Aide à la compréhension sur le cours sur les Probabilités et les variables aléatoires finies

Une première démarche expérimentale

Je vous propose d'engager ce chapitre avec des « séances pratiques » ou expositions de petits problèmes liés à ce chapitre.

Toto rentre de l'école après sa première journée d'école de CM2. Lorsqu'il ouvre la porte de sa maison, sa mère qui l'attendait lui demande comment s'est passée la journée.
– « On est 32 dans la classe », dit Toto.
– « Formidable ! s'exclame la mère. Il y aura sûrement au moins deux élèves qui fêteront le même jour leur anniversaire ! »
– « Ah non ! Je préfère quand j'apporte un gâteau pour moi tout seul » dit Toto un peu déçu.
– « Ne t'inquiète pas, le rassure sa maman, il y a de fortes chances que tu sois le seul qui soit né le 13 mars dans ta classe. »

En effet, la maman a un peu moins d'une chance sur quatre de se tromper lors de sa première affirmation et pour sa seconde réplique, elle a un peu moins d'une chance sur dix de se tromper.

Deux amis discutent entre eux à propos d'un nouveau jeu télévisé.
L'un demande à l'autre : « tu as regardé la nouvelle émisssion *Mettez à l'épreuve votre sens probabiliste* sur France π ? »
– « Moi, tu sais, le samedi soir, je préfère bosser mes intégrales. J'étais justement en train de tester la méthode de Monte-Carlo sur $\displaystyle\int_0^1 e^{-t^2}\,dt$. »
– « Super ! C'est en plein dans les probas, tu sais, mais l'approximation n'est pas terrible. C'est en $\mathscr{O}\left(\dfrac{1}{\sqrt{n}}\right)$ il me semble, où n est le nombre de points, tu m'auras compris. Enfin bref, je te raconte l'émission d'hier soir. Chaque soir, ils présentent un nouveau jeu. Hier, il fallait deviner le gros lot caché derrière l'une des quatre portes. Le gros lot, c'était un stage d'immersion d'une semaine à la Silicon Valley pour visiter là où ils expérimentent leur tout nouvel ordinateur quantique. T'imagines le truc de malade. Chacune des trois autres portes renfermait un lot de consolation : une semaine aux Seychelles ou un truc comme ça. Enfin bref. Le principe du jeu est le suivant. Le joueur sélectionné choisi au pif une des quatre portes.

Dans le pire des cas, il y a au moins deux autres portes perdantes parmi les trois portes non sélectionnées. Vu que l'animateur connaît la bonne porte, il dévoile une mauvaise porte au candidat et il lui propose de rejouer ou non. Eh bien tu sais ce qu'il s'est passé hier ? «
– « Non. Mais je suppose que tu vas me le dire »
– « Exact ! Eh bien, il dit à l'animateur que ce n'est pas la peine et qu'il préfère suivre son premier instinct. Résultat des courses, il a perdu et il se retrouve à passer une semaine aux Caraïbes ou un truc comme ça. Quelle nouille ! »

La meilleur stratégie est ici de modifier son choix parmi les deux autres portes – celle non choisie en premier lieu par le candidat et celle non dévoilée par l'animateur. Par cette stratégie, le candidat aura 1.5 fois plus de chances de gagner. Enfin bon, le voyage aux Antilles, c'est pas mal non plus.

Prenez deux dés à six faces. Lancez-les et répétez l'opération jusqu'à obtenir une somme égale à 5. Notez alors le nombre de lancers effectués ; je précise qu'un lancer est le fait de lancer les deux dés en même temps.
Répétez ce protocole disons 100000 fois. Vous obtenez une séquence de 100000 nombres entiers.
Je ne peux pas prédire quel sera votre $833^{ème}$ entier ni d'ailleurs le $2020^{ème}$, mais je peux prédire que la moyenne obtenue sur vos 100000 nombres sera environ égale à 9.
Vous n'êtes pas convaincu et vous vous dîtes que lancer plus que 100000 fois des dés à six faces prend un peu de temps ?
Heureusement, l'informatique peut aider dans ce cas.
Exécutez ce programme et constatez ...

```
from random import *

def X() :
    """ modélise l'expérience pour obtenir un nombre de lancers """
    compteur=1
    while randint(1,6)+randint(1,6) != 5 :
        compteur+=1
    return compteur

Liste_Nombres=[X() for k in range(100000)] # on a nos 100000 valeurs
print(sum(Liste_Nombres)/100000) # on a la moyenne
```

Personnellement, mon $833^{ème}$ entier vaut 10 et mon $2020^{ème}$ entier vaut 6.

Une personne a été invitée chez son ami pour passer l'après-midi à jouer à des jeux de « hasard ». L'ami que l'on désignera par Jean-Michel par la suite possède un très gros sac où il a mis toutes les matrices inversibles de format 833×833 à coefficients réels. Il faut donc un très gros sac.
– « Regarde, Jean-Claude, dit Jean-Michel à son copain. J'ai mis dans ce cas tous les éléments de $\mathscr{M}_{833}(\mathbb{R})$. J'y ai passé toute la journée d'hier. J'espère ne pas en avoir oublié. À ton avis, y a-t-il plus de matrices inversibles ou plus de matrices non inversibles ? »

– « Tu n'as pas des questions plus difficiles ? Il y a plus de matrices inversibles. »
– « Perdu, répond Jean-Mimi. Il y en a exactement le même nombre : autant de matrices inversibles que de matrices non inversibles. Je te propose alors le jeu suivant. Tu me donnes 100 euros pour pouvoir jouer. Ensuite, une fois que tu m'as donné ton billet de 100 euros, je te laisse mélanger les matrices dans le sac sans tricher et je te laisse piocher deux matrices au hasard. Si l'une des deux matrices que tu as choisies est non inversible, alors je te donne ce billet violet. Tu sais ce que c'est qu'un billet violet ? »
– « Euh non, j'en n'ai jamais vu ! »
– « Eh, bien c'est un billet de 500 euros ! J'ai dû résoudre le DM de maths d'arithmétique du neveu de la banquière pour en avoir un. »
– « Écoute, je veux bien jouer à ton jeu, mais je ne veux pas perdre ton amitié en profitant de ta naïveté. S'il y a autant de matrices inversibles que de matrices non inversibles comme tu me l'as annoncé – en plus, j'ai vérifié pendant que tu me parlais et effectivement, je compte autant de nombres réels que de matrices inversibles tout comme pour les matrices non inversibles, donc il y en a bien autant – je devrais avoir une chance sur deux de piocher une matrice non inversible et donc trois chances sur quatre de piocher au moins une matrice non inversible dans les deux que je prendrai. Tu es perdant. Allez, j'essaie pour du beurre. »
Jean-Claude mélange, pioche deux matrices A et B et s'aperçoit qu'en plus d'être inversibles, les deux matrices A et B piochées sont diagonalisables sur le corps $\mathbb{C}$. Jean-Claude a passé toute sa journée du lendemain à trouver les matrices de passage dans $GL_{833}(\mathbb{R})$.
– « Finalement, j'ai pas eu de chance sur ce coup-là », conclut Jean-Claude.
– « En es-tu vraiment sûr ? », lui demande Jean-Michel ?
Après un instant de réflexion, Jean-Claude répond.
– « OK, j'ai compris ! Pour prendre ma revanche, je te propose de remplacer tes matrices dans $\mathscr{M}_{833}(\mathbb{R})$ par des matrices dans $\mathscr{M}_{833}(\mathbb{Z}/2\mathbb{Z})$. Tu veux rejouer ? Tu me donnes 100 euros et si tu pioches au moins une matrice non inversible, alors je te file 500 euros ? »
Après quelques secondes de silence, Jean-Michel répond : « Là, c'est toi qui seras encore perdant. Tu perdras en moyenne 358.30071 euros, environ. Je peux te calculer les décimales suivantes si tu veux. Pour avoir un jeu équitable, il faudrait que tu me donnes seulement 109 euros et juste un peu moins de 10 centimes », au cas où je gagne. Ou alors, si tu n'as pas de monnaie, on garde la mise à 100 euros et le gain à 500 euros, mais on prend des matrices dans $\mathbb{Z}/11\mathbb{Z}$. Dans ce cas, je t'assure que tu es légèrement gagnant, mais le jeu est relativement équitable. Dans le corps $\mathbb{Z}/7\mathbb{Z}$, tu serais perdant. »
– « Et si on essayait dans $\mathbb{Z}/833\mathbb{Z}$? »
– « Enfin Jean-Claude ! Qu'est-ce que tu me racontes ! Tu sais bien que $833 = 7^2 \times 17$. Non ? »
– « Ah, oui. J'avais oublié. Et dans $\mathbb{Z}/17\mathbb{Z}$ dans ce cas ? »
– « Bof. Je n'ai pas envie de me retrouver sur la paille... En fait, je connais une activité bien meilleure. On prend des spaghettis et en les lançant sur le carrelage, on peut approximer π. Est-ce que ça te tente ? »
– « OK. Mais attention, crues les spaghettis ! »
– « Non, non ! L'important, c'est leur longueur, pas leur forme dans ce genre de calcul. Tu sais bien ? On jette les spaghettis et on calcule le nombre moyen de points d'intersections avec les lignes du carrelage, celles qui vont du frigo à la table basse du salon. »
– « OK, allons-y dans ce cas. Des spaghettis bio, ça ira ? »
– « Sans problème », s'exclame Jean-Michel.

On l'aura compris, l'objectif du probabiliste est de prédire les choses, d'estimer ses chances de gain ou d'occurrence d'une certaine situation afin d'agir en conséquence.
Tout cela demande un peu de formalisme ...

Les premières notions de probabilités

Dans toute la suite, le symbole « $\cup$ » signifie « réunion d'ensembles », alors que le symbole « $\sqcup$ » signifie « réunion disjointe d'ensembles ».
Si Ω est un ensemble, on rappelle que $\mathscr{P}(\Omega)$ désigne l'ensemble des parties de Ω :

$$\mathscr{P}(\Omega) = \Big\{A \mid A \subset \Omega\Big\}.$$

Le but du début du chapitre est de formaliser le concept d'expérience aléatoire.
Une expérience aléatoire est un protocole expérimental qui permet ensuite de faire des mesures, pas forcément prévisibles. Il faut mettre en place plusieurs ingrédients.
Premièrement, on utilise souvent la notation ω pour désigner une expérience. La lettre ω désignera alors le résultat de cette expérience.
On appelle univers de notation traditionnelle Ω, l'ensemble de toutes les expériences possibles associées à un protocole expérimental. Il s'agit toujours d'un ensemble non vide, car sinon, on raisonne sur un protocole expérimental vide – sans jeu de mots.

Voyons cela à travers des exemples de l'exemple 1. Bien évidemment, comme dans toute modélisation, il y a parfois plusieurs bonnes façons de voir les choses.

• Si on lance deux dés traditionnels à six faces, le plus simple est déjà de différencier les deux dés. En effet, si on ne différencie pas les deux dés, à l'affirmation : « j'ai obtenu un 4 et un 2 lors de mon lancer » on ne sait pas exactement ce qu'il s'est passé puisqu'on ne sait pas lequel des deux dés a fait 4 ou 2. Vous me direz que l'on s'en moque puisqu'ensuite on effectue la somme des deux faces... C'est une façon de voir les choses. Il est tout de même plus précis de différencier les deux dés.
Disons que l'on a deux dés différents : un dé rouge et un dé bleu. Chaque expérience peut donc être modélisée par un couple :

$$\omega = (i, j),$$

où i et j sont deux entiers entre 1 et 6, en décidant que la première composante i concerne la face du dé bleu, alors que la seconde composante concerne la face du dé rouge.
On vient de modéliser sans perte d'information l'expérience. En effet, si $\omega = (3, 3)$, on sait que les deux dés ont fait 3, si $\omega = (1, 6)$, on sait que le dé bleu a fait 1 et que le dé rouge a fait 6 et réciproquement, si le lancer des deux dés a fait : « dé rouge : 5 et dé bleu : 6 », alors on associera à cette expérience $\omega = (6, 5)$.
Bien entendu, on peut toujours considérer qu'il y a une perte d'information entre les deux expériences : $\omega_1 = (1, 6)$ et $\omega_2 = (1, 6)$. Les deux expériences auront donné un dé bleu à 1 et un dé rouge à 6, mais peut-être que lors de la première expérience, les deux dés seront assez proches sur la table, alors que lors de la deuxième expérience, le dé rouge sera tombé de la table et aura fait 6 une fois stabilisé à terre. Peut-être également que ce degré de précision ne nous importe pas dans la suite des calculs...

Dans cette modélisation, chaque expérience ω est un couple appartenant à $[\![1,6]\!]^2$ et l'univers Ω des possibles est l'ensemble :

$$\Omega = [\![1,6]\!]^2.$$

Pour terminer sur ce premier point, le fait d'effectuer la somme des deux faces est modélisé par une variable aléatoire. Ici, il s'agit de la variable aléatoire :

$$X : \left| \begin{array}{ccc} \Omega & \longrightarrow & [\![2,12]\!] \\ \omega = (i,j) & \longmapsto & X(\omega) = i + j \end{array} \right. .$$

On verra plus loin dans ce chapitre ce que l'on fera avec tout cela.
Pour l'instant, on retient qu'une expérience est modélisée par un objet ω dont la nature dépend du protocole expérimental proprement dit, que l'univers des possibles Ω est l'ensemble des résultats ω des expériences réalisées et calculer un nombre à partir de cette expérience revient à effectuer une mesure sur ces expériences et que ces mesures sont modélisables par une variable aléatoire $X : \Omega \longrightarrow \mathbb{R}$, si la mesure effectuée est un nombre réel.
Passons à d'autres situations pour y voir plus clair et avoir un aperçu un peu plus large des situations (innombrables) qui peuvent se présenter.

• Si l'on jette un disque de rayon 1 dans le plan, $\mathbb{R}^2$, on peut modéliser chaque expérience ω par un point (x,y) du plan qui indiquera le centre du disque dans le plan. Connaissant ω, on saura exactement la position du disque. Dans cette modélisation, l'univers des possibles est $\Omega = \mathbb{R}^2$. On effectue ensuite une mesure associée à la variable aléatoire :

$$X : \left| \begin{array}{ccl} \mathbb{R}^2 & \longrightarrow & \mathbb{N} \\ \omega = (x,y) & \longmapsto & \mathrm{Card}\Big(\Big\{(a,b) \in \mathbb{Z}^2 \mid (a-x)^2 + (b-y)^2 \leqslant 1\Big\}\Big) \end{array} \right. .$$

• Pour le troisième point, on lance potentiellement la pièce de monnaie une infinité de fois (même si en pratique ceci est impossible) pour pouvoir calculer ensuite la première occurrence de la séquence « P–P–F ».
Une expérience ω peut donc ici être modélisée par une suite :

$$\omega = (\omega_n)_{n \in \mathbb{N}^*} \in \Big\{P, F\Big\}^{\mathbb{N}^*},$$

suite à valeurs dans l'ensemble à deux éléments $\{P,F\}$ – pour les endormis : « P » pour « pile » et « F » pour « face ».
La modélisation se fait de la manière suivante : connaissant la suite des lancers, on construit la suite ω associée de la manière suivante. Pour tout $n \in \mathbb{N}^*$, l'élément ω_n de la suite ω est le résultat du $n^{\text{ème}}$ lancer de la pièce de monnaie.
Réciproquement, connaissant la suite $\omega = (\omega_n)_{n \in \mathbb{N}^*}$, on sait exactement ce qui s'est passé : au $n^{\text{ème}}$ lancer de la pièce, on a obtenu ω_n, avec $\omega_n \in \{P,F\}$. Bien entendu, on perd toujours un peu d'information dans cette modélisation. Par exemple, combien de temps a-til fallu attendre pour que la pièce se stabilise lors du $n^{\text{ème}}$ lancer ? Information somme toute inutile au vu des mesures que l'on s'apprête à faire, à savoir détecter la première séquence « P–P–F ».
L'univers Ω des possibles est ici :

$$\Omega = \{P,F\}^{\mathbb{N}^*}.$$

La variable aléatoire associée à la mesure faite est ici :

$$X : \left| \begin{array}{ccl} \Omega & \longrightarrow & \mathbb{N}^* \cup \{+\infty\} \\ \omega = (\omega_n)_{n\in\mathbb{N}^*} & \longmapsto & \left\{ \begin{array}{l} \min\Big\{k \in \mathbb{N}^* \mid (\omega_k, \omega_{k+1}, \omega_{k+2}) = (P, P, F)\Big\}, \\ \text{ si l'ensemble } \Big\{k \in \mathbb{N}^* \mid (\omega_k, \omega_{k+1}, \omega_{k+2}) = (P, P, F)\Big\} \text{ n'est pas vide} \\ +\infty, \text{ sinon} \end{array} \right. \end{array} \right. .$$

Dans cette définition, on décide de dire qu'il faut une infinité de lancers pour avoir « P–P–F » pour la première fois si cette séquence ne se produit jamais. Par exemple :

$$X\Big((F)_{n\in\mathbb{N}^*}\Big) = +\infty,$$

indiquant que si l'on a obtenu à chaque fois « face » pour tous les lancers, alors on n'a jamais la bonne séquence. Dans cette situation, on peut alors se poser la question : ne serait-ce pas plutôt notre pièce qui a un problème ? En effet, si l'on a affaire à une pièce normalement constituée, cette pièce donnera « pile » ou « face » avec autant de chances... On prend un peu d'avance sur le concept de probabilités...

• Le dernier point concerne un tirage de boules avec trois protocoles de pioche différents.
Le plus simple est de différencier toutes les boules. On numérote les boules de 1 à n et on dit que l'urne contient r boules blanches. Quitte à reconsidérer notre numérotation, on peut considérer que les boules blanches sont les boules numérotées de 1 à r. On associera à chaque boule un entier entre 1 et n.
On passe en revue les trois protocoles expérimentaux.

▷ Lorsque les tirages sont avec remise, on peut tirer trente-six fois la même boule.

Chaque expérience avec remise est donc un p-uplet à composantes dans $[\![1, n]\!]$. Ainsi, chaque expérience ω est un p-uplet :

$$\omega = (\omega_i)_{1\leqslant i\leqslant p}$$

L'univers des possibles est donc dans cette situation :

$$\Omega = [\![1, n]\!]^p.$$

Cet ensemble Ω comporte n^p éléments.
La mesure faite sur ces tirages est associée à la variable aléatoire :

$$X : \left| \begin{array}{ccl} \Omega & \longrightarrow & \mathbb{N} \\ \omega = (\omega_i)_{1\leqslant i\leqslant p} & \longmapsto & \mathrm{Card}\Big(\Big\{i \in [\![1, p]\!] \mid \omega_i \in [\![1, r]\!]\Big\}\Big) \end{array} \right. .$$

▷ Lorsque les tirages sont sans remise, on ne peut pas tirer deux fois la même boule.

Chaque expérience sans remise est donc un p-uplet à composantes toutes différentes dans $[\![1, n]\!]$. Ainsi, chaque expérience ω est un p-uplet :

$$\omega = (\omega_i)_{1\leqslant i\leqslant p}, \text{ avec } \mathrm{Card}\Big(\Big\{\omega_i\ ;\ 1 \leqslant i \leqslant p\Big\}\Big) = p.$$

L'univers des possibles est donc dans cette situation :

$$\Omega = \left\{ (\omega_i)_{1 \leqslant i \leqslant p},\ \text{avec Card}\Big(\big\{\omega_i\ ;\ 1 \leqslant i \leqslant p\big\}\Big) = p \right\}.$$

Cet ensemble Ω comporte $\displaystyle\prod_{k=0}^{p-1}(n-k) = \frac{n!}{(n-p)!}$ éléments.
La mesure faite sur ces tirages est associée à la variable aléatoire :

$$X : \left| \begin{array}{ccl} \Omega & \longrightarrow & \mathbb{N} \\ \omega = (\omega_i)_{1 \leqslant i \leqslant n} & \longmapsto & \text{Card}\Big(\big\{ i \in [\![1,p]\!] \mid \omega_i \in [\![1,r]\!] \big\}\Big) \end{array} \right. .$$

▷ Lorsque les tirages sont simultanés, on a l'impression que les choses vont se passer de la même façon que pour les tirages sans remise. La seule différence est que dans les tirages sans remise, l'ordre d'apparition des boules intervient alors que dans les tirages simultanés, l'ordre d'apparition des boules n'intervient pas puisqu'elles sont piochées en même temps. On va donc devoir modéliser cette expérience avec un autre type d'objet : les ensembles.
Pour les tirages simultanés, chaque expérience ω est une partie A à p éléments dans l'ensemble $[\![1,n]\!]$. Les éléments de l'ensemble A correspondent à des indices associés aux boules piochées simultanément. On a ainsi gommé l'ordre d'apparition des boules.
De ce fait, l'univers Ω des possibles est l'ensemble :

$$\Omega = \mathscr{P}_p([\![1,n]\!]) = \Big\{ A \subset [\![1,n]\!] \mid \text{Card}(A) = p \Big\}$$

des parties à p éléments de l'ensemble $[\![1,n]\!]$.
Cet ensemble Ω comporte $\displaystyle\binom{n}{p}$ éléments.
La mesure faite sur ces tirages est associée à la variable aléatoire :

$$X : \left| \begin{array}{ccl} \Omega & \longrightarrow & \mathbb{N} \\ \omega = A & \longmapsto & \text{Card}\Big(A \cap [\![1,r]\!]\Big) \end{array} \right. .$$

Au programme de MPSI, on ne s'intéresse qu'à des expériences pour lesquelles l'univers Ω est un ensemble fini. Vous verrez en deuxième année – en fait, un peu à la fin de ce chapitre – les situations où l'univers Ω est dénombrable, c'est-à-dire lorsqu'il existe une bijection $\varphi : \mathbb{N} \longrightarrow \Omega$.
Dans la suite, on ne considère pour l'instant que des situations où l'ensemble Ω est fini. Par exemple, si l'on reprend les points de l'exemple 1, on peut conserver les points numéros 1 et 4 avec les trois types de tirages. Les points numéros 2 et 3 de l'exemple 1 correspondent à un univers Ω tantôt égal à $\mathbb{R}^2$ ou $\{P,F\}^{\mathbb{N}^*}$, mais ces deux univers infinis ne sont pas dénombrables : ils contiennent strictement plus d'éléments que l'ensemble $\mathbb{N}$, mais cela sera vu en deuxième année.

Si on a affaire à un protocole expérimental, chaque expérience est notée ω, l'univers Ω correspond à tous les résultats possibles des expériences. On appelle événement (avec deux accents aigus) toute partie de Ω.

Un événement élémentaire est un singleton. L'événement contraire à un événement A est le complémentaire de A dans Ω : l'événement $\Omega \setminus A$.
Deux événements A et B sont dits incompatibles s'ils sont disjoints.
Lorsque l'on regarde la définition 2, on s'aperçoit que l'on définit des choses déjà connues, avec un vocabulaire différent : d'une part, il y a le vocabulaire du probabiliste, qui utilisera les mots « événement élémentaire », « incompabilité entre deux événements », « on a lancé deux dés à six faces et on a fait un double » et d'autre part, il y a le vocabulaire de l'ensembliste, qui utilisera plutôt les mots « singleton », « disjonction d'ensembles », « on considère la partie $A = \Big\{(i,i) \mid i \in [\![1,6]\!]\Big\}$ ».
On voit donc que l'on pourra soit utiliser les notions du premier chapitre sur les ensembles pour manipuler les événements (version ensembliste), soit utiliser des phrases pour les manipuler (version probabiliste).
Par exemple, considérons le protocole suivant : on lance une pièce de monnaie quatre fois de suite. On peut s'intéresser aux événements suivants selon le point du vue probabiliste ou bien selon le point de vue ensembliste. Il s'agira de deux formulations différentes pour désigner la même chose. Vous constaterez que le point du vue du probabiliste est peut-être plus agréable à lire et à comprendre. Voyons cela sur quelques événements possibles associés à cette expérience de lancer quatre fois de suite une pièce de monnaie. L'univers ici considéré est :

$$\Omega = \{P, F\}^4,$$

ensemble à $2^4 = 16$ éléments.

- ▷ le probabiliste pourra considérer l'événement : A : « on a obtenu des résultats identiques à chaque lancer », alors que l'ensembliste parlera plutôt de l'ensemble :

$$A = \Big\{(P,P,P,P),(F,F,F,F)\Big\}.$$

- ▷ l'ensembliste pourra considérer l'événement :

$$B = \Big\{\omega = (\omega_i)_{1 \leqslant i \leqslant 4} \in \Omega \mid \forall i \in [\![1,3]\!],\ \omega_i \neq \omega_{i+1}\Big\}$$

  alors que le probabiliste utilisera plutôt cette façon de décrire les choses : B : « on n'a jamais obtenu deux résultats consécutifs égaux au cours des quatre lancers. »
- ▷ les formulations

$$C = \Big\{\omega = (\omega_i)_{1 \leqslant i \leqslant 4} \mid \omega_2 = P\Big\} \text{ et } D = \Big\{\omega = (\omega_i)_{1 \leqslant i \leqslant 4} \mid \exists i \in [\![1,4]\!],\ \omega_i = F\Big\}$$

  et C : « le deuxième lancer a donné « pile » », puis D : « on a obtenu au moins un « face » » comportent la même information.

Vous pouvez faire le test par vous-même. Prenez une personne λ. Exposez-lui le même événement en version probabiliste et en version ensembliste puis observez la réaction de cette personne. Si cette personne n'y connaît pas grand chose en maths, elle risque de faire une drôle de tête lorsque vous allez lui exposer la version ensembliste alors que les choses vont normalement plutôt bien se passer pour la version probabiliste. Pourtant, c'est la même information. La version probabiliste a certainement un confort de compréhension, mais formellement en parlant d'événements, on ne parle que d'ensembles.

On verra dans ce chapitre que les deux points de vue sont utiles. On se doute que si l'on veut faire des calculs un peu poussés, on aura intérêt à formaliser les choses et donc à privilégier la version ensembliste. Pour décrire un événement relativement simple, on a tout intérêt à privilégier la version probabiliste, mais tout cela reste une affaire de goût.

Parler d'expérience, d'univers des possibles ou bien d'événements n'est pas suffisant. On a besoin de savoir si un événement se produit « souvent » ou pas. C'est le concept de probabilité. Ce qu'il faut tout de suite comprendre est qu'une probabilité n'est pas un nombre : c'est une fonction $\mathbb{P} : \mathscr{P}(\Omega) \longrightarrow [0,1]$ définie sur l'ensemble des parties de Ω, autrement dit définie sur l'ensemble de tous les événements, cette fonction $\mathbb{P}$ vérifiant les deux conditions exposées en définition 3, à savoir le fait que $\mathbb{P}(\Omega) = 1$ et la σ-additivité : pour tous événements disjoints A et B,

$$\mathbb{P}(A \sqcup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

La probabilité $\mathbb{P}(A)$ d'un événement A va modéliser ses chances de se réaliser.
L'événement Ω est l'événement certain de probabilité 1.
L'événement $\emptyset$ est l'événement impossible, dont on ne sait pas encore que sa probabilité est nulle. On verra cela en propriété 1.
Il peut y avoir une différence entre « être vide » et « être négligeable ». Par exemple, si $\Omega = [0,1]$ et que l'on pioche au hasard un nombre réel entre 0 et 1, l'événement $\left\{\frac{1}{2}\right\}$ qui est l'événement « piocher le nombre $\frac{1}{2}$ » n'est pas l'événement impossible car il ne s'agit pas de l'ensemble vide, et pourtant on a l'impression que la probabilité que cela arrive est égal à 0. Tout cela est un peu subtil...

La donnée $(\Omega, \mathbb{P})$ d'un couple d'un univers Ω et d'une probabilité $\mathbb{P} : \mathscr{P}(\Omega) \longrightarrow [0,1]$ – attention, une probabilité $\mathbb{P}$ n'est pas définie sur Ω mais bien sur $\mathscr{P}(\Omega)$ et l'ensemble Ω est un élément du domaine de définition de la fonction $\mathbb{P}$, donc on peut très bien calculer $\mathbb{P}(\Omega)$ qui vaut 1 par définition de la probabilité – est appelé espace probabilisé. Ici, il ne faut pas confondre espace avec espace vectoriel. L'ensemble Ω n'est pas du tout un espace vectoriel. Dire que $(\Omega, \mathbb{P})$ est un espace probabilisé fini signifie que l'on considère une situation aléatoire d'univers fini Ω et que l'on considère également une probabilité $\mathbb{P} : \mathscr{P}(\Omega) \longrightarrow [0,1]$ sur l'ensemble de tous les événements.

Après toutes ces descriptions et ces exemples, on va commencer à formaliser un peu les choses et acquérir des résultats utiles pour la suite.

On commence par la propriété 1 que l'on démontre point par point. Par hypothèse, on dispose donc d'une probabilité $\mathbb{P} : \mathscr{P}(\Omega) \longrightarrow [0,1]$ associée à un univers fini Ω. On va très fortement utiliser la σ-additivité de la fonction $\mathbb{P}$ dans la suite.

• Soit $A \in \mathscr{P}(\Omega)$ un événement. Les événements A et $\overline{A} = \Omega \setminus A$ sont incompatibles, donc :

$$\begin{aligned} 1 &= \mathbb{P}(\Omega) \\ &= \mathbb{P}(A \sqcup \overline{A}) \\ &= \mathbb{P}(A) + \mathbb{P}(\overline{A}) \qquad [\sigma\text{-additivité}] \end{aligned}$$

donc $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.

Appliquons cette formule à l'événement $A = \Omega$. Alors, $\overline{A} = \overline{\Omega} = \Omega \setminus \Omega = \emptyset$, puis :

$$\mathbb{P}(\emptyset) = \mathbb{P}(\overline{\Omega}) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0.$$

• Soient $A \subset B$ deux événements. On pose $C = B \setminus A$. Il est clair que les événements A et C sont incompatibles et de plus :

$$B = A \sqcup C.$$

Par σ-additivité :

$$\mathbb{P}(B) = \mathbb{P}(A \sqcup C) = \mathbb{P}(A) + \mathbb{P}(C) \geqslant \mathbb{P}(A),$$

car $\mathbb{P}(C) \geqslant 0$.
En particulier, lorsque $A \subset B$, on peut écrire :

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A),$$

formule qui ne marche pas du tout lorsque l'événement A n'est pas inclus dans B.

• On démontre par récurrence la propriété :
$\mathscr{P}(r)$: « si $A_1, \cdots, A_r$ sont r événements deux à deux incompatibles, alors :

$$\mathbb{P}\left(\bigsqcup_{i=1}^{r} A_i\right) = \sum_{i=1}^{r} \mathbb{P}(A_i). \text{ »}$$

- $\longrightarrow$ la propriété $\mathscr{P}(1)$ est évidente ;
- $\longrightarrow$ supposons la propriété $\mathscr{P}(r)$ vérifiée ;
- $\longrightarrow$ soient $A_1, \cdots, A_r, A_{r+1}$ des événements deux à deux incompatibles. On pose :

$$B = \bigsqcup_{i=1}^{r} A_i,$$

de sorte que les deux événements A_{r+1} et B sont incompatibles.
On termine comme suit :

$$\begin{aligned} \mathbb{P}\left(\bigsqcup_{i=1}^{r+1} A_i\right) &= \mathbb{P}\Big(B \sqcup A_{r+1}\Big) \\ &= \mathbb{P}(B) + \mathbb{P}(A_{r+1}) \qquad [\sigma\text{-additivité}] \\ &= \sum_{i=1}^{r} \mathbb{P}(A_i) + \mathbb{P}(A_{r+1}) \qquad [\text{H.R.}] \\ &= \sum_{i=1}^{r+1} \mathbb{P}(A_i). \end{aligned}$$

• Si A et B sont deux événements quelconques, pas forcément disjoints, on pose $C_1 = A \setminus B$, $C_2 = A \cap B$ et $C_3 = B \setminus A$, de sorte que :

$$A \cup B = C_1 \sqcup C_2 \sqcup C_3.$$

On en déduit :

$$\begin{array}{rcll}\mathbb{P}(A \cup B) & = & \mathbb{P}(C_1) + \mathbb{P}(C_2) + \mathbb{P}(C_3) & \\ & = & \mathbb{P}(A \setminus (A \cap B)) + \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus (A \cap B)) & , \text{car } A \setminus B = A \setminus (A \cap B) \\ & = & \Big(\mathbb{P}(A) - \mathbb{P}(A \cap B)\Big) + \mathbb{P}(A \cap B) + \Big(\mathbb{P}(B) - \mathbb{P}(A \cap B)\Big) & , \text{car } \left\{ \begin{array}{l} A \cap B \subset A \\ \text{et} \\ A \cap B \subset B \end{array} \right. \\ & = & \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). & \end{array}$$

On voit que cette formule semble généraliser la formule que l'on connaissait sur les cardinaux finis. D'ailleurs, avec les points précédents, on décelait, il me semble, des similarités avec les cardinaux finis. On expliquera cela bientôt.

