^{n} \ln\left(1 - \frac{1}{3k-1}\right)$$

et cette expression tend vers $-\infty$ car c'est une somme partielle d'une série à termes négatifs divergente. En effet, le terme général de la série est équivalent à $-\dfrac{1}{3k}$. On en déduit donc que $\lim\limits_{n\to+\infty} P(R_1 \cap \cdots \cap R_n) = 0$ puis, comme en **b.**, que $P(R) = 0$.

---

3. On suppose que la probabilité p_n pour qu'un couple ait exactement n enfants est de la forme $p_n = \alpha\, q^n$ pour n entier naturel non nul, où α est un réel strictement positif et $q \in]0, 1[$.

a. Que vaut p_0 ? Quelle condition doit-on imposer à α et q?

b. Soit k un entier naturel. En utilisant le cours sur les séries entières, calculer la somme $\sum\limits_{n=k}^{+\infty} \dbinom{n}{k} x^n$, où x est un réel tel que $|x| < 1$.

c. Sachant qu'à chaque naissance, la probabilité qu'il s'agisse d'une fille ou d'un garçon est la même, quelle est la probabilité pour qu'un couple admette exactement k filles ? *On traitera à part le cas $k = 0$.*

- -

a. Les événements E_n: "le couple a n enfants", pour $n \in \mathbb{N}$, forment un système complet, on doit donc avoir $\sum\limits_{n=0}^{+\infty} p_n = 1$, soit $p_0 = 1 - \sum\limits_{n=1}^{+\infty} p_n = 1 - \alpha \dfrac{q}{1-q} = \dfrac{1-(\alpha+1)q}{1-q}$. On doit par ailleurs avoir $p_0 \geq 0$, soit $(\alpha+1)q \leq 1$.

b. Posons $s(x) = \sum\limits_{n=0}^{+\infty} x^n = \dfrac{1}{1-x}$ pour $x \in]-1, 1[$. Un calcul classique, justifié par le cours sur les séries entières (on peut dériver terme à terme sur l'intervalle de convergence, les séries dérivées ont le même rayon de convergence) montre que, pour tout k entier naturel, on a

$$s^{(k)}(x) = \sum_{n=k}^{+\infty} n(n-1)\cdots(n-k+1)\, x^{n-k} = \sum_{n=k}^{+\infty} \frac{n!}{(n-k)!}\, x^{n-k} = k!\, \sum_{n=k}^{+\infty} \binom{n}{k} x^{n-k}\ .$$

Comme, par ailleurs, $\dfrac{\mathrm{d}^k}{\mathrm{d}x^k}\left(\dfrac{1}{1-x}\right) = \dfrac{k!}{(1-x)^{k+1}}$, on conclut

$$\forall x \in]-1,1[\quad \forall k \in \mathbb{N} \qquad \sum_{n=k}^{+\infty} \binom{n}{k} x^n = \frac{x^k}{(1-x)^{k+1}} .$$

c. Soit F_k l'événement: "le couple a k filles" ($k \in \mathbb{N}$), alors par la formule des probabilités totales,

$$P(F_k) = \sum_{n=0}^{+\infty} P(F_k|E_n)\,P(E_n) = \sum_{n=k}^{+\infty} P(F_k|E_n)\,P(E_n) .$$

Pour $n \geq k$, on a $P(F_k|E_n) = \dbinom{n}{k}\left(\dfrac{1}{2}\right)^k\left(\dfrac{1}{2}\right)^{n-k} = \dfrac{1}{2^n}\dbinom{n}{k}$, c'est une loi binomiale.

- Si $k \geq 1$, alors pour $n \geq k$, on a $P(E_n) = \alpha\,q^n$, donc

$$P(F_k) = \alpha \sum_{n=k}^{+\infty} \frac{1}{2^n}\binom{n}{k} q^n = \alpha \sum_{n=k}^{+\infty} \binom{n}{k}\left(\frac{q}{2}\right)^n = \alpha\,\frac{\left(\frac{q}{2}\right)^k}{\left(1-\frac{q}{2}\right)^{k+1}} = \frac{2\alpha\ q^k}{(2-q)^{k+1}} .$$

- Si $k = 0$, on a

$$\begin{aligned} P(F_0) &= P(E_0) + \sum_{n=1}^{+\infty} P(F_0|E_n)\,P(E_n) \\ &= \frac{1-(\alpha+1)q}{1-q} + \sum_{n=1}^{+\infty}\left(\frac{1}{2}\right)^n \alpha\,q^n \\ &= \frac{1-(\alpha+1)q}{1-q} + \alpha\,\frac{q}{2-q} \\ &= \frac{q^2-(\alpha+3)q+2}{(1-q)(2-q)} . \end{aligned}$$

Le lecteur vérifiera que $\displaystyle\sum_{n=0}^{+\infty} P(F_n) = 1$, ce qui le rassurera.

---

4. Soit (A_n) une suite d'événements sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. On pose $S = \displaystyle\bigcap_{n=0}^{+\infty}\left(\bigcup_{k=n}^{+\infty} A_k\right)$.

a. Montrer que S est un événement, i.e. $S \in \mathcal{A}$, et qu'il est réalisé si et seulement si une infinité des événements A_n sont réalisés.

b. Dans cette question et la suivante, on considère une suite infinie de lancers indépendants d'une pièce, la probabilité d'obtenir "Pile" à chaque lancer étant $p \in]0,1[$. Pour tout $n \in \mathbb{N}$, on considère l'événement A_n : "au cours des $2n$ premiers lancers, on obtient autant de Pile que de Face". Calculer $P(A_n)$ pour tout n.

c. Montrer que, pour tout n entier naturel, on a $\dbinom{2n}{n} \leq 4^n$. En déduire que, si $p \neq \dfrac{1}{2}$, la série $\displaystyle\sum_{n\geq 0} P(A_n)$ converge. Montrer alors que $P(S) = 0$.

- -

a. Une tribu est stable par réunion ou intersection finie ou dénombrable, donc $S \in \mathcal{A}$. Soit par ailleurs $\omega \in \Omega$. On a

$$\omega \in S \iff \forall n \in \mathbb{N} \quad \exists k \geq n \quad \omega \in A_k \ .$$

Cela signifie que l'ensemble des indices k tels que $\omega \in A_k$ est une partie de $\mathbb{N}$ non majorée, ou ce qui revient au même, une partie de $\mathbb{N}$ infinie.

b. On reconnaît un schéma de Bernoulli (répétition de $2n$ épreuves de Bernoulli indépendantes) de paramètres $2n$ et p, la probabilité d'apparition de n "succès" lors de $2n$ épreuves est alors $P(A_n) = \binom{2n}{n} p^n(1-p)^n$, c'est la loi binomiale $\mathcal{B}(2n, p)$.

c. L'inégalité $\binom{2n}{n} \leq 4^n$ est vraie pour $n = 0$ (c'est alors une égalité), et on récurre facilement après avoir vérifié que

$$\frac{\binom{2(n+1)}{n+1}}{\binom{2n}{n}} = \frac{(2n+2)!}{\big((n+1)!\big)^2}\,\frac{(n!)^2}{(2n)!} = \frac{2(2n+1)}{n+1} \leq \frac{4n+4}{n+1} = 4 \ .$$

Si $p \neq \frac{1}{2}$, alors $p(1-p) < \frac{1}{4}$ (étudier les variations de $x \mapsto x(1-x)$ sur $[0,1]$), et alors $0 \leq P(A_n) \leq \big[4p(1-p)\big]^n$ (majoration par une suite géométrique de raison < 1), donc par comparaison de séries à termes positifs, la série $\sum P(A_n)$ est convergente.

Posons $B_n = \bigcup\limits_{k=n}^{+\infty} A_k$, alors $P(B_n) \leq \sum\limits_{k=n}^{+\infty} P(A_k)$ par la propriété de sous-additivité, mais comme le reste d'ordre n d'une série convergente tend vers zéro lorsque n tend vers $+\infty$, on en tire que $\lim\limits_{n\to+\infty} P(B_n) = 0$. Enfin, on a $S \subset B_n$ pour tout n, donc par croissance d'une probabilité, $0 \leq P(S) \leq P(B_n)$ pour tout n. Comme $\lim\limits_{n\to+\infty} P(B_n) = 0$, on déduit que $P(S) = 0$.

Autrement dit, si la pièce est déséquilibrée ($p \neq \frac{1}{2}$), le jeu sera presque sûrement définitivement déséquilibré à partir d'un certain moment. J'essaie de m'expliquer plus clairement: par exemple, si $p > \frac{1}{2}$, alors à chaque lancer, il est plus probable d'obtenir "Pile" que "Face" et, si l'on répète indéfiniment des lancers de cette pièce, il est presque sûr (événement de probabilité 1) qu'à partir d'un certain moment, on comptabilisera toujours strictement plus de "Pile" que de "Face" dans les lancers déjà effectués. Bref, si une équipe est vraiment meilleure qu'une autre à un jeu qu'elles répètent indéfiniment, il est presque certain qu'il arrivera un moment où les deux équipes n'égaliseront plus, l'avantage restant définitivement à l'équipe la plus forte.

On a en fait prouvé le **lemme de Borel-Cantelli**: si une suite d'événements (A_n) est telle que la série $\sum P(A_n)$ converge, alors $P(\limsup A_n) = 0$, i.e. $P\Big(\bigcap\limits_{n\in\mathbb{N}} \Big(\bigcup\limits_{k\geq n} A_k\Big)\Big) = 0$, i.e. il est quasi-impossible qu'une infinité d'événements A_n se réalise (c'est un événement "négligeable", i.e. de probabilité nulle).

5*. Soit (A_n) une suite d'événements indépendants. Montrer que la probabilité qu'aucun des événements A_n ne soit réalisé est majorée par $M = \exp\Big(-\sum_{n=0}^{+\infty} P(A_n)\Big)$.

On pourra utiliser l'inégalité $\forall x \in \mathbb{R} \quad 1 - x \leq e^{-x}$.

- -

Pour tout n, on a, puisque les $\overline{A_n}$ sont aussi indépendants,

$$\textbf{(*)} : \qquad P\Big(\bigcap_{k=0}^{n} \overline{A_k}\Big) = \prod_{k=0}^{n} P(\,\overline{A_k}\,) = \prod_{k=0}^{n} \big(1 - P(A_k)\big) \leq \prod_{k=0}^{n} e^{-P(A_k)} = \exp\Big(-\sum_{k=0}^{n} P(A_k)\Big)\,.$$

Par continuité décroissante, $\lim\limits_{n\to+\infty} P\Big(\bigcap_{k=0}^{n} \overline{A_k}\Big) = P\Big(\bigcap_{k=0}^{+\infty} \overline{A_k}\Big)$.

Par ailleurs, $\lim\limits_{n\to+\infty} \exp\Big(-\sum_{k=0}^{n} P(A_k)\Big) = \exp\Big(-\sum_{n=0}^{+\infty} P(A_n)\Big) = M$ par continuité de l'exponentielle, en convenant que $M = 0$ si la série $\sum P(A_k)$ diverge, ce qui est cohérent. Par passage à la limite dans **(*)**, on obtient alors l'inégalité demandée.

NB.1. On déduit notamment que, si les A_n sont mutuellement indépendants et si la série $\sum P(A_n)$ diverge, alors il est presque sûr qu'aucun moins un des événements A_n se réalise.

