北航中法工程师学院
Centrale Pékin

Probabilités

École Centrale de Pékin

Cours de mathématiques du cycle préparatoire

9 mars 2022

Table des matières

Avant-propos

Vous trouverez au fil de ce cours différents symboles :

— Le symbole “☡”, situé dans la marge, signifie que le point correspondant est un point délicat (il s'agit d'un *virage dangereux*).
— Le symbole “□” est un marqueur signifiant la fin d'une démonstration.
— ☡ Ce cours peut comporter des fautes de frappe, des coquilles, voire des erreurs d'argumentation. Ainsi, il faut toujours être vigilant lorsque vous suivez et que vous travaillez ce cours. Vérifier que les exemples sont justes et que les preuves n'ont pas de fautes est un exercice très utile (et indispensable) en mathématiques pour comprendre les notions et comprendre leurs utilisations.

Vous trouverez aussi des notations mathématiques :

Ω	un ensemble
ω	les éléments de Ω
$\mathcal{P}(\Omega)$	l'ensemble des parties de Ω
$\mathcal{A}$	une sigma-algèbre de Ω
A	un élément de la sigma-algèbre $\mathcal{A}$, un ”événement”.
$\mathcal{B}(\mathbb{R})$	la sigma-algèbre des boréliens de $\mathbb{R}$
$\mathbb{P}$	une mesure de probabilité sur $(\Omega, \mathcal{A})$
X	une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$
$\mathbb{P}(X \in A)$	la probabilité de l'événement ”X appartient à A”
	On a $\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega, \text{ t.q. } X(\omega) \in A\})$.
$\mathbb{E}(X)$	l'espérance d'une v.a. réelle et discrète X
	On a $\mathbb{E}(X) = \sum_{x \in \mathbb{R}} x\mathbb{P}(X = x) = \sum_{x \in X(\Omega)} x\mathbb{P}(X^{-1}(x))$ (si elle existe).
$L^1(\Omega, \mathcal{A}, \mathbb{P})$	l'ensemble des v.a. réelles et discrètes X, sur Ω, qui sont intégrables ($\mathbb{E}(\vert X\vert) < +\infty$)
$L^2(\Omega, \mathcal{A}, \mathbb{P})$	l'ensemble des v.a. réelles et discrètes, sur Ω, telles que X^2 est intégrable
$Var(X)$	la variance d'une v.a. réelle et discrète X de carré intégrable
σ_X	l'écart-type d'une v.a.r. et discrète X de carré intégrable
	On a $\sigma_X^2 = Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.
$Cov(X,Y)$	la covariance entre deux v.a.r. et discrètes X, Y
	On a $Cov(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$.

Chapitre 1 Dénombrement, sommabilité

Table des matières du chapitre

1.1 L'ensemble $\mathcal{P}(\Omega)$

On revoit rapidement dans ce chapitre les éléments de théorie des ensembles et de dénombrement que nous utiliserons par la suite de manière répétée.
On suppose dans toute la suite que Ω est un ensemble.

1.1.1 Ensemble des parties, fonctions indicatrices

Un **sous-ensemble** ou une **partie** de Ω est un ensemble dont tous les éléments sont dans Ω.

Axiome *Soit Ω un ensemble et P une propriété sur les éléments de Ω. Alors la collection des éléments de Ω qui vérifient la propriété P forme un ensemble. On note cet ensemble :*

$$\{x \in \Omega \text{ tq } P(x)\} \quad \text{ou} \quad \{x \in \Omega \mid P(x)\} \quad \text{ou} \quad \{x \in \Omega \ , \ P(x)\}.$$

Remarque 1 — *Cela vous montre comment écrire correctement un ensemble :*
si $A \subset \Omega$, on définit $\overline{A}$ (ou A^C), le ***complémentaire de** A **dans** Ω comme*

$$\overline{A} = \{\omega \in \Omega | \ \omega \notin A\}.$$

Ou encore, pour $w \in \Omega$, le ***singleton*** *$\{\omega\}$ est $\omega = \{\alpha \in \Omega \mid \alpha = \omega\}$.*
Le complémentaire et le singleton sont bien des parties de Ω.

Axiome *La collection des parties d'un ensemble Ω est encore un ensemble noté $\mathcal{P}(\Omega)$:*

$$A \in \mathcal{P}(\Omega) \iff A \subset \Omega.$$

$\mathcal{P}(\Omega)$ est donc l'ensemble de toutes les parties de Ω.

Exemple 2 —

1. *Si $\Omega = \emptyset$, alors $\mathcal{P}(\emptyset) = \{\emptyset\}$ et donc a un élément. Et $\mathcal{P}(\{\emptyset\}) =$?*
   On peut voir les ensembles comme des boîtes.
   L'ensemble vide est une boîte vide. Et $\{\emptyset\}$ est une boîte qui contient une boîte vide !
2. *Si $\Omega = \{1, 2, 3\}$, alors*
$$\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Remarque 3 — *Pour tout ensemble Ω, on a toujours $\emptyset \in \mathcal{P}(\Omega)$ et $\Omega \in \mathcal{P}(\Omega)$.*

Définition 4
Soit I un ensemble et $(\Omega_i)_{i\in I}$ une famille d'ensembles.
*Le **produit cartésien** de la famille $(\Omega_i)_{i\in I}$ est l'ensemble $\prod_{i\in I} \Omega_i$ des familles $(\omega_i)_{i\in I}$, avec $\omega_i \in \Omega_i$.*

Exemple 5 — *Par exemple, $(v_n)_{n\in\mathbb{N}}$ est un élément de $\Omega^{\mathbb{N}}$.*

Remarque 6 — *Sur $\mathcal{P}(E)$ on définit le complémentaire d'une partie, ainsi que l'union et l'intersection d'une famille quelconque de parties. On a par exemple*
$$\Omega = \bigcup_{\omega\in\Omega} \{w\}.$$
On peut montrer que les unions et les intersections sont associatives et commutatives pour une famille $(A_i)_{i\in I}$ de parties de Ω.
La distributivité entre l'union et l'intersection est valable pour des familles quelconques de parties. En particulier, on a
$$A \cap \left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} (A \cap A_i) \text{ et } A \cup \left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} (A \cup A_i).$$
On a aussi
$$\bigcup_{i\in I} \left(\bigcap_{j\in J} A_{i,j}\right) \subset \bigcap_{j\in J} \left(\bigcup_{i\in I} A_{i,j}\right).$$
En effet, si $x \in \bigcup_{i\in I} \left(\bigcap_{j\in J} A_{i,j}\right)$, alors il existe $i_0 \in I$ tel que $\forall j \in J$, $x \in A_{i_0,j}$.
Donc $\forall j \in J$ on a $x \in \bigcup_{i\in I} A_{i,j}$ (l'indice i_0 convient).
Finalement, on a prouvé $x \in \bigcap_{j\in J} \left(\bigcup_{i\in I} A_{i,j}\right)$, ce qui termine la preuve.
Mais a priori, il n'y a pas égalité.

Exemple 7 — *Soit $A_{i,j} = [j + i - 1, i + j[$. On a*
$$\bigcup_{i\in\mathbb{Z}} \left(\bigcap_{j\in\mathbb{N}} A_{i,j}\right) = \bigcup_{i\in\mathbb{Z}} \left(\bigcap_{j\in\mathbb{N}} [j + i - 1, i + j[\right) = \emptyset$$
et
$$\bigcap_{j\in\mathbb{N}} \left(\bigcup_{i\in\mathbb{Z}} A_{i,j}\right) = \bigcap_{j\in\mathbb{N}} \left(\bigcup_{i\in\mathbb{Z}} [j + i - 1, i + j[\right) = \mathbb{R}$$

§ 1. *Fonction indicatrice d'un ensemble*

Nous allons définir sur un ensemble particulier (un espace probabilisé) des fonctions particulières (les variables aléatoires). Avant de définir ces objets, les fonctions indicatrices en sont un exemple fondamental : si A est une partie de Ω, on détermine si $\omega \in \Omega$ appartient ou pas à A. Les fonctions indicatrices permettent de définir des opérations ensemblistes (union, intersection,...) en terme algébrique (produit, somme...)

Définition 8 (**Fonction indicatrice**)
*Soit $A \subset \Omega$. On appelle **fonction indicatrice** de A, notée $\mathbb{1}_A$ (ou χ_A), la fonction $\mathbb{1}_A : \Omega \to \mathbb{R}$ définie par*

$$\mathbb{1}_A(x) = \begin{cases} 1 \text{ si } x \in A \\ \\ 0 \text{ sinon.} \end{cases}$$

Proposition 9
Soient A, $B \in \mathcal{P}(\Omega)$. On a

1. $A \subset B$ *si et seulement si* $\mathbb{1}_A \leq \mathbb{1}_B$.
2. $\mathbb{1}_{\overline{A}} = 1 - \mathbb{1}_A$.
3. $\mathbb{1}_{A \cap B} = \mathbb{1}_A . \mathbb{1}_B$.
4. $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A . \mathbb{1}_B$

Exemple 10 —

1. *Écrire les fonctions indicatrices de $A \setminus B$ et de $A \Delta B$ en fonction de celles de A et B.*
2. *Soit une famille $(A_i)_{i \in [\![1,n]\!]}$ de parties non vides de Ω.*

$$\mathbb{1}_{\bigcap\limits_{i \in [\![1,n]\!]} A_i} = \prod_{i \in [\![1,n]\!]} \mathbb{1}_{A_i} \text{ et } \mathbb{1}_{\bigcup\limits_{i \in [\![1,n]\!]} A_i} = 1 - \prod_{i \in I} (1 - \mathbb{1}_{A_i})$$

Proposition 11
Soit $n \in \mathbb{N}^*$. Une famille $(A_i)_{i \in [\![1,n]\!]}$ de parties non vides de Ω est une partition de Ω si et seulement si

$$\mathbb{1}_\Omega = \sum_{i=1}^{n} \mathbb{1}_{A_i}.$$

Preuve — On a les équivalences

1. Soit $i \neq j$. Alors $x \in A_i \cap A_j$ ssi
$$\sum_{i \in I} \mathbb{1}_{A_i}(x) \geq 2.$$
2. De plus, $x \in \bigcup_{i \in j} A_i$ ssi il existe $i \in J$ tel que $x \in A_i$ ssi $\sum\limits_{i \in I} \mathbb{1}_{A_i}(x) \geq 1$.

Ceci montre que Ω est l'union disjointe de la famille $(A_i)_{i \in I}$ ssi $\mathbb{1}_\Omega = \sum\limits_{i \in I} \mathbb{1}_{A_i}$.

□

Corollaire 12
Une famille $(A_i)_{i \in [\![1,n]\!]}$ de parties non vides de Ω est une partition de Ω si et seulement si

$$\forall A \in \mathcal{P}(\Omega), \ \mathbb{1}_A = \sum_{i \in I} \mathbb{1}_{A_i \cap A}.$$

1.2 Cardinal d'un ensemble

1.2.1 Définition

Définition 13
*On dit que deux ensembles Ω et Ω' ont **même cardinal**, et on note $|\Omega| = |\Omega'|$, s'il existe une bijection de Ω dans Ω'.*
*Si Ω est en bijection avec $\{1, \ldots, n\}$, on dit que Ω est un ensemble fini de **cardinal** n et l'on note $|\Omega| = n$ ou $card(\Omega) = n$.*
Par convention, $card(\emptyset) = 0$.
Un ensemble de cardinal infini est un ensemble qui n'est pas de cardinal fini. On écrira $|\Omega| = \infty$.

Remarque 14 — *Cela nous donne une première technique majeure pour calculer le cardinal Ω d'un ensemble. On définit une bijection g entre Ω et un ensemble Ω' dont on connaît le cardinal.*

Exemple 15 — *Je cherche le cardinal de Ω l'ensemble des suites de 4 chiffres (u_0, u_1, u_2, u_3) à valeurs dans $[\![0, 9]\!]$ telles que $u_1 - u_0 = 2$, $u_2 - u_1 = 1$ et $u_3 - u_2 = 3$.*
On définit l'application

$$\varphi : [\![0, 3]\!] \to \Omega,\ k \mapsto (k, k+2, k+3, k+6).$$

On a bien $\varphi(k) \in \Omega$. φ est injective car $\varphi(k) = \varphi(k')$ implique sur le premier terme $k = k'$. Et φ est surjective car si $(u_0, u_1, u_2, u_3) \in \Omega$, alors $u_3 = u_0 + 6 \leq 9$ implique $u_0 \leq 3$ et $\varphi(u_0) = (u_0, u_1, u_2, u_3)$.
On en déduit que $|\Omega| = 4$.

Proposition 16
Soit Ω un ensemble fini de cardinal n.
Si $F \subset \Omega$, alors F est fini et card$(F) \leq \mathrm{Card}\,(\Omega)$.
De plus, $\mathrm{Card}\,(\Omega) = \mathrm{Card}\,(F)$ *si et seulement si $F = \Omega$.*

Preuve — Si Ω est de cardinal n, on pose $\Omega = \{\omega_1, \cdots, \omega_n\}$. On construit par récurrence une suite (f_n)

1. Si $F \neq \emptyset$, alors $f_0 = \omega_{\inf\{k\ |w_k \in F\}}$
2. Si f_{k-1} est construit et si $F \setminus \{f_0, \cdots, f_{k-1}\} \neq \emptyset$, alors $f_k = \omega_{\inf\{k\ |w_k \in F \setminus \{f_0, \cdots, f_{k-1}\}\}}$
3. Si f_{k-1} est construit et si $F \setminus \{f_0, \cdots, f_{k-1}\} = \emptyset$, alors $F = \{f_0, \cdots, f_{k-1}\}$ et $|F| = k$. Et on s'arrête.

Tous les éléments de Ω sont parcourus ssi $n = k$ et $F = \Omega$. □

Remarque 17 — *On en déduit alors que $\mathbb{N}$ n'est pas un ensemble fini. Par exemple, $\mathbb{N}$ est en bijection avec les nombres pairs en prenant $k \mapsto 2k$, donc $\mathbb{N}$ a le même cardinal qu'un sous-ensemble strict, ce qui n'est pas possible pour un ensemble fini.*

Définition 18
On dit qu'un ensemble Ω est

- ***dénombrable** s'il existe une bijection de Ω dans $\mathbb{N}$.*
- *au plus dénombrable si Ω est fini ou dénombrable.*

Proposition 19
Soient Ω et Ω' deux ensembles.

1. *S'il existe une surjection de Ω dans Ω', alors on a $|\Omega| \geq |\Omega'|$.*
2. *S'il existe une injection de Ω dans Ω', alors on a $|\Omega| \leq |\Omega'|$.*

Proposition 20
Soit Ω un ensemble. Il n'existe pas d'injection de $\mathcal{P}(\Omega)$ dans Ω.

Preuve — On raisonne par l'absurde : supposons qu'il existe une injection $f : \mathcal{P}(\Omega) \to \Omega$, $A \mapsto f(A)$.
Alors $\mathcal{A} = \{A \in \mathcal{P}(\Omega) \mid f(A) \notin A\}$ est un sous-ensemble de $\mathcal{P}(\Omega)$ et $B = f(\mathcal{A})$ est un sous-ensemble de Ω, c'est-à-dire un élément de $\mathcal{P}(\Omega)$. On peut donc calculer l'image de B par f. On a deux possibilités

1. Soit $f(B) \notin B$, donc $B \in \mathcal{A}$. On en déduit $f(B) \in f(\mathcal{A}) = B$, ce qui est contradictoire.
2. Soit $f(B) \in B$, donc $B \notin \mathcal{A}$. De plus, comme $f(\mathcal{A}) = B$, il existe $C \in \mathcal{A}$ tel que $f(B) = f(C)$, mais cela contredit l'injectivité de f.

On en déduit qu'il n'existe pas d'injection de $\mathcal{P}(E)$ dans E. □

Remarque 21 — *La proposition ici est pour Ω de cardinal quelconque. On peut donc l'appliquer à $\mathbb{N}$: il n'existe pas de bijection de $\mathbb{N}$ dans $\mathcal{P}(\mathbb{N})$.*
Tous les ensembles de cardinal infini ne sont donc pas nécessairement dénombrables.
Nous allons souvent étudier les probabilités sur des ensembles Ω dénombrables et nous devrons étudier $\mathcal{P}(\Omega)$, qui n'est plus dénombrable.
Par contre les opérations d'unions, produits d'ensembles dénombrables restent dénombrables, comme nous le rappelons ci-dessous.

1.2.2 Cardinal d'une union, d'un produit,...

Proposition 22
Soit Ω un ensemble et A une partie finie de Ω. Alors

$$\operatorname{card} A = \sum_{x\in\Omega} \mathbb{1}_A(x)$$

Proposition 23
Si A et B sont deux sous-ensembles finis de Ω, alors $A \cup B$ est fini et

$$\operatorname{card}(A\cup B) = \operatorname{Card}(A) + \operatorname{Card}(B) - \operatorname{Card}(A\cap B).$$

Corollaire 24
Soit Ω un ensemble fini et $(A_i)_{i\in I}$ une partition de Ω. Alors on a

$$|\Omega| = \sum_{i\in I} |A_i|.$$

Preuve — Les A_i sont deux à deux distincts non vides, donc $|I| \leq |\Omega|$: pour tout i, on choisit un élément $a_i \in A_i$ et l'application $i \mapsto a_i$ est une injection.
Puis on procède par récurrence sur $|I|$, l'étape $|I| = 1$ implique $A_1 = \Omega$ et l'hérédité résulte immédiatement de la proposition précédente. □

Remarque 25 —
• *Cela nous donne une autre méthode fondamentale pour calculer le cardinal d'un ensemble Ω.*
On cherche une partition de $(A_i)_{i\in I}$ de Ω telle que tous les A_i sont de cardinal connu. La proposition ci-dessus permet ainsi de calculer le cardinal de Ω.
En particulier, pour $f : \Omega \to \Omega'$ une fonction surjective, alors $\left(f^{-1}(\omega')\right)_{\omega'\in\Omega'}$ est une partition de Ω.
Ainsi, si Ω est un ensemble fini, on a $|\Omega| = \sum_{\omega'\in\Omega'} \left|f^{-1}(\omega')\right|$.

Proposition 26
Soient Ω et Ω' deux ensembles finis. Alors $\Omega \times \Omega'$ est fini et

$$\operatorname{card}(\Omega\times\Omega') = \operatorname{Card}(\Omega)\operatorname{Card}(\Omega').$$

Preuve — Il suffit d'écrire $\Omega \times \Omega'$ comme l'union disjointe $\cup_{y\in\Omega'}\Omega \times \{y\}$. □

Proposition 27
Soient Ω et Ω' deux ensembles finis de cardinaux n et p.
Alors l'ensemble des applications de Ω dans Ω', Ω'^{Ω}, est un ensemble fini de cardinal p^n.

Preuve — L'application qui à $f : \Omega \to \Omega'$ associe le n-uplet $(f(x_1), \dots, f(x_n))$ est une bijection. Or Ω'^n est de cardinal p^n d'après la proposition 26, d'où le résultat. □

Proposition 28
Soit Ω un ensemble fini de cardinal n.
Alors $\mathcal{P}(\Omega)$ est fini, de cardinal 2^n.

Preuve — L'application $\psi : \mathcal{P}(\Omega) \to \{0,1\}^{\Omega}$, $A \mapsto \mathbb{1}_A$ est bijective. □

1.3 Ensembles dénombrables

1.3.1 Caractérisation des ensembles dénombrables

Remarque 29 — *Soit Ω un ensemble dénombrable et $\varphi : \mathbb{N} \to \Omega$ une bijection.*
On pose $\varphi(n) = \omega_n$. Alors, les éléments de Ω peuvent être indexés par $\mathbb{N}$: $\Omega = \{\omega_n\}_{n\in\mathbb{N}}$.

*On dit que φ est une **énumération** de Ω.*

PROPOSITION 30
Soit Ω un ensemble dénombrable.
Alors toute partie A de Ω est finie ou dénombrable.

COROLLAIRE 31
Toute partie de $\mathbb{N}$ est soit de cardinal fini, soit dénombrable.

1.3.2 Opérations sur les ensembles dénombrables

PROPOSITION 32
Soient Ω et Ω' deux ensembles dénombrables. Alors on a

1. *$\Omega \cup \Omega'$ est dénombrable.*
2. *$\Omega \times \Omega'$ est dénombrable.*

Preuve — On écrit $\Omega = (\omega_n)_{n\in\mathbb{N}}$.

1. On pose $\Omega'' = \Omega' \setminus \Omega$ et $\Omega \cup \Omega' = \Omega \cup \Omega''$.
Si $|\Omega''| = p$, on numérote les éléments de Ω' par ω''_0, ..., ω''_{p-1} et ainsi $\varphi : \mathbb{N} \to \Omega \cup \Omega'$ définie par
$$\varphi(k) = \begin{cases} \omega''_k \text{ si } k \in [\![0, p-1]\!] \\ \\ \omega_{k-p} \text{ si } k \geq p \end{cases}$$
est bijective.
Si Ω'' est dénombrable, on écrit $\Omega'' = (\omega''_n)_{n\in\mathbb{N}}$ et l'application $\varphi : \mathbb{N} \to \Omega \cup \Omega'$ définie par
$$\varphi(k) = \begin{cases} \omega_{k/2} \text{ si } k \text{ pair} \\ \\ \omega''_{(k-1)/2} \text{ sinon} \end{cases}$$
est bijective.
2. Si $\varphi : \Omega \to \mathbb{N}$ et $\psi : \Omega' \to \mathbb{N}$, alors $\varphi \times \psi : \Omega \times \Omega' \to \mathbb{N} \times \mathbb{N}$ est bijective.
L'application $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ définie come suit est bijective :
$$f : \begin{cases} \mathbb{N}^2 & \to & \mathbb{N} \\ (p,q) & \mapsto & 2^p(2q+1) - 1 \end{cases}.$$
L'application φ est bijective : d'après la décomposition en facteur premier, tout nombre $n \in \mathbb{N}^*$ se décompose de manière unique $2^p m$ avec m impair et donc $q = \dfrac{m-1}{2}$. Donc il existe un unique couple (p, q) tel que $\varphi(p, q) = n - 1$.

□

REMARQUE 33 —
• *La première partie de la preuve du 1/ correspond à l'histoire suivante : Un hôtel possède une infinité dénombrable de chambres, et il est complet. Un car arrive avec 60 passagers.*
La personne de la réception répond qu'il pas de problèm. On décale tout le monde de soixante chambres, et les 60 premières seront libres.
• *La seconde partie de la preuve correspond au cas où chacun client se décale de sorte que seules les chambres paires soient occupées, ce qui laisse donc un nombre infini de chambres libres !*

PROPOSITION 34
Soit Ω un ensemble. Pour tout $n \geq 0$ soit A_n une partie de Ω dénombrable.
Alors, $\cup_{n\geq 0} A_n$ est dénombrable.
Une union dénombrable d'ensembles dénombrables est dénombrable.

REMARQUE 35 — *On déduit des résultats de ce chapitre que $\mathbb{Z}$, $\mathbb{Q}$ sont dénombrables.*
Ce n'est pas le cas de $\{0,1\}^{\mathbb{N}}$ (en bijection avec $\mathcal{P}(\mathbb{N})$), de $\mathbb{R}$, de $\mathbb{C}$, ni des intervalles $[a, b]$ avec $a < b$.

EXEMPLE 36 —

1. *Soit $f : I \to \mathbb{R}$ une fonction monotone (croissante ou décroissante). Alors l'ensemble des points de discontinuité de f est au plus dénombrable.*
   En effet, on suppose que f est croissante. Alors f est discontinue en a ssi $]\lim_{a^-} f, \lim_{a^+} f[$ est un intervalle non-vide. Un intervalle non-vide de $\mathbb{R}$ contient un rationnel. Comme f est croissante, les points de discontinuité a sont donc en bijection avec une partie $\mathbb{Q}$. D'où le résultat.
2. *L'ensemble des racines des polynômes à coefficients entiers est dénombrable.*
   L'ensemble $\mathbb{Z}_n[X]$ des polynômes à coefficients entiers de degré $\leq n$ est en bijection avec $\mathbb{Z}^{n+1}$, qui est dénombrable (produit fini d'ensembles dénombrables).
   Et on a $\mathbb{Z}[X] = \cup_{n \geq 0} \mathbb{Z}_n[X]$, réunion dénombrable d'ensembles dénombrables.
   Un nombre réel qui n'est pas racine d'un polynôme à coefficients entiers est appelé un nombre transcendant. Ainsi, la majorité des nombres réels sont transcendants, même on en connait très peu dans les faits.

1.4 Coefficients binomiaux

1.4.1 p-combinaisons

Définition 37
Soit Ω un ensemble de cardinal n. Soit $p \geq 0$.
On appelle p-combinaison de Ω toute partie de Ω de cardinal p.

Remarque 38 — *On peut montrer que le nombre de p-combinaisons d'un ensemble de cardinal n ne dépend que de p et de n.*

Définition 39
Soient $n \geq 0$, $p \in \mathbb{Z}$.
On note $\binom{n}{p}$ (ou C_n^p) le nombre de p-combinaisons d'un ensemble à n éléments.
*Ce nombre est appelé **coefficient binomial** (ou p **parmi** n).*
On convient que si $p < 0$ et si $p > n$, alors $\binom{n}{p} = 0$.

1.4.2 Formules sur les coefficients binomiaux

Proposition 40 (**Propriété du triangle de Pascal**)
Soient $n, p \in \mathbb{N}$. On a
$$\binom{n+1}{p} = \binom{n}{p-1} + \binom{n}{p}.$$

Preuve — Soit Ω de cardinal $n+1$ et $a \in \Omega$. Alors l'ensemble des p-combinaisons de Ω est égal à l'union disjointe des p combinaisons qui contiennent a et de celles qui ne le contiennent pas. Le premier ensemble a pour cardinal $\binom{n}{p-1}$ et le second de cardinal $\binom{n}{p}$, d'où la formule. □

Proposition 41
Soient $n, p \in \mathbb{N}$. Alors on a
$$\binom{n}{p} = \binom{n}{n-p}.$$

Preuve — Si $p > n$, l'égalité se résume à $0 = 0$. Sinon, soit Ω de cardinal n et $\varphi : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$, $A \mapsto \bar{A}$. Alors φ est involutive, donc bijective et induit donc une bijection des p-combinaisons avec les $n-p$-combinaisons, l'égalité est démontrée. □

Proposition 42

Soient $n, p \in \mathbb{N}$ avec $0 \leq p \leq n$. Alors on a

$$\binom{n}{p} = \frac{n!}{p!(n-p)!}$$

REMARQUE 43 — *On a ainsi* $\binom{n}{0} = \binom{n}{n} = 1$, $\binom{n}{1} = \binom{n}{n-1} = 1$, $\binom{n}{2} = \frac{n(n-1)}{2}$.

EXEMPLE 44 — *Soit $n \in \mathbb{N}^*$ fixé, on cherche $p \in \mathbb{N}$ tel que $\binom{n}{p}$ soit maximal. pour cela, on calcule pour $p \neq 0$*

$$\frac{\binom{n}{p}}{\binom{n}{p-1}} = \frac{n!}{(p)!(n-p)!}\frac{(p-1)!(n-p+1)!}{n!} = \frac{n+1-p}{p} = \frac{n+1}{p} - 1$$

Le quotient est décroissant en p, vaut n en 1, 0 en $n+1$ et vaut 1 ssi $n+1 = 2p$.
On en déduit que $\binom{n}{p}$ est maximal pour $p = \left[\frac{n}{2}\right]$ (si le rapport vaut 1 le maximum est atteint pour deux valeurs de p).

1.4.3 Arrangements

DÉFINITION 45 (Arrangements)
Soient $n, p \in \mathbb{N}^*$.
*On appelle **arrangement** à p éléments de $\{1, \ldots, n\}$ tout p-uplet $(a_1, \cdots, a_p)$ de $\{1, \ldots, n\}^p$ tel que les a_i soient deux à deux distincts.*

Au lieu de prendre seulement p éléments parmi n, un arrangement tient aussi compte de l'ordre des éléments choisis. (par ex $(2,3)$ et $(3,2)$ sont deux arrangements différents à 2 él. de $\{1,2,3\}$)

PROPOSITION 46
Soient $n, p \geq 1$.
Si $1 \leq p \leq n$, le nombre d'arrangements à p éléments de $\{1, \cdots, n\}$ vaut $A_p^n = \dfrac{n!}{(n-p)!} = p!\binom{n}{p}$.
Si $p > n$, le nombre d'arrangements à p éléments de $\{1, \cdots, n\}$ vaut $A_p^n = 0 = p!\binom{n}{p}$.

Preuve — Le nombre d'arrangements de p éléments dans I_n revient à la donnée d'un élément a_1 dans I_n, puis d'un élément $a_2 \in I_n \setminus \{a_1\}$, etc puis $a_p \in I_n \setminus \{a_1, \cdots, a_{p-1}\}$... et donc $A_p^n = n(n-1)\cdots(n-p+1)$, ce que nous voulions. Mais cette écriture, bien que convaincante, n'est pas entièrement rigoureuse. Reprenons le raisonnement. On note $\mathcal{A}$ l'ensemble des p-arrangements de $[\![1, n]\!]$.

1. Pour tout $i \in [\![0, n]\!]$, on pose A_i l'ensemble des arrangements qui commencent par i. Les A_i forment une partition de $\mathcal{A}$, donc $|\mathcal{A}| = \sum_{i=1}^{n} |A_i|$. Si σ_i est la permutation $(1\ i)$, $\varphi_i : A_1 \to A_i$, $(1, a_2, \cdots, a_p) \mapsto (\sigma(1), \sigma(a_2), \cdots, \sigma(a_p))$ est une bijection (donnez la réciproque !). Donc $|A_i| = |A_1|$ et $|\mathcal{A}| = n|A_1|$ car il y a n ensembles A_i.
2. Puis on pose, pour $i \neq j$, $A_{(i,j)}$ l'ensemble des permutations qui commencent par (i, j) : $(i, j, a_3, \cdots, a_n)$. On montre que tous les $A_{(i,j)}$ ont même cardinal et les $A_{(i,j)}$ pour $j \in [\![1, n]\!] \setminus \{i\}$ forment une partition de A_i. On en déduit que $|A_i| = (n-1)|A_{(1,2)}|$
3. Avec les notations précédentes, $|A_{(a_i, \cdots, a_r)}| = (n-r)|A_{(a_i, \cdots, a_{r+1})}|$ tant que $r \leq p-1$.