• Ce point est très important pour construire des probabilités lorsque l'univers Ω est fini. Voyons en détail cette construction.
On considère un univers fini :

$$\Omega = \Big\{\omega_1, \cdots, \omega_s\Big\}$$

On considère un s-uplet de nombres réels $(p_1, \cdots, p_s)$ tous positifs et de somme 1 :

$$\forall k \in [\![1, s]\!],\ p_k \geqslant 1 \text{ et } \sum_{k=1}^{s} p_k = 1.$$

On considère maintenant l'application suivante :

$$\mathbb{Q} : \left| \begin{array}{ccc} \mathscr{P}(\Omega) & \longrightarrow & [0,1] \\ A & \longmapsto & \displaystyle\sum_{k=1 \text{ et } \omega_k \in A}^{n} p_k \end{array} \right. .$$

Premièrement, l'application $\mathbb{Q}$ est bien définie puisque si A est dans $\mathscr{P}(\Omega)$, alors la somme $\mathbb{Q}(A)$ est constituée de réels positifs, donc la somme est positive. De plus, la somme est inférieure à la somme où tous les p_k sont pris :

$$\mathbb{Q}(A) \leqslant \sum_{k=1}^{s} p_k = 1.$$

Ainsi, on a toujours l'appartenance $\mathbb{Q}(A) \in [0, 1]$ ce qui est en soit une bonne nouvelle.
Ensuite,

$$\mathbb{Q}(\Omega) = \sum_{k=1}^{s} p_k = 1,$$

par hypothèse sur les réels p_k choisis.
Enfin, on teste la σ-additivité de la fonction $\mathbb{Q}$. Soient A et B deux événements incompatibles. Alors :

$$\mathbb{Q}(A \sqcup B) \quad = \quad \sum_{k=1 \text{ et } \omega \in A \sqcup B}^{s} p_k$$

$$\begin{aligned} &= \sum_{k=1 \text{ et } \omega\in A}^{s} p_k + \sum_{k=1 \text{ et } \omega\in B}^{s} p_k, \text{ car la réunion est disjointe} \\ &= \mathbb{Q}(A) + \mathbb{Q}(B). \end{aligned}$$

Résultat des courses : à l'aide d'une famille $(p_1, \cdots, p_s)$ de $s = \text{Card}(\Omega)$ réels positifs de somme 1, on peut construire assez facilement une probabilité.
Question subsidiaire : existe-t-il des probabilités $\mathbb{P} : \mathscr{P}(\Omega) \longrightarrow [0,1]$ qui ne sont pas de cette forme ? Autrement dit, y a-t-il des probabilités dont l'expression est un peu plus tordue que la formule pour la probabilité $\mathbb{Q}$ ci-dessus ?
La réponse à cette question est : non, ce qui est une très bonne nouvelle. On connaît donc la tête de toutes les probabilités associées à un univers fini. Mais ne vendons pas la peau de l'ours avant de l'avoir tué. Il s'agit quand même de montrer la réponse à la question subsidiaire énoncée plus haut.
Soit donc $\mathbb{P} : \mathscr{P}(\Omega) \longrightarrow [0,1]$ une probabilité, où l'univers $\Omega = \{\omega_1, \cdots, \omega_s\}$ est fini de cardinal s par exemple.
On pose pour tout $k \in [\![1, s]\!]$, le nombre :

$$\pi_k = \mathbb{P}(\{\omega_k\}).$$

En tant que probabilité, le réel π_k est dans $[0,1]$. De plus, comme :

$$\Omega = \bigsqcup_{k=1}^{s} \{\omega_k\},$$

par σ-additivité, on peut écrire :

$$\begin{aligned} 1 = \mathbb{P}(\Omega) &= \mathbb{P}\left(\bigsqcup_{k=1}^{s} \{\omega_k\}\right) \\ &= \sum_{k=1}^{s} \mathbb{P}(\{\omega_k\}) \\ &= \sum_{k=1}^{s} \pi_k. \end{aligned}$$

La famille $(\pi_k)_{1\leqslant k\leqslant s}$ est une famille de réels positifs et de somme égale à 1.
De plus, si A est un événement, on peut écrire :

$$A = \bigsqcup_{k=1 \text{ et } \omega_k\in A}^{s} \{\omega_k\},$$

donc par σ-additivité :

$$\begin{aligned} \mathbb{P}(A) &= \sum_{k=1 \text{ et } \omega_k\in A}^{s} \mathbb{P}(\{\omega_k\}) \\ &= \sum_{k=1 \text{ et } \omega_k\in A}^{s} \pi_k. \end{aligned}$$

Conclusion, la probabilité $\mathbb{P}$ est l'application :

$$\mathbb{P} : \left| \begin{array}{ccc} \mathscr{P}(\Omega) & \longrightarrow & [0,1] \\ A & \longmapsto & \displaystyle\sum_{k=1 \text{ et } \omega_k \in A}^{s} \pi_k \end{array} \right. .$$

La probabilité $\mathbb{P}$ est donc de la forme voulue associée à une famille de s réels positifs de somme 1.
Il y a donc autant de probabilités lorsque l'univers Ω est de cardinal s que de s-uplets $(a_1, \cdots, a_s)$ de réels positifs de somme 1, c'est-à-dire un sacré paquet : autant que de nombres réels.

On l'a compris, dès que l'on a un ensemble Ω à s éléments, dès que l'on a une famille $(p_\omega)_{\omega\in\Omega}$ de réels positifs et de somme égale à 1, on dispose d'une probabilité $\mathbb{P} : A \longmapsto \displaystyle\sum_{\omega\in A} p_\omega$ et toutes les probabilités $\mathbb{Q} : \mathscr{P}(\Omega) \longrightarrow \mathbb{R}$ sont entièrement déterminées par les images des singletons $\{\omega\}$. Plus précisément, la probabilité $\mathbb{Q}$ sera associée à la famille $\Big(\pi_\omega = \mathbb{Q}(\{\omega\})\Big)_{\omega\in\Omega}$. Il suffit donc de connaître une probabilité sur les singletons. En disant cela, je dis trente-six fois la même chose, mais c'est important.

Appliquons ceci à des probabilités « de référence ».
$\longrightarrow$ La première probabilité de référence est la mesure de Dirac.
Donnons-nous un ensemble Ω fini (uniquement pour être dans le programme, mais ce n'est pas très important ici que l'ensemble Ω soit fini).
Fixons un élément ω dans Ω. La famille $(\delta_{\alpha,\omega})_{\alpha\in\Omega}$ [où $\delta_{\alpha,\omega}$ est le symbole de Kronecker ; on rappelle que $\delta_{\alpha,\omega}$ vaut 1 si et seulement si $\alpha = \omega$ et sinon, le symbole de Kronecker vaut 0] est une famille de réels positifs de somme égale à 1. Cette famille est associée à une probabilité : la mesure de Dirac en ω. Il s'agit de la probabilité suivante :

$$\delta_\omega : \left| \begin{array}{ccc} \mathscr{P}(\Omega) & \longrightarrow & [0,1] \\ A & \longmapsto & \mathbf{1}_A(\omega) = \begin{cases} 1, \text{ si } \omega \in A \\ 0, \text{ si } \omega \notin A \end{cases} \end{array} \right. .$$

Si Ω est un ensemble fini de cardinal s, si $\mathbb{P} : \mathscr{P}(\Omega) \longrightarrow [0,1]$ est une probabilité, en notant pour tout $\omega \in \Omega$:

$$p_\omega = \mathbb{P}(\{\omega\}),$$

alors on peut écrire pour tout événement $A \in \mathscr{P}(\Omega)$:

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}\left(\bigsqcup_{\omega\in A} \{\omega\}\right) \\ &= \sum_{\omega\in A} \mathbb{P}(\{\omega\}) \qquad [\sigma\text{-additivité}] \\ &= \sum_{\omega\in A} p_\omega \\ &= \sum_{\omega\in A} p_\omega \times 1 + \sum_{\omega\in\Omega\setminus A} p_\omega \times 0 \end{aligned}$$

$$\begin{aligned} &= \sum_{\omega \in A} p_\omega \times \delta_\omega(A) + \sum_{\omega \in \Omega \setminus A} p_\omega \times \delta_\omega(A) \\ &= \sum_{\omega \in \Omega} p_\omega \cdot \delta_\omega(A). \end{aligned}$$

Par conséquent,

$$\mathbb{P} = \sum_{\omega \in \Omega} p_\omega \cdot \delta_\omega.$$

$\longrightarrow$ On peut proposer en applications numériques :

- $\triangleright$ lorsque $\Omega = [\![1, n]\!]$, avec $n \geqslant 1$ et la famille $\left(p_\omega = \dfrac{1}{n}\right)_{\omega \in \Omega}$ de réels positifs de somme 1, on obtient la loi uniforme et notée $\mathscr{U}_n$ ainsi définie :

$$\mathscr{U}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_i : \left| \begin{array}{ccc} \mathscr{P}([\![1, n]\!]) & \longrightarrow & [0, 1] \\ A & \longmapsto & \displaystyle\sum_{i \in A} \frac{1}{n} = \frac{\operatorname{Card}(A)}{n} \end{array} \right. .$$

  Cette loi uniforme explique pourquoi les formules de la propriété 1 généralisent les formules sur les cardinaux finis.

- $\triangleright$ Lorsque l'univers Ω est un ensemble à 2 éléments $\{0, 1\}$, en fixant un réel p entre 0 et 1, alors la famille $(1 - p, p)$ est un couple de réels positifs de somme 1 et on associe à cette famille la probabilité qui est la loi de Bernoulli de paramètre p et notée $\mathscr{B}(p)$ ainsi définie :

$$\mathscr{B}(p) = (1-p) \cdot \delta_0 + p \cdot \delta_1 : \left| \begin{array}{ccc} \mathscr{P}(\{0, 1\}) & \longrightarrow & [0, 1] \\ \emptyset & \longmapsto & 0 \\ \{0\} & \longmapsto & 1 - p \\ \{1\} & \longmapsto & p \\ \{0, 1\} & \longmapsto & 1 \end{array} \right. .$$

- $\triangleright$ Lorsque n est un entier naturel, en prenant l'univers $\Omega = [\![0, n]\!]$, puis la famille

$$\left(\pi_k = \binom{n}{k} p^k \, (1-p)^{n-k}\right)_{0 \leqslant k \leqslant n},$$

  alors chaque réel de cette famille est positif et par le binôme de Newton dans l'anneau commutatif $\mathbb{R}$, la somme vaut :

$$\sum_{k=0}^{n} \pi_k = \sum_{k=0}^{n} \binom{n}{k} p^k \, (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1.$$

  On associe à cette famille la probabilité appelée loi binomiale de paramètres n et p et notée $\mathscr{B}(n, p)$ ainsi définie :

$$\mathscr{B}(n, p) = \sum_{k=0}^{n} \pi_k \cdot \delta_k : \left| \begin{array}{ccc} \mathscr{P}([\![0, n]\!]) & \longrightarrow & [0, 1] \\ A & \longmapsto & \displaystyle\sum_{k \in A} \binom{n}{k} p^k \, (1-p)^{n-k} \end{array} \right. .$$

▷ Pour terminer ce catalogue non exhaustif, on peut se placer dans le cas où l'univers n'est composé que de deux éléments $\Omega = \{-1, 1\}$. On considère un peu comme pour la loi de Bernoulli, un paramètre $p \in [0,1]$ et on considère la famille à deux éléments $(1-p, p)$, qui est une famille de réels positifs et de somme égale à 1. On définit ainsi la loi de Rademacher et notée $\mathscr{R}(p)$ [prononcer à l'« allemande »] qui est la probabilité :

$$\mathscr{R}(p) = p \cdot \delta_1 + (1-p) \cdot \delta_{-1} : \left| \begin{array}{ccc} \mathscr{P}(\{-1,1\}) & \longrightarrow & [0,1] \\ \emptyset & \longmapsto & 0 \\ \{-1\} & \longmapsto & 1-p \\ \{1\} & \longmapsto & p \\ \{-1,1\} & \longmapsto & 1 \end{array} \right. .$$

Les probabilités conditionnelles et l'indépendance

Considérons un espace probabilisé fini $(\Omega, \mathbb{P})$. Autrement dit, l'ensemble Ω est fini et l'application $\mathbb{P} : \mathscr{P}(\Omega) \longrightarrow [0,1]$ est une probabilité.
Considérons un événement B non négligeable, c'est-à-dire de probabilité strictement positive. Dans certaines situations, on souhaite calculer la probabilité d'événements A, sachant que l'événement B est réalisé. Dans cette configuration, il est plus précis non pas de calculer $\mathbb{P}(A)$ en laissant de côté l'information selon laquelle on sait que B se réalise, mais plutôt de calculer la proportion $\dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ qui indique la part d'apparition de $A \cap B$ dans l'événement B, puisque l'événement B se réalise.
Formalisons un peu les choses.

La définition 5 met en place une fonction $\mathbb{P}_B : \mathscr{P}(\Omega) \longrightarrow [0,1]$. Montrons que cette application est déjà bien définie, puis qu'il s'agit d'une probabilité.
En effet, si A est un événement appartenant donc à l'ensemble $\mathscr{P}(\Omega)$, alors comme on a l'inclusion $A \cap B \subset B$, alors on sait depuis la propriété 1 que :

$$\mathbb{P}(A \cap B) \subset \mathbb{P}(B).$$

Tout est positif et en divisant par $\mathbb{P}(B) > 0$, on obtient :

$$\mathbb{P}_B(A) \in [0,1].$$

Ensuite, il reste à montrer que l'application $\mathbb{P}_B$ vérifie les deux conditions pour avoir une probabilité.

- Premièrement,
$$\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$
- Deuxièmement, si A_1 et A_2 sont deux événements incompatibles, alors il en est de même des événements $A_1 \cap B$ et $A_2 \cap B$.
On utilise maintenant la σ-additivité de la probabilité $\mathbb{P}$, pour pouvoir écrire :
$$\mathbb{P}_B(A_1 \sqcup A_2) \quad = \quad \frac{\mathbb{P}((A_1 \sqcup A_2) \cap B)}{\mathbb{P}(B)}$$

$$
\begin{aligned}
&= \frac{\mathbb{P}((A_1 \cap B) \sqcup (A_2 \cap B))}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(A_1 \cap B)}{\mathbb{P}(B)} + \frac{\mathbb{P}(A_2 \cap B)}{\mathbb{P}(B)} \\
&= \mathbb{P}_B(A_1) + \mathbb{P}_B(A_2).
\end{aligned}
$$

L'application $\mathbb{P}_B$ est bien une probabilité. On la nomme probabilité conditionnelle sachant B. Il existe deux notations différentes, à savoir :

$$\mathbb{P}_B(A) = \mathbb{P}(A \mid B) \qquad \text{[lire « probabilité de } A \text{ sachant } B\text{ »]}$$

Personnellement, j'utiliserai quasi-systématiquement la notation $\mathbb{P}(A \mid B)$ dont la lecture est plus confortable. Le problème avec l'autre notation est qu'à la longue, on peut confondre A et B et ne plus savoir quel événement doit être mis en indice.

La définition 6 est également très importante. Un système complet d'événements [en abrégé S.C.E] est une famille finie $(A_1, \cdots, A_r)$ d'événements qui forment un découpage de l'ensemble Ω. La seule différence avec une partition est que dans une partition, chaque morceau doit être non vide, alors que dans un S.C.E, on autorise aux événements A_i d'être vides. Les deux choses importantes à vérifier pour avoir un SCE est que les événements soient deux à deux incompatibles [pas de débordements entre événements] et que la réunion soit totale [pour tout ω dans Ω, l'expérience ω appartient à au moins un événement A_i [version ensembliste] ; on dit aussi [version probabiliste] que pour toute expérience ω, l'un au moins des événements A_i se réalise.

La première formule à retenir est la formule des probabilités composées. Elle permet d'exprimer la probabilité d'une intersection d'événements.
Démontrons la propriété 2.
Sous les hypothèses, on lit dans la formule à démontrer une probabilité conditionnelle sachant l'événement $A_1 \cap \cdots \cap A_k$. Il faut absolument que la probabilité de cet événement soit strictement positive pour pouvoir parler de cette probabilité conditionnelle. Or, si l'entier k est fixé entre 1 et $p-1$, on peut écrire :

$$\bigcap_{i=1}^{p-1} A_i \subset \bigcap_{i=1}^{k} A_i,$$

donc en prenant les probabilités :

$$\mathbb{P}\left(\bigcap_{i=1}^{p-1} A_i\right) \leqslant \mathbb{P}\left(\bigcap_{i=1}^{k} A_i\right),$$

et par hypothèse, la probabilité de gauche est strictement positive : celle de droite aussi.
Finalement, le produit dans la formule fait apparaître un produit télescopique. On détaille la simplification :

$$\mathbb{P}(A_1) \cdot \prod_{k=1}^{p-1} \mathbb{P}(A_{k+1} \mid A_1 \cap \cdots \cap A_k) \quad = \quad \mathbb{P}(A_1) \cdot \prod_{k=1}^{p-1} \frac{\mathbb{P}(A_1 \cap \cdots \cap A_k \cap A_{k+1})}{\mathbb{P}(A_1 \cap \cdots \cap A_k)}$$

$$
\begin{aligned}
&= \mathbb{P}(A_1) \cdot \prod_{k=1}^{p-1} \frac{\mathbb{P}(A_1 \cap \cdots \cap A_{k+1})}{\mathbb{P}(A_1 \cap \cdots \cap A_k)} \\
&= \mathbb{P}(A_1) \times \frac{\mathbb{P}(A_1 \cap \cdots \cap A_p)}{\mathbb{P}(A_1)} \\
&= \mathbb{P}(A_1 \cap \cdots \cap A_p).
\end{aligned}
$$

Avant d'engager des exemples, on démontre la formule des probabilités totales, qui sert un peu plus souvent que la formule précédente. Ne confondez pas la dénomination de ces deux formules : elles n'ont pas le même usage.
Pour la formule des probabilités totales, l'ingrédient principal est un système complet d'événements. À chaque fois que vous devrez utiliser cette formule des probabilités totales, il faudra toujours expliciter le SCE que vous considérez.
Donnons-nous n'importe quel SCE $(A_1, \cdots, A_r)$ et n'importe quel événement B. Alors, on peut écrire :

$$
\begin{aligned}
\mathbb{P}(B) &= \mathbb{P}(B \cap \Omega) \\
&= \mathbb{P}\left(B \cap \bigsqcup_{i=1}^{r} A_i\right) \\
&= \mathbb{P}\left(\bigsqcup_{i=1}^{r} (A_i \cap B)\right) \qquad \text{[lois de Morgan : chap 1.]} \\
&= \sum_{i=1}^{r} \mathbb{P}(A_i \cap B) \qquad [\sigma\text{-additivité}]
\end{aligned}
$$

La formule de Bayes est très simple si l'on omet la formule monstrueuse finale, surtout qui n'est pas à retenir mais à savoir retrouver.
Sous les hypothèses, on écrit :

$$
\begin{aligned}
\mathbb{P}(A_k \mid B) &= \frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(B \cap A_k)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(B \mid A_k) \cdot \mathbb{P}(A_k)}{\mathbb{P}(B)}.
\end{aligned}
$$

Pour avoir la dernière formule – je répète surtout à ne pas retenir – il suffit d'utiliser la formule des probabilités totales exprimant $\mathbb{P}(B)$ à l'aide du SCE $(A_i)_{1 \leqslant i \leqslant r}$.
La formule de Bayes est surtout utile pour renverser l'expression d'une probabilité conditionnelle sachant un événement en fonction d'une probabilité conditionnelle sachant un autre événement.

Mettons tout cela en pratique à travers l'exemple 3.

On met en place certaines notations en guise de modélisation :

- l'univers Ω correspondant à l'ensemble des personnes assurées ; la probabilité $\mathbb{P}$ n'est pas explicite mais on connaît la probabilité de certaines événements
- événement C_1 : « la personne a moins de 25 ans », événement C_2 : « la personne a entre 25 et 50 ans » et l'événement C_3 : « la personne a plus de 50 ans ». Ainsi, $\mathbb{P}(C_1) = 0.2$, $\mathbb{P}(C_2) = 0.5$ et $\mathbb{P}(C_3) = 0.3$;
- événement A : « la personne a déclaré au moins un accident ». On connaît alors les probabilités conditionnelles ; $\mathbb{P}(A \mid C_1) = 0.1$, $\mathbb{P}(A \mid C_2) = 0.05$ et $\mathbb{P}(A \mid C_3) = 0.08$.

1. On doit calculer $\mathbb{P}(A)$. Pour ce faire, on va utiliser le SCE (C_1, C_2, C_3) et la formule des probabilités totales :

$$\begin{aligned} \mathbb{P}(A) &= \sum_{i=1}^{3} \mathbb{P}(A \mid C_i) \cdot \mathbb{P}(C_i) \\ &= 0.1 \times 0.2 + 0.05 \times 0.5 + 0.08 \times 0.3 \\ &= 0.069 = 6.9\ \%. \end{aligned}$$

2. On doit calculer $\mathbb{P}(C_1 \mid A)$ en utilisant la formule de Bayes :

$$\begin{aligned} \mathbb{P}(C_1 \mid A) &= \frac{\mathbb{P}(A \cap C_1)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A \mid C_1) \cdot \mathbb{P}(C_1)}{\mathbb{P}(A)} \\ &= \frac{0.2 \times 0.1}{0.069} \\ &\simeq 28.99\ \%. \end{aligned}$$

De la même façon,

$$\mathbb{P}(C_3 \mid A) = \frac{\mathbb{P}(A \mid C_3) \cdot \mathbb{P}(C_3)}{\mathbb{P}(A)} \simeq 34.78\ \%.$$

3. Dans cette question, on veut calculer $\mathbb{P}(A \mid C_2 \cup C_3)$.
On en déduit :

$$\begin{aligned} \mathbb{P}(A \mid C_2 \cup C_3) &= \frac{\mathbb{P}(A \cap (C_2 \cup C_3))}{\mathbb{P}(C_2 \cup C_3)} \\ &= \frac{\mathbb{P}(A \cap C_2) + \mathbb{P}(A \cap C_3)}{\mathbb{P}(C_2) + \mathbb{P}(C_3)} \\ &\simeq 6.1\ \%. \end{aligned}$$

4. On veut ici calculer $\mathbb{P}(\overline{C_2} \mid \overline{A})$.
Comme $\mathbb{P}(\cdot \mid \overline{A})$ est une probabilité, alors :

$$\mathbb{P}(\overline{C_2} \mid \overline{A}) = 1 - \mathbb{P}(C_2 \mid \overline{A}).$$

Or,

$$\begin{aligned}
\mathbb{P}(C_2 \mid \overline{A}) &= \frac{\mathbb{P}(\overline{A} \cap C_2)}{1-\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(\overline{A} \mid C_2)\cdot \mathbb{P}(C_2)}{1-\mathbb{P}(A)} \\
&= \frac{(1-\mathbb{P}(A \mid C_2))\cdot \mathbb{P}(C_2)}{1-\mathbb{P}(A)} \\
&\simeq 51.02\ \%.
\end{aligned}$$

Pour terminer, il faut calculer $\mathbb{P}(C_3 \mid \overline{A})$. On en déduit :

$$\begin{aligned}
\mathbb{P}(C_3 \mid \overline{A}) &= \frac{\mathbb{P}(C_3 \cap \overline{A})}{1-\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(\overline{A} \cap C_3)}{1-\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(\overline{A} \mid C_3)\cdot \mathbb{P}(C_3)}{1-\mathbb{P}(A)} \\
&= \frac{(1-\mathbb{P}(A \mid C_3))\ \cdot \mathbb{P}(C_3)}{1-\mathbb{P}(A)} \\
&\simeq 29.65\ \%.
\end{aligned}$$

On aborde maintenant l'indépendance entre événements.
Intuitivement, deux événements A et B sont indépendants si le fait de savoir que l'événement B est réalisé n'influe pas sur le calcul de la probabilité de l'événement A.
Lorsque B est de probabilité strictement positive, on obtient alors cette formulation intuitive de l'indépendance :

$$\mathbb{P}(A) = \mathbb{P}(A \mid B),$$

ce qui se réécrit :

$$\mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

ou encore :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).$$

L'avantage de cette formulation est qu'il n'est plus nécessaire d'avoir $\mathbb{P}(B)$ strictement positif.
Ceci explique la première définition de la définition 7 pour l'indépendance deux à deux.
On définit ensuite l'indépendance mutuelle entre plusieurs événements. La probabilité de l'intersection est égale au produit des probabilités, mais l'intersection étant uniquement prise sur les indices k dans une partie I, pour toute partie I incluse dans $[\![1, p]\!]$. L'erreur classique est de confondre la définition avec le fait que la probabilité de l'intersection sur tous les événements est égale au produit des probabilités. Il n'est pas suffisant de prendre $I = [\![1, p]\!]$.

On démontre la propriété 5.

• Soient $A_1, \cdots, A_p$ des événements mutuellement indépendants. Soient i et j deux indices différents entre 1 et p. En prenant la partie $I = \{i, j\}$, on peut écrire :

$$\begin{aligned}
\mathbb{P}(A_i \cap A_j) &= \mathbb{P}\left(\bigcap_{k \in I} A_k\right) \\
&= \prod_{k \in I} \mathbb{P}(A_k) \\
&= \mathbb{P}(A_i) \times \mathbb{P}(A_j).
\end{aligned}$$

Les événements A_i et A_j sont indépendants.
La réciproque est fausse. L'exemple 4 nous fournit un contre-exemple. Considérons l'univers $\Omega = \{1, 2, 3, 4\}$, la probabilité uniforme $\mathbb{P} = \mathscr{U}_4$, puis les événements $A = \{1, 2\}$, $B = \{1, 3\}$ et $C = \{2, 3\}$.
Comme pour tout événement F, on a :

$$\mathbb{P}(F) = \frac{\mathrm{Card}(F)}{4},$$

alors :

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2},$$

puis :

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(C \cap A) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2}$$

et les événements A, B et C sont bien deux à deux indépendants.
Cependant, les trois événements A, B et C ne sont pas mutuellement indépendants puisque :

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0,$$

alors que :

$$\mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C) = \left(\frac{1}{2}\right)^3 = \frac{1}{8} \neq \mathbb{P}(A \cap B \cap C).$$

• On va montrer le deuxième point en utilisant une récurrence que l'on ne va formaliser – les choses deviendront visibles au fur et à mesure.
Dans la suite, on dit qu' une famille de r événements $(C_1, \cdots, C_r)$ vérifie la propriété $\mathscr{P}$ si :

$$\mathbb{P}(C_1 \cap \cdots \cap \mathbb{C}_r) = \prod_{k=1}^{r} \mathbb{P}(C_k).$$

On va déjà montrer que si la famille $(A_1, \cdots, A_p)$ vérifie la propriété $\mathscr{P}$ et si la famille $(A_1, \cdots, A_{p-1})$ vérifie aussi la propriété $\mathscr{P}$, alors en remplaçant l'événement A_p par son événement contraire $\overline{A_p}$, la nouvelle famille obtenue $(B_1, \cdots, B_p)$ vérifie encore la propriété $\mathscr{P}$.
En effet, supposons que la famille $(A_1, \cdots, A_p)$ vérifie la propriété $\mathscr{P}$, ainsi que la famille $(A_1, \cdots, A_{p-1})$. On considère la famille $(B_1, \cdots, B_p)$ de p événements définis par :

$$\forall i \in [\![1, p-1]\!],\ B_i = A_i \text{ et } B_p = \overline{A_p}.$$

On remarque facilement que l'on a la réunion disjointe :

$$A_1 \cap \cdots \cap A_{p-1} = \Big(A_1 \cap \cdots \cap A_p \Big) \sqcup \Big(B_1 \cap \cdots \cap B_p \Big),$$

car $A_p \sqcup B_p = \Omega$ puis en prenant l'intersection $A_1 \cap \cdots \cap A_{p-1}$ dans cette égalité.
Par σ-additivité, on en déduit :

$$\begin{aligned}
\mathbb{P}(B_1 \cap \cdots \cap B_p) &= \mathbb{P}(A_1 \cap \cdots \cap A_{p-1}) - \mathbb{P}(A_1 \cap \cdots \cap A_p) \\
&= \prod_{k=1}^{p-1} \mathbb{P}(A_k) - \prod_{k=1}^{p} \mathbb{P}(A_k) \\
&= \prod_{k=1}^{p-1} \mathbb{P}(A_k) \times (1 - \mathbb{P}(A_p)) \\
&= \prod_{k=1}^{p-1} \mathbb{P}(A_k) \times \mathbb{P}(\overline{A_p}) \\
&= \prod_{k=1}^{p} \mathbb{P}(B_k).
\end{aligned}$$

On montre maintenant que si la famille $(A_1, \cdots, A_p)$ est mutuellement indépendante, en remplaçant l'événement A_p par son événement contraire, la nouvelle famille obtenue $(B_1, \cdots, B_p)$ reste mutuellement indépendante.
En effet, sous l'hypothèse de mutuelle indépendance de la famille $(A_1, \cdots, A_p)$, soit I une partie incluse dans $[\![1, p]\!]$.
$\longrightarrow$ Si la partie I ne contient pas l'élément p, alors :

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{k \in I} B_k \right) &= \mathbb{P}\left(\bigcap_{k \in I} A_k \right) \\
&= \prod_{k \in I} \mathbb{P}(A_k) \\
&= \prod_{k \in I} \mathbb{P}(B_k).
\end{aligned}$$

$\longrightarrow$ Si la partie I contient l'élément p, on pose $I = \{i_1, \cdots, i_r, p\}$, avec $i_1 < i_2 < \cdots < i_r < p$, l'entier r étant éventuellement nul si $I = \{p\}$. Dans ce cas, les familles $(A_{i_1}, \cdots, A_{i_r})$ et $(A_{i_1}, \cdots, A_{i_r}, A_p)$ vérifient la propriété $\mathscr{P}$ et par le début, on sait que la famille $(A_{i_1}, \cdots, A_{i_r}, \overline{A_p}) = (B_{i_1}, \cdots, B_{i_r}, B_p)$ vérifie également la propriété $\mathscr{P}$.
Résultat des courses : quoiqu'il arrive :

$$\mathbb{P}\left(\bigcap_{k \in I} B_k \right) = \prod_{k \in I} \mathbb{P}(B_k).$$

Ceci étant valable pour toute partie I incluse dans $[\![1, p]\!]$, alors la famille $(B_1, \cdots, B_p)$ reste mutuellement indépendante.

Ensuite, comme l'intersection d'ensembles et le produit de nombres réels sont des opérations commutatives, si la famille $(A_1, \cdots, A_p)$ est une famille d'événements mutuellement indépendants, alors pour toute permutation $\sigma \in \mathfrak{S}_p$, la famille $(A_{\sigma(1)}, \cdots, A_{\sigma(p)})$ reste mutuellement indépendante.
En utilisant des permutations, on voit alors facilement que si la famille $(A_1, \cdots, A_p)$ est une famille d'événements mutuellement indépendants, si l'on remplace l'un des événements A_i par son événement contraire $\overline{A_i}$ (et plus forcément le dernier A_p remplacé par $\overline{A_p}$), alors la nouvelle famille obtenue reste mutuellement indépendante.
Il est alors facile de voir maintenant que si l'on réitère cette opération de remplacer un événement de la famille initiale $(A_1, \cdots, A_p)$ mutuellement indépendante par son événement contraire, toutes les nouvelles familles obtenues restent mutuellement indépendantes.
Étant donné qu'il faut un nombre fini de telles opérations pour passer de la famille mutuellement indépendante $(A_1, \cdots, A_p)$ à la famille $(B_1, \cdots, B_p)$ où chaque B_k vaut arbitrairement A_k ou $\overline{A_k}$, cette nouvelle famille $(B_1, \cdots, B_p)$ reste de toutes façons mutuellement indépendante, ce qui achève la démonstration de la propriété 5.

Les variables aléatoires finies

On l'a déjà vu dans le préambule, les variables aléatoires réelles modéliseront les mesures de nombres réels faites à partir d'expériences aléatoires ω. Au programme de MPSI, on ne considèrera que des variables aléatoires réelles finies $X : \Omega \longrightarrow \mathbb{R}$ (en abrégé v.a.r. ou simplement v.a. pour désigner une variable aléatoire $X : \Omega \longrightarrow E$ où E est n'importe quel ensemble, pas forcément les nombres réels lorsque l'on effectue une mesure autre que numérique) définies sur un univers fini Ω. L'ensemble $X(\Omega)$ des valeurs prises par la fonction f est donc lui aussi un ensemble fini.

On voit que l'on a deux notations différentes selon le point de vue probabiliste ou le point de vue ensembliste.
Par exemple, si $X : \Omega \longrightarrow \mathbb{R}$ est une v.a.r. et si A est une partie de $\mathbb{R}$, on notera :
$$X^{[-1]}(A) = \Big\{\omega \in \Omega \mid X(\omega) \in A\Big\} = \{X \in A\}$$
De même, on notera :
$$\mathbb{P}\Big(X^{[-1]}(A)\Big) = \mathbb{P}(X \in A)$$
Par exemple, la probabilité de l'ensemble :
$$\Big\{\omega \in \Omega \mid X(\omega) \geqslant 1\Big\}$$
sera notée :
$$\mathbb{P}(X \geqslant 1).$$
La probabilité que la variable X prenne une valeur entière et soit comprise entre (-6) et 9 pourra être notée :
$$\mathbb{P}\Big(\{X \in \mathbb{Z}\} \cap \{X \in [-6, 9]\}\Big) = \mathbb{P}\Big(X \in \mathbb{Z}, X \in [-6, 9]\Big)$$
la virgule dans cette dernière expression faisant office d'intersection.
Dans toute la suite, on privilégiera au maximum la notation probabiliste qui permet un certain confort de lecture.