NB.2. On est tout près du **deuxième lemme de Borel-Cantelli**, dit aussi **loi du zéro-un de Borel**, qui dit que, si une suite d'événements (A_n) mutuellement indépendants est telle que la série $\sum P(A_n)$ diverge, alors $P(\limsup A_n) = 1$, i.e. $P\bigg(\bigcap_{n\in\mathbb{N}} \Big(\bigcup_{k\geq n} A_k\Big)\bigg) = 1$, i.e. il est presque sûr qu'une infinité d'événements A_n se réalise. Il suffit d'appliquer ce qui précède à la sous-suite $(A_k)_{k\geq n}$, on obtient alors $P\Big(\bigcap_{k\geq n} \overline{A_k}\Big) = 0$. Par sous-additivité, une union dénombrable d'ensembles négligeables est négligeable, donc $P\bigg(\bigcup_{n\in\mathbb{N}} \Big(\bigcap_{k\geq n} \overline{A_k}\Big)\bigg) = 0$, puis le résultat par passage à l'événement contraire.

Variables aléatoires discrètes.

6. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On suppose qu'il existe $k \in]0,1[$ tel que

$$\textbf{(*)} : \qquad \forall n \in \mathbb{N} \qquad P(X = n) = k\,P(X \geq n)\,.$$

Déterminer la loi de X.

- -

On a, pour tout $n \in \mathbb{N}$, $P(X = n) = k\,\big(P(X = n) + P(X > n)\big)$, soit

$$(1-k)\,P(X=n)=k\,P(X>n)=k\,P(X\geq n+1)=P(X=n+1)$$

en réappliquant la relation **(*)** après décalage des indices. La suite $\big(P(X=n)\big)_{n\in\mathbb{N}}$ est donc géométrique de raison $1-k$, d'où, pour tout n entier naturel, $P(X=n)=(1-k)^n\,P(X=0)$. Mais, toujours grâce à **(*)**, $P(X=0)=k\,P(X\geq 0)=k\,P(\Omega)=k$. Finalement,

$$\forall n\in\mathbb{N}\qquad P(X=n)=k\,(1-k)^n\ .$$

On ne reconnaît pas exactement une loi usuelle, toutefois en posant $Y=X+1$, on note que $Y(\Omega)=\mathbb{N}^*$ et, pour tout $n\in\mathbb{N}^*$,

$$P(Y=n)=P(X=n-1)=k\,(1-k)^{n-1}\ .$$

Donc la variable $Y=X+1$ suit la loi géométrique $\mathcal{G}(k)$.

---

7. Deux variables aléatoires indépendantes X et Y suivent des lois géométriques $\mathcal{G}(p)$ et $\mathcal{G}(q)$ respectivement. Calculer $P(X<Y)$.

- -

On décompose $\{X<Y\}=\displaystyle\bigsqcup_{k=1}^{+\infty}\big(\{X=k\}\cap\{Y>k\}\big)$. Par σ-additivité d'une probabilité, puis indépendance des variables X et Y, on obtient

$$P(X<Y)=\sum_{k=1}^{+\infty}P(X=k)\,P(Y>k)=\sum_{k=1}^{+\infty}p(1-p)^{k-1}\,(1-q)^k=p\,(1-q)\sum_{k=0}^{+\infty}\big((1-p)(1-q)\big)^k\ .$$

Cela donne $P(X<Y)=\dfrac{p(1-q)}{1-(1-p)(1-q)}=\dfrac{p-pq}{p+q-pq}$.

---

8. Soient X et Y deux variables aléatoires indépendantes suivant des lois de Poisson de paramètres λ et μ respectivement. Soit n un entier naturel. Déterminer la loi conditionnelle de X sachant $\{X+Y=n\}$.

- -

Posons $Z=X+Y$. Alors $P(X=k|Z=n)=0$ si $k>n$ et, si $0\leq k\leq n$, alors

$$P(X=k|Z=n)=\frac{P(X=k\ ,\ Z=n)}{P(Z=n)}=\frac{P(X=k\ ,\ Y=n-k)}{P(Z=n)}=\frac{P(X=k)\,P(Y=n-k)}{P(Z=n)}\ ,$$

la dernière égalité résultant de l'indépendance de X et Y. Donc

$$P(X=k|Z=n)=e^{-\lambda}\frac{\lambda^k}{k!}e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}e^{\lambda+\mu}\frac{n!}{(\lambda+\mu)^n}=\binom{n}{k}\frac{\lambda^k\,\mu^{n-k}}{(\lambda+\mu)^n}=\binom{n}{k}\Big(\frac{\lambda}{\lambda+\mu}\Big)^k\Big(\frac{\mu}{\lambda+\mu}\Big)^{n-k}\ .$$

On observe donc que la loi conditionnelle de X sachant $\{Z=n\}$ est la loi binomiale de paramètres n et $\dfrac{\lambda}{\lambda+\mu}$.

---

9. À la gare de péage d'une autoroute, le nombre de voitures circulant dans le sens Nord-Sud suit une loi de Poisson de paramètre a, et le nombre de voitures circulant dans le sens Sud-Nord suit une loi de Poisson de paramètre b. Soient n et k deux entiers tels que $0 \leq k \leq n$. Sachant qu'à un instant donné, n voitures attendent au péage, quelle est la probabilité pour que k parmi elles circulent dans le sens Nord-Sud ?

- -

Notons X et Y les variables aléatoires représentant le nombre de voitures circulant respectivement dans le sens Nord-Sud et dans le sens Sud-Nord, et posons $Z = X + Y$. On cherche à déterminer $P(X = k|Z = n)$. Les variables X et Y étant indépendantes (*ce n'est pas dit dans l'énoncé, mais cela semble une hypothèse raisonnable ?... et puis, disons-nous qu'aucun calcul n'est possible si on ne fait pas cette hypothèse!*), on sait alors que Z suit une loi de Poisson $\mathcal{P}(a + b)$. Alors

$$P(X = k|Z = n) = \frac{P(X = k\ ,\ Z = n)}{P(Z = n)} = \frac{P(X = k\ ,\ Y = n - k)}{P(Z = n)} = \frac{P(X = k)\,P(Y = n - k)}{P(Z = n)}\ ,$$

la dernière égalité résultant de l'indépendance présumée de X et Y. Donc

$$P(X = k|Z = n) = e^{-a}\frac{a^k}{k!}e^{-b}\frac{b^{n-k}}{(n-k)!}e^{a+b}\frac{n!}{(a+b)^n} = \binom{n}{k}\frac{a^k\,b^{n-k}}{(a+b)^n} = \binom{n}{k}\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^{n-k}.$$

On observe donc que la loi conditionnelle de X sachant $\{Z = n\}$ est la loi binomiale de paramètres n et $\dfrac{a}{a+b}$.

---

10. Soient X et Y deux variables aléatoires réelles définies sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$, indépendantes, et toutes deux suivant une loi géométrique de même paramètre $p \in]0, 1[$. Pour $\omega \in \Omega$, on définit la matrice $M(\omega) = \begin{pmatrix} X(\omega) & Y(\omega) \\ Y(\omega) & X(\omega) \end{pmatrix}$. Déterminer la probabilité pour que cette matrice $M(\omega)$ soit inversible.

- -

Soit A l'événement: "la matrice M est inversible". Alors $A = \{X^2 - Y^2 \neq 0\}$, ce qui au passage montre que A est bien un événement puisque $X^2 - Y^2$ est une variable aléatoire sur $(\Omega, \mathcal{A})$. Il est plus simple de travailler d'abord sur l'événement contraire $\overline{A} = \{X^2 - Y^2 = 0\}$. En effet, les variables X et Y étant à valeurs dans $\mathbb{N}^*$, on a (*union disjointe*)

$$\overline{A} = \{X^2 = Y^2\} = \{X = Y\} = \bigsqcup_{n=1}^{+\infty}\big(\{X = n\} \cap \{Y = n\}\big)\ .$$

Par σ-additivité, et par indépendance des variables X et Y, on déduit

$$P(\overline{A}) = \sum_{n=1}^{+\infty} P(X = n)\,P(Y = n) = \sum_{n=1}^{+\infty} p^2(1-p)^{2(n-1)} = \sum_{n=0}^{+\infty} p^2(1-p)^{2n}.$$

On reconnaît une série géométrique de raison $(1-p)^2$, donc $P(\overline{A}) = \dfrac{p^2}{1-(1-p)^2} = \dfrac{p}{2-p}$, puis

$$P(A) = 1 - \frac{p}{2-p} = \frac{2(1-p)}{2-p} \ .$$

---

11. Soit r un entier supérieur ou égal à 2, soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On note

$$X_r = X\ (X-1)\ \cdots\ (X-r+1) = \prod_{k=0}^{r-1} (X-k) \ .$$

a. Calculer l'espérance de X_r lorsque X suit la loi de Poisson de paramètre $\lambda > 0$.

b. Calculer $\mathrm{E}(X_r)$ lorsque X suit la loi géométrique de paramètre $p \in]0,1[$.

- -

a. Par le théorème de transfert, il faut s'assurer de la convergence absolue de la série

$$\sum_{n\geq 0} n(n-1)\cdots(n-r+1)e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n\geq r} \frac{\lambda^n}{(n-r)!} \ .$$

La convergence n'est pas un souci (série exponentielle), et par un décalage d'indice,

$$\mathrm{E}(X_r) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^{k+r}}{k!} = e^{-\lambda}e^{\lambda}\lambda^r = \lambda^r \ .$$

b. Posons $q = 1-p$. On est amené ici à considérer la série $\displaystyle\sum_{n\geq r} n(n-1)\cdots(n-r+1)pq^{n-1}$, qui converge absolument (d'Alembert par exemple). Puis,

$$\mathrm{E}(X_r) = p\, q^{r-1} \sum_{n=r}^{+\infty} \frac{n!}{(n-r)!}\, q^{n-r} \ .$$

Mais, pour $x \in]-1,1[$, on a $\displaystyle\sum_{n=r}^{+\infty} \frac{n!}{(n-r)!} x^{n-r} = \frac{\mathrm{d}^r}{\mathrm{d}x^r}\left(\frac{1}{1-x}\right) = \frac{r!}{(1-x)^{r+1}}$ par une récurrence immédiate. Finalement,

$$\mathrm{E}(X_r) = p\, q^{r-1} \frac{r!}{(1-q)^{r+1}} = r!\, \frac{(1-p)^{r-1}}{p^r} \ .$$

---

12. Soit $N \in \mathbb{N}^*$. Soit $p \in]0,1[$. On pose $q = 1-p$. On considère N variables aléatoires $X_1, \cdots, X_N$ définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$, indépendantes et de même loi géométrique de paramètre p.

a. Soit $i \in [\![1, N]\!]$, soit $n \in \mathbb{N}^*$. Déterminer $P(X_i \leq n)$, puis $P(X_i > n)$.

b. On considère la variable aléatoire Y définie par $Y = \min\limits_{1\leq i\leq N} X_i$, c'est-à-dire

$$\forall \omega \in \Omega \qquad Y(\omega) = \min\left\{X_1(\omega), \cdots, X_N(\omega)\right\} \ .$$

Soit $n \in \mathbb{N}^*$, calculer $P(Y > n)$. En déduire $P(Y \leq n)$, puis $P(Y = n)$.

c. Reconnaître la loi de Y. En déduire $\mathrm{E}(Y)$.