$\vdots$

4. Si $p = r$, il est clair que $|A_{(a_1, \cdots, a_p)}| = 1$ et on en déduit le résultat.

En probabilité, la formule des probabilités composées nous permettra de clarifier ce type de raisonnements. Mais il est préférable ici de faire, par exemple, une simple récurrence sur p pour n fixé en reprenant l'étape 1 ci-dessus.
Une autre méthode est de considérer l'application φ surjective de $\mathcal{A}_n^p$, l'ensemble des arrangements de p éléments de I_n, dans $\mathcal{C}_n^p$, l'ensemble des p-combinaison de I_n, définie par

$$\varphi : \mathcal{A}_n^p \to \mathcal{C}_n^p,\ (a_1, \cdots, a_p) \mapsto \{a_1, \cdots, a_p\}.$$

Pour toute p combinaison $\{a_1, \cdots, a_p\}$,

$$\varphi^{-1}(\{a_1, \cdots, a_p\}) = \{(a_{\sigma(1)}, \cdots, a_{\sigma(p)}),\ \sigma \in S_p\}$$

On en déduit que $|\varphi^{-1}(\{a_1, \cdots, a_p\})| = p!$. Or

$$|\mathcal{A}_n^p| = \sum_{\{a_1, \cdots, a_p\} \in \mathcal{C}_n^p} |\varphi^{-1}(\{a_1, \cdots, a_p\})| = p!\,|\mathcal{C}_n^p| = \frac{n!}{(n-p)!}.$$

□

1.5 Exemples de dénombrement

Exemple 47 — *Soit M un mot de n lettres, composé des lettres $a_1, \dots, a_l$, qui apparaissant respectivement $\alpha_1, \dots, \alpha_l$ fois. On a donc $n = \alpha_1 + \cdots + \alpha_l$.*
Alors, le nombre d'anagrammes (mots obtenus par permutation des lettres) de M, noté $N(n, \alpha_1, \cdots, \alpha_l)$, est égal à

$$N(n, \alpha_1, \cdots, \alpha_l) = \frac{n!}{\alpha_1! \cdots \alpha_l!}.$$

On démontre cela par récurrence sur $l \geq 1$.
Initialisation : *Pour $l = 1$, il n'y a qu'un seul anagramme du mot.*
Hérédité : *Supposons la proposition vraie pour $l - 1 \geq 1$.*
La lettre a_l doit apparaître α_l fois dans l'anagramme. Soit $\mathcal{C}_n^{\alpha_l}$, l'ensemble des α_l-combinaisons dans I_n. Soit $s \in \mathcal{C}_n^{\alpha_l}$. On note Ω_s l'ensemble des anagrammes tels que la position de la lettre a_l corresponde au choix s. Les ensembles Ω_s forment une partition de Ω.
Or, chaque Ω_s correspond aux anagrammes du mot d'origine privé de la lettre a_l et donc est de cardinal $N(n - \alpha_l, \alpha_1, \cdots, \alpha_{l-1})$. Comme $|\mathcal{C}_n^p| = \binom{n}{\alpha_l}$, on obtient

$$N(n, \alpha_1, \cdots, \alpha_l) = \binom{n}{\alpha_l} N(n - \alpha_l, \alpha_1, \cdots, \alpha_{l-1}) = \frac{n!}{\alpha_l!(n - \alpha_l)!} \times \frac{(n - \alpha_l)!}{\alpha_1! \cdots \alpha_{l-1}!} = \frac{n!}{\alpha_1! \cdots \alpha_l!}$$

ce qui termine la preuve.
• *Le nombre $N(n, \alpha_1, \cdots, \alpha_l)$ est appelé* ***coefficient multinomial****.*
Quand $n = 2$, on a $N(2, \alpha_1, \alpha_2) = \binom{\alpha_1 + \alpha_2}{\alpha_1}$, on retrouve les coefficients binomiaux.

Les coefficients binomiaux apparaissent dans le développement de $(a + b)^n$ (pour a, b des éléments d'un anneau commutatif).
On peut généraliser ce résultat avec les coefficients multinomiaux.

Proposition 48
Soient $p \geq 2$ et $a_1, \dots, a_p$ des éléments d'un anneau commutatif A. Alors, on a

$$(a_1 + \cdots + a_p)^n = \sum_{\alpha_1 + \cdots \alpha_p = n} \frac{n!}{\alpha_1! \cdots \alpha_p!} a_1^{\alpha_1} \cdots a_p^{\alpha_p}, \forall n \geq 0.$$

Preuve — On calcule le coefficient de $a_1^{\alpha_1} \cdots a_p^{\alpha_p}$ sachant que la puissance totale est n car on a n facteurs. Cela correspond à écrire un mot de n lettres avec l'alphabet $\{a_1, \cdots, a_p\}$. La proposition ci-dessus permet de conclure.

Par exemple

$$(a_1 + \cdots + a_p)^2 = a_1^2 + \cdots + a_p^2 + 2a_1a_2 + 2a_1a_3 + \cdots + 2a_{n-1}a_n.$$
$$(a_1 + a_2 + a_3)^3 = a_1^3 + a_2^3 + a_3^3 + 3a_1^2a_2 + 3a_1^2a_3 + 3a_2^2a_1 + 3a_2^2a_3 + 3a_3^2a_1 + 3a_3^2a_2 + 6a_1a_2a_3.$$

□

Exemple 49 — *Soit E un ensemble fini à n éléments, et soit $A \subset E$ de cardinal p $(0 \leq p \leq n)$. Dénombrer l'ensemble*

$$\Omega = \{(X, Y) \in \mathcal{P}(E)^2 \mid A \subset X \cap Y \text{ et } X \cup Y = E\}$$

Méthode 1 : *On simplifie d'abord le problème. Par hypothèse, A est inclus dans X et Y, en posant $X' = X \setminus A$ et $Y' = Y \setminus A$, on cherche tous les couples $(X', Y') \in \mathcal{P}(E \setminus A)^2$ tels que $X' \cup Y' = E' = E \setminus A$.*

$$\Omega' = \{(X', Y') \in \mathcal{P}(E')^2 \mid X' \cup Y' = E\}$$

On a une application

$$\varphi : \Omega \to \Omega', \ (X, Y) \mapsto (X \setminus A, Y \setminus A).$$

On vérifie facilement que φ est surjective et que $\varphi^{-1}(X', Y') = (X' \cup A, Y' \cup A)$. Donc $|\Omega| = |\Omega'|$. Pour calculer $|\Omega'|$, on pose Ω_k l'ensemble des couples $(X', Y') \in \Omega'$ tels que X' a k éléments dans E'. Les Ω_k forment une partition de Ω'.
Calculons $|\Omega_k|$: pour X' fixé, soit $\Omega_{k,X'}$ l'ensemble des couples $(X', Y') \in \Omega'$, c'est-à-dire ssi Y' contient

le complémentaire de X' dans E' : $Y' = \overline{X'} \cup Z$, $Z \in \mathcal{P}(X')$ qui est de cardinal 2^k.
Or les $\Omega_{k,X'}$ avec X' de cardinal k forment une partition de Ω_k et il y en a $\binom{n-p}{k}$.
On en déduit que

$$|\Omega_k| = \sum_{X' \in \mathcal{P}(E'),\ |X'|=k} |\Omega_{k,X'}| = \binom{n-p}{k} 2^k.$$

Enfin k varie entre 0 et $n-p$ donc

$$|\Omega| = |\Omega'| = \sum_{k=0}^{n-p} |\Omega_k| = \sum_{k=0}^{n-p} \binom{n-p}{k} 2^k = 3^{n-p}.$$

Méthode 2 : *La simplicité du résultat nous suggère que l'on a peut-être une méthode plus efficace qui traduit la remarque suivante : pour tout $a \in E \setminus A$, on a trois possibiltés*

1. *a appartient à X et pas à Y*
2. *a appartient à Y et pas à X*
3. *a appartient à X et Y*

On peut le rédiger de la manière suivante : posons $\psi : \Omega \to B$, $(X,Y) \mapsto (\mathbb{1}_X, \mathbb{1}_Y)$ avec

$$B = \left\{ f : E \to \{0,1\}^2 \mid \forall a \in A,\ f(a) = (1,1) \text{ et } \forall b \in E \setminus A, f(b) \in \{(0,1),(1,0),(1,1)\} \right\}.$$

On montre que ψ est bijective. Enfin, B est en bijection avec $\{(1,0),(0,1),(1,1)\}^{E\setminus A}$. Donc $|\Omega| = |B| = 3^{n-p}$ car $|E \setminus A| = n-p$.

Exemple 50 — *Ce dernier exemple est plus complexe, mais on peut le résoudre simplement si on "modélise" correctement l'ensemble Ω.*
Soient $1 \leq p \leq n$ et soit Ω l'ensemble des répartitions de n boules dans p trous, numérotés T_1, ..., T_p.
Faites un dessin pour bien comprendre.
On considère que

1. *si T_1 a k_1 boules blanches, on écrit $T_1 = \underbrace{1, \cdots, 1}_{k_1 \text{ fois}}$.*

$\vdots$

p. *si T_p a k_p boules blanches, on écrit $T_p = \underbrace{1, \cdots, 1}_{k_p \text{ fois}}$.*

Puis on définit l'application

$$\begin{array}{rccc} \varphi : & \Omega & \to & [\![0,1]\!]^{n+p-1} \\ & (T_1, \cdots, T_p) & \mapsto & (T_1, 0, T_2, 0, \cdots, 0, T_p) \end{array}$$

L'application φ est clairement injective.
Déterminons l'image de φ : nous avons introduit $p-1$ zéros.
De plus, soit $(a_1, \cdots, a_{n+p-1}) \in [\![0,1]\!]^{n+p-1}$ tel qu'il existe $1 \leq j_1 < j_2 < \cdots < j_{p-1} \leq n$ tels que $a_{j_1} = \cdots = a_{j_{p-1}} = 0$ et pour tout $i \in [\![0, n-1]\!]$, i différent j_1, ..., j_{p-1}, alors $a_i = 1$:

$$(a_1, \cdots, a_{n+p-1}) = (1, \cdots, 1, \underset{j_1}{0}, 1, \cdots, 1, \underset{j_2}{0}, 1 \cdots, 1, \underset{j_{p-1}}{0}, 1 \cdots, 1).$$

On pose

1. *$T_1 = \underbrace{1, \cdots, 1}_{j_1 - 1}$.*

$\vdots$

2. *$T_k = \underbrace{1, \cdots, 1}_{j_k - j_{k-1}}$ si $k \in [\![2, p-1]\!]$.*

$\vdots$

p. si $T_p = \underbrace{1, \cdots, 1}_{n-j_{p-1}}$.

Par construction, $\varphi(T_1, \cdots, T_P) = (a_1, \cdots, a_{n+p-1})$.
On en déduit que $\varphi(\Omega)$ *est en bijection avec l'ensemble des* $p-1$ *combinaisons parmi* n.
On obtient donc, $|\Omega| = \binom{n+p-1}{n} = \binom{n+p-1}{p-1}$.
Application : *Si* x_1, ..., x_p *sont des nombres complexes, combien existe-t-il de monômes* $x_1^{\alpha_1} \cdots x_p^{\alpha_p}$ *tels que* $\alpha_1 + \cdots + \alpha_p = n$ *?*

1.6 Familles sommables

Dans le cours d'Analyse 3, nous avons vu comment étudier les séries, c'est-à-dire la limite de $\sum_{k=0}^{n} u_k$.
Si nous changeons l'ordre des termes, la valeur de la limite change-t-elle ? Peut-on sommer en regroupant les termes par paquets éventuelllement infinis ?
En probabilités, nous aurons besoin de bien maitriser ces questions.
Nous commençons par étudier les séries semi-convergentes, qui montrent que l'ordre de la somme peut changer la valeur de la limite de la somme.

1.6.1 Séries semi-convergentes

Une série $\sum u_n$ **semi-convergente** est une série qui est convergente mais qui n'est pas absolument convergente ($\sum |u_n|$ diverge).

Exemple 51 — *La série harmonique alternée* $H_n^- = \sum \dfrac{(-1)^n}{n}$ *est une série semi-convergente.*

Proposition 52
Soit $\sum u_n$ *une série semi-convergente.*
Pour tout $l \in \mathbb{R}$, *il existe une bijection* $f : \mathbb{N} \to \mathbb{N}$ *telle que* $\sum u_{f(n)}$ *converge vers* l.

Démonstration. Notons que si la suite (u_n) n'a qu'un nombre fini de termes positifs ou négatifs, alors la suite étant convergente, elle serait absolument convergente. On peut donc définir deux sous suites de (u_n) :

1. $(u_{\varphi(k)})$ la sous-suite des termes positifs ou nuls.
2. $(u_{\psi(k)})$ la sous-suite des termes négatifs.

On a $\sum_{k=0}^{\varphi(n)} u_k = \sum_{k=0}^{n} u_{\varphi(k)} + \sum_{k=0}^{\varphi(n)-n-1} u_{\psi(k)}$, car sur les $\varphi(n)+1$ premiers termes, $n+1$ sont positifs, les autres sont donc négatifs. On en déduit que si $\left(\sum_{k=0}^{n} u_{\varphi(k)}\right)$ converge, alors $\left(\sum_{k=0}^{\varphi(n)-n-1} u_{\psi(k)}\right)$ converge.
Mais on a aussi $\sum_{k=0}^{\varphi(n)} |u_k| = \sum_{k=0}^{n} u_{\varphi(k)} - \sum_{k=0}^{\varphi(n)-n-1} u_{\psi(k)}$, donc $\left(\sum_{k=0}^{\varphi(n)} |u_k|\right)$ converge, ce qui contredit l'hypothèse non absolument convergente. On en déduit que les suites $\sum u_{\varphi(k)}$ et $\sum u_{\psi(k)}$ tendent respectivement vers $+\infty$ et $-\infty$.
On construit f par récurrence sur n :
Initialisation : On pose $f(0) = 0$ et $S_0 = u_0$, $\alpha_0 = |l - u_0|$.
Hérédité : On suppose que pour tout $k \leq n$, $f(k)$ est connu et qu'il existe a_{n-1} et b_{n-1}, α_{n-1} et S_{n-1} tels que

$$\{f(0), \cdots, f(n-1)\} = \{\varphi(0), \cdots, \varphi(a_{n-1})\} \cup \{\psi(0), \cdots, \psi(b_{n-1})\},$$

$S_{n-1} = \displaystyle\sum_{k=0}^{n-1} u_{f(k)} = \sum_{k=0}^{a_{n-1}} u_{\varphi(n)} + \sum_{k=0}^{b_{n-1}} u_{\psi(k)}$ et $|S_{n-1} - l| \leq \alpha_{n-1}$.
On obtient a_n, b_n, $f(n)$, $\alpha(n)$ et S_n de la manière suivante :

1. Si $S_{n-1} \leq l$, alors $a_n = a_{n-1} + 1$, $b_n = b_{n-1}$ et $f(n) = \varphi(a_n)$
2. Si $S_{n-1} > l$, alors $a_n = a_{n-1}$, $b_n = b_{n-1} + 1$ et $f(n) = \psi(b_n)$.
3. On pose $S_n = \displaystyle\sum_{k=0}^{n} u_{f(k)}$. Si $S_n - l$ et $S_{n-1} - l$ sont de signe opposé $\alpha_n = |u_{f(n)}|$ et sinon $\alpha_n = \alpha_{n-1}$.

Par construction, soit $|S_n - l| \leq |S_{n-1} - l| \leq \alpha_{n-1} = \alpha_n$ (cas où $S_{n-1} - l$ et $S_n - l$ de même signe) soit $|S_n - l| \leq |u_{f(n)}| = \alpha_n$ (cas où $S_{n-1} - l$ et $S_n - l$ de signe opposé). On en déduit que pour tout $n \in \mathbb{N}$, $|S_n - l| \leq \alpha_n$.

Montrons que pour tout $N \in \mathbb{N}$, la suite $(S_{N+n} - l)$ n'est pas de signe constant, c'est-à-dire que les suites (a_n) et (b_n) tendent toutes les deux vers $+\infty$. Pour cela, on procède par l'absurde : si $S_n - l$ est de signe constant à partir d'un certain rang N, par exemple $S_n - l \geq 0$ pour tout $n > N$, alors $S_n = S_N + \displaystyle\sum_{k=b_N+1}^{n} u_{\psi(k)}$, mais $\displaystyle\sum_{k=b_N+1}^{n} u_{\psi(k)}$ tend vers $-\infty$, ce qui contredit $S_n \geq l$. De même si $S_n - l < 0$ pour tout $n > N$, alors $S_n = S_N + \displaystyle\sum_{k=a_{N+1}}^{n} u_{\varphi(k)}$, mais $\displaystyle\sum_{k=a_{N+1}}^{n} u_{\varphi(k)}$ tend vers ∞, ce qui contredit $S_n < l$.
On en déduit que la suite $(f(n))$ tend vers $+\infty$ et comme $\sum u_n$ converge, $u_{f(n)}$ tend vers 0. Enfin, comme $S_n - l$ change de signe une infinité de fois, on en déduit que (α_n) tend vers 0.
Or, nous avions montré que $|S_n - l| \leq \alpha_n$, donc la suite (S_n) converge vers l. □

Nous allons étudier la notion de famille sommable. Nous verrons entre autres que pour les séries absolument convergentes, la valeur de la somme est la même pour toute permutation $f : \mathbb{N} \to \mathbb{N}$.

1.6.2 Familles sommables de réels positifs

§ *1. Définition et propriétés*

Définition 53
Soit I un ensemble non vide et $(\alpha_i)_{i\in I}$ une famille de nombres réels positifs. On pose

$$\sum_{i\in I} \alpha_i = \sup_{J\subset I,\ |J|<+\infty} \sum_{i\in J} \alpha_i$$

avec la convention que la borne supérieure vaut $+\infty$ si l'ensemble n'est pas majoré.
*Si la borne supérieure est finie, on dit que la famille $(\alpha_i)_i$ est **sommable**.*

Remarque 54 — *On pourrait aussi considérer la somme dans $\overline{\mathbb{R}}$, avec les $\alpha_i \in [0, +\infty]$.*

Proposition 55
Soit I un ensemble non vide et $(\alpha_i)_{i\in I}$ et $(\beta_i)_{i\in I}$ deux familles de réels positifs. On a

1. *Si $\alpha_i \leq \beta_i$, $\forall i \in I$, alors on a*
$$\sum_{i\in I} \alpha_i \leq \sum_{i\in I} \beta_i.$$
2. *Pour $a \geq 0$, on a*
$$\sum_{i\in I} (a\alpha_i) = a \sum_{i\in I} \alpha_i$$
3. $$\sum_{i\in I} (\alpha_i + \beta_i) = \sum_{i\in I} \alpha_i + \sum_{i\in I} \beta_i.$$

Preuve — Nous ne traitons que le point 3/ : Pour tout J fini, par définion de la borne supérieure (majorant)

$$\sum_{i\in J}(\alpha_i+\beta_i)=\sum_{i\in J}\alpha_i+\sum_{i\in J}\beta_i\leq\sum_{i\in I}\alpha_i+\sum_{i\in I}\beta_i.$$

Et encore par définition de la borne (plus petit des majorants)

$$\sum_{i\in I}(\alpha_i+\beta_i)\leq\sum_{i\in I}\alpha_i+\sum_{i\in I}\beta_i.$$

Réciproquement, pour tous J et J' de cardinal fini

$$\sum_{i\in J}\alpha_i+\sum_{i\in J'}\beta_i\leq\sum_{i\in J\cup J'}(\alpha_i+\beta_i)\leq\sum_{i\in I}(\alpha_i+\beta)$$

On applique la définition de la borne supérieure aux deux sommes de gauche successivement et on obtient

$$\sum_{i\in I}\alpha_i+\sum_{i\in I}\beta_i\leq\sum_{i\in I}(\alpha_i+\beta)$$

□

PROPOSITION 56
Soit $(\alpha_i)_{i\in I}$ une famille de réels positifs et soit $A\subset I$ une partie de I (pas forcément finie). Alors on a

$$\sum_{i\in A}\alpha_i=\sum_{i\in I}\mathbb{1}_A(i)\,\alpha_i.$$

COROLLAIRE 57
Soit $(\alpha_i)_{i\in I}$ une famille de réels positifs et soit $A,B\subset I$ deux parties (pas forcément finies). Alors on a

$$A\subset B\Rightarrow\sum_{i\in A}\alpha_i\leq\sum_{i\in B}\alpha_i.$$

$$A\cap B=\emptyset\Rightarrow\sum_{i\in A\cup B}\alpha_i=\sum_{i\in A}\alpha_i+\sum_{i\in B}\alpha_i.$$

§ 2. Lien avec les séries à termes positifs

PROPOSITION 58
Soient $I=\mathbb{N}$ et $(\alpha_i)_{i\in\mathbb{N}}$ une famille de réels positifs.
La famille $(\alpha_i)_i$ est sommable si et seulement si la suite $\left(\sum_{i=0}^{n}\alpha_i\right)_{n\in\mathbb{N}}$ est convergente.
On a dans ce cas :

$$\sum_{i\in I}\alpha_i=\sum_{k=0}^{+\infty}\alpha_k.$$

Preuve — La différence est qu'avec la nouvelle définition on ne tient pas compte de l'ordre des α_i tandis que dans la seconde si ! Rappelons que les α_i sont positifs.
Si $\left(\sum_{i=0}^{n}\alpha_i\right)_{n\in\mathbb{N}}$ converge, la suite est croissante et majorée par sa limite S.
Soit $J\subset\mathbb{N}$ de cardinal fini. Alors J admet un maximum N et

$$\sum_{i\in J}\alpha_i\leq\sum_{j=0}^{N}\alpha_j\leq S$$

Un ensemble non vide majoré de $\mathbb{R}$ admet une borne supérieure majorée par S, donc $\sum_{i\in\mathbb{N}}\alpha_i$ existe. De plus comme pour $J_n=[0,n]$, $\sum_{i\in J_n}\alpha_i=\sum_{i=0}^{n}\alpha_i$ tend vers S, on a bien égalité.

Réciproquement, si $\sum_{i\in\mathbb{N}}\alpha_i$ existe et vaut S, alors $\sum_{i=0}^{n}\alpha_i=\sum_{i\in[0,n]}\alpha_i$ est majoré par S et donc converge, et la première partie de la preuve montre qu'alors $\sum_{i=0}^{+\infty}\alpha_i=S$. □

Remarque 59 — *On a donc montré que lorsqu'on somme des réels positifs, l'ordre de la somme ne change pas le résultat.*
On va montrer que ceci est encore vrai pour les séries absolument convergentes.
On a vu que le résultat est complètement faux pour les séries semi-convergentes.

§ *3. Somme et dénombrabilité*

Proposition 60
Soit I un ensemble non vide et $(\alpha_i)_{i\in I}$ une famille de réels positifs qui est sommable ($\sum_i \alpha_i < +\infty$).
Alors l'ensemble des indices i tels que $\alpha_i > 0$ est fini ou dénombrable.
La famille $(\alpha_i)_{i\in I}$ possède un nombre au plus dénombrable de termes non-nuls.

Preuve — Soit l'ensemble J des indices tels que $\alpha_i > 0$. Pour tout $n > 1$ on pose J_n l'ensemble des indices de J tels que $\dfrac{1}{n-1} > \alpha_i \geq \dfrac{1}{n}$ et J_1 est l'ensemble des α_i tels que $\alpha_i \geq 1$. Comme $\displaystyle\sum_{i\in I} \alpha_i$ converge, on en déduit que J_n est de cardinal fini.
Or $J = \bigcup_{i\in\mathbb{N}} J_i$, donc J est au plus dénombrable comme union de parties finies, ce qui termine la preuve. □

§ *4. Théorème de sommation par paquets*

Théorème 61 (**Théorème de sommation par paquets**)
Soit $(\alpha_i)_{i\in I}$ une famille de réels positifs et soit $(A_j)_{j\in J}$ une partition de I.
Alors on a

$$\sum_{i\in I} \alpha_i = \sum_{j\in J} \left(\sum_{i\in A_j} \alpha_i \right).$$

Preuve — Pour toute famille finie F on pose J_F l'ensemble (fini) des indices j tels que $A_j \cap F \neq \emptyset$. Alors

$$\sum_{i\in F} \alpha_i = \sum_{j\in J_F} \left(\sum_{i\in A_j\cap F} \alpha_i \right) \leq \sum_{j\in J} \left(\sum_{i\in A_j} \alpha_i \right),$$

et par définition de la borne supérieure $\displaystyle\sum_{i\in I} \alpha_i \leq \sum_{j\in J} \left(\sum_{i\in A_j} \alpha_i \right)$.
Réciproquement, pour toute partie finie J_0 de J,

$$\sum_{j\in J_0} \left(\sum_{i\in A_j} \alpha_i \right) = \sum_{i\in \cup_{j\in J_0} A_j} \alpha_i \leq \sum_{i\in I} \alpha_i,$$

ce qui montre l'inégalité inverse. □

1.6.3 Familles sommables de nombres complexes

§ *1. Définition*

Soit $x \in \mathbb{R}$.
On note $x^+ = \max(0, x)$ (partie positive de x) et $x^- = \max(0, -x)$ (partie négative de x).
On a alors $x = x^+ - x^-$ et $|x| = x^+ + x^-$.

Définition 62
Soient I un ensemble non-vide et $(\alpha_i)_{i\in I}$ une famille de réels.
On dit que $(\alpha_i)_{i\in I}$ est ***sommable*** *si les familles $(\alpha_i^+)_{i\in I}$ et $(\alpha_i^-)_{i\in I}$ sont sommables en tant que familles de réels positifs.*
On définit alors la somme

$$\sum_{i\in I} \alpha_i = \sum_{i\in I} \alpha_i^+ - \sum_{i\in I} \alpha_i^-.$$

On dit qu'une famille de nombres complexes $(\beta_i)_{i\in I}$ est ***sommable*** *si la famille des parties réelles et la famille des parties imaginaires sont sommables. On définit alors la somme*

$$\sum_{i\in I} \beta_i = \sum_{i\in I} Re(\beta_i) + i \sum_{i\in I} Im(\beta_i).$$

Remarque 63 — *Pour tout nombre complexe $z \in \mathbb{C}$, on a*

$$\max(|\mathrm{Re}\, z|, |\mathrm{Im}\, z|) \leq |z| \leq |\mathrm{Re}\, z| + |\mathrm{Im}\, z|$$

On en déduit qu'une famille de nombres complexes $(z_i)_{i\in I}$ est sommable si et seulement si la famille $(|z_i|)_{i\in I}$ est sommable.

Proposition 64
Soit $(\alpha_i)_{i\in I}$ une famille de réels ou de nombres complexes. Soit J l'ensemble des indices j tels que $\alpha_j \neq 0$.
- *Si $(\alpha_i)_{i\in I}$ est sommable, alors l'ensemble J est au plus dénombrable.*
- *Si J est infini dénombrable, pour $(j_n)_{n\in\mathbb{N}}$ une énumération de J, on a alors que la famille $(\alpha_i)_{i\in I}$ est sommable ssi la série $\sum_{k=0}^{n} \alpha_{j_k}$ est absolument convergente.*

Dans ce cas, on obtient :

$$\sum_{i\in I} \alpha_i = \sum_{j\in J} \alpha_j = \sum_{k=0}^{\infty} \alpha_{j_k}.$$

Preuve — On a déjà montré ce résultat pour les séries à termes positifs et $|x| = x^+ + x^-$. Pour une famille de réels quelconque, comme $x = x^+ - x^-$, on a

$$\sum_{k=0}^{\infty} \alpha_{j_k} = \sum_{k=0}^{\infty} \alpha_{j_k}^{+} - \sum_{k=0}^{\infty} \alpha_{j_k}^{-}.$$

Enfin, pour les complexes, on applique le résultat à la partie réelle et la partie imaginaire. □

Remarque 65 — *On a donc démontré que si $\sum_{k\geq 0} u_k$ est une série absolument convergente, alors sa somme ne dépend pas de l'ordre dans lequel on somme les termes.*
Si une série est convergente mais pas absolument convergente, on a montré que cela était faux.
Dans le reste du cours, quand on sera face à une série $\sum_k u_k$, il sera primordial de regarder si la somme $\sum_k |u_k|$ est convergente ou non, afin de pouvoir appliquer les résultats de ce chapitre (comme le théorème de sommation par paquets).

§ 2. Propriétés des familles sommables

Proposition 66
Soit I un ensemble non vide et soient $(\alpha_i)_{i\in I}, (\beta_i)_{i\in I}$ deux familles de nombres complexes qui sont sommables.
Alors on a :

1. *Si les α_i, β_i sont tous réels, avec $\alpha_i \leq \beta_i$ $\forall i \in I$, on a*
$$\sum_{i\in I} \alpha_i \leq \sum_{i\in I} \beta_i.$$
2. *Pour $a \in \mathbb{C}$, on a*
$$\sum_{i\in I} (a\alpha_i) = a \sum_{i\in I} \alpha_i.$$
3. $$\sum_{i\in I} (\alpha_i + \beta_i) = \sum_{i\in I} \alpha_i + \sum_{i\in I} \beta_i.$$

Preuve — Nous ne traitons que le point 3/ et pour les familles de réels : La famille $(\alpha_i + \beta_i)$ est sommable car majorée en valeur absolue par $(|\alpha_i| + |\beta_i|)$ qui est sommable par définition.
Pour montrer l'égalité des sommes on doit partitionner l'ensemble des indices I de sorte que α_i, β_i et $\alpha_i + \beta_i$ sont de signe constant :

1. $I_{+,+} = \{i \in I \mid \alpha_i \geq 0,\ \beta_i \geq 0\}$
2. $I_{-,-} = \{i \in I \mid \alpha_i < 0,\ \beta_i < 0\}$
3. $I^+_{+,-} = \{i \in I \mid \alpha_i \geq 0,\ \beta_i < 0, \alpha_i + \beta_i \geq 0\}$
4. $I^-_{+,-} = \{i \in I \mid \alpha_i \geq 0,\ \beta_i < 0, \alpha_i + \beta_i < 0\}$
5. $I^+_{-,+} = \{i \in I \mid \alpha_i < 0,\ \beta_i \geq 0, \alpha_i + \beta_i \geq 0\}$
6. $I^-_{-,+} = \{i \in I \mid \alpha_i < 0,\ \beta_i \geq 0, \alpha_i + \beta_i < 0\}$

Sur chacun des ces ensembles on a l'égalité des sommes et on en déduit l'égalité finale car on somme 6 (fini) parties deux à deux disjointes. □

§ 3. Théorème de sommation par paquets

THÉORÈME 67 (**Théorème de sommation par paquets**)
Soit $(\alpha_i)_{i\in I}$ une famille de nombres complexes et soit $(A_j)_{j\in J}$ une partition de I.
La famille $(\alpha_i)_{i\in I}$ est sommable si et seulement si chacune des familles $(\alpha_i)_{i\in A_j}$ est sommable, et si la famille $\left(\sum_{i\in A_j} |\alpha_i|\right)_{j\in J}$ est elle aussi sommable.
Dans ce cas, on a alors

$$\sum_{i\in I} \alpha_i = \sum_{j\in J} \left(\sum_{i\in A_j} \alpha_i\right).$$

Preuve — C'est une conséquence directe du théorème de sommation par paquets pour les familles de réels positifs.
On l'applique à la famille $(|\alpha_i|)_{i\in I}$. Comme par définition, $(\alpha_i)_{i\in I}$ est sommable ssi $(|\alpha_i|)_{i\in I}$, on a la premère équivalence du théorème.
De plus, on applique encore le théorème de sommation par paquets aux familles (x_i^+), (x_i^-), (y_i^+), (y_i^-) où $\alpha_k = x_k + iy_k$, x_k, y_k réels.
Enfin, d'après la proposition 66, on a

$$\sum_{k\in I} \alpha_k = \sum_{k\in I} x_k^+ - \sum_{k\in I} x_k^- + i\sum_{k\in I} x_k^+ - i\sum_{k\in I} x_k^-$$

on obtient l'égalité de la seconde partie du théorème. □

REMARQUE 68 — *Pour que la famille $(\alpha_i)_{i\in I}$ soit sommable, il ne suffit pas que chacune des familles $(\alpha_i)_{i\in A_j}$ soit sommable et que la famille $\left(\sum_{i\in A_j} \alpha_i\right)_{j\in J}$ soit sommable.*
***Contre-exemple :** si $A_j = \{2j, 2j+1\}$ pour tout $j \in \mathbb{N}$ et $\alpha_j = (-1)^j$, la famille $(1, -1, 1, -1, ...)$ indexée par $\mathbb{N}$ n'est pas sommable.*
Il faut bien vérifier que $\left(\sum_{i\in A_j} |\alpha_i|\right)_{j\in J}$ est une famille sommable.

