Probabilités

Table des matières

1 Espaces probabilisés

Définition. On fixe un ensemble Ω qui sera ensuite appelé **l'univers**.
On appelle tribu, ou σ-algèbre sur Ω tout ensemble $\mathcal{F}$ de parties de Ω (ainsi $\mathcal{F} \subset \mathcal{P}(\Omega)$) vérifiant :
i) $\Omega \in \mathcal{F}$,
ii) $\mathcal{F}$ est stable par passage au complémentaire : si $F \in \mathcal{F}$ alors $\Omega \setminus F \in \mathcal{F}$.
iii) $\mathcal{F}$ est stable par réunion dénombrable : si $(F_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$, alors $\bigcup_{n\in\mathbb{N}} F_n \in \mathcal{F}$.

Vocabulaire spécifique aux probabilités :
Avec les notations précédentes,
◇ Les éléments de $\mathcal{F}$ s'appellent les **événements**.
◇ Si $\{\omega\} \in \mathcal{F}$, on dit que c'est un **événement élémentaire**.
◇ $\emptyset$ est l'événement impossible.
◇ Si A est un événement, $\Omega \setminus A$ est l'événement contraire de A.
◇ Si A et B sont deux événements, $A \cap B$ est l'événement "A et B", $A \cup B$ est l'événement "A ou B". Lorsque $A \cap B = \emptyset$, les deux événement A et B sont dits incompatibles.

Exemple. On lance successivement deux dés. On peut modéliser cette situation en prenant comme univers $\Omega = \{1, \dots, 6\}^2$.
— L'événement « le second lancer est un 6 » est égal à $\{(m, 6)/m \in \{1, \dots, 6\}\}$.
— L'événement « le premier lancer est supérieur au second »
est égal à $\{(m, n) \in \Omega/m \geq n\}$.
— L'événement « la somme des deux lancers est paire »
est égal à $\{(m, n) \in \Omega/2|(m + n)\}$.

Définition. Soit $\mathcal{F}$ une tribu sur un univers Ω. On appelle système complet d'événements toute famille $(A_i)_{i\in I}$ (où I est fini ou dénombrable) d'événements 2 à 2 disjoints dont la réunion vaut Ω.

Exemple. Dans une urne, on a des cubes rouges ou verts et des boules rouges ou vertes. On tire un de ces objets :
Soit A_1 ="L'objet tiré est un cube" et A_2="L'objet tiré est une boule". Alors (A_1, A_2) est un système complet d'événements.
Soit B_1 ="L'objet tiré est rouge" et B_2="L'objet tiré est vert". Alors (B_1, B_2) est un système complet d'événements.
Soit C_1="L'objet tiré est une boule rouge", C_2 ="L'objet tiré est un cube rouge" et C_3 ="L'objet tiré est vert". Alors (C_1, C_2, C_3) est un système complet d'événements.

Définition. Si $\mathcal{F}$ est une tribu sur un univers Ω, on dit que $(\Omega, \mathcal{F})$ est un espace probabilisable.

Définition. Soit $(\Omega, \mathcal{F})$ un espace probabilisable. On dit que P est une probabilité sur $(\Omega, \mathcal{F})$ si et seulement si P est une application de $\mathcal{F}$ dans $[0, 1]$ telle que
i) $P(\Omega) = 1$ et

ii) pour toute suite $(F_n)_{n\in\mathbb{N}}$ d'événements de $\mathcal{F}$ deux à deux disjoints,
$P\left(\bigcup_{n=0}^{\infty} F_n\right) = \sum_{n=0}^{\infty} P(F_n)$.
Dans ce cas, le triplet $(\Omega, \mathcal{F}, P)$ est appelé un espace probabilisé.

Propriété. Avec les notations précédentes, pour $F, G, H, F_n \in \mathcal{F}$ on a :
◇ $P(\emptyset) = 0$,
◇ Si $F_0, \dots, F_p$ sont $p+1$ événements deux à deux disjoints, où $p \geq 1$,
alors $P\left(\bigcup_{n=0}^{p} F_n\right) = \sum_{n=0}^{p} P(F_n)$.
◇ $P(\overline{F}) = 1 - P(F)$ (où $\overline{F}$ désigne $\Omega \setminus F$),
◇ si $G \subset H$, $P(H \setminus G) = P(H) - P(G)$.
◇ si $G \subset H$, $P(G) \leq P(H)$ (on dit que P est croissante),
◇ $P(G \cup H) = P(G) + P(H) - P(G \cap H)$,
◇ **Inégalité de Boole :** $P\left(\bigcup_{n=0}^{\infty} F_n\right) \leq \sum_{n=0}^{\infty} P(F_n)$.

Notation. On notera souvent $P(G, H) \stackrel{\Delta}{=} P(G \cap H)$.
Ainsi, l'avant-dernière formule devient : $P(G \cup H) = P(G) + P(H) - P(G, H)$.

Démonstration.
◇ Posons, pour tout entier n, $F_n = \emptyset$. Alors $(F_n)_{n\in\mathbb{N}}$ est une suite d'événements de $\mathcal{F}$ deux à deux disjoints, donc $P\left(\bigcup_{n=0}^{\infty} F_n\right) = \sum_{n=0}^{\infty} P(F_n)$, ce qui donne $P(\emptyset) = \sum_{n=0}^{\infty} P(\emptyset)$,
donc $\sum_{n=1}^{\infty} P(\emptyset) = 0$ et $P(\emptyset) \in [0, 1]$: nécessairement $P(\emptyset) = 0$.
◇ Posons, pour tout $n > p$, $F_n = \emptyset$. Ainsi
$P\left(\bigcup_{n=0}^{p} F_n\right) = P\left(\bigcup_{n=0}^{\infty} F_n\right) = \sum_{n=0}^{\infty} P(F_n) = \sum_{n=0}^{p} P(F_n)$.
◇ Appliquons la propriété précédente avec $p = 2$, $F_1 = F$ et $F_2 = \overline{F}$.
Ainsi, $P(F) + P(\overline{F}) = P(F \cup \overline{F}) = P(\Omega) = 1$.
◇ En adaptant le raisonnement précédent, lorsque $G \subset H$, on montre que
$P(H) = P(G) + P(H \setminus G)$, donc $P(H) \geq P(G)$ et $P(H \setminus G) = P(H) - P(G)$.
◇ $G \cup H = G \cup (H \setminus (H \cap G))$, or G et $H \setminus (H \cap G)$ sont disjoints,
donc $P(G \cup H) = P(G) + (P(H) - P(H \cap G))$.
◇ Pour tout $n \geq 1$, posons $G_n = F_n \setminus \bigcup_{k=0}^{n-1} F_k$ et posons $G_0 = F_0$.
Ainsi $(G_n)_{n\in\mathbb{N}}$ est une suite d'événements de $\mathcal{F}$ deux à deux disjoints,
donc $P\left(\bigcup_{n=0}^{\infty} G_n\right) = \sum_{n=0}^{\infty} P(G_n)$, ce qui donne $P\left(\bigcup_{n=0}^{\infty} F_n\right) = \sum_{n=0}^{\infty} P(G_n) \leq \sum_{n=0}^{\infty} P(F_n)$,
car $G_n \subset F_n$. □

Propriété : Probabilité sur un univers dénombrable. Lorsque Ω est fini ou dénombrable, on prendra toujours $\mathcal{F} = \mathcal{P}(\Omega)$. Dans ce cas, pour se donner une probabilité P sur $(\Omega, \mathcal{F})$, il faut et il suffit de donner une famille sommable $(p_\omega)_{\omega\in\Omega}$ de réels positifs telle que $\sum_{\omega\in\Omega} p_\omega = 1$. On définit alors P par :

pour tout $F \in \mathcal{F}$, $P(F) = \sum_{\omega\in F} p_\omega$.

En particulier, pour tout $\omega \in \Omega$, $p_\omega = P(\{\omega\})$.

Dans ce cas, la famille $\Big(P(\{\omega\})\Big)_{\omega\in\Omega}$ est appelée le germe de la probabilité P.

Démonstration.

$\diamond$ Supposons d'abord que P est une probabilité sur $(\Omega, \mathcal{F})$. Posons pour tout $\omega \in \Omega$, $p_\omega = P(\{\omega\})$. Ainsi $(p_\omega)_{\omega\in\Omega}$ est une famille de réels positifs.

Soit $F \in \mathcal{P}(\Omega)$. F est fini ou dénombrable.

Si F est fini, on a $P(F) = P(\bigcup_{\omega\in F}\{\omega\}) = \sum_{\omega\in F} P(\{\omega\}) = \sum_{\omega\in F} p_\omega$.

Si F est dénombrable, il existe une bijection $n \longmapsto \omega_n$ de $\mathbb{N}$ dans F.

Alors $P(F) = P(\bigcup_{n\in\mathbb{N}}\{\omega_n\}) = \sum_{n\in\mathbb{N}} p_{\omega_n}$.

D'après le cours sur la sommabilité (théorème page 5), ceci démontre que la famille de réels positifs $(p_\omega)_{\omega\in F}$ est sommable et que $\sum_{\omega\in F} p_\omega = \sum_{n\in\mathbb{N}} p_{\omega_n} = P(F)$.

En particulier, avec $F = \Omega$, $(p_\omega)_{\omega\in\Omega}$ est sommable et $\sum_{\omega\in\Omega} p_\omega = 1$.

$\diamond$ Réciproquement, supposons que $(p_\omega)_{\omega\in\Omega}$ est une famille sommable de réels positifs telle que $\sum_{\omega\in\Omega} p_\omega = 1$.

Lorsque $F \in \mathcal{F}$ est une partie dénombrable, $\sum_{\omega\in F} p_\omega \leq \sum_{\omega\in\Omega} p_\omega = 1$, donc la famille $(p_\omega)_{\omega\in F}$ est sommable et $\sum_{\omega\in F} p_\omega \in [0,1]$.

On peut donc définir l'application P de $\mathcal{F}$ dans $[0,1]$ par :

pour tout $F \in \mathcal{F}$, $P(F) = \sum_{\omega\in F} p_\omega$.

En particulier, lorsque $F = \{\omega\}$, on a bien $P(F) = p_\omega$.

Il reste à démontrer que P est une probabilité.

Clairement $P(\Omega) = \sum_{\omega\in\Omega} p_\omega = 1$.

Soit $(F_n)_{n\in\mathbb{N}}$ une suite d'événements de $\mathcal{F}$ deux à deux disjoints. Posons $F = \bigcup_{n=0}^{\infty} F_n$.

Alors $(F_n)_{n\in\mathbb{N}}$ est une partition de F (même si certains F_n peuvent être vides), donc on peut appliquer le théorème de sommation par paquets pour des familles de réels positifs, ce qui permet d'écrire $\sum_{\omega\in F} p_\omega = \sum_{n\in\mathbb{N}} \sum_{\omega\in F_n} p_\omega$, donc on a bien $P(F) = \sum_{n\in\mathbb{N}} P(F_n)$.

□

Définition. Supposons que Ω est de cardinal fini. On dit que P est la probabilité uniforme lorsque tous les événements élémentaires sont équiprobables. Dans ce cas, avec les notations de la propriété précédente, pour tout $\omega \in \Omega$, $p_\omega = \dfrac{1}{\mathrm{Card}(\Omega)}$, et pour tout $F \in \mathcal{F}$, $P(F) = \dfrac{\mathrm{Card}(F)}{\mathrm{Card}(\Omega)}$.

Remarque. Dans le cas où le cardinal de Ω est infini, la probabilité uniforme n'est plus définie car la série de terme général constant $c > 0$ est toujours divergente.

Exemple. Supposons que dans une ville donnée il y a 7 accidents par semaine. Alors durant la quasi-totalité des semaines, certains jours verront plusieurs accidents.
En effet, on peut modéliser la situation par $\Omega = \{1, \dots, 7\}^7$, l'événement élémentaire $(j_1, \dots, j_7)$ représentant la liste des jours où surviennent les accidents, ce qui correspond à une situation de tirage avec remise dans une urne contenant les chiffres de 1 à 7.
Notons A l'événement "les 7 accidents surviennent sur des jours différents de la semaine (donc un accident par jour exactement)", ce qui correspond à un tirage sans remise.
On a $P(A) = \dfrac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} = \dfrac{7!}{7^7} = 0,00612 \pm 10^{-5}$.
Cela signifie qu'un tel événement n'aura lieu en moyenne qu'environ une fois tous les trois ans !

Exemple. Le paradoxe des anniversaires :
Supposons que 23 personnes se trouvent dans la même salle. Quelle est la probabilité qu'au moins deux d'entre elles aient leur anniversaire le même jour ?
On peut modéliser cette situation, en première approximation, par un tirage aléatoire avec remise dans l'ensemble $\{1, \dots, 365\}$, avec la mesure uniforme ; un modèle plus réaliste devrait prendre en compte les années bissextiles, ainsi que les variations saisonnières du taux de natalité (sous nos latitudes, le nombre de naissances est plus élevé en été qu'en hiver, par exemple), etc. Pour le modèle précédent, la probabilité qu'au moins deux des $k = 23$ personnes aient leur anniversaire le même jour est donnée par $1 - \dfrac{365 \times 364 \times \cdots \times (365 - k + 1)}{365^{23}} = 0,507 \pm 10^{-3}$: il y a plus d'une chance sur deux que ça ait lieu !
Cette probabilité est de 97% s'il y a 50 personnes, et de 99,99996% pour 100 personnes.

Exercice. Deux joueurs A et B lancent un dé non pipé à tour de rôle (en commençant par A). Le premier qui obtient 6 a gagné et le jeu s'arrête.

1°) Quel est l'univers Ω pour cette expérience ?

2°) Calculer les probabilités des événements "A gagne", "B gagne", "ni A ni B ne gagne".

Solution :

1°) Ω est l'ensemble des suites finies $(x_0, \dots, x_n)$ où $x_n = 6$ et pour tout $i < n$, $x_i \in \{1, \dots, 5\}$, auxquelles on ajoute l'événement Δ correspondant à l'événement "ni A ni B ne gagne".
Nous reprendrons cet exercice plus loin, avec une solution plus simple.

2°) $P(A \text{ et } B \text{ perd}) = P(\Delta) = 1 - \sum_{n=0}^{+\infty} P(\{1, \dots, 5\}^n \times \{6\}) = 1 - \sum_{n=0}^{+\infty} (\frac{5}{6})^n \frac{1}{6} = 0.$

$P(A \text{ gagne}) = \sum_{n=0}^{+\infty} P(\{1, \dots, 5\}^{2n} \times \{6\}) = \frac{1}{6} \frac{1}{1 - (\frac{5}{6})^2} = \frac{6}{11}.$

Ce calcul est correct mais manque de justification. Nous y reviendrons.

Exemple. Une probabilité définie sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$:

Soit a un réel strictement positif fixé. On pose : $\forall\ k \in \mathbb{N}, \quad p_k = \dfrac{e^{-a} a^k}{k!}$.

Comme les p_k sont bien positifs et que $\sum_{k=0}^{+\infty} \dfrac{e^{-a} a^k}{k!} = e^{-a} \sum_{k=0}^{+\infty} \dfrac{a^k}{k!} = e^{-a} e^a = 1$, la famille (p_k) définit bien une probabilité, en posant pour tout $E \subset \mathbb{N}$: $P(E) = \sum_{k \in E} p_k$: on l'appelle loi de Poisson de paramètre a.
Calculons par exemple, $P(2\mathbb{N})$:

$P(2\mathbb{N}) = \sum_{k \in 2\mathbb{N}} p_k = \sum_{l=0}^{+\infty} \dfrac{e^{-a} a^{2l}}{(2l)!} = e^{-a} \text{cha} = \dfrac{1 + e^{-2a}}{2}.$

Une conséquence de ce résultat est : si l'on tire un nombre entier au hasard suivant une loi de Poisson, la probabilité qu'il soit pair est strictement supérieure à $\frac{1}{2}$.

Remarque. L'infini est une notion fondamentale des mathématiques. En remplaçant un nombre très grand par l'infini, on obtient souvent un modèle mathématique plus pratique à l'usage, même si sa construction nécessite une théorie plus conséquente.
Ainsi, dans l'exemple précédent, il serait plus pratique de considérer deux joueurs qui lancent le dé indéfiniment. L'univers Ω devient simplement $\{1, \dots, 6\}^{\mathbb{N}}$.
Mais Ω est alors un univers non dénombrable. Dans ce cas, pour des raisons théoriquement délicates que nous ne développerons pas, il n'est pas possible de construire des probabilités intéressantes sur $(\Omega, \mathcal{P}(\Omega))$ lorsque Ω n'est pas dénombrable. On sait le faire en restreignant $\mathcal{P}(\Omega)$ à une tribu bien choisie de Ω. C'est la raison d'être de la définition sophistiquée d'une tribu : un détour théorique conséquent pour construire des outils mathématiques pratiques.