- -

a. On a $P(X_i = k) = q\,p^{k-1}$ pour $k \in \mathbb{N}^*$, donc

$$P(X_i \leq n) = \sum_{k=1}^{n} P(X_i = k) = \sum_{k=1}^{n} pq^{k-1} = p\,\sum_{k=0}^{n-1} q^k = p\,\frac{1-q^n}{1-q} = 1 - q^n\ .$$

Par événement contraire, $P(X_i > n) = 1 - P(X_i \leq n) = q^n = (1-p)^n$.

b. Si $\omega \in \Omega$, on a $Y(\omega) > n \iff \min\big\{X_1(\omega), \cdots, X_N(\omega)\big\} > n \iff \forall i \in [\![1, N]\!]\ \ X_i(\omega) > n$, donc $\{Y > n\} = \displaystyle\bigcap_{i=1}^{N} \{X_i > n\}$ et, par indépendance des variables X_i,

$$P(Y > n) = \prod_{i=1}^{N} P(X_i > n) = (q^n)^N = q^{nN}\ .$$

Par événement contraire de nouveau, $P(Y \leq n) = 1 - P(Y > n) = 1 - q^{nN}$, puis

$$P(Y = n) = P(Y > n-1) - P(Y > n) = q^{(n-1)N} - q^{nN} = q^{(n-1)N}(1 - q^N)\ .$$

c. On a $Y(\Omega) = \mathbb{N}^*$ et, pour tout $n \in \mathbb{N}^*$, $P(Y = n) = (1 - q^N)\,(q^N)^{n-1}$, on reconnaît une loi géométrique de paramètre $1 - q^N$. Donc, d'après le cours, $\mathrm{E}(Y) = \dfrac{1}{1-q^N}$.

---

13. Chez un marchand de journaux, on peut acheter des pochettes contenant chacune une image. La collection complète comporte N images distinctes, dont l'obtention à chaque achat est équiprobable. On note X_k le nombre d'achats nécessaires pour avoir obtenu au moins k images différentes. Ainsi, $X_1 = 1$ et X_N est le nombre d'achats nécessaires à l'obtention de la collection complète.

a. Quelle est la loi de la variable $X_{k+1} - X_k$, avec $k \in [\![1, N-1]\!]$?

b. En déduire l'espérance de X_N.

- -

a. Si le collectionneur a déjà k images distinctes en sa possession, chaque achat d'une nouvelle pochette est une épreuve de Bernoulli où la probabilité de succès est $p_k = \dfrac{N-k}{N}$. La variable $X_{k+1} - X_k$ représente le temps d'attente du premier succès, elle suit donc une loi géométrique de paramètre p_k. En particulier, $\mathrm{E}(X_{k+1} - X_k) = \dfrac{1}{p_k} = \dfrac{N}{N-k}$.

b. Par linéarité de l'espérance, de la relation $X_N = X_1 + \displaystyle\sum_{k=1}^{N-1} (X_{k+1} - X_k)$, on tire

$$\mathrm{E}(X_N) = 1 + \sum_{k=1}^{N-1} \frac{N}{N-k} = 1 + N\,\sum_{j=1}^{N-1} \frac{1}{j} = N\,H_N$$

en posant, comme d'habitude, $H_N = \displaystyle\sum_{k=1}^{N} \frac{1}{k}$. On en déduit que $\mathrm{E}(X_N) \underset{N \to +\infty}{\sim} N\,\ln(N)$.

---

14. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires deux à deux indépendantes, suivant toutes la même loi de Poisson de paramètre a. Soit Y_n la variable aléatoire telle que $Y_n(\omega) = 1$ si $X_n(\omega) = 0$, et $Y_n(\omega) = 0$ sinon. Soit enfin $Z_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} Y_k$.

a. Montrer que $\mathrm{V}(Z_n) \leq \dfrac{1}{4n}$.

b. À partir de quelle valeur de n peut-on affirmer qu'il y a 90% de chance au moins pour que $|Z_n - e^{-a}|$ soit inférieur à $\dfrac{1}{10}$?

- -

a. Pour tout n, la variable Y_n suit une loi de Bernoulli de paramètre $p = P(X_n = 0) = e^{-a}$. On a alors $\mathrm{E}(Y_n) = p = e^{-a}$ et $\mathrm{V}(Y_n) = p(1-p) = e^{-a}\big(1-e^{-a}\big)$. Pour tout n, la variable Y_n est une fonction de X_n, précisément $Y_n = f(X_n)$, avec $f : \mathbb{N} \to \{0, 1\}$ fonction indicatrice de $\{0\}$. Les X_n étant supposées deux à deux indépendantes, les Y_n le sont alors aussi, elles sont donc décorrélées. On a alors

$$\mathrm{V}(Z_n) = \frac{1}{n^2}\,\mathrm{V}\Big(\sum_{k=1}^{n} Y_k\Big) = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{V}(Y_k) = \frac{1}{n}e^{-a}(1-e^{-a})\ .$$

Comme $\forall p \in [0,1] \quad p(1-p) \leq \dfrac{1}{4}$ (*petite étude de fonction*), on déduit $\mathrm{V}(Z_n) \leq \dfrac{1}{4n}$.

b. Par linéarité, on a $\mathrm{E}(Z_n) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}\mathrm{E}(Y_k) = e^{-a}$, il suffit donc d'appliquer l'inégalité de Bienaymé-Tchebychev. Il est clair que $Z_n \in L^2(\Omega, \mathbb{R})$, i.e. Z_n^2 est d'espérance finie, puisque Z_n prend ses valeurs dans l'ensemble fini $\Big\{\dfrac{k}{n}\ ;\ 0 \leq k \leq n\Big\}$. Avec $\varepsilon = \dfrac{1}{10}$, on a

$$P\Big(|Z_n - e^{-a}| \geq \varepsilon\Big) \leq \frac{\mathrm{V}(Z_n)}{\varepsilon^2} = 100\,\mathrm{V}(Z_n) \leq \frac{25}{n}$$

en utilisant l'inégalité obtenue en **a.** Pour avoir $P\Big(|Z_n - e^{-a}| < \dfrac{1}{10}\Big) > \dfrac{9}{10}$, soit par événement contraire $P\Big(|Z_n - e^{-a}| \geq \dfrac{1}{10}\Big) \leq \dfrac{1}{10}$, il suffit donc que l'on ait $\dfrac{25}{n} \leq \dfrac{1}{10}$, ce qui est vrai dès que $n \geq 250$.

---

15. On modélise une partie de pile ou face infinie par l'univers $\Omega = \{0,1\}^{\mathbb{N}}$ des suites infinies de 0 et de 1 (par exemple, "Pile" est codé par 0, "Face" est codé par 1). On suppose qu'à chaque lancer, l'apparition de "Face"=1 a une probabilité $p \in]0,1[$, et que les lancers sont indépendants. On admet l'existence d'une tribu sur Ω et d'une probabilité P telle que, si on note $\omega = (\omega_n)_{n\in\mathbb{N}}$ un élément de Ω, on ait $P(\omega_n = 1) = p$ pour tout n. Les lancers étant supposés indépendants, si on se donne des entiers naturels **distincts** $n_1, \cdots, n_k, m_1, \cdots, m_l$, on a alors

$$P\big(\omega_{n_1} = 1, \cdots, \omega_{n_k} = 1, \omega_{m_1} = 0, \cdots, \omega_{m_l} = 0\big) = p^k(1-p)^l\ .$$

Toute suite $\omega \in \Omega$ se compose d'une succession de blocs formés de 0 et de blocs formés de 1. On note $X(\omega)$ la longueur du premier bloc de la suite ω, et $Y(\omega)$ la longueur du deuxième bloc. Ainsi, pour $\omega = (1,1,1,0,0,1,1,0,1,...)$, on a $X(\omega) = 3$ et $Y(\omega) = 2$.

a. Déterminer les lois des variables aléatoires X et Y. Ont-elles la même loi ?

b. Déterminer la loi conjointe du couple $U = (X, Y)$. Dans quel cas les variables X et Y sont-elles indépendantes ?

c. Calculer les espérances des variables X et Y.

d. Étudier les variations de $\mathrm{E}(X)$ en fonction de p. Comment choisir p pour avoir $\mathrm{E}(X) = \dfrac{5}{2}$?

- -

Pour tout $n \in \mathbb{N}^*$, on introduit les événements suivants:

S_n: "on obtient Face (=1, =succès) au n-ème lancer" ;

A_n: "le premier bloc est constitué de n succès" ;

B_n: "le premier bloc est constitué de n échecs".

a. • Pour $n \in \mathbb{N}^*$, on a $\{X = n\} = A_n \sqcup B_n$ (réunion disjointe).

Donc $P(X = n) = P(A_n) + P(B_n)$.

Or, $A_n = S_1 \cap S_2 \cap \cdots \cap S_n \cap \overline{S_{n+1}}$, donc $P(A_n) = p^n(1-p)$ par indépendance.

Par symétrie, $P(B_n) = (1-p)^n p$, donc $\forall n \in \mathbb{N}^* \quad P(X = n) = (1-p)p^n + p(1-p)^n$.

• La famille constituée des événements A_k, $k \geq 1$, et des B_k, $k \geq 1$ est un système quasi-complet d'événements puisque la réunion de tous ces ensembles donne l'univers Ω privé des deux seules suites constituées uniquement de 0 ou uniquement de 1, qui sont de probabilité nulle. *On peut aussi vérifier par le calcul que* $\displaystyle\sum_{n=1}^{+\infty} P(X = n) = 1$*, ce qui montre que le premier bloc est presque sûrement de longueur finie.* On applique donc la formule des probabilités totales:

$$P(Y = n) = \sum_{k=1}^{+\infty} \Big(P(Y = n | A_k)\, P(A_k) + P(Y = n | B_k)\, P(B_k) \Big) .$$

Or, si l'événement A_k est réalisé (i.e. les k premières épreuves sont des succès et la $k+1$-ème est un échec), alors $Y = n$ si et seulement si les épreuves $k+2$, $k+3$, $\cdots$, $k+n$ conduisent à des échecs alors que l'épreuve $k+n+1$ est un succès. On en déduit que $P(Y = n | A_k) = (1-p)^{n-1} p$. De même, $P(Y = n | B_k) = p^{n-1}(1-p)$. Donc

$$\begin{aligned} P(Y = n) &= \sum_{k=1}^{+\infty} (1-p)^{n-1} p\, p^k (1-p) + \sum_{n=1}^{+\infty} p^{n-1}(1-p)\,(1-p)^k p \\ &= p^2(1-p)^n \sum_{k=1}^{+\infty} p^{k-1} + (1-p)^2 p^n \sum_{k=1}^{+\infty} (1-p)^{k-1} \\ &= p^2(1-p)^{n-1} + (1-p)^2 p^{n-1} . \end{aligned}$$

• Si $p = \dfrac{1}{2}$, alors $P(X = n) = P(Y = n) = \dfrac{1}{2^n}$, donc X et Y ont la même loi. Réciproquement, si X et Y ont la même loi, alors en particulier $P(X = 1) = P(Y = 1)$, soit $2p(1-p) = 2(1-p)^2$, d'où $p = 1 - p$, puis $p = \dfrac{1}{2}$, la condition est nécessaire et suffisante.

b. • Soient k et l deux entiers naturels non nuls, alors l'événement $\{U = (k,l)\} = \{X = k\} \cap \{Y = l\}$ est la réunion disjointe des événements $S_1 \cap \cdots \cap S_k \cap \overline{S_{k+1}} \cap \cdots \cap \overline{S_{k+l}} \cap S_{k+l+1}$ (les k premières épreuves sont des succès, les l suivantes sont des échecs, la $k+l+1$-ème est un succès) et $\overline{S_1} \cap \cdots \cap \overline{S_k} \cap S_{k+1} \cap \cdots \cap S_{k+l} \cap \overline{S_{k+l+1}}$ (le contraire). Donc

$$\forall (k,l) \in (\mathbb{N}^*)^2 \qquad P(X = k\ ,\ Y = l) = p^{k+1}(1-p)^l + (1-p)^{k+1}p^l\ .$$