1.6.4 Somme double

On s'interroge sur la possibilité d'intervertir deux sommes indexées par des ensembles dénombrables :
Soit $(u_{i,j})_{(i,j)\in\mathbb{N}\times\mathbb{N}}$ une famille de complexes.

Suppposons que pour j fixé la somme $s_j = \sum_{i=0}^{+\infty} u_{i,j}$ existe, ainsi que la somme

$$\sum_{j=0}^{+\infty} s_j = \sum_{j=0}^{+\infty} \left(\sum_{i=0}^{+\infty} u_{i,j}\right).$$

Peut-on en déduire que la série $s'_i = \sum_{j=0}^{+\infty} u_{i,j}$ est convergente ?

La réponse est **non**.

Et même si s'_i converge et que la somme $\sum_{i=0}^{+\infty} s'_i$ converge, alors alors $\sum_{i=0}^{+\infty} s'_i$ n'est pas nécessairement égal à $\sum_{j=0}^{+\infty} s_j$.

Le théorème de Fubini pour les familles sommables nous donne les conditions à vérifier

Théorème 69 (**Théorème de Fubini**)
Soit $(u_{i,j})_{(i,j)\in\mathbb{N}^2}$ une famille de nombres complexes.
Alors, la famille $(u_{i,j})_{i,j}$ est sommable si et seulement si $\sum_{i=0}^{+\infty}\left(\sum_{j=0}^{+\infty}|u_{i,j}|\right) < +\infty$ est convergente, ou si $\sum_{j=0}^{+\infty}\left(\sum_{i=0}^{+\infty}|u_{i,j}|\right) < +\infty$.
On a alors

$$\sum_{(i,j)\in\mathbb{N}} u_{i,j} = \sum_{i=0}^{+\infty}\left(\sum_{j=0}^{+\infty} u_{i,j}\right) = \sum_{j=0}^{+\infty}\left(\sum_{i=0}^{+\infty} u_{i,j}\right).$$

Preuve — C'est exactement le théorème de sommation par paquets appliqué à $A_j = \{(i,j),\ i\in\mathbb{N}\}$ pour $j\in\mathbb{N}$.

□

Remarque 70 — *Pour montrer qu'une famille $(u_{i,j})_{(i,j)\in\mathbb{N}}$ est sommable, on va ainsi montrer que la série $\sum_i |u_{i,j}|$ est convergente pour tout j, puis que la série $\sum_j\left(\sum_i |u_{i,j}|\right)$ est convergente.*
On pourra alors appliquer le théorème de Fubini pour intervertir les deux sommes de séries.
Le théorème de Fubini est le théorème principal qui permet d'intervertir les sommes de séries (ou série et intégrale en analyse).

Corollaire 71 (Produit de séries absolument convergentes)
Soient $\sum_{i\geq 0} a_i$ et $\sum_{i\geq 0} b_i$ deux séries absolument convergentes.
Alors la famille $(u_{i,j} = a_i b_j)_{(i,j)\in\mathbb{N}^2}$ est sommable, et on a

$$\sum_{(i,j)\in\mathbb{N}} a_i b_j = \sum_{i\in\mathbb{N}} a_i \sum_{i\in\mathbb{N}} b_i.$$

Remarque 72 — *Si l'on pose $c_n = \sum_{k=0}^{n} a_k b_{n-k}$ et que l'on utilise le théorème de sommation par paquets, on retrouve le fait que le produit de Cauchy de deux séries absolument convergentes est absolument convergent, et que la somme du produit est le produit des sommes.*
Avec des séries entières il est courant de faire des produit $\sum a_n x^n \sum b_n x^n = \sum c_n x^n$.
Tant que les séries sont absolument convergentes, on pourra parler aussi bien de produit de familles sommables que de produit de Cauchy.

Exemple 73 — *Montrer que la famille $(\alpha_{m,n}) = \left(\frac{(-1)^{mn}}{m^2n^2}\right)_{m,n\in\mathbb{N}^*}$ est sommable et calculer sa somme S.*
On sait que $A = \sum_{n>0}\frac{1}{n^2} = \frac{\pi^2}{6}$ et que si P est l'ensemble des entiers pairs non nuls et I les entiers impairs,

alors

$$U = \sum_{p \in P} \frac{1}{p^2} = \sum_{n \in \mathbb{N}^*} \frac{1}{(2n)^2} = \frac{1}{4} A$$

et

$$V = \sum_{i \in I} \frac{1}{i^2} = \sum_{n \in \mathbb{N}^*} \frac{1}{n^2} - \sum_{p \in P} \frac{1}{p^2} = \frac{3}{4} A.$$

On a

$$\sum_{(m,n) \in (N^*)^2} |\alpha_{m,n}| = \sum_{(m,n) \in (N^*)^2} \frac{1}{m^2 n^2} = \left(\sum_{n>0} \frac{1}{n^2} \right)^2$$

donc la famille est bien sommable.
De plus le produit $m.n$ est impair ssi m et n sont impairs.
On en déduit que

$$S = U^2 + 2UV - V^2 = -\frac{\pi^4}{288}.$$

Exemple 74 — *Démontrer que la famille* $(\alpha_{n,p}) = \left(\frac{1}{n^p} \right)_{n,p \geq 2}$ *est sommable et calculer sa somme S.*
En effet, pour $n \geq 2$ fixé, la série géométrique $\sum_{p \geq 2} \frac{1}{n^p}$ *converge et vaut* $\frac{1}{n^2} \times \frac{1}{1 - \frac{1}{n}} = \frac{1}{n-1} - \frac{1}{n}$ *qui est le terme général d'une série télescopique convergente vers 1.*
La famille est bien sommable de somme 1 et on a ainsi montré que

$$\sum_{p=2}^{+\infty} \left(\sum_{n=2}^{+\infty} \frac{1}{n^p} \right) = 1 = \sum_{p=2} (\zeta(p) - 1) = 1$$

où la fonction zeta de Riemann vaut $\zeta(z) = \sum_{k=1}^{+\infty} \frac{1}{n^z}$ *pour* $|z| > 1$.

Exemple 75 —

1. *Calculer la somme* $\sum_{n=0}^{+\infty} \left(\sum_{k \geq n} \frac{1}{k!} \right)$.
2. *On pose* $a_{n,p} = \frac{1}{n^2 - p^2}$ *si $n \neq p$ et 0 sinon.*
   (a) *Expliquer pourquoi la famille n'est pas sommable.*
   (b) $\sum_n \sum_p a_{n,p}$ *et* $\sum_p \sum_n a_{n,p}$.

1. *La somme inversée vaut*
   $\sum_{k=0}^{+\infty} \sum_{n=0}^{k} \frac{1}{k!} = \sum_{k \geq 0} \frac{k+1}{k!} = \sum_{k=0}^{+\infty} \frac{1}{k!} + \sum_{k=1}^{+\infty} \frac{1}{(k-1)!} = 2e$. *Comme la famille est à termes positifs, on en déduit que la famille est sommable et que la somme demandée vaut encore $2e$.*
2. (a) *La somme des* $a_{n,n-1} \sim \frac{1}{2n}$ *tend vers* $+\infty$

   (b) *Il faut calculer pour tout $n \in \mathbb{N}$, la somme* $\sum_{p=0}^{+\infty} \frac{1}{n^2 - p^2}$.

   Pour $n = 0$, la somme $-\sum_{p=0}^{+\infty} \frac{1}{p^2} = -\frac{\pi^2}{6}$.

Si $n \neq 0$, *on remarque que* $\dfrac{1}{n^2-p^2} = \dfrac{1}{2n}\left(\dfrac{1}{n+p} + \dfrac{1}{n-p}\right)$, *donc pour* N *grand*

$$\begin{aligned}\sum_{p=0, p\neq n}^{N} \left(\frac{1}{n+p} + \frac{1}{n-p}\right) &= \sum_{k\geq n, k\neq 2n}^{N+n} \frac{1}{k} + \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{N-n} \frac{1}{k} \\ &= \frac{1}{n} - \frac{1}{2n} + \sum_{k=1}^{N+n} \frac{1}{k} - \sum_{k=1}^{N-n} \frac{1}{k} \\ &= \frac{1}{2n} + \sum_{k=N-n+1}^{N+n} \frac{1}{k}\end{aligned}$$

et on obtient

$$\sum_{p=0}^{+\infty} \frac{1}{n+p} + \frac{1}{n-p} = \lim_{N\to+\infty} \sum_{p=0}^{N} \frac{1}{n+p} + \frac{1}{n-p} = \frac{1}{2n}.$$

En remplaçant dans l'expression

$$\sum_n \sum_p a_{n,p} = \sum_{n\neq 0} \frac{1}{4n^2} - \frac{\pi^2}{6} = -\frac{\pi^2}{8} = -\sum_p \sum_n a_{n,p}.$$

Chapitre 2 Espaces probabilisés

Table des matières du chapitre

2.1 Le langage des probabilités

2.1.1 Expériences aléatoires

Définition 76
On appelle expérience aléatoire une expérience $\mathcal{E}$ qui, reproduite dans des conditions identiques, peut conduire à plusieurs résultats possibles, et dont on ne peut prévoir le résultat par avance.

Définition 77
L'ensemble de tous les résultats possibles d'une expérience aléatoire est appelé ensemble d'états ou ***univers****.*
Il est noté Ω.
Un élément de Ω est noté généralement ω ($\omega \in \Omega$).
On dit que ω est un résultat possible de l'expérience aléatoire.

Exemple 78 —

1. *On lance une pièce :* $\Omega = \{P, F\}$*, assimilé à* $\Omega = \{0, 1\}$.
2. *On lance un dé :* $\Omega = \{1, 2, 3, 4, 5, 6\}$.
3. *Génotype d'un individu :* $\Omega = \{AA, Aa, aa\}$.
4. *On étudie n individus :* $\Omega = \{AA, Aa, aa\}^n$.
5. *On étudie la durée de vie d'une bactérie :* $\Omega = [0, +\infty[$.
6. *On étudie la durée d'une communication téléphonique :* $\Omega = [0, +\infty[$.
7. *On envoie une fléchette sur une cible circulaire de 30 cm de diamètre :* $\Omega = \{(x, y), x^2 + y^2 \leq 15\}$.
8. *Cours d'un actif financier sur un intervalle de temps $[t_1, t_2]$:* $\Omega = \mathcal{C}^0([t_1, t_2], \mathbb{R}_+^*)$.

Remarque 79 — *Cette longue liste d'exemples montre que l'espace Ω peut varier énormément dans sa structure, d'une expérience à l'autre. Cela permet de réaliser la richesse de la théorie qu'il faut mettre en place pour créer un modèle qui englobe tous les cas.*

2.1.2 Événements aléatoires

Remarque 80 — *Quelle information pouvons-nous tirer de l'expérience ? Dans le jeu de fléchettes, on s'intéresse à la chance de tomber dans une des couronnes ou un des secteurs de la cible.*
Les résultats du jeu peuvent se décrire à l'aide de parties du disque. Pas la température de la pièce...

Définition 81
Soit Ω un ensemble associé à une expérience aléatoire.
On appelle **événement aléatoire** *(associé à l'expérience $\mathcal{E}$) un sous-ensemble de $A \subset \Omega$ dont on peut dire au vu de l'expérience s'il est réalisé ou non.*

Exemple 82 —

1. $\Omega = \{0, 1\}$. *"La pièce tombe sur Pile" : $A = \{0\}$.*
2. $\Omega = \{0, 1\}^n$, $\omega = (\omega_1, \cdots, \omega_n)$. *"Le nombre de Faces est supérieur au nombre de Piles" :*
$$A = \{\omega \in \Omega, \sum_{i=1}^{n} \omega_i \geq \frac{n}{2}\}.$$
3. *Dans un lancer de deux dés*
$$A = \{\textit{La somme des deux dés est inférieure à } 4\}$$
est un évènement aléatoire mais
$$B = \{\textit{Le résultat du premier dé lancé est inférieur à } 4\}$$
n'en est pas si Ω ne contient que les résultats non ordonnés des tirages.

Un événement aléatoire A est un sous-ensemble, donc il est caractérisé par l'ensemble des ω qu'il contient (l'ensemble des résultats tels que l'événement se réalise).

2.1.3 Opérations ensemblistes et description des événements aléatoires

Comme les événements sont des sous-ensembles, on peut effectuer des opérations ensemblistes dessus, et donner des interprétations.
Soient A et B sont deux événements d'une expérience aléatoire. (deux parties d'un ensemble Ω) On a :

1. A n'est pas réalisé : $\overline{A}$ (le complémentaire de A, noté aussi A^C)
2. A **et** B sont réalisés : $A \cap B$
3. A **ou** B sont réalisés : $A \cup B$
4. A réalisé $\Rightarrow$ B réalisé : $A \subset B$.
5. A et B sont incompatibles : $A \cap B = \emptyset$.
6. Toujours vrai : Ω est l'événement certain (il arrive toujours).
7. Jamais vrai : $\emptyset$ est l'événement impossible (il n'arrive jamais).

On note $\mathcal{A}$ l'ensemble de tous les événements possibles dans l'expérience aléatoire.
C'est un ensemble de parties de Ω.
Cet ensemble modélise l'information que l'on peut obtenir à partir des résultats de l'expérience.
Il est important de comprendre que l'on n'a pas toujours $\mathcal{A} = P(\Omega)$, ensemble de toutes les parties de Ω : dans le jeu de fléchettes, on ne considère pas que pour gagner la flèche doit avoir des coordonnées rationnelles dans un repère centré au milieu du disque.

Remarque 83 — *Pour que la modélisation soit cohérente avec l'intuition, l'ensemble $\mathcal{A}$ doit être stable par les opérations ensemblistes : si A, $B \in \mathcal{A}$, alors on doit avoir $A \cap B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$, $\overline{A} \in \mathcal{A}$, mais aussi $\Omega \in \mathcal{A}$, $\emptyset \in \mathcal{A}$.*

2.1.4 Probabilités : approche intuitive

Remarque 84 — *Comment savoir si la pièce est truquée dans un jeu de Pile ou Face ?*

Approche intuitive Considérons une expérience aléatoire donnée $\mathcal{E}$ et un événement A pour cette expérience.
Le but : associer à chaque événement A un nombre $\mathbb{P}(A)$ compris entre 0 et 1, qui représente la chance a priori que cet événement soit réalisé.
Ce nombre réel est appelé **probabilité de l'événement** A.
Supposons que l'on répète n fois l'expérience $\mathcal{E}$. On note n_A le nombre de fois où l'événement A s'est réalisé. Alors,

$$f_n(A) = \frac{n_A}{n}$$

donne la fréquence des réalisations de A sur ces n essais.
On remarque alors que

1. $f_n(A) \in [0,1]$;
2. $f_n(\Omega) = 1$ et $f_n(\emptyset) = 0$;
3. Si $A \cap B = \emptyset$, on a
$$f_n(A \cup B) = f_n(A) + f_n(B).$$
4. Si $A \subset B$ on a $f_n(A) \leq f_n(B)$;
5. Intuitivement, on imagine avoir
$$\mathbb{P}(A) = \lim_{n\to+\infty} f_n(A).$$

Cette conception de la probabilité d'un événement A comme fréquence d'apparition de l'événement est l'approche intuitive que l'on a de la notion.
Pour une expérience où on lance un dé à 6 face (dé équilibré, non truqué), on a l'intuition que la probabilité de l'événement $\{\text{Obtenir } 4\}$ est égale à $\frac{1}{6}$, car cette quantité représente la fréquence à laquelle les lancers de dé vont donner un 6.

Exemple 85 (Expérience aléatoire avec une infinité de résultats possibles) — *On considère un jeu de Pile ou Face infini, avec pièce équilibrée. L'ensemble des résultats est ainsi* $\Omega = \{0,1\}^{\mathbb{N}}$.
On veut regarder l'événement

$$A = \{\textit{on ne tire jamais Pile}\}.$$

Que pourrait valoir la probabilité de l'événement A, $\mathbb{P}(A)$?
Si l'on regrde les n premiers tirages, on est sur l'ensemble fini $\Omega_n = \{0,1\}^n$ et, on regarde l'événement $A_n = \{\textit{on ne tire jamais Pile en n lancers}\}$.
Alors, on a$\mathbb{P}(A_n) = \dfrac{1}{2^n}$.
Il semble alors évident d'écrire que

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcap_n An\right) = \lim_{n\to+\infty} \mathbb{P}(A_n) = 0.$$

Cependant, pour qu'un tel résultat soit vrai, il faut que la probabilité $\mathbb{P}$ possèdent une propriété de passage à la limite. Cela implique aussi que l'ensemble des événements $\mathcal{A}$ posède une condition de stabilité pour des intersections dénombrables (on a $A = \cap_{n\geq 1} A_n$).

Avec cette section, nous avons donné quelques exemples d'expériences aléatoires, nous avons vu un peu de vocabulaire du monde des probabilités (expérience, événement, réalisation, probabilité), et nous avons vu que deux objets semblaient importants (l'ensemble des événements, la probabilité de chaque événement).
Il est temps de définir mathématiquement ces objets.

2.2 Définitions générales en probabilités

Pour avoir une structure mathématique qui permet de modéliser facilement et fidèlement les expériences aléatoires que l'on veut étudier dans le monde réel, nous allons avoir besoin de trois objets fondamentaux : les **sigma-algèbres** (ou tribus), les **mesures de probabilités**, et les **variables aléatoires**.
Dans ce chapitre, nous allons définir les sigma-algèbres et les mesures de probabilités, sur un ensemble Ω donné.
Les variables aléatoires seront définies au prochain chapitre.
Ces définitions sont courtes, mais extrêmement importantes. De ces définitions découlent toutes les propriétés que nous utiliserons dans nos raisonnemens et nos calculs.
Nous verrons des exemples de mesures de probabilités sur des ensembles finis ou dénombrables, et avec les propriétés des mesures de probabilités nous utiliserons les techniques de dénombrement pour calculer des probabilités.

2.2.1 Tribu | σ-algèbre

Commençons par l'objet de base, la sigma-algèbre (ou tribu).

Définition 86
Soit Ω un ensemble. Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$ un ensemble de parties de Ω.
*On dit que $\mathcal{A}$ est une **σ-algèbre** (ou une **tribu**) si elle vérifie les propriétés suivantes :*

1. *On a $\emptyset \in \mathcal{A}$ et $\Omega \in \mathcal{A}$.*
2. *$\mathcal{A}$ est stable par passage au complémentaire :*
   Pour tout $A \in \mathcal{A}$ on a $\overline{A} \in \mathcal{A}$.
3. *$\mathcal{A}$ est stable par réunion et intersection dénombrables :*
   Pour $(A_n)_{n\in\mathbb{N}}$ une famille d'éléments de $\mathcal{A}$, on a
   $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$ et $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Remarque 87 —

1. *Une σ-algèbre $\mathcal{A}$ toujours non vide.*
2. *Pour A, $B \in \mathcal{A}$, alors $A \cup B \in \mathcal{A}$ et $A \cap B \in \mathcal{A}$.*
   Une σ-algèbre est stable par intersection et par réunion (finies).
3. *L'ensemble $\mathcal{A}$ est un ensemble de parties de Ω, et pas un sous-ensemble de Ω.*
   Cet ensemble est un sous-anneau de $(\mathcal{P}(\Omega), \cup, \cap)$, et une $\mathbb{Z}/2\mathbb{Z}$-algèbre, avec une propriété de "dénombrabilité" en plus.
   Le nom de "sigma-algèbre" fait référence à cette structure algébrique, et à cette propriété en plus.
4. *Notons également que la propriété 3. n'implique pas que $\mathcal{A}$ soit stable par réunion ou intersection infinie non dénombrable.*
5. *La σ-algèbre représente l'ensemble des parties de Ω que l'on va chercher à mesurer, dont on va chercher à donner une probabilité (pour tout $A \in \mathcal{A}$, on donnera un sens à $\mathbb{P}(A)$, la probabilité de la partie A). C'est le premier élément fondamental pour modéliser les expériences aléatoires.*

Exemple 88 —

1. *$\mathcal{A} = \{\emptyset, \Omega\}$ est une σ-algèbre.*
   On l'appelle la tribu grossière, ou triviale. C'est la plus petite tribu de Ω.
2. *L'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω est une σ-algèbre sur Ω.*

Proposition-Définition 89
Soit Ω un ensemble. Soit $C \subset P(\Omega)$.
On appelle tribu engendrée par C la plus petite tribu contenant C.
Cette tribu existe toujours, car d'une part $\mathcal{P}(\Omega)$ est une tribu contenant C, et d'autre part l'intersection d'une famille quelconque de tribus est une tribu.
Ainsi, la tribu engendrée par C est l'intersection de toutes les tribus contenant C.

Exemple 90 —

1. *La tribu engendrée par l'ensemble* $\{A\}$ *est* $\{\emptyset, A, \overline{A}, \Omega\}$.
2. *Si* $(A_i)_{i\in I}$ *est une partition finie ou dénombrable de* Ω*, la tribu engendrée par* $\{A_i, i \in I\}$ *est l'ensemble des réunions* $B_J = \bigcup_{i\in J} A_i$*, pour tout* $J \subset I$.

Proposition 91
Soit Ω *un ensemble fini ou dénombrable.*
Soit $\mathcal{A}$ *une* σ*-algèbre sur* Ω *qui contienne tous les singletons :* $\forall\omega \in \Omega$*, on a* $\{\omega\} \in \mathcal{A}$.
Alors, on a $\mathcal{A} = \mathcal{P}(\Omega)$.

Preuve — Soit B une partie de Ω. On a $B = \cup_{\omega\in B}\{\omega\}$.
Comme Ω est dénombrable, l'ensemble B est fini ou dénombrable.
Comme tous les singletons $\{\omega\}$ sont dans $\mathcal{A}$, on en déduit que leur réunion finie ou dénombrable est dans $\mathcal{A}$; donc B est dans $\mathcal{A}$.
On a donc bien que $\mathcal{A} = \mathcal{P}(\Omega)$. ☐

Remarque 92 — *A chaque fois que l'ensemble* Ω *sera fini ou dénombrable, on choisira comme* σ*-algèbre* $\mathcal{A}$ *l'ensemble* $\mathcal{P}(\Omega)$.
En effet, c'est la seule σ*-algèbre de* Ω *qui contienne tous les singletons.*

Remarque 93 — *Quand* Ω *est infini non dénombrable (par exemple* $\mathbb{R}$*), la* σ*-algèbre engendrée par les singletons est différente de* $\mathcal{P}(\Omega)$.
De plus, la tribu $\mathcal{P}(\Omega)$ *sera trop grande pour que l'on puisse définir la probabilité de tous ses éléments de façon simple.*

Définition 94
Soit $\Omega = \mathbb{R}$.
On appelle ***tribu borélienne*** *la tribu de* $\mathbb{R}$ *engendrée par l'ensemble de tous les intervalles.*
On la note $\mathcal{B}(\mathbb{R})$.

Proposition 95
La tribu borélienne de $\mathbb{R}$ *est la tribu engendrée par les intervalles de la forme* $]-\infty, a]$ *pour* $a \in \mathbb{Q}$.
De plus, cette tribu contient tous les singletons.

Preuve — • Rappelons que toute tribu est stable par passage au complémentaire, par réunion ou intersection dénombrable. Puisque $]-\infty, a]$ est le complémentaire de l'intervalle ouvert $]a, +\infty[$, il appartient à la tribu borélienne, et donc la tribu C engendrée par ces intervalles est incluse dans la tribu borélienne.
Réciproquement, soit $]x, y[$ un intervalle ouvert de $\mathbb{R}$. Soit $(x_n)_{n\geq 0}$ une suite de rationnels décroissant vers x et $(y_n)_{n\geq 0}$ une suite de rationnels croissant strictement vers y.
On a :

$$]x, y[= \bigcup_{n\geq 0} \left( ]-\infty, y_n] \cap \overline{]-\infty, x_n]} \right).$$

Nous en déduisons que tout intervalle ouvert appartient à C, d'où le résultat.
• Soit $x \in \mathbb{R}$. Alors le complémentaire de $\{x\}$ est $\mathbb{R} \setminus \{x\}$.
Or, les intervalles $]-\infty, x[$ et $]x, +\infty[$ sont dans $\mathcal{B}(\mathbb{R})$. Donc, par réunion et passage au complémentaire, on en déduit que $\{x\}$ est dans $\mathcal{B}(\mathbb{R})$. ☐

Remarque 96 —

1. *La tribu borélienne de* $\mathbb{R}$ *contient tous les singletons* $\{x\}$.
2. *Cette* σ*-algèbre* $\mathcal{B}(R)$ *contient strictement la* σ*-algèbre engendrée par tous les singletons. Elle est aussi strictement incluse dans* $\mathcal{P}(\mathbb{R})$.
   C'est-à-dire qu'il existe des éléments dans $\mathcal{B}(\mathbb{R})$ *qu'on ne peut pas engendrer simplement à partir de singletons, et qu'il existe des parties* E *de* $\mathbb{R}$ *qui ne sont pas dans la tribu borélienne.*
3. *Dans la suite du cours, les ensembles* Ω *sur lesquels nous prendrons des tribus seront en général soit finis, soit dénombrables, soit égaux à* $\mathbb{R}$.
   - *Si* Ω *est fini ou dénombrable, on aura comme tribu* $\mathcal{A} = \mathcal{P}(\Omega)$.

- *Si $\Omega = \mathbb{R}$, on aura comme tribu $\mathcal{B}(\mathbb{R})$.*

Ces cas paraissent simples, mais ils permettent déjà de construire énormément de variables aléatoires, et d'obtenir beaucoup de phénomènes différents. La majorité des exemples d'expériences aléatoires que vous connaissez peuvent se modéliser avec ces ensembles et ces tribus.

2.2.2 Mesure de probabilité, probabilité

Maintenant que nous avons les sigma-algèbres, nous pouvons donner la définition d'une mesure de probabilité, et celle d'une probabilité.

Définition 97
Soient Ω un ensemble, et $\mathcal{A}$ une tribu sur Ω. Soit $\mathbb{P} : \mathcal{A} \to [0,1]$ une fonction.
On dit que $\mathbb{P}$ est une ***mesure de probabilité*** *sur le couple $(\Omega, \mathcal{A})$ si elle vérifie les propriétés suivantes :*

1.
$$\mathbb{P}(\Omega) = 1. \quad (\textit{mesure totale de l'ensemble})(*)$$

2. *Pour toute famille dénombrable $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$ deux-à-deux disjoints, on a*
$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n). \quad (\textit{sigma-additivité})(**)$$

Le nombre réel $\mathbb{P}(A)$ est appelé ***probabilité*** *de la partie A.*

Remarque 98 —

1. *La propriété $(**)$ est appelée σ-additivité.*
Cette propriété implique que la série de terme général $\mathbb{P}(A_n)$ est convergente.
En effet, c'est une série à termes positifs et majorée par 1.
De plus, sa somme est égale à $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right)$.
2. *Tout comme on distingue un polynôme $P(X)$ de son évaluation en x, $P(X)$, et la fonction dérivée f' du nombre dérivé en x, $f'(x)$, on distingue bien la* ***mesure de probabilité*** *$\mathbb{P}$ de la* ***probabilité*** *d'une partie A, $\mathbb{P}(A)$.*
Le premier est une fonction, le second un nombre réel entre 0 et 1.
3. *Pour que la propriété 1) ait un sens, il était nécessaire que Ω soit un élément de $\mathcal{A}$.*
Pour que la propriété 2) ait un sens, il était nécessaire que $\mathcal{A}$ soit stable par réunion dénombrable.
4. *La mesure de probabilité $\mathbb{P}$ n'est définie que sur $\mathcal{A}$.*
Ainsi, la probabilité $\mathbb{P}(A)$ a un sens uniquement pour les éléments de la sigma-algèbre $\mathcal{A}$.
Si une partie $E \subset A$ n'est pas dans la sigma-algèbre $\mathcal{A}$, parler de "probabilité de E" ***n'a pas de sens****.*
Par exemple, pour $\Omega = \mathbb{R}$ et $\mathcal{A}$ la sigma-algèbre des boréliens, il existe des parties E de $\mathbb{R}$ qui ne sont pas dans $\mathcal{A}$, et pour lesquelles on ne pourra pas parler de "probabilité" (pas avec les mesures de probabilités les plus classiques sur $\mathbb{R}$).
5. *La mesure de probabilité est le deuxième objet mathématique essentiel pour modéliser les expériences aléatoires.*

Exemple 99 — *Soit Ω un ensemble, et $\mathcal{A}$ une sigma-algèbre sur Ω.*
Soit $\omega_0 \in \Omega$.
On définit la fonction $\delta_{\omega_0} : A \in \mathcal{A} \mapsto \begin{cases} 1 \text{ si } \omega_0 \in A \\ 0 \text{ sinon} \end{cases}$.
Alors, δ_{ω_0} est une mesure de probabilité. (Le montrer)
Cette mesure est appelée ***mesure de Dirac*** *en ω_0.*
C'est un des exemples les plus simples de mesure de probabilité que l'on peut construire. On le reverra par la suite, car il est en fait très utile.