Propriété. de continuité : dans un espace probabilisé $(\Omega, \mathcal{F}, P)$,

si (F_n) est une suite croissante d'événements, $P\left(\bigcup_{n=0}^{\infty} F_n\right) = \lim_{n \to +\infty} P(F_n)$.

Si (F_n) est une suite décroissante d'événements, $P\left(\bigcap_{n=0}^{\infty} F_n\right) = \lim_{n \to +\infty} P(F_n)$.

Remarque. Lorsque (F_n) est croissante, pour tout $N \in \mathbb{N}$, $\bigcup_{n=0}^{N} F_n = F_N$, donc $\bigcup_{n=0}^{\infty} F_n$ représente informellement la limite de la suite (F_n).

Lorsque (F_n) est décroissante la limite informelle des F_n est maintenant $\bigcap_{n=0}^{\infty} F_n$.

Ainsi la propriété précédente peut informellement s'écrire :

pour une suite (F_n) monotone d'événements, $P\left(\lim_{n\to+\infty} F_n\right) = \lim_{n\to+\infty} P(F_n)$.

Cela fournit un moyen mnémotechnique simple et cela explique l'appellation de cette propriété.

Démonstration.

◇ Pour tout $n \geq 1$, posons $G_n = F_n \setminus F_{n-1}$ et posons $G_0 = F_0$. Alors $P\left(\bigcup_{n=0}^{\infty} F_n\right) = P\left(\bigcup_{n=0}^{\infty} G_n\right) = \sum_{n=0}^{\infty} P(G_n)$ car les G_n sont deux à deux disjoints, donc $P\left(\bigcup_{n=0}^{\infty} F_n\right) = P(F_0) + \sum_{n=1}^{\infty}(P(F_n) - P(F_{n-1}))$.

Ainsi la série télescopique $\sum(P(F_n)-P(F_{n-1}))$ converge, donc la suite $(P(F_n))$ converge également et $P\left(\bigcup_{n=0}^{\infty} F_n\right) = \lim_{n\to+\infty} P(F_n)$.

◇ Passons aux complémentaires : la suite $(\overline{F_n})$ étant croissante,

$P\left(\bigcap_{n=0}^{\infty} F_n\right) = 1 - P\left(\bigcup_{n=0}^{\infty} \overline{F_n}\right) = 1 - \lim_{n\to+\infty} P(\overline{F_n}) = 1 - \lim_{n\to+\infty}(1 - P(F_n))$. □

Un peu de vocabulaire :

On dit que l'événement F est négligeable si et seulement si $P(F) = 0$.

On dit que l'événement F est presque sûr si et seulement si $P(F) = 1$.

Si $\mathcal{Q}$ est une propriété dépendant de $\omega \in \Omega$, lorsque $\{\omega \in \Omega / \mathcal{Q}(\omega)\}$ est un événement presque sûr, on dit que "$\mathcal{Q}(\omega)$ presque sûrement".

Propriété. Une réunion finie ou dénombrable d'événements négligeables est négligeable.

Démonstration.

Dans un espace probabilisé $(\Omega, \mathcal{F}, P)$, considérons une suite (F_n) d'événements négligeables.

D'après l'inégalité de Boole, $0 \leq P(\bigcup_{n\in\mathbb{N}} F_n) \leq \sum_{n\in\mathbb{N}} P(F_n) = 0$. □

Remarque. Par passage au complémentaire, on en déduit qu'une intersection finie ou dénombrable d'événements presque sûrs est presque sûre.

> **Exercice. Borel-Cantelli, première partie :**
> Si $(G_n)_{n\in\mathbb{N}}$ est une suite d'événements, on note $\limsup_{n\to+\infty} G_n$ l'événement "il y a une infinité de n pour lesquels G_n se réalise".

1°) Montrer que $\limsup_{n\to+\infty} G_n = \bigcap_{n=0}^{\infty} \bigcup_{m=n}^{\infty} G_m$.

2°) Montrer que si $\sum P(G_n)$ converge alors $\limsup_{n\to+\infty} G_n$ est un événement négligeable.

Solution :
Attention au vocabulaire probabiliste : la phrase "G_n se réalise" signifie, pour $x \in \Omega$ donné (x est un événement élémentaire), "$x \in G_n$".

1°) Soit $x \in \Omega$: $\{n \in \mathbb{N}/x \in G_n\}$ est un ensemble d'entiers naturels, donc il est infini si et seulement si il n'est pas majoré, c'est-à-dire si et seulement si pour tout $n \in \mathbb{N}$, il existe $m \geq n$ tq $x \in G_m$. Ainsi,
$x \in \limsup_{n\to+\infty} G_n \iff \{n \in \mathbb{N}/x \in G_n\}$ est infini

$$\iff \forall n \in \mathbb{N},\ \exists m \geq n\ \ x \in G_m \iff x \in \bigcap_{n=0}^{\infty} \bigcup_{m=n}^{\infty} G_m.$$

2°) Posons $H_n = \bigcup_{m=n}^{\infty} G_m$. (H_n) est une suite décroissante, donc d'après la propriété de continuité, $P(\limsup_{n\to+\infty} G_n) = \lim_{n\to+\infty} P(H_n)$, mais $P(H_n) \leq \sum_{m\geq n} P(G_m)$, donc $P(\limsup_{n\to+\infty} G_n) \leq \lim_{n\to+\infty} \sum_{m\geq n} P(G_m) = 0$.

2 Probabilité conditionnelle et indépendance

De nombreuses affirmations prennent la forme "si G a lieu, alors la probabilité de H est p », où G et H sont des événements (tels "il pleuvra demain", et "le bus sera à l'heure", respectivement).
Afin de motiver la définition de la probabilité conditionnelle d'un événement H étant connue la réalisation d'un événement G, utilisons l'interprétation fréquentiste des probabilités (qui sera "formalisée" plus loin avec la loi faible des grands nombres).
On considère deux événements H et G. On désire déterminer la fréquence de réalisation de l'événement H lorsque l'événement G a lieu. Une façon de procéder est la suivante : on répète l'expérience un grand nombre de fois N. On note le nombre N_G de tentatives lors desquelles G est réalisé, et le nombre $N_{H\cap G}$ de ces dernières tentatives lors desquelles H est également réalisé. La fréquence de réalisation de H parmi les tentatives ayant donné lieu à G est alors donnée par $\dfrac{N_{H\cap G}}{N_G} = \dfrac{N_{H\cap G}}{N} \times \dfrac{N}{N_G}$.
Lorsque N devient grand, on s'attend à ce que le terme de gauche converge vers la probabilité de H conditionnellement à la réalisation de l'événement G, alors que le terme de droite devrait converger vers $\dfrac{P(H \cap G)}{P(G)}$. Ceci motive la définition suivante.

Définition. Si $P(G) > 0$, $\boxed{P(H|G) \triangleq \frac{P(H \cap G)}{P(G)}}$: c'est la probabilité conditionnelle de H sachant que G est réalisé.
L'application $H \longmapsto P(H|G)$ est une probabilité sur l'univers probabilisé de travail. Cette dernière probabilité est souvent notée P_G. Ainsi,
$P(H|G) = P_G(H) = \frac{P(H \cap G)}{P(G)}$.

Remarque. Lorsque $x \in G$, on a $[x \in H \iff x \in H \cap G]$, mais $H \longmapsto P(H \cap G)$ n'est pas une probabilité : on divise par $P(G)$ pour cela. En particulier, $P(G|G) = 1$.

Exemple. On jette deux dés non pipés. Sachant que le premier jet nous donne 3, quelle est la probabilité que la somme soit strictement supérieure à 6 ?
On veut calculer $P(A|B)$, où $B = \{(3,k)/k \in \{1,\dots,6\}\}$
et $A = \{(a,b) \in \{1,\dots,6\}^2 / a+b > 6\}$. Or $A \cap B = \{(3,4),(3,5),(3,6)\}$, donc $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\#(A \cap B)}{\#B} = \frac{3}{6} = \frac{1}{2}$

Exemple. On choisit une famille au hasard parmi toutes les familles ayant deux enfants et dont au moins un est un garçon. Quelle est la probabilité que les deux enfants soient des garçons ?
On veut calculer $P(A|B)$, où $B = \{(G,G),(F,G),(G,F)\}$ et $A = A \cap B = \{(G,G)\}$.
Ainsi $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3}$.
Formule des probabilités composées : $P(H \cap G) = P(G) \times P(H|G)$.
Plus généralement, pour $k \geq 2$, si $G_1, \dots, G_k$ sont k événements
tels que $P(G_1 \cap \dots \cap G_{k-1}) > 0$, alors

$$\boxed{P(\bigcap_{i=1}^{k} G_i) = P(G_1) \times P(G_2|G_1) \times P(G_3|G_1 \cap G_2) \times \dots \times P(G_k|G_1 \cap \dots \cap G_{k-1})}.$$

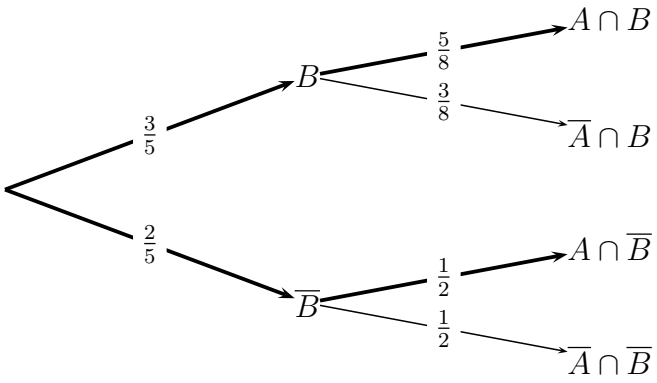
Démonstration.
Par récurrence sur k, car $P(\bigcap_{i=1}^{k} G_i) = P(\bigcap_{i=1}^{k-1} G_i) P(G_k | \bigcap_{i=1}^{k-1} G_i)$. □

Exemple. On tire 4 cartes d'un jeu de 52 cartes. Quelle est la probabilité p que l'on ait tiré les 4 as ?
Ici, on choisit pour univers Ω l'ensemble des quadruplets de cartes, muni de la probabilité uniforme. On note A_i l'événement "la i-ème carte tirée est un as". On cherche donc à calculer la probabilité de $A_1 \cap A_2 \cap A_3 \cap A_4$.
On a $P(A_1) = \frac{4}{52}$, $P(A_2 \mid A_1) = \frac{3}{51}$, $P(A_3 \mid A_2 \cap A_1) = \frac{2}{50}$, et $P(A_4 \mid A_1 \cap A_2 \cap A_3) = \frac{1}{49}$.
Le résultat cherché est donc $p = P(A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{4.3.2.1}{52.51.50.49} = 0,00000369$.
Ou bien, en prenant comme univers l'ensemble des mains (i.e ensembles) de 4 cartes, la probabilité cherchée est celle d'un événement élémentaire et l'on retrouve $p = \frac{1}{\binom{52}{4}}$.

Formule des probabilités totales :
si $(G_i)_{i\in I}$ est un système complet d'événements, où I est fini ou dénombrable, et si pour tout $i \in I$, $P(G_i) > 0$, alors $P(G) = \sum_{i\in I} P(G|G_i)P(G_i)$.

Démonstration.
$P(G) = P(G \cap \bigcup_{i\in I} G_i) = P(\bigcup_{i\in I}[G \cap G_i]) = \sum_{i\in I} P(G \cap G_i) = \sum_{i\in I} P(G|G_i)P(G_i)$. □

Exemple. On se donne deux urnes. La première contient deux balles rouges et trois balles bleues ; la seconde trois rouges et quatre bleues. Une balle est tirée au hasard de la première urne et elle est placée dans la seconde. On tire ensuite au hasard une balle de la seconde urne : quelle est la probabilité qu'elle soit bleue ?
Soit A l'événement « la balle tirée de la seconde urne est bleue », et B l'événement « la balle déplacée de la première urne à la seconde est bleue ». Puisque B et $\overline{B}$ forment une partition de Ω, une application de la formule des probabilités totales donne $P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$.
Or, $P(A|B) = P(A|$ la 2ème urne contient trois balles rouges et cinq bleues$) = \frac{5}{8}$ et $P(A|\overline{B}) = P(A|$ la 2ème urne contient quatre balles rouges et quatre bleues$) = \frac{1}{2}$.
Puisque $P(B) = \frac{3}{5}$ et $P(\overline{B}) = \frac{2}{5}$, on obtient $P(A) = \frac{23}{40}$.
On représente souvent une situation de ce type par l'arbre suivant :



Exemple. : **Le paradoxe de Monty Hall :**
Un bienfaiteur vous propose le jeu suivant. Il va vous présenter 3 enveloppes fermées ; 2 d'entre elles contiennent du papier journal, la dernière un chèque d'un million d'euros. Vous devrez choisir une enveloppe, sans l'ouvrir. Il ouvrira ensuite une des deux enveloppes restantes et vous montrera qu'elle contient du papier journal. Vous aurez alors le choix entre conserver l'enveloppe choisie initialement, ou bien changer pour celle qui reste. Quelle est la meilleure stratégie ?
Notons G_1 l'événement "l'enveloppe choisie contient le chèque" et G_2 l'événement "l'enveloppe qui reste contient le chèque". L'habitude des situations équiprobables pourrait mener à l'intuition que $P(G_1) = P(G_2) = \frac{1}{2}$. Mais il est pourtant clair que, avant l'intervention du bienfaiteur, $P(G_1) = \frac{1}{3}$, donc d'après la formule des probabilités totales,

après l'intervention du bienfaiteur :
$P(G_2) = P(G_2|G_1)P(G_1) + P(G_2|\overline{G}_1)P(\overline{G_1}) = 0 \times \frac{1}{3} + 1\frac{2}{3} = \frac{2}{3}$: insistons sur le fait que dans cette égalité, les probabilités conditionnelles $P(G'|G'')$ représentent la probabilité que G' soit vrai après l'action du bienfaiteur sachant que G'' est vrai avant son intervention.
Ainsi après l'intervention du bienfaiteur, on a toujours $P(G_1) = \frac{1}{3}$.
En conclusion, vous avez deux fois plus de chances de gagner si vous changez d'enveloppe.

Formule de Bayes :

Si $P(G) \in]0,1[$ et si $P(H) > 0$, alors $\boxed{P(G|H) = \dfrac{P(H|G)P(G)}{P(H|G)P(G) + P(H|\overline{G})P(\overline{G})}}$.

Plus généralement, si $(G_i)_{i\in I}$ est un système complet d'événements, si pour tout $i \in I$, $P(G_i) > 0$, et si $P(H) > 0$, alors $\boxed{P(G_i|H) = \dfrac{P(H|G_i)P(G_i)}{\sum_{j\in I} P(H|G_j)P(G_j)}}$.

Démonstration.
$P(G_i|H) = \dfrac{P(H \cap G_i)}{P(H)} = \dfrac{P(H|G_i)P(G_i)}{P(H)}$ et on conclut en utilisant la formule des probabilités totales. □

Exercice. Considérons un test de dépistage d'une certaine maladie qui, selon les indications du fabricant, lorsqu'il est pratiqué sur une personne effectivement malade, indique "positif" dans 95% des cas, et lorsqu'il est pratiqué sur une personne non affectée de cette maladie, indique "positif" dans seulement 1% des cas.
Dans l'hypothèse où seulement $0,1\%$ de la population est atteinte de cette maladie, calculer la probabilité qu'une personne ne soit pas malade alors que le résultat du test est positif.

Solution : Notons M l'événement "être malade" et P l'événement "le résultat du test est positif".
On nous demande de calculer $Pr(\overline{M}|P)$, sachant d'après l'énoncé que
$Pr(P|M) = 0,95$, $Pr(P|\overline{M}) = 0,01$ et $Pr(M) = 0,001$.
D'après la formule de Bayes,
$$Pr(\overline{M}|P) = \frac{Pr(P|\overline{M})Pr(\overline{M})}{Pr(P|M)Pr(M) + Pr(P|\overline{M})Pr(\overline{M})} = \frac{0,01 \times 0,999}{0,95 \times 0,001 + 0,01 \times 0,999},$$
donc $Pr(\overline{M}|P) = \dfrac{0,00999}{0,01094} = 91,316\%$.
On peut s'interroger dans ces conditions sur l'intérêt du test.
Un calcul similaire donne
$$Pr(M|\overline{P}) = \frac{Pr(\overline{P}|M)Pr(M)}{Pr(\overline{P}|M)Pr(M) + Pr(\overline{P}|\overline{M})Pr(\overline{M})} = \frac{0,05 \times 0,001}{0,05 \times 0,001 + 0,99 \times 0,999},$$
donc $Pr(M|\overline{P}) = \dfrac{5 \times 10^{-5}}{0,98906} = (4.94 \times 10^{-5} \pm 10^{-7})\%$.