• Si $p = \dfrac{1}{2}$, alors

$$P(X = k\ ,\ Y = l) = \frac{1}{2^{k+l}} = \frac{1}{2^k}\ \frac{1}{2^l} = P(X = k)\ P(Y = l)\ ,$$

donc les variables X et Y sont indépendantes. Réciproquement, si X et Y sont indépendantes, alors on a en particulier

$$P(X = 1\ ,\ Y = 1) = P(X = 1)\ P(Y = 1)\ ,$$

soit

$$p^2(1-p) + (1-p)^2 p = 2p(1-p)\ 2(1-p)^2\ ,$$

ou encore $1 = 4(1-p)^2$, donc $p = \dfrac{1}{2}$, la condition est nécessaire et suffisante.

c. • Calculons: en utilisant, pour $|x| < 1$, la relation $\displaystyle\sum_{n=1}^{+\infty} nx^{n-1} = \frac{\mathrm{d}}{\mathrm{d}x}\Big(\frac{1}{1-x}\Big) = \frac{1}{(1-x)^2}$, on a

$$\begin{aligned} \mathrm{E}(X) &= (1-p)p \sum_{n=1}^{+\infty} np^{n-1} + p(1-p) \sum_{n=1}^{+\infty} n(1-p)^{n-1} \\ &= \frac{(1-p)p}{(1-p)^2} + \frac{p(1-p)}{p^2} \ = \ \frac{p}{1-p} + \frac{1-p}{p}\ . \end{aligned}$$

Le résultat dépend de p et intuitivement on comprend bien que, si le jeu est très déséquilibré, i.e. si p est très proche de 0 (resp. très proche de 1), alors il y a de fortes chances pour que la suite d'épreuves commence par un très grand nombre (un "bloc") d'échecs (resp. de succès). Ceci est confirmé par le fait que $\mathrm{E}(X)$ *tend vers* $+\infty$ *lorsque p tend vers* 0^+ *ou vers* 1^-.

• De façon semblable,

$$\begin{aligned} \mathrm{E}(Y) &= (1-p)^2 \sum_{n=1}^{+\infty} np^{n-1} + p^2 \sum_{n=1}^{+\infty} n(1-p)^{n-1} \\ &= \frac{(1-p)^2}{(1-p)^2} + \frac{p^2}{p^2} \ = \ 2\ . \end{aligned}$$

Le résultat ne dépend pas de p, cela peut sembler surprenant, je ne sais pas l'expliquer intuitivement.

d. Posons $f(p) = \mathrm{E}(X) = \dfrac{p}{1-p} + \dfrac{1-p}{p} = \dfrac{1}{p(1-p)} - 2$, alors $f'(p) = \dfrac{2p-1}{p^2(1-p)^2}$, donc f est décroissante sur $\Big]0, \dfrac{1}{2}\Big]$, puis croissante sur $\Big[\dfrac{1}{2}, 1\Big[$. On a $\displaystyle\lim_{p\to 0^+} f(p) = \lim_{p \to 1^-} f(p) = +\infty$.

L'espérance de X est minimale pour $p = \dfrac{1}{2}$ et vaut alors 2. On a évidemment la symétrie $f(1-p) = f(p)$, qui se traduit par la symétrie de la courbe par rapport à la droite verticale d'équation $x = \dfrac{1}{2}$. On obtient facilement

$$\mathrm{E}(X) = \frac{5}{2} \iff p = \frac{1}{3} \text{ ou } p = \frac{2}{3} \; .$$

---

16. Un péage autoroutier comporte deux barrières. Le nombre de voitures arrivant à ce péage par jour suit une loi de Poisson de paramètre λ, et chaque voiture choisit arbitrairement de franchir l'une ou l'autre des deux barrières, ces choix étant indépendants. On note X_1 et X_2 les variables aléatoires représentant le nombre de voitures franchissant chacune des deux barrières dans une journée.

a. Déterminer la loi de X_1.

b. En exploitant $X_1 + X_2$, calculer la covariance de X_1 et X_2.

c. Montrer que les variables X_1 et X_2 sont indépendantes.

- -

a. Posons $X = X_1 + X_2$ (nombre de voitures arrivant au péage). Alors, si k et n sont deux entiers tels que $k \geq n$, on a

$$P(X_1 = n | X = k) = \binom{k}{n} \left(\frac{1}{2}\right)^k \; :$$

en effet, la loi conditionnelle de X_1 sachant $\{X = k\}$ est binomiale de paramètres k et $\dfrac{1}{2}$. Si l'on fixe $n \in \mathbb{N}$, on a alors

$$\begin{aligned}
P(X_1 = n) &= \sum_{k=n}^{+\infty} P(X_1 = n | X = k)\, P(X = k) \\
&= \sum_{k=n}^{+\infty} \binom{k}{n} \frac{1}{2^k}\, e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=n}^{+\infty} \frac{\left(\frac{\lambda}{2}\right)^n \left(\frac{\lambda}{2}\right)^{k-n}}{n!\,(k-n)!} \\
&= \frac{e^{-\lambda} \left(\frac{\lambda}{2}\right)^n}{n!} \sum_{p=0}^{+\infty} \frac{\left(\frac{\lambda}{2}\right)^p}{p!} \\
&= \frac{e^{-\lambda} \left(\frac{\lambda}{2}\right)^n}{n!}\, e^{\frac{\lambda}{2}} = e^{-\frac{\lambda}{2}} \frac{\left(\frac{\lambda}{2}\right)^n}{n!} \; .
\end{aligned}$$

On conclut que X_1 (et de même X_2 évidemment) suit une loi de Poisson de paramètre $\dfrac{\lambda}{2}$.

b. La covariance étant une forme bilinéaire symétrique sur $\mathrm{L}^2(\Omega, \mathbb{R})$, on a

$$\mathrm{V}(X) = \mathrm{V}(X_1 + X_2) = \mathrm{Cov}(X_1 + X_2, X_1 + X_2) = \mathrm{V}(X_1) + \mathrm{V}(X_2) + 2\,\mathrm{Cov}(X_1, X_2) \; .$$

Comme on sait que la variance d'une variable de Poisson de paramètre λ vaut λ, on a donc $\lambda = \dfrac{\lambda}{2} + \dfrac{\lambda}{2} + 2\,\mathrm{Cov}(X_1, X_2)$, d'où $\mathrm{Cov}(X_1, X_2) = 0$: les variables X_1 et X_2 sont décorrélées (*ce qui n'entraîne pas a priori qu'elles sont indépendantes*).

c. Soient k et l deux entiers naturels, on calcule:

$$\begin{aligned}
P(X_1 = k\ ,\ X_2 = l) &= P(X_1 = k\ ,\ X = k + l) \ = \ P(X_1 = k | X = k + l)\, P(X = k + l) \\
&= \binom{k+l}{k} \frac{1}{2^{k+l}}\, e^{-\lambda} \frac{\lambda^{k+l}}{(k+l)!} \\
&= \frac{e^{-\lambda}\, \lambda^{k+l}}{k!\, l!\, 2^{k+l}} \ = \ e^{-\frac{\lambda}{2}} \frac{\left(\frac{\lambda}{2}\right)^k}{k!}\ e^{-\frac{\lambda}{2}} \frac{\left(\frac{\lambda}{2}\right)^l}{l!} \\
&= P(X_1 = k)\, P(X_2 = l)\ .
\end{aligned}$$

Les variables X_1 et X_2 sont donc indépendantes.

---

17. Un joueur dispose de N dés équilibrés. Il lance une première fois ceux-ci et on note X_1 le nombre de “six” obtenus. Il met de côté les dés correspondants et relance les autres dés (s'il en reste). On note X_2 le nombre de “six” obtenus et on répète l'expérience définissant ainsi une suite de variables aléatoires $X_1,\ X_2,\ \cdots$ La variable $S_n = X_1 + X_2 + \cdots + X_n$ correspond alors au nombre de “six” obtenus après n lancers.

a. Vérifier que S_n suit une loi binomiale dont on précisera les paramètres.

b. Montrer qu'il est presque sûr qu'il existe un rang n pour lequel $S_n = N$.

c. On définit alors la variable aléatoire

$$T = \min\Big(\{n \geq 1 | S_n = N\}\ \cup\ \{+\infty\}\Big)\ .$$

Déterminer la loi de T.

d. Vérifier que la variable T est d'espérance finie, et donner une formule exprimant celle-ci. Calculer cette espérance pour $N = 1$ et $N = 2$.

- -

a. Il est clair que chaque variable S_n prend ses valeurs dans $[\![0, N]\!]$, on va donc montrer par récurrence que S_n suit une loi binomiale de paramètres N et p_n, avec $p_n \in]0, 1[$ dépendant de n, et que l'on va chercher à déterminer.

• Pour $n = 1$, la variable $S_1 = X_1$ suit clairement la loi binomiale $\mathcal{B}\Big(N, \dfrac{1}{6}\Big)$, ce qui initialise avec $p_1 = \dfrac{1}{6}$.

• Soit $n \in \mathbb{N}^*$ et $k \in [\![0, N]\!]$, on a alors $\{S_{n+1} = k\} = \displaystyle\bigsqcup_{j=0}^{k} \big(\{S_n = j\}\ \cap\ \{X_{n+1} = k - j\}\big)$, donc

$$P(S_{n+1} = k)\ =\ \sum_{j=0}^{k} P(S_n = j)\, P(X_{n+1} = k - j \mid S_n = j)\ .$$

• Si l'événement $\{S_n = j\}$ est réalisé, alors, au $(n+1)$-ème lancer, on lance $N-j$ dés. La loi conditionnelle de X_{n+1} sachant $\{S_n = j\}$ est donc binomiale de paramètres $N-j$ et $\dfrac{1}{6}$, soit $P(X_{n+1} = k-j \mid S_n = j) = \dbinom{N-j}{k-j} \left(\dfrac{1}{6}\right)^{k-j} \left(\dfrac{5}{6}\right)^{N-k}$. On fait l'hypothèse (de récurrence) que S_n suit la loi $\mathcal{B}(N, p_n)$, alors $P(S_n = j) = \dbinom{N}{j} p_n^j (1-p_n)^{N-j}$. Donc

$$P(S_{n+1} = k) = \sum_{j=0}^{k} \binom{N}{j} p_n^j (1-p_n)^{N-j} \binom{N-j}{k-j} \left(\frac{1}{6}\right)^{k-j} \left(\frac{5}{6}\right)^{N-k}.$$

Par ailleurs, $\dbinom{N}{j} \dbinom{N-j}{k-j} = \dfrac{N!}{j!(N-j)!} \dfrac{(N-j)!}{(k-j)!(N-k)!} \times \dfrac{k!}{k!} = \dbinom{N}{k} \dbinom{k}{j}$. Donc

$$\begin{aligned} P(S_{n+1} = k) &= \binom{N}{k} \left(\frac{5}{6}\right)^{N-k} (1-p_n)^{N-k} \left[\sum_{j=0}^{k} \binom{k}{j} p_n^j \left(\frac{1-p_n}{6}\right)^{k-j}\right] \\ &= \binom{N}{k} \left(\frac{5-5p_n}{6}\right)^{N-k} \left(p_n + \frac{1-p_n}{6}\right)^{k} \\ &= \binom{N}{k} p_{n+1}^k (1-p_{n+1})^{N-k}, \end{aligned}$$

en posant $p_{n+1} = p_n + \dfrac{1-p_n}{6} = \dfrac{1+5p_n}{6}$. Ainsi, S_{n+1} suit la loi $\mathcal{B}(N, p_{n+1})$, ce qui achève la récurrence.