On peut associer à n'importe quelle v.a.r $X : \Omega \longrightarrow \mathbb{R}$ une probabilité appelée la loi de la variable X. C'est l'application notée $\mathbb{P}_X$ ainsi définie :

$$\mathbb{P}_X : \left| \begin{array}{ccc} \mathscr{P}(X(\omega)) & \longrightarrow & [0,1] \\ A & \longmapsto & \mathbb{P}(X \in A) \end{array} \right. .$$

On montre que l'application $\mathbb{P}_X$ est bien une probabilité. Elle est bien définie car $\mathbb{P}(X \in A)$ est toujours dans $[0,1]$, puisque $\mathbb{P}$ est une probabilité.
Deuxièmement, comme $\{X \in X(\Omega)\} = \Omega$, car on le rappelle :

$$\{X \in X(\Omega)\} = \Big\{\omega \in \Omega \mid X(\omega) \in X(\Omega)\Big\},$$

alors :

$$\mathbb{P}_X(X(\Omega)) = \mathbb{P}(X \in X(\Omega)) = \mathbb{P}(\Omega) = 1.$$

Troisièmement, si A et B sont deux parties disjointes dans $\mathscr{P}(X(\Omega))$, on remarque que les ensembles $\{X \in A\}$ et $\{X \in B\}$ sont deux parties incompatibles dans $\mathscr{P}(\Omega)$. Par σ-additivité de la probabilité $\mathbb{P}$, on peut alors écrire :

$$\begin{array}{rcl} \mathbb{P}_X(A \sqcup B) & = & \mathbb{P}(\{X \in A\} \sqcup \{X \in B\}) \\ & = & \mathbb{P}(X \in A) + \mathbb{P}(X \in B) \\ & = & \mathbb{P}_X(A) + \mathbb{P}_X(B). \end{array}$$

L'application $\mathbb{P}_X$ est bien σ-additive.
On se souvient que l'ensemble Ω est fini, ainsi que l'ensemble $X(\Omega)$. On sait alors que toute probabilité $\mu : \mathscr{P}(X(\Omega)) \longrightarrow [0,1]$ est entièrement déterminée sur les singletons. En notant $X(\Omega) = \{x_1, \cdots, x_s\}$, la loi $\mathbb{P}_X$ de la v.a. X est étroitement associée au s-uplet

$$\Pi = \Big(\pi_k = \mathbb{P}(X = x_k)\Big)_{1 \leqslant k \leqslant s}$$

de nombres réels positifs et de somme égale à 1. Connaissant cette famille Π, on aura alors directement la loi par la formule :

$$\mathbb{P}_X : \left| \begin{array}{ccc} \mathscr{P}(X(\Omega)) & \longrightarrow & [0,1] \\ A & \longmapsto & \displaystyle\sum_{k=1 \text{ tel que } x_k \in A}^{s} \pi_k \end{array} \right. .$$

Ceci explique la méthode de calcul d'une variable aléatoire X.
Pour détailler un peu plus cette méthode, il est parfois plus facile non pas de déterminer explicitement les valeurs prises par la fonction $X : \Omega \longrightarrow \mathbb{R}$, mais de déterminer un ensemble fini F inclus dans $\mathbb{R}$ et contenant $X(\Omega)$.
Ensuite, pour tout $\alpha \in F$, il faut calculer $\mathbb{P}(X = \alpha)$ en détaillant en général l'événement $\{X = \alpha\}$. Lorsque le nombre α appartient à $F \setminus X(\Omega)$, comme α n'est pas une valeur prise par la variable X, alors l'événement $\{X = \alpha\}$ est impossible (c'est l'ensemble vide) et donc dans ce cas : $\mathbb{P}(X = \alpha) = 0$.

Examinons plusieurs situations de calcul à travers l'exemple 5.

• Dans ce premier point, l'univers considéré est :

$$\Omega = \mathscr{P}(E).$$

L'expression « on choisit au hasard une partie $\omega \in \Omega$ » sous-entend que la probabilité retenue est la loi uniforme $\mathbb{P}$ ainsi définie :

$$\mathbb{P} : \left| \begin{array}{ccc} \mathscr{P}(\Omega) & \longrightarrow & [0,1] \\ A & \longmapsto & \dfrac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} = \dfrac{\mathrm{Card}(A)}{2^n} \end{array} \right. .$$

La variable aléatoire considérée est l'application :

$$X : \left| \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{N} \\ \omega & \longmapsto & \mathrm{Card}(\omega) \end{array} \right. ,$$

en insistant sur le fait que l'expérience ω est en fait une partie finie de E.
L'ensemble des valeurs prises par X est :

$$X(\Omega) = [\![0, n]\!],$$

car le cardinal de ω pris dans Ω est un entier compris entre 0 (si $\omega = \emptyset$) et n (si $\omega = \Omega$).
Ensuite, si k est un entier dans $X(\Omega)$, alors l'événement $\{X = k\}$ est en fait l'événement : « avoir pioché une partie de E à k éléments ».
Il y a $\begin{pmatrix} n \\ k \end{pmatrix}$ telles parties ω. On en déduit :

$$\mathbb{P}(X = k) = \begin{pmatrix} n \\ k \end{pmatrix} \frac{1}{2^n}.$$

On remarque que la loi de X est en fait la loi binomiale $\mathscr{B}\left(n, \dfrac{1}{2}\right)$.

• Dans ce protocole expérimental, on peut tout à fait calculer explicitement l'ensemble $X(\Omega)$ des valeurs prises. Le plus simple est d'utiliser un ensemble fini qui contiendra l'ensemble $X(\Omega)$ pour minimiser les calculs.
Une chose est sûre, on a toujours l'inclusion :

$$X(\Omega) \subset [\![0, q]\!],$$

car on pioche q boules et on ne peut piocher strictement plus de q boules blanches.

Considérons la situation où les tirages sont avec remise.
Si k est un entier entre 0 et q, l'événement $\{X = k\}$ consiste à piocher q boules avec remise avec k boules blanches et $q - k$ boules non blanches.
On compte le nombre de tirages favorables, vu que l'on est dans une situation d'équiprobabilité de chaque tirage.
Il y a $\begin{pmatrix} q \\ k \end{pmatrix}$ façons de placer les k boules blanches dans les q boules piochées. Une fois les k boules blanches placées (et donc aussi les $q - k$ boules non blanches), on peut choisir les boules blanches ou non piochées. Cela fait $p^k\ (N - p)^{q-k}$ choix.

Il y a donc $\begin{pmatrix} q \\ k \end{pmatrix} p^k \ (N-p)^{q-k}$ tirages favorables à l'événement $\{X=k\}$.
Il y a en tout N^q tirages possibles, car il y a N choix possibles pour chaque tirage.
La probabilité de piocher k boules blanches vaut donc :

$$\mathbb{P}(X=k) = \begin{pmatrix} q \\ k \end{pmatrix} \frac{p^k \ (N-p)^{q-k}}{N^q} = \begin{pmatrix} q \\ k \end{pmatrix} \alpha^k \cdot (1-\alpha)^{q-k},$$

où $\alpha = \dfrac{p}{N}$. On s'aperçoit que la variable X suit la loi binomiale $\mathscr{B}(q,\alpha)$.

Lorsque les tirages sont avec remise, il y a $\displaystyle\prod_{i=0}^{q}(N-i) = \frac{N!}{(N-q)!}$ tirages possibles.
En fixant un entier k dans $[\![0,q]\!]$, on compte le nombre de tirages favorables à l'événement $\{X=k\}$. Il faut choisir la place des p boules blanches parmi les N tirages, ce qui donne $\begin{pmatrix} q \\ k \end{pmatrix}$ choix de placements. Ensuite, une fois que le placement est fixé, il suffit de compléter les cases (k cases à remplir avec des boules blanches et $(q-k)$ cases à remplir avec des boules autres).
Pour les choix des k boules blanches, il y a $\displaystyle\prod_{i=0}^{k-1}(p-i) = \frac{p!}{(p-k)!}$ choix possibles et pour les choix des $q-k$ boules autres que blanches, il y a $\displaystyle\prod_{i=0}^{q-k-1}(N-p-i) = \frac{(N-p)!}{(N-p-q+k)!}$, lorsque ce calcul est possible (autrement dit lorsque $N-p-q+k \geqslant 0$) et 0 sinon.
On en déduit que la probabilité de l'événement $\{X=k\}$ vaut lorsque le calcul est possible (autrement dit, lorsqu'il s'agit d'entiers naturels à chaque fois qu'apparaît une factorielle) :

$$\mathbb{P}(X=k) = \begin{pmatrix} q \\ k \end{pmatrix} \cdot \frac{\frac{p!}{(p-k)!} \times \frac{(N-p)!}{(N-p-q+k)!}}{\frac{N!}{(N-q)!}} = \frac{\begin{pmatrix} p \\ k \end{pmatrix} \times \begin{pmatrix} N-p \\ q-k \end{pmatrix}}{\begin{pmatrix} N \\ q \end{pmatrix}}.$$

En ce qui concerne les tirages simultanés, il y a $\begin{pmatrix} N \\ q \end{pmatrix}$ tirages possibles, car l'ordre de pioche n'intervient pas.
Ensuite, en fixant un entier k dans $[\![0,q]\!]$, on compte le nombre de tirages simultanés favorables à l'événement $\{X=k\}$.
Pour les compter, comme l'ordre n'intervient plus, il suffit de choisir les k boules blanches prises parmi les boules blanches et les $q-k$ boules non blanches prises parmi les boules autres que blanches. Cela fait : $\begin{pmatrix} p \\ k \end{pmatrix} \cdot \begin{pmatrix} N-p \\ q-k \end{pmatrix}$ tirages simultanés favorables.
En conclusion, la probabilité de l'événement $\{X=k\}$ vaut :

$$\mathbb{P}(X=k) = \frac{\begin{pmatrix} p \\ k \end{pmatrix} \times \begin{pmatrix} N-p \\ q-k \end{pmatrix}}{\begin{pmatrix} N \\ q \end{pmatrix}}.$$

On retrouve la même loi que lors des tirages sans remise, ce qui est plutôt une bonne nouvelle car intuitivement, même si le protocole expérimental change légèrement, procéder à des tirages sans remise ou simultanés revient souvent au même. Dans les tirages sans remise, on aligne les boules piochées dans un certain ordre, alors que dans les tirages simultanés, on prend les boules en même temps, mais ici, le nombre de boules blanches piochées ne dépend pas de l'ordre de pioche.

Quelques remarques pour terminer sur cet exemple qui est à bien comprendre.
Premièrement, il est plus facile de procéder au calcul lors de tirages simultanés que lors de tirages sans remise.
Deuxièmement, dans le cas de tirages avec remise, lorsque $1 \leqslant p < N$, on remarque que pour tout entier k dans $[\![0, q]\!]$, la probabilité calculée :

$$\mathbb{P}(X = k) = \left(\begin{array}{c} q \\ k \end{array} \right) \alpha^k \ (1 - \alpha)^{q-k}$$

est strictement positive. De ce fait, l'ensemble des valeurs vraiment prises par la variable X est :

$$X(\Omega) = [\![0, q]\!].$$

Pour les tirages simultanés (c'est la même chose pour les tirages sans remise), toujours dans le cas $1 \leqslant p < N$, tous les entiers k entre 0 et q ne sont pas forcément des valeurs prises par la fonction X ; en tout cas, pour certains entiers k entre 0 et q, l'événement $\{X = k\}$ peut être de probabilité nulle. En effet, pour que la valeur k soit effectivement une valeur prise par la fonction X, il faut déjà que l'entier k soit inférieur à p car on ne peut pas piocher strictement plus de boules qu'il n'y a de boules blanches en tout.
Ensuite, il faut également que le nombre $q-k$ de boules non blanches soit inférieur ou égal au nombre total $N - p$ de boules autres que blanches. Cela impose à l'entier k d'être supérieur ou égal à $N - p - q$.
En résumé, l'entier k doit appartenir à l'ensemble :

$$\Big[\!\!\Big[\max\{0, N - p - q\}, \min\{p, q\} \Big]\!\!\Big].$$

Réciproquement, si k appartient à cet ensemble, il est facile de voir qu'il existe une expérience ω telle que $X(\omega) = k$ lorsque ω est l'ensemble où l'on réunit n'importe quelles k boules blanches et n'importe quelles $q - k$ boules autres que blanches par exemple.
Par conséquent, les valeurs prises par la variable aléatoire X avec une probabilité strictement positives forment l'ensemble

$$\Big[\!\!\Big[\max\{0, N - p - q\}, \min\{p, q\} \Big]\!\!\Big].$$

Le plus simple est de dire que l'ensemble des valeurs prises par X est inclus dans $[\![0, q]\!]$ et que certaines valeurs de cet ensemble $[\![0, q]\!]$ sont peut-être des valeurs prises par la variable X avec une probabilité nulles (donc pas vraiment une valeur prise ; le probabiliste fonctionne de toutes façons de façon « presque sûre »...)

Démontrons maintenant la propriété 6.
• **Le premier point est extrêmement important. À retenir de toute urgence.**

On considère une variable aléatoire finie $X : \Omega \longrightarrow \mathbb{R}$. On considère la famille

$$\mathscr{F} = \Big(\{X = x\}\Big)_{x \in X(\Omega)}.$$

On va montrer que la famille $\mathscr{F}$ est un SCE.
En effet, si x et y sont deux éléments différents dans $X(\Omega)$, l'intersection $\{X = x\} \cap \{X = y\}$ est vide car sinon, il existerait un élément ω dans l'intersection $\{X = x\} \cap \{X = y\}$, ce qui imposerait :

$$X(\omega) = x \text{ et } X(\omega) = y \text{ , donc } x = y.$$

Les événements de la famille $\mathscr{F}$ sont donc deux à deux incompatibles.
Ensuite, on montre que :

$$\bigcup_{x \in X(\Omega)} \{X = x\} = \Omega.$$

L'inclusion « $\subset$ » est évidente car on rappelle la notation ensembliste pour l'ensemble $\{X = x\}$:

$$\{X = x\} = \Big\{\omega \in \Omega \mid X(\omega) = x\Big\}.$$

Chaque ensemble $\{X = x\}$ est donc une partie de Ω : leur réunion également.
Pour l'autre inclusion « $\supset$ », on considère un élément ω dans Ω. On pose $\xi = X(\omega)$. Alors :

$$\omega \in \{X = \xi\} \subset \bigcup_{x \in X(\Omega)} \{X = x\}.$$

On a l'inclusion réciproque.
Il faut bien faire attention à vos manipulations d'objets. Si $X : \Omega \longrightarrow \mathbb{R}$ est une variable aléatoire, l'ensemble $\{X = 1\}$ par exemple est bien un ensemble d'expériences $\omega \in \Omega$ telles que $X(\omega) = 1$. Il ne faut pas perdre de vue le fait que dans la notation $\{X = 1\}$ se cache de manière implicite un ensemble relatif aux éléments $\omega \in \Omega$, même si ces éléments ω n'apparaissent pas explicitement dans la notation $\{X = 1\}$ par exemple.

• Pour le deuxième point, considèrons une variable aléatoire finie $X : \Omega \longrightarrow \mathbb{R}$, puis $f : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction quelconque.
Déjà, la composition $f \circ X : \Omega \longrightarrow \mathbb{R}$ est tout à fait possible et on peut noter avec toujours les mêmes versions ensembliste /probabiliste :

$$f \circ X = f(X).$$

La fonction $f(X) : \Omega \longrightarrow \mathbb{R}$ est donc une variable aléatoire finie puisque l'ensemble Ω est fini, donc l'ensemble $f(X(\Omega))$ également.
La loi $\mathbb{P}_{f(X)}$ de cette v.a.r. $f(X)$ est uniquement déterminée par les valeurs prises sur les singletons. Plus précisément, si y est dans $f(X(\Omega))$, alors :

$$\begin{aligned}\mathbb{P}_{f(X)}(\{y\}) &= \mathbb{P}(f(X) = y) \\ &= \mathbb{P}\Big(X \in f^{[-1]}(\{y\})\Big).\end{aligned}$$

Ce deuxième point est « vide » en terme d'idées nouvelles. On est juste en train de manipuler des notations probabilistes ou ensemblistes, rien de plus.

On va associer à toute variable aléatoire finie $X : \Omega \longrightarrow \mathbb{R}$ un nombre que l'on appellera espérance de la variable aléatoire X et notée $\mathbb{E}(X)$.
Nous allons démontrer les deux formules proposées pour l'espérance de X. Retenez dès à présent plutôt la formule :

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x),$$

car l'autre formule n'est valable que lorsque l'univers Ω est fini, ce qui ne sera pas forcément le cas en deuxième année. Lorsqu'en deuxième année l'ensemble $X(\Omega)$ est infini dénombrable, alors on aura affaire à une série plutôt qu'à une somme finie...

Nous allons – comme d'habitude dans bon nombre de démonstrations à venir – utiliser le SCE $\Big(\{X = x\}\Big)_{x \in X(\Omega)}$ pour découper la somme :

$$A = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\}).$$

Voici le détail des calculs :

$$\begin{aligned}
A &= \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\}) \\
&= \sum_{x \in X(\Omega)} \left(\sum_{\omega \in \Omega \text{ et } X(\omega) = x} X(\omega) \cdot \mathbb{P}(\{\omega\}) \right) \\
&= \sum_{x \in X(\Omega)} \left(\sum_{\omega \in \Omega \text{ et } X(\omega) = x} x \cdot \mathbb{P}(\{\omega\}) \right) \\
&= \sum_{x \in X(\Omega)} x \cdot \left(\sum_{\omega \in \Omega \text{ et } X(\omega) = x} \mathbb{P}(\{\omega\}) \right) \\
&= \sum_{x \in X(\Omega)} x \cdot \mathbb{P} \left(\bigsqcup_{\omega \in \Omega \text{ et } X(\omega) = x} \{\omega\} \right) \\
&= \sum_{x \in X(\Omega)} x \cdot \mathbb{P}\left(X = x\right).
\end{aligned}$$

Une variable est centrée si son espérance est nulle. On verra plus tard la faible utilité de cela.

Expliquons la méthode en haut de la page 7.
Pour calculer l'espérance d'une variable aléatoire $X : \Omega \longrightarrow \mathbb{R}$, il y a plusieurs possibilités.
La première démarche à suivre est de calculer la loi de X. On rappelle que cela nécessite de trouver un ensemble fini F contenant l'ensemble $X(\Omega)$ des valeurs prises par la variable X, puis pour tout $x \in F$ de calculer la probabilité $\mathbb{P}(X = x)$. Une fois que l'on a la loi de la variable X, on peut théoriquement calculer la somme :

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x) = \sum_{x \in F} x \cdot \mathbb{P}(X = x),$$

car on rappelle que les éléments x dans $F \setminus X(\Omega)$ sont associés à des probabilités $\mathbb{P}(X = x)$ nulles.
Une deuxième approche un peu plus subtile est d'interpréter la variable aléatoire $X : \Omega \longrightarrow \mathbb{R}$ comme une somme d'autres variables aléatoires plus simples.
Par exemple, si A est un événement – donc une partie de l'univers Ω, on rappelle que la fonction indicatrice de l'événement A est la fonction :

$$\mathbf{1}_A : \left| \begin{array}{ccl} \Omega & \longrightarrow & \{0,1\} \\ \omega & \longmapsto & \left\{ \begin{array}{l} 1, \text{ si } \omega \in A \\ 0, \text{ si } \omega \notin A \end{array} \right. \end{array} \right. .$$

Il s'agit d'une variable aléatoire $\mathbf{1}_A : \Omega \longrightarrow \{0,1\}$. De plus, on a directement les égalités :

$$\{\mathbf{1}_A = 1\} = A \text{ et } \{\mathbf{1}_A = 0\} = \overline{A}.$$

Il est facile de calculer l'espérance de cette variable aléatoire. Voici le calcul très simple mais **très important** pour l'espérance de l'indicatrice :

$$\begin{aligned} \mathbb{E}(\mathbf{1}_A) &= \sum_{x \in \mathbf{1}_A(\Omega)} x \cdot \mathbb{P}(\mathbf{1}_A = x) \\ &= \sum_{k \in \{0,1\}} k \cdot \mathbb{P}(\mathbf{1}_A = k) \\ &= 0 \cdot \mathbb{P}(\mathbf{1}_A = 0) + 1 \cdot \mathbb{P}(\mathbf{1}_A = 1) \\ &= \mathbb{P}(A). \end{aligned}$$

On démontrera à travers la propriété 7 que l'espérance est linéaire. On en déduit que si l'on arrive à interpréter la variable aléatoire $X : \Omega \longrightarrow \mathbb{R}$ comme une combinaison linéaire de fonctions indicatrices :

$$X = \sum_{k=1}^{n} \lambda_k \cdot \mathbf{1}_{A_k}$$

alors on peut maintenant accéder à l'espérance de la variable X sans avoir recours à la loi de X, selon les calculs suivants :

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left(\sum_{k=1}^{n} \lambda_k \cdot \mathbf{1}_{A_k} \right) \\ &= \sum_{k=1}^{n} \lambda_k \cdot \mathbb{E}(\mathbf{1}_{A_k}) \qquad \text{[linéarité de l'espérance]} \\ &= \sum_{k=1}^{n} \lambda_k \cdot \mathbb{P}(A_k) \qquad , \text{ car } \mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A) \end{aligned}$$

Démontrons maintenant la propriété 7 qui regroupe les propriétés fondamentales à retenir pour l'instant sur l'espérance d'une variable aléatoire finie.

• Soit $(\Omega, \mathbb{P})$ un espace probabilisé fini.
L'ensemble des variables aléatoires finies $X : \Omega \longrightarrow \mathbb{R}$ n'est autre que l'espace vectoriel $\mathscr{F}(\Omega, \mathbb{R})$ de toutes les fonctions définies de Ω et à valeurs réelles.
Soient maintenant $X : \Omega \longrightarrow \mathbb{R}$ et $Y : \Omega \longrightarrow \mathbb{R}$ deux variables aléatoires définies sur l'univers fini Ω. On se donne un scalaire réel λ. Il est plus facile ici d'utiliser la première formule donnant l'espérance d'une variable aléatoire finie ; on pourra voir plus tard dans le chapitre comment il est possible d'utiliser la seconde formule.
On peut écrire successivement :

$$\begin{aligned}
\mathbb{E}(\lambda \cdot X + Y) &= \sum_{\omega \in \Omega} \Big(\lambda \cdot X + Y\Big)(\omega) \cdot \mathbb{P}(\{\omega\}) \\
&= \sum_{\omega \in \Omega} \Big(\lambda \cdot X(\omega) + Y(\omega)\Big) \cdot \mathbb{P}(\{\omega\}) \\
&= \lambda \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\}) + \sum_{\omega \in \Omega} Y(\omega) \cdot \mathbb{P}(\{\omega\}) \\
&= \lambda \cdot \mathbb{E}(X) + \mathbb{E}(Y).
\end{aligned}$$

• Soient $X : \Omega \longrightarrow \mathbb{R}$ et $Y : \Omega \longrightarrow \mathbb{R}$ deux variables aléatoires telles que $X \leqslant Y$, autrement dit :

$$\forall \omega \in \Omega, \;\; X(\omega) \leqslant Y(\omega).$$

On en déduit :

$$\begin{aligned}
\mathbb{E}(Y) - \mathbb{E}(X) &= \mathbb{E}(Y - X) \qquad \text{[linéarité de l'espérance]} \\
&= \sum_{\omega \in \Omega} \Big(Y(\omega) - X(\omega)\Big) \cdot \mathbb{P}(\{\omega\}) \qquad \bigstar \\
&\geqslant 0,
\end{aligned}$$

car la somme $\bigstar$ est une somme dont chaque terme est positif.
On remarque une analogie avec l'intégrale : l'intégrale est linéaire et si $a \leqslant b$ sont deux réels, puis $f : [a, b] \longrightarrow \mathbb{R}$ et $g : [a, b] \longrightarrow \mathbb{R}$ sont deux fonctions continues telles que $f \leqslant g$ sur le segment $[a, b]$, alors :

$$\int_a^b f \leqslant \int_a^b g.$$

On sait qu'une somme comme définissant l'espérance est conceptuellement une intégrale discrète... Il est donc normal de trouver une telle analogie.

• Le point suivant est également très important et connu sus le nom de formule de transfert.
Démontrons cette formule de transfert.
Soient $X : \Omega \longrightarrow \mathbb{R}$ une v.a.r. et $f : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction (la fonction f pourrait être également définie de $X(\Omega)$ vers $\mathbb{R}$).

Alors, on fait une nouvelle fois intervenir le SCE $\Big(\{X = x\}\Big)_{x \in X(\Omega)}$:

$$\begin{aligned}
\mathbb{E}(f(X)) &= \sum_{y \in f(X(\Omega))} y \cdot \mathbb{P}(f(X) = y) \qquad \text{[formule utile de l'espérance de } f(X)\text{]} \\
&= \sum_{y \in f(X(\Omega))} y \cdot \left(\sum_{x \in X(\Omega)} \mathbb{P}(f(X) = y, X = x) \right) \qquad \text{[formule des probabilités totales]}
\end{aligned}$$

Fixons un élément y dans $f(X(\Omega))$ et un élément x dans $X(\Omega)$.
On distingue deux cas :

- premier cas : $y \neq f(x)$. On va montrer alors que :

  $$\{f(X) = y, X = x\} = \emptyset.$$

  On rappelle que la virgule dans un ensemble fait office d'intersection. Si l'ensemble $\{f(X) = y, X = x\} = \{f(X) = y\} \cap \{X = x\}$ était non vide, on trouverait un élément ω dans l'intersection. On aurait alors d'une part :

  $$f(X)(\omega) = f(X(\omega)) = y$$

  et d'autre part :

  $$X(\omega) = x,$$

  donc $y = f(X(\omega)) = f(x)$, ce qui a été exclu. L'intersection $\{f(X) = y\} \cap \{X = x\}$ est bien vide et donc :

  $$\mathbb{P}(f(X) = y, X = x) = \mathbb{P}(\emptyset) = 0.$$

- second cas : $y = f(x)$. On va montrer alors que :

  $$\{f(X) = y, X = x\} = \{X = x\}.$$

  $\triangleright$ Soit ω dans l'ensemble de gauche. Alors, $f(X(\omega)) = y$ et $X(\omega) = x$, donc $X(\omega) = x$ et l'élément ω appartient au second ensemble.
  $\triangleright$ Soit ω dans l'ensemble de droite. Alors, $X(\omega) = x$, puis $f(X(\omega)) = f(x) = y$, et donc ω appartient à la fois à l'ensemble $\{X = x\}$ et à l'ensemble $\{f(X) = y\}$ et donc à l'intersection de droite.

On poursuit alors le calcul de la somme en séparant la sommation sur x selon que x vérifie ou non l'égalité $f(x) = y$:

$$\begin{aligned}
\mathbb{E}(f(X)) &= \sum_{y \in f(X(\Omega))} y \cdot \left(\sum_{x \in X(\Omega)} \mathbb{P}(f(X) = y, X = x) \right) \\
&= \sum_{y \in f(X(\Omega))} y \cdot \left(\sum_{x \in X(\Omega) \text{ et } f(x) = y} \mathbb{P}(f(X) = y, X = x) \right) + \\
&\qquad \sum_{y \in f(X(\Omega))} y \cdot \left(\sum_{x \in X(\Omega) \text{ et } f(x) \neq y} \mathbb{P}(f(X) = y, X = x) \right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{y\in f(X(\Omega))} y\cdot\left(\sum_{x\in X(\Omega)\text{ et } f(x)=y} \mathbb{P}(X=x)\right) + \sum_{y\in f(X(\Omega))} y\cdot\left(\sum_{x\in X(\Omega)\text{ et } f(x)\neq y} \mathbb{P}(\emptyset)\right)\\
&= \sum_{y\in f(X(\Omega))} y\cdot\left(\sum_{x\in X(\Omega)\text{ et } f(x)=y} \mathbb{P}(X=x)\right)\\
&= \sum_{y\in f(X(\Omega))} \left(\sum_{x\in X(\Omega)\text{ et } f(x)=y} f(x)\cdot\mathbb{P}(X=x)\right)\\
&= \sum_{(x,y)\in X(\Omega)\times f(X(\Omega))\text{ et } f(x)=y} f(x)\cdot\mathbb{P}(X=x)\\
&= \sum_{x\in X(\Omega)} \left(\sum_{y\in f(X(\Omega))\text{ et } y=f(x)} f(x)\cdot\mathbb{P}(X=x)\right)\\
&= \sum_{x\in X(\Omega)} f(x)\cdot\mathbb{P}(X=x),
\end{aligned}
$$

car la sommation $\displaystyle\sum_{y\in f(X(\Omega))\text{ et } y=f(x)} \star$ ne comporte qu'un seul terme correspondant au seul indice $y=f(x)$.

Voici en couleur une démonstration sûrement plus compacte et digeste de cette formule de transfert.
On se donne une variable aléatoire finie $X:\Omega\longrightarrow\mathbb{R}$ et une fonction $f:\mathbb{R}\longrightarrow\mathbb{R}$. On va commencer par montrer la formule :

$$f(X)=\sum_{x\in X(\Omega)} f(x)\cdot\mathbf{1}_{\{X=x\}}.$$

On fixe donc un élément ω dans Ω.
On pose $\xi=X(\omega)$, de sorte que pour tout x dans $X(\Omega)$:

$\longrightarrow$ ou bien $x=\xi$ auquel cas :

$$\mathbf{1}_{\{X=x\}}(\omega)=1,$$

$\longrightarrow$ ou bien $x\neq\xi$ auquel cas :

$$\mathbf{1}_{\{X=x\}}(\omega)=0.$$

On en déduit :

$$
\begin{aligned}
\left(\sum_{x\in X(\Omega)} f(x)\cdot\mathbf{1}_{\{X=x\}}\right)(\omega) &= \sum_{x\in X(\Omega)} f(x)\cdot\mathbf{1}_{\{X=x\}}(\omega)\\
&= f(\xi)\cdot 1\\
&= f(\xi)\\
&= f(X(\omega))\\
&= f(X)(\omega).
\end{aligned}
$$

On a bien démontré cette formule.
Il suffit maintenant de prendre l'espérance, ce qui donne :

$$\begin{aligned}
\mathbb{E}(f(X)) &= \mathbb{E}\left(\sum_{x\in X(\Omega)} f(x)\cdot \mathbf{1}_{\{X=x\}}\right) \\
&= \sum_{x\in X(\Omega)} f(x)\cdot \mathbb{E}\Big(\mathbf{1}_{\{X=x\}}\Big) \\
&= \sum_{x\in X(\Omega)} f(x)\cdot \mathbb{P}(X=x)
\end{aligned}$$

et on obtient ce qu'il faut.

• Si a et b sont deux réels fixés, on va appliquer la formule de transfert à la fonction $f : x \longmapsto ax+b$, ce qui donne :

$$\begin{aligned}
\mathbb{E}(aX+b) &= \mathbb{E}(f(X)) \\
&= \sum_{x\in X(\Omega)} f(x)\cdot \mathbb{P}(X=x) \\
&= \sum_{x\in X(\Omega)} (ax+b)\cdot \mathbb{P}(X=x) \\
&= a\cdot \sum_{x\in X(\Omega)} x\cdot \mathbb{P}(X=x) + b\cdot \sum_{x\in X(\Omega)} \mathbb{P}(X=x) \\
&= a\cdot \mathbb{E}(X)+b,
\end{aligned}$$

car comme $\Big(\{X=x\}\Big)_{x\in X(\Omega)}$ est un SCE, alors :

$$\begin{aligned}
1 &= \mathbb{P}(\Omega) \\
&= \mathbb{P}\left(\bigsqcup_{x\in X(\Omega)} \{X=x\}\right) \\
&= \sum_{x\in X(\Omega)} \mathbb{P}(X=x).
\end{aligned}$$

• Soient $a>0$ un nombre réel, puis $X : \Omega \longrightarrow \mathbb{R}$ une variable aléatoire finie.
On remarque que l'on a l'inégalité :

$$a\cdot \mathbf{1}_{\{|X|\geqslant a\}} \leqslant |X|.$$

En effet, on va le vérifier pour tout élément ω de l'univers Ω.
Soit ω un élément de Ω. On distingue deux cas :
$\longrightarrow$ si $|X(\omega)|<a$, alors :

$$a\cdot \mathbf{1}_{\{|X|\geqslant a\}}(\omega) = a\cdot 0 = 0 \leqslant |X(\omega)| = |X|(\omega)$$

$\longrightarrow$ si $|X(\omega)| \geqslant a$, alors :

$$a \cdot \mathbf{1}_{\{|X| \geqslant a\}}(\omega) = a \cdot 1 = a \leqslant |X(\omega)| = |X|(\omega).$$

Quoiqu'il arrive, on a bien :

$$\forall \omega \in \Omega, \ a \cdot \mathbf{1}_{\{|X| \geqslant a\}}(\omega) \leqslant |X|(\omega).$$

On prend maintenant l'espérance dans cette inégalité, ce qui donne – compte tenu du fait que si $U : \Omega \longrightarrow \mathbb{R}$ et $V : \Omega \longrightarrow \mathbb{R}$ sont deux variables aléatoires telles que $U \leqslant V$, alors : $\mathbb{E}(U) \leqslant \mathbb{E}(V)$:

$$\mathbb{E}\Big(a \cdot \mathbf{1}_{\{|X| \geqslant a\}}\Big) \leqslant \mathbb{E}\Big(|X|\Big),$$

ou encore par linéarité de l'espérance et le fait que pour tout événement $A \subset \Omega$, $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$:

$$a \cdot \mathbb{P}(|X| \geqslant a) \leqslant \mathbb{E}\Big(|X|\Big).$$

Il suffit maintenant de tout diviser par le réel strictement positif a, sans modifier le sens de l'inégalité. On obtient alors l'inégalité voulue.

Il est bien de retenir l'ensemble des démonstrations de cette propriété 7, où l'on commence à utiliser ce que l'on sait sur les probabilités et ce que l'on apprend sur les variables aléatoires. En particulier, lorsque l'on doit montrer qu'une certaine probabilité est inférieure à une expression faisant intervenir une espérance, comme dans l'inégalité de Markov et d'autres inégalités du même style à venir, on retiendra que l'on mettra en place une inégalité avec une indicatrice pour ensuite prendre l'espérance dans cette inégalité.

Mettons tout ceci en pratique et détaillons plusieurs situations à travers les exemples 6 ou 7.

• On note $X : \Omega \longrightarrow \mathbb{N}$ la variable comptant le nombre aléatoire de boules blanches piochées. On va calculer l'espérance de X de deux manières différentes.