Ici, quel est l'univers Ω ?

Définition. $\boxed{H \text{ et } G \text{ sont indépendants si et seulement si } P(G \cap H) = P(G)P(H)}$.
Lorsque $P(H) > 0$, H et G sont indépendants si et seulement si $P(G) = P(G|H)$ (i.e : la connaissance de la réalisation de H n'apporte aucune information à propos de celle de G).

Propriété. Si H et G sont indépendants, alors H et $\overline{G}$ sont aussi indépendants.

Démonstration.
$P(\overline{G} \cap H) = P(H \setminus (H \cap G)) = P(H) - P(H \cap G) = P(H) - P(H)P(G)$,
donc $P(\overline{G} \cap H) = P(H)(1 - P(G)) = P(H)P(\overline{G})$. □

Remarque.
Un événement A est indépendant de lui-même si et seulement si $P(A) \in \{0, 1\}$.

Démonstration.
$P(A \cap A) = P(A) \times P(A) \iff P(A) = P(A)^2 \iff P(A) \in \{0, 1\}$. □

Définition. I étant un ensemble quelconque, les événements de la famille $(G_i)_{i\in I}$ sont mutuellement indépendants si et seulement si pour toute partie J de I, finie et non vide, $P\left(\bigcap_{i\in J} G_i\right) = \prod_{i\in J} P(G_i)$.

Remarque. Soit $(G_i)_{i\in I}$ une famille d'événements. Ils sont mututellement indépendants si et seulement si toute sous-famille finie $(G_i)_{i\in J}$ (avec $J \subset I$ et J finie et non vide) est constituée d'événements mutuellement indépendants.

Propriété. Soit $(G_i)_{i\in I}$ une famille d'événements mutuellement indépendants. Si l'on remplace dans cette famille certains G_i par leur conjugué $\overline{G_i}$, alors la famille est encore une famille d'événements mutuellement indépendants.

Démonstration.
⋄ Soit $n \in \mathbb{N}$ avec $n \geq 2$ et soit $b = (G_i)_{1\leq i\leq n}$ une famille de n événements mutuellement indépendants. Posons $b' = (\overline{G_1}, G_2, \ldots, G_n)$ et montrons que ces événements sont mutuellement indépendants.
Renommons $b' = (G'_1, G'_2, \ldots, G'_n)$.
Soit $J \subset \{1, \ldots, n\}$ avec J non vide.
Premier cas : Supposons que $1 \notin J$.
Alors pour tout $i \in J$, $G_i = G'_i$, donc $P\left(\bigcap_{i\in J} G'_i\right) = \prod_{i\in J} P(G'_i)$.
Second cas : Supposons que $1 \in J$. Alors $P\left(\bigcap_{i\in J} G'_i\right) = P\left(\overline{G_1} \cap \bigcap_{i\in J\setminus\{1\}} G_i\right)$, or G_1 et $\bigcap_{i\in J\setminus\{1\}} G_i$ sont indépendants, donc $\overline{G_1}$ et $\bigcap_{i\in J\setminus\{1\}} G_i$ sont également indépendants.
On en déduit que $P\left(\bigcap_{i\in J} G'_i\right) = \prod_{i\in J} P(G'_i)$.

◇ Par récurrence sur $p \in \{2, \dots, n\}$, on en déduit que $\overline{G_1}, \dots, \overline{G_p}, G_{p+1}, \dots, G_n$ sont mutuellement indépendants.
◇ Notons maintenant $f = (G_i)_{i\in I}$ et f' la famille déduite de f en remplaçant certains G_i par leur conjugué $\overline{G_i}$. Alors le point précédent prouve que toute sous-famille finie de f' est constituée d'événements mutuellement indépendants, donc c'est aussi le cas pour la famille f'. □

Remarque. "mutuellement indépendants" $\Longrightarrow$ "2 à 2 indépendants", mais la réciproque est fausse.
En effet, lançons trois fois une pièce. L'univers Ω est égal à $\{F, P\}^3$ et chaque événement élémentaire a pour probabilité $\frac{1}{8}$.
Notons $A_{i,j}$ l'événement "le ième lancer donne le même résultat que le jème".
$P(A_{1,2}) = \frac{1}{2} = P(A_{1,3}) = P(A_{2,3})$ et $P(A_{1,2} \cap A_{1,3}) = \frac{1}{4}$, donc $A_{1,2}$ et $A_{1,3}$ sont indépendants. Par symétrie, les événements $A_{1,2}$, $A_{1,3}$ et $A_{2,3}$ sont deux à deux indépendants.
Cependant, $P(A_{1,2} \cap A_{1,3} \cap A_{2,3}) = P(A_{1,2} \cap A_{1,3}) = \frac{1}{4} \neq P(A_{1,2})P(A_{1,3})P(A_{2,3})$, donc les 3 événements ne sont pas mutuellement indépendants. C'est informellement évident car les deux premiers événements impliquent le troisième.

Exercice. Soit $n \in \mathbb{N}^*$. On munit $\mathbb{N}_n$ de la probabilité uniforme.
Pour tout $k \in \mathbb{N}_n$ tel que $k \mid n$, on pose $A_k = \{m \in \mathbb{N}_n \ / \ k \mid m\}$.
Lorsque $k_1, \dots, k_r \in \mathbb{N}^*$ sont r diviseurs de n deux à deux premiers entre eux, montrer que les événements $A_{k_1}, \dots, A_{k_r}$ sont mutuellement indépendants.
En déduire que $\varphi(n) = n \prod_{\substack{p\in\mathbb{P}\\ p \mid n}} (1 - \frac{1}{p})$, où φ désigne l'indicatrice d'Euler.

Solution :
◇ Soit $k_1, \dots, k_r$ r diviseurs de n deux à deux premiers entre eux.
Soit $m \in \mathbb{N}_n$. D'après cette dernière hypothèse, m est un multiple de $k_1, \dots, k_r$ si et seulement si m est un multiple de $\prod_{i=1}^{r} k_i$, donc $A_{k_1} \cap \dots \cap A_{k_r} = A_{k_1 \times \dots \times k_r}$.
D'autre part, si d est un diviseur de n, il existe $e \in \mathbb{N}^*$ tel que $n = de$. Alors $A_d = \{d, 2d, \dots, ed\}$, donc $|A_d| = e = \frac{n}{d}$, puis $P(A_d) = \frac{1}{d}$.
On en déduit que $P(A_{k_1} \cap \dots \cap A_{k_r}) = \dfrac{1}{k_1 \times \dots \times k_r} = \prod_{i=1}^{r} P(A_{k_i})$.
La famille $(k_1, \dots, k_r)$ étant quelconque parmi les familles de diviseurs de n deux à deux premiers entre eux, cette dernière égalité est encore valable pour toute sous-famille de $(k_1, \dots, k_r)$. Ceci prouve que les événements $A_{k_1}, \dots, A_{k_r}$ sont mutuellement indépendants.
◇ On sait que $\varphi(n)$ est le cardinal de $B = \{h \in \mathbb{N}_n \ / \ h \wedge n = 1\}$,
or pour tout $h \in \mathbb{N}_n$,
$h \in B \iff [\forall p \in \mathbb{P},\ (p \mid n \Longrightarrow h \wedge p = 1)]$
$\iff [\forall p \in \mathbb{P},\ (p \mid n \Longrightarrow h \notin A_p),]$

donc $B = \bigcap_{\substack{p\in\mathbb{P}\\ p\ |\ n}} \overline{A_p}$. D'après le cours, il s'agit d'une intersection d'événements mutuellement indépendants, donc $P(B) = \prod_{\substack{p\in\mathbb{P}\\ p\ |\ n}} P(\overline{A_p}) = \prod_{\substack{p\in\mathbb{P}\\ p\ |\ n}} (1-\frac{1}{p})$.

Or $P(B) = \frac{|B|}{n}$, donc $\varphi(n) = n \prod_{\substack{p\in\mathbb{P}\\ p\ |\ n}} (1-\frac{1}{p})$.

Exercice. Lemme de Borel-Cantelli, deuxième partie :

1°) Soit $(G_n)_{n\in\mathbb{N}}$ une suite d'événements mutuellement indépendants.
On suppose que $\sum_{n=0}^{+\infty} P(G_n) = +\infty$.
Montrer que $P(G_n$ est réalisé pour une infinité d'entiers n$) = 1$.

2°) Une expérience consiste à lancer de manière répétée un dé équilibré. Montrer que si on prolonge indéfiniment l'expérience, alors il est certain que la valeur 6 sortira une infinité de fois.

3°) On dispose d'une suite infinie d'urnes. L'urne numéro n contient n boules numérotées de 1 à n. Une seule boule est tirée de chacune des urnes (les tirages sont indépendants). Quelle est la probabilité que la suite des numéros tirés comporte une infinité de 1 ?

Solution :

1°) Notons H l'événement "G_n est réalisé pour une infinité d'entiers n". On a déjà vu que $H = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} G_k$. Par continuité décroissante des probabilités,
$P(H) = \lim_{n\to+\infty} P(\bigcup_{k\geq n} G_k) = \lim_{n\to+\infty} [1 - P(\bigcap_{k\geq n} \overline{G_k})]$.
Toujours par continuité décroissante des probabilités,
$P(\bigcap_{k\geq n} \overline{G_k}) = \lim_{N\to+\infty} P(\bigcap_{k=n}^{N} \overline{G_k})$, puis d'après l'indépendance des $\overline{G_k}$,
$P(\bigcap_{k\geq n} \overline{G_k}) = \lim_{N\to+\infty} \prod_{k=n}^{N} P(\overline{G_k})$, donc
$P(H) = 1 - \lim_{n\to+\infty} \lim_{N\to+\infty} \prod_{k=n}^{N} [1 - P(G_k)]$.
Premier cas : Supposons que $\{n \in \mathbb{N}/P(G_n) = 1\}$ est infini. Alors, pour tout $n \in \mathbb{N}$, il existe $k \geq n$ tel que $P(G_k) = 1$, donc $\lim_{N\to+\infty} \prod_{k=n}^{N} [1 - P(G_k)] = 0$ et $P(H) = 1$.

Deuxième cas : Sinon, il existe $n_0 \in \mathbb{N}$ tel que $\forall n \geq n_0,\ \ P(G_n) < 1$. Ainsi, pour $n \geq n_0$, $\prod_{k=n}^{N}[1-P(G_k)] = \exp(\sum_{k=n}^{N}\ln[1-P(G_k)])$.
Si la série $\sum_{n\geq n_0}\ln(1-P(G_n))$ converge, alors $\ln(1-P(G_n)) \underset{n\to+\infty}{\longrightarrow} 0$, donc $P(G_n) = 1 - e^{\ln(1-P(G_n))} \underset{n\to+\infty}{\longrightarrow} 0$, puis $\ln(1-P(G_n)) \sim -P(G_n)$, donc $\sum P(G_n)$ converge, ce qui est faux par hypothèse. Ainsi $\sum \ln(1-P(G_n))$ diverge, mais c'est une série à termes négatifs, donc $\sum_{k=n}^{N}\ln[1-P(G_k)] \underset{N\to+\infty}{\longrightarrow} -\infty$,
puis $\prod_{k=n}^{N}[1-P(G_k)] \underset{N\to+\infty}{\longrightarrow} 0$, ce qui montre que $P(H) = 1$.

2°) On note G_n l'événement "au lancer n, la valeur du dé est 6". Les G_n sont indépendants et $P(G_n) = \dfrac{1}{6}$. On a bien $\sum P(G_n)$ diverge, donc d'après la première question, presque sûrement, la valeur 6 sortira une infinité de fois.

3°) On note G_n l'événement "la boule tirée dans l'urne n a pour numéro 1". Les G_n sont indépendants et $P(G_n) = \dfrac{1}{n}$. On sait que $\sum P(G_n)$ diverge, donc d'après la première question, presque sûrement, la valeur 1 est présente une infinité de fois dans la suite des numéros tirés.

Exercice. Urne de Polya :
Une urne contient r boules rouges et b boules blanches. On tire une boule et on note sa couleur. On remet alors la boule tirée dans l'urne et on ajoute a boules de la même couleur. La même procédure est répétée $n-1$ fois et l'on fait ainsi n tirages.
Soient R_j l'événement "la j-ème boule tirée est rouge" et B_j l'événement "la j-ème boule tirée est blanche".

1°) Calculer $P(R_1)$, $P(B_1)$, $P(R_2)$ et $P(B_2)$.

2°) En procédant par récurrence, calculer $P(R_j)$ et $P(B_j)$ pour tout $j \in \{2,\dots,n\}$.

3°) Interpréter le cas où $a = -1$.

Solution :

1°) $P(R_1) = \dfrac{r}{r+b}$ et $P(B_1) = \dfrac{b}{r+b}$. D'après la formule des probabilités totales,
$$\begin{aligned} P(R_2) &= P(R_2|R_1)P(R_1) + P(R_2|B_1)P(B_1) \\ &= \frac{r+a}{r+a+b}\times\frac{r}{r+b} + \frac{r}{r+a+b}\times\frac{b}{r+b}, \end{aligned}$$
donc $P(R_2) = \dfrac{r}{r+b}$ et de même on calcule que $P(B_2) = P(B_1) = \dfrac{b}{r+b}$.

2°) Notons $S(j)$ la propriété suivante :

pour tous r, a, b tels que $r+b \neq 0$, $P(R_j) = \dfrac{r}{r+b}$ et $P(B_j) = \dfrac{b}{r+b}$.
On a $S(1)$ et $S(2)$ d'après la première question.
Soit $j \geq 2$. On suppose $S(j-1)$.
$P(R_j) = P(R_j|R_1)P(R_1) + P(R_j|B_1)P(B_1)$.
Si R_1 se produit, après le premier tirage l'urne contient $r+a$ boules rouges et b boules blanches, si bien que le jème tirage correspond au $(j-1)$ième tirage d'une urne possédant au départ $r+a$ boules rouges et b boules blanches. Ainsi d'après $S(j-1)$, $P(R_j|R_1) = \dfrac{r+a}{r+a+b}$.
De même $P(R_j|B_1) = \dfrac{r}{r+a+b}$.
Ainsi $P(R_j) = \dfrac{r+a}{r+a+b} \times \dfrac{r}{r+b} + \dfrac{r}{r+a+b} \times \dfrac{b}{r+b} = \dfrac{r}{r+b}$.
On procède de même pour $P(B_j)$ ce qui prouve $S(j)$.

3°) Les calculs restent valables pour $a = -1$ car il est toujours possible de retirer une boule rouge si l'on tire une boule rouge, mais cela ne fonctionne plus avec $a = -2$. Le cas où $a = -1$ correspond au cas du tirage sans remise. On a donc prouvé que, dans une urne contenant au départ r boules rouges et b boules blanches, la probabilité de tirer une boule rouge au $j$ème tirage vaut $\dfrac{r}{r+b}$, que le tirage se fasse avec remise (le résultat est alors évident) ou qu'il se fasse sans remise.

Remarque. En toute rigueur, on devrait avant tout déterminer quel est l'univers probabilisé $(\Omega, \mathcal{F}, P)$ sous-jacent, or ce dernier n'est pas précisé dans la solution ci-dessus.
En convenant que R signifie "Rouge" et B signifie "Bleu", on peut considérer que $\Omega = \{R, B\}^n$, $\mathcal{F} = \mathcal{P}(\Omega)$, mais pour P, c'est plus compliqué :
Si $\omega = (x_1, \ldots, x_n) \in \Omega$, $P(\{\omega\}) = p_\omega = \displaystyle\prod_{i=1}^{n} q_i$, où pour tout $i \in \{1, \ldots, n\}$, en notant k_i le nombre de $j \in \{1, \ldots, i-1\}$ tel que $x_j = R$, $q_i = \begin{cases} \frac{r+k_i a}{r+b+(i-1)a} \text{ si } x_i = R \\ \frac{b+(i-1-k_i)a}{r+b+(i-1)a} \text{ si } x_i = B \end{cases}$
Lorsque $a = -1$, on voit ainsi qu'il faut imposer la condition $n \leq r+b$. En effet, lorsque $n > r+b$, la formule donnant p_ω impose une division par 0.
À part pour cette remarque, cette approche est lourde et sans intérêt. Il est préférable de considérer que Ω est l'ensemble de toutes les réalisations ω possibles du jeu, ce qui peut rester très vague.
On note alors X_j l'application de Ω dans $\{R, B\}$ telle que $X_j(\omega)$ représente la couleur de la j-ième boule tirée.
Ainsi, l'événement R_j s'écrit $R_j = \{\omega \in \Omega / X_j(\omega) = R\} = (X_j = R)$.
Un tel formalisme permet de s'affranchir d'un choix précis et laborieux de l'univers Ω. Il s'agit de la notion de variables aléatoires.