2.2.3 Mesure de probabilité uniforme, cas fini

PROPOSITION-DÉFINITION 100
Soit Ω un ensemble fini.
On appelle ***mesure de probabilité uniforme*** *sur Ω la mesure de probabilité telle que*

$$\mathbb{P}(\{\omega\}) = \frac{1}{\text{card}\,(\Omega)}, \ \forall \omega \in \Omega.$$

Pour toute partie $A \in \mathcal{P}(\Omega)$, on a alors

$$\mathbb{P}(A) = \frac{\text{card}\,(A)}{\text{card}(\Omega)}.$$

Preuve — On pose la fonction $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ définie par $\mathbb{P}(A) = \dfrac{\text{card}\,(A)}{\text{card}(\Omega)}$.
Il faut montrer que cette fonction est une mesure de probabilité, afin de prouver le résultat.
• Tout d'abord, on peut remarquer que $\mathbb{P}$ est bien définie : $\mathbb{P}(A)$ est compris entre 0 et 1.
• Ensuite, on remarque que $\mathbb{P}(\Omega) = 1$. La fonction vérifie donc la propriété $(*)$.
• Maintenant, soit $(A_n)_{n\geq 0}$ une famille d'éléments de la sigma-algèbre $\mathcal{P}(\Omega)$ qui soient deux à deux disjoints.
Comme l'ensemble Ω est fini, cette famille de parties de Ω a seulement un nombre fini de parties qui sont non-vides.
Quitte à réordonner, on peut supposer que $A_0, \ldots, A_m$ sont non-vides, et que $A_n = \emptyset$, pour tout $n \geq m+1$.
On a alors :

$$\mathbb{P}(\cup_{n\geq 0} A_n) = \mathbb{P}(\cup_{n=0}^m A_n) = \frac{\text{card}\,(\cup_{n=0}^m A_n)}{\text{card}(\Omega)}$$

$$\mathbb{P}(\cup_{n\geq 0} A_n) = \frac{\sum_{n=0}^m \text{card}\,(A_n)}{\text{card}(\Omega)} = \sum_{n=0}^m \mathbb{P}(A_n).$$

Cela montre que $\mathbb{P}$ vérifie la propriété $(**)$, donc que $\mathbb{P}$ est une mesure de probabilité.
On a bien construit la mesure de probabilité uniforme ! □

REMARQUE 101 —

1. *Cette probabilité décrit mathématiquement l'expression intuitive de "au hasard" (tirage au hasard d'une carte, lancer au hasard d'un dé, etc).*
   C'est-à-dire, que l'on a plusieurs résultats possibles pour une expérience aléatoire (les 6 faces d'un dé, les 52 cartes d'un jeu,...), et qu'aucun résultat n'est avantagé par rapport aux autres.
   Autrement dit, tous les résultats de l'expérience ont une probabilité identique d'arriver. Cette probabilité est donc : $\frac{1}{\textit{nombre de résultats possibles}}$.
2. *Pour calculer la probabilité d'un événement A avec la mesure de probabilité uniforme, il faut calculer $Card(A)$, c'est-à-dire dénombrer A (compter le nombre d'éléments de A).*
3. *Ainsi, le calcul des probabilités (avec la mesure uniforme) se ramène à du calcul combinatoire (au contenu du premier chapitre).*
   La difficulté est de bien décrire et dénombrer l'ensemble total Ω et a partie A qui nous intéresse.

Il existe beaucoup d'autres mesures de probabilités sur un ensemble fini. Mais avant de continuer, intéressons-nous aux propriétés de ces fonctions.

2.2.4 Propriétés d'une mesure de probabilité

La définition d'une mesure de probabilité est courte, mais la propriété de sigma-additivité $(**)$ engendre beaucoup d'autres propriétés qui sont extrêmement utiles pour le calcul de probabilités.
Ces propriétés font intervenir toutes les propriétés des ensembles classiques (union, intersection, complémentaire, inclusion), et toutes les propriétés d'une sigma-algèbre (union dénombrable, intersection dénombrable).

PROPOSITION 102 (**Propriétés élémentaires**)
Soient Ω un ensemble, et $\mathcal{A}$ une tribu sur Ω.
Soit $\mathbb{P}$ une mesure de probabilité. Soient $A, B \in \mathcal{A}$. Alors, on a les résultats suivants :

1. $\mathbb{P}(\emptyset) = 0$ *;*

2. $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$;
3. *Si* $A \subset B$, *alors* $\mathbb{P}(A) \leq \mathbb{P}(B)$.
   La fonction $\mathbb{P}$ *est croissante pour l'inclusion.*
4. $\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cup B)$

Preuve — • Commençons par prouver le point 3).
On découpe B en deux sous-ensembles : $B \cap A = A$ et $B \cap \bar{A}$.
Ces deux sous-ensembles sont inclus dans $\mathcal{A}$, et sont disjoints.
On a donc par sigma-additivité de $\mathbb{P}$ que $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap \bar{A}) \geq \mathbb{P}(A)$.
• Prouvons ensuite le point 4).
On sépare l'ensemble $A \cup B$ en trois morceaux : $A_1 = A \setminus (A \cap B)$, $A_2 = A \cap B$, $A_3 = B \setminus (A \cap B)$.
Ces sous-ensembles appartiennent tous à la tribu $\mathcal{A}$ comme réunions/intersections/complémentaires d'éléments de $\mathcal{A}$. (par exemple, $A \setminus (A \cap B) = A \cap \overline{(A \cap B)}$)
De plus, ces sous-ensembles sont deux à deux disjoints.
Ainsi, la propriété de sigma-addiivité de la mesure de probabilité $\mathbb{P}$ donne :

$$\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A_1 \cup A_2) + \mathbb{P}(A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_2) + \mathbb{P}(A_3)$$
$$\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A_1 \cup A_2 \cup A_3) + \mathbb{P}(A_2) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cup B).$$

• On obtient maintenant le point 2) en utilisant 4) avec $B = \bar{A}$.
• On obtient maintenant le point 1) en utilisant 2) avec $A = \Omega$. □

Remarque 103 —

1. *On voit à nouveau dans cette Proposition que les propriétés de la sigma-algèbre A sont utilisées.*
   Le point 1) *n'aurait pas de sens si* $\mathcal{A}$ *ne contenait pas l'ensemble vide* $\emptyset$.
   Le point 2) *n'aurait pas de sens si* $\mathcal{A}$ *n'était pas stable par passage au complémentaire.*

Le résultat suivant est lui aussi extrêmement utile dans la pratique.

Proposition 104 (**Probabilités et suites croissantes/décroissantes**)
Soient Ω *un ensemble, et* $\mathcal{A}$ *une tribu sur* Ω.
Soit $\mathbb{P}$ *une mesure de probabilité. Alors, on a les propriétés suivantes :*

(i) Pour toute famille $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ *croissante pour l'inclusion, on a*

$$\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \lim_{n \to +\infty} \mathbb{P}(A_n).$$

(ii) Pour toute famille $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ *décroissante pour l'inclusion, on a*

$$\mathbb{P}(\bigcap_{n \in \mathbb{N}} A_n) = \lim_{n \to +\infty} \mathbb{P}(A_n).$$

Preuve —
(i) Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'éléments de $\mathcal{A}$ croissante pour l'inclusion. Notons $A = \bigcup_{n \geq 0} A_n$.
Posons $B_0 = A_0$, et définissons par récurrence $B_n = A_n \setminus B_{n-1}$, pour $n \geq 1$.
Comme $A_n = \bigcup_{p \leq n} B_p$, $\bigcup_{n \in \mathbb{N}} B_n = A$ et comme les B_n sont deux-à-deux disjoints, nous avons

$$\mathbb{P}(A) = \sum_{n \geq 0} \mathbb{P}(B_n) = \lim_{n \to +\infty} \sum_{p=0}^{n} \mathbb{P}(B_p) = \lim_{n \to +\infty} \mathbb{P}(A_n).$$

(ii) Si $(A_n)_n$ est une suite décroissante pour l'inclusion, alors la suite des complémentaires $C_n = \bar{A_n}$ est croissante pour l'inclusion.
En appliquant le point (i) et en utilisant la propriété $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$, on en déduit le résultat. □

Remarque 105 —

1. *Ce résultat entraîne en particulier que si* $(A_n)_{n \in \mathbb{N}}$ *est une suite croissante ou décroissante d'événements, la suite des probabilités* $(\mathbb{P}(A_n))_{n \geq 0}$ *admet une limite quand* n *tend vers l'infini.*
2. *Pour que ce résultat ait un sens, il fallait que la sigma-algèbre* $\mathcal{A}$ *soit stable par intersection dénombrable.*
   Et voilà, toutes les propriétés d'une sigma-algèbre ont été nécessaires pour obtenir des propriétés sur les mesures de probabilités.

3. *Par définition d'une mesure de probabilité, on peut calculer la probabilité d'une réunion dénombrable de parties disjointes.*
Avec la propriété précédente, on peut calculer la probabilité d'une réunion d'ensembles formant une suite croissante (ou intersection d'ensembles formant une suite décroissante).
On remarque dans la preuve de la proposition que la condition de suite croissante/décroissante était absolument nécessaire pour obtenir le résultat.

Si les parties A_n ne sont pas deux-à-deux disjointes, et ne forment pas de suite croissante/décroissante, on ne peut pas appliquer les propriétés précédentes.
Par contre, nous avons la majoration suivante, très utile dans la pratique.

Proposition 106 (**Probabilité d'une réunion quelconque**)
Soient Ω un ensemble, $\mathcal{A}$ une tribu sur Ω, et $\mathbb{P}$ une mesure de probabilité sur $(\Omega, \mathcal{A})$.
Soit $(A_n)_{n\in I}$ une famille finie ou dénombrable d'élements de $\mathcal{A}$ (une famille d'événements). On a alors

$$\mathbb{P}\left(\bigcup_{n\in I} A_n\right) \leq \sum_{n\in I} \mathbb{P}(A_n).$$

Preuve — • Supposons d'abord l'ensemble I fini. Il s'agit de montrer que pour tout entier k,

$$\mathbb{P}(A_1 \cup \cdots \cup A_k) \leq \mathbb{P}(A_1) + \cdots + P(Ak).$$

Nous montrons cette propriété par récurrence sur k :
Initialisation : Elle est évidente pour $k = 1$.
Hérédité : Supposons la propriété vraie pour $k-1$, avec $k \geq 2$.
Posons $B = A_1 \cup \cdots \cup A_{k-1}$ et $C = B \cup A_k$.
Nous savons que

$$\mathbb{P}(C) + \mathbb{P}(B \cap A_k) = \mathbb{P}(B) + \mathbb{P}(A_k),$$

donc $\mathbb{P}(C) \leq \mathbb{P}(B) + \mathbb{P}(A_k)$, et nous en déduisons immédiatement la proposition au rang k.
• Considérons maintenant le cas où I est dénombrable.
Nous pouvons supposer sans restriction que $I = \mathbb{N}$, d'après les résultats du premier chapitre.
Posons $B_n = \bigcup_{i=0}^n A_i$, qui est une suite croissante, qui converge en croissant vers l'ensemble $C = \bigcup_{n\in\mathbb{N}} A_n$.
D'après la première partie de la preuve, nous avons

$$\mathbb{P}(B_n) \leq \sum_{i=0}^n \mathbb{P}(A_i).$$

Mais le membre de gauche ci-dessus croît vers $\mathbb{P}(C)$ en vertu de la proposition précédente, tandis que le membre de droite croît vers $\displaystyle\sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$.
En passant à la limite, nous obtenons le résultat. □

Et voilà, nous avons énoncé ici toutes les propriétés les plus fondamentales d'une mesure de probabilité. Avec les deux objets que sont la sigma-algèbre (ou tribu) et la mesure de probabilité, on regarde en général le triplet suivant :

Définition 107
Soient Ω un ensemble, $\mathcal{A}$ une tribu sur Ω, et $\mathbb{P}$ une mesure de probabilité sur $(\Omega, \mathcal{A})$.
*Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé un **espace probabilisé** (ou **espace de probabilité**).*

Un espace de probabilité est tout simplement un ensemble que l'on a muni d'une mesure de probabilité. Et pour définir une mesure de probabilité, il faut aussi définir une sigma-algèbre.

La définition suivante est fondamentale en théorie des probabilités.
Elle introduit une notion de "vrai ou faux", qui dépend de la probabilité choisie sur la sigma-algèbre $\mathcal{A}$ (sur l'ensemble de tous les événements que l'on va considérer).

Définition 108
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
Soient $A, B \in \mathcal{A}$.
— *Si on a $\mathbb{P}(A) = 0$, on dit alors que A est un **événement négligeable** (un événement de probabilité nulle).*

— *Si on a $\mathbb{P}(B) = 1$, on dit alors que B est un événement* ***vrai*** $\mathbb{P}$***-presque-sûrement*** *(un événement de probabilité 1).*
Cela veut dire que l'ensemble des $\omega \in \Omega$ tels que $\omega \notin B$ est un ensemble de probabilité nulle. (un événement négligeable).

REMARQUE 109 — *Une chose qui est importante est que pour une mesure de probabilité $\mathbb{P}$ donnée, on peut avoir plusieurs parties A telles que $\mathbb{P}(A) = 0$ (pas seulement $A = \emptyset$).*
Une partie de probabilité nulle modélise un événement qui n'arrivera jamais. Ainsi, il est important de savoir identifier les événements de probabilité nulle, afin de les écarter dans les calculs.

Pour bien comprendre ce qu'est une mesure de probabilité, et comment s'en servir, comment utiliser ses propriétés, nous allons plonger dans des exemples.
Commençons par le cas où l'ensemble Ω est fini.

2.3 Probabilités sur un ensemble fini - Calcul combinatoire

Dans ce paragraphe, nous supposons que l'ensemble Ω est fini.
Nous rappelons que dans ce cas, nous choisirons toujours $\mathcal{A} = \mathcal{P}(\Omega)$.
Comme l'ensemble Ω est fini, nous rappelons aussi qu'en posant $n = Card(\Omega)$, on pourra écrire $\Omega = \{\omega_1, \ldots, \omega_n\}$.
L'ensemble Ω est alors en bijection avec $\{1, \ldots, n\}$, donc on pourra aussi se ramener à considérer des mesures de probabilités sur $\{1, \ldots, n\}$.

2.3.1 Caractérisation des mesures de probabilités dans le cas fini

Avant de passer à d'autres exemples, nous allons démontrer la proposition suivante, qui permet de caractériser toutes les mesures de probabilité sur les ensembles finis.

PROPOSITION 110
Soit $\Omega = \{\omega_1, \cdots, \omega_n\}$ un ensemble fini.

(i) Soit $\mathbb{P}$ une mesure de probabilité sur Ω, et soit $A \in \mathcal{P}(\Omega)$. Nous avons alors

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

(ii) Soit $\mathbb{P}$ une mesure de probabilité sur Ω.
La fonction $\mathbb{P}$ est entièrement caractérisée par ses valeurs sur les singletons, c'est-à-dire par la famille des

$$\{\mathbb{P}(\{\omega_i\}),\ \omega_i \in \Omega\}.$$

(iii) Soit $(p_i)_{1 \leq i \leq n}$ une famille de nombres réels incluse dans $[0, 1]$ et telle que $\sum_{i=1}^{n} p_i = 1$.
Alors, il existe une unique mesure de probabilité $\mathbb{P}$ telle que l'on ait $p_i = P(\{\omega_i\})$ pour tout $\omega_i \in \Omega$.
On a : $\mathbb{P}(A) = \sum_{i=1}^{n} p_i . \chi_A(\omega_i) = \sum_{i=1}^{n} p_i \delta_{\omega_i}(A)$.

Preuve —

1. Comme l'ensemble Ω est fini, l'esemble A est fini.
   Donc, dans l'écriture $A = \cup_{\omega \in A}\{\omega\}$, on a une réunion finie.
   Par propriété de la mesure de probabilité $\mathbb{P}$, on a donc :
   $$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}).$$
2. Soient $\mathbb{P}$ et Q deux mesures de probabilités sur Ω telles que $\mathbb{P}(\{\omega_i\}) = Q(\{\omega_i\})$, pour tout $1 \leq i \leq n$.
   Alors, pour toute partie A dans $\mathcal{P}(\Omega)$, on a
   $$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A} Q(\{\omega\}) = Q(A).$$
   On en déduit donc que l'on a $\mathbb{P} = Q$.

3. Soit $(p_i)_{1 \leq i \leq n}$ avec $p_i \in [0,1]$ et $\sum_{i=1}^{n} p_i = 1$.
On définit la fonction $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ par $\mathbb{P}(A) = \sum_{i=1}^{n} p_i \chi_A(\omega_i)$.
• On remarque en premier lieu que $\mathbb{P}$ est bien définie : on a toujours $\mathbb{P}(A) \in [0,1]$.
• On a bien $\mathbb{P}(\{\omega_i\}) = p_i$ par construction de $\mathbb{P}$.
• On a bien $\mathbb{P}(\Omega) = \sum_i p_i = 1$, donc $\mathbb{P}$ vérifie la condition $(*)$.
• Pour la preuve de la propriété $(**)$, cela est identique à la preuve du même point pour la mesure de proba. uniforme.
• La preuve de l'unicité vient du point (ii).

□

Remarque 111 —
— *Ainsi, pour définir une mesure de probabilité sur $\{1, \ldots, n\}$, il faut et suffit simplement de choisir des nombres réels $p_1, \ldots, p_n$ dans $[0,1]$ et dont la somme vaut 1.*
Avec cette proposition, nous pouvons donc construire très facilement toutes les mesures de probabilités que l'on veut sur un ensemble de mesure finie.
La partie difficile est ensuite le calcul de la probabilité $\mathbb{P}(A)$ pour une partie A donnée.
— ***Remarque de notation :*** *Pour une mesure de probabilité $\mathbb{P}$ et pour $\omega \in \Omega$, on calcule la probabilité du singleton $\{\omega\}$.*
La fonction $\mathbb{P}$ est définie sur $\mathcal{P}(\Omega)$ et pas sur Ω, donc parler de "$P(\omega)$" n'a pas de sens !

Exemple 112 — *Loi de Bernoulli de paramètre $p \in [0,1]$: $\mathcal{B}(p)$*
On prend un ensemble Ω à deux éléments, et un nombre réel $p \in [0,1]$. On pose

$$\Omega = \{\omega_1, \omega_2\} \text{ et } p_{\omega_1} = p,\ p_{\omega_2} = 1 - p.$$

Le singleton $\{p_1\}$ sera ainsi de probabilité p, et le singleton $\{p_2\}$ de probabilité $1-p$.
La mesure de probabilité associée modélise en particulier la chance pour une pièce de tomber sur Pile (ou Face) dans un jeu de pile ou face.
Dans ce cas $\Omega = \{P, F\}$ peut être assimilé à $\{0,1\}$.
• *Pour un lancer de pièce avec une pièce "équilibrée", le nombre réel p sera égal à $\frac{1}{2}$.*
On retrouve alors une mesure de probabilité uniforme.

Définition 113
Soit $\Omega = \{\omega_1, \ldots, \omega_n\}$ un ensemble fini, et $\mathbb{P}$ une mesure de probabilité sur Ω.
L'ensemble $\{p_i = \mathbb{P}(\{\omega_i\}),\ \omega_i \in \Omega\}$ est appelé ***loi de probabilité*** *de $\mathbb{P}$.*

Nous avons vu que la loi de probabilité de $\mathbb{P}$ caractérise la fonction $\mathbb{P}$.
Nous reparlerons plus tard dans le cours des lois de probabilités et de leur intérêt.

2.3.2 Modèles de tirages dans une urne

Un grand exemple de mesure de probabilité est celui de l'urne contenant des boules de couleur.
Le modèle général est le suivant : Une urne contient N boules de k couleurs différentes, réparties en N_1 boules de couleur 1, N_2 boules de couleur 2, ..., N_k boules de couleur k.
Nous appelons

$$p_i = \frac{N_i}{N}$$

la proportion de boules de couleur i.
Tirons au hasard n boules de cette urne, $n \leq N$, et intéressons-nous à la répartition des couleurs dans l'échantillon obtenu.
Nous notons par $P_{n_1 n_2 \cdots n_k}$ la probabilité d'obtenir n_1 boules de couleur 1, n_2 boules de couleur 2,... , n_k boules de couleur k, avec bien sûr

$$n_1 + n_2 + \ldots + n_k = n.$$

Nous allons considérer trois façons de tirer les boules au hasard : tirage avec remise, tirage sans remise, tirage simultané.
Nous verrons que chaque tirage donnera lieu à un calcul de probabilité et à un résultat différent.

Remarque 114 — *Les modèles d'urnes correspondent au cas où l'on souhaite récolter des données en statistiques.*
Le problème du choix de la façon de tirer un échantillon est à chaque fois présent.

Tirage exhaustif ou simultané - La loi hypergéométrique
Nous tirons les n boules en même temps.
L'ensemble Ω est alors l'ensemble de toutes les parties possibles de n éléments distincts, et le nombre de résultats possibles est $\binom{N}{n}$.
Le nombre de cas favorables donnant la bonne répartition des couleurs est

$$\binom{N_1}{n_1}\cdots\binom{N_k}{n_k}.$$

Donc la probabilité recherchée est

$$\widehat{p}_{n_1 n_2 \cdots n_k} = \frac{\binom{N_1}{n_1}\cdots\binom{N_k}{n_k}}{\binom{N}{n}}$$

Cette loi de probabilité (ou distribution) s'appelle la distribution polygéométrique.
Dans le cas de deux couleurs elle vaudra

$$\widehat{p}_{n_1, n-n_1} = \frac{\binom{N_1}{n_1}\binom{N-N_1}{n-n_1}}{\binom{N}{n}},$$

et on l'appelle aussi loi (ou distribution) hypergéométrique.

Exemple 115 — *Ainsi, si dans une usine de fabrication de pièces, nous savons que parmi N pièces usinées il y en a M qui sont à mettre au rebut, et si nous choisissons au hasard et simultanément un échantillon de n pièces, alors la probabilité pour que cet échantillon contienne k pièces défectueuses est*

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{M}}.$$

Tirage avec remise - La loi binomiale

Les tirages sont successifs. Nous replaçons la boule tirée dans l'urne avant le tirage suivant. Nous pouvons donc tirer plusieurs fois la même boule.
L'ensemble Ω est alors l'ensemble de tous les n-uplets d'éléments de l'urne.
Comme on a N boules au total, on a donc $Card(\Omega) = N^n$.
Nous munissons l'ensemble Ω de sa probabilité uniforme.
Le nombre de façons de déterminer les places des k couleurs parmi n est égal au nombre de façons de partager le nombre n en k parties de tailles n_i. Cela est donc égal au coefficient multinomial $\dfrac{n!}{n_1!n_2!...n_k!}$.
Une fois la place des k couleurs choisies, nous avons N_i possibilités pour chaque boule de couleur i.
Le nombre de n-uplets qui donnent la répartition n_1, n_2, ..., n_k est alors égal à

$$\frac{n!}{n_1!n_2!...n_k!}N_1^{n_1}\cdots N_k^{n_k}.$$

Nous avons donc finalement

$$P_{n_1 n_2 \cdots n_k} = \frac{n!}{n_1!n_2!...n_k!}\frac{N_1^{n_1}\cdots N_k^{n_k}}{N^n}$$

Cette mesure de probabilité est appelée une **distribution multinomiale**.
Dans le cas particulier où $k = 2$, on pose $p = \dfrac{N_1}{N} = p_1$. La probabilité vaudra alors

$$p_{n_1, n-n_1} = \binom{n}{n_1} p^{n_1}(1-p)^{n-n_1}.$$

On l'appelle probabilité **binomiale de paramètres** n **et** p : $\mathcal{B}(n,p)$.
Cette mesure de probabilité ne dépend pas du nombre de boules N, juste de la proportion p de boules de couleur n°1.

REMARQUE 116 — *Pour une mesure probabilité binomiale de paramètres n et p, les paramètres n et p ne sont pas interchangeables : $n \in \mathbb{N}$, $p \in [0,1]$.*
On a $n+1$ répartitions de couleurs possibles dans un tirage de n boules (de 0 boules de couleur 1 à n boules de couleur 1).

Tirage sans remise

Nous tirons maintenant successivement les boules de l'urne, mais sans les replacer dans l'urne après tirage.
L'ensemble Ω est alors l'ensemble des familles de n éléments distincts parmi N, et le nombre de cas possibles sera égal au nombre d'arrangements :

$$N(N-1)\cdots(N-n+1) = A_N^n.$$

En raisonnant comme dans le cas avec remise, nous pouvons montrer que le nombre de cas favorables vaut

$$\frac{n!}{n_1!n_2!...n_k!}A_{N_1}^{n_1}...A_{N_k}^{n_k} = \frac{\binom{N_1}{n_1}\cdots\binom{N_k}{n_k}}{\binom{N}{n}}.$$

Cela donne finalement la même mesure de probabilité que celle du cas de tirage simultané.
Ainsi, il y a équivalence du tirage sans remise et du tirage simultané, du point de vue de la composition de l'échantillon. On peut donc se permettre de ne pas prendre en compte l'ordre des individus dans un tel tirage.

Cas d'une urne dont le nombre de boules est infini

Nous nous plaçons dans les hypothèses du tirage simultané, avec $k=2$ couleurs. On suppose de plus que N et N_1 tendent vers l'infini, de telle manière que $\dfrac{N_1}{N}$ converge vers un nombre réel $0 < p < 1$.
On a vu que la mesure de probabilité du tirage sans remise est :

$$\widehat{p}_{n_1,n-n_1} = \frac{\binom{N_1}{n_1}\binom{N-N_1}{n-n_1}}{\binom{N}{n}}$$

Fixons n_1 le nombre de boules de couleur 1 que l'on veut trouver dans le tirage. En supposant que N_1 tend vers l'infini, on a

$$\binom{N_1}{n_1} = \frac{N_1(N_1-1)\cdots(N_1-n_1+1)}{n_1!} \sim_{N_1\to+\infty} \frac{N_1^{n_1}}{n_1!}$$

On en déduit donc que

$$\widehat{p}_{n_1,n-n_1} \underset{N\to\infty,\frac{N_1}{N}\to p}{\sim} \binom{n}{n_1}p^{n_1}(1-p)^{n-n_1}$$

REMARQUE 117 — *Ce résultat est intuitivement naturel car si le nombre de boules devient aussi grand que l'on veut (avec des proportions de couleurs qui restent "stables"), les tirages de boules avec ou sans remise deviennent presque équivalents : on a une chance très infime de tomber deux fois sur la même boule.*

Exemples

EXEMPLE 118 — *Les yeux bandés, vous manipulez au hasard 7 fiches où sont écrites les lettres E, E, T, B, R, L, I. Quelle est la probabilité que vous écriviez le mot LIBERTE ?*
Solution : $\frac{\text{nb de cas favorables}}{\text{nb de cas possibles}} = \dfrac{2}{7!} = \dfrac{1}{2520}$
Le fait de dire "au hasard", et de dire que l'on manipule toutes les fiches de la même façon indique que la mesure de probabilité que l'on choisit pour modéliser l'expérience est la mesure uniforme.

EXEMPLE 119 — *On tire au hasard quatre cartes d'un jeu de cinquante-deux cartes.*
Quelle est la probabilité pour que, parmi ces quatre cartes, il y ait exactement deux rois ?
Solution : L'hypothèse au hasard amène à modéliser cette expérience comme un tirage uniforme dans un certain ensemble Ω qu'il faut préciser.
Ici, on prend pour Ω l'ensemble des parties à 4 éléments de l'ensemble de 52 cartes. Le cardinal de Ω est donc $\binom{52}{4}$, et $\mathbb{P}$ est la probabilité uniforme sur Ω.
Les résultats favorables sont les tirages qui contiennent exactement 2 rois, à savoir 2 rois et 2 cartes parmi les 48 cartes autres que des rois. Ainsi, la probabilité cherchée vaut $\frac{\textit{nb de cas favorables}}{\textit{nb de cas possibles}} = \frac{\binom{4}{2}\binom{48}{2}}{\binom{52}{4}}$.

EXEMPLE 120 — *On lance trois dés parfaitement équilibrés.*
Montrer que la probabilité que la somme des points dépasse strictement dix est égale à la probabilité que cette somme ne dépasse pas dix. (Cela permettra de construire un jeu parfaitement équitable.)
Solution : L'ensemble Ω est ici l'ensemble des familles (a_1, a_2, a_3) de 3 nombres compris entre 1 et 6, $\{1, .., 6\}^3$, muni de la probabilité $\mathbb{P}$ uniforme.
On remarque que

$$a_1 + a_2 + a_3 > 10 \iff (7 - a_1) + (7 - a_2) + (7 - a_3) \leq 10.$$

Ainsi, si A désigne l'événement "la somme des points est strictement supérieure à 10", nous remarquons que l'application $(a_1, a_2, a_3) \to (7 - a_1, 7 - a_2, 7 - a_3)$ est une bijection de A sur $\overline{A}$.
Les événements A et $\overline{A}$ ont donc même cardinal, et donc même probabilité de réalisation (qui vaut donc $\frac{1}{2}$).

REMARQUE 121 — *Une difficulté majeure dans ce genre de calculs combinatoires est de bien préciser le modèle probabiliste (l'ensemble et la mesure de probabilité que l'on choisit).*
De célèbres paradoxes sont nés de cette difficulté.

EXEMPLE 122 — *Rappelons le problème du chevalier de Méré. Ce personnage marquant de la cour de Louis XIV qui "avait très bon esprit, mais n'était pas très bon géomètre" (cf. lettre de Pascal à Fermat du 29 juillet 1654) était un joueur impénitent, toujours à la recherche de règles cachées lui permettant d'avoir un avantage sur ses adversaires. Voici deux de ses règles.*

1. *Il est avantageux de parier sur l'apparition d'au moins un 6 en lançant un dé 4 fois de suite.*

   Cette règle est bonne puisque la probabilité de l'événement qui nous intéresse vaut

   $$1 - \left(\frac{5}{6}\right)^4 \simeq 0.5177 > \frac{1}{2}.$$

   La différence avec $\dfrac{1}{2}$ est faible, mais apte à fournir à long terme des gains assurés : le chevalier devait jouer souvent...
2. *Il est avantageux de parier sur l'apparition d'au moins un double-six en lançant deux dés 24 fois de suite.*

   Cette règle est mauvaise, puisque la probabilité de l'événement cherché vaut :

   $$1 - \left(\frac{35}{36}\right)^{24} = 0.4914 < \frac{1}{2}.$$

   Le Chevalier était donc moins heureux avec cette règle qu'avec la précédente. En fait, il s'était laissé abuser par un soi-disant argument d'homothétie : en lançant un dé, il y a 6 résultats possibles, en lançant deux dés, il y en a $6^2 = 36$, soit 6 fois plus.
   Comme il est avantageux de parier sur l'apparition d'au moins un 6 en lançant un dé 4 fois de suite, il doit être avantageux de parier sur l'apparition d'un double-six en lançant deux dés $4 \times 6 = 24$ fois de suite.

2.4 Probabilités sur un ensemble dénombrable

On suppose dans cette section Ω est dénombrable.
L'ensemble est en bijection avec $\mathbb{N}$, donc nous pouvons numéroter ses éléments : $\Omega = \{\omega_i,\ i \in \mathbb{N}\}$.