• De plus, la suite (p_n) est arithmético-géométrique, l'équation $l = \dfrac{1+5l}{6}$ admet pour racine $l = 1$, donc la suite (q_n) définie par $q_n = 1 - p_n$, est géométrique de raison $\dfrac{5}{6}$ avec $q_1 = 1 - p_1 = \dfrac{5}{6}$, ce qui donne $q_n = \left(\dfrac{5}{6}\right)^n$, puis $p_n = 1 - \left(\dfrac{5}{6}\right)^n$.

Remarque. Pour identifier la loi de S_n, on peut aussi dire que l'on lance n fois chacun des N dés, et que l'on appelle "succès" (pour chaque dé) le fait qu'il soit tombé au moins une fois sur la face 6 au cours des n lancers. On retrouve ainsi le fait que S_n suit une loi binomiale dont les paramètres sont N et la probabilité de "succès" pour chaque dé, soit $1 - \left(\dfrac{5}{6}\right)^n$.

b. Pour $n \in \mathbb{N}^*$ donné, on a alors $P(S_n = N) = \dbinom{N}{N} p_n^N (1-p_n)^0 = p_n^N = \left(1 - \left(\dfrac{5}{6}\right)^n\right)^N$, donc $\lim\limits_{n \to +\infty} P(S_n = N) = 1$.

Soit E l'événement: "il existe $n \in \mathbb{N}^*$ pour lequel $S_n = N$". Alors E est la réunion (croissante) des événements $\{S_n = N\}$ pour $n \in \mathbb{N}^*$. En particulier, E contient chaque événement $\{S_n = N\}$, i.e. $\forall n \in \mathbb{N}^* \quad \{S_n = N\} \subset E$. Par croissance d'une probabilité, on déduit que $\forall n \in \mathbb{N}^* \quad P(S_n = N) \leq P(E) \leq 1$. Comme $\lim\limits_{n \to +\infty} P(S_n = N) = 1$, on a $P(E) = 1$.

c. Notons que T est une variable aléatoire à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$, mais on a prouvé en **b.** que $P(\overline{E}) = P(T = +\infty) = 0$, on peut donc considérer que T est à valeurs dans $\mathbb{N}$.

Notons, pour tout $n \in \mathbb{N}^*$, l'identité $\{S_n = N\} = \{T \leq n\}$, donc $P(T \leq n) = P(S_n = N)$ puis, pour tout $n \in \mathbb{N}^*$,

$$P(T = n) = P(T \leq n) - P(T \leq n-1) = \left(1 - \left(\frac{5}{6}\right)^n\right)^N - \left(1 - \left(\frac{5}{6}\right)^{n-1}\right)^N .$$

Remarque. Il peut être intéressant aussi de remarquer qu'une expérience aléatoire équivalente consiste à lancer séparément chacun des N dés jusqu'à obtention de la face 6, et à noter le maximum des N temps d'attente. Ainsi, si l'on note $T_1, \cdots, T_N$ les temps d'attente pour chaque dé, les variables T_k sont mutuellement indépendantes, chacune suit la loi géométrique de paramètre $\frac{1}{6}$, et $T = \max\limits_{1 \leq k \leq N} T_k$.

d. Comme T est à valeurs dans $\mathbb{N}$, utilisons la formule

$$\mathrm{E}(T) = \sum_{n=1}^{+\infty} P(T \geq n) = \sum_{n=0}^{+\infty} P(T > n) .$$

Or,

$$P(T > n) = 1 - P(T \leq n) = 1 - P(S_n = N) = 1 - \left(1 - \left(\frac{5}{6}\right)^n\right)^N = 1 - \exp\left[N \ln\left(1 - \left(\frac{5}{6}\right)^n\right)\right] .$$

Comme $\lim\limits_{n \to +\infty} \left(\frac{5}{6}\right)^n = 0$, on a $\ln\left(1 - \left(\frac{5}{6}\right)^n\right) \underset{n \to +\infty}{\sim} -\left(\frac{5}{6}\right)^n$ qui tend aussi vers 0, puis $1 - \exp\left[N \ln\left(1 - \left(\frac{5}{6}\right)^n\right)\right] \underset{n \to +\infty}{\sim} N \left(\frac{5}{6}\right)^n$, ce qui justifie la convergence de la série de terme général $P(T \geq n)$, donc la variable aléatoire T est d'espérance finie.

- Pour $N = 1$, on a simplement $P(T > n) = \left(\frac{5}{6}\right)^n$, donc $\mathrm{E}(T) = \sum\limits_{n=0}^{+\infty} \left(\frac{5}{6}\right)^n = 6$, en fait dans ce cas c'est évident puisque T suit simplement la loi géométrique de paramètre $\frac{1}{6}$.
- Pour $N = 2$, on a $P(T > n) = 2 \times \left(\frac{5}{6}\right)^n - \left(\frac{5}{6}\right)^{2n}$, puis $\mathrm{E}(T) = \frac{96}{11} \simeq 8,7$.

---

18. Soit T une variable aléatoire à valeurs dans $\mathbb{N}$. On suppose que $\forall n \in \mathbb{N} \quad P(T > n) > 0$. On appelle **taux de panne** associé à T la suite $(\theta_n)_{n \in \mathbb{N}}$ définie par

$$\theta_n = P(T = n \mid T \geq n) .$$

Typiquement, si T est la variable aléatoire indiquant l'instant où un matériel tombe en panne, la quantité θ_n indique la probabilité qu'il tombe en panne à l'instant n sachant qu'il était encore fonctionnel jusque là.

a. Montrer que $\theta_n \in [0, 1[$ pour tout n.

b. Exprimer $P(T \geq n)$ à l'aide des θ_k. En déduire que la série $\sum \theta_k$ diverge.

c*. Inversement, soit (θ_n) une suite d'éléments de $[0, 1[$ telle que la série $\sum \theta_n$ diverge. Montrer que la suite $(\theta_n)_{n \in \mathbb{N}}$ est un taux de panne associé à une certaine variable aléatoire T.

- -

a. Le réel θ_n est une probabilité (conditionnelle) donc appartient à $[0,1]$. S'il valait 1, on aurait $P(T=n)=P(T\geq n)$, donc par différence $P(T>n)=0$, ce qui est exclu.

b. On a, pour tout n entier naturel,

$$\theta_n=\frac{P\big(\{T=n\}\,\cap\,\{T\geq n\}\big)}{P(T\geq n)}=\frac{P(T=n)}{P(T\geq n)}=\frac{P(T\geq n)-P(T\geq n+1)}{P(T\geq n)}=1-\frac{P(T\geq n+1)}{P(T\geq n)}\ .$$

Donc $\dfrac{P(T\geq n+1)}{P(T\geq n)}=1-\theta_n$ pour tout n. Partant de $P(T\geq 0)=1$, on a immédiatement

$$\forall n\in\mathbb{N}\qquad P(T\geq n)=\prod_{k=0}^{n-1}(1-\theta_k)\ .$$

Mais $P(T\geq n)=\displaystyle\sum_{k=n}^{+\infty}P(T=k)\underset{n\to+\infty}{\longrightarrow}0$ car c'est le reste d'ordre $n-1$ d'une série convergente. Donc $\ln\big(P(T\geq n)\big)=\displaystyle\sum_{k=0}^{n-1}\ln(1-\theta_k)\underset{n\to+\infty}{\longrightarrow}-\infty$. La série à termes négatifs $\sum\ln(1-\theta_k)$ diverge donc.

Deux cas se présentent alors:

- si θ_k ne tend pas vers 0, alors évidemment la série $\sum\theta_k$ diverge (grossièrement) ;
- si θ_k tend vers 0, alors $\theta_k\sim-\ln(1-\theta_k)$ et, par le critère des équivalents (les séries considérées sont à termes positifs), la série $\sum\theta_k$ est encore divergente.

c. **Analyse**: si T est une variable aléatoire solution, alors

$$P(T=n)=P(T\geq n)-P(T\geq n+1)=\prod_{k=0}^{n-1}(1-\theta_k)-\prod_{k=0}^{n}(1-\theta_k)=\theta_n\prod_{k=0}^{n-1}(1-\theta_k)\ .$$

La loi de T est donc déterminée.

Synthèse: Posons $p_n=\theta_n\displaystyle\prod_{k=0}^{n-1}(1-\theta_k)$ pour tout n entier naturel. Alors les p_n sont des réels positifs et, pour $n\in\mathbb{N}$, on a $p_n=q_n-q_{n+1}$ en posant $q_n=\displaystyle\prod_{k=0}^{n-1}(1-\theta_k)$. Comme la série $\sum\theta_k$ diverge, il en est de même de la série à termes négatifs $\sum\ln(1-\theta_k)$ par un raisonnement analogue à celui de la fin de la question **b.**, donc ses sommes partielles tendent vers $-\infty$, soit $\ln(q_n)\underset{n\to+\infty}{\longrightarrow}-\infty$, puis $\displaystyle\lim_{n\to+\infty}q_n=0$. La série télescopique $\sum p_n=\sum(q_n-q_{n+1})$ est alors convergente et $\displaystyle\sum_{n=0}^{+\infty}p_n=\sum_{n=0}^{+\infty}(q_n-q_{n+1})=q_0-\lim_{n\to+\infty}q_n=q_0=1$. La suite $(p_n)_{n\in\mathbb{N}}$ est donc une distribution de probabilités sur l'ensemble $\mathbb{N}$, il existe alors une variable aléatoire T (sur un certain espace probabilisé), à valeurs dans $\mathbb{N}$, et telle que $P(T=n)=p_n$ pour tout n.

Enfin, $P(T \geq n) = \sum_{k=n}^{+\infty} p_k = \sum_{k=n}^{+\infty} (q_k - q_{k+1}) = q_n$, puis

$$P(T = n | T \geq n) = \frac{P(T = n)}{P(T \geq n)} = \frac{p_n}{q_n} = \theta_n \ .$$

La suite (θ_n) est donc le taux de panne associé à la variable aléatoire T.

---

19. Soient X et Y deux variables aléatoires sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $\mathbb{N}$. On suppose que $Y \leq X$ et que $\mathrm{E}(X) < +\infty$. On suppose de plus que, pour tout $x \in \mathbb{N}$, la loi de Y conditionnellement à l'événement $\{X = x\}$ est la loi uniforme sur $[\![0, x]\!]$.

a. Donner une relation simple entre $\mathrm{E}(X)$ et $\mathrm{E}(Y)$.

b. Montrer que

$$\forall n \in \mathbb{N} \qquad P(X = n) = (n+1)\big(P(Y = n) - P(Y = n+1)\big) \ .$$

c. Montrer que les variables $X - Y$ et Y ont la même loi.