3 Variables aléatoires discrètes

Définition. Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Une variable aléatoire à valeurs dans un ensemble E muni d'une tribu $\mathcal{E}$ est une fonction $X : \Omega \longrightarrow E$ telle que, pour tout $A \in \mathcal{E}$, $X^{-1}(A) \in \mathcal{F}$.

Remarque. Lorsque $E = \mathbb{R}$, on dit que X est une variable aléatoire réelle.
Lorsque $E = \mathbb{N}$, on dit que X est une variable aléatoire entière.

Notation. En probabilité, avec les notations précédentes,
l'événement $X^{-1}(A) = \{\omega \in \Omega / X(\omega) \in A\}$ est souvent noté $(X \in A)$ ou $\{X \in A\}$,
si bien que $P(X \in A) = P(X^{-1}(A)) = P(\{\omega \in \Omega / X(\omega) \in A\})$.
En particulier, si X est une variable aléatoire réelle, pour tout $x \in \mathbb{R}$, on note
$P(X < x) = P(\{\omega \in \Omega / X(\omega) < x\})$. On définit de même $P(X = x)$, $P(X \geq x)$ ...

Propriété. Toujours avec les notations précédentes, si l'on pose, pour tout $A \in \mathcal{E}$, $P_X(A) = P(X \in A)$, **alors P_X est une probabilité sur $(E, \mathcal{E})$ que l'on appelle la loi de X.**
En pratique, $(\Omega, \mathcal{F}, P)$ est rarement connu (on sait seulement qu'il existe), mais $(E, \mathcal{E}, P_X)$ est souvent donné en même temps que la fonction X.

Remarque. Supposons que $X(\Omega) \in \mathcal{E}$. Alors, pour tout $A \in \mathcal{E}$,
$P_X(A) = P(X^{-1}(A)) = P(X^{-1}(A \cap X(\Omega))) = P_X(A \cap X(\Omega))$: on dit que P_X est *portée* par $X(\Omega)$.

Définition. Si B est un événement de Ω, la loi de X conditionnée par B désigne l'application $A \longmapsto P(X \in A|B) = \dfrac{P((X \in A) \cap B)}{P(B)}$, de $\mathcal{E}$ dans $[0, 1]$. C'est encore une probabilité sur $(E, \mathcal{E})$.

Remarque. Le programme de MP ne prévoit que l'étude des variables aléatoires discrètes, ce que nous supposerons donc dorénavant.

Définition. On dit qu'une variable aléatoire X est discrète si et seulement si $X(\Omega)$ est fini ou dénombrable et si $\mathcal{E} = \mathcal{P}(E)$.

Propriété. Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et X une application de Ω dans un ensemble quelconque E. X est une variable aléatoire discrète si et seulement si $X(\Omega)$ est fini ou dénombrable et si, pour tout $d \in X(\Omega)$, $X^{-1}(\{d\}) \in \mathcal{F}$.
Dans ce cas, la loi de X est entièrement déterminée par la famille $(P(X = d))_{d \in X(\Omega)}$.

Démonstration.
cf la propriété des probabilités sur un univers dénombrable vue page 4. En particulier,
$P_X(E) = P_X(X(\Omega) \cap E) = P(\Omega) = 1 = P(\bigsqcup_{d \in X(\Omega)} X^{-1}(d)) = \sum_{d \in X(\Omega)} P(X = d)$. □

Remarque. Toute variable aléatoire entière est discrète.

Définition. Soit X une variable aléatoire discrète de Ω dans E et f une application de E dans un ensemble F. Alors $Y = f(X) \overset{\Delta}{=} f \circ X$ est une nouvelle variable aléatoire discrète dont la loi est donnée par : $\forall y \in F,\ P_Y(y) = P(X \in f^{-1}(\{y\})) = \sum_{\substack{x \in X(\Omega) \\ f(x)=y}} P_X(x)$.

Démonstration.
◇ $Y(\Omega) = \bigcup_{x \in X(\Omega)} \{f(x)\}$ est fini ou dénombrable.
◇ Soit $y \in Y(\Omega)$. Pour tout $\omega \in \Omega$,
$\omega \in Y^{-1}(\{y\}) \iff f(X(\omega)) = y \iff X(\omega) \in f^{-1}(\{y\})$,
donc $Y^{-1}(\{y\}) = X^{-1}(\bigcup_{\substack{x \in X(\Omega) \\ x \in f^{-1}(\{y\})}} \{x\}) = \bigcup_{\substack{x \in X(\Omega) \\ x \in f^{-1}(\{y\})}} X^{-1}(\{x\})$. Ainsi $Y^{-1}(\{y\}) \in \mathcal{F}$, ce qui prouve que Y est une variable aléatoire discrète
et $P_Y(y) = P(Y^{-1}(\{y\})) = \sum_{\substack{x \in X(\Omega) \\ x \in f^{-1}(\{y\})}} P(X^{-1}(\{x\})) = \sum_{\substack{x \in X(\Omega) \\ f(x)=y}} P_X(x)$. □

Définition. Soit X une variable aléatoire de Ω dans un ensemble E de cardinal fini. On dit que X suit une loi uniforme (souvent notée $\mathcal{U}$) si et seulement si P_X est la probabilité uniforme, c'est-à-dire si et seulement si pour tout $k \in E$, $P(X = k) = \dfrac{1}{\mathrm{Card}(E)}$.

Définition. On fixe une variable aléatoire X à valeurs dans $\mathbb{N}$.
Les lois classiques au programme sont les suivantes :

- Loi de dirac, lorsqu'il existe $n_0 \in \mathbb{N}$ tel que $P(X = n_0) = 1$ et $P(X = n) = 0$ pour tout $n \neq n_0$. On dit alors que X est une variable aléatoire déterministe, ou bien constante.
- **Loi de Bernoulli** de paramètre $p \in [0, 1]$, notée $\mathcal{B}(p)$:
  $\boxed{P(X = 1) = p \text{ et } P(X = 0) = 1 - p}$.
  C'est le cas lorsque X représente le succès $(X = 1)$ ou l'échec $(X = 0)$ d'une épreuve.
- **Loi binomiale** de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$, notée $\mathcal{B}(n, p)$:
  Pour tout $k \in \{0, \dots, n\}$, $\boxed{P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}}$ (et $P(X = m) = 0$ pour $m \notin \{0, \dots, n\}$).
  Nous verrons que c'est par exemple le cas lorsque X désigne le nombre de succès parmi une suite de n épreuves indépendantes de loi de Bernoulli de paramètre p.
  On remarquera que $\mathcal{B}(p) = \mathcal{B}(1, p)$.
- **Loi géométrique** de paramètre $p \in]0, 1[$, notée $\mathcal{G}(p)$:
  Pour tout $n \in \mathbb{N}^*$, $\boxed{P(X = n) = (1-p)^{n-1} p}$ (et $P(X = 0) = 0$).
  Nous verrons que c'est par exemple le cas lorsque X représente l'instant du premier succès lors d'une suite d'épreuves indépendantes de loi de Bernoulli de paramètre p.
- **Loi de Poisson** de paramètre $\lambda \in \mathbb{R}_+^*$, notée $\mathcal{P}(\lambda)$:

pour tout $n \in \mathbb{N}$, $\boxed{P(X=n) = e^{-\lambda}\dfrac{\lambda^n}{n!}}$.

— **Hors programme : Loi hypergéométrique** de paramètres $k, m, n \in \mathbb{N}^*$: C'est par exemple le cas lorsque X désigne le nombre de boules rouges obtenue à la suite de k tirages sans remise dans une urne contenant au départ m boules rouges et n boules noires. Ainsi en convenant que $\binom{b}{a}$ est nul dès que $a \notin \{0, \dots, b\}$, pour tout $r \in \mathbb{N}$, $\boxed{P(X=r) = \dfrac{\binom{m}{r}\binom{n}{k-r}}{\binom{m+n}{k}}}$.

Remarque. Ces lois classiques sont présentées en utilisant la notion de variable aléatoire, pour expliquer ce qu'elles modélisent le plus souvent, mais chacune de ces lois n'est rien d'autre qu'une probabilité sur $\mathbb{N}$, donnée par $(p_n)_{n\in\mathbb{N}} \in \mathbb{R}_+^{\mathbb{N}}$ telle que $\sum_{n=0}^{+\infty} p_n = 1$.

Notation. On utilisera la notation $X \sim \mathcal{L}$ pour indiquer que la variable aléatoire X suit la loi $\mathcal{L}$ et la notation $X \sim Y$ pour indiquer que les deux variables aléatoires suivent la même loi.

Propriété. X est une variable aléatoire à valeurs dans $\{0,1\}$ si et seulement si il existe un événement A tel que X est la fonction indicatrice de A, notée 1_A. Dans ce cas, on a $X = 1_A \sim \mathcal{B}(p)$ où $p = P(A)$.

Démonstration.
Si X est une variable aléatoire à valeurs dans $\{0,1\}$, c'est la fonction indicatrice de $A = X^{-1}(\{1\})$. □

Hors programme : Convergence en loi :
Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et soit $(X_k)_{k\in\mathbb{N}}$ une suite de variables aléatoires discrètes de Ω dans $\mathbb{N}$. Soit X une autre variable aléatoire de Ω dans $\mathbb{N}$.
On dit que X_k converge en loi vers X lorsque k tend vers $+\infty$ si et seulement si pour tout $n \in \mathbb{N}$, $P(X_k = n) \underset{k\to+\infty}{\longrightarrow} P(X = n)$: on notera

$$\boxed{X_k \xrightarrow[k\to+\infty]{\mathcal{L}} X \iff [\forall n \in \mathbb{N},\ \ P(X_k = n) \underset{k\to+\infty}{\longrightarrow} P(X = n)]}.$$

Définition. Soit X une variable aléatoire réelle (pas nécessairement discrète). La fonction de répartition de X est l'application $x \longmapsto P(X \leq x)$, de $\mathbb{R}$ dans $\mathbb{R}$.

Définition. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles et X une autre variable aléatoire réelle. On dit que X_n converge en loi vers X et on note $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ si et seulement si , pour tout $x \in \mathbb{R}$ tel que $P(X = x) = 0$, $P(X_n \leq x) \underset{n\to+\infty}{\longrightarrow} P(X \leq x)$.
Ainsi, $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ si et seulement si la suite des fonctions de répartition de X_n converge simplement vers la fonction de répartition de X sur $\{x \in \mathbb{R} \ / \ P(X = x) = 0\}$.

Propriété. Cette définition est bien une généralisation de celle donnée lorsque X_n et X sont à valeurs dans $\mathbb{N}$.

Démonstration.
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires entières et X une autre variable aléatoire entière. Notons (1) : $\forall n \in \mathbb{N},\ P(X_k = n) \underset{k\to+\infty}{\longrightarrow} P(X = n)$
et (2) : $\forall x \in \mathbb{R},\ P(X = x) = 0 \Longrightarrow P(X_k \leq x) \underset{k\to+\infty}{\longrightarrow} P(X \leq x)$.
Il s'agit de montrer que (1) $\Longleftrightarrow$ (2).
$\diamond$ Supposons (2). Soit $n \in \mathbb{N} \cup \{-1\}$. Posons $x = n + \frac{1}{2}$: $P(X = x) = 0$, car X est à valeurs dans $\mathbb{N}$, donc d'après (2), $P(X_k \leq x) \underset{k\to+\infty}{\longrightarrow} P(X \leq x)$.
Or $P(X \leq x) = P(X \leq n + \frac{1}{2}) = P(X \leq n)$. Ainsi, $P(X_k \leq n) \underset{k\to+\infty}{\longrightarrow} P(X \leq n)$.
Soit $n \in \mathbb{N}$: $P(X = n) = P((X \leq n) \setminus (X \leq n-1)) = P(X \leq n) - P(X \leq n-1)$ et de même, pour tout $k \in \mathbb{N}$, $P(X_k = n) = P(X_k \leq n) - P(X_k \leq n-1)$. On en déduit que $P(X_k = n) \underset{k\to+\infty}{\longrightarrow} P(X = n)$.
$\diamond$ Réciproquement, supposons (1). Soit $x \in \mathbb{R}$ tel que $P(X = x) = 0$.
$P(X \leq x) = P(X \leq \lfloor x \rfloor) = \sum_{n=0}^{\lfloor x \rfloor} P(X = n)$.
De même, pour tout $k \in \mathbb{N}$, $P(X_k \leq x) = \sum_{n=0}^{\lfloor x \rfloor} P(X_k = n)$,
donc d'après (1), $P(X_k \leq x) \underset{k\to+\infty}{\longrightarrow} P(X \leq x)$. $\square$

Remarque. Une situation assez usuelle est le cas d'une suite de variables aléatoires discrètes qui convergent en loi vers une variable aléatoire réelle non discrète : cf le théorème de la limite centrée, où une certaine suite de variables aléatoires discrètes (Z_n) converge en loi vers une variable aléatoire réelle suivant une loi normale (gaussienne), c'est-à-dire : $\forall x \in \mathbb{R},\ P(Z_n \leq x) \underset{n\to+\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du$.

Revenons dans le cadre du programme.

Propriété. Pour les variables aléatoires entières, les lois géométriques sont les seules lois sans mémoire. Plus précisément, si X est une variable aléatoire à valeurs dans $\mathbb{N}^*$, elle est sans mémoire, c'est-à-dire qu'elle vérifie pour tout $(n,k) \in \mathbb{N}^2$
$P(X > n+k | X > n) = P(X > k)$, si et seulement si il existe $p \in]0,1[$ tel que $X \sim \mathcal{G}(p)$.

Remarque. Cette propriété dit par exemple que même si le numéro 6 n'est pas sorti pendant 50 semaines consécutives à la loterie, cela ne rend pas sa prochaine apparition plus probable : $P(X > 51 | X > 50) = P(X > 1)$, où X est l'instant de première apparition du 6.

Démonstration.
$\diamond$ Supposons d'abord qu'il existe $p \in]0,1[$ tel que $X \sim \mathcal{G}(p)$ et soit $(n,k) \in \mathbb{N}^2$.
$P(X > n+k | X > n) = \dfrac{P(X > n+k)}{P(X > n)}$, or

$P(X > n) = \sum_{k=n+1}^{+\infty} P(X = k) = \sum_{k=n+1}^{+\infty} (1-p)^{k-1}p = p(1-p)^n \frac{1}{1-(1-p)} = (1-p)^n$,
donc $P(X > n+k | X > n) = \frac{(1-p)^{n+k}}{(1-p)^n} = (1-p)^k = P(X > k)$.

$\diamond$ Réciproquement, supposons que X est sans mémoire.
Si l'on pose $u_n = P(X > n)$, on a donc, pour tout $(n,k) \in \mathbb{N}^{*2}$, $u_{n+k} = u_n u_k$. En particulier, pour tout $n \in \mathbb{N}^*$, $u_{n+1} = u_n u_1$. On en déduit par récurrence que, pour tout $n \in \mathbb{N}^*$, $u_n = u_1^n$.
On a aussi $u_0 = P(X > 0) = 1 = u_1^0$.
L'énoncé suppose implicitement que pour tout $n \in \mathbb{N}$, $P(X > n) > 0$, sinon les probabilités conditionnelles de l'énoncé ne seraient pas définies,
donc $u_1 = P(X > 1) > 0$. De plus $u_1 < 1$, sinon pour tout $n \in \mathbb{N}$, $P(X > n) = u_1^n = 1$,
donc $P(X \leq n) = 0$ puis $1 = P(X \in \mathbb{N}^*) = \sum_{n=1}^{+\infty} P(X = n) = 0$, ce qui est faux.
Ainsi $u_1 \in]0,1[$. Posons $p = 1 - u_1 \in]0,1[$.
Pour tout $n \in \mathbb{N}^*$, $(X = n) = (X > n-1) \setminus (X > n)$,
donc $P(X = n) = P(X > n-1) - P(X > n) = (1-p)^{n-1} - (1-p)^n = p(1-p)^{n-1}$. □

4 Variables aléatoires indépendantes

4.1 Lois conjointes et lois marginales

Définition. Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé.
Soit $n \in \mathbb{N}^*$. Pour tout $i \in \{1, \dots, n\}$, on considère une variable aléatoire X_i discrète de Ω dans un ensemble E_i.
Alors, en posant pour tout $\omega \in \Omega$, $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$, on définit une variable aléatoire discrète $X \triangleq (X_1, \dots, X_n)$ de Ω dans $E_1 \times \cdots \times E_n$.

On dit que la loi de X est la loi conjointe des variables aléatoires $X_1, \dots, X_n$.
Pour tout $i \in \{1, \dots, n\}$, la loi de X_i est appelée la $i$ème loi marginale de X.
En résumé, un n-uplet de variables aléatoires discrètes est une variable aléatoire discrète.