La proposition suivante généralise au cas dénombrable la proposition vue dans le cas fini.
On rappelle que la sigma-algèbre considérée sur Ω est l'ensemble $\mathcal{P}(\Omega)$.

PROPOSITION 123
Soit Ω un ensemble dénombrable.

1. *Soit $(p_n)_{n\geq 0}$ une suite de nombres réels tels que $0 \leq p_n \leq 1$ et $\sum_{n=0}^{+\infty} p_n = 1$.*
   Alors il existe une unique mesure de probabilité $\mathbb{P}$ telle que pour tout $A \subset \Omega$, on ait
   $$\mathbb{P}(A) = \sum_{\omega_n \in A} p_n = \sum_{\omega_n \in A} \mathbb{P}(\{\omega_n\}).$$
2. *Soit $\mathbb{P}$ une probabiité sur $(\Omega, \mathcal{P}(\Omega))$.*
   La fonction $\mathbb{P}$ est entièrement caractérisée par ses valeurs sur les singletons, c'est-à-dire par la famille des
   $$\{\mathbb{P}(\{\omega_i\}, \, i \geq 0\}.$$

Preuve —

1. **Existence :** On définit la fonction $\mathbb{P}$ sur $\mathcal{P}(\Omega)$ par $\mathbb{P}(A) = \sum_{\omega_n \in A} p_n$.
   Rappelons que cette somme infinie vaut :
   $$\sum_{\omega_n \in A} p_n = \sup_{B \subset A, \ |B| < \infty} \sum_{\omega_n \in B} p_n$$
   Cette somme de termes positifs est majorée par $\sum_{n=0}^{+\infty} p_n = 1$, donc elle est finie.
   • Ainsi, la fonction $\mathbb{P}$ est bien définie.
   • On remarque que l'on a $\mathbb{P}(\Omega) = \sum_{n=0}^{+\infty} p_n = 1$. Donc la propriété $(*)$ est vérifiée.
   Si A est un ensemble fini, on en déduit par additivité de $\mathbb{P}$ que
   $$\mathbb{P}(A) = \sum_{\omega_n \in A} \mathbb{P}(\{\omega_n\}).$$
   Enfin, si $A \subset \Omega$ est dénombrable, alors A correspond à une suite extraite $(\omega_{\rho(n)})_{n\geq 0}$ de $(\omega_n)_{n\geq 0}$.
   Par σ-additivité
   $$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{n=0}^{+\infty} \{\omega_{\rho(n)}\}\right) = \sum_{n=0}^{+\infty} p_{\rho(n)}.$$
   Ceci montre que si $\mathbb{P}$ existe, elle est uniquement déterminée. Il reste à montrer que $\mathbb{P}$ vérifie bien les axiomes d'une probabilité :
   • Soit $(A_n)_{n\geq 0}$ une famille dénombrable d'éléments de $\mathcal{A}$ deux à deux disjoints. On note A leur union.
   $$\mathbb{P}(A) = \sum_{\omega_n \in A} p_n = \sum_{k=0}^{+\infty} \left(\sum_{\omega_n \in A_k} p_n \right) = \sum_{k \geq 0} \mathbb{P}(A_k).$$
   Cela découle du théorème de sommation par paquets (voir chapitre Dénombrement). Ainsi la fonction $\mathbb{P}$ vérifie la propriété $(**)$, c'est donc une mesure de probabilité. **Unicité :** Si on avait $\mathbb{P}$ et Q deux mesures de probabilité telles que
   $$\mathbb{P}(A) = \sum_{\omega_n \in A} p_n = Q(A),$$
   alors on a $\mathbb{P} = Q$ car ces fonctions sont égales sur $\mathcal{P}(\Omega)$.
2. On pose $p_n = \mathbb{P}(\{\omega_n\})$.
   Alors la famille $(p_n)_{n\geq 0}$ vérifie les conditions du point 1).
   Par unicité de la mesure de probabilité associée aux $(p_n)_n$, on en déduit que pour toute partie $A \in \mathcal{P}(\Omega)$, on a
   $$\mathbb{P}(A) = \sum_{\omega_n \in A} \mathbb{P}(\{\omega_n\}),$$
   ce qui montre que $\mathbb{P}$ est entièrement déterminée par sa valeur en les singletons.

□

EXEMPLE 124 — *Soit* $\theta > 0$ *et* $p_n = e^{-\theta}\dfrac{\theta^n}{n!}$.
Il est facile de vérifier que $0 \leq p_n \leq 1$ *et que*

$$\sum_{n=0}^{+\infty} p_n = e^{-\theta} \sum_{n=0}^{+\infty} \frac{\theta^n}{n!} = 1$$

La suite $(p_n)_{n\in\mathbb{N}}$ *définit une probabilité sur* $\mathbb{N}$, *appelée* ***loi de Poisson de paramètre*** θ : $\mathcal{P}(\theta)$.

EXEMPLE 125 — *Soit* $\Omega = \Omega = \{\omega_0, \omega_1, \cdots, \omega_n, \cdots\}$ *un ensemble dénombrable* $\mathbb{P}$ *une mesure probabilité sur* Ω, *et* $p_n = \mathbb{P}(\{\omega_n\})$. *Alors pour tout* $A \in \mathcal{A}$, *on a*

$$\mathbb{P}(A) = \sum_{n\in\mathbb{N}} p_n \delta_{\omega_n}(A).$$

La mesure de probabilité $\mathbb{P}$ *peut donc s'écrire comme somme de la série de fonctions :*

$$\mathbb{P} = \sum_{n=0}^{+\infty} p_n \delta_{\omega_n}.$$

Ainsi, toute mesure de probabilité sur un ensemble fini ou dénombrable peut s'écrire comme une "combinaison convexe" de mesures de Dirac. (une combinaison convexe est une combinaison linéaire avec tous les coefficients positifs ou nuls, et dont la somme vaut 1*)*

2.5 Conditionnement et indépendance

2.5.1 Probabilités conditionnelles

La notion de conditionnement est l'une des plus fructueuses de la théorie des probabilités (de regarder des résultats "à condition que", "sachant que").
L'idée de base qui permet de comprendre de cette notion est la suivante : une information supplémentaire sur l'expérience modifie la vraisemblance que l'on accorde à l'événement étudié.

EXEMPLE 126 — *Cherchons, pour un lancer de deux dés équilibrés, la probabilité de l'événement "la somme est supérieure ou égale à 10". Elle vaut*

- $\dfrac{1}{6}$ *sans information supplémentaire.*
- $\dfrac{1}{2}$ *si l'on sait que le résultat du second dé est* 6
- 0 *si l'on sait a priori que le résultat d'un des dés est* 2.

Pour obtenir ces résultats, nous avons dans chaque cas calculé le rapport du nombre de résultats favorables sur le nombre de cas possibles.
Il est à chaque fois indispensable de bien définir l'espace probabilisé associé à l'expérience en tenant compte des informations a priori.
On remarque que l'information a priori change la valeur de la probabilité de l'événement aléatoire. (c'est le même événement, mais regardé ici sur des espaces probabilisés un peu différents)

L'approche pour donner un sens mathématique à cette notion se base à nouveau sur la notion de fréquence d'apparition.
Cela donne la définition suivante.

DÉFINITION 127
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé.*
Soit $A, B \in \mathcal{A}$ *deux événements, avec* $\mathbb{P}(B) > 0$.
On appelle la ***probabilité conditionnelle*** *de* A *sachant* B *le nombre*

$$\mathbb{P}(A|B) = \mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Pour B donné, la fonction $\mathbb{P}(.|B)$ définit encore une mesure de probabilité.

Proposition 128
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé.*
Soit $B \in \mathcal{A}$ *une partie telle que* $\mathbb{P}(B) > 0$. *Alors,*

1. *La fonction* $\mathbb{P}(.|B) : A \in \mathcal{A} \mapsto \mathbb{P}(A|B) = \frac{\mathbb{P}(A)}{\mathbb{P}(B)} \in [0,1]$ *est une mesure de probabilité sur* Ω.
   On l'appelle ***mesure de probabilité conditionnelle*** *sachant* B.
2. *Soit* $A \in \mathcal{A}$. *Si* $\mathbb{P}(A) > 0$ *et* $\mathbb{P}(B) > 0$, *on a*

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

Preuve — • Comme $\mathbb{P}(B) > 0$, il est clair que $0 \leq \mathbb{P}(A|B) \leq 1$, donc $\mathbb{P}(.|B)$ est bien définie.
• De même, on a $\mathbb{P}(\Omega|B) = 1$.
• Soit $(A_n)_{n\geq 0}$ une famille d'élements de $\mathcal{A}$ qui sont deux à deux disjoints.
Alors $(A_n \cap B)_{n\geq 0}$ est encore une famille d'éléments de $\mathcal{A}$ deux à deux disjoints.
Par σ additivité de $\mathbb{P}$, on obtient

$$\mathbb{P}(\bigcup_{n\geq 0} A_n|B) = \frac{\mathbb{P}\left(\bigcup_{n\geq 0}(A_n \cap B)\right)}{\mathbb{P}(B)} = \frac{\sum_{n\geq 0}\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n\geq 0}\mathbb{P}(A_n|B).$$

Cela montre bien que $\mathbb{P}(.|B)$ est une mesure de probabilités.
Le point 2) découle de la définition de la probabilité conditionnelle. □

Proposition 129 (Formule des probabilités composées)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé.*
Soient $A_1, \ldots, A_n \in A$ *tels que* $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$.
Alors, on a

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)\cdots\mathbb{P}(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

Preuve — Par hypothèse, on a $\mathbb{P}(A_1 \cap \ldots \cap A_k) > 0$ pour tout $1 \leq k \leq n$.
On peut alors écrire le télescopage suivant :

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \frac{\Pi_{i=1}^{n}\mathbb{P}(A_1 \cap \ldots \cap A_i)}{\Pi_{j=1}^{n-1}\mathbb{P}(A_1 \cap \ldots \cap A_j)} = \mathbb{P}(A_1)\Pi_{k=2}^{n}\frac{\mathbb{P}(A_1 \cap \ldots \cap A_k}{\mathbb{P}(A_1 \cap \ldots A_{k-1})}.$$

Cela donne le résultat. □

Pour $A, B \in \mathcal{A}$ deux événéments sur Ω, regarder la probabilité conditionnelle de A sachant B ne donne qu'une information partielle sur la probabilité de A.
Avec suffisamment d'événements B bien choisis (en utilisant des partitions), on peut arriver à retrouver $\mathbb{P}(A)$.

Définition 130
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé.*
Soit $(B_i)_{i\in I}$ *une partition finie ou dénombrable de* Ω, *telle que* $B_i \in A$ *et* $\mathbb{P}(B_i) > 0$ *pour chaque* i.
Une telle famille est appelée ***système complet d'évènements***.

Proposition 131 (**Formule des probabilités totales**)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé, et soit* $(B_i)_{i\in I}$ *un système complet d'événements de* Ω.
Alors, pour tout $A \in \mathcal{A}$, *on a*

$$\mathbb{P}(A) = \sum_{i\in I}\mathbb{P}(A \cap B_i) = \sum_{i\in I}\mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Preuve — Nous avons $A = \bigcup_{i\in I}(A \cap B_i)$. Par hypothèse, les ensembles $(A \cap B_i)$ sont deux-à-deux disjoints, et par ailleurs $\mathbb{P}(A \cap B_i) = \mathbb{P}(A|B_i)\mathbb{P}(B_i)$. La σ-additivité donne le résultat. □

Théorème 132 (**Formule de Bayes**)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé, et soit* $(B_i)_{i\in I}$ *un système complet d'événements de* Ω.
Soit $A \in \mathcal{A}$ *tel que* $\mathbb{P}(A) > 0$. *Alors, on a*

$$\mathbb{P}(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j\geq 0}\mathbb{P}(A|B_j)P(B_j)},\ \forall i \in I.$$

Preuve — On sait d'après la proposition précédente que $\sum_{j\geq 0} \mathbb{P}(A|B_j)P(B_j) = \mathbb{P}(A)$, et par définition d'une probabilité conditionnelle on a

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A\cap B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)}.$$

□

Remarque : Notre intuition habituelle est très mauvaise quand il s'agit d'estimer certaines probabilités conditionnelle !
La formule de Bayes est le résultat qui permet de mettre cela en évidence.

Exemple 133 — *Un individu est tiré au hasard dans une population où l'on trouve une proportion* 10^{-4} *de séropositifs. On lui fait passer un test de détection de la séropositivité.*
Par ailleurs, des essais antérieurs ont permis de savoir que la probabilité d'avoir un résultat positif lors du test si l'individu est séropositif est $0,99$ *(c'est la sensibilité du test, la proba de vrais positifs), et que celle d'avoir un résultat positif si l'individu n'est pas séropositif est de* $0,001$ *(*$0,999 = 1 - 0,001$ *est la spécificité du test, la proba de vrais négatifs).*
Sachant que le test donne un résultat positif, quelle est la probabilité pour que l'individu soit vraiment séropositif ?
Solution : Considérons les événements A "l'individu est séropositif", et B "le test de détection donne un résultat positif".
Les données de l'énoncé fournissent $\mathbb{P}(A) = 10^{-4}$*, donc* $\mathbb{P}(\overline{A}) = 0,9999$*, ainsi que* $\mathbb{P}(B|A) = 0,99$ *(vrai positif) et* $\mathbb{P}(B|\overline{A}) = 0,001$ *(faux positif).*
Nous trouvons alors

$$\begin{aligned}\mathbb{P}(A|B) &= \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}\\ &= \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A)+\mathbb{P}(B|\overline{A})\mathbb{P}(\overline{A})}\\ &= \frac{0,99\times 10^{-4}}{0,99\times 10^{-4}+0,001\times 0,9999} \simeq 0,09\end{aligned}$$

Contrairement à l'intuition, cette probabilité est plutôt faible ($0.09 = \frac{9}{100} = 9\%$*).*
La proportion de gens séropositifs est très faible, ce qui fait que même si le test détecte très bien la maladie, le volume de faux positifs est finalement bien plus important que le volume de vrais positifs.
Dans cette population, ce test, bien que très efficace, n'est pas extrêmement fiable (le test seul n'est pas suffisant pour vraiment savoir si on est séropositif ou pas).

Exemple 134 — *On classe les gérants de portefeuilles en deux catégories, les bien informés et les autres. Lorsqu'un gérant bien informé achète une valeur boursière pour son client, on peut montrer par une étude préalable que la probabilité que le cours de cette valeur monte est de* $0,8$.
Si le gérant est mal informé, la probabilité que le cours descende est de $0,6$.
On sait par ailleurs que si l'on choisit au hasard un gérant de portefeuille, il y a une chance sur 10 *que celui-ci soit un gérant bien informé.*
Un client choisit au hasard un gérant dans l'annuaire, et lui demande d'acheter une valeur. Sachant que le cours de cette valeur est monté, cherchons la probabilité pour que le gérant soit mal informé.
Solution : Notons M l'événement "la valeur monte" et I l'événement "le gérant est bien informé".
Par la formule des probabilités totales, la probabilité que la valeur monte vaut

$$\mathbb{P}(M) = \mathbb{P}(M|I)\mathbb{P}(I) + \mathbb{P}(M|\overline{I})\mathbb{P}(\overline{I}) = 0,8\times 0,1 + 0,4\times 0,9 = 0,44.$$

La formule de Bayes donne alors

$$\mathbb{P}(\overline{I}|M) = \frac{\mathbb{P}(M|\overline{I})\mathbb{P}(\overline{I})}{\mathbb{P}(M)} = \frac{0,4\times 0,9}{0,44} \simeq 0,818.$$

2.5.2 Evénements indépendants

La notion d'indépendance est un outil absolument fondamentale en probabilités.

Intuitivement, deux événements A et B sont indépendants si le fait de savoir que A est réalisé ne donne aucune information sur la réalisation de B, et réciproquement.
L'indépendance est modélisée mathématiquement par cette définition.

Définition 135
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé, et* $A, B \in \mathcal{A}$ *deux événements.*
On dit que les événements A *et* B ***sont indépendants*** *si* $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Remarque 136 —

1. *Si* $\mathbb{P}(A) > 0$ *et* $\mathbb{P}(B) > 0$*, alors*
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \Leftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A) \Leftrightarrow \mathbb{P}(B|A) = \mathbb{P}(B).$$
Ainsi, si A *est indépendant de* B*, la probabilité de voir* A *réalisé ne dépend pas de la réalisation de* B*, et réciproquement.*
2. *La notion d'indépendance est une notion qui dépend totalement de la mesure de probabilité* $\mathbb{P}$.
Cette notion n'a rien à voir avec les opérations ensemblistes dans $\mathcal{P}(\Omega)$.
Par exemple, cela n'a rien à voir avec le fait que A *et* B *soient disjoints ou non. (Cf. Exemple ci-dessous).*

Exemple 137 —

1. *On lance* 3 *fois de suite un dé équilibré.*
Si A_i *est un événement qui ne dépend que du* i*-ème lancer, alors* A_1, A_2, A_3 *sont indépendants (pour la mesure uniforme).*
2. *Si deux événements* A *et* B *sont disjoints mais pas de probabilité nulle, on a alors* $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$ *et* $\mathbb{P}(A)\mathbb{P}(B) > 0$.
Donc A *et* B *ne sont pas indépendants pour la mesure de probas* $\mathbb{P}$*. (par exemple* $A = \{$*faire Pile*$\}$ *et* $B = \{$*faire Face*$\}$ *dans un jeu de Pile ou Face équilibré)*
3. *On tire une carte au hasard dans un jeu de* 52 *cartes.*
$$A = \{\textit{la carte est une dame}\}\,; \quad B = \{\textit{la carte est un coeur}\}.$$
Il est facile de voir que $\mathbb{P}(A) = \frac{4}{52}$, $\mathbb{P}(B) = \frac{13}{52}$*, et*
$$\mathbb{P}(A \cap B) = \mathbb{P}(\{\textit{la carte est la dame de coeur}\}) = \frac{1}{52} = \mathbb{P}(A)\mathbb{P}(B).$$
Ainsi, les événements A *et* B *sont indépendants pour la mesue uniforme* $\mathbb{P}$.
4. *Supposons maintenant que le jeu de cartes soit trafiqué.*
Soit $\mathbb{Q}$ *la nouvelle mesure de probabilité correspondant au tirage de cartes. Supposons que*
$$\mathbb{Q}(\{\textit{dame de coeur}\}) = \frac{1}{2},\ \mathbb{Q}(\{\textit{autre carte}\}) = \frac{1}{2} \times \frac{1}{51} = \frac{1}{102}.$$
Alors
$$\mathbb{Q}(A \cap B) = \frac{1}{102} \neq \mathbb{Q}(A)\mathbb{Q}(B) = \frac{2}{51} \times \frac{13}{102}.$$
Les événements A *et* B *ne sont pas indépendants pour la mesure de probas* $\mathbb{Q}$.

Proposition 138
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé, et* $A, B \in \mathcal{A}$ *deux événements.*
Si A *et* B *sont indépendants, alors il en est de même de* A *et* $\overline{B}$, $\overline{A}$ *et* B, $\overline{A}$ *et* $\overline{B}$.

Preuve — Supposons A et B indépendants.
$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) \Rightarrow \mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(\overline{B})$$
Donc A et $\overline{B}$ sont indépendants.
$$\mathbb{P}(\overline{A}) = \mathbb{P}(\overline{A} \cap \overline{B}) + \mathbb{P}(\overline{A} \cap B) \Rightarrow \mathbb{P}(\overline{A} \cap \overline{B}) = \mathbb{P}(\overline{A}) - \mathbb{P}(\overline{A})\mathbb{P}(B) = \mathbb{P}(\overline{A})\mathbb{P}(\overline{B})$$
où l'on a appliqué la première partie de la preuve à $\overline{A}$ et B.
Les deux derniers cas s'en déduisent immédiatement. □

La notion d'indépendance se généralise à une famille finie ou dénombrable d'événements de la manière suivante.

Définition 139
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé, et* $(A_n)_{n\geq 0} \in \mathcal{A}^{\mathbb{N}}$ *une famille d'événements.*
On dit que cete famille est ***indépendante*** *si*

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$$

pour toute famille finie finie $(i_1, \cdots, i_k)$ *d'entiers, avec* $i_1 < i_2 < \ldots < i_k$.

Remarque 140 — *Il faut faire très attention avec cette définition.*

1. *Pour que la suite* (A, B, C) *soit indépendante, la propriété doit être vérifiée pour toutes les intersections de deux ensembles et l'intersection des* 3 *ensembles. Il ne suffit pas d'avoir*
$$\mathbb{P}(A \cap B \cap C) = P(A)P(B)P(C).$$
Par exemple, prenons un lancer de 1 dé avec $A = \{1, 2, 3\}$, $B = \{2, 4, 6\}$ *et* $C = \{1, 2, 4, 5\}$.
Nous avons $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{1}{2}$, $\mathbb{P}(C) = \frac{2}{3}$.
Ainsi, nous avons bien $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$, *mais* $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B)$, *donc la famille* (A, B, C) *n'est pas indépendante.*
2. *Il ne suffit pas non plus que les événements soient indépendants deux à deux.*
Par exemple, on joue 2 fois à Pile ou Face (avec pièce équilibrée) et on considère les événements $A = \{$ *Face au premier lancer* $\}$, $B = \{$ *Face au deuxième lancer* $\}$ *et* $C = \{$ *les deux tirages donnent le même résultat* $\}$.
On vérifie que ces événements sont deux à deux indépendants, mais par contre on a $\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$, *donc la famille* (A, B, C) *n'est pas indépendante.*

2.5.3 Lemme de Borel-Cantelli

On termine ce chapitre par un résultat aux conséquences vraiment surprenantes et qui permet de mieux comprendre la notion d'indépendance.
Nous commençons par définir la limite supérieure et inférieure d'une suite de parties d'un ensemble.

Définition 141
Soient Ω *un ensemble,* $\mathcal{A}$ *une* σ*-algèbre, et* $(A_n)_{n\in\mathbb{N}}$ *une suite d'éléments de* $\mathcal{A}$.
On définit la ***limite supérieure*** *de la famille* $(A_n)_n$ *comme l'ensemble*

$$\limsup_n A_n = \bigcap_p \left(\bigcup_{n\geq p} A_n \right) \in \mathcal{A},$$

et la ***limite inférieure*** *de la famille* $(A_n)_n$ *comme l'ensemble*

$$\liminf_n A_n = \bigcup_p \left(\bigcap_{n\geq p} A_n \right) \in \mathcal{A}.$$

Remarque 142 —

1. *L'ensemble* $\limsup_n A_n$ *est l'ensemble des* ω *qui apparaissent une infinité de fois parmi les* A_n.
Inversement, on a $\omega \notin \limsup_n A_n$ *ssi* ω *appartient à au plus un nombre fini de* A_n.
On peut remarquer que la suite des $(\bigcup_{n\geq p} A_n)_p$ *est une suite décroissante.*
L'ensemble $\limsup_n(A_n)$ *est donc une limite de suite décroissante pour l'inclusion.*
2. *L'ensemble* $\liminf_n A_n$ *est l'ensemble des* ω *qui apparaissent dans tous les* A_n *à partir d'un certain rang* p *(pour tout* $n \geq p$*).*
On peut remarquer que la suite des $(\bigcap_{n\geq p} A_n)_p$ *est croissante.*
L'ensemble $\liminf_n(A_n)$ *est donc une limite de suite croissante pour l'inclusion.*
3. *On peut aussi remarquer que* $\liminf_n(A_n) \subset \limsup_n(A_n)$.
Ces ensembles ne sont en général pas égaux.

4. *Comme le passage au complémentaire change les unions en intersections, et les intersections en unions, on montre facilement que* $\liminf_n(A_n) = \limsup_n(\overline{A_n})$.

Le lemme de Borel-Cantelli nous permet de déterminer facilement la probabilité de $\limsup_n(A_n)$ dans certaines situations.

THÉORÈME 143 (**Lemme de Borel-Cantelli**)
Soient Ω *un ensemble,* $\mathcal{A}$ *une* σ*-algèbre, et* $(A_n)_{n\in\mathbb{N}}$ *une suite d'éléments de* $\mathcal{A}$.

1. *Si on a* $\displaystyle\sum_{n\geq 0} \mathbb{P}(A_n) < +\infty$*, alors* $\mathbb{P}(\limsup_n A_n) = 0$.
2. *Si la famille* $(A_n)_{n\geq 0}$ *est indépendante, alors on a*
$$\sum_{n\geq 0} \mathbb{P}(A_n) = +\infty \ \textit{ implique } \ \mathbb{P}(\limsup_n A_n) = 1.$$

Preuve —

1. Comme la suite $\left(\bigcup_{n\geq p} A_n\right)_p$ est décroissante, par propriétés de la mesure $\mathbb{P}$ on a
$$\mathbb{P}(\limsup_n A_n) = \lim_{p\to+\infty} \mathbb{P}\left(\bigcup_{n\geq p} A_n\right) \leq \lim_{p\to+\infty} \sum_{n\geq p} \mathbb{P}(A_n).$$
Si la série $\displaystyle\sum_{n\geq 0} \mathbb{P}(A_n)$ est convergente, le reste de cette série tend vers 0. Donc, on a $\mathbb{P}(\limsup A_n) = 0$.
2. Supposons maintenant que les A_n soient indépendants et que la série $\displaystyle\sum_{n\geq 0} \mathbb{P}(A_n)$ diverge.
Soit m un nombre entier. Nous avons
$$\mathbb{P}\left(\bigcup_{i=p}^{m} A_i\right) = 1 - \mathbb{P}\left(\bigcap_{i=p}^{m} \overline{A_i}\right) = 1 - \prod_{i=p}^{m} \mathbb{P}(\overline{A_i}) = 1 - \prod_{i=p}^{m}(1 - \mathbb{P}(A_i)) \geq 1 - e^{-\sum_{i=p}^{m} \mathbb{P}(A_i)}$$
grâce à l'inégalité $1 - x \leq e^{-x}$ pour $x \geq 0$.
Ainsi, en passant à la limite on obtient
$$\mathbb{P}\left(\bigcup_{i=p}^{+\infty} A_i\right) \geq 1 - e^{-\sum_{i=p}^{+\infty} \mathbb{P}(A_i)} = 1.$$
On a donc $\mathbb{P}(\bigcup_{i=p}^{+\infty} A_i) = 1$ pour tout $p \geq 1$. Comme la suite des $(\bigcup_{i=p}^{+\infty} A_i)_p$ est décroissante et converge vers $\limsup_n(A_n)$, les propriétés de $\mathbb{P}$ nous donnent $\mathbb{P}(\limsup_n A_n) = \lim_{p\to+\infty} \mathbb{P}\left(\bigcup_{n\geq p} A_n\right) = 1$.

□

REMARQUE 144 —

1. *Il est clair que le point* 2) *est totalement faux dans le cas où la famille n'est pas indépendante. On peut prendre, par exemple, tous les* A_n *égaux à un même événement* A *de probabilité* $\mathbb{P}(A) \in]0,1[$.
2. *Le théorème montre que si la suite* $(A_n)_{n\geq 0}$ *est indépendante, alors* $\limsup_n A_n$ *est de probabilité* 0 *ou* 1 *suivant que la série* $\displaystyle\sum_{n\geq 0} \mathbb{P}(A_n)$ *converge ou diverge.*
3. *Le lemme de Borel-Cantelli porte sur* $\limsup_n(A_n)$. *Attention à ne pas confondre* lim sup *et* lim inf *!*
4. *On peut parfois calculer* $\mathbb{P}(\liminf_n(A_n))$ *avec Borel-Cantelli en utilisant le fait que* $\liminf_n(A_n) = \limsup_n(\overline{A_n})$.

EXEMPLE 145 — *Supposons que vous vous installez les yeux bandés devant votre clavier et que vous tapez indéfiniment (de façon dénombrable) sur les touches au hasard (probabilité uniforme).*
Prenons M *un mot de longueur* l*, et pour* $k \geq 1$ *définissons l'évènement* A_k *: "les lettres* lk *à* $k(l+1) - 1$ *forment le mot* M*"*
En prenant $\mathbb{P}$ *la mesure de probabilité uniforme sur notre ensemble dénombrable, cette famille d'événements est alors indépendante.*
On a de plus $\mathbb{P}(A_1) \geq \frac{1}{l!} > 0$*, et* $\mathbb{P}(A_k) = \mathbb{P}(A_1)$ *(la proba de* A_k *ne dépend pas de* k*).*

Cela donne donc $\sum_{k=0}^{+\infty} \mathbb{P}(A_k) = +\infty$.
On peut donc appliquer le lemme de Borel-Cantelli pour obtenir que la probabilité que le mot M *apparaisse une infinité de fois est de 1. (c'est un événement* $\mathbb{P}$*-presque sûr, et l'événement contraire est de probabilité nulle)*
Cela signifie qu'en tapant assez longtemps vous êtes certains d'écrire autant de fois que vous le voudrez le corrigé de toutes les épreuves de ECPK de toutes les matières depuis l'existence de l'école. Il faudra juste être très patient ! (on prend pour M *le mot correspondant à toute cette littérature).*

Donc courage !

Chapitre 3 Variables aléatoires

Table des matières du chapitre

3.1 Variables aléatoires

3.1.1 Définition

Maintenant que nous avons défini les espaces probabilisés $(\Omega, \mathcal{A}, \mathbb{P})$, nous allons pouvoir définir les variables aléatoires.

Définition 146
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $(F, \mathcal{F})$ où un ensemble muni d'une σ-algèbre.
Soit $X : \Omega \to F$ une fonction.
On dit que X est une ***variable aléatoire*** *de $(\Omega, \mathcal{A}, \mathbb{P})$ dans $(F, \mathcal{F})$ si on a $X^{-1}(B) \in \mathcal{A}$, $\forall B \in \mathcal{F}$.*

Remarque 147 — *Le nom donné, qui est utilisé maintenant couramment, n'est pas le mieux choisi : une variable aléatoire, malgré son nom, n'est pas une variable, mais une fonction (une fonction en la variable $\omega \in \Omega$).*
Une variable aléatoire est une fonction !
On peut abréger le nom "variable aléatoire" en ***v.a.**.*

Faisons un exemple.

Exemple 148 — *Étudions un lancer de deux dés équilibrés.*
Dans ce cas, l'ensemble des états est $\Omega = \{(i,j) : 1 \leq i \leq 6; 1 \leq j \leq 6\}$.
On a alors aussi $\mathcal{A} = \mathcal{P}(\Omega)$. *Puisque les dés sont équilibrés, on prend pour* $\mathbb{P}$ *la mesure de probabilité uniforme.*
Pour $A \subset \Omega$ *un événement, on a donc*

$$\mathbb{P}(A) = \frac{\text{card}\, A}{36}.$$

L'application $X : \Omega \to \{1, 2, \cdots, 12\}$ *définie par*

$$X(i,j) = i + j$$

est la variable aléatoire "somme des résultats des deux dés". Elle a pour loi

$$\mathbb{P}_X(B) = \frac{\textit{nombre de couples } (i,j) \textit{ tels que } i+j \in B}{36}.$$

3.1.2 Loi d'une variable aléatoire

On peut associer à une variable aléatoire X une mesure de probabilité, de la façon suivante.