- -

a. Déjà, par comparaison, X est d'espérance finie, puis en calculant dans le monde des bisounours (c'est-à-dire dans $[0, +\infty]$ où toutes les interversions sont permises):

$$\begin{aligned}
\mathrm{E}(Y) &= \sum_{k=0}^{+\infty} k\, P(Y = k) \\
&= \sum_{k=0}^{+\infty} k \left(\sum_{n=k}^{+\infty} P(Y = k \mid X = n)\, P(X = n) \right) \\
&= \sum_{k=0}^{+\infty} k \left(\sum_{n=k}^{+\infty} \frac{P(X = n)}{n+1} \right) = \sum_{n=0}^{+\infty} \frac{P(X = n)}{n+1} \left(\sum_{k=0}^{n} k \right) \\
&= \sum_{n=0}^{+\infty} \frac{P(X = n)}{n+1} \times \frac{n(n+1)}{2} = \frac{1}{2}\, \mathrm{E}(X) \ .
\end{aligned}$$

b. On a d'abord, pour tout n entier naturel,

$$P(Y = n) = \sum_{k=n}^{+\infty} P(Y = n \ , \ X = k) = \sum_{k=n}^{+\infty} P(Y = n \mid X = k)\, P(X = k) = \sum_{k=n}^{+\infty} \frac{P(X = k)}{k+1} \ .$$

On en déduit que $P(Y = n) - P(Y = n+1) = \dfrac{P(X = n)}{n+1}$, puis la relation demandée.

c. Pour tout n entier naturel, on a

$$\begin{aligned}
P(X - Y = n) &= \sum_{l=0}^{+\infty} P(X - Y = n \ , \ Y = l) = \sum_{l=0}^{+\infty} P(X = n + l \ , \ Y = l) \\
&= \sum_{l=0}^{+\infty} P(Y = l \mid X = n + l)\, P(X = n + l)
\end{aligned}$$

$$= \sum_{l=0}^{+\infty} \frac{P(X=n+l)}{n+l+1} = \sum_{k=n}^{+\infty} \frac{P(X=k)}{k+1} = P(Y=n)$$

par une translation d'indice, et en comparant avec la première ligne de la question **b.**

---

20. Soient $X_1, \cdots, X_n$ des variables aléatoires réelles discrètes admettant un moment d'ordre deux, i.e. telles que $X_1^2, \cdots, X_n^2$ soient d'espérance finie. On appelle **matrice des covariances** de la famille $(X_1, \cdots, X_n)$ la matrice carrée d'ordre n:

$$S = \big(\mathrm{Cov}(X_i, X_j)\big)_{1 \le i,j \le n} .$$

Soient $a_1, \cdots, a_n$ des réels. Exprimer la variance de $X = a_1 X_1 + \cdots + a_n X_n$ à l'aide de la matrice S. En déduire que la matrice symétrique S est positive.

- -

La covariance étant une forme bilinéaire symétrique sur l'espace vectoriel $L^2(\Omega, \mathbb{R})$, on a

$$\mathrm{V}(X) = \mathrm{Cov}\Big(\sum_{i=1}^n a_i X_i, \sum_{j=1}^n a_j X_j\Big) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \ \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \ s_{i,j} .$$

Soit alors la matrice-colonne $A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$. On reconnaît que $\mathrm{V}(X) = A^\top \, S \, A$.

On a donc $A^\top S A = \mathrm{V}\Big(\sum_{i=1}^n a_i X_i\Big) \ge 0$ pour tout $A \in \mathcal{M}_{n,1}(\mathbb{R})$, d'où $S \in \mathcal{S}_n^+(\mathbb{R})$.

Fonctions génératrices.

21. On dispose d'un premier dé normal (non pipé, faces numérotées de 1 à 6), et d'un deuxième dé (non pipé) dont les faces ne portent aucune inscription. On note X la variable aléatoire correspondant au résultat du lancer du dé normal. En utilisant des fonctions génératrices, déterminer de quelle façon il est possible de numéroter les faces du deuxième dé afin que, lors du lancer simultané de ces deux dés, la somme des deux nombres obtenus prenne toutes les valeurs de 1 à 12 avec équiprobabilité.

- -

La variable X suit la loi uniforme sur l'intervalle $[\![1, 6]\!]$, sa fonction génératrice est donc le polynôme

$$G_X(t) = \frac{1}{6}(t + t^2 + t^3 + t^4 + t^5 + t^6) = \frac{t}{6}\,\frac{1-t^6}{1-t} \qquad (\text{si } t \neq 1) .$$

Notons Y la variable aléatoire correspondant au résultat du lancer du deuxième dé ; ce dernier étant non pipé, si $n_1, \cdots, n_6$ sont les numéros (non nécessairement distincts) inscrits sur les faces (et supposés entiers naturels), on a

$$G_Y(t) = \frac{1}{6} \sum_{k=1}^{6} t^{n_k} .$$

Soit enfin $Z = X+Y$. Les variables X et Y étant indépendantes, on a $G_Z(t) = G_X(t)G_Y(t)$. On voudrait que Z suive la loi uniforme sur $[\![1, 12]\!]$, c'est-à-dire que $G_Z(t)$ soit égal à $\dfrac{1}{12}\displaystyle\sum_{k=1}^{12} t^k$, soit encore (pour $t \neq 1$), à $\dfrac{t}{12}\,\dfrac{1-t^2}{1-t}$. Cette condition est réalisée si et seulement si (pour $t \notin \{0,1\}$)

$$G_Y(t) = \frac{G_Z(t)}{G_X(t)} = \frac{1}{2}\,\frac{1-t^{12}}{1-t^6} = \frac{1}{2}(1+t^6) = \frac{1}{2}t^0 + \frac{1}{2}t^6\ .$$

Il est donc nécessaire et suffisant que les valeurs 0 et 6 apparaissent de façon équiprobable lors du lancer du deuxième dé, autrement dit que trois de ses faces portent le numéro 0, et les trois autres le numéro 6.

---

22. On effectue une suite de lancers d'une pièce de monnaie équilibrée. On suppose que les résultats des lancers sont indépendants. On note X la variable aléatoire égale au nombre de lancers nécessaires pour obtenir deux “Pile” consécutifs. Par exemple, pour la suite de lancers PFFPFPPFFF..., on a $X = 7$. De même, on note Y la variable aléatoire égale au nombre de lancers nécessaires pour obtenir “Face” immédiatement suivi de “Pile”. Dans l'exemple ci-dessus, on a donc $Y = 4$. On pose enfin $x_n = P(X = n)$ et $y_n = P(Y = n)$ pour tout n entier naturel non nul.

a. Prouver la relation $\forall n \geq 3 \quad x_n = \dfrac{1}{2}x_{n-1} + \dfrac{1}{4}x_{n-2}$.

b. En déduire la fonction génératrice de la variable aléatoire X. Calculer $\mathrm{E}(X)$.

c. Avec des méthodes analogues, calculer l'espérance de la variable Y. *Ici, on pourra aussi expliciter y_n qui a une expression plus simple que x_n.*

- -

a. Discutons suivant le résultat du premier lancer. Notons F_1 l'événement “le premier lancer donne Face”. On a alors, pour $n \geq 3$,

$$\begin{aligned}
x_n = P(X = n) &= P(\{X = n\} \cap F_1) + P(\{X = n\} \cap \overline{F_1}) \\
&= P(\{X = n\} \cap F_1) + P(\{X = n\} \cap \overline{F_1} \cap F_2) \\
&= P(X = n \mid F_1)\ P(F_1) + P(X = n \mid \overline{F_1} \cap F_2)\ P(\overline{F_1} \cap F_2) \\
&= x_{n-1} \times \frac{1}{2} + x_{n-2} \times \frac{1}{4}\ .
\end{aligned}$$

On a par ailleurs $x_1 = 0$ et $x_2 = \dfrac{1}{4}$.

b. On multiplie par t^n avec $t \in [-1, 1]$, et on somme, pour n de 3 à l'infini, la relation démontrée en **a.**:

$$\sum_{n=3}^{+\infty} x_n t^n = \frac{1}{2}\sum_{n=3}^{+\infty} x_{n-1}t^n + \frac{1}{4}\sum_{n=3}^{+\infty} x_{n-2}t^n\ ,$$

soit

$$G_X(t) - \frac{1}{4}t^2 = \frac{t}{2}\sum_{n=2}^{+\infty} x_n t^n + \frac{t^2}{4}\sum_{n=1}^{+\infty} x_n t^n\ ,$$

donc

$$G_X(t)-\frac{1}{4}t^2=\frac{t}{2}\,G_X(t)+\frac{t^2}{4}\,G_X(t)\ ,$$

et finalement $G_X(t)=\dfrac{t^2}{4-2t-t^2}$.

On a $G_X'(t)=\dfrac{2t(4-t)}{(4-2t-t^2)^2}$, puis $\mathrm{E}(X)=G_X'(1)=6$, c'est le temps d'attente moyen pour voir apparaître deux "Pile" consécutifs.

c. Ici,

$$y_n=P(Y=n)=P(\{Y=n\}\ \cap\ F_1)+P(Y=n|\overline{F_1})\,P(\overline{F_1})=\frac{1}{2^n}+y_{n-1}\times\frac{1}{2}\ .\quad \textbf{(*)}$$

En effet, l'événement $\{Y=n\}\ \cap\ F_1$ correspond à la réalisation de $n-1$ "Face" suivies d'un "Pile", et est donc de probabilité $\dfrac{1}{2^n}$. On peut en déduire que $2^ny_n=2^{n-1}y_{n-1}+1$, donc que la suite (2^ny_n) est arithmétique de raison 1, puis $2^ny_n=2^1y_1+(n-1)=n-1$, enfin $y_n=\dfrac{n-1}{2^n}$. On peut aussi déduire de la relation de récurrence **(*)** ci-dessus que $\displaystyle\sum_{n=2}^{+\infty}y_nt^n=\sum_{n=2}^{+\infty}\left(\frac{t}{2}\right)^n+\frac{1}{2}\sum_{n=2}^{+\infty}y_{n-1}t^n$, soit $G_Y(t)=\dfrac{t^2}{4}\,\dfrac{1}{1-\frac{t}{2}}+\dfrac{t}{2}\,G_Y(t)$, donc $G_Y(t)=\left(\dfrac{t}{2-t}\right)^2$. Enfin, $G_Y'(t)=\dfrac{4t}{(2-t)^3}$, puis $\mathrm{E}(Y)=G_Y'(1)=4$. *Ayant calculé explicitement les y_n, on peut aussi calculer directement l'espérance par* $\displaystyle\mathrm{E}(Y)=\sum_{n=2}^{+\infty}ny_n=\frac{1}{4}\sum_{n=2}^{+\infty}n(n-1)\Big(\frac{1}{2}\Big)^{n-2}=\frac{1}{4}\times\left[\frac{2}{(1-t)^3}\right]_{t=\frac{1}{2}}=4$ (*on aura effectivement reconnu la dérivée seconde de* $t\mapsto\dfrac{1}{1-t}$ *évaluée pour* $t=\dfrac{1}{2}$).

---

23. On répète indéfiniment le lancer d'une pièce équilibréee et l'on note la suite des résultats obtenus. On appelle **série** toute séquence de résultats identiques consécutifs. Par exemple, si les huit premiers lancers donnent, dans l'ordre: PPFFFPFF, on aura obtenu quatre séries, à savoir 'PP', 'FFF', 'P' et 'FF'. Pour tout n entier naturel non nul, on note X_n la variable aléatoire donnant le nombre de séries obtenues lors des n premiers lancers. Dans l'exemple précédent, $X_1=X_2=1$, $X_3=X_4=X_5=2$, $X_6=3$, $X_7=X_8=4$, $\cdots$

a. Pour $n\geq 2$ et $1\leq k\leq n$, prouver la relation

$$P(X_n=k)=\frac{1}{2}\Big(P(X_{n-1}=k)+P(X_{n-1}=k-1)\Big)\ .$$

b. En déduire, par récurrence, la fonction génératrice G_n de la variable X_n.

c. En déduire que la variable $Y_n=X_n-1$ suit une loi binomiale dont on précisera les paramètres.

- -

a. Pour $n\geq 2$, notons E_n l'événement: "les lancers $n-1$ et n donnent le même résultat". Alors

$$\begin{aligned}\{X_n = k\} &= \Big(\{X_n = k\} \,\cap\, E_n\Big) \sqcup \Big(\{X_n = k\} \,\cap\, \overline{E_n}\Big)\\ &= \Big(\{X_{n-1} = k\} \,\cap\, E_n\Big) \sqcup \Big(\{X_{n-1} = k-1\} \,\cap\, \overline{E_n}\Big)\ ,\end{aligned}$$

On a donc

$$P(X_n = k) = P(E_n|X_{n-1} = k)\, P(X_{n-1} = k) + P(\overline{E_n}|X_{n-1} = k-1)\, P(X_{n-1} = k-1)\ .$$

La pièce étant équilibrée (*sinon, il faudrait décomposer plus!*), on a $P(E_n|X_{n-1} = k) = P(E_n) = \dfrac{1}{2}$, de même pour $P(\overline{E_n}|X_{n-1} = k-1)$. On obtient donc

$$P(X_n = k) = \frac{1}{2}\Big(P(X_{n-1} = k) + P(X_{n-1} = k-1)\Big)\ .$$

b. Notons que $X_n(\Omega) = [\![1, n]\!]$ pour tout $n \in \mathbb{N}^*$.