$\longrightarrow$ la méthode standard en calculant la loi. On sait que $X(\Omega) = [\![0, q]\!]$ et pour tout $k \in X(\Omega)$:

$$\mathbb{P}(X = k) = \begin{pmatrix} q \\ k \end{pmatrix} \alpha^k \ (1-\alpha)^{q-k},$$

où $\alpha = \dfrac{p}{N}$ est la proportion de boules blanches dans le sac.
À l'aide de cette loi, on dispose du calcul classique de l'espérance :

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{q} k \ \mathbb{P}(X = k) \\
&= \sum_{k=0}^{q} k \begin{pmatrix} q \\ k \end{pmatrix} \alpha^k \ (1-\alpha)^{q-k} \\
&= \sum_{k=1}^{q} k \begin{pmatrix} q \\ k \end{pmatrix} \alpha^k \ (1-\alpha)^{q-k} \\
&= \sum_{k=1}^{q} q \begin{pmatrix} q-1 \\ k-1 \end{pmatrix} \alpha^k \ (1-\alpha)^{q-k}
\end{aligned}$$

$$\begin{aligned}
&= q \sum_{\ell=0}^{q-1} \binom{q-1}{\ell} \alpha^{\ell+1} (1-\alpha)^{q-1-\ell} \\
&= q\,\alpha \sum_{\ell=0}^{q-1} \binom{q-1}{\ell} \alpha^{\ell} (1-\alpha)^{q-1-\ell} \\
&= q\,\alpha\,(\alpha + (1-\alpha))^{q-1} \qquad \text{[binôme dans } \mathbb{R} \text{ commutatif]} \\
&= q\,\alpha = q\,\frac{p}{N}.
\end{aligned}$$

$\longrightarrow$ on peut interpréter X comme une somme d'indicatrices. On note pour tout i entre 1 et q, l'événement A_i : « la $i^{ème}$ boule piochée est blanche ».
On remarque alors que l'on peut exprimer la variable X selon :

$$X = \sum_{i=1}^{q} \mathbf{1}_{A_i},$$

car si ω est un élément de Ω, alors la somme $\displaystyle\sum_{i=1}^{q} \mathbf{1}_{A_i}(\omega)$ est une somme composée de termes valant 0 ou 1, la valeur globale de la somme valant le nombre de termes valant 1, autrement dit correspondant au nombre total de boules blanches piochées.
On en déduit par linéarité de l'espérance :

$$\mathbb{E}(X) = \sum_{i=1}^{q} \mathbb{E}(\mathbf{1}_{A_i}) = \sum_{i=1}^{q} \mathbb{P}(A_i).$$

Or, si i est un entier entre 1 et q, alors la probabilité de l'événement A_i est égale à la proportion de boules blanches dans le sac :

$$\forall i \in [\![1, q]\!],\ \mathbb{P}(A_i) = \frac{p}{N}.$$

Conclusion,

$$\mathbb{E}(X) = \sum_{i=1}^{q} \frac{p}{N} = q\,\frac{p}{N}.$$

• On munit l'univers $\Omega = \mathscr{P}([\![1, n]\!])$ de la probabilité uniforme $\mathbb{P} = \mathscr{U}_{2^n}$. On note

$$X : \left| \begin{array}{ccc} \mathscr{P}([\![1, n]\!]) & \longrightarrow & \mathbb{N} \\ A & \longmapsto & \mathrm{Card}(A) \end{array} \right.$$

la variable aléatoire considérée.
Là encore, on propose les deux méthodes pour calculer l'espérance de la variable X.
$\longrightarrow$ On commence par calculer la loi de X. Le calcul a déjà été fait en exemple 5.
Ainsi, $X(\Omega) = [\![0, n]\!]$ et pour tout entier k entre 0 et n :

$$\mathbb{P}(X = k) = \binom{n}{k} \frac{1}{2^n}.$$

On en déduit le calcul de l'espérance :

$$\mathbb{E}(X) = \sum_{k=0}^{n} k \begin{pmatrix} n \\ k \end{pmatrix} \frac{1}{2^n} = \frac{1}{2^n} \sum_{k=0}^{n} k \begin{pmatrix} n \\ k \end{pmatrix}.$$

On peut encore utiliser la formule pour tout $k \in [\![1, n]\!]$,

$$k \begin{pmatrix} n \\ k \end{pmatrix} = n \begin{pmatrix} n-1 \\ k-1 \end{pmatrix},$$

pour avoir :

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{2^n} \sum_{k=0}^{n} k \begin{pmatrix} n \\ k \end{pmatrix} \\ &= \frac{1}{2^n} \sum_{k=1}^{n} k \begin{pmatrix} n \\ k \end{pmatrix} \\ &= \frac{1}{2^n} \sum_{\ell=0}^{n-1} n \begin{pmatrix} n-1 \\ \ell \end{pmatrix} \\ &= \frac{1}{2^n}\, n\, (1+1)^{n-1} \\ &= \frac{n}{2}. \end{aligned}$$

On peut aussi interpréter le terme k comme issu d'une dérivation. On pose la fonction :

$$f : t \longmapsto \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} t^k = (1+t)^n,$$

donc en dérivant :

$$f' : t \longmapsto \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} k\, t^{k-1} = n\, (1+t)^{n-1}.$$

Conclusion,

$$\mathbb{E}(X) = \frac{1}{2^n}\, f'(1) = \frac{1}{2^n}\, n\, 2^{n-1} = \frac{n}{2}.$$

$\longrightarrow$ On peut également interpréter la variable X comme une combinaison linéaire de fonctions indicatrices. Pour tout entier i entre 1 et n, on note A_i l'événement : « la partie piochée contient l'élément i ».

On remarque alors que la somme $\sum_{i=1}^{n} \mathbf{1}_{A_i}$ compte exactement le nombre aléatoire d'éléments dans la partie piochée ω.

Conclusion,

$$X = \sum_{i=1}^{n} \mathbf{1}_{A_i},$$

donc :

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Or, parmi toutes les parties possibles ω dans $\Omega = \mathscr{P}(\llbracket 1, n \rrbracket)$, il y a exactement autant de parties contenant l'élément i que de parties ne contenant pas l'élément i, par la bijection :

$$\varphi : \omega \longmapsto \omega \setminus \{i\}$$

de bijection réciproque $\varphi^{-1} : \omega \longmapsto \omega \cup \{i\}$.
Conclusion,

$$\mathbb{P}(A_i) = \frac{1}{2},$$

et finalement :

$$\mathbb{E}(X) = \sum_{i=1}^{n} \frac{1}{2} = \frac{n}{2}.$$

• Pour l'exemple 7, premier point, en notant X le nombre de colonnes entièrement remplies par des croix, on verra plus tard dans le chapitre que le nombre moyen demandé correspond à l'espérance de la variable aléatoire $X : \Omega \longrightarrow \mathbb{N}$.
Au lieu de calculer la loi de la variable X qui est un peu pénible à calculer, on va décomposer la variable X en une somme d'indicatrices. Pour ce faire, pour tout indice i entre 1 et n, on note A_i l'événement : « la $i^{ème}$ colonne du tableau est entièrement remplie de croix ».
On remarque alors que l'on a la formule :

$$X = \sum_{i=1}^{n} \mathbf{1}_{A_i},$$

et donc en prenant l'espérance :

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Fixons un entier i entre 1 et n et calculons $\mathbb{P}(A_i)$. Comme le placement des croix se fait « au hasard » – comprendre « de manière équiprobable » – et que les croix sont indiscernables, il y a $\begin{pmatrix} 2n \\ p \end{pmatrix}$ placements possibles.
On compte maintenant le nombre de placements favorables à l'événement A_i. Pour cela, il faut déjà placer deux croix dans la $i^{ème}$ colonne. Il reste à placer comme on veut les $(p-2)$ autres croix dans les $2n-2$ autres cases. Il y a donc $\begin{pmatrix} 2n-2 \\ p-2 \end{pmatrix}$ placements favorables.
Conclusion,

$$\mathbb{P}(A_i) = \frac{\begin{pmatrix} 2n-2 \\ p-2 \end{pmatrix}}{\begin{pmatrix} 2n \\ p \end{pmatrix}} = \frac{p(p-1)}{2n(2n-1)}$$

et donc :

$$\mathbb{E}(X) = \frac{p(p-1)}{2(2n-1)}.$$

On peut s'amuser à interpréter cette formule, ou tout du moins à vérifier si elle tient la route. Par exemple, si $p = 1$, il n'y a absolument aucune chance de remplir entièrement ne serait-ce qu'une colonne : le nombre moyen est égal à 0, conformément à la formule trouvée.
Lorsque $p = 2n$, on remplit toutes les colonnes et le nombre moyen de colonnes remplies vaut n : ça marche !
Lorsque $p = 2n-1$, on remplit toutes les cases, sauf une seule. Quoiqu'il arrive, il y a toujours $n-1$ colonnes totalement remplies. Cela coïncide également avec la formule trouvée. Celle-ci semble être juste. Ce genre d'interprétation est souhaitable à l'oral d'un concours. On peut ainsi mettre en évidence notre « sens probabiliste » exactement de la même façon que vous démontrez votre « sens physique » lorsque vous explorez les cas limites dans une formule que vous venez d'établir.

• Pour le deuxième point de l'exemple 7, on ne va pas s'amuser à calculer la loi de X. On va plutôt interpréter la variable X comme une somme de p variables plus simples. On différencie les dés que l'on lance en les numérotant de 1 à p.
On note $X_i : \Omega \longrightarrow [\![1,6]\!]$ la variable aléatoire donnant le numéro de la face supérieure du $i^{ème}$ dé. De ce fait, il est clair que :

$$X = \sum_{i=1}^{p} X_i.$$

Par linéarité de l'espérance, on obtient :

$$\mathbb{E}(X) = \sum_{i=1}^{p} \mathbb{E}(X_i).$$

Il suffit maintenant de calculer l'espérance de X_i, où l'entier i est fixé entre 1 et p. On se doute que cette espérance ne dépendra pas de l'entier i.
Pour calculer $\mathbb{E}(X_i)$, on calcule la loi de X_i qui est facile à établir :

- ▷ premièrement, $X_i(\Omega) = [\![1,6]\!]$
- ▷ deuxièmenent, comme chaque dé est parfaitement équilibré, pour tout $k \in X(\Omega)$,

$$\mathbb{P}(X_i = k) = \frac{1}{6}.$$

Conclusion,

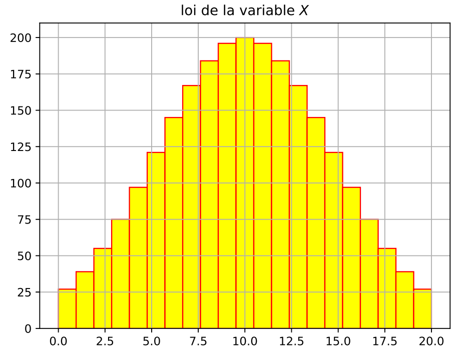
$$\mathbb{E}(X_i) = \sum_{k=1}^{6} k\ \mathbb{P}(X_i = k) = \sum_{k=1}^{6} k\ \frac{1}{6} = \frac{7}{2}.$$

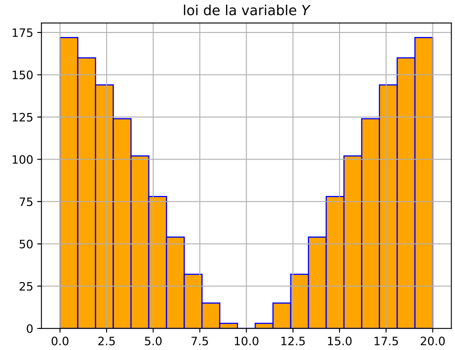
En définitive,

$$\mathbb{E}(X) = p\ \frac{7}{2}.$$

On remarque qu'il n'a pas été utile d'expliciter l'univers Ω ni même la probabilité $\mathbb{P}$, même si on peut tout exprimer. Il n'est pas rare dans un exercice de calculer une probabilité du type $\mathbb{P}(X \in A)$ sans exprimer explicitement la probabilité $\mathbb{P} : \mathscr{P}(\Omega) \longrightarrow [0,1]$, car l'univers considéré peut être un peu pénible, ainsi que la probabilité $\mathbb{P}$ associée. On consolidera tout cela au fil des exercices...

Voici deux histogrammes donnant la loi de deux variables aléatoires $X : \Omega \longrightarrow \mathbb{R}$ et $Y : \Omega \longrightarrow \mathbb{R}$.





Il semble clair au vu de la symétrie des histogrammes par rapport à la valeur centrale 10 que les variables aléatoires X et Y ont une même espérance égale à 10. Cependant, il est manifeste que les variables X et Y n'ont pas du tout la même loi, la distribution des valeurs prises par X étant plus resserrée autour de la moyenne $\mathbb{E}(X) = 10$ et la distribution des valeurs prises par Y étant cette fois-ci plus étalée autour de cette même moyenne $\mathbb{E}(Y) = 10$.
La moyenne est un critère de position qui indique autour de quelle valeur s'étalent les valeurs prises par une variable aléatoire, mais ne renseignent pas l'« étalement » autour de cette valeur.
Pour différencier les deux histogrammes, nous allons définir un critère de dispersion qui est l'écart-type ou la variance. L'écart-type de la variable X sera plus petit que l'écart-type de la variable Y.

On va plutôt manipuler la variance que l'écart-type ; c'est plus commode sans la racine carrée. Comme la variable $Y = \Big(X - \mathbb{E}(X)\Big)^2$ prend des valeurs positives, il est clair que l'espérance de Y est un réel positif, ce qui justifie que la variance de n'importe quelle variable aléatoire finie est toujours un nombre positif et ce qui justifie également la bonne définition de l'écart-type, le passage à la racine carrée étant par conséquent autorisé.

Démontrons les principales propriétés relatives à la variance et regroupées en proposition 8.
On considère dans la suite une variable aléatoire finie $X : \Omega \longrightarrow \mathbb{R}$.

• On démontre la formule de Huyghens qui deviendra la formule à dorénavant utiliser en première intention pour calculer une variance.
On pose $m = \mathbb{E}(X)$ qui est un scalaire. On utilise la fonction :

$$f : t \longmapsto (t - m)^2 = t^2 - 2mt + m^2,$$

ce qui donne :

$$\begin{aligned}
V(X) &= \mathbb{E}(f(X)) \\
&= \mathbb{E}(X^2 - 2mX + m^2) \\
&= \mathbb{E}(X^2) - 2m\,\mathbb{E}(X) + \mathbb{E}(m^2) \qquad \text{[linéarité de l'espérance]} \\
&= \mathbb{E}(X^2) - 2m \times m + m^2, \text{ car si } Y = c \text{ est constante, } \mathbb{E}(Y) = c \\
&= \mathbb{E}(X^2) - m^2 \\
&= \mathbb{E}(X^2) - \Big(\mathbb{E}(X)\Big)^2.
\end{aligned}$$

Prenez garde à la disposition des parenthèses, qui ont leur importance : le carré de l'espérance n'est pas l'espérance du carré...

• Soient a et b deux nombres réels. On pose $Y = a \cdot X + b$. On peut écrire $\mathbb{E}(Y) = a\ \mathbb{E}(X) + b$, puis

$$\begin{aligned} V(Y) &= \mathbb{E}\left(\left(Y - \mathbb{E}(Y)\right)^2\right) \\ &= \mathbb{E}\left(\left(a\ X + b - a\ \mathbb{E}(X) - b\right)^2\right) \\ &= \mathbb{E}\left(\left(a\ X - a\ \mathbb{E}(X)\right)^2\right) \\ &= \mathbb{E}\left(\left(a\ (X - \mathbb{E}(X))\right)^2\right) \\ &= a^2\ \mathbb{E}\left(\left(X - \mathbb{E}(X)\right)^2\right) \\ &= a^2 V(X). \end{aligned}$$

En passant à la racine carrée, on a la formule $\sigma(a\ X + b) = |a|\ \sigma(X)$.

• Si $\sigma(X) > 0$, alors on peut écrire en posant $m = \mathbb{E}(X)$ et $\sigma = \sigma(X)$:

$$\mathbb{E}\left(\frac{X - \mathbb{E}(X)}{\sigma(X)}\right) = \frac{1}{\sigma}\ \left(\mathbb{E}(X) - m\right) = 0,$$

et :

$$\begin{aligned} V\left(\frac{X - \mathbb{E}(X)}{\sigma(X)}\right) &= V\left(\frac{1}{\sigma}\ X - \frac{m}{\sigma}\right) \\ &= \frac{1}{\sigma^2} V(X) \\ &= \frac{1}{\sigma^2}\ \sigma^2 = 1. \end{aligned}$$

• On termine par l'inégalité de Bienaymé-Tchebychev, qui ressemble beaucoup à une inégalité de type Markov.
Soient $a > 0$ et $X : \Omega \longrightarrow \mathbb{R}$ une variable aléatoire finie. On pose dans la suite $m = \mathbb{E}(X)$.
On remarque qu'on a l'inégalité :

$$a^2\ \mathbf{1}_{\{|X-m| \geqslant a\}} \leqslant (X - m)^2.$$

En effet, si ω est un élément quelconque de l'univers Ω, on distingue deux cas :
$\longrightarrow$ si $|X(\omega) - m| \geqslant a$, alors :

$$a^2\ \mathbf{1}_{\{|X-m| \geqslant a\}}(\omega) = a^2 \leqslant \left(|X(\omega) - m|\right)^2 = (X(\omega) - m)^2$$

$\longrightarrow$ si $|X(\omega) - m| < a$, alors :

$$a^2 \ \mathbf{1}_{\{|X-m|\geqslant a\}}(\omega) = 0 \leqslant \Big(|X(\omega) - m|\Big)^2 = (X(\omega) - m)^2.$$

Il suffit maintenant de prendre l'espérance dans cette inégalité, ce qui donne :

$$\mathbb{E}\Big(a^2 \ \mathbf{1}_{\{|X-m|\geqslant a\}}\Big) \leqslant \mathbb{E}\Big((X - m)^2\Big)$$

ou encore :

$$a^2 \ \mathbb{P}(|X - m| \geqslant a) \leqslant V(X),$$

et il suffit finalement de diviser par la quantité strictement positive a^2 pour obtenir ce qu'il faut.

On passe à l'exemple 8.
Soit $X : \Omega \longrightarrow \mathbb{R}$ une variable aléatoire finie de variance nulle.
On pose $m = \mathbb{E}(X)$. On pose $Y = (X - m)^2$, de sorte que :

$$V(X) = \mathbb{E}(Y) = 0.$$

Or, la variable Y prend des valeurs positives et on peut écrire :

$$\mathbb{E}(Y) = \sum_{y\in Y(\Omega)} y \cdot \mathbb{P}(Y = y).$$

Soit y une valeur non nulle dans $Y(\Omega)$. Comme la somme définissant $\mathbb{E}(Y)$ est nulle et est exclusivement composée de termes positifs, alors chaque terme est nul.
Ainsi, la quantité $y \cdot \mathbb{P}(Y = y)$ est nulle, ainsi que $\mathbb{P}(Y = y)$, puisque y est non nul.
En conclusion,

$$\begin{array}{rcl} \mathbb{P}(Y \neq 0) & = & \mathbb{P}\left(\displaystyle\bigsqcup_{y\in Y(\Omega)\setminus\{0\}} \{Y = y\}\right) \\ & = & \displaystyle\sum_{y\in Y(\Omega)\setminus\{0\}} \mathbb{P}(Y = y) \\ & = & \displaystyle\sum_{y\in Y(\Omega)\setminus\{0\}} 0 \\ & = & 0. \end{array}$$

On en déduit en passant à l'événement contraire :

$$\mathbb{P}(Y = 0) = 1,$$

ce qui se traduit par le fait que la variable Y est presque sûrement nulle.
Autrement dit, la variable X est presque sûrement égale à la constante m (qui vaut l'espérance de X).

La réciproque non demandée dans l'exemple 8 est vraie : si $X = c$ presque sûrement, alors $(X-c)^2 = 0$ presque sûrement et donc :

$$V(X) = \mathbb{E}\left((X-c)^2\right) = \mathbb{E}(0) = 0.$$

On peut énoncer le résultat « important » suivant : si $U : \Omega \longrightarrow \mathbb{R}$ et $V : \Omega \longrightarrow \mathbb{R}$ sont deux variables aléatoires finies égales presque sûrement, alors les variables U et V ont la même loi donc la même espérance ou la même variance. C'est ce que l'on montre à travers la propriété 9.

Donnons-nous deux variables aléatoires réelles X et Y définies sur le même univers fini Ω telles que les variables X et Y soient égales presque sûrement. On en déduit :

$$\mathbb{P}(X = Y) = 1 \text{ ou encore } \mathbb{P}(X \neq Y) = 0.$$

Soit F un ensemble fini contenant $X(\Omega) \cup Y(\Omega)$. Il suffit de montrer que pour tout $x \in F$, on a :

$$\mathbb{P}(X = x) = \mathbb{P}(Y = x)$$

pour conclure que les variables X et Y ont la même loi.
En fait, on va le montrer pour tout réel x.
$\longrightarrow$ Soit x un réel. On remarque l'inclusion :

$$\{X = x\} \setminus \{Y = x\} \subset \{X \neq Y\},$$

car si ω appartient à $\{X = x\} \setminus \{Y = x\}$, alors $X(\omega) = x$ et $Y(\omega) \neq x$, donc $X(\omega) \neq Y(\omega)$.
Par conséquent, $\mathbb{P}\Big(\{X = x\} \setminus \{Y = x\}\Big) \leqslant \mathbb{P}(X \neq Y) = 0$, donc $\mathbb{P}\Big(\{X = x\} \setminus \{Y = x\}\Big) = 0$.
En intervertissant les rôles de X et de Y, on peut également écrire :

$$\mathbb{P}\Big(\{Y = x\} \setminus \{X = x\}\Big) = 0.$$

Ainsi, en utilisant le SCE $\Big(\{Y = x\}, \{Y \neq x\}\Big)$, puis le SCE $\Big(\{X = x\}, \{X \neq x\}\Big)$

$$\begin{aligned}
\mathbb{P}(X = x) &= \mathbb{P}(X = x, Y = x) + \mathbb{P}(X = x, Y \neq x) \\
&= \mathbb{P}(X = x, Y = x) + 0 \\
&= \mathbb{P}(Y = x, X = x) + 0 \\
&= \mathbb{P}(Y = x, X = x) + \mathbb{P}(Y = x, X \neq x) \\
&= \mathbb{P}(Y = x).
\end{aligned}$$

$\longrightarrow$ On en déduit alors :

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x) \\
&= \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x) + \sum_{x \in Y(\Omega) \setminus X(\Omega)} x \cdot \mathbb{P}(X = x), \text{ car si } x \notin X(\Omega),\ \mathbb{P}(X = x) = 0
\end{aligned}$$

$$\begin{aligned}
&= \sum_{x\in X(\Omega)\cup Y(\Omega)} x\cdot \mathbb{P}(X=x)\\
&= \sum_{x\in X(\Omega)\cup Y(\Omega)} x\cdot \mathbb{P}(Y=x)\\
&= \sum_{x\in Y(\Omega)} x\cdot \mathbb{P}(Y=x) + \sum_{x\in X(\Omega)\setminus Y(\Omega)} x\cdot \mathbb{P}(Y=x)\\
&= \sum_{x\in Y(\Omega)} x\cdot \mathbb{P}(Y=x), \text{ car si } x\notin Y(\Omega),\ \mathbb{P}(Y=x)=0\\
&= \mathbb{E}(Y).
\end{aligned}$$

$\longrightarrow$ Soit une fonction $f:\mathbb{R}\longrightarrow\mathbb{R}$. On remarque l'inclusion :

$$\{X=Y\}\subset\{f(X)=f(Y)\},$$

car si ω est dans l'ensemble de gauche, alors $X(\omega)=Y(\omega)$, puis :

$$f(X(\omega))=f(Y(\omega)),$$

donc l'élément ω appartient à l'ensemble de droite.
On en déduit :

$$1=\mathbb{P}(X=Y)\leqslant\mathbb{P}(f(X)=f(Y))\in[0,1],$$

et donc :

$$\mathbb{P}(f(X)=f(Y))=1.$$

Les variables $f(X)$ et $f(Y)$ sont encore égales presque sûrement. Par le point précédent, ces deux variables ont la même espérance.

$\longrightarrow$ On peut appliquer ceci à la fonction $f:t\longmapsto t^2$, ce qui donne :

$$\mathbb{E}(X^2)=\mathbb{E}(Y^2).$$

Ainsi,

$$V(X)=\mathbb{E}(X^2)-(\mathbb{E}(X))^2=\mathbb{E}(Y^2)-(\mathbb{E}(Y))^2=V(Y).$$

$\longrightarrow$ On verra un contre-exemple dans le premier point de l'exemple 9.

• Passons justement à l'exemple 9, premier point. On rappelle que la variable $X:\Omega\longrightarrow\{0,1\}$ considérée suit la loi :

$$\mathbb{P}(X=0)=\mathbb{P}(X=1)=\frac{1}{2}.$$

On en déduit :

$$\mathbb{P}(Y=0)=\mathbb{P}(1-X=0)=\mathbb{P}(X=1)=\frac{1}{2}$$

et :

$$\mathbb{P}(Y=1)=\mathbb{P}(1-X=1)=\mathbb{P}(X=0)=\frac{1}{2}.$$

Résultat des courses, la variable Y prend presque sûrement ses valeurs dans l'ensemble $\{0,1\}$: $\mathbb{P}(Y \in \{0,1\}) = 1$ et la variable Y suit la loi de Bernoulli de paramètre $\dfrac{1}{2}$: les variables X et Y suivent la même loi.
Cependant, les variables X et Y ne sont pas égales presque sûrement. Ce serait d'ailleurs plutôt le contraire : elles sont presque sûrement différentes. En fait, elles sont toujours différentes : $\{X \neq Y\} = \Omega$.

• Plaçons-nous sous les hypothèses du deuxième point.
▷ L'implication (1) $\Longrightarrow$ (2) est évidente car toute fonction polynomiale est une fonction.
▷ On suppose que pour toute fonction polynomiale $f : \mathbb{R} \longrightarrow \mathbb{R}$, on a :

$$\mathbb{E}(f(X)) = \mathbb{E}(f(Y)).$$

On pose l'ensemble fini $F = X(\Omega) \cup Y(\Omega) = \{\alpha_1, \cdots, \alpha_r\}$, ensemble à r éléments.
On peut associer aux r réels différents $\alpha_1, \cdots, \alpha_r$ les r polynômes de Lagrange $L_1, \cdots, L_r$ dans l'espace $\mathbb{R}_{r-1}[X]$.
Fixons un entier i entre 1 et r. On remarque que l'on a l'égalité :

$$L_i(X) = \mathbf{1}_{\{X=\alpha_i\}} \text{ et de même } L_i(Y) = \mathbf{1}_{\{Y=\alpha_i\}}.$$

Détaillons la première égalité.
Soit ω un élément de Ω. On distingue deux cas :
$\longrightarrow$ si $X(\omega) = \alpha_i$, alors

$$L_i(X)(\omega) = L_i(X(\omega)) = L_i(\alpha_i) = 1 = \mathbf{1}_{\{X=\alpha_i\}}(\omega)$$

$\longrightarrow$ si $X(\omega) \neq \alpha_i$, comme $X(\omega)$ appartient de toutes façons à l'ensemble F, on pose $X(\omega) = \alpha_j$, avec donc $j \neq i$. Par conséquent :

$$L_i(X)(\omega) = L_i(X(\omega)) = L_i(\alpha_j) = 0 = \mathbf{1}_{\{X=\alpha_i\}}(\omega).$$

Il suffit maintenant d'utiliser l'hypothèse avec la fonction polynomiale $f : t \longmapsto L_i(t)$, ce qui donne :

$$\begin{aligned}
\mathbb{P}(X = \alpha_i) &= \mathbb{E}\Big(\mathbf{1}_{\{X=\alpha_i\}}\Big) \\
&= \mathbb{E}(L_i(X)) \\
&= \mathbb{E}(f(X)) \\
&= \mathbb{E}(f(Y)) \\
&= \mathbb{E}(L_i(Y)) \\
&= \mathbb{E}\Big(\mathbf{1}_{\{Y=\alpha_i\}}\Big) \\
&= \mathbb{P}(Y = \alpha_i).
\end{aligned}$$

Ainsi, pour tout $x \in F$, on a :

$$\mathbb{P}(X = x) = \mathbb{P}(Y = x),$$

et dès que le réel x n'appartient pas à F, alors x n'appartient ni à $X(\Omega)$ ni à $Y(\Omega)$, d'où

$$\mathbb{P}(X = x) = 0 = \mathbb{P}(Y = x).$$

Quoiqu'il arrive, les variables X et Y suivent la même loi.

$\triangleright$ On suppose finalement que les variables X et Y ont la même loi. Soit $f : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction. Par la formule de transfert, on en déduit :

$$\begin{aligned}
\mathbb{E}(f(X)) &= \sum_{x \in X(\Omega)} f(x) \cdot \mathbb{P}(X = x) \\
&= \sum_{x \in X(\Omega) \cup Y(\Omega)} f(x) \cdot \mathbb{P}(X = x) \qquad \text{[ajout de termes nuls]} \\
&= \sum_{x \in X(\Omega) \cup Y(\Omega)} f(x) \cdot \mathbb{P}(Y = x) \qquad \text{, car } \mathbb{P}_X = \mathbb{P}_Y \\
&= \sum_{x \in Y(\Omega)} f(x) \cdot \mathbb{P}(Y = x) \qquad \text{[suppression de termes nuls]} \\
&= \mathbb{E}(f(Y)).
\end{aligned}$$

• On termine cette section par cette dernière équivalence.
On suppose que $X = Y$ presque sûrement. Soit A un événement. On remarque l'inclusion :

$$\{X \cdot \mathbf{1}_A \neq Y \cdot \mathbf{1}_A\} \subset \{X \neq Y\},$$

donc :

$$0 \leqslant \mathbb{P}\Big(X \cdot \mathbf{1}_A \neq Y \cdot \mathbf{1}_A\Big) \leqslant \mathbb{P}(X \neq Y) = 0,$$

car les variables X et Y sont presque sûrement égales. Par conséquent, les variables $X \cdot \mathbf{1}_A$ et $Y \cdot \mathbf{1}_A$ restent presque sûrement égales : ces deux variables ont donc la même loi et la même espérance.
On suppose réciproquement que pour tout événement A, on a :

$$\mathbb{E}(X \cdot \mathbf{1}_A) = \mathbb{E}(Y \cdot \mathbf{1}_A).$$

On choisit de prendre l'événement :

$$A = \{Y > X\}.$$

On pose Z la variable :

$$Z = (Y - X) \cdot \mathbf{1}_A.$$

Il est visible que la variable aléatoire Z est centrée puisque par linéarité de l'espérance :

$$\mathbb{E}(Z) = \mathbb{E}(Y \cdot \mathbf{1}_A - X \cdot \mathbf{1}_A) = \mathbb{E}(Y \cdot \mathbf{1}_A) - \mathbb{E}(X \cdot \mathbf{1}_A) = 0.$$

De plus, la variable Z prend uniquement des valeurs positives.

En effet, si ω est un élément de Ω, soit ω appartient à A, auquel cas : $Y(\omega) > X(\omega)$, puis :

$$Z(\omega) = (Y - X)(\omega) = Y(\omega) - X(\omega) > 0,$$

soit ω n'appartient pas à A auquel cas : $\mathbf{1}_A(\omega) = 0$ et

$$Z(\omega) = 0.$$

On va maintenant montrer que la variable centrée Z à valeurs dans $[0, +\infty[$ est presque sûrement nulle – ce qui est en fait un résultat général sur les variables aléatoires positives d'espérance nulle qui sont toujours presque sûrement nulles.
En effet, l'expression de l'espérance de Z donne :

$$0 = \mathbb{E}(Z) = \sum_{z \in Z(\Omega)} z \cdot \mathbb{P}(Z = z).$$

Cette somme nulle constituée exclusivement de termes positifs est en fait constituée de termes nuls :

$$\forall z \in Z(\Omega),\ z \cdot \mathbb{P}(Z = z) = 0$$

et donc :

$$\forall z \in Z(\Omega) \setminus \{0\},\ \mathbb{P}(Z = z) = 0.$$

Par conséquent,

$$\begin{aligned} \mathbb{P}(Z \neq 0) &= \mathbb{P}\Big(Z \in Z(\Omega) \setminus \{0\}\Big) \\ &= \sum_{z \in Z(\Omega)\setminus\{0\}} \mathbb{P}(Z = z) \\ &= 0. \end{aligned}$$

En prenant l'événement contraire, on obtient :

$$\mathbb{P}(Z = 0) = 1.$$

On revient alors à l'événement A. On va montrer que $\mathbb{P}(A) = 0$.
On remarque que l'on a l'inclusion :

$$A \subset \{Z \neq 0\}$$

car on a déjà montré que pour tout ω dans A, on avait $Z(\omega) > 0$.
On en déduit :

$$\mathbb{P}(A) \leqslant \mathbb{P}(Z \neq 0) = 0, \text{ donc } \mathbb{P}(A) = 0.$$

En intervertissant les rôles de X et de Y, on obtiendrait :

$$\mathbb{P}(X > Y) = 0.$$

Conclusion,

$$\mathbb{P}(X \neq Y) = \mathbb{P}\Big(\{X > Y\} \sqcup \{X < Y\}\Big) = \mathbb{P}(X > Y) + \mathbb{P}(X < Y) = 0.$$

En revenant à l'événement contraire, on conclut que l'événement $\{X=Y\}$ est presque sûr – *i.e. de probabilité égale à* 1 – et donc que les variables X et Y sont égales presque sûrement.

Les vecteurs aléatoires

Étant donné un espace probabilisé fini $(\Omega, \mathbb{P})$, il est parfois utile non pas d'effectuer une seule mesure numérique modélisable par une variable aléatoire $X : \Omega \longrightarrow \mathbb{R}$ mais plusieurs mesures numériques. Introduisons les notions à venir à travers deux exemples, avant de procéder à la formalisation.

Considérons un entier $N \geqslant 1$, puis choisissons au hasard un entier entre 1 et N. L'espace probabilisé modélisant cette situation est l'univers $\Omega = [\![1, N]\!]$ muni de la probabilité uniforme $\mathbb{P} = \mathscr{U}_N = \dfrac{1}{N}\displaystyle\sum_{k=1}^{N} \delta_k$ dont on rappelle l'expression :

$$\mathbb{P} : \left| \begin{array}{ccc} \mathscr{P}([\![1, N]\!]) & \longrightarrow & [0,1] \\ A & \longmapsto & \dfrac{\mathrm{Card}(A)}{N} \end{array} \right. .$$

Une fois l'entier ω pioché dans Ω, on s'intéresse à deux mesures différentes sur cet entier ω :

- le nombre $X(\omega)$ de facteurs premiers différents divisant l'entier ω. Par exemple, $X(1) = 0$, alors que $X(833) = 2$ ou $X(2^{2020}) = 1$;
- la somme $Y(\omega)$ des chiffres de l'entier ω en base 2. Par exemple, $Y(1) = 1$, alors que $Y(833) = 4$ ou $Y(2^{2020}) = 1$.