Proposition-Définition 149
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé et* $(F, \mathcal{F})$ *un ensemble avec une* σ*-algèbre. Soit* $X : \Omega \to F$ *une variable aléatoire.*
On définit la fonction $\mathbb{P}_X : \mathcal{F} \to [0,1]$ *par*

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) =_{def} \mathbb{P}(X \in B).$$

Alors, la fonction $\mathbb{P}_X$ *est une mesure de probabilité.*
Cette mesure de probabilité est appelée ***loi de la variable aléatoire*** X.
Dans le langage probabiliste, on notera aussi $\mathbb{P}_X(B) =_{def} \mathbb{P}(X \in B)$.

Preuve — Comme X est une variable aléatoire, tous les $X^{-1}(B)$ sont des éléments de $\mathcal{A}$, donc la onction $\mathbb{P}_X$ est bien définie. On rappelle que

$$X^{-1}(\emptyset) = \emptyset,\ X^{-1}(F) = \Omega,\ X^{-1}(\overline{B}) = \overline{X^{-1}(B)},$$
$$X^{-1}(\bigcap_i A_i) = \bigcap_i X^{-1}(A_i),\ X^{-1}(\bigcup_i A_i) = \bigcup_i X^{-1}(A_i).$$

Comme $\mathcal{F}$ est une σ-algèbre sur F, on montre alors facilement que $\mathbb{P}_X$ est une mesure de probabilité. (elle vérifie $(*)$ et $(**)$)
□

La loi d'une variable aléatoire X donne énormément d'informations sur la variable aléatoire X, tout comme la loi d'une mesure de probabilité $\mathbb{P}$ (si Ω fini ou dénombrable) donne énormément d'informations sur $\mathbb{P}$.

Exemple 150 — *Dans l'exemple précédent du lancer de deux dés équilibrés, et pour* X *la variable aléatoire "somme des deux faces obtenues", la loi de probabilité de* X, $\mathbb{P}_X$, *est une mesure de probas sur l'ensemble* $\{2, \ldots, 12\}$.
On a par exemple

$$\mathbb{P}_X(\{2\}) = \mathbb{P}_X(\{12\}) = \frac{1}{36},\ \mathbb{P}_X(\{3\}) = \frac{2}{36},\ \mathbb{P}_X(\{5\}) = \frac{4}{36}.$$

Remarque 151 — *Pour une expérience aléatoire donnée, en faire une modélisation mathématique implique de trouver une* σ*-algèbre* $\mathcal{F}$ *sur l'ensemble d'arrivée* F *telle que* $X^{-1}(B) \in \mathcal{A}$, *pour tout* $B \in \mathcal{F}$.
Cette σ*-algèbre* $\mathcal{F}$ *peut être a priori diffcile à décrire.*

Donnons quelques exemples d'expériences aléatoires classiques que l'on va chercher à modéliser en mathématiques avec des espaces probabilisés et des variables aléatoires.

Exemple 152 —

1. *Le nombre de 6 obtenus dans un lancer de 3 dés équilibrés.*
   Les espaces sont $\Omega = \{1, \dots, 6\}^3$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}$ *la mesure uniforme sur* Ω, $F = \{0, 1, 2, 3\}$, $\mathcal{F} = \mathcal{P}(F)$.
   La variable aléatoire est $X : (a_1, a_2, a_3) \in \Omega \mapsto \chi_6(a_1) + \chi_6(a_2) + \chi_6(a_3) \in F$.
2. *Le nombre d'appels dans un central téléphonique pendant une heure* $F = \mathbb{N}$.
3. *La distance du point atteint par une flèche flèche par rapport centre de la cible (cible de* 15 *cm de rayon) :* $F = [0, 15]$.
4. *La valeur maximale du prix d'un actif sur un intervalle de temps donné :* $F = \mathbb{R}_+$.

En général, l'ensemble F *sera un ensemble fini ou dénombrable, ou* $\mathbb{R}$ *ou* $\mathbb{R}^d$*, ou un ensemble un peu plus particulier.*

Remarque 153 —
• *Si l'ensemble* Ω *est fini ou dénombrable, on utilise* $\mathcal{A} = \mathcal{P}(\Omega)$.
Ainsi, pour toute fonction $X : \Omega \to F$ *et pour tout* $B \in \mathcal{F}$*, on a* $X^{-1}(B) \in \mathcal{P}(\Omega)$.
On a donc montré que toute fonction sur un ensemble fini ou dénombrable est une variable aléatoire !
• *Si cette fois l'ensemble* $X(\Omega)$ *est fini ou dénombrable (ce qui sera souvent le cas dans le cours), alors la* σ*-algèbre* $\mathcal{F}$ *contient* $\mathcal{P}(X(\Omega))$.
De plus, la condition $X^{-1}(B) \in \mathcal{A}$, $\forall B \in \mathcal{F}$*, se réduit à*

$$\forall x \in F, \; X^{-1}(\{x\}) \in \mathcal{A}.$$

Ceci nous conduit à la notion de variables aléatoires discrètes

3.2 Variables aléatoires discrètes

Définition 154
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé et* $(F, \mathcal{F})$ *un ensemble avec une* σ*-algèbre.*
Soit $X : \Omega \to F$ *une variable aléatoire.*
Si $X(\Omega)$ *est un ensemble fini ou dénombrable, on dit que* X *est une* ***variable aléatoire discrète****.*

Remarque 155 —
• *Une variable aléatoire discrète est une variable aléatoire qui prend un nombre fini ou dénombrable de valeurs.*
• *Pour* $X : \Omega \to F$ *une fonction telle que* $X(\Omega)$ *est fini ou dénombrable,* X *est une variable aléatoire si et seulement si* $X^{-1}(\{x\}) \in \mathcal{A}$*, pour tout* $x \in X(\Omega)$.
• *Si* Ω *est fini ou dénombrable, on sait alors automatiquement que toute fonction* $X : \Omega \to F$ *est une variable aléatoire discrète.*

Exemple 156 (**Fonction indicatrice**) —
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $A \in \mathcal{A}$.
On définit $\mathbb{1}_A$ *(ou* χ_A*) la* ***fonction indicatrice*** *de* A*, par* $\mathbb{1}_A(\omega) = 1$ *si* $\omega \in A$ *et* 0 *sinon.*
Alors la fonction $\mathbb{1}_A$ *est une variable aléatoire discrète sur* $(\Omega, \mathcal{A}, \mathbb{P})$.
Ces variables aléatoires sont les v.a. les plus simples que l'on puisse construire (avec les v.a. constantes). Elles sont extrêmement utiles dans les calculs. (pour des sommes, produits, découpages en partition)
On a par exemple que $\mathbb{P}_{\mathbb{1}_A}(\{1\}) = \mathbb{P}(\mathbb{1}_A^{-1}(\{1\}) = \mathbb{P}(A)$.

On rappelle que quand Ω est fini ou dénombrable, une mesure de probabilité $\mathbb{P}$ sur Ω sera caractérisée par les $p_\omega = \mathbb{P}(\{\omega\})$ (par sa loi).
Cela n'est pas vrai quand Ω est infini non dénombrable. Mais, si X est une v.a. discrète sur Ω, on peut quand même déterminer la mesure $\mathbb{P}_X$ avec la probabilité de singletions.
C'est ce que nous donne le résultat suivant.

Proposition 157
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $X : \Omega \to F$ *une variable aléatoire.*

— *Si Ω est fini ou dénombrable, alors pour tout $y \in F$, on a*

$$\mathbb{P}_X(\{y\}) = \mathbb{P}(X^{-1}(\{y\}))\mathbb{P}(\{\omega \ t.q. \ X(\omega) = y\}) = \sum_{\omega, \ X(\omega)=y} \mathbb{P}(\{\omega\}).$$

Dans le langage probabiliste, on notera aussi $\mathbb{P}_X(\{y\}) =_{def} \mathbb{P}(X = y)$.
Quand Ω est fini ou dénombrable, on peut calculer toutes les probabilités de la forme $\mathbb{P}(X = A)$ en utilisant la probabilité de tous les singletons $\{\omega\}$.

— *Si F est dénombrable, alors la loi de la v.a. X est caractérisée par la famille des $(\mathbb{P}_X(\{y\}))_{\ y_i \in F}$.*

Exemple 158 — *Une variable aléatoire X de loi uniforme sur $\{1, \cdots, n\}$ a pour loi la famille $(\frac{1}{n})_{1 \leq k \leq n}$.*

3.3 Espérance des v.a. discrètes réelles

Définition 159
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(F, \mathcal{F})$ un ensemble avec une σ-algèbre.
Soit $X : \Omega \to F$ une variable aléatoire.
Si F est inclus dans $\mathbb{R}$ on dit que X est une ***variable aléatoire réelle****.*
On pourra abréger ce nom en ***v.a.r.****.*

Dans la majorité des exemples que nous avons vus, les variables aléatoires étaient réelles.
Les variables aléatoires étaient aussi discrètes.
Nous allons donc nous intéresser aux v.a.r. qui sont discrètes.

3.3.1 Espérance d'une variable aléatoire

Motivation : Considérons X une variable aléatoire réelle, définie sur un ensemble Ω fini ou dénombrable. On peut en général répéter l'expérience aléatoire associée à X autant de fois que l'on veut. Pour n répétitions de l'expérience X_1, ... , X_n les valeurs successives prises par X.
Pour avoir une idée du comportement de la variable X, il est naturel de considérer leur moyenne arithmétique

$$M_n = \frac{1}{n}(X_1 + ... + X_n).$$

En regroupant suivant les différents résultats y de l'expérience, nous obtenons

$$M_n = \sum_{y \in X(\Omega)} f_n(\{y\})y,$$

où $f_n(\{y\})$ est la fréquence de réalisation du résultat $\{y\}$ au cours des n expériences, c'est-à-dire de la fréquence de réalisation de l'événement $X^{-1}(\{y\})$ (dans l'ensemble de départ Ω).
D'après le précédent raisonnement intuitif sur $\mathbb{P}(X^{-1}(\{y\}))$, vue comme une mesure de la fréquence de réalisation de cet événement, on peut supposer que $f_n(\{y\})$ converge vers $\mathbb{P}(X^{-1}(\{y\})) = \mathbb{P}_X(\{y\})$ que l'on note aussi $\mathbb{P}(X = y)$.
Et si on peut de plus intervertir la somme et la limite dans l'expression ci-dessus (par exemple vrai si X prend un nombre fini de valeurs), alors la suite $(M_n)_{n \in \mathbb{N}^*}$ converge vers

$$\sum_{y \in X(\Omega)} f_n(\{y\})y = \sum_{y \in F} f_n(\{y\})y.$$

L'espérance d'une variable aléatoire, ou moyenne, est à percevoir comme la limite de ses moyennes arithmétiques, lorsque le nombre d'expériences tend vers l'infini.

Définition 160
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $X : \Omega \to F$ une variable aléatoire réelle discrète.

Si la somme $\sum_{y\in F} |y|\mathbb{P}(X=y)$ *est finie, on dit que* X *est* ***d'espérance finie*** *(pour la mesure* $\mathbb{P}$*).*
On appelle ***espérance de la v.a.*** X *le nombre* $\mathbb{E}(X) = \sum_{y\in F} y\mathbb{P}(X=y)$.

Remarque 161 —

1. *On a besoin de supposer que* X *est une v.a. discrète pour que la famille des* $(y\mathbb{P}(X=y))_{y\in F}$ *possède au plus un nombre dénombrable de termes non-nuls.*
2. *On peut remarquer que le nombre réel* $\mathbb{E}(X)$ *ne dépend que de la loi de* X *(de la famille* $(\mathbb{P}_X(\{y\})_{y\in F}$.
3. *L'hypothèse de convergence absolue de* $\sum_{y\in F} y\mathbb{P}(X=y)$*, permet de s'assurer que la somme est indépendante de l'ordre de sommation.*
4. *Le terme d'espérance (introduit par Pascal) fait référence aux problèmes de jeux et d'espérance de gain. (au fait d'espérer gagner de l'argent en jouant longtemps à un jeu de hasard)*
5. *Si la variable aléatoire* X *est d'espérance finie, alors la fonction* $|X|$*, qui est aussi une v.a.r., est d'espérance finie.*
   En effet, on a $\mathbb{E}(|X|) = \sum_{y\in F} |y|\mathbb{P}(X=y) < +\infty$.

Les v.a. réelles sont des fonctions à valeurs dans $\mathbb{R}$. On peut donc les additionner $(X+Y)$, les multiplier par une constante (aX), mais aussi les multiplier entre elles (XY), ou les composer par une fonction réelle $(f(X)$, pour $f:\mathbb{R}\to\mathbb{R})$.
Dans le cas où X, Y sont discrètes, toutes ces opérations définissent encore des v.a.r. discrètes. (A vérifier.)
On peut alors étudier ce qui se passe par rapport à l'espérance, et par rapport au fait d'être intégrable.

3.3.2 Propriétés de l'espérance des v.a. discrètes

Les sommes finies ou dénombrables qui apparaissent dans l'espérance sont liées aux intégrales. La définition suivante va permettre de relier ces notions.

Définition 162
Soit $(\Omega,\mathcal{A},\mathbb{P})$ *un espace probabilisé. Soit* X *une v.a. réelle discrète sur* Ω.
Si X *est d'espérance finie, on dit aussi que* X *est* ***intégrable****.*
On note $L^1(\Omega,\mathcal{A},\mathbb{P})$ *l'ensemble de toutes les v.a. réelles discrètes intégrables (sur* $(\Omega,\mathcal{A},\mathbb{P})$*).*

Avec ce point de vue, nous allons continuer à étudier les ensembles de v.a.r. discrètes comme des ensembles de fonctions.
Quand l'ensemble Ω et la mesure de probabilité $\mathbb{P}$ sont clairs, on utilisera parfois l'abréviation L^1 pour $L^1(\Omega,\mathcal{A},\mathbb{P})$.
Attention, dans ce chapitre toutes les v.a. que l'on considèrera intégrables/de carré intégrable/etc seront **discrètes**.

Proposition 163
Soit $(\Omega,\mathcal{A},\mathbb{P})$ *un espace probabilisé. Alors*

1. $L^1(\Omega,\mathcal{A},\mathbb{P})$ *est un* $\mathbb{R}$*-espace vectoriel.*
2. *L'espérance* $\mathbb{E}: L^1(\Omega,\mathcal{A},\mathbb{P}) \to \mathbb{R}$ *est une forme linéaire. On a* $\mathbb{E}(aX+bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$, $\forall X, Y \in L^1$, $\forall a, b \in \mathbb{R}$.
3. *On a* $X \in L^1(\Omega,\mathcal{A},\mathbb{P})$ *ssi* $|X| \in L^1(\Omega,\mathcal{A},\mathbb{P})$*. De plus, on a* $|\mathbb{E}(X)| \le \mathbb{E}(|X|)$.
4. *L'espérance est positive : Si* $X \ge 0$ *(i.e.* $X(\omega) \ge 0\ \forall\omega$*) et* $X \in L^1$*, alors on a* $\mathbb{E}(X) \ge 0$.
5. *Soient* $X, Y \in L^1$ *avec* $X \le Y$*. Alors on a* $\mathbb{E}(X) \le \mathbb{E}(Y)$.
6. $L^1(\Omega,\mathcal{A},\mathbb{P})$ *contient toutes les variables aléatoires réelles bornées. (les fonctions* X *telles que* $|X| \le b$ *pour un* $b \in \mathbb{R}$*).*
7. *Si* X *est une v.a.r. constante (*$X = a$ *pour un* $a \in \mathbb{R}$*), alors* $\mathbb{E}(X) = a$.

8. *Si Ω est fini, alors $L^1(\Omega, \mathcal{A}, \mathbb{P})$ contient toutes les v.a. réelles, ce qui est aussi égal à l'ensemble de toutes les fonctions de Ω vers $\mathbb{R}$.*

Preuve — On démontre chaque point en utilisant la définition de l'espérance et la définition des v.a. discrètes. Aucun de ces résultat n'est difficile à obtenir. □

Remarque 164 —

— *Ces propriétés font fortement penser à celles des espaces vectoriels normés (voir Analyse 4), en utilisant la fonction $X \mapsto \mathbb{E}(|X|)$.*
Mais, en général, cette fonction n'est pas une norme.
On montre facilement que l'on a $\mathbb{E}(|X|) = 0$ si et seulement si ($X(\omega) = 0$ ou $\mathbb{P}(\{\omega\}) = 0$, pour tout $\omega \in \Omega$), si et seulement si $\mathbb{P}(X = 0) = 1$.(A vérifier.)
Ainsi, si la probabilité de certains singletons vaut 0, il existe des v.a. X dont l'espérance vaut 0.
Par exemple, sur $\{1, 2, 3\}$ si on prend $\mathbb{P}$ de loi $(0, \frac{1}{2}, \frac{1}{2})$, alors la fonction indicatrice $X = \delta_1(.)$ est une fonction qui est non-nulle mais telle que $\mathbb{E}(|X|) = 0$.

— *La fonction $\mathbb{E}(|.|)$ est appelée une **semi-norme**. Elle vérifie toutes les propriétés d'une norme, sauf celle pour le cas nul.*

— *Si Ω n'est pas fini, on peut avoir des v.a.r. discrètes qui ne sont pas intégrables.*
Pour $\Omega = \mathbb{N}$ et $\mathbb{P}$ de loi $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots)$, la fonction $X : \omega \mapsto 2^\omega$ est bien définie et est une v.a.r. discrète (car Ω est dénombrable).
Par contre, son espérance est infinie car $\mathbb{P}(\{n\}).X(n) = \frac{2^n}{2^{n+1}} = \frac{1}{2}$ (et la famille $(\frac{1}{2})_{n\geq 0}$ n'est pas sommable). Donc X n'est pas intégrable.

Exemple 165 — *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $A \in \mathcal{P}(\Omega)$. Pour $\mathbb{1}_A$ la fonction indicatrice de A, cette v.a. réelle est bornée, donc intégrable, et on a*

$$\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A).$$

Cela donne donne un lien très utile entre la probabilité d'un événement et l'espérance d'une variable aléatoire.

La définition d'espérance utilise une somme sur l'espace d'arrivée F, somme qui est dénombrable car la v.a. X est discrète.
Mais si l'ensemble Ω est fini ou dénombrable, ne peut-on pas décomposer chaque terme $\mathbb{P}(X = y)$ en une somme sur des parties de Ω, et exprimer l'espérance comme une somme sur chaque $\omega \in \Omega$?
C'est ce que nous allons démontrer.

3.3.3 Lemme de transfert, théorème de transfert

Théorème 166 (**Lemme de transfert**)
Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé fini ou dénombrable. Soit $X : \Omega \to F$ une v.a. réelle sur Ω, qui est d'espérance finie.
On a alors la formule fondamentale suivante :

$$\mathbb{E}(X) = \sum_{y\in F} y\mathbb{P}(X = y) = \sum_{\omega\in\Omega} p_\omega X(\omega).$$

Preuve — Par hypothèse d'espérance finie, la famille des $y\mathbb{P}(X = y)$ est sommable. Comme Ω est dénombrable, la famille des $p_\omega X(\omega)$ est dénombrable.
On pose alors les ensembles $A_y = X^{-1}(y)$. Ces ensembles (pour $y \in X(\Omega)$) forment une partition de Ω (les autres A_y sont vides), et le théorème de sommation par paquets nous permet d'obtenir le résultat. □

Exemple 167 — *Un nombre m est choisi au hasard uniforme entre 1 et 10, et nous devons deviner ce nombre en posant des questions auxquelles il ne sera répondu que par oui ou par non.*
Calculons l'espérance du nombre N de questions nécessaires dans les cas suivants :

1. *Premier cas : la question numéro i "Est-ce que $m = i$?".*
Avec ce choix de questions, on obtient

$$\mathbb{P}(N = k) = \mathbb{P}(\text{ le nombre } k \text{ a été choisi}) = \frac{1}{10}.$$

Ainsi, l'espérance de N vaut :

$$\mathbb{E}(N) = \sum_{k=1}^{10} k\mathbb{P}(N = k) = \frac{10(10+1)}{2} \times \frac{1}{10} = \frac{11}{2}$$

2. *Deuxième cas : Avec chaque question, nous essayons d'éliminer à peu près la moitié des réponses possibles, avec le protocole suivant : Est-ce que $m \leq 5$? $m \leq 2$? (resp. $m \leq 7$?), $m \leq 4$? (resp. $m \leq 9$?).*
Alors, il faut 3 questions pour trouver 1, 2, 5, 6, 7 et 10. Et il faut 4 questions pour trouver 3, 4, 8 et 9.
L'espérance de N dans ce cas vaut donc

$$\mathbb{E}(N) = 3 \times \frac{6}{10} + 4 \times \frac{4}{10} = \frac{17}{5}$$

L'espérance dans le second cas est strictement inférieure. L'interprétation est que la seconde stratégie va "en moyenne" permettre de trouver le nombre m en moins de questions qu'avec la première stratégie.
L'espérance donne le nombre "moyen" de questions qu'il faudra poser pour trouver m. Elle ne dit par contre rien sur le nombre minimal ni le nombre maximal de questions que l'on peut avoir à poser pour trouver m.

Remarque 168 —
• *Dans le lemme de transfert, il est absolument nécessaire que Ω soit fini ou dénombrable.*
• *La somme des $p_\omega X(\omega)$ peut avoir un sens si Ω n'est pas dénombrable (par exemple quand $\Omega = \mathbb{R}$ et $\mathcal{A} = \mathcal{B}(\mathbb{R})$), mais elle ne sera en général pas égale à $\mathbb{E}(X)$.*
• *La raison : Quand est Ω non-dénombrable, il existe des mesures de probabilité $\mathbb{P}$ telles que $\mathbb{P}(\{\omega\}) = 0$ pour tout $\omega \in \Omega$ (la probabilité de chaque singleton est nulle).*
Mais comme la mesure doit vérifier $\mathbb{P}(\Omega) = 1$, on se retrouve avec $1 = \mathbb{P}(\Omega) \neq \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 0$.
• *Dans la théorie générale des probabilités discrètes, on remplace la deuxième somme du théorème de transfert par une intégrale.*
Cependant il faut d'abord avoir construit cette intégrale, et construit beaucoup d'outils en plus, et ce n'est pas du tout l'objectif de ce cours.

Soit maintenant $f : F \to \mathbb{R}$ une fonction. Ainsi $Y = f(X)$ est encore une v.a.r. discrète. Cela permet de généraliser le lemme de transfert.

Théorème 169 (**Théorème de transfert**)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé fini ou dénombrable.
Soient $X : \Omega \to F$ une v.a. intégrable, et $f : F \to \mathbb{R}$.
Si la v.a. discrète $f(X)$ et intégrable, on a alors

$$\mathbb{E}(f(X)) = \sum_{x_i\in F} f(x_i)\mathbb{P}(X = x_i) = \sum_{\omega\in\Omega} f(X(\omega))p_\omega.$$

Preuve — Ceci est encore une conséquence du théorème de sommation par paquets. □

3.3.4 Variance et écart-type

Définition 170
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
On définit l'ensemble $L^2(\Omega, \mathcal{A}, \mathbb{P})$ comme l'ensemble des v.a. réelles discrètes X telles que X^2 est intégrable.
Si $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, on dit que X est ***de carré sommable****.*

Si l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ est clair, on pourra noter L^2 à la place de $L^2(\Omega, \mathcal{A}, \mathbb{P})$.
Attention à bien voir que l'ensemble $L^2(\Omega, \mathcal{A}, \mathbb{P})$ est un ensemble de v.a. **réelles** et **discrètes**.

Proposition 171
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé.*
L'ensemble $L^2(\Omega, \mathcal{A}, \mathbb{P})$ *est un sous-espace vectoriel de*

$$L^1(\Omega, \mathcal{A}, \mathbb{P})$$

.
Pour tout $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, *on a*

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|) \leq \sqrt{\mathbb{E}(X^2)}.$$

Preuve — Soient X et Y deux variables aléatoires réelles et $a \in \mathbb{R}$. Si X et Y sont dans L^2, l'inégalité

$$(aX+Y)^2 \leq 2a^2X^2 + 2Y^2$$

montre que que $aX + Y \in L^2$: L^2 est bien un espace vectoriel.
L'inclusion $L^2 \subset L^1$ découle de $|X| \leq 1 + X^2$.
La première inégalité a déjà été vue.
Pour la seconde, on peut supposer X est positive.
Soit alors $a = \mathbb{E}(X)$ et $Y = X - a$. Par linéarité

$$\mathbb{E}(Y^2) = \mathbb{E}(X^2) - 2a\mathbb{E}(X) + a^2 = E(X^2) - a^2,$$

et $\mathbb{E}(Y^2) \geq 0$. Donc $a^2 \leq \mathbb{E}(X^2)$, ce qui est le résultat cherché. □

Définition 172
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$.
On définit la ***variance*** *de* X *par :*

$$\text{Var}\,(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \sum_{x_i \in F} (x_i - \mathbb{E}(X))^2 p_i^X.$$

On note aussi $\sigma_X = \sqrt{Var(X)}$, ***l'écart-type*** *de* X.

Remarque 173 —
• *En développant le carré* $(X - \mathbb{E}(X))^2$ *on obtient*

$$Var(X) = \sigma_X^2 == \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

ce qui permet de montrer que ce nombre réel est bien défini quand X *est de carré intégrable.*
• *Par définition, on constate que* $Var(X) \geq 0$. *Cela montre donc que* $\mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0$.
• *L'écart-type est une grandeur qui mesure une distance de la v.a* X *par rapport à son espérance* $\mathbb{E}(X)$. *(penser à* $d(x,y) = \sqrt{\langle x-y, x-y\rangle}$ *pour les espaces euclidiens)*
Elle mesure, dans un sens, à quel point la v.a. X *s'écarte en moyenne de* $\mathbb{E}(X)$.

Exemple 174 (Un jeu de loto) —
Le joueur coche 6 *numéros sur une grille qui en comporte* 49. *Les* 6 *numéros gagnants sont déterminés par tirage au sort. Soit* n *le nombre de numéros gagnants d'une grille.*
Pour une mise de 2 *Euros, on reçoit le gain* $G = g(n)$ *suivant :*

n *numéros gagnants*	*gain* $g(n)$	*probabilité*
6	2 132 885 E	$7,2\ 10^{-8}$
5	3 575 E	$7,8\ 10^{-5}$
4	94 E	$9,7\ 10^{-4}$
3	11 E	$7,8\ 10^{-2}$

Le gain moyen est donc de

$$\begin{aligned}
\mathbb{E}(G) &= \sum_n g(n)\mathbb{P}(N=n) \\
&= 11 \times 7.8\ 10^{-2} + 94 \times 9.7\ 10^{-4} + 3575 \times 1.8\ 10^{-5} + 2132885 \times 7.2\ 10^{-8} \\
&= 1,16\ E.
\end{aligned}$$

Ainsi le bénéfice moyen du joueur, qui vaut $\mathbb{E}(G) - 2 = -0.84$*, est négatif, et le jeu est défavorable au joueur.*
On peut calculer aussi que l'écart-type de ce jeu vaut 572*. La grande valeur de l'écart-type vient du fait que ce jeu peut rapporter énormément d'argent (même si cela est très très rare), alors qu'en moyenne chaque joueur perd un peu d'argent à chaque partie.*
Beaucoup de jeux de hasard sont basés sur ce ptincipe : gros gains avec très faibles probabilités (grande variance), et gains moyens légèrement négatifs (espérance légèrement négative).

Remarque 175 — *Pour* X *une v.a.r. discrète, on sait que* X^2 *intégrable implique* X *intégrable.*
La réciproque est par contre fausse en général.
Contre-exemple : *On prend* $\Omega = \mathbb{N}$ *et* $\mathbb{P}$ *la mesure dont la loi est* $(\frac{1}{2^{n+1}})_{n\geq 0}$*. On pose* $X : \mathbb{N} \to \mathbb{R}$ *avec* $X(n) = \sqrt{2}^{n+1}$*.*
Alors X *est une v.a. discrète, positive, et* $\mathbb{E}(X) = \sum_{n\geq 0} \frac{1}{\sqrt{2}^{n+1}} < +\infty$*. (la famille* $(X(n)\mathbb{P}(X = n))_n$ *est sommable, et on applique la formule de transfert)*
Par contre, on a $X^2(n) = 2^{n+1}$*, ce qui donne* $\mathbb{E}(X^2) = \sum_{n\geq 0} 1 = +\infty$*. (la famille* $(X^2(n)\mathbb{P}(X = n))_n$ *n'est pas sommable)*

On peut généraliser la définition d'intégrabilité pour toutes les puissances de X.

Définition 176
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé,* X *une v.a.r. discrète, et* $k \geq 1$*.*
Si la v.a. X^k *est intégrable (si* $\mathbb{E}(|X^k|) < +\infty$*), on dit que* X *possède un* ***moment d'ordre*** k*.*
Le moment d'ordre k *de* X *est la quantité* $\mathbb{E}(X^k)$*.*

3.3.5 Fonction génératrice d'une v.a. à valeurs dans $\mathbb{N}$

On va montrer que pour X une v.a. discrète, sa loi $\mathbb{P}_X$ peut être caractérisée par une fonction, appelée fonction génératrice, définie sur $[0, 1]$ et indéfiniment dérivable sur $[0, 1[$.
Comme X est une v.a. réelle discrète, à bijection près on peut considérer que X est à valeurs dans $\mathbb{N}$. Soit une variable aléatoire X à valeurs dans $\mathbb{N}$, dont la loi est caractérisée par les nombres $p_n = p_n^X = \mathbb{P}(X = n)$.