Démonstration.
$X(\Omega) \subset \prod_{i=1}^{n} X_i(\Omega)$ est fini ou dénombrable et, pour tout $(k_1, \dots, k_n) \in \prod_{i=1}^{n} X_i(\Omega)$,
$X^{-1}(\{(k_1, \dots, k_n)\}) = \bigcap_{i=1}^{n} X_i^{-1}(\{k_i\}) \in \mathcal{F}$, donc X est bien une variable aléatoire discrète. □

Définition. Un tel n-uplet est appelé un **vecteur aléatoire discret** lorsque les variables aléatoires X_i sont toutes réelles.

Exemple. Soit $X = (X_1, X_2)$ un couple de variables aléatoires entières. On note $(p_{1,k}) = (P(X_1 = k))$ la première loi marginale de X et $(p_{2,k}) = (P(X_2 = k))$ la seconde loi marginale.
On note également $c_{h,k} = P(X = (h, k))$ la loi conjointe.
Alors $p_{1,k} = \sum_{h\in\mathbb{N}} c_{k,h}$ et $p_{2,k} = \sum_{h\in\mathbb{N}} c_{h,k}$.

Démonstration.
$p_{1,k} = P(X_1 = k) = P(\bigcup_{h\in\mathbb{N}} [X = (k, h)]) = \sum_{h\in\mathbb{N}} P(X = (h, k))$. □

Remarque. La connaissance des lois marginales ne permet pas de retrouver la loi conjointe.

> **Exercice.** Un sac contient 4 boules numérotées de 1 à 4. On tire deux boules avec remise. On note X_1 le numéro de la première boule, X_2 le numéro de la deuxième boule et Y le plus grand des deux numéros obtenus. Représenter dans un tableau la loi conjointe du couple (X_1, Y) ainsi que leurs lois marginales.
>
> **Solution :**
> $P([X_1 = 1] \cap [Y = 3]) = P([X_1 = 1] \cap [X_2 = 3]) = P([X_1 = 1])P([X_2 = 3])$ car, le tirage étant avec remise, les deux événements $X_1 = i$ et $X_2 = j$ sont indépendants, donc $P([X_1 = 1] \cap [Y = 3]) = \frac{1}{16}$.
> De même, $P([X_1 = 3] \cap [Y = 3]) = P([X_1 = 3] \cap [X_2 \in \{1, 2, 3\}]) = 3/16$, ...En effectuant tous les calculs, on obtient le tableau suivant qui résume la loi conjointe du couple (X_1, Y) ainsi que leurs lois marginales.

$X_1 \setminus Y$	1	2	3	4	P_{X_1}
1	1/16	1/16	1/16	1/16	4/16
2	0	2/16	1/16	1/16	4/16
3	0	0	3/16	1/16	4/16
4	0	0	0	4/16	4/16
P_Y	1/16	3/16	5/16	7/16	

Définition. Soit $X = (X_1, X_2)$ un couple de variables aléatoires discrètes. Pour tout $h \in X_2(\Omega)$ tel que $P(X_2 = h) > 0$, la loi conditionnelle de X_1 sachant que $X_2 = h$ désigne la probabilité $A \longmapsto P(X_1 \in A | X_2 = h)$ (définie sur $\mathcal{P}(X_1(\Omega))$). Elle est caractérisée par la suite des $(P(X_1 = k|X_2 = h))_{k\in\mathbb{N}}$.
On définit de même la loi conditionnelle de X_2 sachant que $X_1 = k$.

Exemple. Avec les notations de l'exemple précédent, $P(X_1 = k|X_2 = h) = \dfrac{c_{k,h}}{p_{2,h}}$.

4.2 Indépendance

Définition. Soit $X = (X_1, \ldots, X_n)$ un n-uplet de variables discrètes. On dit qu'elles sont mutuellement indépendantes si et seulement si

pour tout $k = (k_1, \dots, k_n) \in \prod_{i=1}^{n} X_i(\Omega)$, $P(X = k) = \prod_{i=1}^{n} P(X_i = k_i)$.

Remarque. Avec les notations précédentes, si $k = (k_1, \dots, k_n) \in \prod_{i=1}^{n} X_i(\Omega)$ et $k \notin X(\Omega)$, alors $P(X = k) = 0$.

Propriété. Avec les notations précédentes, $X_1, \dots, X_n$ sont indépendantes si et seulement si pour toute famille $K_1, \dots, K_n$ de parties de $X_1(\Omega), \dots, X_n(\Omega)$,

$$P(X_1 \in K_1, \dots, X_n \in K_n) = \prod_{i=1}^{n} P(X_i \in K_i),$$

c'est-à-dire si et seulement si les événements $X_i \in K_i$ pour $i \in \{1, \dots, n\}$ sont mutuellement indépendants.

Démonstration.
L'implication "$\Longleftarrow$" est évidente. Montrons la réciproque.
On suppose que $X_1, \dots, X_n$ sont indépendantes. Alors

$$\begin{aligned} P(X_1 \in K_1, \dots, X_n \in K_n) &= \sum_{k_1 \in K_1, \dots, k_n \in K_n} P(X_1 = k_1, \dots, X_n = k_n) \\ &= \sum_{k_1 \in K_1, \dots, k_n \in K_n} P(X_1 = k_1) \times \cdots \times P(X_n = k_n) \\ &= \left(\sum_{k_1 \in K_1} P(X_1 = k_1) \right) \times \cdots \times \left(\sum_{k_n \in K_n} P(X_n = k_n) \right) \\ &= \prod_{i=1}^{n} P(X_i \in K_i). \end{aligned}$$

L'avant-dernière égalité est claire lorsque les ensembles K_i sont de cardinaux finis mais elle mérite quelques précisions dans le cas général :
Pour tout $i \in \{1, \dots, n\}$, K_i étant fini ou dénombrable, il existe une suite croissante $(J_{i,h})_{h \in \mathbb{N}}$ de parties finies de K_i dont la réunion vaut K_i.
Alors la suite $(J_{1,h} \times \cdots \times J_{n,h})_{h \in \mathbb{N}}$ est adaptée à $K_1 \times \cdots \times K_n$, donc d'après le cours sur les familles sommables de réels positifs,

$$\begin{aligned} \prod_{i=1}^{n} P(X_i \in K_i) &= \lim_{h \to +\infty} \left(\sum_{k_1 \in J_{1,h}} P(X_1 = k_1) \right) \times \cdots \times \left(\sum_{k_n \in J_{n,h}} P(X_n = k_n) \right) \\ &= \lim_{h \to +\infty} \sum_{k_1 \in J_{1,h}, \dots, k_n \in J_{n,h}} P(X_1 = k_1) \times \cdots \times P(X_n = k_n) \\ &= \sum_{k_1 \in K_1, \dots, k_n \in K_n} P(X_1 = k_1) \times \cdots \times P(X_n = k_n). \end{aligned}$$

□

Remarque. Si $X_1, \dots, X_n$ sont des variables aléatoires mutuellement indépendantes, alors elles sont 2 à 2 indépendantes, mais la réciproque est fausse.

Démonstration.
On adapte la démonstration établie dans le cadre des événements indépendants :

Prenons $\Omega = \{0,1\}^3$, chaque événement élémentaire ayant pour probabilité $\frac{1}{8}$.
Posons $X = (X_1, X_2, X_3) = Id_\Omega$.
$|X_2 - X_1|$ et $|X_3 - X_1|$ sont des variables aléatoires entières indépendantes car pour $(\alpha, \beta) \in \{0,1\}^2$, en étudiant les quatre valeurs possibles du couple (α, β),
$P(|X_2 - X_1| = \alpha, |X_3 - X_1| = \beta) = \frac{1}{4} = P(|X_2 - X_1| = \alpha) \times P(|X_3 - X_1| = \beta)$.
Ainsi, par permutation circulaire, les 3 variables aléatoires $|X_2 - X_1|, |X_3 - X_1|$ et $|X_3 - X_2|$ sont 2 à 2 indépendantes.
Cependant elles ne sont pas mutuellement indépendantes, car
$P(|X_2 - X_1| = 0, |X_3 - X1| = 0, |X_3 - X_2| = 0) = \frac{1}{4}$ alors que
$P(|X_2 - X_1| = 0) \times P(|X_3 - X1| = 0) \times P(|X_3 - X_2| = 0) = \frac{1}{8}$. □

Définition. Si $(X_i)_{i \in I}$ est une famille de variables aléatoires discrètes, avec I de cardinal infini, on dit que ces variables aléatoires sont mutuellement indépendantes si et seulement si pour toute partie finie J incluse dans I, les variables aléatoires X_j pour $j \in J$ sont mutuellement indépendantes.

Propriété. Soit X et Y deux variables aléatoires discrètes indépendantes de Ω dans E et F respectivement. Soit $f : E \longmapsto E'$ et $g : F \longmapsto F'$ deux fonctions. Alors $f(X)$ et $g(Y)$ sont encore deux variables aléatoires discrètes indépendantes.

Démonstration.
Soit $x' \in f \circ X(\Omega)$ et $y' \in g \circ Y(\Omega)$.
$P(f(X) = x', g(Y) = y') = P(X \in f^{-1}(\{x'\}), Y \in g^{-1}(\{y'\}))$, or X et Y sont indépendantes, donc

$$\begin{aligned} P(f(X) = x', g(Y) = y') &= P(X \in f^{-1}(\{x'\})) \times P(Y \in g^{-1}(\{y'\})) \\ &= P(f(X) = x') \times P(g(Y) = y'). \end{aligned}$$

□

Remarque. On peut généraliser l'énoncé et la démonstration au cas suivant : Si $(X_i)_{i \in I}$ est une famille de variables aléatoires mutuellement indépendantes, alors pour toute famille de fonctions $(f_i)_{i \in I}$ correctement définies, $(f_i(X_i))_{i \in I}$ est encore une famille de variables aléatoires mutuellement indépendantes.

Corollaire. Lemme de coalition. Soit $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ des variables aléatoires discrètes mutuellement indépendantes. Alors pour toutes fonctions f et g correctement définies, les variables aléatoires $f(X_1, \ldots, X_m)$ et $g(Y_1, \ldots, Y_n)$ sont indépendantes.

Démonstration.
$X = (X_1, \ldots, X_m)$ et $Y = (Y_1, \ldots, Y_n)$ sont indépendantes, donc il suffit d'appliquer la propriété précédente. □

Remarque. Là encore, on peut généraliser ...

Propriété. Soit $X_1, \ldots, X_m$ des variables aléatoires entières mutuellement indépendantes. On suppose qu'il existe $p \in [0,1]$ tel que, pour tout $i \in \{1, \ldots, m\}$, $X_i \sim \mathcal{B}(n_i, p)$, où $n_i \in \mathbb{N}^*$ (p ne dépend pas de i).
Alors $X_1 + \cdots + X_m \sim \mathcal{B}(n_1 + \cdots + n_m, p)$.

Démonstration.
◇ Commençons par le cas où $m = 2$, en changeant un peu les notations :
On suppose que X et Y sont deux variables aléatoires entières indépendantes avec $X \sim \mathcal{B}(n,p)$ et $Y \sim \mathcal{B}(m,p)$. On doit montrer que $X+Y \sim \mathcal{B}(n+m,p)$.
Soit $k \in \{0,\dots,n+m\}$:
$P(X+Y=k) = P(\bigcup_{i+j=k} (X=i)\cap(Y=j)) = \sum_{i+j=k} P((X=i)\cap(Y=j))$.
X et Y étant indépendantes,
$P(X+Y=k) = \sum_{\substack{i+j=k \\ 0\le i\le n,\ 0\le j\le m}} P(X=i)P(Y=j)$,

donc $P(X+Y=k) = \sum_{\substack{i+j=k \\ 0\le i\le n,\ 0\le j\le m}} \binom{n}{i} p^i (1-p)^{n-i} \binom{m}{j} p^j (1-p)^{m-j}$,

puis $P(X+Y=k) = p^k(1-p)^{n-k+m} \sum_{\substack{i+j=k \\ 0\le i\le n,\ 0\le j\le m}} \binom{n}{i}\binom{m}{j}$.

Ainsi, il reste à montrer que $\sum_{\substack{i+j=k \\ 0\le i\le n,\ 0\le j\le m}} \binom{n}{i}\binom{m}{j} = \binom{n+m}{k}$, ce qui est immédiat,
car pour choisir k éléments parmi un ensemble E de $n+m$ éléments, en ayant préalablement partionné E en $E = F \cup G$ avec F de cardinal n et G de cardinal m, il suffit de choisir i éléments dans F et j éléments dans G avec $i+j=k$.
◇ Démontrons maintenant le cas général par récurrence sur m.
Notons $R(m)$ la propriété à démontrer. Elle est évidente pour $m=1$ et on vient de démontrer $R(2)$.
Pour $m \ge 3$, supposons $R(m-1)$ et montrons $R(m)$.
D'après $R(m-1)$, $X_1+\cdots+X_{m-1} \sim \mathcal{B}(n_1+\cdots+n_{m-1},p)$, mais d'après une propriété précédente, $X_1+\cdots+X_{m-1}$ et X_m sont indépendantes, donc d'après $R(2)$,
$X_1+\cdots+X_m = (X_1+\cdots+X_{m-1})+X_m \sim \mathcal{B}(n_1+\cdots+n_m,p)$. □

Remarque. Vous verrez en seconde année une démonstration plus rapide qui utilise les fonctions génératrices.

Remarque. Considérons m épreuves indépendantes, chaque épreuve ayant une probabilité de succès égale à $p \in [0,1]$, avec p indépendant de l'épreuve. Pour la i-ème épreuve, notons X_i la variable aléatoire à valeurs dans $\{0,1\}$ telle que
$X_i = 1 \iff$ [la i-ème épreuve est un succès]. Ainsi $X_1,\dots,X_m$ est une suite de m variables aléatoires indépendantes de Bernoulli de paramètre p.
On sait que $\mathcal{B}(p) = \mathcal{B}(1,p)$, donc la propriété précédente montre que
$X = X_1+\cdots+X_m \sim \mathcal{B}(m,p)$. Or X désigne le nombre de succès parmi les m épreuves. On a donc montré que le nombre de succès parmi une suite de m épreuves indépendantes de loi de Bernoulli de paramètre p suit une loi binomiale de paramètres m et p.

Exemple. Considérons la distribution aléatoire de r balles dans n urnes. Quelle est la probabilité qu'une urne donnée contienne exactement k balles ?
Première modélisation, à l'aide d'événements :
Posons $\Omega = \{1, \dots, n\}^r$ et convenons que l'événement élémentaire $\omega = (b_1, \dots, b_r)$ correspond au fait que pour tout $i \in \mathbb{N}_r$, la i-ième balle soit dans l'urne de numéro b_i.
On se donne $h \in \{1, \dots, n\}$ et on cherche $P(A)$ où
$A = \{(b_1, \dots, b_r) \in \Omega / \text{Card}(\{i \in \{1, \dots, r\} / b_i = h\}) = k\}$.
Pour choisir un élément de A, on peut d'abord choisir les k balles présentes dans l'urne h, parmi les r balles, puis on choisit la répartition des $r - k$ balles restantes dans les $n - 1$ urnes disponibles. Il s'ensuit que la probabilité en question est donnée par
$P(A) = \dfrac{\binom{r}{k} \times (n-1)^{r-k}}{n^r}$.
Seconde modélisation, à l'aide de variables aléatoires :
On désigne par B_i la variable aléatoire représentant le numéro de l'urne contenant la i-ème balle : $P(B_i = h) = \frac{1}{n}$. Les variables aléatoires $B_1, \dots, B_r$ sont mutuellement indépendantes.
Posons maintenant $T_i = 1_{[B_i = h]}$. Ainsi $T_i \sim \mathcal{B}(\frac{1}{n})$ et T_i "teste" si la i-ème balle est bien dans l'urne numéro h.
Or on a vu que $T_1 + \dots + T_r \sim \mathcal{B}(r, \frac{1}{n})$,
donc $P(A) = P(T_1 + \dots + T_r = k) = \binom{r}{k} (\frac{1}{n})^k (1 - \frac{1}{n})^{r-k}$.

Exercice. Soit $X_1, \dots, X_m$ des variables aléatoires entières mutuellement indépendantes telles que chaque X_i suit une loi de Poisson de paramètre $\lambda_i > 0$. Montrer que $X = X_1 + \dots + X_m$ suit une loi de Poisson de paramètre $\lambda = \lambda_1 + \dots + \lambda_m$.