Voici un script Python permettant le codage récursif de ces deux variables :

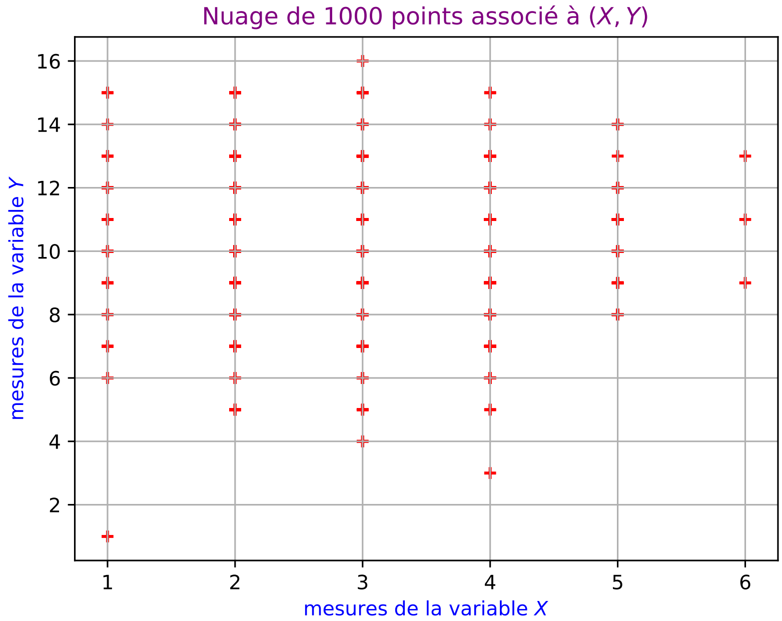
```
def X(k) :
    """ définition récursive de la variable X """
    if k==1 :
        return 0
    else :
        p=2
        while k%p!=0 :
            p+=1
        #p est le plus petit nombre premier divisant k
        exposant=0
        k_temp=k
        while k_temp%p==0 :
            k_temp=k_temp//p
            exposant+=1
        # exposant est la p-valuation dans k
        # k_temp est devenu l'entier sans les facteurs premiers égaux à p
        return 1+X(k_temp)

def Y(k):
    """ définition récursive de la variable Y """
    if k==0 :
        return 0
```

```
    else :
        return (k%2)+Y(k//2)
```

On prend par exemple $N = 1000000$ et on effectue $q = 1000$ mesures, en réalisant q fois l'expérience suivante : « piocher aléatoirement ω dans Ω, calculer $X(\omega)$ et $Y(\omega)$; tracer dans le plan affine euclidien habituel le point $M(\omega) = \Big(X(\omega), Y(\omega)\Big)$ ». Ceci nous donne ce qu'on appelle un nuage de q points.
Le script suivant permet de réaliser ce nuage de points :

```
from pylab import *
# randint(a,b) -> renvoie un entier aléatoire dans [a,b[
import os
os.chdir('D:\\Annee_2019-2020\SUIVI_A_DISTANCE\Cours_Suivi')
def Trace_Nuage(N,q) :
    """ trace le nuage de q points """
    L_omega=[randint(1,N+1) for k in  range(q)]
    LX=[X(omega) for omega in L_omega]
    LY=[Y(omega) for omega in L_omega]
    figure()
    grid()
    scatter(LX,LY,color='red',marker='+')
    xlabel(r'mesures de la variable $X$',color='blue')
    ylabel(r'mesures de la variable $Y$',color='blue')
    title(r'Nuage de '+str(q)+' points associé à $(X,Y)$',color='purple')
    savefig('nuage.eps',format='eps')
    show()

N=1000000
q=1000
Trace_Nuage(N,q)
```

Voici le nuage de points obtenu en conséquence :



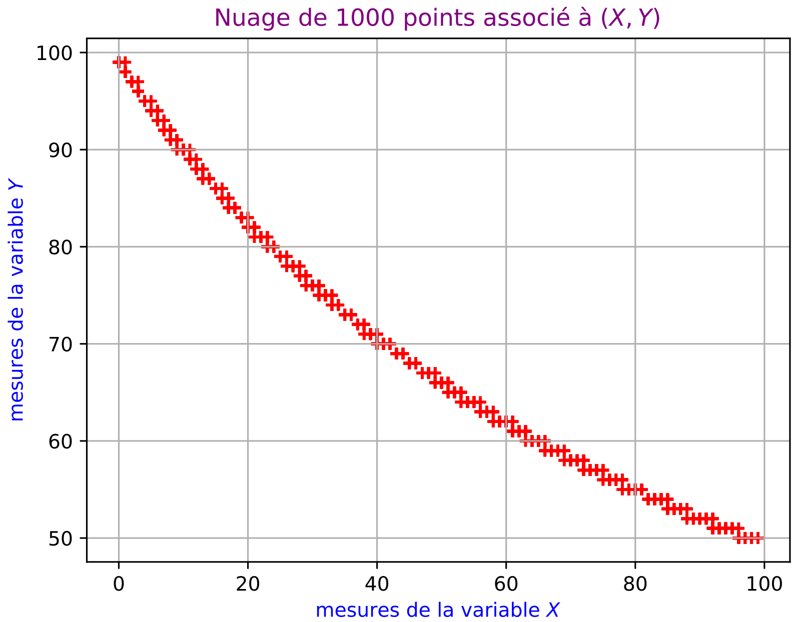
Considérons maintenant la nouvelle situation suivante : on pioche aléatoirement un nombre ω dans l'ensemble $\Omega =]0,1[$. Dans cette situation, on considère une situation complètement hors programme de première et de deuxième année, puisque l'univers considéré ici n'est pas du tout fini ni dénombrable. On pose $X(\omega) = \lfloor 100 \cdot \omega \rfloor$ et $Y(\omega) = \left\lfloor 100 \cdot \dfrac{1}{1+\omega} \right\rfloor$. Autrement dit, on vient de définir deux variables aléatoires X et Y sur l'intervalle $]0,1[$, en disant que pour tout ω dans $\Omega =]0,1[$, le nombre $X(\omega)$ est l'entier formé par les deux premières décimales après la virgule de ω, alors que le nombre $Y(\omega)$ est l'entier formé par les deux premières décimales après la virgule de $\dfrac{1}{1+\omega}$.

On effectue le tracé d'un nuage de $q = 1000$ points avec le script ci-dessous :

```
def X(omega) :
    return floor(100*omega)


def Y(omega):
    return floor(100/(1+omega))

from pylab import *
# uniform(0,1) -> renvoie un nombre aléatoire entre 0 et 1
import os
os.chdir('D:\\Annee_2019-2020\SUIVI_A_DISTANCE\Cours_Suivi')

def Trace_Nuage_bis(q) :
    """ trace le nuage de q points """
    L_omega=[uniform(0,1) for k in  range(q)]
    LX=[X(omega) for omega in L_omega]
    LY=[Y(omega) for omega in L_omega]
```

```
    figure()
    grid()
    scatter(LX,LY,color='red',marker='+')
    xlabel(r'mesures de la variable $X$',color='blue')
    ylabel(r'mesures de la variable $Y$',color='blue')
    title(r'Nuage de '+str(q)+' points associé à $(X,Y)$',color='purple')
    savefig('nuage_bis.eps',format='eps')
    show()

q=1000
Trace_Nuage_bis(q)
```

Voici le nuage de points obtenu :



On remarque que l'on a deux nuages de points de formes vraiment différentes. Dans la première situation, on a l'impression qu'il n'y a pas franchement de lien entre les variables X et Y, alors que dans la deuxième situation, le lien est visuellement clair entre X et Y : les valeurs de Y décroissent globalement en fonction des valeurs de X.
On dira dans la première situation que les variables X et Y semblent indépendantes avec une covariance $\mathrm{Cov}(X,Y)$ proche de 0, alors que dans la deuxième situation, les variables X et Y ne sont pas indépendantes et la covariance $\mathrm{Cov}(X,Y)$ sera strictement négative.

Formalisons tout cela.

Étant donné un espace probabilisé fini $(\Omega,\mathbb{P})$ et un entier $d \geqslant 1$, on peut considérer une variable aléatoire $\vec{X} : \Omega \longrightarrow \mathbb{R}^d$ à valeurs dans l'espace vectoriel $\mathbb{R}^d$. Pour tout $\omega \in \Omega$, on pose le vecteur :

$$\vec{X}(\omega) = \Big(X_1(\omega),\cdots,X_d(\omega)\Big).$$

La variable aléatoire $\vec{X} : \Omega \longrightarrow \mathbb{R}^d$ aussi appelée vecteur aléatoire peut être mis sous la forme :

$$\vec{X} = (X_1, \cdots, X_d),$$

pour signifier que l'on a :

$$\forall \omega \in \Omega,\ \vec{X}(\omega) = (X_1, \cdots, X_d)(\omega) = \Big(X_1(\omega), \cdots, X_d(\omega)\Big).$$

On dispose alors de d variables aléatoires $X_i : \Omega \longrightarrow \mathbb{R}$ à valeurs réelles.
Lorsque $d = 2$, on parle alors de couple aléatoire $(X, Y) : \Omega \longrightarrow \mathbb{R}^2$ qui est une fonction de la forme :

$$(X, Y) : \left| \begin{array}{ccl} \Omega & \longrightarrow & \mathbb{R}^2 \\ \omega & \longmapsto & (X, Y)(\omega) = \Big(X(\omega), Y(\omega)\Big) \end{array} \right. .$$

Tout ce qui sera dit sur les couples (X, Y) de variables aléatoires peut facilement être généralisé à un d-uplet $(X_1, \cdots, X_d)$ de variables aléatoires. On ne détaillera essentiellement ce qui se passe que pour les couples aléatoires, le plus souvent.

Soient $X : \Omega \longrightarrow \mathbb{R}$ et $Y : \Omega \longrightarrow \mathbb{R}$ deux variables aléatoires finies. On pose :

$$X(\Omega) = \{x_1, \cdots, x_p\} \text{ et } Y(\Omega) = \{y_1, \cdots, y_q\}.$$

On définit ainsi le couple aléatoire :

$$(X, Y) \left| \begin{array}{ccl} \Omega & \longrightarrow & X(\Omega) \times Y(\Omega) \\ \omega & \longmapsto & \Big(X(\omega), Y(\omega)\Big) \end{array} \right. .$$

On peut de nouveau parler de la loi $\mathbb{P}_{(X,Y)}$ du couple (X, Y). Il s'agit de la probabilité suivante :

$$\mathbb{P}_{(X,Y)} : \left| \begin{array}{ccl} \mathscr{P}\Big(X(\Omega) \times Y(\Omega)\Big) & \longrightarrow & [0, 1] \\ A & \longmapsto & \mathbb{P}\Big((X, Y) \in A\Big) \end{array} \right. .$$

Comme l'ensemble $X(\Omega) \times Y(\Omega)$ est fini et compte $p \times q$ éléments – c'est-à-dire les couples (x_i, y_j), lorsque (i, j) parcourt $[\![1, p]\!] \times [\![1, q]\!]$, la probabilité $\mathbb{P}_{(X,Y)}$ est uniquement déterminée par les images des singletons.
En posant pour tout $(i, j) \in [\![1, p]\!] \times [\![1, q]\!]$,

$$\begin{array}{rcl} \pi_{i,j} & = & \mathbb{P}_{(X,Y)}(\{(x_i, y_j)\}) \\ & = & \mathbb{P}\Big((X, Y) \in \{(x_i, y_j)\}\Big) \\ & = & \mathbb{P}\Big((X, Y) = (x_i, y_j)\Big) \\ & = & \mathbb{P}\Big(\{X = x_i\} \cap \{Y = y_j\}\Big) \\ & = & \mathbb{P}(X = x_i, Y = y_j) \end{array}$$

alors la famille $\Pi = \Big(\pi_{i,j}\Big)_{1\leqslant i\leqslant p, 1\leqslant j\leqslant q}$ est une famille de nombres réels de somme égale à 1 et pour tout événement $A \subset X(\Omega)\times Y(\Omega)$, on a :

$$\mathbb{P}_{(X,Y)}(A) = \mathbb{P}\Big((X,Y)\in A\Big) = \sum_{1\leqslant i\leqslant p,\ 1\leqslant j\leqslant q \text{ et } (x_i,y_j)\in A} \pi_{i,j}.$$

On peut ranger les nombres de la famille Π dans un tableau come suit :

$X \setminus Y$	y_1	$\cdots$	y_j	$\cdots$	y_q
x_1	$\pi_{1,1}$	$\cdots$	$\pi_{1,j}$	$\cdots$	$\pi_{1,q}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	$\pi_{i,1}$	$\cdots$	$\pi_{i,j}$	$\cdots$	$\pi_{i,q}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_p	$\pi_{p,1}$	$\cdots$	$\pi_{p,j}$	$\cdots$	$\pi_{p,q}$

On remarque que les familles $\mathscr{F} = \Big(\{X = x_i\}\Big)_{1\leqslant i\leqslant p}$ et $\mathscr{G} = \Big(\{Y = y_j\}\Big)_{1\leqslant j\leqslant q}$ constituent deux systèmes complets d'événements. En utilisant la formule des probabilités totales, on peut écrire :

- pour tout $i \in [\![1,p]\!]$,

$$\begin{aligned}\mathbb{P}(X = x_i) &= \sum_{j=1}^{q} \mathbb{P}(X = x_i, Y = y_j)\\ &= \sum_{j=1}^{q} \pi_{i,j}.\end{aligned}$$

- pour tout $j \in [\![1,q]\!]$,

$$\begin{aligned}\mathbb{P}(Y = y_j) &= \sum_{i=1}^{p} \mathbb{P}(Y = y_j, X = x_i)\\ &= \sum_{i=1}^{p} \mathbb{P}(X = x_i, Y = y_j)\\ &= \sum_{i=1}^{p} \pi_{i,j}.\end{aligned}$$

On en déduit que l'on obtient la loi de X en effectuant des sommations par lignes et que l'on obtient la loi de Y en effectuant des sommations par colonnes. Ces informations se trouvent donc en marge du tableau de la loi $\mathbb{P}_{(X,Y)}$:

$X \setminus Y$	y_1	$\cdots$	y_j	$\cdots$	y_q	loi de X
x_1	$\pi_{1,1}$	$\cdots$	$\pi_{1,j}$	$\cdots$	$\pi_{1,q}$	$\mathbb{P}(X = x_1) = \sum_{j=1}^{q} \pi_{1,j}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	$\pi_{i,1}$	$\cdots$	$\pi_{i,j}$	$\cdots$	$\pi_{i,q}$	$\mathbb{P}(X = x_i) = \sum_{j=1}^{q} \pi_{i,j}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_p	$\pi_{p,1}$	$\cdots$	$\pi_{p,j}$	$\cdots$	$\pi_{p,q}$	$\mathbb{P}(X = x_p) = \sum_{j=1}^{q} \pi_{p,j}$
loi de Y	$\mathbb{P}(Y = y_1) = \sum_{i=1}^{p} \pi_{i,1}$	$\cdots$	$\mathbb{P}(Y = y_j) = \sum_{i=1}^{p} \pi_{i,j}$	$\cdots$	$\mathbb{P}(Y = y_q) = \sum_{i=1}^{p} \pi_{i,q}$	

La probabilité $\mathbb{P}_{(X,Y)}$ s'appelle la loi conjointe du couple (X, Y), alors que les lois $\mathbb{P}_X$ et $\mathbb{P}_Y$ des variables X et Y s'appellent les lois marginales : elles se lisent en marge du tableau de la loi conjointe.
Il y a plus d'information dans la loi conjointe que dans les lois marginales. Par exemple, si X et Y suivent la loi de Rademacher de paramètre $\frac{1}{2}$, on peut avoir les deux tableaux suivants pour le couple (X, Y) avec la loi de Rademacher $\mathscr{R}\left(\frac{1}{2}\right)$ pour chaque loi marginale $\mathbb{P}_X$ et $\mathbb{P}_Y$:

$X \setminus Y$	-1	1
-1	0	$\frac{1}{2}$
1	$\frac{1}{2}$	0

ou bien

$X \setminus Y$	-1	1
-1	$\frac{1}{3}$	$\frac{1}{6}$
1	$\frac{1}{6}$	$\frac{1}{3}$

.

La somme sur chaque ligne ou colonne pour les deux tableaux vaut invariablement $\frac{1}{2}$, les variables X et Y prenant leurs valeurs dans l'ensemble à deux éléments $\{-1, 1\}$. Il y a énormément d'autres exemples possibles.
On peut aussi parler de loi conditionnelle de X sachant l'événement $\{Y \in B\}$, à condition que la probabilité de l'événement $\{Y \in B\}$ soit non nulle comme la probabilité :

$$\mathbb{P}_{X \mid \{Y \in B\}} : \left| \begin{array}{ccc} \mathscr{P}(X(\Omega)) & \longrightarrow & [0,1] \\ A & \longmapsto & \dfrac{\mathbb{P}(X \in A, Y \in B)}{\mathbb{P}(Y \in B)} \end{array} \right. .$$

On définit ensuite la covariance $\mathrm{Cov}(X, Y)$ comme :

$$\mathrm{Cov}(X, Y) = \mathbb{E}\Big((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\Big).$$

La propriété 10 nous dit qu'on a une formule un peu plus simple pour calculer cette covariance – FORMULE DE HUYGHENS.

Posons $m = \mathbb{E}(X)$ et $\mu = \mathbb{E}(Y)$, de sorte que :

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}\Big((X-m)(Y-\mu)\Big) \\
&= \mathbb{E}\Big(XY - m\cdot Y - \mu\cdot X + m\cdot\mu\Big) \\
&= \mathbb{E}(XY) - m\,\mathbb{E}(Y) - \mu\,\mathbb{E}(X) + \mathbb{E}(m\cdot\mu) \qquad \text{[linéarité de l'espérance]} \\
&= \mathbb{E}(XY) - m\,\mu - \mu\,m + m\cdot\mu \qquad \text{, car si } Z = c \text{ est constante, } \mathbb{E}(Z) = \mathbb{E}(c) = c \\
&= \mathbb{E}(XY) - m\,\mu \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y).
\end{aligned}$$

La définition 10 fait écho à la définition 7.
On dira que deux variables $X : \Omega \longrightarrow \mathbb{R}$ et $Y : \Omega \longrightarrow \mathbb{R}$ sont indépendantes si pour tout $(x,y) \in X(\Omega)\times Y(\Omega)$, les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants.
Comme pour tout $x \in \mathbb{R}\setminus X(\Omega)$, on a $\mathbb{P}(X = x) = 0$, puis $\mathbb{P}(X = x, Y = y) = 0$, et de même pour tout $y \in \mathbb{R}\setminus Y(\Omega)$, on a $\mathbb{P}(Y = y) = 0$, puis $\mathbb{P}(X = x, Y = y) = 0$, la définition de l'indépendance entre deux variables X et Y donnée en définition 10 est équivalente à :

$$\forall (x,y) \in \mathbb{R}^2,\ \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\cdot\mathbb{P}(Y = y).$$

La définition de la mutuelle indépendance entre plusieurs variables aléatoires peut être déroutante. On se souvient que pour l'indépendance mutuelle entre p événements $A_1, \cdots, A_p$, il faut prendre n'importe quelle partie I incluse dans $[\![1,p]\!]$ puis montrer que :

$$\mathbb{P}\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I}\mathbb{P}(A_i).$$

Or, dans la définition de la mutuelle indépendance de p variables aléatoires $X_1, \cdots, X_p$, rien de tel, pas de partie I incluse dans $[\![1,p]\!]$.
Pour dissiper le doute, on va déjà montrer que la définition de la mutuelle indépendance donnée en définition 10 (notée propriété $\mathscr{P}$ dans la suite) est équivalente à la propriété $\mathscr{Q}$:

$\mathscr{Q}$: pour toute partie finie I incluse dans $[\![1,p]\!]$, pour tout uplet $(x_i)_{i\in I}$ de réels, alors :

$$\mathbb{P}\left(\bigcap_{i\in I}\{X_i = x_i\}\right) = \prod_{i\in I}\mathbb{P}(X_i = x_i).$$

On commence par démontrer le plus facile, à savoir $\boxed{\mathscr{Q} \Longrightarrow \mathscr{P}.}$
Supposons la propriété $\mathscr{Q}$ vérifiée. Prenons alors la partie $I = [\![1,p]\!]$, puis n'importe quels réels x_i avec $x_i \in X_i(\Omega)$, pour tout i entre 1 et p. Alors :

$$\mathbb{P}\Big((X_1,\cdots,X_p) = (x_1,\cdots,x_p)\Big) = \mathbb{P}\left(\bigcap_{i\in I}\{X_i = x_i\}\right)$$

$$= \prod_{i\in I} \mathbb{P}(X_i = x_i)$$
$$= \prod_{i=1}^{p} \mathbb{P}(X_i = x_i).$$

On obtient sans difficulté la propriété $\mathscr{P}$.

On montre maintenat l'autre implication $\boxed{\mathscr{P} \Longrightarrow \mathscr{Q}.}$
Supposons la propriété $\mathscr{P}$.
On fixe une partie I incluse dans $[\![1,p]\!]$.
On fixe un uplet $(x_i)_{i\in I}$ de nombres réels.
On distingue deux cas.

- S'il existe $i \in I$ tel que $x_i \notin X_i(\Omega)$, alors l'événement $\{X_i = x_i\}$ est impossible, donc vide et de probabilité nulle. D'autre part, on a bien évidemment l'inclusion :
$$\bigcap_{k\in I} \{X_k = x_k\} \subset \{X_i = x_i\},$$
donc dans ce cas :
$$\mathbb{P}\left(\bigcap_{k\in I} \{X_k = x_k\}\right) = 0 = \prod_{k\in I} \mathbb{P}(X_k = x_k),$$
car ce dernier produit contient un facteur nul : le facteur $\mathbb{P}(X_i = x_i)$.
- On se place maintenant dans la situation où pour tout $i \in I$, on a :
$$x_i \in X_i(\Omega).$$
La suite de la démonstration va être un peu technique. Pour alléger les choses, on va poser :
$$J = [\![1,p]\!] \setminus I \text{ et le produit cartésien } F = \prod_{j\in J} X_j(\Omega).$$
On remarque alors que l'on a l'égalité suivante :
$$\bigcap_{i\in I} \{X_i = x_i\} = \bigsqcup_{(y_j)_{j\in F}} \Big\{\forall i \in I,\ X_i = x_i \text{ et } \forall j \in J,\ X_j = y_j\Big\}.$$
Par σ-additivité de la probabilité $\mathbb{P}$, on peut écrire :
$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i\in I} \{X = x_i\}\right) &= \mathbb{P}\left(\bigsqcup_{(y_j)_{j\in F}} \Big\{\forall i \in I,\ X_i = x_i \text{ et } \forall j \in J,\ X_j = y_j\Big\}\right) \\
&= \sum_{(y_j)\in F} \mathbb{P}\Big((X_i)_{i\in I} = (x_i)_{i\in I} \text{ et } (X_j)_{j\in J} = (y_j)_{j\in F}\Big) \\
&= \sum_{(y_j)\in F} \left(\prod_{i\in I} \mathbb{P}(X_i = x_i)\right) \times \left(\prod_{j\in J} \mathbb{P}(X_j = y_j)\right) \qquad \text{[propriété } \mathscr{P}\text{]}
\end{aligned}$$

$$
\begin{aligned}
&= \left(\prod_{i\in I}\mathbb{P}(X_i=x_i)\right)\times\left(\sum_{(y_j)\in F}\left(\prod_{j\in J}\mathbb{P}(X_j=y_j)\right)\right)\\
&= \left(\prod_{i\in I}\mathbb{P}(X_i=x_i)\right)\times\left(\prod_{j\in J}\sum_{y_j\in X_j(\Omega)}\mathbb{P}(X_j=y_j)\right) \qquad \left[\begin{array}{c}\text{somme rectangulaire}\\ =\\ \text{produit des sommes}\end{array}\right]\\
&= \left(\prod_{i\in I}\mathbb{P}(X_i=x_i)\right)\times\left(\prod_{j\in J}\mathbb{P}(X_j\in X_j(\Omega))\right) \qquad [\sigma\text{-additivité}]\\
&= \left(\prod_{i\in I}\mathbb{P}(X_i=x_i)\right)\times\left(\prod_{j\in J}1\right) \qquad \text{, car } \{X_j\in X_j(\Omega)\}=\Omega\\
&= \left(\prod_{i\in I}\mathbb{P}(X_i=x_i)\right).
\end{aligned}
$$

On obtient quoiqu'il arrive la propriété $\mathscr{Q}$.

Il est moins pénible de vérifier seulement la propriété $\mathscr{P}$, ce qu'on fera systématiquement. Pour finir sur le sujet, on remarque que la propriété $\mathscr{Q}$ implique que la formule de la propriété $\mathscr{P}$ n'est pas seulement valable pour tout p-uplet $(x_1,\cdots,x_p)\in\displaystyle\prod_{i=1}^{p}X_i(\Omega)$ mais pour tout p-uplet $(x_1,\cdots,x_p)$ dans $\mathbb{R}^p$.

On démontre maintenant la propriété 11, point par point.
• Soient A une partie de $X(\Omega)$ et B une partie de $Y(\Omega)$.
On en déduit :

$$
\begin{aligned}
\mathbb{P}\Big((X,Y)\in A\times B\Big) &= \mathbb{P}(X\in A, Y\in B)\\
&= \mathbb{P}\left(\bigsqcup_{(a,b)\in A\times B}\{X=a,Y=b\}\right)\\
&= \sum_{(a,b)\in A\times B}\mathbb{P}(X=a,Y=b) \qquad [\sigma\text{-additivité}]\\
&= \sum_{(a,b)\in A\times B}\mathbb{P}(X=a)\cdot\mathbb{P}(Y=b) \qquad [\text{indépendance}]\\
&= \sum_{a\in A}\sum_{b\in B}\mathbb{P}(X=a)\cdot\mathbb{P}(Y=b)\\
&= \sum_{a\in A}\mathbb{P}(X=a)\cdot\left(\sum_{b\in B}\mathbb{P}(Y=b)\right)\\
&= \sum_{a\in A}\mathbb{P}(X=a)\cdot\mathbb{P}\left(\bigsqcup_{b\in B}\{Y=b\}\right) \qquad [\sigma\text{-additivité}]\\
&= \sum_{a\in A}\mathbb{P}(X=a)\cdot\mathbb{P}\left(Y\in B\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{P}(Y \in B) \cdot \sum_{a \in A} \mathbb{P}(X = a) \\
&= \mathbb{P}(Y \in B) \cdot \mathbb{P}\left(\bigsqcup_{a \in A} \{X = a\} \right) \qquad [\sigma\text{-additivité}] \\
&= \mathbb{P}(Y \in B) \cdot \mathbb{P}(X \in A) \\
&= \mathbb{P}(X \in A) \cdot \mathbb{P}(Y = B).
\end{aligned}
$$

• Ensuite, on va montrer la formule de transfert pour les vecteurs aléatoires. On considère un vecteur aléatoire $\vec{Z} = (Z_1, \cdots, Z_d) \longrightarrow \mathbb{R}^d$, où les variables aléatoires $Z_i : \Omega \longrightarrow \mathbb{R}$ sont finies mais pas forcément indépendantes. On considère une fonction $f : \mathbb{R}^d \longrightarrow \mathbb{R}$, la fonction f pouvant en fait être uniquement définie sur $\vec{Z}(\Omega)$.
On va utiliser le SCE : $\left(\left\{ \vec{Z} = (z_1, \cdots, z_d) \right\} \right)_{(z_1, \cdots, z_d) \in \vec{Z}(\Omega)}$ et commencer par montrer la formule :

$$f(\vec{Z}) = \sum_{\vec{z} \in \vec{Z}(\Omega)} f(\vec{z}) \cdot \mathbf{1}_{\{\vec{Z} = \vec{z}\}}.$$

En effet, soit ω un élément de Ω. On pose le vecteur $\vec{\tau} = \vec{Z}(\omega)$.
On fixe maintenant un autre vecteur $\vec{z}$ dans $\vec{Z}(\Omega)$. On distingue deux cas.
$\longrightarrow$ Si $\vec{z} = \vec{\tau}$, alors :

$$\mathbf{1}_{\{\vec{Z} = \vec{z}\}}(\omega) = 1 \text{ et } f(\vec{z}) = f\Big(\vec{Z}(\omega)\Big) = f(\vec{Z})(\omega).$$

$\longrightarrow$ Si $\vec{z} \neq \vec{\tau}$, alors :

$$\mathbf{1}_{\{\vec{Z} = \vec{z}\}}(\omega) = 0.$$

On en déduit que :

$$
\begin{aligned}
\left(\sum_{\vec{z} \in \vec{Z}(\Omega)} f(\vec{z}) \cdot \mathbf{1}_{\{\vec{Z} = \vec{z}\}} \right)(\omega) &= \sum_{\vec{z} \in \vec{Z}(\Omega)} f(\vec{z}) \cdot \mathbf{1}_{\{\vec{Z} = \vec{z}\}}(\omega) \\
&= f(\tau) \cdot \mathbf{1}_{\{\vec{Z} = \vec{\tau}\}}(\omega) \\
&= f(\tau) \\
&= f(\vec{Z})(\omega).
\end{aligned}
$$

Il suffit alors de prendre l'espérance dans cette égalité, ce qui donne :

$$
\begin{aligned}
\mathbb{E}\Big(f(\vec{Z})\Big) &= \mathbb{E}\left(\sum_{\vec{z} \in \vec{Z}(\Omega)} f(\vec{z}) \cdot \mathbf{1}_{\{\vec{Z} = \vec{z}\}} \right) \\
&= \sum_{\vec{z} \in \vec{Z}(\Omega)} f(\vec{z}) \cdot \mathbb{E}\Big(\mathbf{1}_{\{\vec{Z} = \vec{z}\}}\Big) \qquad [\text{linéarité de l'espérance}] \\
&= \sum_{\vec{z} \in \vec{Z}(\Omega)} f(\vec{z}) \cdot \mathbb{P}\Big(\vec{Z} = \vec{z}\Big).
\end{aligned}
$$

Lorsque $d = 2$, on obtient la formule de transfert pour les couples de variables aléatoires.
On peut alors redémontrer la linéarité de l'espérance grâce à cette formule de transfert. Voici le détail : si U et V sont deux v.a.r. définies de Ω vers $\mathbb{R}$, si $\lambda \in \mathbb{R}$ est un scalaire, en posant la fonction

$$f : \left| \begin{array}{ccc} \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x,y) & \longmapsto & \lambda \cdot x + y \end{array} \right. ,$$

on en déduit :

$$\begin{array}{rcll}
\mathbb{E}(\lambda \cdot U + V) & = & \mathbb{E}\Big(f(U,V)\Big) & \\
& = & \displaystyle\sum_{(x,y)\in(U,V)(\Omega)} f(x,y) \cdot \mathbb{P}(U = x, V = y) & \text{[formule de transfert]} \\
& = & \displaystyle\sum_{(x,y)\in U(\Omega)\times V(\Omega)} f(x,y) \cdot \mathbb{P}(U = x, V = y) & \text{[ajout de termes nuls]} \\
& = & \displaystyle\sum_{(x,y)\in U(\Omega)\times V(\Omega)} (\lambda \cdot x + y) \cdot \mathbb{P}(U = x, V = y) & \\
& = & \lambda \displaystyle\sum_{x\in U(\Omega)} x \left(\sum_{y\in V(\Omega)} \mathbb{P}(U = x, V = y) \right) + \sum_{y\in V(\Omega)} y \left(\sum_{x\in U(\Omega)} \mathbb{P}(V = y, U = x) \right) & \\
& = & \lambda \displaystyle\sum_{x\in U(\Omega)} x \cdot \mathbb{P}(U = x) + \sum_{y\in V(\Omega)} y \cdot \mathbb{P}(V = y) & \text{[formule probabilités totales]} \\
& = & \lambda \cdot \mathbb{E}(U) + \mathbb{E}(V). &
\end{array}$$

On démontre maintenant la chose importante suivante : lorsque l'on a affaire à deux variables indépendantes, alors l'espérance du produit est toujours égale au produit des espérances.
En effet, en utilisant le fait que les variables aléatoires finies $X : \Omega \longrightarrow \mathbb{R}$ et $Y : \Omega \longrightarrow \mathbb{R}$ sont indépendantes et la formule de transfert avec la fonction $f : \left| \begin{array}{ccc} \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x,y) & \longmapsto & x \cdot y \end{array} \right.$:

$$\begin{array}{rcll}
\mathbb{E}(X \cdot Y) & = & \mathbb{E}\Big(f(X,Y)\Big) & \\
& = & \displaystyle\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} f(x,y) \cdot \mathbb{P}(X = x, Y = y) & \text{[formule de transfert]} \\
& = & \displaystyle\sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} x \cdot y \cdot \mathbb{P}(X = x) = \mathbb{P}(Y = y) & \text{[indépendance]} \\
& = & \displaystyle\sum_{x\in X(\Omega)} \left(\sum_{y\in Y(\Omega)} x \cdot y \cdot \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y) \right) & \\
& = & \displaystyle\sum_{x\in X(\Omega)} x \cdot \mathbb{P}(X = x) \cdot \left(\sum_{y\in Y(\Omega)} y \cdot \mathbb{P}(Y = y) \right) & \\
& = & \displaystyle\sum_{x\in X(\Omega)} x \cdot \mathbb{P}(X = x) \cdot \mathbb{E}(Y) &
\end{array}$$

$$\begin{aligned} &= \mathbb{E}(Y) \cdot \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x) \\ &= \mathbb{E}(Y) \cdot \mathbb{E}(X) \\ &= \mathbb{E}(X) \cdot \mathbb{E}(Y). \end{aligned}$$

Par conséquent, dès que deux variables X et Y sont indépendantes, alors la covariance est nulle en utilisant la formule de Huyghens – propriété 10 – et ce que l'on vient de montrer.
On verra en exemple 11 que la réciproque est fausse. Il est tout à fait possible d'avoir un couple (U, V) de variables aléatoires finies avec $\text{Cov}(U, V) = 0$ mais U et V non indépendantes.