Définition 177
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* X *une v.a.r. à valeurs dans* $\mathbb{N}$*.*
On appelle ***fonction génératrice*** *de* X*, la fonction* $G_X : [0, 1] \to \mathbb{R}_+$ *définie par*

$$G_X(s) = \sum_{n=0}^{\infty} s^n \mathbb{P}(X = n) = \sum_{n=0}^{\infty} s^n p_n, \, \forall s \in [0, 1].$$

Remarque 178 —

1. *Cette quantité est la somme d'une série entière à termes positifs, dont tous les termes sont majorés par* 1*. Son rayon de convergence est donc d'au moins* 1*.*
   On sait aussi que pour $s = 1$ *la série est convergente, de somme* 1*. La fonction* G_X *est donc bien définie sur* $[0, 1]$*.*
2. *Pour tout* $s \in [0, 1]$*, la fonction* $s^X : \omega \in \Omega \mapsto s^{X(\omega)} \in \mathbb{R}_+$ *est bien définie et est une v.a. discrète (c'est la composée d'une v.a. discrète par une fonction). La fonction génératrice* G_X *s'écrit alors*
   $$G_X(s) = \mathbb{E}(s^X).$$
3. *La fonction génératrice ne dépend que de la famille de probabilités* $(\mathbb{P}(X = n))_{n\geq 0}$*, c'est-à-dire de la loi de* X*.*

Proposition 179
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* X *une v.a.r. à valeurs dans* $\mathbb{N}$*.*
La fonction génératrice G_X *est continue sur* $[0, 1]$ *et infiniment dérivable (de classe* C^∞*) sur* $[0, 1[$*.*
De plus, on peut retrouver la loi de X *à partir de* G_X*.*

Preuve — La fonction G_X est la somme d'une série entière à termes positifs qui converge normalement sur $[0,1]$, puisque $\sum_{n=0} \mathbb{P}(X = n) = 1$.
Les propriétés de continuité et de dérivabilité en découlent. (voir Analyse 4)
De plus, on sait que
$$\mathbb{P}(X = n) = p_n = \frac{G_X^{(n)}(0)}{n!}.$$
Ainsi, on peut retrouver la famille $(\mathbb{P}(X = n))_n$ (la loi de X) à partir de la fonction G_X.
On dit aussi que la fonction G_X **caractérise la loi** de X. □

La fonction génératrice G_X ne donne pas seulement toutes les informations sur la loi de X, elle permet aussi de dire si X est intégrable, et de calculer son espérance.

Proposition 180
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit X une v.a.r. à valeurs dans $\mathbb{N}$.
La v.a. X est intégrable si et seulement si G_X est dérivable à gache en $s = 1$.
Dans ce cas, on a $\mathbb{E}(X) = G'_X(1)$.

Preuve — La fonction G_X est dérivable sur $[0,1[$ et
$$G'_X(s) = \sum_{n\geq 0} np_n s^{n-1}.$$
Si la variable aléatoire X est intégrable alors $\mathbb{E}(X) = \sum_{n\geq 0} np_n$ est convergente et la série entière définissant G'_X est normalement convergente sur $[0,1]$, donc G'_X admet une limite en 1^-.
0n en déduit que G_X est de classe $\mathcal{C}^1$ sur $[0,1]$ et que
$$G'_X(1) = \mathbb{E}(X).$$
Si X n'est pas intégrable, alors $\left(\sum_{k=0}^{n} kp_k\right)_{n\geq 0}$ tend vers $+\infty$.
Supposons que $G'_X(s)$ admet une limite A en 1^-. La série étant à termes positifs, la fonction est croissante en la variable s et
$$\forall n \in \mathbb{N},\ \sum_{k=0}^{n} kp_k s^{k-1} \leq \sum_{k=0}^{+\infty} kp_k s^{k-1} \leq A.$$
Par passage à la limite en 1, on obtient $\sum_{k=0}^{n} kp_k \leq A$, puis $\sum_{k=0}^{+\infty} kp_k \leq A$, ce qui est absurde. Donc $G'_X(s)$ n'admet pas de limite 1^- et comme la fonction est croissante,
$$\lim_{s\to 1^-} G'_X(s) = +\infty.$$
On en déduit que G_X n'est pas dérivable en 1. □

Plus généralement, la même démonstration prouve que

Proposition 181
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, X une v.a.r. à valeurs dans $\mathbb{N}$,et $p \geq 1$.
La v.a. $X(X-1)...(X-p)$ est intégrable si et seulement si G_X est $p+1$ fois dérivable à gauche en $s = 1$.
Dans ce cas, on a $\mathbb{E}(X(X-1)...(X-p)) = G_X^{(p+1)}(1)$.
En particulier on a $\mathbb{E}(X(X-1)) = G''_X(1)$, d'où $\mathrm{Var}(X) = G''_X(1) - (G'_X(1))^2 + G'_X(1)$.

Preuve — On procède par récurrence sur p, le cas $p = 0$ étant déjà traité.
$$\forall s \in [0,1[,\ G_X^{(p+1)}(s) = \sum_{n\geq 0} n(n-1)\cdots(n-p)\, p_n s^{n-p-1}$$
et
$$\mathbb{E}(X(X-1)...(X-p)) = \sum_{n\geq 0} n(n-1)\cdots(n-p)\, p_n.$$
Un raisonnement similaire au cas $p = 0$, montre que $G_X^{(p+1)}$ admet une limite en 1^- ssi $\mathbb{E}(X(X-1)...(X-p))$ existe et alors on a l'égalité attendue. □

Remarque 182 —
• *On peut en fait montrer avec la dernière proposition que X^p est intégrable ssi G_X est $(p+1)$ fois dérivable à gauche en $s = 1$. (On utilise le fait que X^p est une combinaison linéaire des $X(X-1)\ldots(X-k)$, pour*

$0 \leq k \leq p$.)
Ainsi, en étudiant la fonction génératrice G_X, on peut dire si la v.a. X possède des moments d'ordre p (si $X^p \in L^1$).
• *Si X a un moment d'ordre p, comme X est à valeurs dans $\mathbb{N}$ on a $\mathbb{E}(X^p) = \sum_{n \geq 0} n^p \mathbb{P}(X = n)$.*
On peut alors calculer la somme de cette série à l'aide des dérivées de la fonction G_X.
Même lorsque $k = 1, 2$ (pour calculer l'espérance ou la variance), il peut être beaucoup plus rapide et simple d'utiliser les dérivées de la fonction génératrice plutôt qu'un calcul direct.

3.4 Variables aléatoires discrètes usuelles

Dans cette section, nous présentons des v.a.r. discrètes usuelles. Ces v.a. sont à chaque fois à valeurs dans $\mathbb{N}$. ($X : \Omega \to \mathbb{N}$)
Pour une v.a. discrète, on l'a dit, la v.a. X est totalement déterminée par sa mesure de probabilité associée $\mathbb{P}_X$.
Ainsi, pour décrire une v.a. discrète, il n'est pas nécessaire de vraiment décrire l'espace probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

3.4.1 Variable aléatoire de Bernoulli

Définition 183
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{N}$ une v.a.
Soit $p \in [0, 1]$.
*Si on a $\mathbb{P}(X = 1) = p$ et $\mathbb{P}(X = 0) = 1 - p$, on dit que X est une **variable aléatoire de Bernouilli**, de paramètre p.*

Proposition 184
Soient $p \in [0, 1]$ et X une v.a. de Bernouilli de paramètre p.
Alors, on a

$$\begin{aligned} \mathbb{E}(X) &= p \\ \mathrm{Var}\,(X) &= p(1-p) \\ G_X(s) &= (1 - p + ps). \end{aligned}$$

Preuve — On calcule

$$\mathbb{E}(X) = p \times 1 + 0 \times (1 - p) = p$$

et

$$\mathrm{Var}\,(X) = p - p^2 = p(1 - p).$$

Enfin, on a $G_X(s) = (1 - p)s^0 + ps^1$. □

Remarque 185 — *Quand on modélise le jeu du pile ou pace avec une pièce, en supposant que face (1) apparaît avec la probabilité p et pile (0) avec la probabilité $1 - p$, on obtient une v.a. de Bernoulli de paramètre p.*

3.4.2 Variable aléatoire binomiale

Définition 186
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{N}$ une v.a.
Soient $p \in [0, 1]$ et $n \geq 1$.
*Si on a $X(\Omega) = \{1, \ldots, n\}$ avec $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, pour tout $0 \leq k \leq n$, on dit que X est une **variable aléatoire binomiale**, de paramètres n et p.*
On note sa loi de probabilités $\mathbb{P}_X = B(n, p)$.

Remarque 187 —

1. *On retrouve cette v.a. (et sa mesure de probas associée) dans le modèle des urnes : on tire n boules parmi des boules de 2 couleurs (blanc ou noir), sachant que la probabilité de choisir une boule noire est p. Si X donne le nombre de boules noires, alors X est une v.a. binomiale.*
2. *On peut aussi considérer n lancers de Pile ou Face, sachant que la probabilité d'obtenir Face est p, et X la v.a. compte le nombre de Faces au bout de n lancers.*
3. *Pour X une v.a. binomiale, on dit aussi que sa loi de probabilité est une* ***loi binomiale****.*
4. *La loi de probas $B(1,p)$ est égale à la loi de probas de Bernoulli de paramètre p.*

Proposition 188
Soient $n \geq 1$, $p \in [0,1]$, et X une variable binomiale de loi $B(n,p)$.
Alors, on a

$$\begin{array}{rcl} G_X(s) & = & (1-p+ps)^n \\ \mathbb{E}(X) & = & np \\ \mathrm{Var}(X) & = & np(1-p). \end{array}$$

Preuve — Ici, on va calculer l'espérance et la variance de X grâce à la fonction génératrice G_X.
On calcule :

$$G_X(s) = \sum_{k=0}^{n} \binom{n}{k} p^k s^k (1-p)^{n-k} = (1-p+ps)^n.$$

En dérivant G_X et en calculant $G'_X(1)$, sachant que $k\binom{n}{k} = n\binom{n-1}{k-1}$ pour $k \neq 0$, on obtient

$$\mathbb{E}(X) = \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = np(1-p+p\times 1)^{n-1} = np$$

De même,

$$G''_X(1) = n(n-1)p^2(1-p+p\times 1)^{n-2} = n(n-1)p^2.$$

et donc

$$\begin{array}{rcl} \mathrm{Var}(X) & = & \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2 \\ & = & n(n-1)p^2 + np - (np)^2 \\ & = & np(1-p) \end{array}$$

□

Exemple 189 — *Aux jeux olympiques de Vancouver (2010), 86 médailles d'or ont été mises en jeu. Nous faisons l'hypothèse que la probabilité qu'un pays remporte une médaille est proportionnelle à sa population. Soit X le nombre de médailles prévues pour la France. X va suivre une loi binomiale $B(86,p)$, où*

$$p = \frac{\textit{population France}}{\textit{population monde}} = \frac{60\times 10^6}{6000\times 10^6} = 0,01.$$

Ainsi l'espérance de X sera égale à $86 \times 0,01 = 0,86$.
Cherchons la probabilité pour que le nombre de médailles soit inférieur à 3. Elle vaut

$$\mathbb{P}(X \leq 3) = \mathbb{P}(X=0) + \mathbb{P}(X=1) + \mathbb{P}(X=2) + \mathbb{P}(X=3),$$

avec pour tout $k \in \{0, \cdots, 86\}$

$$\mathbb{P}(X=k) = \binom{86}{k}(0,01)^k(0,99)^{86-k}.$$

Tous calculs faits, nous trouvons

$$\mathbb{P}(X \leq 3) = 0,9889.$$

La France a en fait remporté 2 médailles d'or (la France en a obtenu 4 sur 99 en 2015 et 5 sur 103 en 2018).

3.4.3 Variable aléatoire géométrique

Dans un jeu de Pile ou Face (on lance autant de fois que l'on veut, avec $\mathbb{P}(Face) = p$), on considère la variable X qui donne le numéro du premier lancer donnant Face (les précédents étant Pile).
Comme on considère les lancers indépendants, on obtient

$$\mathbb{P}(X = k) = (1-p)^{k-1}p.$$

DÉFINITION 190
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{N}$ une v.a.
Soit $p \in]0, 1[$.
Si on a $\mathbb{P}(X = k) = p(1-p)^{k-1}$ pour tout $k \geq 1$, on dit que X est une ***variable aléatoire géométrique*** *de paramètre p.*

PROPOSITION 191
Soient $p \in]0, 1[$ et X une v.a. géométrique de paramètre p.
Alors, on a

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{p} \\ \mathrm{Var}\,(X) &= \frac{1-p}{p^2} \\ G_X(s) &= \frac{ps}{1-(1-p)s} \end{aligned}$$

Preuve — Le critère de D'Alembert montre que

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} < +\infty.$$

A nouveau, calculons la fonction génératrice, puis déterminons espérance et variance avec G_X.

$$G_X(s) = \sum_{k \geq 1} p(1-p)^{k-1} s^k = \frac{ps}{1-(1-p)s}.$$

Comme $0 < p < 1$, il n'y a pas de problèmes pour les quotients. On obtient alors directement

$$G'_X(s) = \frac{p}{(1-(1-p)s)^2}, \; G''_X(s) = \frac{2p(1-p)}{(1-(1-p)s)^3},$$

et donc $\mathbb{E}(X) = G'_X(1) = \frac{1}{p}$.
On a aussi $G''_X(1) = \frac{2(1-p)}{p^2}$ et donc

$$\mathrm{Var}\,(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

□

REMARQUE 192 — *En d'autres termes, si l'on regarde une expérience de Bernouilli de paramètre p, et qu'on la répète de façon indépendante jusqu'à obtenir un succès (un Pile par exemple), alors le nombre moyen de répétitions à faire est $\frac{1}{p}$.*
faut donc, en moyenne, s'attendre à lancer 6 fois un dé équilibré avant d'obtenir le premier 1.
On retrouve ici un résultat intuitif qui dit que pour A un événement de probabilité p ($0 < p < 1$), il faudra faire en moyenne $\frac{1}{p}$ tentatives pour que l'événement A se réalise. (en moyenne 36 lancers pour obtenir un double-6 avec deux dés équilibrés, en moyenne 52 tirages avec remise pour piocher l'as de coeur, etc)

3.4.4 Variable aléatoire de Poisson

DÉFINITION 193
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{N}$ une v.a.
Soit $\theta > 0$.

Si on a $\mathbb{P}(X = k) = e^{-\theta}\dfrac{\theta^k}{k!}$ *pour tout* $k \in \mathbb{N}$, *on dit que* X *est une* ***variable aléatoire de Poisson*** *de paramètre* $\theta > 0$.
On dit aussi que sa loi de probas $\mathbb{P}_X$ *est une* ***loi de Poisson*** *de paramètre* $\theta > 0$.

Proposition 194
Soient $\theta > 0$ *et* X *une v.a. de Poisson de paramètre* θ.
Alors, on a

$$\begin{array}{rcl} \mathbb{E}(X) & = & \theta \\ \mathrm{Var}\,(X) & = & \theta \\ G_X(s) & = & e^{\theta(s-1)}. \end{array}$$

Preuve — On peut calculer facilement l'espérance de X ainsi que la fonction génératrice G_X.

$$\mathbb{E}(X) = e^{-\theta}\sum_{k=0}^{\infty} k\frac{\theta^k}{k!} = \theta,$$

$$G_X(s) = e^{-\theta}\sum_{k=0}^{\infty}\frac{\theta^k s^k}{k!} = e^{\theta(s-1)}.$$

On calcule facilement $G''_X(1) = \theta^2$ et donc on en déduit la variance de X vaut

$$\mathrm{Var}\,(X) = \theta^2 + \theta - \theta^2 = \theta.$$

□

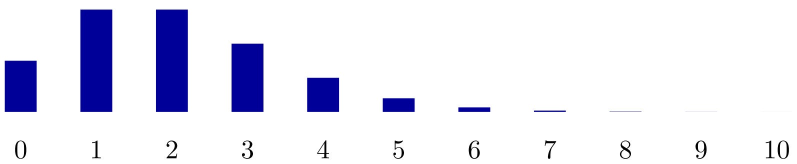
Remarque 195 — *La loi de Poisson est une loi de probabilité discrète. Elle décrit le nombre d'évènements se produisant dans un laps de temps fixé, dans le cas où ces évènements se produisent avec une fréquence moyenne connue, et indépendamment du temps écoulé depuis l'évènement précédent.*

Exemple 196 — *Une société constate en moyenne trois accidents du travail par an. L'effectif total est relativement élevé, aussi considère-t-on que le nombre d'accidents suit une loi de Poisson. Quelle est la probabilité que plus de quatre accidents surviennent dans l'année ?*
On comprend ici que $\theta = 3$.
On calcule alors :

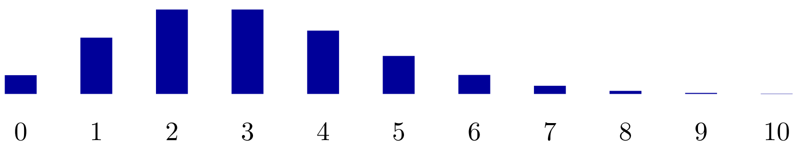
$$\mathbb{P}(X \geq 4) = 1 - e^{-3} \times \sum_{k=0}^{3}\frac{3^k}{k!} \simeq 0.26$$

Diagrammes

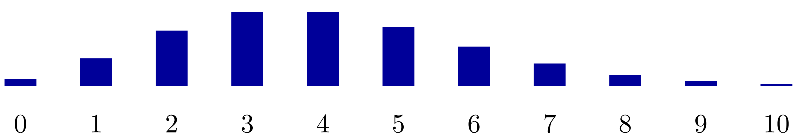
Loi de poisson de paramètre $\theta = 2$: $\mathbb{P}(X = 2) = 0,27$



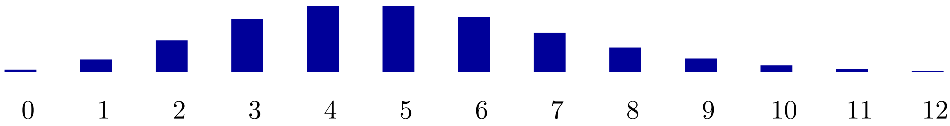
Loi de poisson de paramètre $\theta = 3$: $\mathbb{P}(X = 3) = 0,22$



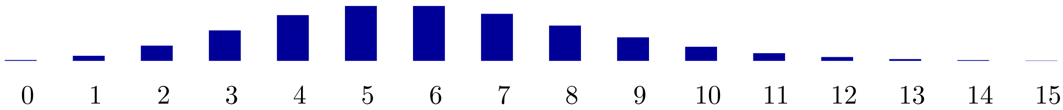
Loi de poisson de paramètre $\theta = 4$: $\mathbb{P}(X = 4) = 0,20$



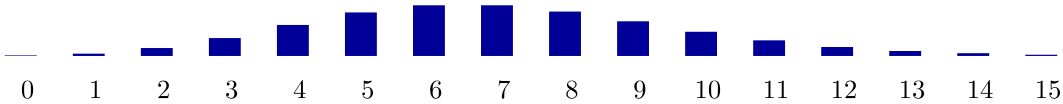
Loi de poisson de paramètre $\theta = 5$: $\mathbb{P}(X = 5) = 0,18$



Loi de poisson de paramètre $\theta = 6$: $\mathbb{P}(X = 6) = 0,16$



Loi de poisson de paramètre $\theta = 7$: $\mathbb{P}(X = 7) = 0,15$



3.5 Variables aléatoires indépendantes

Nous avons vu la notion de probabilités indépendantes. Cette notion, totalement dépendante de la mesure de probas $\mathbb{P}$, donne des informations très utiles sur la réalisation d'événements A et B.
Nous allons généraliser cette notion aux variables aléatoires.

3.5.1 Définition

Définition 197
*Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to F$ et $Y : \Omega \to G$ deux v.a. discrètes.
On dit X et Y sont des* ***variables aléatoires indépendantes*** *si on a*

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y),\ \forall (x, y) \in F \times G.$$

Proposition 198
*Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to F$ et $Y : \Omega \to G$ deux v.a. discrètes.
Alors X et Y sont indépendantes si et seulement si on a*

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B),\ \forall A \subset F,\ \forall B \subset G.$$

Preuve — L'implication $\Leftarrow$ s'obtient en prenant $A = \{x\}$ et $B = \{y\}$.
Pour l'implication réciproque, on écrit que $A \times B$ est l'union disjointe des singletons $\{x, y\}$, $x \in A$, $y \in B$. De plus, comme X et Y sont des v.a. discrètes, il n'y a qu'un nombre fini ou dénombrables d'éléments de A (resp. B) qui sont atteints par X (resp. Y). et

$$\begin{aligned}
\mathbb{P}(X \in A, Y \in B) &= \sum_{(x,y)\in A\times B} \mathbb{P}(X = x, Y = y) \\
&= \sum_{(x,y)\in A\times B} \mathbb{P}(X = x)\mathbb{P}(Y = y) \\
&= \left(\sum_{x\in A} \mathbb{P}(X = x)\right) \times \left(\sum_{y\in B} \mathbb{P}(Y = y)\right) \\
&= \mathbb{P}(A) \times \mathbb{P}(B)
\end{aligned}$$

Notons que les regroupements de somme sont licites puisque nous sommons des réels positifs et que ces sommes sont majorées par 1. □

Remarque 199 — *Pour X, Y deux v.a.d. indépendantes, si posant $Z = (X, Y)$, alors Z est une v.a. discrète et sa loi de probas est donnée par la famille $(\mathbb{P}(X = x)\mathbb{P}(Y = y))_{(x,y)}$.*

On peut aussi généraliser la notion d'indépendances à n variables aléatoires :

DÉFINITION 200
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X_i : \Omega \to F_i$, $1 \leq i \leq n$, n v.a. discrètes.
*On dit que $X_1, \ldots, X_n$ sont des **variables aléatoires indépendantes** si*

$$\forall (x_1, \cdots, x_n) \in F_1 \times \cdots \times F_n,\ \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n).$$

EXEMPLE 201 — *Considérons X_1, ..., X_n n variables aléatoires de Bernoulli de paramètre $p \in]0,1[$ qui sont indépendantes.*
Soit $x_i \in \{0,1\}$, pour $i \in \{1, ..., n\}$.
La probabilité que la suite $(X_1, \cdots, X_n)$ soit égale à $(x_1, \cdots, x_n)$, vaut alors

$$\mathbb{P}(X_i = x_i,\ 1 \leq i \leq n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_i x_i}(1-p)^{n - \sum_i x_i}.$$

On retouve le modèle du tirage dans une urne sans remise (loi binomiale).

3.5.2 Fonctions de transfert et indépendance

PROPOSITION 202
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
Soient X, Y deux v.a. discrètes indépendantes. Soient $f : F \to \mathbb{R}$ et $g : G \to \mathbb{R}$, telles que les v.a. $f(X)$ et $g(X)$ sont intégrables.
Alors le produit $f(X)g(Y)$ est aussi intégrable et vérifie

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

Preuve — On écrit

$$\begin{aligned} \sum_{(x,y) \in F \times G} |f(x)g(y)| \mathbb{P}((X,Y) = (x,y)) &= \sum_{x \in F, y \in G} |f(x_i)|\, |g(y)|\, \mathbb{P}(X = x)\mathbb{P}(Y = y) \\ &= \left(\sum_{x \in F} |f(x)| \mathbb{P}(X = x) \right) \times \left(\sum_{y \in G} |g(y)| \mathbb{P}(Y = y) \right) \end{aligned}$$

Le terme de droite étant fini par hypothèse, on en déduit que la variable aléatoire $f(X)g(Y)$ est intégrable.
Les égalités sont alors valables sans les valeurs absolues, ce qui montre exactement que $f(X)$ et $g(Y)$ sont indépendantes. □

COROLLAIRE 203
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient X, Y deux v.a.d. réelles et indépendantes.
Alors, on a $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ et $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$.

Preuve — On utilise la proposition précédente. □

COROLLAIRE 204
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X_i : \Omega \to F_i$, $1 \leq i \leq n$, n v.a. discrètes. Soient $f_1 : F_1 \times \cdots \times F_k \to G_1$ et $f_2 : F_{k+1} \times \cdots \times E_n \to G_2$ des fonctions.
Si les v.a. X_1, ...,X_n sont indépendantes, alors les v.a. $f_1(X_1, \cdots, X_k)$ et $f_2(X_{k+1}, \cdots, X_n)$ sont indépendantes.

Preuve — On utilise la proposition précédente. □

COROLLAIRE 205
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X_1, \ldots, X_n$ des v.a. réelles discrètes.
Si les v.a. $X_1, \ldots, X_n$ sont indépendantes et ont la même loi de probas, alors on a $\sigma_{X_1 + \cdots + X_n} = \sqrt{n}\sigma_{X_1}$.

Preuve — On applique un corollaire précédent. □

Attention, la notion d'indépendance pour les variables aléatoires a quelques bonnes propriétés, mais certaines manipulations ne préservent pas cette indépendance.
L'intérêt d'avoir des v.a. indépendantes est d'étudier des fonctions en ces v.a.

3.5.3 Sommes de variables aléatoires indépendantes

Proposition 206
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soient* X, Y *deux v.a. à valeurs dans* $\mathbb{Z}$.
On pose $Z = X + Y$. *Alors, on a*

$$\mathbb{P}(Z = i) = \sum_{j \in \mathbb{Z}} \mathbb{P}((X, Y) = (j, i - j)) = \sum_{j \in \mathbb{Z}} \mathbb{P}((X, Y) = (i - j, j)).$$

En particulier, si X *et* Y *sont indépendantes, on a*

$$\mathbb{P}(Z = i) = \sum_{j \in \mathbb{Z}} \mathbb{P}(X = j)\mathbb{P}(Y = i - j) = \sum_{j \in \mathbb{Z}} \mathbb{P}(X = i - j)\mathbb{P}(Y = j).$$

Remarque 207 — *La loi de probabilité définie par* $\mathbb{P}_{X+Y}(\{i\}) = \sum_{j \in \mathbb{Z}} \mathbb{P}(X = i - j)\mathbb{P}(Y = j)$ *s'appelle le* ***produit de convolution*** *des deux mesures de probabilité* $\mathbb{P}_X$ *et* $\mathbb{P}_Y$. *On le note* $\mathbb{P}_X * \mathbb{P}_Y$.
Nous n'allons pas plus étudier la convolution dans ce cours.

Exemple 208 — *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $n \geq 1$ *un entier et soient* $X, Y : \Omega \to \{0, \ldots, n\}$ *deux v.a. qui sont de loi de probas uniforme, et qui sont indépendantes.*
On étudie la loi de la v.a. $Z = X + Y$. *On a :*

$$\mathbb{P}(Z = k) = \sum_{j \in \mathbb{N}} \mathbb{P}(X = j)\mathbb{P}(Y = k - j).$$

Or

$$\mathbb{P}(X = j) = \mathbb{P}(Y = j) = \begin{cases} \dfrac{1}{n+1} \text{ si } j \in \{0, 1, \cdots, n\} \\ \\ 0 \text{ sinon.} \end{cases}$$

On doit donc séparer deux cas :

1. *si* $0 \leq k \leq n$, *alors*
$$\mathbb{P}(Z = k) = \sum_{j=0}^{k} \left(\frac{1}{n+1}\right)^2 = \frac{k+1}{(n+1)^2}$$
2. *si* $n \leq k \leq 2n$, *alors*
$$\mathbb{P}(Z = k) = \sum_{j=k-n}^{n} \left(\frac{1}{n+1}\right)^2 = \frac{2n-k+1}{(n+1)^2}$$

   On vérifie que $\mathbb{P}(Z = 2n - k) = \mathbb{P}(Z = k)$ *pour* $0 \leq k \leq n$.
   Le diagramme en bâton de la loi de probas Z *est en fait un triangle. On l'appelle loi triangulaire.*

Exemple 209 — *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soient* $\lambda_1, \lambda_2 > 0$, *et* $X, Y : \Omega \to \mathbb{N}$ *deux v.a. de Poisson de paramètres* λ_1 *et* λ_2, *et qui sont indépendantes.*
On étudie la loi de la v.a. $Z = X + Y$. *On a :*

$$\begin{aligned} \mathbb{P}(Z = i) &= \sum_{j \in \mathbb{N}} \mathbb{P}(X = j)\,\mathbb{P}(Y = i - j) \\ &= \sum_{j=0}^{i} e^{-(\lambda_1+\lambda_2)} \frac{\lambda_1^j \lambda_2^{i-j}}{j!(i-j)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{i!} \sum_{j=0}^{i} \binom{i}{j} \lambda_1^j \lambda_2^{i-j} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{i!} (\lambda_1 + \lambda_2)^i. \end{aligned}$$

Cela montre que Z *suit encore une loi de Poisson, de paramètre* $\lambda_1 + \lambda_2$.

Remarque 210 — *On peut souvent généraliser les études de sommes de 2 v.a. indépendantes à des sommes de* n *v.a. indépendantes.*

3.5.4 Fonction génératrice et indépendance

PROPOSITION 211
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X, Y : \Omega \to \mathbb{N}$ deux v.a. indépendantes. On pose $Z = X + Y$. Alors, les fonctions génératrices de X, Y, Z vérifient

$$G_Z(s) = G_X(s)G_Y(s), \ \forall s \in [0, 1].$$

Preuve — Il suffit de remarquer que pour $s \in [0, 1]$,

$$G_Z(s) = \mathbb{E}(s^Z) = E(s^{X+Y})$$

et $G_X(s) = \mathbb{E}(s^X)$ et $G_Y(s) = \mathbb{E}(s^Y)$.
La proposition 202 permet de conclure. □

REMARQUE 212 — *Comme la fonction génératrice de la v.a. Z décrit la loi de probas de Z (ainsi que son espérance, sa variance,...), on peut utiliser cette proposition pour calculer dans certains cas très facilement la loi d'une somme de variables aléatoires.*

EXEMPLE 213 —

1. *Soient $n \geq 1$, $p \in [0, 1]$, et $X_1, \ldots, X_n$ des variables de Bernoulli de paramètre p qui sont indépendantes. Alors, la v.a. $X = X_1 + \cdots + X_n$ a pour fonction génératrice $(1 - p + ps)^n$. Ainsi, X est une v.a. binomiale de paramètres n et p.*
2. *Soient X et Y sont des variables aléatoires indépendantes de lois binomiales $B(n, p)$ et $B(m, p)$ respectivement (avec le même $p \in [0, 1]$).*
   Alors, leur somme $Z = X + Y$ vérifie
   $$G_Z(s) = (1 - p + ps)^n(1 - p + ps)^m = (1 - p + ps)^{n+m}.$$
   On en déduit que $X + Y$ est une loi binomiale de paramètres $n + m$ et p.
3. *Soient X, Y des v.a. indépendantes de loi de Poisson de paramètres λ_1 et λ_2 respectivement.*
   Alors, leur somme $Z = X + Y$ vérifie
   $$G_Z(s) = e^{\lambda_1(s-1)}e^{\lambda_2(s-1)} = e^{(\lambda_1+\lambda_2)(s-1)}.$$
   Cela démontre à nouveau (et plus facilement) que $X + Y$ est une v.a. de Poisson de paramètre $\lambda_1 + \lambda_2$.