Solution : Il suffit d'adapter la démonstration précédente. La récurrence est similaire.
On suppose que X et Y sont deux variables aléatoires entières indépendantes avec $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$. On doit montrer que $X + Y \sim \mathcal{P}(\lambda + \mu)$.
Soit $k \in \mathbb{N}$:
$P(X + Y = k) = P(\bigcup_{i+j=k} (X = i) \cap (Y = j)) = \sum_{i+j=k} P(X = i) P(Y = j)$,
donc $P(X + Y = k) = \sum_{i+j=k} e^{-\lambda} \frac{\lambda^i}{i!} e^{-\mu} \frac{\mu^j}{j!} = \frac{e^{-\lambda-\mu}}{k!} \sum_{i+j=k} \frac{k!}{i!j!} \lambda^i \mu^j = e^{-\lambda-\mu} \frac{(\lambda+\mu)^k}{k!}$
d'après la formule du binôme de Newton.

Remarque. Ici encore, les fonctions génératrices vous permettront en seconde année d'obtenir ce résultat plus rapidement.

Propriété. Soit $(p_n) \in]0,1[^{\mathbb{N}}$ telle que $np_n \underset{n \to +\infty}{\longrightarrow} \lambda \in \mathbb{R}_+^*$. Soit (X_n) une suite de variables aléatoires telle que X_n suit la loi binomiale de paramètres n et p_n. Alors X_n converge en loi vers la loi de Poisson de paramètre λ.

Démonstration.
Fixons $k \in \mathbb{N}$. Il s'agit de montrer que $P(X_n = k) \underset{n\to+\infty}{\longrightarrow} \dfrac{\lambda^k e^{-\lambda}}{k!}$. Or

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{1}{k!} \times \frac{n}{n} \times \frac{n-1}{n} \times \cdots \times \frac{n-k+1}{n} (np_n)^k (1-p_n)^n (1-p_n)^{-k} \\ &\underset{n\to+\infty}{\sim} \frac{1}{k!}\lambda^k (1-p_n)^n. \end{aligned}$$

De plus $(1-p_n)^n = e^{n\ln(1-\frac{\lambda}{n}+o(\frac{1}{n}))} = e^{-\lambda+o(1)} \underset{n\to+\infty}{\longrightarrow} e^{-\lambda}$, ce qui permet de conclure. □

Remarque. Vue la démonstration, l'approximation de la loi de X_n par une loi de Poisson est d'autant plus valable que n est grand (par rapport à k) et que λ est petit par rapport à n : il faut donc $k << n$ et $\lambda << n$.

Application : Soit V un volume de N atomes radioactifs. Notons X la variable aléatoire désignant le nombre d'atomes qui se désintègrent par unité de temps. Si p est la probabilité de désintégration d'un atome pendant la même unité de temps, alors $X \sim \mathcal{B}(N,p)$.
Pour $k << N$ et $p << 1$, on peut considérer que $X \sim \mathcal{P}(\lambda)$, où λ est le taux moyen de désintégration par unité de temps.

Application : Loi des événements rares. Présentation informelle :
Dans une file d'attente, supposons que le nombre moyen d'individus arrivant par unité de temps vaut $\lambda > 0$. Quelle est la loi suivie par le nombre N d'individus arrivant pendant une unité de temps ?
Si l'on découpe l'unité de temps en n durées égales, notons N_i le nombre d'individus arrivant entre les temps $\frac{i-1}{n}$ et $\frac{i}{n}$. Les variables aléatoires $N_1, \ldots, N_n$ sont indépendantes de moyenne $\frac{\lambda}{n}$.
Pour n assez grand, on considère qu'au plus un individu peut arriver entre les temps $\frac{i-1}{n}$ et $\frac{i}{n}$ (c'est l'hypothèse des événements rares), donc N_i suit une loi de Bernoulli de paramètre $\frac{\lambda}{n}$. Ainsi N suit une loi binomiale de paramètres $\frac{\lambda}{n}$ et n. Si l'on fait tendre n vers l'infini, on voit que N suit une loi de Poisson de paramètre λ.
Les atomes radioactifs précédents peuvent aussi être modélisés par une file d'attente : les atomes désintégrés rejoignent la file d'attente.

Définition.
Une loi discrète sur un ensemble E est la donnée d'une probabilité P sur E muni de sa tribu pleine $\mathcal{P}(E)$ telle que, en posant $S = \{x \in E / P(x) > 0\}$, on ait $\displaystyle\sum_{x\in S} P(\{x\}) = 1$.
Alors S est au plus dénombrable et, pour tout $B \subset E$, $P(B) = P(B \cap S)$.

Théorème. Soit $(E_n)_{n\in\mathbb{N}}$ une suite d'ensembles et pour tout $n \in \mathbb{N}$, soit $\mathcal{L}_n$ une loi discrète sur E_n. Alors il existe un espace probabilisé $(\Omega, \mathcal{F}, P)$ et une suite (X_n) de variables aléatoires discrètes mutuellement indépendantes telle que, pour tout $n \in \mathbb{N}$, $X_n \sim \mathcal{L}_n$.

Démonstration.
Admis ! La démonstration est hors programme. □

Remarque. Ce théorème est indispensable pour modéliser rigoureusement des processus stochastiques, même très simples. Par exemple, le jeu de pile ou face infini est modélisé par la donnée d'une suite (X_n) de variables aléatoires indépendantes telles que pour tout $n \in \mathbb{N}$, $X_n \sim \mathcal{B}(\frac{1}{2})$. Le théorème fournit un espace Ω inconnu. Pour $\omega \in \Omega$, la suite $(X_n(\omega))_{n\in\mathbb{N}}$ est une réalisation d'un jeu de pile ou face infini. On peut donc faire des probabilités sur un tel modèle.
Cela permet de formaliser dans l'exercice page 14 les questions 2 et 3.

Remarque. De même, ce théorème prouve l'existence d'une suite $(X_n)_{n\geq 1}$ de variables aléatoires indépendantes telles que pour tout $n \in \mathbb{N}^*$, $X_n \sim \mathcal{B}(p)$, où $p \in]0,1[$ ne dépend pas de n. Cette suite modélise une succession infinie d'épreuves indépendantes qui ont toutes la même probabilité de succès, égale à p.
Montrons maintenant que si X désigne l'instant du premier succès, alors $X \sim \mathcal{G}(p)$.
Formellement, pour tout $\omega \in \Omega$, $X(\omega) = \min\{k \in \mathbb{N}^* / X_k(\omega) = 1\}$.
Si $n \in \mathbb{N}^*$, pour tout $\omega \in \Omega$,
$X(\omega) = n \iff [X_n(\omega) = 1 \text{ et } \forall k \in \{1,\dots,n-1\},\ X_k(\omega) = 0]$, mais les X_k sont mutuellement indépendants,
donc $P(X = n) = P(X_n = 1) \times \prod_{k=1}^{n-1} P(X_k = 0) = p(1-p)^{n-1}$.

Nous pouvons maintenant reprendre l'exercice vu en page 5 :

> **Exercice.** Deux joueurs A et B lancent un dé non pipé à tour de rôle (en commençant par A). Le premier qui obtient 6 a gagné et le jeu s'arrête.
>
> **1°)** Quel est l'univers Ω pour cette expérience ?
>
> **2°)** Calculer les probabilités des événements "A gagne", "B gagne", "ni A ni B ne gagne".
>
> **Solution :**
>
> **1°)** Selon un théorème du cours, il existe un espace probabilisé $(\Omega, \mathcal{F}, P)$ et une suite (X_n) de variables aléatoires mutuellement indépendantes telle que, pour tout $n \in \mathbb{N}^*$, $X_n \sim \mathcal{B}(\frac{1}{6})$. On choisira ici cet univers Ω, ce qui revient à considérer que les joueurs continuent (virtuellement) à lancer le dé après l'arrêt du jeu.
>
> **2°)** Notons Y la variable aléatoire égale à l'instant de première apparition d'un 6. D'après le cours, $Y \sim \mathcal{G}(\frac{1}{6})$.
> Ainsi, $P(A \text{ gagne}) = P(Y \text{impair}) = \sum_{n=0}^{+\infty} P(Y = 2n+1) = \sum_{n=0}^{\infty} \frac{1}{6}(\frac{5}{6})^{2n}$,
> donc $P(A \text{ gagne}) = \frac{1}{6}\frac{1}{1-(\frac{5}{6})^2} = \frac{6}{11}$.
> $P(B \text{ gagne}) = P(Y \text{ pair}) = 1 - P(A \text{ gagne}) = \frac{5}{11}$
> et $P(\text{ni } A \text{ ni } B \text{ ne gagne}) = P(Y = +\infty) = 0$.

5 Espérance et variance

5.1 L'espérance

Définition. Soit X une variable aléatoire discrète à valeurs dans $\mathbb{C}$.
$\diamond$ Si X est à valeurs dans $\mathbb{R}_+$, on appelle espérance de X la quantité dans $[0, +\infty]$
$\boxed{E(X) \triangleq \sum_{d \in X(\Omega)} d.P(X = d)}$.
$\diamond$ Sinon, on dit que X est d'espérance finie si et seulement si la famille $(d.P(X = d))_{d \in X(\Omega)}$ est sommable, et dans ce cas, on appelle espérance de X la quantité $E(X) \triangleq \sum_{d \in X(\Omega)} d.P(X = d)$

Remarque. $E(X)$ représente la valeur moyenne de X, puisque c'est la moyenne de ses valeurs pondérées par leurs probabilités respectives. Ce résultat sera formalisé par la loi faible des grands nombres.

Remarque. $E(X)$ ne dépend que de la loi de X.

Remarque. Lorsque $X : \Omega \longrightarrow \mathbb{R}$ est une variable aléatoire réelle, on représente parfois sa loi par un *diagramme en bâtons*. Alors $E(X)$ est la moyenne des $d \in X(\Omega)$, pondérée par les hauteurs des bâtons.

Propriété. Si Ω est fini, alors $E(X) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\})$

Démonstration.
$E(X) = \sum_{d \in X(\Omega)} d \times \sum_{\substack{\omega \in \Omega \\ X(\omega)=d}} P(\{\omega\}) = \sum_{d \in X(\Omega)} \sum_{\substack{\omega \in \Omega \\ X(\omega)=d}} X(\omega)P(\{\omega\}) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\})$, car les $\{\omega \in \Omega / X(\omega) = d\}$ lorsque $d \in X(\Omega)$ constituent une partition de Ω. $\square$

Remarque. Cette relation reste vraie lorsque Ω est dénombrable, grâce à la théorie de la sommabilité, mais ce n'est pas prévu dans le programme officiel.

Exemple. si $X \sim \mathcal{B}(p)$, où $p \in [0, 1]$, $E(X) = 1 \times p + 0 \times (1 - p) = p$.

Propriété. Si A est un événement de l'espace probabilisé $(\Omega, \mathcal{F}, P)$,
alors $P(A) = E(1_A)$, où 1_A désigne la fonction caractéristique de la partie A de Ω.

Démonstration.
On sait déjà que $1_A \sim \mathcal{B}(P(A))$. $\square$

Exemple. Si $X \sim \mathcal{P}(\lambda)$, où $\lambda > 0$,
$E(X) = \sum_{n=0}^{+\infty} n\frac{\lambda^n}{n!}e^{-\lambda} = \sum_{n=1}^{+\infty} \lambda\frac{\lambda^{n-1}}{(n-1)!}e^{-\lambda} = \lambda$.
C'est un résultat attendu en regard de la modélisation des événements rares.

Exemple. Si X est une variable aléatoire suivant une loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$, alors $E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$. D'après la formule comité-président, $E(X) = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{(n-1)-(k-1)} = np(p+(1-p))^{n-1} = np$.
Si l'on effectue n épreuves indépendantes de loi de Bernoulli de paramètre p, le nombre moyen de succès attendu est bien de np.

Exemple. Si $X \sim \mathcal{G}(p)$, où $p \in]0,1[$,
$E(X) = \sum_{n=1}^{+\infty} np(1-p)^{n-1} = p \sum_{n=1}^{+\infty} n x^{n-1}$ où $x = 1-p \in]0,1[$.
Pour tout $N \geq 1$, $\sum_{n=1}^{N} n x^{n-1} = \frac{d}{dx}(\sum_{n=0}^{N} x^n) = \frac{d}{dx}(\frac{1-x^{N+1}}{1-x})$,
donc $\sum_{n=1}^{N} n x^{n-1} = \frac{-(N+1)x^N(1-x)+(1-x^{N+1})}{(1-x)^2} \underset{N\to+\infty}{\longrightarrow} \frac{1}{(1-x)^2}$,
donc $E(X) = p\frac{1}{(1-(1-p))^2} = \frac{1}{p}$.
Si l'on effectue n épreuves indépendantes de loi de Bernoulli de paramètre p, on attend bien $\frac{1}{p}$ pour l'instant moyen du premier succès.

Remarque. Dans le calcul précédent, on peut éviter de passer par N si l'on dispose de la théorie des séries entières.
Les fonctions génératrices fourniront une autre méthode pour calculer les espérances.

Définition. Une variable aléatoire réelle est dite centrée si et seulement si $E(X) = 0$.

> **Exercice.** Montrer qu'une variable aléatoire réelle et positive est centrée si et seulement si elle est nulle presque sûrement.
>
> **Solution :**
> Soit X une variable aléatoire positive telle que $E(X) = 0$.
> Ainsi, $\sum_{d \in X(\Omega)} dP(X=d) = 0$ et pour tout $d \in X(\Omega)$, $dP(X=d) \geq 0$, donc pour tout $d \in X(\Omega)$, $dP(X=d) = 0$. Ainsi, pour tout $d \in X(\Omega) \cap \mathbb{R}_+^*$, $P(X=d) = 0$. On en déduit que $P(X>0) = 0$, donc $P(X=0) = 1$: X est nulle presque sûrement.
> Réciproquement, il est clair que si $P(X=0) = 1$, alors $E(X) = 0$.

Théorème de transfert : Soit X une variable aléatoire discrète définie de Ω dans un ensemble E et soit g une application de E dans $\mathbb{C}$.
$g(X)$ est d'espérance finie si et seulement si la famille $(g(d).P(X=d))_{d \in X(\Omega)}$ est sommable, et dans ce cas, $\boxed{E(g(X)) = \sum_{d \in X(\Omega)} g(d)P(X=d)}$.

Démonstration.
◇ $g(X)$ est d'espérance finie si et seulement si la famille $\Big(|yP(g(X)=y)|\Big)_{y\in g(X)(\Omega)}$ est sommable, or en travaillant dans $\mathbb{R}_+\cup\{+\infty\}$, on peut écrire

$$\begin{aligned}\sum_{y\in g(X(\Omega))}|y|P(g(X)=y) &= \sum_{y\in g(X(\Omega))}|y|P(X\in g^{-1}(\{y\}))\\ &= \sum_{y\in g(X(\Omega))}|y|\sum_{\substack{x\in X(\Omega)\\ tq\ \ g(x)=y}}P(X=x)\\ &= \sum_{y\in g(X(\Omega))}\ \sum_{\substack{x\in X(\Omega)\\ tq\ \ g(x)=y}}|g(x)|P(X=x),\end{aligned}$$

or $X(\Omega)=\bigsqcup_{y\in g(X(\Omega))}\{x\in X(\Omega)\ /\ g(x)=y\}$, donc par sommation par paquets pour des familles de réels positifs, $\sum_{y\in g(X(\Omega))}|y|P(g(X)=y)=\sum_{x\in X(\Omega)}|g(x)|P(X=x)$.
Ainsi $g(X)$ est d'espérance finie si et seulement si la famille $(g(d).P(X=d))_{d\in X(\Omega)}$ est sommable.
◇ Supposons que c'est le cas. Alors on peut écrire

$$\begin{aligned}E(g(X))=\sum_{y\in g(X(\Omega))}yP(g(X)=y) &= \sum_{y\in g(X(\Omega))}yP(X\in g^{-1}(\{y\}))\\ &= \sum_{y\in g(X(\Omega))}y\sum_{\substack{x\in X(\Omega)\\ tq\ \ g(x)=y}}P(X=x)\\ &= \sum_{y\in g(X(\Omega))}\ \sum_{\substack{x\in X(\Omega)\\ tq\ \ g(x)=y}}g(x)P(X=x),\end{aligned}$$

or $X(\Omega)=\bigsqcup_{y\in g(X(\Omega))}\{x\in X(\Omega)\ /\ g(x)=y\}$, donc par sommation par paquets pour des familles **sommables** de réels, $E(g(X))=\sum_{x\in X(\Omega)}g(x)P(X=x)$.

□

Exercice. Paradoxe de Saint-Pétersbourg : On jette une pièce de monnaie jusqu'à l'apparition du premier "face" ; si cela a lieu au T-ème lancer, votre gain sera de 2^T euros.
Pour avoir le droit de jouer, il faut s'acquitter d'une mise. Quelle serait une mise équitable ?