• Le point suivant est un cas particulier du lemme des coalitions. Montrons ce cas particulier pour se mettre dans l'ambiance.
On sait que les variables X et Y sont indépendantes. On se donne deux fonctions $f : \mathbb{R} \longrightarrow \mathbb{R}$ et $g : \mathbb{R} \longrightarrow \mathbb{R}$, ces deux fonctions pouvant comme d'habitude être seulement définies respectivement sur les ensembles $X(\Omega)$ et $Y(\Omega)$.
Soient deux réels a et b. Il s'agit de montrer que :

$$\mathbb{P}\Big(f(X) = a, g(Y) = b\Big) = \mathbb{P}(\Big(f(X) = a\Big) \cdot \mathbb{P}\Big(g(Y) = b\Big).$$

On remarque que l'on a les égalités :

$$\Big\{f(X) = a\Big\} = \Big\{X \in f^{[-1]}(\{a\})\Big\} \text{ et } \Big\{g(Y) = b\Big\} = \Big\{Y \in g^{[-1]}(\{b\})\Big\}.$$

Posons alors les deux parties de $\mathbb{R}$:

$$A = f^{[-1]}(\{a\}) \text{ et } B = g^{[-1]}(\{b\}).$$

On en déduit :

$$\begin{aligned} \mathbb{P}\Big(f(X) = a, g(Y) = b\Big) &= \mathbb{P}(X \in A, Y \in B) \\ &= \mathbb{P}(X \in A) \times \mathbb{P}(Y \in B) \qquad \text{[premier point propriété 11]} \\ &= \mathbb{P}(f(X) = a) \times \mathbb{P}(g(Y) = b). \end{aligned}$$

• Arrêtons-nous un instant – en fait beaucoup plus qu'un instant ...– sur la propriété suivante connue sous le nom du lemme des coalitions.
Écoutons le dialogue qui peut s'établir entre un probabiliste et un ensembliste.
Le probabiliste prend connaissance de l'énoncé du lemme des coalitions et après quelques secondes de réflexion, dit à son interlocuteur :
– « Mais c'est parfaitement évident !! Chaque variable Y_j dépend de certaines variables X_k et si $j_1, \cdots, j_s$ sont des indices différents, les variables X_k qui interviennent dans la définition de Y_{j_1} ou de Y_{j_2}, etc jusqu'à la variable Y_{j_s} sont différentes car on a affaire à une partition. L'information présente à travers Y_{j_1} dépendra d'informations présentes dans certains X_k et l'information présente dans Y_{j_2} dépendra d'autres informations dépendant de certains autres

X_ℓ. Il est donc tout à fait normal de récolter encore des informations indépendantes concernant les Y_j. »
– « Tu es sûr ? Il vaut peut-être mieux faire les calculs. Juste pour être sûr. Cela me rassurerait. »
– « Ok, si ça peut te rassurer. »

Sur ce, l'ensembliste prend un crayon et un papier et commence à effectuer sa démonstration du lemme des coalitions. Voici sa démonstration ci-après retranscrite, agrémentée de commentaires pour la rendre un peu plus intelligible.

La démonstration est assez technique et met en place des notations assez lourdes. Allons-y !

On se donne p variables $X_1, \cdots, X_p$ aléatoires définies de l'univers fini Ω vers $\mathbb{R}$.
On se donne une partition $(U_1, \cdots, U_s)$ de $[\![1, p]\!]$.
Pour tout entier j entre 1 et s, on pose l'ensemble :

$$F_j = \prod_{k \in U_j} X_k(\Omega).$$

On se donne des fonctions $f_1, \cdots, f_s$ définies respectivement de $F_1, \cdots, F_s$ vers $\mathbb{R}$.
On forme ainsi s variables aléatoires $Y_1, \cdots, Y_s$, avec :

$$Y_j = f_j\Big((X_k)_{k \in U_j}\Big) : \Omega \longrightarrow \mathbb{R}.$$

Avec ces notations, pour tout entier j entre 1 et s, on fixe un réel α_j. Il s'agit de montrer que :

$$\mathbb{P}\left(\bigcap_{j=1}^{s} \{Y_j = \alpha_j\}\right) = \prod_{j=1}^{s} \mathbb{P}(Y_j = \alpha_j),$$

pour avoir la mutuelle indépendance des variables $Y_1, \cdots, Y_s$.
Pour tout $j \in [\![1, s]\!]$, on note

$$A_j = f_j^{[-1]}(\{\alpha_j\}).$$

On peut commencer le calcul en remarquant que pour tout $j \in [\![1, s]\!]$:

$$\{Y_j = \alpha_j\} = \Big\{(X_k)_{k \in U_j} \in A_j\Big\}.$$

On en déduit :

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{j=1}^{s} \{Y_j = \alpha_j\}\right) &= \mathbb{P}\left(\bigcap_{j=1}^{s} \Big\{(X_k)_{k \in U_j} \in A_j\Big\}\right) \\
&= \mathbb{P}\left(\bigsqcup_{\substack{(x_1, \cdots, x_p) \in \prod_{i=1}^{p} X_i(\Omega) \\ \text{tel que} \\ \forall j \in [\![1, s]\!],\ \vec{a}_j = (x_k)_{k \in U_j} \in A_j}} \Big\{X_1 = x_1, X_2 = x_2, \cdots, X_p = x_p\Big\}\right)
\end{aligned}$$

$$= \sum_{\substack{(x_1,\cdots,x_p)\in \prod_{i=1}^{p} X_i(\Omega) \\ \text{tel que} \\ \forall j\in [\![1,s]\!],\ \vec{a}_j=(x_k)_{k\in U_j}\in A_j}} \mathbb{P}\Big(X_1=x_1, X_2=x_2,\cdots,X_p=x_p\Big) \qquad \color{red}{[\sigma\text{-additivité}]}$$

$$= \sum_{\substack{(x_1,\cdots,x_p)\in \prod_{i=1}^{p} X_i(\Omega) \\ \text{tel que} \\ \forall j\in [\![1,s]\!],\ \vec{a}_j=(x_k)_{k\in U_j}\in A_j}} \left(\prod_{k=1}^{p} \mathbb{P}(X_k=x_k)\right) \qquad \color{red}{[\text{indépendance des } X_k]}$$

$$= \sum_{(\vec{a}_1,\cdots,\vec{a}_s)\in \prod_{j=1}^{s} A_j} \prod_{j=1}^{s} \left(\prod_{k\in U_j} \mathbb{P}(X_k=x_k)\right) \qquad \color{red}{[\text{utilisation de la partition}]}$$

$$= \sum_{(\vec{a}_1,\cdots,\vec{a}_s)\in \prod_{j=1}^{s} A_j} \prod_{j=1}^{s} \mathbb{P}\left(\Big(X_k\Big)_{k\in U_j}=\vec{a}_j\right) \qquad \color{red}{[\text{indépendance des } X_k]}$$

$$= \prod_{j=1}^{s} \sum_{\vec{a}_j\in A_j} \mathbb{P}\left(\Big(X_k\Big)_{k\in U_j}=\vec{a}_j\right) \qquad \color{red}{[\text{calculs avec sommes rectangulaires}]}$$

$$= \prod_{j=1}^{s} \mathbb{P}\left(\Big(X_k\Big)_{k\in U_j}\in A_j\right) \qquad \color{red}{[\sigma\text{-additivité}]}$$

$$= \prod_{j=1}^{s} \mathbb{P}\left(Y_j=\alpha_j\right).$$

Voilà, le lemme des coalitions est établi.

On en déduit que si r est un entier entre 1 et $p-1$, si $\varphi : \mathbb{R}^r \longrightarrow \mathbb{R}$ et $\psi : \mathbb{R}^{p-r} \longrightarrow \mathbb{R}$ sont des fonctions quelconques et si $X_1,\cdots,X_p$ sont des variables aléatoires finies mutuellement indépendantes, on peut appliquer le lemme des coalitions avec :

$\longrightarrow$ $s=2$

$\longrightarrow$ la partition (U_1,U_2) telle que $U_1=[\![1,r]\!]$ et $U_2=[\![r+1,p]\!]$

$\longrightarrow$ $Y_1=\varphi(X_1,\cdots,X_r)$ et $Y_2=\psi(X_{r+1},\cdots,X_p)$.

Les variables Y_1 et Y_2 sont donc indépendantes.

• On termine par le dernier point.
Soient $X_1,\cdots,X_p$ des variables aléatoires finies mutuellement indépendantes. Pour tout $i\in [\![1,p]\!]$, on choisit une partie A_i de $\mathbb{R}$.
Comme pour tout $i\in[\![1,p]\!]$, on sait que :

$$\mathbb{P}(X_i\in\mathbb{R}\setminus X_i(\Omega))=0,$$

alors :

$$\mathbb{P}(X_i \in A_i) = \mathbb{P}(X_i \in A_i \cap X_i(\Omega)),$$

et de même en posant dans la suite $B_i = A_i \cap X_i(\Omega)$, on peut encore écrire :

$$\mathbb{P}\left(\bigcap_{i=1}^{p}\{X_i \in A_i\}\right) = \mathbb{P}\left(\bigcap_{i=1}^{p}\{X_i \in B_i\}\right).$$

Il suffit alors d'établir que :

$$\mathbb{P}\left(\bigcap_{i=1}^{p}\{X_i \in B_i\}\right) = \prod_{i=1}^{p}\mathbb{P}(X_i \in B_i).$$

Le très gros avantage des ensembles B_i par rapport aux ensembles A_i est que tous les ensembles B_i sont finis, ce qui va autoriser les sommations suivantes :

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{p}\{X_i \in B_i\}\right) &= \mathbb{P}\left(\bigsqcup_{(b_1,\cdots,b_p)\in\prod_{i=1}^{p}B_i}\{X_1 = b_1,\cdots,X_p = b_p\}\right)\\
&= \sum_{(b_1,\cdots,b_p)\in\prod_{i=1}^{p}B_i}\mathbb{P}\Big(X_1 = b_1,\cdots,X_p = b_p\Big) \qquad [\sigma\text{-additivité}]\\
&= \sum_{(b_1,\cdots,b_p)\in\prod_{i=1}^{p}B_i}\prod_{i=1}^{p}\mathbb{P}(X_i = b_i) \qquad [\text{indépendance des } X_i]\\
&= \prod_{i=1}^{p}\sum_{b_i\in B_i}\mathbb{P}(X_i = b_i) \qquad [\text{manipulation des sommes restangulaires}]\\
&= \prod_{i=1}^{p}\mathbb{P}(X_i \in B_i) \qquad [\sigma\text{-additivité}]
\end{aligned}$$

D'autre part, on peut montrer que l'espérance du produit est égale au produit des espérances, soit en utilisant la formule de transfert puis la manipulation des sommes rectangulaires, soit par récurrence sur l'entier p.
On choisit cette récurrence, par exemple.
On note notre hypothèse de récurrence :

$\mathscr{P}(p)$: « pour toutes variables aléatoires $X_1,\cdots,X_p$ définies sur le même univers fini Ω et à valeurs réelles, si les variables $X_1,\cdots,X_p$ sont mutuellement indépendantes, alors :

$$\mathbb{E}\left(\prod_{i=1}^{p}X_i\right) = \prod_{i=1}^{p}\mathbb{E}(X_i) \text{ ».}$$

▷ Lorsque $p = 1$, le résultat à montrer est évident.
▷ Supposons le résultat vrai au rang p.
▷ Au rang $p+1$, on se donne $p+1$ variables aléatoires $X_1, \cdots, X_p, X_{p+1}$ mutuellement indépendantes. On pose la nouvelle variable :

$$Y = \prod_{i=1}^{p} X_i.$$

Par le lemme des coalitions, les variables Y et X_{p+1} sont mutuellement indépendantes.
On sait alors que :

$$\mathbb{E}(Y \cdot X_{p+1}) = \mathbb{E}(Y) \cdot \mathbb{E}(X_{p+1}).$$

Or,

$$Y \cdot X_{p+1} = \prod_{i=1}^{p+1} X_i,$$

et par hypothèse de récurrence,

$$\mathbb{E}(Y) = \prod_{i=1}^{p} \mathbb{E}(X_i),$$

car les variables $X_1, \cdots, X_p$ restent mutuellement indépendantes (en utilisant une nouvelle fois par exemple le principe des coalitions si on veut avec la partition en singletons, puis les fonctions $f_i = \mathrm{id}_{\mathbb{R}}$ pour tout i entre 1 et p et la fonction nulle pour f_{p+1}. Les variables $X_1, \cdots, X_p, 0$ seront mutuellement indépendantes, ainsi que les variables $X_1, \cdots, X_p$ assez rapidement – une ou deux lignes de calculs à faire pour vérifier que la définition est satisfaite).
On conclut maintenant assez directement que la propriété $\mathscr{P}(p+1)$ est vérifiée.

Remarque : la propriété 11 est assez volumineuse en terme de démonstrations. Toutes les démonstrations sont exigibles dans le cas d'un couple de variables aléatoires. En particulier, il faut savoir retrouver les démonstrations des points 1,2 (dans le cas $d = 2$), 3, 4, le lemme des coalitions dans le cas des variables $Y_1 = \varphi(X_1, \cdots, X_r)$ et $Y_2 = \psi(X_{r+1}, \cdots, X_p)$ et le dernier point dans le cas général.
D'un point de vue probabiliste, il faut retenir que le lemme des coalitions doit être conforme à votre intuition.

On examine maintenant l'exemple 10.
Dans l'exemple 10, on a envie de répondre oui à la question et la réponse à la question est ... oui ! Démontrons ceci. On note les deux points (1) et (2).
On commence par montrer l'implication (1) $\Longrightarrow$ (2).
On suppose les événements $A_1, \cdots, A_p$ mutuellement indépendants.
Soient p réels $x_1, \cdots, x_p$.
Si l'un des x_i – par exemple x_{i_0} – n'appartient pas à $\{0, 1\}$, alors les événements $\bigcap_{i=1}^{p} \{\mathbf{1}_{A_i} = x_i\}$ et $\{\mathbf{1}_{A_{i_0}} = x_{i_0}\}$ sont impossibles ce qui donne :

$$\mathbb{P}\left(\bigcap_{i=1}^{p} \{\mathbf{1}_{A_i} = x_i\}\right) = 0 = \prod_{i=1}^{p} \mathbb{P}(\mathbf{1}_{A_i} = x_i).$$

Supposons que tous les x_i appartiennent à $\{0,1\}$. On note I l'ensemble des indices $i \in [\![1,p]\!]$ tels que x_i soit égal à 1 et donc :

$$J = [\![1,p]\!] \setminus I = \Big\{ i \in [\![1,p]\!],\ x_i = 0 \Big\}.$$

On en déduit :

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{p} \{\mathbf{1}_{A_i} = x_i\}\right) &= \mathbb{P}\left(\bigcap_{i\in I} \{\mathbf{1}_{A_i} = 1\} \cap \bigcap_{j\in J} \{\mathbf{1}_{A_j} = 0\}\right)\\
&= \mathbb{P}\left(\bigcap_{i\in I} A_i \cap \bigcap_{j\in J} \overline{A_j}\right)\\
&= \prod_{i\in I} \mathbb{P}(A_i) \times \prod_{j\in J} \mathbb{P}(\overline{A_j}) \qquad \text{[utiliser la propriété 5 – deuxième point]}\\
&= \prod_{i\in I} \mathbb{P}(\mathbf{1}_{A_i} = 1) \times \prod_{j\in J} \mathbb{P}(\mathbf{1}_{A_j} = 0)\\
&= \prod_{i\in I} \mathbb{P}(\mathbf{1}_{A_i} = x_i) \times \prod_{j\in J} \mathbb{P}(\mathbf{1}_{A_j} = x_j)\\
&= \prod_{i=1}^{p} \mathbb{P}(\mathbf{1}_{A_i} = x_i).
\end{aligned}$$

On a bien l'assertion (2).

On regarde maintenant l'implication (2) $\Longrightarrow$ (1).
On suppose que les variables $\mathbf{1}_{A_1}, \cdots, \mathbf{1}_{A_p}$ sont mutuellement indépendantes.
Soit K une partie finie incluse dans $[\![1,p]\!]$.
On considère pour tout i entre 1 et p, la partie :

$$B_i = \begin{cases} \{1\}, \text{ si } i \in K \\ \{0,1\}, \text{ si } i \notin K \end{cases}.$$

En utilisant le dernier point de la propriété 11, on peut écrire :

$$\mathbb{P}\left(\bigcap_{i=1}^{p} \{\mathbf{1}_{A_i} \in B_i\}\right) = \prod_{i=1}^{p} \mathbb{P}(\mathbf{1}_{A_i} \in B_i) \qquad \bigstar$$

Or, si l'indice i appartient à K, on sait que :

$$\{\mathbf{1}_{A_i} \in B_i\} = \{\mathbf{1}_{A_i} = 1\} = A_i$$

et si l'indice i n'appartient pas à l'ensemble K, alors :

$$\{\mathbf{1}_{A_i} \in B_i\} = \{\mathbf{1}_{A_i} \in \{0,1\}\} = \Omega.$$

Résultat des courses, la formule $\bigstar$ devient :

$$\mathbb{P}\left(\bigcap_{i\in K} A_i\right) = \prod_{i\in K} \mathbb{P}(A_i).$$

Ceci étant valable pour toute partie K incluse dans $[\![1,p]\!]$, on a bien l'assertion (1).

Passons à la propriété 12 qui sera une application de ce que l'on sait désormais.
• Si l'ensemble Ω est fini, l'ensemble $\mathscr{F}(\Omega,\mathbb{R})$ est un $\mathbb{R}$-espace vectoriel de dimension finie. Plus précisément, en voici une base :

$$\mathscr{B}=\big(\mathbf{1}_{\{\omega\}}\big)_{\omega\in\Omega}.$$

On vérifie assez facilement que la famille est libre en évaluant toute combinaison linéaire nulle :

$$\sum_{\omega\in\Omega}\lambda_\omega\cdot\mathbf{1}_{\{\omega\}}=0$$

en un élément $\alpha\in\Omega$, ce qui aura pour effet de simplifier la somme en :

$$\lambda_\alpha=0_{\mathbb{R}}.$$

De plus, la famille est génératrice car :

$$\forall X\in\mathscr{F}(\Omega,\mathbb{R}),\ X=\sum_{\omega\in\Omega}X(\omega)\cdot\mathbf{1}_{\{\omega\}}.$$

Examinons l'application suivante :

$$\Phi:\left|\begin{array}{ccc}\mathscr{F}(\Omega,\mathbb{R})^2 & \longrightarrow & \mathbb{R}\\ (X,Y) & \longmapsto & \mathbb{E}(X\cdot Y)\end{array}\right..$$

Il est clair que l'application Φ est symétrique. De plus, elle est linéaire par rapport à la première variable :

$$\begin{aligned}\Phi(\lambda\cdot X_1+X_2,Y) &= \mathbb{E}\Big((\lambda X_1+X_2)\cdot Y\Big)\\ &= \mathbb{E}\Big(\lambda\cdot X_1\,Y+X_2\,Y\Big) && \text{[distributivité du produit]}\\ &= \lambda\,\mathbb{E}(X_1\,Y)+\mathbb{E}(X_2\,Y) && \text{[linéarité de l'espérance]}\\ &= \lambda\cdot\Phi(X_1,Y)+\Phi(X_2,Y).\end{aligned}$$

L'application Φ est donc bilinéaire.
Ensuite, l'application Φ est positive car si $X:\Omega\longrightarrow\mathbb{R}$ est une variable aléatoire, alors

$$\Phi(X,X)=\mathbb{E}(X^2)\geqslant 0,$$

par positivité de l'espérance.
Enfin, testons le caractère défini de la forme Φ. Soit $X:\Omega\longrightarrow\mathbb{R}$ une variable aléatoire telle que :

$$\Phi(X,X)=0.$$

On en déduit par la formule de transfert :

$$0=\mathbb{E}(X^2)=\sum_{x\in X(\Omega)}x^2\,\mathbb{P}(X=x).$$

La somme nulle est à termes positifs : tous les termes sont nuls amenant le fait que pour tout x dans $X(\Omega) \setminus \{0\}$, on a :

$$x^2 \cdot \mathbb{P}(X = x) = 0, \text{ puis } \mathbb{P}(X = x) = 0.$$

Ainsi, $\mathbb{P}(X \neq 0) = 0$ et en passant à l'événement contraire :

$$\mathbb{P}(X = 0) = 1.$$

La variable X est presque sûrement nulle. Le problème est que la variable X peut ne pas être la variable nulle – sous-entendu vraiment nulle partout.
Par exemple, si $\Omega = \{0, 1, 2\}$, muni de la probabilité $\mathbb{P}$ associée à la famille $(1, 0, 0)$ de nombres réels positifs et de somme 1, alors :

$$\forall A \in \mathscr{P}(\Omega), \ \mathbb{P}(A) = \mathrm{Card}(A \cap \{0\}).$$

Considérons maintenant la variable aléatoire $X : \left| \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{R} \\ \omega & \longmapsto & \omega \end{array} \right.$ Il apparaît que la variable X n'est pas du tout la variable nulle.
Cependant,

$$\mathbb{P}(X \neq 0) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = \mathbb{P}(\{1, 2\}) = 0$$

et donc :

$$\Phi(X, X) = \sum_{i=0}^{2} i^2 \cdot \mathbb{P}(X = i) = 0.$$

La forme bilinéaire symétrique positive Φ n'est pas un produit scalaire sur l'espace $\mathscr{F}(\Omega, \mathbb{R})$ de dimension $s = \mathrm{Card}(\Omega)$.
Pourtant, on a l'impression qu'il ne manque pas grand chose à l'application Φ pour être un produit scalaire.
Si on voulait complètement formaliser les choses, voici la trame du contournement de cette difficulté :

- ▷ on définit la relation $\mathscr{R}$ sur l'ensemble $E = \mathscr{F}(\Omega, \mathbb{R})$ de toutes les variables aléatoires $X : \Omega \longrightarrow \mathbb{R}$ comme suit :

  $$X \ \mathscr{R} \ Y \iff \mathbb{P}(X = Y) = 1.$$

  Autrement dit, deux variables sont en relation si elles sont presque sûrement égales.
  La relation est réflexive (facile), symétrique (facile) et transitive (un peu moins facile).
  Allez, on détaille. Si $X \ \mathscr{R} \ Y$ et $Y \ \mathscr{R} \ Z$, alors on remarque l'inclusion :

  $$\{X \neq Z\} \subset \{X \neq Y\} \cup \{Y \neq Z\},$$

  car si $\omega \in \Omega$ appartient à l'ensemble de gauche, alors $X(\omega) \neq Z(\omega)$. Il est donc impossible d'avoir à la fois $X(\omega) = Y(\omega)$ et $Y(\omega) = Z(\omega)$. Par conséquent, l'élément ω appartient bien à l'ensemble de droite.
  Ensuite, par définition de la relation $\mathscr{R}$, on a :

  $$\mathbb{P}(X \neq Y) = \mathbb{P}(Y \neq Z) = 0,$$

  donc :

  $$\mathbb{P}(X \neq Z) \leqslant \mathbb{P}(\{X \neq Y\} \cup \{Y \neq Z\}) = \mathbb{P}(X \neq Y) + \mathbb{P}(Y \neq Z) - \mathbb{P}(X \neq Y, Y \neq Z) = 0.$$

  Ainsi, les variables X et Z sont égales presque sûrement.

▷ On note par exemple $[E] = E/\mathscr{R}$ l'ensemble quotient qui est l'ensemble de toutes les classes d'équivalence.

▷ On vérifie (un peu fastidieux mais pas très compliqué) qu'en notant $[X]$ la classe d'équivalence de la variable $X : \Omega \longrightarrow \mathbb{R}$, alors les opérations suivantes sont toutes bien définies car elles ne dépendent pas réellement des représentants choisis dans les classes d'équivalence :

$\longrightarrow$ addition :

$$\forall([X],[Y]) \in [E]^2,\ [X]+[Y] = [X+Y]$$

$\longrightarrow$ multiplication :

$$\forall([X],[Y]) \in [E]^2,\ [X]\times[Y] = [X\times Y]$$

$\longrightarrow$ multiplication par un scalaire :

$$\forall[X] \in [E],\ \forall\lambda \in \mathbb{R},\ \lambda\cdot[X] = [\lambda\cdot X]$$

$\longrightarrow$ espérance :

$$\forall[X] \in [E],\ \mathbb{E}([X]) = \mathbb{E}(X),$$

ce dernier point étant justifié par le fait que deux variables égales presque sûrement admettent la même espérance (et aussi la même variance ...)

On vérifierait que le quadruplet $([E],+,\times,\cdot)$ est une $\mathbb{R}$-algèbre commutative, non intègre en général (sauf si Ω est un singleton) et de dimension finie.

▷ L'application :

$$\Psi : \left|\begin{array}{ccc} [E]^2 & \longrightarrow & \mathbb{R} \\ ([X],[Y]) & \longmapsto & \mathbb{E}([X]\times[Y]) \end{array}\right.$$

ressemble beaucoup à l'application Φ de tout à l'heure. Enfin bref. On vérifierait assez facilement que l'application Ψ est bilinéaire symétrique positive et si $\Psi([X],[X]) = 0$, alors : $\mathbb{E}(X^2) = 0$, amenant le fait que la variable X est nulle presque sûrement. De là, on peut maintenant écrire :

$$[X] = [0],$$

par la définition de la relation d'équivalence $\mathscr{R}$ du début et on obtient que le vecteur $[X]$ est cette fois-ci vraiment le vecteur nul, dans l'espace $[E]$.

On obtient sans grande surprise toutes les choses déjà connues sur les produits scalaires, en particulier l'inégalité de Cauchy-Schwarz.

On démontre malgré tout cette inégalité puisque l'application Φ du début n'était pas tout à fait un produit scalaire. Le formalisme avec l'application Ψ et le passage au quotient est plutôt hors programmme...

Soient $X : \Omega \longrightarrow \mathbb{R}$ et $Y : \Omega \longrightarrow \mathbb{R}$ deux variables aléatoires finies.

Si la variable Y est constante presque sûrement, égale à c, alors :

$$X\ Y = c\ X, \text{ presque sûrement}$$

et donc :

$$\mathrm{Cov}(X,Y) = \mathbb{E}(X\ Y) - \mathbb{E}(X)\cdot\mathbb{E}(Y) = \mathbb{E}(c\ X) - \mathbb{E}(X)\times c = 0.$$

L'inégalité souhaitée est alors acquise car l'écart-type $\sigma(Y)$ est en fait lui aussi nul ou bien de toutes façons, tous les écarts-types sont positifs ou nuls quoiqu'il arrive.

Si la variable Y n'est pas constante presque sûrement, alors sa variance est strictement positive.

Dans la suite, on pose $X' = X - \mathbb{E}(X)$ et $Y' = Y - \mathbb{E}(Y)$; les variables X' et Y' sont maintenant centrées et la variable Y' n'est pas l'application nulle presque sûrement.
Considérons l'application suivante :

$$f : \left| \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ t & \longmapsto & \mathbb{E}\left(\left(X' + t \cdot Y'\right)^2\right) \end{array} \right. .$$

En développant le carré puis en utilisant la linéarité de l'espérance, et en remarquant que $\mathbb{E}(X'^2) = V(X)$ la variance de X, idem pour Y et :

$$\mathbb{E}(X'\ Y') = \mathrm{Cov}(X, Y)$$

par définition de la covariance, on voit que l'on a :

$$\forall t \in \mathbb{R},\ f(t) = \mathbb{E}(X'^2) + 2t\ \mathbb{E}(X'\ Y') + t^2\ \mathbb{E}(Y'^2) = V(X) + 2t\ \mathrm{Cov}(X, Y) + t^2\ V(Y).$$

La fonction f est polynomiale de degré 2 et ne prend que des valeurs positives : le discriminant associé à ce polynôme est négatif. Or,

$$\Delta = 4\ \left(\mathrm{Cov}(X, Y))^2 - V(X)^2 V(Y)^2\right) \leqslant 0,$$

donc :

$$\Big(\mathrm{Cov}(X, Y)\Big)^2 \leqslant V(X)^2\ V(Y)^2,$$

et on obtient l'inégalité voulue en prenant les racines carrées.

• Le point suivant provient directement de la linéarité de l'espérance. On remarquera que la formule est vraie même si les variables X_i ne sont pas indépendantes.

• Pour la variance d'une somme, on pose comme précédemment, pour tout entier k entre 1 et n :

$$X'_k = X_k - \mathbb{E}(X_k),$$

de sorte que pour tout entier k,

$$\mathbb{E}\left((X'_k)^2\right) = V(X_k),$$

et pour tous indices i et j,

$$\mathbb{E}(X'_i\ X'_j) = \mathrm{Cov}(X_i, X_j).$$

Finalement, en posant la constante :

$$m = \sum_{k=1}^{n} \mathbb{E}(X_k),$$

alors :

$$X_1 + \cdots + X_n = X'_1 + \cdots + X'_n + m,$$

donc les variables $X_1 + \cdots + X_n$ et $X'_1 + \cdots + X'_n$ ont la même variance.

On déroule alors les calculs comme suit :

$$\begin{array}{rcll}
V(X_1+\cdots+X_n) & = & V(X'_1+\cdots+X'_n) & \\
& = & \mathbb{E}\left(\left(\displaystyle\sum_{k=1}^{n} X'_k\right)^2\right) - \left(\mathbb{E}\left(\displaystyle\sum_{k=1}^{n} X'_k\right)\right)^2 & \text{[formule de Huyghens]} \\
& = & \mathbb{E}\left(\left(\displaystyle\sum_{k=1}^{n} X'_k\right)^2\right) & \text{[variables } X'_k \text{ centrées]} \\
& = & \mathbb{E}\left(\displaystyle\sum_{k=1}^{n} {X'_k}^2\right) + 2 \displaystyle\sum_{1\leqslant i<j\leqslant n} \mathbb{E}(X'_i\ X'_j) & \text{[développement du carré]} \\
& = & \displaystyle\sum_{k=1}^{n} \mathbb{E}\left({X'_k}^2\right) + 2 \displaystyle\sum_{1\leqslant i<j\leqslant n} \mathbb{E}(X'_i\ X'_j) & \\
& = & \displaystyle\sum_{k=1}^{n} V(X_k) + 2 \displaystyle\sum_{1\leqslant i<j\leqslant n} \mathrm{Cov}(X_i, X_j). &
\end{array}$$

Lorsque les variables $X_1, \cdots, X_n$ sont deux à deux indépendantes (hypothèse un peu moins forte que l'indépendance mutuelle), pour tous indices $i < j$, alors :

$$\mathrm{Cov}(X_i, X_j) = 0,$$

et on obtient dans le cas d'indépendance que la variance d'une somme est la somme des variances. Ceci ne marche qu'en cas d'indépendance.

Appliquons sur des exemples.

Pour l'exemple 11, on nous donne la loi conjointe du couple (X, Y).
À partir de ce tableau, on peut en déduire les lois marginales.
La variable X prend ses valeurs dans $\{0, 1, 2\}$, avec

$$\mathbb{P}(X=0) = \frac{1}{4},\ \mathbb{P}(X=1) = \frac{1}{2} \text{ et } \mathbb{P}(X=2) = \frac{1}{4}.$$

La variable Y prend ses valeurs dans $\{0, 1\}$, avec

$$\mathbb{P}(Y=0) = \frac{1}{2} = \mathbb{P}(Y=1).$$

On calcule la covariance à l'aide de la formule de Huyghens.
Par la formule de transfert, on obtient :

$$\mathbb{E}(X\ Y) = \sum_{i=0}^{2}\sum_{j=0}^{1} i\ j\ \mathbb{P}(X=i, Y=j) = 2 \times 1 \times \frac{1}{4} = \frac{1}{2}.$$

Comme on connaît les lois marginales, on peut facilement calculer les espérances. On trouve :

$$\mathbb{E}(X) = 1 \text{ et } \mathbb{E}(Y) = \frac{1}{2}.$$

Conclusion,

$$\mathrm{Cov}(X, Y) = \mathbb{E}(X\ Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y) = 0.$$

Cependant, on voit à l'aide de la loi conjointe que les variables X et Y ne sont pas indépendantes. En effet,

$$\mathbb{P}(X = 0, Y = 0) = 0 \neq \mathbb{P}(X = 0)\ \mathbb{P}(Y = 0).$$

En fait, si les variables X et Y étaient indépendantes, dès qu'un zéro apparaît dans le tableau de la loi conjointe, alors cela signifie que :

$$\mathbb{P}(X = x, Y = y) = 0,$$

donc $\mathbb{P}(X = x)\ \mathbb{P}(Y = y) = 0$, puis $\mathbb{P}(X = x) = 0$ ou $\mathbb{P}(Y = y) = 0$ et donc on aurait une ligne ou une colonne remplie de zéros, ce qui n'est manifestement pas le cas dans le tableau proposé.

On termine cette section sur l'indépendance des variables aléatoires, section riche en émotion, par l'exemple 12, traité sous forme d'exercice.
Les résultats que l'on va rencontrer sont à connaître.
On note parfois en abrégé v.a.i.id pour variables aléatoires indépendantes identiquement distribuées.
Traduction : on considère des variables aléatoires $X_n : \Omega \longrightarrow \mathbb{R}$ qui sont mutuellement indépendantes (c'est-à-dire que pour tout $N \in \mathbb{N}$, les variables $X_0, \cdots, X_N$ sont mutuellement indépendantes) et possédant chacune la même loi (autrement dit, pour tout $x \in \mathbb{R}$, la quantité $\mathbb{P}(X_n = x)$ dépend de x mais pas de l'entier n).