On a $X_1 = 1$ (v.a. constante), donc $G_1(t) = t$. Pour $n \geq 2$, on reprend la relation obtenue en **a.**, on multiplie chaque membre par t^n, on somme, et on assaisonne avant de servir:

$$\begin{aligned}G_n(t) &= \sum_{k=1}^{+\infty} P(X_n = k)\, t^n\\ &= \frac{1}{2}\Big(\sum_{k=1}^{+\infty} P(X_{n-1} = k)\, t^k + \sum_{k=2}^{+\infty} P(X_{n-1} = k-1)\, t^k\Big)\\ &= \frac{1}{2}\Big(G_{n-1}(t) + \sum_{k=1}^{+\infty} P(X_{n-1} = k)\, t^{k+1}\Big)\\ &= \frac{1+t}{2}\, G_{n-1}(t)\ .\end{aligned}$$

On a écrit des sommes infinies par commodité, mais ce sont toutes en fait des sommes finies, et il n'y a donc aucun problème de convergence, les fonctions génératrices sont des fonctions polynomiales et sont défiinies sur tout $\mathbb{R}$.

On déduit de ce calcul que $\forall t \in \mathbb{R} \quad \forall n \in \mathbb{N}^* \quad G_n(t) = t\left(\dfrac{1+t}{2}\right)^{n-1}$.

c. On sait que $X_n(\Omega) = [\![1, n]\!]$ et

$$G_n(t) = \frac{t}{2^{n-1}} \sum_{k=0}^{n-1} \binom{n-1}{k}\, t^k = \sum_{k=1}^{n} \binom{n-1}{k-1} \frac{1}{2^{n-1}}\, t^k\ .$$

Par identification, on a $P(X_n = k) = \dfrac{1}{2^{n-1}} \dbinom{n-1}{k-1}$ pour tout $k \in [\![1, n]\!]$.

Si on pose $Y_n = X_n - 1$, alors $Y_n(\Omega) = [\![0, n-1]\!]$ et

$$\forall k \in [\![0, n-1]\!] \quad P(Y_n = k) = P(X_n = k+1) = \frac{1}{2^{n-1}} \binom{n-1}{k} = \binom{n-1}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-1-k}\ .$$

La variable Y_n suit la loi binomiale $\mathcal{B}\Big(n-1, \dfrac{1}{2}\Big)$.

---

24. On répète une expérience aléatoire ayant la probabilité p de réussir et $q = 1 - p$ d'échouer, définissant ainsi une suite de variables de Bernoulli indépendantes $(X_n)_{n\geq 1}$. Pour $m \in \mathbb{N}^*$, on note S_m la variable aléatoire déterminant le nombre d'essais jusqu'à l'obtention de m succès: $S_m = k \iff \big(X_1 + \cdots + X_k = m \quad \text{et} \quad X_1 + \cdots + X_{k-1} < m\big)$.

a. Déterminer la loi et la fonction génératrice de S_1.

b. Même question avec $S_m - S_{m-1}$, pour $m \geq 2$.

c. En déduire la fonction génératrice, puis la loi, de S_m.

- -

a. On reconnaît le temps d'attente du premier succès, donc S_1 suit la loi géométrique de paramètre p. On a alors $G_{S_1}(t) = \dfrac{pt}{1-qt}$, avec un rayon de convergence $R = \dfrac{1}{q} > 1$.

b. La variable $S_m - S_{m-1}$ représente le temps d'attente du m-ème succès, compté à partir du $(m-1)$-ème succès. Cette variable suit aussi la loi géométrique de paramètre p.

c. On a $S_m = S_1 + \displaystyle\sum_{k=2}^{m}(S_k - S_{k-1})$, les variables sommées étant mutuellement indépendantes.

En effet, si on se donne un m-uplet $(n_1, \cdots, n_m)$ d'entiers naturels non nuls, l'événement

$$E = \big\{S_1 = n_1\ ,\ S_2 - S_1 = n_2\ ,\ \cdots\ ,\ S_m - S_{m-1} = n_m\big\}$$

coïncide avec l'événement

$$\big\{X_{n_1} = X_{n_1+n_2} = \cdots = X_{n_1+n_2+\cdots+n_m} = 1\ ;\ X_k = 0 \text{ pour les autres } k \text{ dans } [\![1, n_1 + \cdots + n_m]\!]\big\}\ .$$

Les X_k étant mutuellement indépendantes, on a alors

$$\begin{aligned} P(E) &= p^m\, q^{n_1+\cdots+n_m-m} = \big(pq^{n_1-1}\big)\big(pq^{n_2-1}\big)\cdots\big(pq^{n_m-1}\big) \\ &= P(S_1 = n_1)\, P(S_2 - S_1 = n_2)\ \cdots\ P(S_m - S_{m-1} = n_m)\ , \end{aligned}$$

ce qui prouve l'indépendance mutuelle des variables S_1, $S_2 - S_1$, $\cdots$, $S_m - S_{m-1}$. La fonction génératrice S_m est alors le produit des fonctions génératrices, ce qui donne

$$G_{S_m}(t) = \Big(\frac{pt}{1-qt}\Big)^m = p^m\, t^m\, (1-qt)^{-m}\ .$$

Par le développement en série entière de $(1+x)^\alpha$, on obtient, si $|t| < \dfrac{1}{q}$,

$$\begin{aligned} G_{S_m}(t) &= p^m\, t^m \sum_{k=0}^{+\infty} \frac{(-m)(-m-1)\cdots(-m-k+1)}{k!}\,(-qt)^k \\ &= p^m\, t^m \sum_{k=0}^{+\infty} \frac{(m+k-1)!}{k!\,(m-1)!}\,(qt)^k \\ &= p^m\, t^m \sum_{k=0}^{+\infty} \binom{m+k-1}{m-1}\, q^k\, t^k \end{aligned}$$

$$= \sum_{n=m}^{+\infty} \binom{n-1}{m-1} p^m\, q^{n-m}\, t^n\ .$$

Il est clair que $S_m(\Omega) = [\![m, +\infty [\![$. Pour $n \geq m$, on a donc $P(S_m = n) = \binom{n-1}{m-1} p^m q^{n-m}$.

25.a. Trouver la constante k et une condition sur le réel a pour que la suite $\left(\left(\dfrac{a}{a+1}\right)^n k\right)_{n\in\mathbb{N}}$ soit une distribution de probabilités sur l'ensemble $\mathbb{N}$.

b. Soit (X_n) une suite de variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $\mathbb{N}$, et ayant toutes la loi définie par la distribution de probabilités ci-dessus. Déterminer la fonction génératrice de la variable $S_n = \displaystyle\sum_{k=1}^{n} X_k$.

c. Calculer de deux manières l'espérance et la variance de S_n.

- -

a. Posons $p_n = k\left(\dfrac{a}{a+1}\right)^n$ pour n entier naturel. La condition recherchée est que les réels p_n soient tous positifs et de somme 1. La positivité est réalisée si et seulement si $a \in]-\infty, -1[\ \cup\ [0, +\infty[$. Mais, si $a < -1$, alors $\dfrac{a}{a+1} > 1$, donc la série géométrique $\sum p_n$ diverge. Si $a \geq 0$, alors $0 \leq \dfrac{a}{a+1} < 1$, donc la série géométrique $\displaystyle\sum_{n\geq 0}\left(\dfrac{a}{a+1}\right)^n$ converge et a pour somme $a+1$, qui est un réel strictement positif, donc les conditions seront satisfaites si et seulement si on prend $k = \dfrac{1}{a+1}$.

Bilan. Les conditions recherchées sont $a \geq 0$ et $k = \dfrac{1}{a+1}$.

b. La fonction génératrice de chaque variable aléatoire X_k est donnée par

$$G_{X_k}(t) = \sum_{n=0}^{+\infty} P(X_k = n)\, t^n = \frac{1}{a+1} \sum_{n=0}^{+\infty} \left(\frac{at}{a+1}\right)^n = \frac{1}{a+1-at}$$

avec un rayon de convergence $R = 1 + \dfrac{1}{a} > 1$ ($R = +\infty$ dans le cas particulier $a = 0$). Les variables X_k étant mutuellement indépendantes, on a alors

$$\forall t \in]-R, R[\qquad G_{S_n}(t) = \prod_{k=1}^{n} G_{X_k}(t) = \big(G_{X_1}(t)\big)^n = \frac{1}{(a+1-at)^n}\ .$$

c. • En utilisant la fonction génératrice, $G'_{S_n}(t) = \dfrac{na}{(a+1-at)^{n+1}}$, puis $G''_{S_n}(t) = \dfrac{n(n+1)a^2}{(a+1-at)^{n+2}}$, donc $\mathrm{E}(S_n) = G'_{S_n}(1) = na$, puis

$$\mathrm{V}(S_n) = G'_{S_n}(1) + G''_{S_n}(1) - \big(G'_{S_n}(1)\big)^2 = na + n(n+1)a^2 - (na)^2 = na(a+1)\ .$$

• Calculons d'abord l'espérance de X_1 en utilisant $\displaystyle\sum_{n=1}^{+\infty} nx^{n-1} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2}$ pour $|x| < 1$:

$$\mathrm{E}(X_1)=\sum_{n=1}^{+\infty} n\,P(X_1=n)=\frac{1}{a+1}\,\sum_{n=1}^{+\infty} n\left(\frac{a}{a+1}\right)^n=\frac{a}{(a+1)^2}\left(\frac{1}{1-\dfrac{a}{a+1}}\right)^2=a$$

Par linéarité de l'espérance, $\mathrm{E}(S_n)=n\,\mathrm{E}(X_1)=na$. De la même façon, en utilisant $\displaystyle\sum_{n=2}^{+\infty} n(n-1)x^{n-2}=\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Big(\frac{1}{1-x}\Big)=\frac{2}{(1-x)^3}$ pour $|x|<1$:

$$\mathrm{E}\big(X_1(X_1-1)\big)=\sum_{n=2}^{+\infty} n(n-1)\,P(X_1=n)=2\,\frac{a^2}{(a+1)^3}\left(\frac{1}{1-\dfrac{a}{a+1}}\right)^3=2a^2\ ,$$

puis $\mathrm{E}(X_1^2)=\mathrm{E}(X_1^2-X_1)+\mathrm{E}(X_1)=2a^2+a$, puis $\mathrm{V}(X_1)=\mathrm{E}(X_1^2)-\mathrm{E}(X_1)^2=a(a+1)$, et enfin puisque les variables X_k sont mutuellement indépendantes,

$$\mathrm{V}(S_n)=\sum_{k=1}^{n}\mathrm{V}(X_k)=n\,\mathrm{V}(X_1)=na(a+1)\ .$$

• En fait, il est bien plus simple de remarquer tout de suite que la loi de X “ressemble” à une loi géométrique, et plus précisément que la variable $Y=X+1$ suit la loi géométrique de paramètre $k=\dfrac{1}{a+1}$. On a alors $\mathrm{E}(Y)=\dfrac{1}{k}=a+1$, donc $\mathrm{E}(X_k)=a$ pour tout k, puis $\mathrm{E}(S_n)=na$ par linéarité de l'espérance. Ensuite, $\mathrm{V}(X_k)=\mathrm{V}(Y)=\dfrac{1-k}{k^2}=a(a+1)$ puis, par indépendance, $\mathrm{V}(S_n)=\displaystyle\sum_{k=1}^{n}\mathrm{V}(X_k)=na(a+1)$.