Solution : La mise équitable est l'espérance du gain : $M=E(2^T)$.
On a $T\sim\mathcal{G}(\frac{1}{2})$, donc d'après la formule de transfert,
$M=\sum_{n=1}^{+\infty}2^nP(T=n)=\sum_{n=1}^{+\infty}2^n\frac{1}{2}(\frac{1}{2})^{n-1}=\sum_{n=1}^{+\infty}1=+\infty$.
En conclusion, quelle que soit la mise de départ, on a théoriquement intérêt à jouer. En pratique, accepteriez-vous de miser toute votre fortune ?

Linéarité de l'espérance : Soit X et Y deux variables aléatoires complexes d'espérances finies et soit $\alpha, \beta \in \mathbb{C}$.
Alors $\alpha X + \beta Y$ est encore d'espérance finie et $\boxed{E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)}$.

Démonstration.
Pour tout $\omega \in \Omega$, posons $Z(\omega) = (X(\omega), Y(\omega))$. Ainsi Z est une variable aléatoire discrète de Ω dans $\mathbb{C}^2$. Notons $h : \mathbb{C}^2 \longrightarrow \mathbb{C}$ l'application définie par $h(x,y) = \alpha x + \beta y$. Ainsi $\alpha X + \beta Y = h(Z)$.
Pour tout $(x,y) \in Z(\Omega)$, $h(x,y)P(Z = (x,y)) = (\alpha x + \beta y)P(X = x, Y = y)$, et pour $(x,y) \in (X(\Omega) \times Y(\Omega)) \setminus Z(\Omega)$, $h(x,y)P(Z = (x,y)) = 0 = (\alpha x + \beta y)P(X = x, Y = y)$. D'après la formule de transfert, si la famille $((\alpha x + \beta y)P(X = x, Y = y))_{x \in X(\Omega), y \in Y(\Omega)}$ est sommable, alors

$$\begin{aligned}
E(\alpha X + \beta Y) &= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} h(x,y) P((X,Y) = (x,y)) \\
&= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} (\alpha x + \beta y) P(X = x, Y = y) \\
&= \alpha \sum_{x \in X(\Omega)} x \sum_{y \in Y(\Omega)} P(X = x, Y = y) + \beta \sum_{y \in Y(\Omega)} y \sum_{x \in X(\Omega)} P(X = x, Y = y) \\
&= \alpha \sum_{x \in X(\Omega)} x P(X = x) + \beta \sum_{y \in Y(\Omega)} y P(Y = y) \\
&= \alpha E(X) + \beta E(Y).
\end{aligned}$$

Mais $|(\alpha x + \beta y)P(X = x, Y = y)| \leq (|\alpha||x| + |\beta||y|)P(X = x, Y = y)$, or en travaillant dans $[0, +\infty]$, on peut aussi écrire,
en posant $S = \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} (|\alpha||x| + |\beta||y|)P(X = x, Y = y)$:

$$\begin{aligned}
S &= |\alpha| \sum_{x \in X(\Omega)} |x| \sum_{y \in Y(\Omega)} P(X = x, Y = y) + |\beta| \sum_{y \in Y(\Omega)} |y| \sum_{x \in X(\Omega)} P(X = x, Y = y) \\
&= |\alpha| \sum_{x \in X(\Omega)} |x| P(X = x) + |\beta| \sum_{y \in Y(\Omega)} |y| P(Y = y) \\
&< +\infty, \text{ car } X \text{ et } Y \text{ sont d'espérances finies.}
\end{aligned}$$

L'inégalité précédente permet de conclure. □

Remarque. On en déduit notamment que si X est une variable aléatoire complexe d'espérance finie, pour tout $a, b \in \mathbb{C}$, $aX + b$ est d'espérance finie
et $E(aX + b) = aE(X) + b$.

Linéarité de l'espérance : traduction algébrique.
Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé.
$\boxed{\text{Notons } L^1 \text{ l'ensemble des variables aléatoires discrètes de } \Omega \text{ dans } \mathbb{C} \text{ d'espérance finie.}}$
L^1 est un espace vectoriel et l'espérance E est une forme linéaire sur L^1.

Lemme : Soit $X \in L^1$ une variable aléatoire presque sûrement constante, c'est-à-dire telle qu'il existe $c \in \mathbb{C}$ tel que $P(X = c) = 1$. Alors $E(X) = c$.

Démonstration.
$E(X) = \sum_{d \in X(\Omega)} d.P(X = d)$, or pour tout $d \in X(\Omega) \setminus \{c\}$, $P(X = d) = 0$, donc $E(X) = cP(X = c) = c$. □

Propriété. Soit $X \in L^1$. X est presque sûrement constante si et seulement si X est presque sûrement égale à son espérance.

Démonstration.
Le sens indirect est évident. Pour le sens direct, supposons qu'il existe $c \in \mathbb{C}$ tel que $P(X = c) = 1$. Alors $E(X) = c$, donc $P(X = E(X)) = 1$, ce qui signifie que X est presque sûrement égale à son espérance. □

Propriété. Positivité de l'espérance : Si X est à valeurs dans $\mathbb{R}_+$, alors $E(X) \geq 0$. On peut résumer en : $\boxed{X \geq 0 \Longrightarrow E(X) \geq 0}$.

Démonstration.
Il suffit d'utiliser la définition de l'espérance. □

Propriété : croissance de l'espérance :
Soit $X, Y \in L^1$. On suppose que X et Y sont à valeurs réelles et que $X \leq Y$ (c'est-à-dire que, pour tout $\omega \in \Omega$, $X(\omega) \leq Y(\omega)$). Alors $E(X) \leq E(Y)$.

Démonstration.
$Y - X \geq 0$, donc $E(Y - X) \geq 0$ et la linéarité de l'espérance permet de conclure. □

Propriété : Inégalité triangulaire :
Soit X une variable aléatoire discrète à valeurs complexes.
Alors $X \in L^1 \Longleftrightarrow |X| \in L^1$ et dans ce cas, $|E(X)| \leq E(|X|)$.

Démonstration.
D'après la formule de transfert, $|X|$ est d'espérance finie si et seulement si la famille $(|x|P(X = x))_{x \in X(\Omega)}$ est sommable, donc si et seulement si la famille $(xP(X = x))_{x \in X(\Omega)}$ est sommable, c'est-à-dire si et seulement si $X \in L^1$.
Dans ce cas, $E(X) = \sum_{d \in X(\Omega)} dP(X = d)$, donc par inégalité triangulaire,
$|E(X)| \leq \sum_{d \in X(\Omega)} |d|P(X = d) = E(|X|)$ (toujours d'après la formule de transfert). □

Propriété de comparaison : Soit X une variable aléatoire complexe et Y une variable aléatoire réelle telles que $|X| \leq Y$ avec Y d'espérance finie. Alors X est aussi d'espérance finie.

Démonstration.
Cette propriété ne se déduit pas immédiatement des précédentes, car, tant qu'elle n'est pas démontrée, on ne dispose pas de l'implication $0 \leq X \leq Y \in L^1 \Longrightarrow X \in L^1$. En effet, il faudrait passer par $Y - X$ et utiliser que $E(Y - X) = E(Y) - E(X)$, ce qui ne peut pas se faire en travaillant dans $\mathbb{R}_+ \cup \{+\infty\}$, ce qui donc nécessite de présupposer que $X, Y \in L^1$.

Pour tout $\omega \in \Omega$, posons $Z(\omega) = (X(\omega), Y(\omega))$. Ainsi Z est une variable aléatoire discrète de Ω dans $\mathbb{C} \times \mathbb{R}$. Notons $h : \mathbb{C} \times \mathbb{R} \longrightarrow \mathbb{R}$ l'application définie par $h(x, y) = |x| - y$. Ainsi $|X| - Y = h(Z)$. D'après la formule de transfert, si la famille $((|x| - y)P(X = x, Y = y))_{(x,y)\in Z(\Omega)}$ est sommable, alors $|X| - Y \in L^1(\Omega, P)$, puis par linéarité, $|X| = Y + (|X| - Y) \in L^1(\Omega, P)$, ce qu'il fallait démontrer.
Il reste donc à montrer que $\sum_{(x,y)\in Z(\Omega)} ||x| - y|P(X = x, Y = y) < +\infty$.
Si $(x, y) \in Z(\Omega)$, alors il existe $\omega \in \Omega$ tel que $(x, y) = Z(\omega) = (X(\omega), Y(\omega))$, donc $|x| = |X(\omega)| \leq Y(\omega) = y$. Ainsi $||x| - y| = y - |x| \leq y$. On peut donc écrire, dans le cadre des familles dénombrables de réels positifs :

$$\begin{aligned}\sum_{(x,y)\in Z(\Omega)} ||x| - y|P(X = x, Y = y) &\leq \sum_{(x,y)\in Z(\Omega)} yP(X = x, Y = y)\\ &= \sum_{x\in X(\Omega), y\in Y(\Omega)} yP(X = x, Y = y)\\ &= \sum_{y\in Y(\Omega)} yP(Y = y) = E(Y) < +\infty.\end{aligned}$$

□

Formule. Inégalité de Markov : X désigne une variable aléatoire réelle et $a > 0$.

$$\boxed{\text{Si } X \geq 0, \text{ alors } P(X \geq a) \leq \frac{E(X)}{a}}.$$

Démonstration.
$E(X) = \sum_{d\in X(\Omega)} dP(X = d) \geq \sum_{d\in X(\Omega) \text{ et } d\geq a} dP(X = d)$, car X est à valeurs positives,
donc $E(X) \geq \sum_{d\in X(\Omega) \text{ et } d\geq a} aP(X = d) = a \sum_{d\in X(\Omega) \text{ et } d\geq a} P(X = d) = aP(X \geq a)$. □

Théorème. Si $X_1, \ldots, X_k$ sont k variables aléatoires discrètes complexes d'espérances finies et **mutuellement indépendantes**, alors $X_1 \times \cdots \times X_k$ est d'espérance finie et $E(X_1 \times \cdots \times X_k) = E(X_1) \times \cdots \times E(X_k)$.
La réciproque est fausse.

Démonstration.
On le démontre seulement lorsque $k = 2$, le passage au cas quelconque étant simple par récurrence sur k.
Notons $h : \mathbb{C}^2 \longrightarrow \mathbb{C}$ l'application définie par $h(x, y) = xy$.
D'après la formule de transfert,

$$\begin{aligned}E(X_1X_2) &= \sum_{(x,y)\in X_1(\Omega)\times X_2(\Omega)} h(x, y)P((X_1, X_2) = (x, y))\\ &= \sum_{(x,y)\in X_1(\Omega)\times X_2(\Omega)} xyP(X_1 = x)P(X_2 = y) \text{ (car } X_1 \text{ et } X_2 \text{ sont indépendantes)}\\ &= \sum_{x\in X_1(\Omega)} xP(X_1 = x) \sum_{y\in X_2(\Omega)} yP(X_2 = y)\\ &= E(X_1)E(X_2).\end{aligned}$$

Comme dans des démonstrations précédentes, ce calcul est clair lorsque les sommes sont finies, et dans le cas de sommes dénombrables, la théorie de la sommabilité permet de le justifier.
Montrons maintenant que la réciproque est fausse en construisant un contre-exemple : Considérons $\Omega = \{-1, 0, 1\}$ muni de la distribution uniforme. On définit deux variables aléatoires par $X(\omega) = \omega$ et $Y(\omega) = |\omega|$.
Alors, $E(X) = 0$, $E(Y) = \frac{2}{3}$ et $E(XY) = 0$ (car $XY = X$).
Par conséquent $E(XY) = E(X)E(Y)$, mais elles ne sont pas indépendantes,
car $P(X = 0, Y = 0) = \frac{1}{3} \neq P(X = 0)P(Y = 0)$. □

5.2 La variance

Définition. Soit $k \in \mathbb{N}^*$ et X une variable aléatoire réelle. Si X^k est d'espérance finie, on dit que $E(X^k)$ est le moment d'ordre k de X.

Notation. On note $L^2(\Omega, P)$ l'ensemble des variables aléatoires X discrètes à valeurs réelles possédant un moment d'ordre 2, définies sur l'espace probabilisé $(\Omega, \mathcal{F}, P)$.

Lemme : Si $X_1, X_2 \in L^2(\Omega, P)$, alors $X_1X_2 \in L^1(\Omega, P)$.

Démonstration.
$2|X_1X_2| \leq X_1^2 + X_2^2$. La propriété de comparaison permet de conclure □

Corollaire.
◇ $L^2(\Omega, P) \subset L^1(\Omega, P)$.
◇ $L^2(\Omega, P)$ est un $\mathbb{R}$-espace vectoriel .

Démonstration.
◇ Notons $\mathbf{1}$ la fonction constante égale à 1 définie sur Ω.
$\mathbf{1}$ est une variable aléatoire positive et $E(\mathbf{1}^2) = 1$, donc $\mathbf{1} \in L^2(\Omega, P)$.
Si $X \in L^2(\Omega, P)$, on peut donc écrire $X = X \times \mathbf{1} \in L^1(\Omega, P)$.
◇ Soit $X, Y \in L^2(\Omega, P)$. Soit $\alpha \in \mathbb{R}$.
$(\alpha X + Y)^2 = \alpha^2 X^2 + Y^2 + 2\alpha XY$, or $X^2, Y^2, XY \in L^1(\Omega, P)$ et $L^1(\Omega, P)$ est un espace vectoriel, donc $(\alpha X + Y)^2 \in L^1(\Omega, P)$, puis $\alpha X + Y \in L^2(\Omega, P)$. Ainsi $L^2(\Omega, P)$ est non vide et stable par combinaison linéaire, donc c'est un sous-espace vectoriel de l'ensemble des fonctions de Ω dans $\mathbb{R}$. □

Définition. Si $X_1, X_2 \in L^2(\Omega, P)$, on appelle covariance entre X_1 et X_2
la quantité $\boxed{Cov(X_1, X_2) = E[(X_1 - E(X_1))(X_2 - E(X_2))]}$.
Les propriétés précédentes montrent que cette quantité est effectivement définie.

Propriété. Cov est une forme bilinéaire symétrique positive sur $L^2(\Omega, P)$.

Démonstration.
D'aprés la linéarité de l'espérance, si $X_1, X_2, X_3 \in L^2(\Omega, P)$, pour $\alpha \in \mathbb{R}$,

$$\begin{aligned} Cov(\alpha X_1 + X_2, X_3) &= E((\alpha X_1 + X_2 - \alpha E(X_1) - E(X_2))(X_3 - E(X_3))) \\ &= E(\alpha(X_1 - E(X_1))(X_3 - E(X_3)) + (X_2 - E(X_2))(X_3 - E(X_3))) \\ &= \alpha Cov(X_1, X_3) + Cov(X_2, X_3). \end{aligned}$$

□

Remarque. Cov n'est cependant pas un produit scalaire sur $L^2(\Omega, P)$, car elle n'est pas définie positive : en effet, si $X \in L^2(\Omega, P)$, d'après un exercice précédent, $Cov(X, X) = 0 \iff E((X - E(X))^2) = 0 \iff P((X - E(X))^2 = 0) = 1$,
donc $Cov(X, X) = 0 \iff P(X = E(X)) = 1$.
Ainsi, $Cov(X, X) = 0$ si et seulement si X est presque sûrement constante.
Cov est donc presque un produit scalaire, au caractère "définie" près. Même si ce n'est pas au programme, il sera utile dans ce chapitre de savoir que beaucoup de propriétés propres aux produits scalaires se généralisent à cette situation. Cela donne une interprétation géométrique à ces propriétés.

Définition. Si $X \in L^2(\Omega, P)$, on appelle variance de X la quantité
$\boxed{Var(X) = E[(X - E(X))^2]}$, qui représente l'écart quadratique moyen entre X et son espérance.
On appelle écart type de X la quantité $\sigma(X) = \sqrt{Var(X)}$.

Remarque.
◇ La variance de X mesure donc la dispersion de la variable aléatoire X autour de sa valeur moyenne.
◇ L'écart-type de X est une "pseudo-norme" euclidienne de X.
◇ $Var(X) = 0$ si et seulement si X est presque sûrement constante.
◇ Pour tout $X \in L^1(\Omega, P)$, $E(X - E(X)) = 0$, par linéarité de l'espérance.

Définition. Si X est une variable aléatoire réelle, on dit qu'elle est réduite si et seulement si $X \in L^2(\Omega, P)$ et $Var(X) = 1$.

Propriété. Formule de Koenig-Huygens :
Si $X \in L^2(\Omega, P)$, alors $\boxed{Var(X) = E(X^2) - E(X)^2}$.
Si $X_1, X_2 \in L^2(\Omega, P)$, alors $Cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$: en particulier, si deux variables aléatoires de $L^2(\Omega, P)$ sont indépendantes, elles sont orthogonales au sens de Cov (la réciproque est fausse).