On pose par exemple $X = X_0$ qui est une variable aléatoire qui a la même loi que toutes les variables X_n, pour tout $n \in \mathbb{N}$.

1. Soit $n \in \mathbb{N}$. Alors :

$$\begin{aligned}
\mathbb{E}(X_n) &= \sum_{x \in X_n(\Omega)} x\ \mathbb{P}(X_n = x) \\
&= \sum_{x \in X_n(\Omega) \cup X(\Omega)} x\ \mathbb{P}(X_n = x) \qquad \text{[ajout de termes nuls]} \\
&= \sum_{x \in X_n(\Omega) \cup X(\Omega)} x\ \mathbb{P}(X = x) \qquad \text{, car } \mathbb{P}_{X_n} = \mathbb{P}_X \\
&= \sum_{x \in X(\Omega)} x\ \mathbb{P}(X = x) \qquad \text{[suppression de termes nuls]} \\
&= \mathbb{E}(X).
\end{aligned}$$

De la même façon,

$$\begin{aligned}
\mathbb{E}(X_n^2) &= \sum_{x \in X_n(\Omega)} x^2\ \mathbb{P}(X_n = x) \\
&= \sum_{x \in X_n(\Omega) \cup X(\Omega)} x^2\ \mathbb{P}(X_n = x) \qquad \text{[ajout de termes nuls]} \\
&= \sum_{x \in X_n(\Omega) \cup X(\Omega)} x^2\ \mathbb{P}(X = x) \qquad \text{, car } \mathbb{P}_{X_n} = \mathbb{P}_X \\
&= \sum_{x \in X(\Omega)} x^2\ \mathbb{P}(X = x) \qquad \text{[suppression de termes nuls]} \\
&= \mathbb{E}(X^2)
\end{aligned}$$

puis :

$$V(X_n) = \mathbb{E}(X_n^2) - (\mathbb{E}(X_n))^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = V(X).$$

Toutes les variables X_n ont la même espérance et la même variance et donc le même écart-type, en prenant la racine carrée.

2. Si $\sigma = 0$, alors pour tout $n \in \mathbb{N}$, $\mathbb{E}\left((X_n - m)^2\right) = 0$ et on a déjà vu dans ce cas que la variable X_n était égale à m presque sûrement.
3. On suppose que σ est strictement positif.
   On fixe un entier $n \geqslant 1$. On fixe un réel $\varepsilon > 0$.
   Dans la suite, on pose :
   $$X_k' = X_k - m$$
   qui est une variable centrée.
   De plus,
   $$\frac{X_1 + \cdots + X_n}{n} - m = \frac{X_1' + \cdots + X_n'}{n}.$$
   On reconnaît une inégalité de type Bienaymé-Tchebychev, que l'on redémontre.
   On dispose des égalités :
   $$\begin{aligned}
   A &= \left\{ \left| \frac{X_1 + \cdots + X_n}{n} - m \right| \geqslant \varepsilon \right\} \\
   &= \left\{ \left| X_1' + \cdots + X_n' \right| \geqslant n\ \varepsilon \right\} \\
   &= \left\{ \left(X_1' + \cdots + X_n' \right)^2 \geqslant (n^2\ \varepsilon^2) \right\}.
   \end{aligned}$$
   On dispose en outre de l'inégalité :
   $$n^2\ \varepsilon^2\ \mathbf{1}_A \leqslant \left(X_1' + \cdots + X_n' \right)^2,$$
   comme d'habitude en prenant un élément ω quelconque dans Ω et en distinguant les cas $\omega \in A$ ou $\omega \notin A$.

On prend alors l'espérance dans cette inégalité, ce qui donne :

$$n^2\ \varepsilon^2\ \mathbb{P}(A) \leqslant \mathbb{E}\left(\left(X'_1 + \cdots + X'_n\right)^2\right).$$

Or,

$$\begin{aligned}
\mathbb{E}\left(\left(X'_1 + \cdots + X'_n\right)^2\right) &= V(X_1 + \cdots + X_n) \qquad \text{[définition de la variance]}\\
&= \sum_{k=1}^{n} V(X_k) \qquad \text{[indépendance]}\\
&= \sum_{k=1}^{n} \sigma^2 \qquad \text{[question 1.]}\\
&= n\ \sigma^2.
\end{aligned}$$

Conclusion,

$$\mathbb{P}(A) \leqslant \frac{n\ \sigma^2}{n^2\ \varepsilon^2} = \frac{\sigma^2}{n\ \varepsilon^2},$$

ce qui correspond à ce qu'il fallait.
On aurait pu directement appliquer l'inégalité de Bienaymé-Tchebychev à la variable $Y = \dfrac{X_1 + \cdots + X_n}{n}$. On a sûrement – presque sûrement dirait le probabiliste – trop détaillé.

4. Si $\varepsilon > 0$ est fixé, il suffit d'appliquer le théorème des gendarmes, puisque :

$$\lim_{n \longrightarrow +\infty} \frac{\sigma^2}{n\ \varepsilon^2} = 0.$$

Ce résultat est connu sous le nom de loi faible des grands nombres et le résultat est à connaître et à savoir redémontrer dans l'idéal.

Ouvrons une parenthèse hors programme dont le résultat me paraît important.
Le résultat suivant est connu sous la loi forte des grands nombres– en abrégé LFGN.
Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a.i.id (variables finies ayant toutes la même loi et étant mutuellement indépendantes), en notant $m = \mathbb{E}(X_n)$ l'espérance commune à tous les X_n, alors presque sûrement, on a :

$$\lim_{n \longrightarrow +\infty} \frac{X_1 + \cdots + X_n}{n} = m.$$

Ce résultat est hors de portée en terme de démonstration. Nous n'allons pas la faire ici.

Imaginons que l'on ait une suite $(A_n)_{n\in\mathbb{N}^*}$ d'événements mutuellement indépendants – *i.e. pour tout N, les événements $A_1, \cdots, A_N$ sont mutuellement indépendants* – et tous de même probabilité p.
Alors, en utilisant l'exemple 10, la famille $(\mathbf{1}_{A_n})_{n\in\mathbb{N}^*}$ est une suite de variables mutuellement indépendantes et chaque variable $\mathbf{1}_{A_n}$ suit la loi de Bernoulli de paramètre $p = \mathbb{P}(A_n)$.

Nous sommes dans les conditions d'application de la loi forte des grands nombres, ce qui donne que si l'on choisit un élément ω selon la probabilité $\mathbb{P}$, alors presque sûrement on a :

$$\lim_{n \longrightarrow +\infty} \frac{\mathbf{1}_{A_1}(\omega) + \cdots + \mathbf{1}_{A_n}(\omega)}{n} = p.$$

Or, la quantité $\dfrac{\mathbf{1}_{A_1}(\omega) + \cdots + \mathbf{1}_{A_n}(\omega)}{n}$ correspond à la proportion de réalisations de l'événement A de probabilité p, lorsque l'on effectue des essais indépendants dans les mêmes conditions.
La probabilité $p = \mathbb{P}(A)$ apparaît comme une proportion limite de réalisation de l'événement A.
Lorsque l'on a affaire à une famille $(X_n)_{n \in \mathbb{N}^*}$ de va.i.id et de même loi qu'une variable $X : \Omega \longrightarrow \mathbb{R}$, alors presque sûrement, on a :

$$\lim_{n \longrightarrow +\infty} \frac{X_1 + \cdots + X_n}{n} = \mathbb{E}(X).$$

L'espérance $\mathbb{E}(X)$ apparaît comme la valeur moyenne limite sur n réalisations de la variable X (donnant les X_k) ces réalisations étant effectuées de manière indépendante et dans les mêmes conditions.

Je vous suggère de tester par vous-même l'interface PYTHON à l'adresse :

`http://833duparc.free.fr/Nouveaux_Cours_MPSI/Interface_Probabilites.py`

Il s'agit d'un script PYTHON dans lequel vous pouvez visualiser la stabilisation des proportions aléatoires de réalisations indépendantes d'une variable aléatoire X.
Vous pouvez modifier la définition de cette variable à l'intérieur du script lui-même.
Par défaut, la variable X compte le nombre de 1 lorsqu'on lance 20 dés à six faces parfaitement équilibrés.

Les lois usuelles finies

Nous passons dans ce paragraphe les trois lois usuelles finies qui sont à votre programme de première année : la loi uniforme, la loi de Bernoulli et la loi binomiale.
À cette liste de trois lois, il faudra ajouter deux autres lois usuelles en deuxième année.

Détaillons dans cette couleur la loi uniforme.
Étant donné un univers fini Ω de cardinal s, la loi uniforme $\mathscr{U}_\Omega$ est la probabilité :

$$\mathscr{U}_\Omega = \frac{1}{s} \sum_{\omega \in \Omega} \delta_\omega,$$

où le symbole δ_ω désigne la mesure de Dirac concentrée en le singleton $\{\omega\}$.
La caractéristique de la loi uniforme est que la probabilité de chaque singleton est identique.
Autrement dit, pour toute partie $A \in \mathscr{P}(\Omega)$, on a :

$$\mathscr{U}_\Omega(A) = \frac{1}{s} \sum_{\omega \in \Omega} \delta_\omega(A) = \frac{\mathrm{Card}(A)}{s}.$$

La probabilité d'un événement A se calcule alors de la façon suivante :

- compter le nombre total d'expériences possibles : c'est l'entier s qui est le cardinal de l'univers Ω
- compter le nombre d'expériences favorables à l'événement A : c'est le cardinal de l'ensemble A
- la probabilité de l'événement A est donc le quotient du nombre d'expériences favorables à l'événement A par le nombre total d'expériences possibles.

Si l'on transpose tout ceci aux variables aléatoires, lorsque $X : \Omega \longrightarrow \mathbb{R}$ est une variable aléatoire finie, on dira que X suit la loi uniforme si :

$$\boxed{\mathbb{P}_X = \mathscr{U}_{X(\Omega)}}$$

L'univers considéré ici est bien l'ensemble $X(\Omega)$ des valeurs prises par la variable X.
Ainsi,

$$\mathbb{P}_X = \frac{1}{\text{Card}(X(\Omega))} \sum_{x \in X(\Omega)} \delta_x,$$

ou encore pour toute partie A incluse dans $X(\Omega)$, on a :

$$\mathbb{P}(X \in A) = \mathbb{P}_X(A) = \frac{1}{\text{Card}(X(\Omega))} \sum_{x \in X(\Omega)} \delta_x(A) = \frac{\text{Card}(A)}{\text{Card}(X(\Omega))},$$

ou encore de manière un peu plus générale, pour toute partie A incluse dans $\mathbb{R}$ – la partie A étant finie ou infinie, peu importe, comme on a l'égalité des événements :

$$\Big\{X \in A\Big\} = \Big\{X \in A \cap X(\Omega)\Big\},$$

on en déduit :

$$\forall A \in \mathscr{P}(\mathbb{R}),\ \mathbb{P}(X \in A) = \mathbb{P}_X(A) = \frac{\text{Card}(A \cap X(\Omega))}{\text{Card}(X(\Omega))}.$$

On l'aura compris, une variable $X : \Omega \longrightarrow \mathbb{R}$ qui suit une lui uniforme est une variable pour laquelle chaque valeur prise a autant de chance de se produire que les autres. On est dans une situation d'équiprobabilité.

Démontrons la propriété 13 qui nous donne l'espérance et la variance d'une variable $X : \Omega \longrightarrow [\![1, n]\!]$, dont la loi est $\mathscr{U}_n$, où la quantité n est un entier strictement positif.
Dans cette situation, on a :

$$\forall k \in [\![1, n]\!],\ \mathbb{P}(X = k) = \frac{1}{n}.$$

On obtient assez directement en utilisant une somme arithmétique :

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=1}^{n} k\ \mathbb{P}(X = k) \\ &= \sum_{k=1}^{n} k\ \frac{1}{n} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{k=1}^{n} k \\
&= \frac{1}{n}\,\frac{n(n+1)}{2} \\
&= \frac{n+1}{2}.
\end{aligned}
$$

La valeur moyenne d'une telle variable X est donc $\dfrac{n+1}{2}$.
En ce qui concerne la variance $V(X)$, on applique la formule de Huyghens en commençant par calculer $\mathbb{E}(X^2)$. Voici le calcul :

$$
\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=1}^{n} k^2\,\mathbb{P}(X=k) && \text{[formule de transfert]} \\
&= \sum_{k=1}^{n} k^2\,\frac{1}{n} \\
&= \frac{1}{n}\sum_{k=1}^{n} k^2 \\
&= \frac{1}{n}\,\frac{n(n+1)(2n+1)}{6} && \text{[formule de la somme des carrés]} \\
&= \frac{(n+1)(2n+1)}{6}.
\end{aligned}
$$

En conclusion, on écrit :

$$
\begin{aligned}
V(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \\
&= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\
&= \frac{n+1}{12} \times \Big(2(2n+1) - 3(n+1)\Big) \\
&= \frac{n+1}{12} \times (n-1) \\
&= \frac{n^2-1}{12}.
\end{aligned}
$$

Passons aux exemples pour illustrer tout cela.
• Le premier point a déjà été largement abordé auparavant sur la situation d'équiprobabilité. Metttons cela en pratique à travers un exemple concret.
Prenez un sac contenant n boules numérotées de 1 à n. Piochez aléatoirement une boule du sac. En notant ω cette expérience, on définit $X(\omega)$ le numéro de la boule piochée.
Eh bien la variable X suit la loi uniforme $\mathscr{U}_n$. Les choses sont intuitivement et expérimentalement évidentes sur ce point.

Réalisez maintenant la nouvelle expérience suivante : prenez un entier $N \in \mathbb{N}$ assez grand ; disons $N = 10^5$.
Réalisez l'expérience suivante N fois de manière indépendante. L'indépendance est expérimentalement facile à établir : vous piochez la première boule dans le sac et vous notez $X_1(\omega)$ le nombre obtenu.
Vous remettez la boule et vous mélangez puis vous repiochez une deuxième fois. Vous obtenez un deuxième nombre $X_2(\omega)$.
Vous répétez ce processus $N-1$ fois, pour obtenir N nombres $X_1(\omega), \cdots, X_N(\omega)$.
Les variables $X_1, \cdots, X_N$ ainsi définies sont des va.i.id. On est dans le champ d'application de la loi forte des grands nombres. Lorsque l'on calcule la moyenne

$$\frac{X_1(\omega) + \cdots + X_N(\omega)}{N}$$

des nombres obtenus, vous constituez une suite aléatoire de nombres et cette suite converge presque sûrement vers $\mathbb{E}(X) = \dfrac{n+1}{2}$. Autrement dit, la probabilité que la suite ne converge pas vers $\dfrac{n+1}{2}$ est de 0 %. Autrement dit, si – lorsque vous faites vos calculs – vous constatez que la suite ne converge pas vers cette valeur, rien d'inquiétant. L'événement est négligeable mais pas impossible. Il faudra tout de même se poser des questions sur la teneur de vos calculs...
Allez jouer au loto : vous aurez une infinité de fois plus de chances de gagner le gros lot.
Vous pouvez tout à fait utiliser l'interface PYTHON dont l'adresse est donnée page **73** pour réaliser à moindres frais cette vérification expérimentale.

• Détaillons le deuxième point.
La variable X définie ici est une variable qui prend ses valeurs dans l'ensemble :

$$[\![0, s]\!]$$

car on pioche s boules simultanément. Toutes les valeurs entières entre 0 et s ne sont pas forcément des valeurs prises ; cela dépend de l'entier p ainsi que de l'entier N.
La variable X ne suit pas du tout une loi uniforme, mais pour calculer la loi de X, on va devoir utiliser la situation d'équiprobabilité suivante : tous les tirages simultanés sont équiprobables.
Fixons un entier k entre 0 et s.
Il y a déjà $\begin{pmatrix} N \\ s \end{pmatrix}$ tirages possibles.
On compte maintenant le nombre de tirages favorables à l'événement $\{X = k\}$. Pour cela, il faut choisir les k boules blanches piochées parmi les p boules blanches en totalité, puis il ne faut pas oublier de choisir les $s-k$ boules autres que blanches présentes également dans le tirages de s boules.
Il y a donc $\begin{pmatrix} p \\ k \end{pmatrix} \times \begin{pmatrix} N-p \\ s-k \end{pmatrix}$ tirages favorables.
On en déduit la loi de la variable aléatoire X :

$$\forall k \in [\![0, s]\!],\ \mathbb{P}(X = k) = \frac{\begin{pmatrix} p \\ k \end{pmatrix} \times \begin{pmatrix} N-p \\ s-k \end{pmatrix}}{\begin{pmatrix} N \\ s \end{pmatrix}}.$$

Cette loi porte un nom : c'est la loi hypergéométrique, qui est une loi usuelle mais pas à votre programme. C'est bien de connaître le nom de cette loi, mais vous n'avez pas à connaître davantage de choses sur cette loi.
Pour être tout à fait complet, on peut écrire par σ-additivité de la probabilité $\mathbb{P}$ et par le fait que la famille $\Big(\{X = k\}\Big)_{0 \leqslant k \leqslant s}$ est un système complet d'événements – avec potentiellement des événements vides dans cette famille, cela dépend des entiers p et N :

$$\sum_{k=0}^{s} \mathbb{P}(X = k),$$

donc en multipliant par le dénominateur $\left(\begin{array}{c} N \\ s \end{array}\right)$, on aboutit à la formule de Van der Monde :

$$\sum_{k=0}^{s} \left(\begin{array}{c} p \\ k \end{array}\right) \times \left(\begin{array}{c} N-p \\ s-k \end{array}\right) = \left(\begin{array}{c} N \\ s \end{array}\right).$$

• Pour le troisième point, les choses sont un peu plus subtiles et laisse un peu le champ libre à notre interprétation probabiliste.
On part du principe – l'énoncé suggère d'apporter des renseignements supplémentaires afin de finaliser le calcul – que les mesures forment un n-uplet que l'on note $(a_1, \cdots, a_n)$ de composantes toutes différentes. Tout se passe comme si on avait un ensemble de n mesures et que l'on piochait les mesures les unes après les autres.
On note $b_1, \cdots, b_n$ les mesures effectuées et rangées dans l'ordre strictement croissant.
Les mesures effectives $a_1, \cdots, a_n$ sont associées à la permutation $\sigma \in \mathfrak{S}_n$ telle que :

$$\forall i \in [\![1, n]\!],\ a_i = b_{\sigma(i)}.$$

Tout se passe comme si l'on choisissait une permutation τ aléatoirement dans l'ensemble $\mathfrak{S}_n$ puis que l'on regardait les mesures $b_{\tau(1)}, \cdots, b_{\tau(n)}$ et que l'on testait si la $j^{\text{ème}}$ année est un record, autrement dit si pour tout entier i entre 1 et $j-1$, on a :

$$b_{\tau(i)} < b_{\tau(j)},$$

ce qui se traduit par ces conditions plus facilement exploitables sur la permutation σ :

$$\forall i \in [\![1, j-1]\!],\ \tau(i) < \tau(j).$$

On vient de transposer le problème à la nouvelle situation suivante :
« On munit l'ensemble $\mathfrak{S}_n$ de la probabilité uniforme :

$$\mathbb{P} = \frac{1}{n!} \sum_{\tau \in \mathfrak{S}_n} \delta_\tau.$$

On fixe un entier j entre 1 et n, correspondant à la $j^{\text{ème}}$ année, donc à la $j^{\text{ème}}$ mesure.
On veut calculer la probabilité de l'événement :

$$A_j = \Big\{\tau \in \mathfrak{S}_n \mid \forall i \in [\![1, j-1]\!],\ \tau(i) < \tau(j)\Big\}. »$$

Pour calculer cette probabilité, on va exploiter l'équiprobabilité relative aux permutations τ.
Il y a $n!$ permutations τ au total.
On compte le nombre de permutations τ favorables à l'événement A_j.
Pour tout entier k entre 1 et n, on note :

$$B_{j,k} = \Big\{\tau \in A_j \mid \tau(j) = k\Big\}.$$

Il est clair que l'on a l'égalité d'ensembles :

$$A_j = \bigsqcup_{k=1}^{n} B_{j,k}.$$

On fixe un entier k entre 1 et n et on calcule le cardinal de l'ensemble $B_{j,k}$.
Pour compter les permutations τ dans l'ensemble $B_{j,k}$, il n'y a qu'un seul choix possible pour l'image $\tau(j)$: on doit choisir $\tau(j) = k$.
Ensuite, il s'agit de compter les choix des images $\tau(1), \cdots, \tau(j-1)$, qui sont nécessairement des images dans l'ensemble $[\![1, k-1]\!]$.
Comme les images $\tau(1), \cdots, \tau(j-1)$ sont toutes différentes, cela impose d'avoir $j-1 \leqslant k-1$.
Par conséquent, si $k < j$, alors l'ensemble $B_{j,k}$ est vide.
Plaçons-nous dans le cas où l'entier k est supérieur ou égal à j. Dans cette situation, il y a $k-1$ choix pour $\tau(1)$, puis $k-2$ choix pour $\tau(2)$, ..., jusqu'à $k-j+1$ choix pour $\tau(j-1)$.
Cela fait donc $\displaystyle\prod_{\ell=1}^{j-1}(k-\ell) = \frac{(k-1)!}{(k-j)!}$ choix possibles pour les $j-1$ premières images de la permutation τ.
Il reste à choisir arbitrairement les $n-j$ dernières images $\tau(j+1), \cdots, \tau(n)$ dans l'ensemble à $n-j$ entiers qui restent. Cela fait $(n-j)!$ choix.
Il y a donc :

$$\frac{(k-1)!}{(k-j)!}\,(n-j)!$$

éléments dans l'ensemble $B_{j,k}$, lorsque $k \geqslant j$ et 0 élément dans $B_{j,k}$, lorsque $k < j$.
On en déduit le nombre de permutations favorables :

$$\begin{aligned}
\mathrm{Card}(A_j) &= \sum_{k=1}^{n} \mathrm{Card}(B_{j,k}) \\
&= \sum_{k=j}^{n} \frac{(k-1)!}{(k-j)!}\,(n-j)! \\
&= (n-j)! \sum_{k=0}^{n-j} \frac{(k+j-1)!}{k!} \\
&= (j-1)! \times (n-j)! \sum_{k=0}^{n-j} \binom{k+j-1}{j-1}.
\end{aligned}$$

On montre facilement par récurrence sur l'entier n, avec $n \geqslant j$, la formule de Pascal permettant d'avoir l'hérédité :

$$\sum_{k=0}^{n-j} \binom{k+j-1}{j-1} = \binom{n}{j}.$$

Conclusion,

$$\mathbb{P}(A_j) = \binom{n}{j} \times (j-1)! \times (n-j)! \times \frac{1}{n!} = \frac{1}{j}.$$

La probabilité que la $j^{ème}$ année soit une année record vaut $\frac{1}{j}$.
On peut interpréter cette formule.
Il semble assez normal que cette probabilité décroisse en fonction de j, puisque plus les années passent, plus l'échantillon des mesures a de chances de contenir une grande valeur de crue.
Lorsque $j = 1$, la première mesure est nécessairement une mesure record. Il est donc normal de trouver $\mathbb{P}(A_1) = 1$.

Passons maintenant à la loi de Bernoulli. Il s'agit de la loi usuelle la plus simple et parmi toutes les probabilités, ce sont les plus simples juste après les mesures de Dirac.
Pour définir une loi de Bernoulli, on a besoin d'un paramètre p appartenant à $[0, 1]$.
La loi de Bernoulli est alors la probabilité :

$$\mathscr{B}(p) = (1-p)\ \delta_0 + p\ \delta_1.$$

On dit qu'une variable aléatoire $X : \Omega \longrightarrow \mathbb{R}$ suit la loi de Bernoulli de paramètre p si :

$$\boxed{\mathbb{P}_X = \mathscr{B}(p)}$$

Dans ce cas, on notera $X \sim \mathscr{B}(p)$ et pour toute partie A incluse dans $\mathbb{R}$, que la partie A soit finie ou infinie, peu importe, on peut en déduire :

$$\mathbb{P}(X \in A) = (1-p) \cdot \mathbf{1}_A(0) + p \cdot \mathbf{1}_A(1),$$

cette probabilité valant ou bien 0 – si $A \cap \{0, 1\} = \emptyset$, ou bien $1-p$ – si $A \cap \{0, 1\} = \{0\}$, ou bien p – si $A \cap \{0, 1\} = \{1\}$, ou bien 1 – si $\{0, 1\} \subset A$.

La propriété 14 nous indique les formules à connaître impérativement pour l'espérancec et la variance d'une variable aléatoire X telle que $X \sim \mathscr{B}(p)$.
Supposons donc que la variable $X : \Omega \longrightarrow \{0, 1\}$ suive la loi de Bernoulli de paramètre $p \in [0, 1]$.
On calcule directement l'espérance :

$$\mathbb{E}(X) = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = p.$$

On calcule ensuite la variance en calculant $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = 0^2\ \mathbb{P}(X = 0) + 1^2\ \mathbb{P}(X = 1) = p.$$

En fait, on aurait pu remarquer que la variable X prend presque sûrement ses valeurs dans l'ensemble à deux éléments $\{0,1\}$ et donc presque sûrement, on a l'égalité :

$$X^2 = X.$$

Il est donc normal de trouver la même espérance pour les variables X et X^2.
Il suffit finalement d'appliquer la formule de Huyghens :

$$V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = p - p^2 = p(1-p).$$

Il n'y a pas grand chose à dire d'autre sur cette loi simple. Passons à des exemples d'applications à travers l'exemple 14.
• Dès que l'on a une situation avec deux issues possibles [Vrai/Faux], [Pile/Face], l'une des deux issues étant codée par 0 et l'autre par 1, on a affaire à une variable aléatoire qui suit automatiquement une loi de Bernoulli.

• Le deuxième point est **très important** et a déjà été utilisé plusieurs fois dans les feuilles précédentes, en particulier pour démontrer la formule de transfert sur les vecteurs ou la formule standard.
Si A est un événement dans $\mathscr{P}(\Omega)$, il est clair que l'indicatrice $\mathbf{1}_A$ ne prend que deux valeurs possibles : 0 ou 1. De plus,

$$\{\mathbf{1}_A = 0\} = \overline{A} \text{ et } \{\mathbf{1}_A = 1\} = A,$$

donc :

$$\mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(A) \text{ et } \mathbb{P}(\mathbf{1}_A = 0) = 1 - \mathbb{P}(A).$$

Par conséquent :

$$\mathbf{1}_A \sim \mathscr{B}\Big(\mathbb{P}(A)\Big).$$

On a alors immédiatement,

$$\mathbb{E}(1_A) = \mathbb{P}(A)$$

résultat fondamental dans l'application de la loi forte des grands nombres pour interpréter la quantité $\mathbb{P}(A)$ comme une proportion limite de réalisations de l'événement A sur un grand nombre d'essais indépendants.

• Le dernier point est plus subtil.

1. La variable M prend ses valeurs dans l'ensemble des matrices à coefficients dans $\mathbb{Z}/2\mathbb{Z}$. Ces matrices dans $\mathscr{M}_n(\mathbb{Z}/2\mathbb{Z})$ sont au nombre de 2^{n^2}. En effet, pour les choisir, il s'agit de choisir pour chacun des n^2 coefficients le coefficient entre 0 et 1, ce qui laisse à chaque fois deux possibilités.
   On compte maintenant le nombre de matrices favorables à l'événement A : « la matrice choisie est inversible ».
   On va compter le nombre de matrices inversibles dans $\mathscr{M}_n(\mathbb{Z}/2\mathbb{Z})$ en choisissant les coefficients colonne par colonne.
   Pour choisir une matrice P inversible dans $\mathscr{M}_n(\mathbb{Z}/2\mathbb{Z})$, il s'agit de choisir n'importe quelle colonne non nulle pour la première colonne C_1, ce qui donne $2^n - 1$ possibilités pour C_1.

Une fois la colonne C_1 fixée, il y a $2^n - 2$ choix pour la colonne C_2 de façon à ce que la colonne C_2 ne soit pas colinéaire à la colonne C_1. En effet, dans l'espace $\mathscr{M}_{n,1}(\mathbb{Z}/2\mathbb{Z})$, il n'y a que deux vecteurs dans la droite vectorielle $\mathrm{Vect}(C_1)$ car comme le corps est $K = \mathbb{Z}/2\mathbb{Z}$, alors :

$$\mathrm{Vect}(C_1) = \{0, C_1\}.$$

Imaginons que les k premières colonnes $C_1, \cdots, C_k$ soient construites de façon à ce qu'elles forment une famille libre.
On peut choisir pour la colonne C_{k+1} n'importe quelle colonne dans l'ensemble

$$\mathscr{M}_{n,1}(\mathbb{Z}/2\mathbb{Z}) \setminus \mathrm{Vect}(C_1, \cdots, C_k).$$

L'ensemble $\mathrm{Vect}(C_1, \cdots, C_k)$ compte exactement 2^k éléments car l'application suivante :

$$\varphi : \left| \begin{array}{ccc} (\mathbb{Z}/2\mathbb{Z})^k & \longrightarrow & \mathrm{Vect}(C_1, \cdots, C_k) \\ (\lambda_1, \cdots, \lambda_k) & \longmapsto & \displaystyle\sum_{i=1}^{k} \lambda_i \cdot C_i \end{array} \right.$$

est un isomorphisme donc une bijection.
Il y a donc $2^n - 2^k$ choix possibles pour la colonne C_{k+1}.
En définitive, il y a $\displaystyle\prod_{k=0}^{n-1} \left(2^n - 2^k\right)$ choix possibles de matrices inversibles.
La probabilité que la matrice M soit inversible est donc égale à :

$$\begin{aligned} \mathbb{P}(M \text{ inversible }) &= \frac{1}{2^{n^2}} \prod_{k=0}^{n-1} \left(2^n - 2^k\right) \\ &= \prod_{k=0}^{n-1} \frac{\left(2^n - 2^k\right)}{2^n} \\ &= \prod_{k=0}^{n-1} \left(1 - \frac{1}{2^{n-k}}\right) \\ &= \prod_{k=1}^{n} \left(1 - \frac{1}{2^k}\right). \end{aligned}$$

2. Un programme Python nous indique que la probabilité que la matrice M soit inversible lorsque $n = 100$ vaut environ :

$$p \simeq 0.2887$$

On note maintenant notre gain algébrique – donc potentiellement négatif en cas de perte – $X : \Omega \longrightarrow \mathbb{R}$ au jeu qu'on nous propose.
Par les règles du jeu, en notant A l'événement : « la matrice piochée est inversible dans $\mathscr{M}_{100}(\mathbb{Z}/2\mathbb{Z})$ », alors :

$$X = -100 + 500 \cdot \mathbf{1}_A.$$

On en déduit par linéarité de l'espérance :

$$\mathbb{E}(X) = -100 + 500\ \mathbb{E}(\mathbf{1}_A) = -100 + 500 \cdot \mathbb{P}(A) \simeq 44.39 \ \text{ euros}.$$

Donc, la réponse est : « Oui, il vaut mieux jouer à ce jeu ! »

En effet, la loi forte des grands nombres nous dit qu'en moyenne, on aura un gain algébrique d'environ 44 euros. En moyenne, on est gagnant.

Si l'on jouait un grand nombre de fois à ce jeu, presque sûrement notre gain après n parties serait équivalent à $\mathbb{E}(X)\ n$, lorsque n tend vers $+\infty$, donc presque sûrement, notre gain algébrique tend vers $+\infty$, si l'on joue une infinité de fois à ce jeu.

Terminons la liste avec la loi binomiale.

Pour constituer une loi binomiale, nous avons besoin de deux ingrédients : d'une part, un entier naturel n, et d'autre part un réel p dans $[0, 1]$.
La loi binomiale de paramètres n et p est la probabilité notée $\mathscr{B}(n, p)$ telle que :

$$\mathscr{B}(n, p) = \sum_{k=0}^{n} \binom{n}{k}\ p^k\ (1-p)^{n-k}\ \delta_k.$$

Parmi les lois usuelles, (de première et de deuxième année), c'est en général la plus difficile à détecter dans une situation concrète donnée.
Une variable aléatoire $X : \Omega \longrightarrow \mathbb{R}$ suit la loi binomiale de paramètres n et p si :

$$\boxed{\mathbb{P}_X = \mathscr{B}(n, p)}$$

Dans ce cas, la variable X prend presque sûrement ses valeurs dans $[\![0, n]\!]$ et pour toute partie A incluse dans $\mathbb{R}$, on peut écrire :

$$\mathbb{P}(X \in A) = \sum_{k \in A \cap [\![0,n]\!]} \binom{n}{k}\ p^k\ (1-p)^{n-k}.$$

• Le premier point de la propriété 15 est **très important** car il permet de faire le lien entre la loi binomiale et des lois de Bernoulli.
On considère n variables aléatoires $X_1, \cdots, X_n$ toutes définies sur le même univers Ω, mutuellement indépendantes et suivant toutes la loi de Bernoulli de paramètre p.
Comme chaque X_i prend ses valeurs dans $\{0, 1\}$, alors la somme $X_1 + \cdots + X_n$ que l'on note Y prend ses valeurs dans l'ensemble $[\![0, n]\!]$. En effet, si ω est un élément de Ω, alors la quantité

$$Y(\omega) = \sum_{i=1}^{n} X_i(\omega)$$

est une somme de termes valant 0 ou 1 : la valeur $Y(\omega)$ de cette somme est égale au nombre de termes valant 1, ce qui donne assurément un entier entre 0 et n.
Fixons maintenant un entier k entre 0 et n et calculons la probabilité $\mathbb{P}(Y = k)$.
On remarque que la famille $\mathscr{F} = \left(\left\{X_1 = \varepsilon_1, \cdots, X_n = \varepsilon_n\right\}\right)_{(\varepsilon_1, \cdots, \varepsilon_n) \in \{0,1\}^n}$ est un système complet d'événements.