26*. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On définit sa **fonction caractéristique** Φ_X par $\Phi_X(t)=\mathrm{E}(e^{itX})$ pour tout t réel.

a. Montrer que Φ_X est définie et continue sur $\mathbb{R}$.

b. Pour $k\in\mathbb{N}$, calculer l'intégrale $I_k=\dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi}\Phi_X(t)\,e^{-ikt}\,\mathrm{d}t$. Que dire de deux variables aléatoires X et Y, à valeurs dans $\mathbb{N}$, telles que $\Phi_X=\Phi_Y$?

c. On suppose que X est d'espérance finie. Montrer que la fonction Φ_X est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et déterminer $\Phi'_X(0)$.

- -

a. On a donc $\Phi_X(t)=\mathrm{E}(e^{itX})=\displaystyle\sum_{n=0}^{+\infty}P(X=n)\,e^{int}$ pour tout réel t tel que cette série soit convergente. Mais si l'on pose $u_n(t)=P(X=n)\,e^{int}$, on a $|u_n(t)|=P(X=n)$, donc $\|u_n\|_\infty=P(X=n)$, donc $\displaystyle\sum_{n=0}^{+\infty}\|u_n\|_\infty=1$, ce qui garantit la convergence normale sur $\mathbb{R}$ de la série de fonctions $\sum u_n$, d'où l'existence de $\Phi_X(t)=\displaystyle\sum_{n=0}^{+\infty}u_n(t)$ pour tout t, et la continuité sur $\mathbb{R}$ (puisque chacune des fonctions u_n est continue sur $\mathbb{R}$ et qu'il y a CVN).

b. Fixons $k \in \mathbb{N}$. On a alors $I_k = \displaystyle\int_{-\pi}^{\pi} \sum_{n=0}^{+\infty} v_n(t)\,\mathrm{d}t$ où l'on pose $v_n(t) = \dfrac{1}{2\pi} P(X = n) e^{i(n-k)t}$.

On est bigrement tenté d'intervertir série et intégrale! Or il se trouve que les fonctions v_n sont continues sur le segment $[-\pi, \pi]$ et que la série de fonctions $\sum v_n$ converge normalement sur ce **segment** puisque $\|v_n\|_\infty = \dfrac{1}{2\pi} P(X = n)$, terme général d'une série convergente, nous sommes donc autorisés à faire cette interversion. Donc

$$I_k = \sum_{n=0}^{+\infty} \int_{-\pi}^{\pi} v_n(t)\,\mathrm{d}t = \frac{1}{2\pi} \sum_{n=0}^{+\infty} P(X = n) \int_{-\pi}^{\pi} e^{i(n-k)t}\,\mathrm{d}t = P(X = k)$$

puisque $\displaystyle\int_{-\pi}^{\pi} e^{i(n-k)t}\,\mathrm{d}t = 2\pi\,\delta_{n,k}$ (symbole de Kronecker).

On en déduit que, si $\Phi_X = \Phi_Y$, alors X et Y ont la même loi (*Attention! Ne pas répondre* $X = Y$). De même que la fonction génératrice G_X, la fonction caractéristique caractérise la loi de la variable aléatoire.

c. Reprenons les fonctions u_n introduites en **a.**, elles sont $\mathcal{C}^1$ et $u_n'(t) = in\,P(X = n)\,e^{int}$, donc $\|u_n'\|_\infty = n\,P(X = n)$. Si on suppose $X \in L^1(\Omega)$, alors la série $\sum \|u_n'\|_\infty$ converge, autrement dit la série de fonctions $\sum u_n'$ converge normalement sur $\mathbb{R}$, ce qui autorise à affirmer (sachant qu'il y a convergence simple de la série $\sum u_n$) que la fonction $\Phi_X = \displaystyle\sum_{n=0}^{+\infty} u_n$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}$, et à intervertir série et dérivée. Ainsi,

$$\forall t \in \mathbb{R} \qquad \Phi_X'(t) = \sum_{n=0}^{+\infty} u_n'(t) = i \sum_{n=1}^{+\infty} n\,P(X = n)\,e^{int}\ .$$

En particulier, $\Phi_X'(0) = i\,\mathrm{E}(X)$.

---

27. On note X une variable aléatoire à valeurs dans $\mathbb{N}$. On se donne une suite $(X_i)_{i\geq 1}$ de variables aléatoires indépendantes sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, de même loi que X. Soit d'autre part N une variable aléatoire à valeurs dans $\mathbb{N}$, que l'on suppose indépendante des X_i. Pour tout $\omega \in \Omega$, on pose

$$T(\omega) = \prod_{i=1}^{N(\omega)} X_i(\omega)\ .$$

On conviendra que $T(\omega) = 1$ si $N(\omega) = 0$.

On suppose que X est d'espérance finie notée m. En utilisant la fonction génératrice G_N de la variable N, donner une condition nécessaire et suffisante pour que T soit d'espérance finie, et exprimer dans ce cas $\mathrm{E}(T)$. Étudier le cas particulier où N et X sont des variables de Poisson.

- -

La variable T étant positive, calculons dans $[0, +\infty]$ où toutes les interversions de sommations sont permises par le programme. Notons que, pour tout couple (k, n) d'entiers naturels, on a l'égalité des événements

$$\{T=k\}\ \cap\ \{N=n\}\ =\ \{N=n\}\ \cap\ \Big\{\prod_{i=1}^{n} X_i = k\Big\}$$

et que, les variables N, $X_1, \cdots, X_n$ étant indépendantes, on déduit du lemme des coalitions l'indépendance des variables N et $\displaystyle\prod_{i=1}^{n} X_i$. Donc

$$P(T=k, N=n) = P(N=n)\ P\Big(\prod_{i=1}^{n} X_i = k\Big)\ .$$

C'est parti, calculons!

$$\begin{aligned}
\mathrm{E}(T) &= \sum_{k=0}^{+\infty} k\ P(T=k)\ =\ \sum_{k=0}^{+\infty} k\ \Big(\sum_{n=0}^{+\infty} P(T=k, N=n)\Big) \\
&= \sum_{k=0}^{+\infty} k\ \Big(\sum_{n=0}^{+\infty} P(N=n)\ P\Big(\prod_{i=1}^{n} X_i = k\Big)\Big) \\
&= \sum_{n=0}^{+\infty} P(N=n)\ \Big(\sum_{k=0}^{+\infty} k\ P\Big(\prod_{i=1}^{n} X_i = k\Big)\Big) \\
&= \sum_{n=0}^{+\infty} P(N=n)\ \mathrm{E}\Big(\prod_{i=1}^{n} X_i\Big) \\
&= \sum_{n=0}^{+\infty} P(N=n)\ \big(\mathrm{E}(X)\big)^n\ .
\end{aligned}$$

On en déduit que T est d'espérance finie si et seulement si la série $\displaystyle\sum_{n\geq 0} P(N=n)\, m^n$ est convergente et que, dans ce cas, $\mathrm{E}(T) = G_N\big(\mathrm{E}(X)\big) = G_N(m)$.

Si $X \sim \mathcal{P}(\lambda)$ et $N \sim \mathcal{P}(\mu)$ avec $\lambda > 0$ et $\mu > 0$, alors la fonction génératrice de N est une série entière de rayon de convergence infini, donc $\mathrm{E}(T) < +\infty$ et, précisément, $\mathrm{E}(T) = G_N(\lambda) = e^{\mu(\lambda-1)}$.

Exercices avec Python.

28. Une pièce a la probabilité $p \in]0,1[$ de tomber sur pile. On la lance jusqu'à avoir obtenu deux fois pile, et on note X le nombre de fois où elle est tombée sur "face".

a. Loi de X. Montrer que X est d'espérance finie et la calculer.

Lorsque $X = n$, on place $n+1$ boules numérotées de 0 à n dans une urne, et on en tire une au hasard. On note Y le numéro de la boule tirée.

b. Utiliser Python pour simuler cette expérience aléatoire.

c. Loi de Y. Montrer que Y est d'espérance finie et la calculer.

- -

a. Pour tout n entier naturel, l'événement $\{X = n\}$ est réalisé si et seulement si la séquence des $n+2$ premiers lancers est l'une des $n+1$ suivantes, comportant n fois F et deux fois P (le dernier lancer est un P=pile!):

$$FFF...FFPP\ ,\ FFF...FPFP\ ,\ \cdots\ ,\ FFPF...FFFP\ ,\ FPFF...FFFP\ ,\ PFFF...FFFP\ .$$

Sur tous les $n+2$ premiers lancers possibles, chacune de ces séquences a une probabilité égale à p^2q^n, avec $q = 1-p$. On en déduit que

$$\forall n \in \mathbb{N} \qquad P(X = n) = (n+1)\, p^2 q^n\ .$$

NB: On vérifiera par un calcul facile que $\displaystyle\sum_{n=0}^{+\infty} P(X = n) = 1$, on a donc bien un système quasi-complet d'événements.

Le calcul de l'espérance nous amène à considérer la série de terme général $n\,P(X = n)$, soit $\sum n(n+1)p^2q^n$. Cette série converge (règle de d'Alembert par exemple), et on obtient

$$\mathrm{E}(X) = \sum_{n=1}^{+\infty} n(n+1)p^2q^n = p^2q \sum_{n=1}^{+\infty} n(n+1)q^{n-1} = p^2q \sum_{n=2}^{+\infty} n(n-1)q^{n-2}\ .$$

Pour $x \in]-1,1[$, posons $s(x) = \displaystyle\sum_{n=0}^{+\infty} x^n = \frac{1}{1-x}$. Alors s est de classe $\mathcal{C}^\infty$, et sa dérivée seconde sur $]-1,1[$ s'exprime de deux manières: $s''(x) = \displaystyle\sum_{n=2}^{+\infty} n(n-1)x^{n-2} = \frac{2}{(1-x)^3}$. On en conclut que

$$\mathrm{E}(X) = \frac{2p^2q}{(1-q)^3} = 2\,\frac{q}{p} = 2\,\frac{1-p}{p}\ .$$

b. *cf.* script.

c. Comme $\big(\{X = n\}\big)_{n\in\mathbb{N}}$ est un système quasi-complet d'événements, la formule des probabilités totales donne, pour tout k entier naturel,

$$P(Y = k) = \sum_{n=k}^{+\infty} P(Y = k \mid X = n)\ P(X = n) = \sum_{n=k}^{+\infty} \frac{1}{n+1}(n+1)p^2q^n = p^2\,\frac{q^k}{1-q} = p\,q^k\ .$$

On en déduit

$$\mathrm{E}(Y) = \sum_{k=0}^{+\infty} k\,P(Y = k) = pq \sum_{k=1}^{+\infty} k\,q^{k-1}\ ,$$

on reconnaît cette fois-ci $\displaystyle\sum_{k=1}^{+\infty} k\,x^{k-1} = \frac{\mathrm{d}}{\mathrm{d}x}\Big(\sum_{k=0}^{+\infty} x^k\Big) = \frac{\mathrm{d}}{\mathrm{d}x}\Big(\frac{1}{1-x}\Big) = \frac{1}{(1-x)^2}$, calcul valable sur $]-1,1[$, donc $\mathrm{E}(Y) = \dfrac{q}{p} = \dfrac{1-p}{p}$.