Démonstration.
D'après la linéarité de l'espérance,
$Var(X) = E(X^2 - 2E(X)X + E(X)^2) = E(X^2) - 2E(X)E(X) + E(X)^2$. □

Propriété. Pour $a, b \in \mathbb{R}$ et $X \in L^2(\Omega, P)$, $Var(aX + b) = a^2Var(X)$.

Démonstration.
$Var(aX + b) = E((aX + b - E(aX + b))^2) = E((aX - aE(X))^2) = a^2Var(X)$. □

Propriété. Si $X \in L^2(\Omega, P)$ avec $\sigma(X) \neq 0$, alors $\dfrac{X - E(X)}{\sigma(X)}$ est une variable aléatoire centrée et réduite.

Propriété.
◇ Si $X_1, X_2 \in L^2(\Omega, P)$, $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$.
◇ Si $X_1, \dots, X_k \in L^2(\Omega, P)$, $Var(X_1 + \cdots + X_k) = \displaystyle\sum_{i=1}^{k} Var(X_i) + 2 \sum_{1 \leq i < j \leq k} Cov(X_i, X_j)$.

◇ Si $X_1, \dots, X_k$ sont k variables aléatoires de $L^2(\Omega, P)$ que l'on suppose **deux à deux indépendantes**, alors $Var(X_1 + \cdots + X_k) = Var(X_1) + \cdots + Var(X_k)$.

Démonstration.
$Var(X_1+\cdots+X_k) = Cov(X_1+\cdots+X_k, X_1+\cdots+X_k)$. Utilisons d'abord la linéarité de Cov selon son premier argument : $Var(X_1+\cdots+X_k) = \sum_{i=1}^{k} Cov(X_i, X_1+\cdots+X_k))$, puis utilisons la linéarité de Cov selon son deuxième argument :

$$Var(X_1 + \cdots + X_k) = \sum_{i=1}^{k}\sum_{j=1}^{k} Cov(X_i, X_j) = \sum_{i=1}^{k} Var(X_i) + 2\sum_{1\le i<j\le k} Cov(X_i, X_j). \ \square$$

Propriété. Inégalité de Cauchy-Schwarz :
pour tout $X, Y \in L^2(\Omega, P)$, $\boxed{E(XY)^2 \le E(X^2)E(Y^2)}$.
Il y a égalité si et seulement si il existe $\alpha, \beta \in \mathbb{R}$ tel que $(\alpha, \beta) \neq (0, 0)$ et $\alpha X + \beta Y$ est presque sûrement nulle.

Démonstration.
◇ *Première méthode :*
Pour tout $a \in \mathbb{R}$, $0 \le E((aX + Y)^2) = a^2E(X^2) + 2aE(XY) + E(Y^2)$.
Premier cas : Supposons que $E(X^2) \neq 0$.
Alors $a \longmapsto a^2E(X^2) + 2aE(XY) + E(Y^2)$ est un polynôme de degré 2 toujours positif, donc en prenant son discriminant, $4E(XY)^2 - 4E(Y^2)E(X^2) \le 0$, ce qui prouve l'inégalité de Cauchy-Schwarz.
Si c'est une égalité, alors $a \longmapsto a^2E(X^2)+2aE(XY)+E(Y^2)$ possède une racine réelle double $\alpha \in \mathbb{R}$, auquel cas $E((\alpha X + Y)^2) = 0$, donc (cf exercice page 30) $\alpha X + Y$ est presque sûrement nulle.
Réciproquement, s'il existe $\alpha, \beta \in \mathbb{R}$ tel que $(\alpha, \beta) \neq (0, 0)$ et $\alpha X + \beta Y$ est presque sûrement nulle, alors $E((\alpha X + \beta Y)^2) = 0$, donc $\beta \neq 0$ (sinon $E(\alpha^2X^2) = 0$ mais $E(X^2) \neq 0$, donc $\alpha = 0$ et $(\alpha, \beta) = (0, 0)$) et $\frac{\alpha}{\beta}$ est une racine réelle de
$a \longmapsto a^2E(X^2)+2aE(XY)+E(Y^2)$, donc l'inégalité de Cauchy-Schwarz est une égalité.
Deuxième cas : On suppose que $E(X^2) = 0$. Alors $X = 0$ presque sûrement, donc $E(XY)^2 = 0 = E(X^2)E(Y^2)$: on est dans le cas d'égalité.
On a bien, avec $\alpha = 1$ et $\beta = 0$: $(\alpha, \beta) \neq (0, 0)$ et $\alpha X + \beta Y = X$ est presque sûrement nulle.
◇ *Seconde méthode (seulement son ébauche)* : Sur $L^2(\Omega, P)$, on définit la relation binaire R en convenant que $X \ R \ Y$ si et seulement si $X = Y$ presque sûrement.
R est une relation d'équivalence. De plus, $X \ R \ Y \iff Y - X \in E$ où E est l'ensemble des variables aléatoires presque sûrement nulles. E est un sous-espace vectoriel de $L^2(\Omega, P)$, donc on peut naturellement structurer $L^2(\Omega, P)/R = L^2(\Omega, P)/E$ comme un $\mathbb{R}$-espace vectoriel , que l'on notera $\mathcal{L}^2(\Omega, P)$. Lorsque $\overline{X}, \overline{Y} \in \mathcal{L}^2(\Omega, P)$, on pose $\langle \overline{X}, \overline{Y} \rangle = E(XY)$. On vérifie que $\langle ., . \rangle$ est un produit scalaire sur $\mathcal{L}^2(\Omega, P)$. On en déduit l'inégalité de Cauchy-Schwarz dans $L^2(\Omega, P)$ avec cas d'égalité à partir de l'inégalité de Cauchy-Schwarz dans $\mathcal{L}^2(\Omega, P)$. $\square$

Corollaire. Inégalité de Cauchy-Schwarz :

pour tout $X, Y \in L^2(\Omega, P)$, $\boxed{Cov(X,Y)^2 \leq Var(X)Var(Y)}$.
Il y a égalité si et seulement si il existe $\alpha, \beta \in \mathbb{R}$ tel que $(\alpha, \beta) \neq (0,0)$ et $\alpha X + \beta Y$ est presque sûrement constante.

Démonstration.
On applique l'inégalité de Cauchy-Schwarz à $X - E(X)$ et $Y - E(Y)$. □

Définition. (hors programme) :
Soient $X, Y \in L^2(\Omega, P)$ telles que $Var(X)Var(Y) > 0$.
Le coefficient de corrélation linéaire entre X et Y est $Corr(X,Y) \triangleq \dfrac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$.

Propriété. $Corr(X,Y) \in [-1,1]$.

Démonstration.
C'est le corollaire de l'inégalité de Cauchy-Schwarz. □

Propriété. $|Corr(X,Y)| = 1$ si et seulement si il existe $(a,b) \in \mathbb{R}^2$ tel que $P(Y = aX + b) = 1$.

Démonstration.
C'est le cas d'égalité de Cauchy-Schwarz, dans le premier cas de sa démonstration, car par hypothèse, $E((X - E(X))^2) \neq 0$ (et $E((Y - E(Y))^2) \neq 0$). □

Remarque. Ce qui précède suggère que $Corr(X,Y)$ indique dans quelle mesure Y dépend **linéairement** de X, mais $Corr(X,Y)$ ne mesure pas les dépendances non linéaires (on peut avoir par exemple $Corr(X, X^2) = 0$).

Formule. Espérance et variance pour les lois au programme.
◇ **Loi de Bernoulli** de paramètre $p \in [0,1]$: $P(X=1) = p$ et $P(X=0) = 1-p$.
$\boxed{E(X) = p \text{ et } Var(X) = p(1-p)}$.
◇ **Loi binomiale** de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$: Pour tout $k \in \{0, \dots, n\}$, $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ (et $P(X=m) = 0$ pour $m \notin \{0, \dots, n\}$).
$\boxed{E(X) = np \text{ et } Var(X) = np(1-p)}$.
◇ **Loi géométrique** de paramètre $p \in]0,1[$: Pour tout $n \in \mathbb{N}^*$,
$P(X=n) = (1-p)^{n-1}p$ (et $P(X=0) = 0$).
$\boxed{E(X) = \dfrac{1}{p} \text{ et } Var(X) = \dfrac{1-p}{p^2}}$.
◇ **Loi de Poisson** de paramètre $\lambda \in \mathbb{R}_+^*$: pour tout $n \in \mathbb{N}$, $P(X=n) = e^{-\lambda}\dfrac{\lambda^n}{n!}$.
$\boxed{E(X) = \lambda = Var(X)}$.
◇ (hors programme) :
Loi hypergéométrique de paramètres $k, m, n \in \mathbb{N}^*$: $P(X=r) = \dfrac{\binom{m}{r}\binom{n}{k-r}}{\binom{m+n}{k}}$.
$\boxed{E(X) = kp \text{ et } Var(X) = kpq\dfrac{m+n-k}{m+n-1}, \text{ où } p = \dfrac{m}{m+n} \text{ et } q = 1-p}$.

Démonstration.
◇ Pour la loi de Bernoulli, $X^2 = X$ et $E(X) = p$,
donc $Var(X) = E(X^2) - E(X)^2 = E(X) - E(X)^2 = p(1-p)$.
◇ Pour la loi binomiale : $Var(X)$ ne dépend que de la loi de X (d'après la formule de transfert), donc on peut choisir pour le calcul n'importe quelle variable aléatoire X suivant une loi binomiale. Or c'est le cas lorsque $X = X_1 + \cdots + X_n$ où $X_1, \ldots, X_n$ sont des variables aléatoires de Bernoulli de même paramètre p, mutuellement indépendantes. Ainsi $Var(X) = Var(X_1) + \cdots + Var(X_n) = np(1-p)$.
◇ Pour la loi géométrique : comme pour le calcul de son espérance,
$E(X(X-1)) = \sum_{n=1}^{+\infty} n(n-1)p(1-p)^{n-1} = px\sum_{n=2}^{+\infty} n(n-1)x^{n-2}$ où $x = 1-p \in]0,1[$.
Pour tout $N \geq 2$, $\sum_{n=2}^{N} n(n-1)x^{n-2} = \frac{d^2}{dx^2}(\sum_{n=0}^{N} x^n) = \frac{d^2}{dx^2}(\frac{1-x^{N+1}}{1-x})$,
donc après calculs et passage à la limite, $\sum_{n=2}^{+\infty} n(n-1)x^{n-2} = \frac{2}{(1-x)^3}$,
donc $E(X(X-1)) = p(1-p)\frac{2}{p^3} = \frac{2(1-p)}{p^2}$,
puis $E(X^2) = E(X(X-1)) + E(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}$.
D'après la formule de Koenig-Huygens,
$Var(X) = E(X^2) - E(X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$.
◇ Pour la loi de Poisson :
$E(X(X-1)) = \sum_{n=0}^{+\infty} n(n-1)\frac{\lambda^n}{n!}e^{-\lambda} = \sum_{n=2}^{+\infty} \lambda^2 \frac{\lambda^{n-2}}{(n-2)!}e^{-\lambda} = \lambda^2$,
donc $Var(X) = E(X^2) - E(X)^2 = E(X(X-1)) + E(X) - E(X)^2 = \lambda$. □

6 Propriétés de convergence

Formule. Inégalité de Bienaymé-Tchebychev : Soit X une variable aléatoire réelle. Alors, pour tout $\varepsilon > 0$, $\boxed{P(|X - E(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}}$.

Démonstration.
Posons $Y = (X - E(X))^2$: c'est une variable aléatoire réelle positive à laquelle on peut appliquer l'inégalité de Markov.
Ainsi $P(|X - E(X)| \geq \varepsilon) = P((X - E(X))^2 \geq \varepsilon^2) \leq \frac{E(Y)}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}$. □

Définition. (hors programme) Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires et soit X une variable aléatoire . On dit que X_n converge vers X en probabilité si et seulement si pour tout $\varepsilon > 0$, $P(|X_n - X| \geq \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0$.

Remarque. (Admis) La convergence en proba implique la convergence en loi, mais la réciproque est fausse.

Définition. On écrit qu'une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ est i.i.d (pour indépendantes et identiquement distribuées) si et seulement si elles sont mutuellement indépendantes et si elles ont toutes la même loi.

Théorème. Loi faible des grands nombres :
Soit (X_n) une suite de variables aléatoires dans $L^2(\Omega, P)$ que l'on suppose toutes de même loi et deux à deux indépendantes. Posons $\mu = E(X_n)$, qui est indépendante de n. Alors la suite de variables aléatoires

$\boxed{\frac{X_1 + \cdots + X_n}{n} \text{ converge en probabilité vers la variable aléatoire constante égale à } \mu}$,

ce qui signifie que, pour tout $\varepsilon > 0$, $P(|\frac{X_1 + \cdots + X_n}{n} - \mu| \geq \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0$.

Démonstration : Fixons $\varepsilon > 0$. Appliquons l'inégalité de Bienaymé-Tchebychev à la variable aléatoire $Y = \frac{X_1 + \cdots + X_n}{n}$: $P(|Y - E(Y)| \geq \varepsilon) \leq \frac{Var(Y)}{\varepsilon^2}$,
mais $E(Y) = \mu$ et $Var(Y) = \frac{1}{n^2} Var(X_1 + \cdot + X_n) = \frac{1}{n^2}(Var(X_1) + \cdot + Var(X_n))$, car les X_i sont deux à deux indépendantes.
Comme de plus elles ont même loi, pour tout i, $Var(X_i) = Var(X_1)$,
donc $P(|\frac{X_1 + \cdots + X_n}{n} - \mu| \geq \varepsilon) \leq \frac{Var(X_1)}{n\varepsilon^2} \underset{n\to+\infty}{\longrightarrow} 0$. □

Remarque. Ce qu'affirme la loi faible des grands nombres, c'est que pour une précision ε donnée, la probabilité que l'espérance et la moyenne empirique diffère de plus de ε peut être rendue aussi petite que l'on désire en considérant un échantillon suffisamment grand. En ce sens, elle justifie a posteriori l'axiomatique de la théorie de probabilités, en faisant le lien avec la notion intuitive de fréquence de réalisation d'un événement. En effet, considérons une expérience aléatoire décrite par un espace probabilisé $(\Omega, \mathcal{F}, P)$, que l'on répète N fois, de façon indépendante, afin d'obtenir une suite de résultats $(\omega_1, \omega_2, \ldots, \omega_N)$.
Cette suite de N résultats appartient à l'espace probabilisé $(\Omega^N, \mathcal{F}_N, P_N)$, où sans entrer dans tous les détails, $\mathcal{F}_N$ est une tribu contenant tous les $A_1 \times \cdots \times A_N$ avec $A_i \in \mathcal{F}$ pour tout i, et où $P_N(A_1 \times \cdots \times A_N) = P(A_1) \times \cdots \times P(A_N)$.
Alors, pour tout événement $A \in \mathcal{F}$, les variables aléatoires $Y_k(\omega'_1, \omega'_2, \ldots, \omega'_N) = 1_A(\omega'_k)$ sont indépendantes et ont toutes la même loi, avec $E(Y_k) = P(A)$.
Par conséquent, si l'on note $N(A) = \#\{k \in \{1, \ldots, N\}/\omega_k \in A\}$ le nombre d'expériences lors desquelles l'événement A est réalisé, on a $\frac{N(A)}{N} = \frac{1}{N}\sum_{k=1}^{N} Y_k$, donc en adaptant la démonstration de la loi faible des grands nombres,
pour tout $\varepsilon > 0$, $P_N(|\frac{N(A)}{N} - P(A)| \leq \varepsilon) \underset{N\to+\infty}{\longrightarrow} 1$.

Propriété. Soit (X_n) une suite i.i.d dans $L^2(\Omega, P)$.
Posons $\mu = E(X_n)$ et $\sigma^2 = Var(X_n)$, qui sont indépendantes de n.
Pour tout $n \in \mathbb{N}^*$, notons $S_n = X_1 + \cdots + X_n$ puis $\boxed{Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}}$.
Alors, pour tout $n \in \mathbb{N}$, $E(Z_n) = 0$ et $Var(Z_n) = 1$. On a aussi $Z_n = \dfrac{S_n - E(S_n)}{\sqrt{Var(S_n)}}$.

Démonstration.
C'est un calcul analogue à celui développé lors de la démonstration de la loi faible des grands nombres. □

Théorème de la limite centrée (admis) : Avec les notations précédentes,
$\boxed{Z_n \text{ converge en loi vers la loi normale standard}}$, ce qui signifie que, pour tout $x \in \mathbb{R}$,
$P(Z_n \leq x) \underset{n\to+\infty}{\longrightarrow} \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{u^2}{2}} du.$

Remarque. Informellement $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} \sim \mathcal{N}$ lorsque n est grand, donc $\dfrac{S_n}{n} - \mu \sim \dfrac{\sigma}{\sqrt{n}}\mathcal{N}$: c'est plus précis que la loi faible des grands nombres.