De plus, on remarque que l'on a pour tout $(\varepsilon_1, \cdots, \varepsilon_n)$ dans $\{0,1\}^n$:

$$\Big\{X_1 = \varepsilon_1, \cdots, X_n = \varepsilon_n\Big\} = \left\{Y = \sum_{i=1}^{n} \varepsilon_i, X_1 = \varepsilon_1, \cdots, X_n = \varepsilon_n\right\}.$$

On en déduit :

$$\begin{aligned}
\mathbb{P}(Y = k) &= \sum_{(\varepsilon_1,\cdots,\varepsilon_n)\in\{0,1\}^n} \mathbb{P}(Y = k, X_1 = \varepsilon_1, \cdots, X_n = \varepsilon_n) \qquad \text{[formule des probabilités totales]} \\
&= \sum_{\left[(\varepsilon_1,\cdots,\varepsilon_n)\in\{0,1\}^n \text{ et } \sum\limits_{i=1}^{n} \varepsilon_i = k\right]} \mathbb{P}(Y = k, X_1 = \varepsilon_1, \cdots, X_n = \varepsilon_n) + \\
&\qquad \sum_{\left[(\varepsilon_1,\cdots,\varepsilon_n)\in\{0,1\}^n \text{ et } \sum\limits_{i=1}^{n} \varepsilon_i \neq k\right]} \mathbb{P}(Y = k, X_1 = \varepsilon_1, \cdots, X_n = \varepsilon_n) \qquad [\sigma\text{-additivité}] \\
&= \sum_{\left[(\varepsilon_1,\cdots,\varepsilon_n)\in\{0,1\}^n \text{ et } \sum\limits_{i=1}^{n} \varepsilon_i = k\right]} \mathbb{P}(X_1 = \varepsilon_1, \cdots, X_n = \varepsilon_n) \\
&= \sum_{\left[(\varepsilon_1,\cdots,\varepsilon_n)\in\{0,1\}^n \text{ et } \sum\limits_{i=1}^{n} \varepsilon_i = k\right]} \prod_{i=1}^{n} \mathbb{P}(X_i = \varepsilon_i) \qquad \text{[indépendance des } X_i\text{]} \\
&= \sum_{\left[(\varepsilon_1,\cdots,\varepsilon_n)\in\{0,1\}^n \text{ et } \sum\limits_{i=1}^{n} \varepsilon_i = k\right]} \left(p^k\,(1-p)^{n-k}\right) \qquad \left[\begin{array}{c}\text{pour } k \text{ indices } i, \text{ on a } \varepsilon_i = 1 \\ \text{et pour les } (n-k) \text{ autres indices } j \\ \text{on a } \varepsilon_j = 0\end{array}\right] \\
&= \binom{n}{k} p^k\,(1-p)^{n-k}
\end{aligned}$$

cette dernière égalité provenant du fait que la somme $\displaystyle\sum_{\left[(\varepsilon_1,\cdots,\varepsilon_n)\in\{0,1\}^n \text{ et } \sum\limits_{i=1}^{n} \varepsilon_i = k\right]} \star$ compte autant de termes qu'il y a de façons de choisir k indices parmi n indices, les k indices choisis correspondant aux k indices i tels que $\varepsilon_i = 1$, qui est bien la seule possibilité pour avoir :

$$\sum_{i=1}^{n} \varepsilon_i = k.$$

On voit donc que la variable Y suit la loi binomiale $\mathscr{B}(n,p)$.

Une remarque à propos de cette sommation et en général de la distinction de cas en probabilités.
Bien souvent, on effectue une disjonction de cas en probabilités pour mettre en place un système complet d'événements, et c'est ce que l'on fait d'ailleurs ici pour avoir $Y \sim \mathscr{B}(n,p)$.
En terminale, peut-être manipuliez-vous des arbres pour faire vos distinctions de cas. Si vous en avez l'habitude, vous pouvez les conserver dans vos raisonnements au brouillon.
Vous remarquerez que dans ce cours, je n'en fais jamais usage. Il est clairement dit dans les rapports officiels de concours que le fait d'utiliser un arbre à l'écrit ne rapporte aucun point.
Il faut mentionner la propriété utilisée – souvent la formule des probabilités totales ou la formule des probabilités composées – en faisant apparaître le cas échéant le SCE utilisé – pour la formule des probabilités totales.
Si cela vous aide d'utiliser des arbres, utilisez-les uniquement sur votre brouillon.

• Le point suivant nous donne les formules de l'espérance et de la variance d'une variable aléatoire suivant la loi binomiale.
On peut tout à fait utiliser la formule de l'espérance ou de la variance pour démontrer ces formules. On pourra interpréter les sommes :

$$\sum_{k=0}^{n} k \binom{n}{k} t^{k-1} \text{ et } \sum_{k=0}^{n} k(k-1) \binom{n}{k} t^{k-2},$$

comme les deux premières dérivées de la fonction :

$$f : t \longmapsto \sum_{k=0}^{n} \binom{n}{k} t^k = (1+t)^n.$$

Je vous suggère de faire les calculs par vous-même pour manipuler ces sommations et retrouver les formules données.
Nous allons utiliser le premier point, pour plus d'efficacité.
Considérons n variables aléatoires $X_1, \cdots, X_n$, mutuellement indépendantes et suivant chacune la loi $\mathscr{B}(p)$.
On sait que la variable $X = X_1 + \cdots + X_n$ suit la loi binomiale $\mathscr{B}(n,p)$ et que l'espérance et la variance ne dépendent pas de la variable considérée mais plutôt de sa loi.
Par linéarité de l'espérance, on en déduit :

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) \\
&= \sum_{i=1}^{n} \mathbb{E}(X_i) \\
&= \sum_{i=1}^{n} p \text{ , car } X_i \sim \mathscr{B}(p) \\
&= n\ p.
\end{aligned}$$

Pour la variance, la variance de la somme des X_i est la somme des variances par indépendance deux à deux – qui est impliquée par la mutuelle indépendance :

$$\begin{aligned}
V(X) &= V\left(\sum_{i=1}^{n} X_i\right) \\
&= \sum_{i=1}^{n} V(X_i) \qquad \text{[indépendance deux à deux]} \\
&= \sum_{i=1}^{n} p\,(1-p)\ , \text{ car } X_i \sim \mathscr{B}(p) \\
&= n\,p\,(1-p).
\end{aligned}$$

• Pour le dernier point, on considère deux variables aléatoires $X : \Omega \longrightarrow \mathbb{R}$ et $Y : \Omega \longrightarrow \mathbb{R}$ indépendantes, telles que :

$$X \sim \mathscr{B}(n,p) \text{ et } Y \sim \mathscr{B}(m,p),$$

le deuxième paramètre p étant commun aux deux lois binomiales, et n puis m étant deux entiers strictement positifs.
On va utiliser le résultat hors programme mais utile suivant : si $(\Omega, \mathbb{P})$ est un espace probabilisé, si $Z : \Omega \longrightarrow \mathbb{R}$ est une variable aléatoire finie dont on note $\mathscr{L}$ la loi de probabilité – la loi $\mathscr{L}$ est donc une probabilité $\mathscr{L} : \mathscr{P}(Z(\Omega)) \longrightarrow [0,1]$, on peut construire un espace probabilisé $(\Omega', \mathbb{P}')$ et une suite $\left(Z_k : \Omega' \longrightarrow \mathbb{R}\right)_{k\in\mathbb{N}}$ de variables aléatoires mutuellement indépendantes telle que chaque variable Z_k suive la même loi que la variable Z du début.
Ce résultat est difficile à montrer et peut-être intuitivement compréhensible. Il suffit de considérer des réalisations de la mesure Z une infinité de fois, ces mesures étant indépendantes et toujours réalisées dans les mêmes conditions.
Par exemple, si on lance un dé à n faces numérotées de 1 à n, si on effectue une infinité de lancers de ce dé et que l'on note au $k^{\text{ème}}$ lancer le numéro Z_k obtenu, on obtient une suite $(Z_k)_{k\in\mathbb{N}}$ de va.i.id de loi $\mathscr{U}_n$. Le nouvel univers Ω' est grosso modo ici un ensemble du type :

$$\Omega' = \Omega^{\mathbb{N}^*},$$

ensemble des suites indexées par $\mathbb{N}^*$ et à valeurs dans Ω.
On voit que l'univers Ω' est en général infini mais pas dénombrable – dès que Ω compte au moins deux éléments. Bref, les choses sont difficiles à montrer, hors de portée ici.

On revient au sujet.
On considère $n+m$ variables aléatoires $Z_1, \cdots, Z_n, Z_{n+1}, \cdots, Z_{n+m}$ mutuellement indépendantes, chaque variable $Z_i : \Omega' \longrightarrow \{0,1\}$ étant définie sur un même univers Ω' avec une probabilité $\mathbb{P}'$ suivant la loi de Bernoulli de paramètre p. L'existence est donc admise.
On pose $X' = Z_1 + \cdots + Z_n$ et $Y' = Z_{n+1} + \cdots + Z_{n+m}$. D'après le lemme des coalitions, les variables X' et Y' sont indépendantes.
D'après le premier point déjà démontré de la propriété 15, on sait que la variable X' suit la loi binomiale $\mathscr{B}(n,p)$ et que la variable Y' suit la loi binomiale $\mathscr{B}(m,p)$ car les variables $Z_1, \cdots, Z_n$ d'une part et les variables $Z_{n+1}, \cdots, Z_{n+m}$ d'autre part restent mutuellement indépendantes.

Toujours par ce premier point, comme la variable $X'+Y'$ est une somme de $n+m$ variables mutuellement indépendantes suivant chacune la loi de Bernoulli de paramètre p, alors la variable $X'+Y'$ suit la loi binomiale $\mathscr{B}(n+m,p)$.
On va calculer la loi de $X+Y$ en fixant un réel x et en calculant $\mathbb{P}(X+Y=x)$.
Voici le détail du calcul :

$$\begin{aligned}
\mathbb{P}(X+Y=x) &= \sum_{k=0}^{n} \mathbb{P}(X+Y=x, X=k) \quad \left[\begin{array}{c}\text{utiliser le SCE } \Big(\{X=k\}\Big)_{0\leqslant k\leqslant n} \\ \text{et la formule des probabilités totales}\end{array}\right] \\
&= \sum_{k=0}^{n} \mathbb{P}(X=k, Y=x-k) \quad \text{, car } \{X+Y=x, X=k\} = \{X=k, Y=x-k\} \\
&= \sum_{k=0}^{n} \mathbb{P}(X=k)\ \mathbb{P}(Y=x-k) \quad [\text{indépendance}] \\
&= \sum_{k=0}^{n} \mathbb{P}_X(\{k\}) \cdot \mathbb{P}_Y(\{x-k\}) \\
&= \sum_{k=0}^{n} \mathscr{B}(n,p)(\{k\}) \cdot \mathscr{B}(m,p)(\{x-k\}) \quad \text{, car } \mathbb{P}_X = \mathscr{B}(n,p) \text{ et } \mathbb{P}_Y = \mathscr{B}(m,p) \\
&= \sum_{k=0}^{n} \mathbb{P}'(X'=k) \cdot \mathbb{P}'(Y'=x-k) \quad \text{, car } \mathbb{P}'_{X'} = \mathbb{P}_X \text{ et } \mathbb{P}'_{Y'} = \mathbb{P}_Y \\
&= \sum_{k=0}^{n} \mathbb{P}'(X'=k, Y'=x-k) \quad [\text{indépendance}] \\
&= \mathbb{P}'(X'+Y'=x) \quad \left[\begin{array}{c}\text{utiliser le SCE } \Big(\{X'=k\}\Big)_{0\leqslant k\leqslant n} \\ \text{et la formule des probabilités totales}\end{array}\right] \\
&= \mathbb{P}'_{X'+Y'}(\{x\}) \\
&= \mathscr{B}(n+m,p)(\{x\}).
\end{aligned}$$

Résultat des courses, on a bien :

$$X+Y \sim \mathscr{B}(n+m,p).$$

Détaillons maintenant les points de l'exemple 15.

Dans un sac contenant une proportion de boules blanches égale à $p \in [0,1]$, on effectue n tirages avec remise et on note X le nombre de boules blanches obtenues. Immédiatement, la variable X suit la loi binomiale $\mathscr{B}(n,p)$.
Pour s'en convaincre, on peut noter pour tout $i \in [\![1,n]\!]$, l'événement A_i : « *piocher une boule blanche au $i^{ème}$ tirage* ».
On remarque alors que :

$$X = \sum_{i=1}^{n} \mathbf{1}_{A_i}.$$

Comme les tirages sont avec remise, alors tous les tirages sont indépendants, ce qui assure la mutuelle indépendance des fonctions indicatrices $\mathbf{1}_{A_i}$ et finalement, chaque indicatrice $\mathbf{1}_{A_i}$ suit la loi de Bernoulli de paramètre p. On a bien $X \sim \mathscr{B}(n,p)$ dans ce cas.
Le nombre moyen de boules blanches piochées vaut $n\,p = \mathbb{E}(X)$, ce qui est plutôt conforme à l'intuition.

Pour l'exemple 16 qui est rédigé comme un exercice, voici ce que cela donne.

1. Lorsque l'on regarde l'agencement des questions, on se dit que l'on va commencer par calculer l'espérance $\mathbb{E}(Y)$ sans expliciter la loi de la variable Y qui est demandée en deuxième question. On va donc plutôt interpréter la variable Y comme une somme de variables plus simples.
   Pour tout entier i entre 1 et n et pour tout entier k entre 1 et p, on pose l'événement :

   $A_{i,k}$: « *l'élément i appartient à la partie X_k* ».

   On pose ensuite pour tout entier i entre 1 et n, la variable :

   $$Z_i = \prod_{k=1}^{p} \mathbf{1}_{A_{i,k}} = \mathbf{1}_{\left(\bigcap_{k=1}^{p} A_{i,k}\right)}.$$

   On remarque que l'on a l'égalité :

   $$Y = \sum_{i=1}^{n} Z_i, \qquad \bigstar$$

   car si ω est un élément de Ω, alors la quantité $\displaystyle\sum_{i=1}^{n} Z_i(\omega)$ est une somme composée de n termes valant 0 ou 1. La valeur de cette somme est égal au nombre de termes valant 1, autrement dit le nombre d'entiers i entre 1 et n tels que :

   $$i \in \bigcap_{k=1}^{p} X_k(\omega).$$

   Il s'agit bien du nombre d'éléments dans la partie $\displaystyle\bigcap_{k=1}^{p} X_k(\omega)$: c'est $Y(\omega)$.

   Par linéarité de la somme, on peut écrire :

   $$\mathbb{E}(Y) = \sum_{i=1}^{n} \mathbb{E}\left(\prod_{k=1}^{p} \mathbf{1}_{A_{i,k}}\right).$$

   Comme pour tout entier i, les variables $\mathbf{1}_{A_{i,1}}, \cdots, \mathbf{1}_{A_{i,n}}$ sont mutuellement indépendantes, d'après le lemme des coalitions et que l'indicatrice $\mathbf{1}_{A_{i,k}}$ ne dépend que de la variable X_k, alors :

   $$\mathbb{E}(Y) = \sum_{i=1}^{n} \prod_{k=1}^{p} \mathbb{E}\left(\mathbf{1}_{A_{i,k}}\right) = \sum_{i=1}^{n} \prod_{k=1}^{p} \mathbb{P}(A_{i,k}).$$

Or, la probabilité de l'événement $A_{i,k}$ peut se calculer par équiprobabilité.

Il y a en tout 2^n parties dans $\mathscr{P}(\llbracket 1, n \rrbracket)$.

Il y a exactement 2^{n-1} parties qui ne contiennent pas l'élément i, car ce sont toutes les parties incluses dans $\llbracket 1, n \rrbracket \setminus \{i\}$. Il y a donc $2^n - 2^{n-1} = 2^{n-1}$ parties contenant i et finalement :

$$\mathbb{P}(A_{i,k}) = \frac{1}{2}$$

puis :

$$\mathbb{E}(Y) = \sum_{i=1}^{n} \frac{1}{2^p} = \frac{n}{2^p}.$$

2. On ne peut pas exploiter immédiatement l'égalité ★, car les variables Z_i, même si elles suivent toutes une loi de Bernoulli ne sont pas forcément indépendantes car chaque Z_i dépend de toutes les variables $X_1, \cdots, X_p$.

   On fait autrement en faisant appel à notre sens probabiliste.

   Comme les variables X_k sont mutuellement indépendantes et de loi uniforme, pour former chaque partie aléatoire X_k, tout se passe comme si on passait en revue les entiers i entre 1 et n. Pour chaque tel entier i, on prend une pièce de monnaie parfaitement équilibrée. Si la pièce fait « *pile* » alors on incorpore l'élément i dans la partie aléatoire X_k et sinon on passe à l'entier suivant $i + 1$.

   En adoptant ce point de vue, tous les lancers sont indépendants les uns des autres en changeant d'indice i mais aussi en procédant à la construction d'une autre partie aléatoire X_ℓ.

   De ce point de vue, la variable Y s'interprète de la manière suivante : on considère l'entier $i = 1$. On procède à p lancers indépendants de pièce de monnaie. Si on n'a eu que des « *pile* », alors on compte $+1$ à un compteur initialisé à 0 et on recommence avec l'entier $i = 2$. Une fois tous les entiers i passés en revue, la valeur finale du compteur vaut $Y(\omega)$.

   De ce point de vue, la variable Y s'interprète comme une somme de p variables aléatoires indépendantes (car les lancers sont indépendants), chacune de ces p variables suivant la loi de Bernoulli de paramètre $\frac{1}{2^p}$ (égal à la probabilité de faire p fois « *pile* »).

   Immédiatement, on a :

$$Y \sim \mathscr{B}\left(n, \frac{1}{2^p}\right).$$

   On retrouve la valeur de l'espérance donnée en première question.

On termine le chapitre avec l'exercice 17, qui est un exercice récapitulatif.

1. Une chose est sûre : la variable X_2 dépend de la variable X_1.

   Il est clair que la variable X_1 suit la loi binomiale $\mathscr{B}(n, p)$.

   Ensuite, il est également assez clair que l'on a l'inclusion :

$$X_2(\Omega) \subset \llbracket 0, n \rrbracket.$$

   On remarque que dans cette expérience, l'univers Ω est assez compliqué à expliciter. On ne le fait pas et d'ailleurs, il est inutile d'expliciter Ω : on ne calcule que les lois de variables.

On va utiliser le SCE $\mathscr{F} = \Big(\{X_1 = k\}\Big)_{0 \leqslant k \leqslant n}$.
On fixe un entier i entre 0 et n. Alors :

$$\begin{array}{rcll} \mathbb{P}(X_2 = i) & = & \displaystyle\sum_{k=0}^{n} \mathbb{P}(X_2 = i, X_1 = k) & \text{[formule des probabilités totales]} \\ & = & \displaystyle\sum_{k=0}^{n} \mathbb{P}(X_2 = i \mid X_1 = k) \cdot P(X_1 = k) & \\ & = & \displaystyle\sum_{k=0}^{n} \mathscr{B}(k,p)(\{i\}) \cdot \mathbb{P}(X_1 = k) & \end{array}$$

car la probabilité de la variable X_2 sachant l'événement $\{X_1 = k\}$ est la loi binomiale de paramètres k et p, puisque sachant l'événement $\{X_1 = k\}$, on lance à la deuxième étape k pièces de monnaie.
On poursuit le calcul :

$$\begin{array}{rcl} \mathbb{P}(X_2 = i) & = & \displaystyle\sum_{k=i}^{n} \mathscr{B}(k,p)(\{i\}) \cdot \mathbb{P}(X_1 = k) \\ & = & \displaystyle\sum_{k=i}^{n} \binom{k}{i} \; p^i \; q^{k-i} \cdot \binom{n}{k} \; p^k \; q^{n-k} \\ & = & \displaystyle\sum_{k=i}^{n} \frac{n!}{i! \; (n-k)! \; (k-i)!} \; p^{k+i} \; q^{n-i} \\ & = & \displaystyle\binom{n}{i} \; p^i \; q^{n-i} \; \sum_{k=i}^{n} \binom{n-i}{k-i} p^k \\ & = & \displaystyle\binom{n}{i} \; p^i \; q^{n-i} \; \sum_{\ell=0}^{n-i} \binom{n-i}{\ell} \; p^{\ell+i} \\ & = & \displaystyle\binom{n}{i} \; p^i \; q^{n-i} \; p^i \; (1+p)^{n-i} \\ & = & \displaystyle\binom{n}{i} \; (p^2)^i \; (q(1+p))^{n-i}. \end{array}$$

En posant $\alpha = p^2$, alors :

$$1 - \alpha = 1 - p^2 = (1-p)(1+p) = q(1+p).$$

On en déduit que la probabilité $\mathbb{P}(X_2 = i)$ est égale à :

$$\mathbb{P}(X_2 = i) = \binom{n}{i} \; \alpha^i \; (1-\alpha)^{n-i}.$$

Autrement dit, la variable X_2 suit la loi binomiale $\mathscr{B}\,(n, p^2)$.

2. On note pour tout entier i entre 1 et n, le numéro Z_i de l'étape où la pièce numéro i donnera « *face* » pour la première fois, avec $Z_i = +\infty$ si la $i^{ème}$ pièce donne toujours « *pile* ».

   On remarque que l'on a :

$$X_k = \sum_{i=1}^{n} \mathbf{1}_{\{Z_i \geqslant k+1\}}.$$

   En effet, si ω est un élément de Ω, alors $X_k(\omega)$ est le nombre de pièces encore présentes à l'issue de la $k^{ème}$ étape.

   D'autre part, la quantité $\displaystyle\sum_{i=1}^{n} \mathbf{1}_{\{Z_i \geqslant k+1\}}(\omega)$ est le nombre d'entiers i entre 1 et n tels que $Z_i(\omega) \geqslant k+1$, autrement dit, le nombre de pièces qui ont donné « *pile* » aux k premières étapes.

   Il s'agit du même nombre.

   Ensuite, chaque indicatrice $\mathbf{1}_{\{Z_i \geqslant k+1\}}$ suit la loi de Bernoulli de paramètre :

$$\mathbb{P}(Z_i \geqslant k+1).$$

   La probabilité de l'événement $\{Z_i \geqslant k+1\}$ est la probabilité d'obtenir k fois « *pile* », probabilité égale à p^k.

   Finalement, les indicatrices $\mathbf{1}_{\{Z_i \geqslant k+1\}}$ ne dépendent que de lancers disjoints. Par le lemme des coalitions, ces indicatrices sont mutuellement indépendantes.

   Résultat des courses, la variable X_k suit la loi binomiale de paramètres n et p^k.

   On retrouve le résultat de la question précédente.

3. Avec les notations déjà introduites, on peut écrire :

$$Y = \max\{Z_1, \cdots, Z_n\}.$$

   Cette question empiète largement sur le programme de deuxième année. Je vous laisse lire l'annexe de cours pour consulter la définition d'une probabilité dans le cas d'un univers infini. En particulier, la définition de la σ-additivité n'est pas tout à fait la même et prolonge celle connue sur les univers finis.

   On a clairement :

$$Y(\Omega) \subset \mathbb{N}^* \cup \{+\infty\}.$$

   Fixons un entier k dans $\mathbb{N}^*$.

   On remarque que l'on a les égalités suivantes :

$$\begin{aligned} \{Y = k\} &= \Big\{\max\{Z_1, \cdots, Z_n\} = k\Big\} \\ &= \Big\{\max\{Z_1, \cdots, Z_n\} \leqslant k\Big\} \setminus \Big\{\max\{Z_1, \cdots, Z_n\} \leqslant k-1\Big\}. \end{aligned}$$

   Or, pour tout entier $\ell \in \mathbb{N}$, l'événement $\Big\{\max\{Z_1, \cdots, Z_n\} \leqslant \ell\Big\}$ est exactement l'événement :

$$\bigcap_{i=1}^{n} \{Z_i \leqslant \ell\}.$$

La probabilité de cet événement vaut par indépendance des Z_i :

$$\prod_{i=1}^{n} \mathbb{P}(Z_i \leqslant \ell) = \prod_{i=1}^{n} (1 - \mathbb{P}(Z_i > \ell)) = (1 - p^{\ell})^n.$$

Par conséquent,

$$\mathbb{P}(Y = k) = (1 - p^k)^n - (1 - p^{k-1})^n.$$

La famille $\mathscr{F} = \Big(\{Y = k\}\ ;\ k \in \mathbb{N}^*; \{Y = +\infty\}\Big)$ forme un système complet d'événements.

On en déduit par σ-additivité :

$$1 = \sum_{k=1}^{+\infty} \mathbb{P}(Y = k) + \mathbb{P}(Y = +\infty).$$

La série convergente $\displaystyle\sum_{k=1}^{+\infty} \mathbb{P}(Y = k) = \sum_{k=1}^{+\infty} \left((1 - p^k)^n - (1 - p^{k-1})^n\right)$ est une série télescopique de somme égale à $\displaystyle\lim_{k \longrightarrow +\infty} (1 - p^k)^n - (1 - p^0)^n = 1$, car $|p| < 1$.

On en déduit :

$$\mathbb{P}(Y = +\infty) = 0.$$

On a complètement la loi de la variable Y.

La probabilité que le jeu ne s'arrête jamais vaut $\mathbb{P}(Y = +\infty) = 0$. Ceci doit être conforme à votre intuition probabiliste.

Les probabilités sur les univers infinis

Vous verrez en deuxième année que lorsque l'univers Ω est infini, on peut définir une probabilité $\mathbb{P}$ sur autre chose que l'ensemble $\mathscr{P}(\Omega)$, mais passons. Ceci n'est pas très important.

Ce qui est important, ce sont les choses suivantes :

- une expérience aléatoire reste notée ω
- l'ensemble Ω de toutes les expériences possibles forme un ensemble éventuellement infini. C'est le cas typiquement lorsqu'on lance une infinité de fois une pièce de monnaie où l'univers Ω est l'ensemble $\{P, F\}^{N^*}$ des suites à valeurs dans l'ensemble à deux éléments $\{P, F\}$
- un événement A est une partie de Ω. Vous verrez en deuxième année que l'ensemble de tous les événements forme ce qu'on appelle une tribu, qui peut être différente de l'ensemble $\mathscr{P}(\Omega)$ de toutes les parties de Ω. Dans cette annexe, on part du principe que la tribu des événements envisagée est l'ensemble $\mathscr{P}(\Omega)$ de toutes les parties de Ω. Ce point n'est pas très important, de toutes façons ;
- une probabilité $\mathbb{P}$ reste dans cette configuration une application $\mathbb{P} : \mathscr{P}(\Omega) \longrightarrow [0, 1]$ telle que :
  $\longrightarrow$ la probabilité de l'événement certain Ω vaut 1

$\longrightarrow$ la σ-additivité est vérifie ; celle-ci consiste à dire que pour toute famille $(A_n)_{n\in\mathbb{N}}$ d'événements deux à deux disjoints, alors la série à termes positifs $\sum_n \mathbb{P}(A_n)$ est toujours convergente et la somme de cette série vaut la probabilité de la réunion :

$$\mathbb{P}\left(\bigsqcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n).$$

Cette définition prolonge la définition connue lorsque l'univers Ω est fini. En effet, en prenant la famille $(\emptyset)_{n\in\mathbb{N}}$ des événements vides, alors cette famille est à événements deux à deux disjoints, donc la série $\sum_n \mathbb{P}(\emptyset)$ converge.

Or, cette série convergente est de terme général constant égal à $\mathbb{P}(\emptyset)$ et ce terme doit tendre vers 0 par convergence de la série. Résultat des courses :

$$\mathbb{P}(\emptyset) = 0.$$

On en déduit que si A et B sont deux événements disjoints, en construisant la famille $(A_n)_{n\in\mathbb{N}}$ de telle sorte que :

$$A_0 = A,\ A_1 = B \text{ et } \forall n \geqslant 2,\ A_n = \emptyset,$$

alors la réunion des A_n forme $A \sqcup B$ et la somme de la série convergente vaut uniquement la somme de ses deux premiers termes, d'où :

$$\mathbb{P}(A \sqcup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

La propriété 16 fait écho au dernier point de la propriété 1. Vous verrez que la σ-additivité est un peu technique à montrer... je ne développe pas, cela nous entraînerait trop loin.

L'exemple 18 peut être détaillé.

• En utilisant la propriété 16 admise, on remarque que la famille $\left(p_k = \dfrac{1}{2^{k+1}}\right)_{k\in\mathbb{N}}$ est une famille de réels positifs de série convergente égale à :

$$\sum_{k=0}^{+\infty} \frac{1}{2^{k+1}} = \frac{1}{2} \times \frac{1}{1-\frac{1}{2}} = 1.$$

La fonction $\mathbb{P}$ proposée est directement reliée à cette famille sommable positive de somme égale à 1.

• Là encore, la suite $\left(\pi_k = \dfrac{1}{\zeta(s)}\ \dfrac{1}{k^s}\right)_{k\in\mathbb{N}^*}$ est une suite de réels positifs de série convergente, et la somme de la série vaut bien 1, puisque :

$$\sum_{k=1}^{+\infty} \frac{1}{\zeta(s)}\ \frac{1}{k^s} = \frac{1}{\zeta(s)} \sum_{k=1}^{+\infty} \frac{1}{k^s} = \frac{1}{\zeta(s)}\ \zeta(s) = 1.$$

On résout alors les deux questions posées.

1. Montrer que la famille infinie $(A_m)_{m\in\mathscr{Q}}$ d'événements sont mutuellement indépendants revient à montrer que pour tout entier $N \geqslant 2$, la famille **finie** $(A_m)_{m\in\mathscr{Q}\cap[\![1,N]\!]}$ est mutuellement indépendante.

   Cela revient à montrer que pour toute partie **finie** I incluse dans $\mathscr{Q}$, on a :

$$\mathbb{P}\left(\bigcap_{m\in I} A_m\right) = \prod_{m\in I} \mathbb{P}(A_m).$$

   Fixons une partie finie $I = \{p_1, \cdots, p_r\}$ incluse dans $\mathscr{Q}$. Les nombres p_i sont donc des nombres premiers différents.

   On va commencer par calculer la probabilité de l'événement A_m, pour tout entier $m \geqslant 1$.

   On en déduit par définition de la probabilité $\mathbb{P}$ de l'énoncé :

$$\begin{aligned}
\mathbb{P}(A_m) &= \sum_{k\in A_m} \frac{1}{\zeta(s)} \frac{1}{k^s} \\
&= \sum_{\ell=1}^{+\infty} \frac{1}{\zeta(s)} \frac{1}{(\ell m)^s} \\
&= \frac{1}{\zeta(s)} \sum_{\ell=1}^{+\infty} \frac{1}{\ell^s} \frac{1}{m^s} \\
&= \frac{1}{\zeta(s)} \frac{1}{m^s} \sum_{\ell=1}^{+\infty} \frac{1}{\ell^s} \\
&= \frac{1}{\zeta(s)} \frac{1}{m^s} \zeta(s) \\
&= \frac{1}{m^s}.
\end{aligned}$$

   Or, on remarque que l'événement A_m n'est autre que l'événement « *être multiple de l'entier m* » comme dirait le probabiliste.

   On en déduit que l'événement $\bigcap_{m\in I} A_m$ est : « *être multiple des nombres premiers $p_1, \cdots, p_r$* », ce qui est tout à fait équivalent grâce au théorème de Gauss à l'événement : « *être multiple de l'entier $N = \prod_{k=1}^{s} p_k$* ».

   En d'autres termes :

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{m\in I} A_m\right) &= \mathbb{P}(A_N) \\
&= \frac{1}{N^s} \\
&= \prod_{k=1}^{s} \frac{1}{p_k^s}
\end{aligned}$$

$$\begin{aligned} &= \prod_{k=1}^{s} \mathbb{P}(A_{p_k}) \\ &= \prod_{m\in I} \mathbb{P}(A_m). \end{aligned}$$

On a bien l'indépendance mutuelle des événements A_m.

2. On remarque que l'entier 1 est le seul entier divisible par aucun nombre premier.
Par conséquent,
$$\{1\} = \bigcap_{m\in\mathscr{Q}} \overline{A_m}.$$
En utilisant le deuxième point de la propriété 5, on observe que la famille $(\overline{A_m})_{m\in\mathscr{Q}}$ reste mutuellement indépendante.
Pour tout entier $N \geqslant 2$, on en déduit :
$$\begin{aligned} \mathbb{P}\left(\bigcap_{m\in\mathscr{Q}\cap[\![1,N]\!]} \overline{A_m}\right) &= \prod_{m\in\mathscr{Q}\cap[\![1,N]\!]} \mathbb{P}(\overline{A_m}) \\ &= \prod_{m\in\mathscr{Q}\cap[\![1,N]\!]} \left(1-\frac{1}{m^s}\right). \end{aligned}$$
Vous verrez en deuxième année que l'on pourra passer à la limite lorsque N tend vers $+\infty$, dans cette égalité, ce qui donne :
$$\mathbb{P}\left(\bigcap_{m\in\mathscr{Q}} \overline{A_m}\right) = \prod_{m\in\mathscr{Q}} \left(1-\frac{1}{m^s}\right).$$
Autrement dit :
$$\mathbb{P}(\{1\}) = \prod_{m\in\mathscr{Q}} \left(1-\frac{1}{m^s}\right).$$
Or, $\mathbb{P}(\{1\}) = \dfrac{1}{\zeta(s)}$, ce qui conduit à la formule désirée en passant à l'inverse.

En ce qui concerne les variables aléatoires discrètes $X : \Omega \longrightarrow \mathbb{R}$, il n'y a pas grand chose de neuf concernant la loi $\mathbb{P}_X$.

En ce qui concerne l'espérance d'une variable aléatoire discrète $X : \Omega \longrightarrow \mathbb{R}$ prenant un nombre fini ou dénombrable de valeurs possibles, la variable X peut ne pas admettre une espérance. Cela dépend de l'absolue convergence de la série $\displaystyle\sum_{x\in X(\Omega)} |x|\ \mathbb{P}(X=x)$.
Viennent alors se greffer des problèmes de convergence pour également les variances, les formules de transfert, etc.

Je laisse le reste à la deuxième année...