3.5.5 Une appplication en arithmétique

EXEMPLE 214 — *Soit $n > 1$ un entier fixé. On choisit de manière équiprobable un entier x dans $\{1, \ldots, n\}$. Pour tout entier $m \leq n$, on note A_m l'événement "m divise x". On note également B l'événement "x est premier avec n". Enfin, on note $p_1, \ldots, p_r$ les diviseurs premiers de n.*

1. *Exprimer B en fonction des A_{p_k}.*
2. *Pour tout $m \leq n$ qui divise n, calculer la probabilité de A_m.*
3. *Montrer que les événements $A_{p_1}, \ldots, A_{p_r}$ sont mutuellement indépendants.*
4. *En déduire la probabilité de B.*
5. *En déduire que $\phi(n)$ le nombre d'éléments premiers avec n (inversibles de $\mathbb{Z}/n\mathbb{Z}$). Démontrer que*
   $$\phi(n) = n \prod_{k=1}^{r} \left(1 - \frac{1}{p_k}\right).$$

1. *On sait que x est premier avec n si et seulement si aucun des diviseurs premiers de n ne divise x. On a donc :*
   $$B = A_{p_1}^c \cap \cdots \cap A_{p_r}^c.$$
2. *Il suffit de calculer le cardinal de A_m. Mais si $n = km$, alors les multiples de m qui sont inférieurs ou égaux à n sont $m, 2m, \ldots, km$. On a donc*
   $$\mathbb{P}(A_m) = \frac{k}{n} = \frac{1}{m}.$$

3. *Soit $i_1 < \cdots < i_m$ des entiers distincts choisis dans $\{1, \ldots, r\}$. On doit prouver que*

$$\mathbb{P}(A_{p_{i_1}}) \ldots \mathbb{P}(A_{p_{i_m}}) = \mathbb{P}(A_{p_{i_1}} \cap \cdots \cap A_{p_{i_m}}).$$

Mais,

$$\mathbb{P}(A_{p_{i_1}}) \ldots \mathbb{P}(A_{p_{i_m}}) = \prod_{j=1}^{m} \frac{1}{p_{i_j}}.$$

D'autre part, puisque $p_{i_1}, \ldots, p_{i_m}$ sont premiers entre eux deux à deux, un entier est multiple de $p_{i_1} \ldots p_{i_m}$ si et seulement s'il est multiple de chaque p_{i_j}, $j = 1, \ldots, m$. On en déduit que

$$A_{p_{i_1}} \cap \cdots \cap A_{p_{i_m}} = A_{p_{i_1} \ldots p_{i_m}},$$

soit

$$\mathbb{P}(A_{p_{i_1}} \cap \cdots \cap A_{p_{i_m}}) = \frac{1}{p_{i_1} \cdots p_{i_m}},$$

ce qui prouve le résultat voulu.

4. *Les événements $A^c_{p_1}, \ldots, A^c_{p_r}$ sont également indépendants. On en déduit que*

$$\mathbb{P}(B) = \prod_{j=1}^{r} \mathbb{P}(A^c_{p_j}) = \prod_{j=1}^{r} \left(1 - \frac{1}{p_j}\right).$$

5. *On a donc*

$$\mathbb{P}(B) = \frac{\phi(n)}{n}$$

ce qui, grâce à la question précédente, donne le résultat voulu.

3.5.6 Indépendance et produits cartésiens

Quand on modélise l'expérience de deux jets de dés équilibrés consécutifs, on considère comme ensemble d'états le produit $\Omega = \{1, \cdots, 6\} \times \{1, .., 6\}$ et on prend pour probabilité la probabilité uniforme sur $(\Omega, \mathcal{P}(\Omega))$.
On vérifie alors facilement que les variables aléatoires X_1 et X_2 qui donnent le résultat du premier jet (resp. second jet) sont des variables aléatoires de loi uniforme sur $\{1, \cdots, 6\}$ qui sont indépendantes.
On se pose la question inverse : On prend $(E_i, \mathcal{A}_i, \mathbb{P}_i)_{i \in I}$ une famille d'espaces probabilisés dont les mesures $\mathbb{P}_i$ sont discrètes. Peut-on construire un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et une famille de v.a. discrètes et indépendantes $(X_i)_{i \in I}$, avec X_i à valeurs dans $(E_i, \mathcal{A}_i)$, tels que la loi de X_i soit $\mathbb{P}_i$?
Ce problème est fondamental pour pour la construction des modèles probabilistes, et on dépasse très vite le cadre dénombrable.
Par exemple un lancer de pièces infini (dénombrable) conduit à un espace $\Omega = \{0, 1\}^{\mathbb{N}}$ non dénombrable.
On se limite donc au cas où I est fini.

PROPOSITION 215
Soit $n \geq 1$. Soient $(\Omega_1, \mathcal{P}(\Omega_i), \mathbb{P}_i)$, $1 \leq i \leq n$ des espaces probabilités, tels que Ω_i est dénombrable et $\mathbb{P}_i$ une mesure de probabilité de loi (x^i_j, p^i_j), $j \in \mathbb{N}$.
On pose $\Omega = \prod_{i=1}^n E_i$ et $X_i : \Omega \to \Omega_i$ la projection de Ω sur Ω_i.
Alors, la famille

$$\left((x^1_{i_1}, \cdots, x^n_{i_n}), p^1_{i_1} \cdots p^n_{i_n}\right)_{(i_1, \cdots, i_n) \in \mathbb{N}^n}$$

définit une mesure de probabilité $\mathbb{P}$ sur Ω telle que les variables X_i soient indépendantes.
*Cette mesure de probas est appelée **mesure produit** de $\mathbb{P}_1, \ldots, \mathbb{P}_n$.*

Preuve — On vérifie que

$$\sum_{(x^1_{i_1}, \cdots, x^n_{i_n}) \in \Omega} p^1_{i_1} \cdots p^n_{i_n} = \prod_{i=1}^{n} \left(\sum_{j \geq 0} p^i_j \right) = 1$$

donc on a bien définit une probabilité.
De plus, de la même manière

$$(X_{i_0} = x^{i_0}_j) = \bigcup_{(a_1, \cdots, a_n) \in \Omega,\ a_{i_0} = x^{i_0}_j} \{(a_1, \cdots, a_n)\}$$

L'union étant disjointe, en sommant, on trouve

$$\mathbb{P}(X_{i_0} = x_j^{i_0}) = p_j^{i_0}.$$

Enfin, pour tout $(i_1, \cdots, i_n) \in \mathbb{N}$, on a par définition de $\mathbb{P}$

$$\mathbb{P}\left(\bigcap_{j=1}^{n} (X_j = x_{i_j}^j)\right) = \prod_{j=1}^{n} \mathbb{P}(X_j = x_{i_0}^j).$$

Les variables X_i sont bien indépendantes pour $\mathbb{P}$. □

THÉORÈME 216 (**Théorème de Kolmogorov**)
Soient $(\Omega_1, \mathcal{P}(\Omega_i), \mathbb{P}_i)$, $i \in \mathbb{N}$ des espaces probabilités, tels que Ω_i est dénombrable et $\mathbb{P}_i$ une mesure de probabilité.
Soient $\Omega = \prod_{i=1}^{\infty} \Omega_i$ et $X_i : \Omega \to \Omega_i$ la projection de Ω sur E_i.
Soit $\mathcal{T}$ la sigma-algèbre sur Ω engendrée par

$$(A_n)_{n\in\mathbb{N}} \in (\mathcal{P}(\Omega_i))^{\mathbb{N}} \ \ t.q. \ \ \exists n_0 \ \ avec \ \ A_n = \Omega_n, \forall n \geq n_0.$$

Alors il existe une mesure de probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{T})$ telle que

$$\mathbb{P}\left(\prod_{i=1}^{+\infty} A_n\right) = \prod_{i=1}^{+\infty} \mathbb{P}_i\left(A_n\right).$$

Les v.a. X_i sont alors indépendantes.

REMARQUE 217 — *Ce théorème nous permet par exemple de modéliser le jeu de pile ou face infini avec la probabilité p d'obtenir face. Ainsi, la probabilité d'obtenir face au i-ème lancer reste p.*
Remarquons que, même dans ce cas élémentaire, la tribu $\mathcal{T}$ n'est pas dénombrable.
La probabilité d'obtenir une suite lancers donnés, par exemple $(P, F, P, F, P\cdots)$, est nulle.

EXEMPLE 218 — *Soit $p \in]0, 1[$. On considère une suite de parties de pile ou face de paramètre p, où les parties sont indépendantes.*
Pour $n \geq 1$, on définit la fonction T_n qui donne le numéro de la partie à laquelle on obtient exactement n fois pile (et $T_n(\omega) - n$ fois face).
Enfin, on pose les fonctions $A_1 = T_1$ et $A_n = T_n - T_{n-1}$.

1. *Quelle est la loi de la v.a. T_1 ?*
   Donner la valeur de son espérance.
2. *Soit $n \geq 2$.*
   Montrer que les v.a. $A_1, \ldots, A_n$ sont indépendantes et ont la même loi de probas.

1. *La variable aléatoire T_1 est le temps d'attente du premier pile ; elle suit la loi géométrique de paramètre p, donc d'espérance $1/p$.*
2. *Notons X_n la variable aléatoire égale à 1 si la partie numéro n amène pile et 0 sinon. Les variables X_n sont des variables aléatoires de Bernoulli indépendantes de même paramètre p. Soit $(i_1, \ldots, i_n) \in \mathbb{N}^n$. L'événement $(A_1 = i_1, \ldots, A_n = i_n)$ s'écrit aussi :*

   $$X_1 = \cdots = X_{i_1-1} = 0, X_{i_1} = 1, X_{i_1+1} = \cdots = X_{i_1+i_2-1} = 0, X_{i_1+i_2} = 1, \ldots, X_{i_1+\cdots+i_n} = 1.$$

   Donc, en posant $q = 1 - p$, on a :

   $$P(A_1 = i_1, \ldots, A_n = i_n) = q^{i_1-1}pq^{i_2-1}p \ldots q^{i_n-1}p.$$

   En sommant pour $(i_1, \ldots, i_{n-1})$ parcourant $(\mathbb{N}^*)^{n-1}$, on a :

   $$P(A_n = i_n) = q^{i_n-1}p.$$

   (A_n) suit bien une loi géométrique de paramètre p. De plus l'expression ci-dessus prouve que :

   $$P(A_1 = i_1, \ldots, A_n = i_n) = P(A_1 = i_1) \ldots P(A_n = i_n),$$

   ce qui montre que les variables $A_1, \ldots, A_n$ sont indépendantes.

3.6 Fonction de répartition

Nous avons défini la fonction génératrice G_X d'une variable aléatoire discrète à valeurs dans $\mathbb{N}$. Celle-ci caractérise la loi de X et permet de calculer l'espérance et plus généralement les moments d'ordre p pour tout p.

3.6.1 Définition

Nous considérons ici des variables aléatoires réelles :

Définition 219
Soit X une variable aléatoire et $\mathbb{P}_X$ sa loi. On appelle fonction de répartition de X la fonction suivante :

$$\forall x \in \mathbb{R},\ F_X(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X \leq x)$$

Exemple 220 — *Si $\mathbb{P}_X$ est la mesure de Dirac en 0, c'est-à-dire $\mathbb{P}(X = 0) = 1$, la fonction de répartition de X est la fonction de Heaviside en 0 : $H(x) = 0$ si $x < 0$ et $H(x) = 1$ si $x \geq 0$.*

Remarque 221 — *L'exemple ci-dessus montre qu'en général, une fonction de répartition n'est pas une fonction continue. L'étude générale des fonctions de répartition et en particulier des points de continuité est hors programme. Cependant quelques propriétés élémentaires sont à connaître.*

Proposition 222
Avec les notations précédentes, on a les propriétés suivantes

1. *F_X est croissante*
2. *F_X est continue à droite : $\lim\limits_{y \to x^+} F(y) = F(x)$*
3. *$\lim\limits_{x \to -\infty} F_X(x) = 0$, $\lim\limits_{x \to +\infty} F_X(x) = 1$.*

Preuve — La croissance de F est immédiate. Pour montrer 2/, on considère une suite $(x_n)_{n \geq 0}$ décroissant strictement vers x. Alors la suite $(]-\infty, x_n])_{n \geq 0}$ décroît vers $]-\infty, x]$. Par limite décroissante, $F_X(x_n)$ décroît vers $F_X(x)$. Comme F est monotone, on en déduit que F admet une limite à droite.
De même, si $]-\infty, x]$ décroît vers $\emptyset$ (resp. croît vers $\mathbb{R}$) lorsque x décroît vers $-\infty$ (resp. croît vers $+\infty$) et la monotonie de F permet de conclure. □

Remarque 223 — *Comme F est croissante, elle admet une limite à gauche en chaque point notée $F(x-)$. En remarquant que $]-\infty, y[= \bigcup\limits_{n \in \mathbb{N}}]-\infty, y_n]$ si (y_n) croît strictement vers y, on obtient facilement que pour $x < y$,*

1. *$\mathbb{P}_X(]x, y]) = \mathbb{P}(x < X \leq y) = F(y) - F(x)$,*
2. *$\mathbb{P}_X(]x, y[) = \mathbb{P}(x < X < y) = F(y-) - F(x)$,*
3. *$\mathbb{P}_X([x, y]) = \mathbb{P}(x \leq X \leq y) = F(y) - F(x-)$,*
4. *$\mathbb{P}_X([x, y[) = P(x \leq X < y) = F(y-) - F(x-)$.*

En particulier,

$$\mathbb{P}_X(\{x\}) = F(x) - F(x-)$$

est le saut de la fonction F au point x. Nous avons donc

Proposition 224

$$\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x) = 0 \Leftrightarrow F \text{est continue en } x.$$

3.6.2 Fonction de répartition d'une variable aléatoire discrète

§ *1. Variable aléatoire constante* Si X est identiquement égale au réel $a \in \mathbb{R}$, alors sa loi est la mesure de Dirac en a et sa fonction de répartition est $F(x) = \mathbb{1}_{[a,\infty[}(x)$.

§ *2. Variable aléatoire à valeurs dans* $\mathbb{N}$ Si X prend ses valeurs dans $\mathbb{N}$, alors sa loi $\mathbb{P}_X$ est caractérisée par la suite $p_n = \mathbb{P}_X(\{n\}) = \mathbb{P}(X = n)$. La fonction de répartition F de X vaut alors

$$\begin{array}{rcl} F : \mathbb{R} & \to & [0,1] \\ x & \mapsto & \left\{ \begin{array}{l} 0 \text{ si } x < 0 \\ p_0 + \ldots + p_n \text{ si } n \leq x < n+1, n \in \mathbb{N} \end{array} \right. \end{array}.$$

La fonction F est une fonction en escalier qui saute de l'amplitude pn au point n. Puisque $F(x) \in [0,1]$, F admet au plus k sauts de taille supérieure à $\frac{1}{k}$, pour tout $k \in \mathbb{N}^*$.

Exemple 225 — *Soit X une variable géométrique et F sa fonction de répartition. Alors*

$$\forall n \in \mathbb{N}^*,\ F(n) = \sum_{k=1}^{n} p(1-p)^{k-1} = p \times \frac{1-(1-p)^n}{1-(1-p)} = 1-(1-p)^n$$

et donc $F(x) = 1 - (1-p)^{E(x)}$ si $x \geq 0$ et 0 sinon.

§ *3. Cas général* Plus généralement, si X prend ses valeurs dans une partie finie ou dénombrable E de $\mathbb{R}$, la loi $\mathbb{P}_X$ de X est caractérisée pour tout $x_i \in E$ par $p_i = \mathbb{P}_X(\{x_i\}) = P(X = x_i)$. La fonction de répartition F de X est alors

$$\begin{array}{rcl} F : \mathbb{R} & \to & [0,1] \\ x & \mapsto & \displaystyle\sum_{x_i \leq x} p_i \end{array}.$$

avec la convention qu'une somme "vide" vaut 0.

Proposition 226
La loi d'une variable aléatoire discrète est uniquement déterminée par sa fonction de répartition.

Preuve — La loi d'une variable discrète est par définition la distribution $(x_i, p_i)_{i\in\mathbb{N}}$. D'après la remarque 223, $p_i = \mathbb{P}(X = x_i) = F(x_i) - F(x_i-)$, ce qui montre bien que F détermine uniquement la loi de X. La réciproque étant par définition. □

Remarque 227 — *Notons aussi que l'ensemble E, quoiqu'au plus dénombrable, peut être dense dans $\mathbb{R}$, par exemple il peut être égal à l'ensemble des rationnels $\mathbb{Q}$. Dans ce cas, si $q_i > 0$ pour tout $i \in \mathbb{Q}$, la fonction F nous donne un exemple de fonction discontinue en tout nombre rationnel, et continue partout ailleurs.*

§ *4. Loi sans mémoire*

Définition 228
On dit qu'une variable aléatoire discrète $X : \Omega \to \mathbb{N}^*$ est sans mémoire si

$$\forall n, m \in \mathbb{N},\ \mathbb{P}(X > m+n | X > n) = \mathbb{P}(X > m).$$

Remarque 229 — *La variable X est sans mémoire ou sans viellissement car si $\mathbb{P}$ est la probabilité qu'un processus dure n unités de temps, alors si le processus a déjà duré n unités de temps, sa probabilité qu'il dure encore m unités de temps est la même que s'il partait de l'instant 0.*

Proposition 230
Si X est une variable aléatoire discrète sans mémoire $X : \Omega \to \mathbb{N}^*$, alors X est une variable géométrique.

Preuve — Posons $p = \mathbb{P}(X = 1)$. D'après la proposition 226, il suffit de montrer que la fonction de répartition F de X coïncide avec celle d'une variable géométrique : on pose $p = \mathbb{P}(X = 1)$ et on veut montrer que pour tout $n \in \mathbb{N}$, $F(n) = \mathbb{P}(X \leq n) = 1 - (1-p)^n$ ce qui est équivalent à $\mathbb{P}(X > n) = (1-p)^n$.
On a bien :

$$\mathbb{P}(X > 1) = 1 - \mathbb{P}(X = 1) = 1 - p$$

puisque X à valeurs dans $\mathbb{N}^*$.
Enfin,

$$\mathbb{P}(X > k+1 | X > k) = \mathbb{P}(X > 1) \iff \mathbb{P}(X > k+1) = (1-p) \times \mathbb{P}(X > k).$$

Ce qui détermine bien une suite géométrique de raison $(1-p)$. □

On termine ce paragraphe avec une propriété sur l'espérance des variables aléatoires à valeurs dans $\mathbb{N}$ qui peut être fort utile

Proposition 231
Si X est une variable aléatoire discrète $X : \Omega \to \mathbb{N}^*$ intégrable, alors

$$\mathbb{E}(X) = \sum_{n \geq 1} \mathbb{P}(X \geq n).$$

Preuve — Soit (k, p_k) la distribution de X. Alors

$$\mathbb{P}(X \geq n) = \sum_{k \geq n} p_k$$

et

$$\sum_{n \geq 1} \mathbb{P}(X \geq n) = \sum_{n \geq 1} \left(\sum_{k \geq n} p_k \right) = \sum_{k \geq 1} \left(\sum_{n=1}^{k} p_k \right) = \sum_{k \geq 1} k p_k$$

la dernière égalité résulant du théorème de sommation par paquets :

$$I = \bigcup_{n \in \mathbb{N}} \{(n,k)\ ,\ k \geq n\} = \bigcup_{k \in \mathbb{N}} \{(n,k)\ ,\ n \in [\![1, k]\!]\}$$

la somme indexée par la dernière partition étant sommable par hypothèse. □

3.7 L'ensemble $L^2(\Omega, \mathcal{A}, \mathbb{P})$

Nous revenons à l'étude de l'espace vectoriel $L^2(\Omega, \mathcal{A}, \mathbb{P})$.
Le but est de progresser dans la comparaison de variables aléatoires en s'aidant de quantités supplémentaires (covariance, approximation).

Remarque 232 — *Nous avons déjà vu que $L^2(\Omega, \mathcal{A}, \mathbb{P})$, l'ensemble des v.a. réelles et discrètes de carré intégrable sur $(\Omega, \mathcal{A}, \mathbb{P})$, est un espace vectoriel.*

3.7.1 Covariance

Proposition 233
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
• *La fonction $(X, Y) \mapsto \mathbb{E}(XY)$ définit une forme bilinéaire symétrique positive sur $L^2(\Omega, \mathcal{A}, \mathbb{P})$.*
• *L'inégalité de Cauchy-Schwarz donne*

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

Preuve — On vérifie avec les propriétés de l'espérance toutes les propriétés de l'énoncé. □

Corollaire 234
En prenant $Y = 1$, on retrouve l'inégalité

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

Remarque 235 — *La forme bilin. $(X, Y) \mapsto \mathbb{E}(XY)$ n'est en général pas un produit scalaire. pas définie positive.*
Pour X une v.a.r. discrète, de loi de probas $(x_i, \mathbb{P}(X = x_i))$, on a

$$\mathbb{E}(X^2) = \sum_{i \geq 0} x_i^2 \, \mathbb{P}(X = x_i).$$

Selon la mesure de probas $\mathbb{P}$, on peut avoir $\mathbb{P}(X = x_i) = 0$ même si $x_i \neq 0$.
On a en général seulement :

$$\mathbb{E}(X^2) = 0 \Leftrightarrow \mathbb{P}(X = 0) = 1.$$

C'est-à-dire, $\mathbb{E}(X^2) = 0$ si et seulement si la v.a. X est $\mathbb{P}$-presque sûrement égale à 0.

Proposition-Définition 236
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
*• La fonction **covariance** Cov : $(X, Y) \mapsto \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big)$ définit une forme bilinéaire symétrique positive sur $L^2(\Omega, \mathcal{A}, \mathbb{P})$.*
• L'inégalité de Cauchy-Schwarz donne

$$\mathrm{Cov}(X, Y)^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y).$$

On peut remarquer que pour $X \in L^2$, on a $Cov(X, X) = \mathrm{Var}(X)$.

Preuve — L'application est clairement bilinéaire symétrique positive puisque $\mathbb{E}$ est linéaire, positive.
Il faut cependant justifier que la covariance est bien définie : développons

$$(X - \mathbb{E}(X))(Y - \mathbb{E}(Y) = XY - \mathbb{E}(X)Y - \mathbb{E}(Y)X + \mathbb{E}(X)\mathbb{E}(Y)$$

qui est somme de variables intégrables puisque $L^2 \subset L^1$ et la variable constante est intégrable. □

Corollaire 237
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. On a

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

Remarque 238 —
• La covariance n'est pas un produit scalaire non plus.
En effet, on a $Cov(X, X) = 0$ ssi $Var(X) = 0$ ssi $\mathbb{E}((X - \mathbb{E}(X))^2) = 0$, ssi $\mathbb{P}(X - \mathbb{E}(X) = 0) = 1$, ssi $\mathbb{P}(X = \mathbb{E}(X)) = 1$.
Une v.a.r. discrète X vérifie $Cov(X, X) = 0$ si et seulement si X est $\mathbb{P}$-presque sûrement égale à son espérance.
• La covariance est un outil très important dans l'étude des v.a. (de carré intégrable). Elle permet de quantifier à quel point des v.a. X_1, X_2 ne sont pas indépendantes.

Proposition 239 (Variance d'une somme de v.a.)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X_1, \ldots, X_n$ des v.a. de carré intégrable. On a alors

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j).$$

Preuve — Cela résulte immédiatement du fait que cov est une forme bilinéaire symétrique et que $\mathrm{Var}(X) = \mathrm{cov}(X, X)$. □

Définition 240
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ de variance non-nulle.
*On définit le **coefficient de corrélation** des v.a. X et Y come*

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Pour des produits scalaires, cela revient à écrire $\frac{\langle X,Y\rangle}{\|X\|\|Y\|}$.

Proposition 241
Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ de variance non-nulle. On a alors

$$-1 \leq \rho(X,Y) \leq 1$$

De plus, on a

$$|\rho(X,Y)| = 1 \iff \exists a,b,c \in \mathbb{R} \text{ tels que } \mathbb{P}(aX + bY + c = 0) = 1.$$

Preuve — On applique l'inégalité de Cauchy-Schwarz au produit bilinéaire symétrique positif cov. Le cas d'égalité est équivalent à $\mathrm{cov}(Y,Y) = 0$ ou $\mathrm{cov}(X+bY, X+bY) = 0$.
D'après la remarque 235

$$\mathbb{E}\Big(\big(X + bY - \mathbb{E}(X+bY)\big)^2\Big) = 0 \Leftrightarrow \mathbb{P}\Big(X + bY - \mathbb{E}(X+bY) = 0\Big) = 1$$

et de même si $\mathrm{cov}(Y,Y) = 0$, alors $\mathbb{P}(Y - \mathbb{E}(Y) = 0) = 1$. □

Proposition 242
Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, et $a, b, a', b' \in \mathbb{R}$. Alors on a

$$\mathrm{cov}(aX + b, a'Y + b') = aa'\mathrm{cov}(X,Y).$$

En particulier, cela donne

$$\mathrm{var}(aX + b) = a^2\mathrm{Var}(X).$$

Preuve — Comme la covariance est une forme bilinéaire symétrique, il suffit en fait de développer l'expression par bilinéarité, et de montrer que $Cov(X,b') = Cov(b,Y) = 0$ (la covariance entre une v.a. et une v.a. constante vaut 0).
Cela donne :

$$\begin{aligned}\mathbb{E}((aX+b)(a'Y+b')) - \mathbb{E}(aX+b)\mathbb{E}(a'Y+b') &= aa'\mathbb{E}(XY) + ab'\mathbb{E}(X) + a'b\mathbb{E}(Y) + bb' \\ &- \left[aa'\mathbb{E}(X)\mathbb{E}(Y) + ab'\mathbb{E}(X) + a'b\mathbb{E}(Y) + bb'\right] \\ &= aa'\left[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\right]\end{aligned}$$

Ce qui montre l'égalité désirée. □

Corollaire 243
Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ de variance non-nulle et $a, a', b, b' \in \mathbb{R}$.
Si $aa' > 0$, alors on a $\rho(X,Y) = \rho(aX + b, a'Y + b')$.
Les v.a X et Y, et $aX + b$ et $a'Y + b'$ ont le même coefficient de corrélation linéaire.

Preuve — D'après la proposition ci-dessus, on a

$$\rho(aX+b, a'Y+b') = \frac{aa'\mathrm{cov}(X,Y)}{a\sqrt{\mathrm{cov}(X,X)}a'\sqrt{\mathrm{cov}(Y,Y)}} = \rho(X,Y)$$

□

Remarque 244 — *Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ une v.a. de carré intégrable, dont l'écart-type est strictement positif ($\sigma_X > 0$).*
Alors la variable aléatoire $Y = \dfrac{X - \mathbb{E}(X)}{\sigma_X}$ est une v.a. de L^2, qui est d'espérance nulle et d'écart-type 1.
On dit que Y est une v.a. ***centrée*** *($\mathbb{E}(Y) = 0$) et* ***réduite*** *($\sigma_Y = 1$).*

Remarque 245 — *L'inégalité de Minkowski pour la forme bilin $Var(.,.)$ (ou inégalité triangulaire) montre que si X et Y sont de carré intégrable, alors on a*

$$\sigma_{X+Y} \leq \sigma_X + \sigma_Y.$$

3.7.2 Approximation linéaire

Dans l'étude des variables aléatoires, il est fréquent de connaître certaines informations sur la v.a. X (son espérance, sa variance, un échantillon de valeurs $X(\omega)$,...) mais pas toutes les informations.

On peut alors chercher à approximer la v.a. X avec les informations que l'on connaît.
Quelles sont les v.a. (les fonctions) les plus si

Nous considérons X, Y deux v.a. de carré intégrable dont on connaît que les variances et la covariance.
On suppose que la v.a. X est entièrement connue, mais qu Y ne l'est pas vraiment.
On veut ainsi chercher à approximer Y par une v.a. qui dépende de X, avec les informations que l'on connaît.
L'approximation la plus fondamentale que l'on peut faire en mathématiques est l'approximation linéaire, l'approximation par des fonctions affines.
On va ainsi chercher une v.a. de la forme $aX + b$ qui soit la plus "proche" de notre v.a. Y.
Plus proche au quel sens ? On choisira ici au sens **des moindres carrés**, c'est à dire qui minimise la quantité $\mathbb{E}((Y - (aX + b))^2)$.
On retrouve ici des question de minimisation de "distance" sur des espaces vectoriels (voir chapitre Orthogonalité).

Comment trouver ces coefficients a et b ? En utilisant les mêmes idées que celles du projeté orthogonal !
On a $\{aX + b,\, a, b \in \mathbb{R}\}$ qui est un sous-ev de v.a. de dimension 1 ou 2.
On suppose ici que $Var(X) \neq 0$, donc en particulier que X n'est pas une v.a. constante (sinon l'étude se simplifie trop).
On va chercher à "orthonormaliser" la famille $(1, X)$.
bullet La v.a. 1 vérifie $\mathbb{E}(1^2) = 1$, c'est bon.
• On cherche une v.a. de la forme $X - \lambda.1$ telle que $\mathbb{E}(1 \cdot (X - \lambda.1)) = 0$.
On obtient alors $\lambda = \mathbb{E}(X)$, c'est-à-dire $X' = X - \mathbb{E}(X)$.
• On veut alors "renormaliser" X'. On a $\mathbb{E}((X')^2) = Var(X) = \sigma_X^2 \neq 0$.
Donc, la v.a. $X'' = \frac{X'}{\sigma_X} = \dfrac{X - \mathbb{E}(X)}{\sigma_X}$ convient.
Posons $\beta = \mathbb{E}(Y.1)$ et $\alpha = \mathbb{E}(YX'')$.
Avec ce choix de coefficients, la bilinéarité de $(Z_1, Z_2) \mapsto \mathbb{E}(Z_1.Z_2)$ nous donne :

$$\mathbb{E}((Y - (\alpha X'' + \beta)).(aX + b)) = a.0 + b.0 = 0,\ \forall a, b \in \mathbb{R}.$$

Ainsi, d'après le théorème de Pythagore appliqué à $(Z_1, Z_2) \mapsto \mathbb{E}(Z_1.Z_2)$, on a donc :

$$\mathbb{E}((Y - (aX+b))(Y - (aX+b))) = \mathbb{E}([Y - (\alpha X''+\beta) + (\alpha X''+\beta - aX - b)][Y - (\alpha X''+\beta) + (\alpha X''+\beta - aX - b)]),$$

$$\mathbb{E}((Y - (aX + b))(Y - (aX + b))) = \mathbb{E}((Y - (\alpha X'' + \beta))^2) + \mathbb{E}((\alpha X'' + \beta - aX - b)^2),$$

$$\mathbb{E}((Y - (aX + b))(Y - (aX + b))) \geq \mathbb{E}((Y - (\alpha X'' + \beta))^2).$$

On en déduit donc que la meilleure approximation de Y par une v.a. de la forme $aX + b$, au sens des moindres carrés, est $\alpha X'' + \beta$.
Reste à calculer α et β.
On a $\beta = \mathbb{E}(Y)$, et $\alpha = \frac{\mathbb{E}(YX - Y\mathbb{E}(X)}{\sigma_X}$.
En ré-exprimant $\alpha X'' + \beta$ en fonction de X,σ_Y, $Cov(X, Y)$, $Var(Y)$, on obtient :

Proposition 246
Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, avec $\sigma_X \neq 0$.
La meilleure approximation de Y par une fonction de la forme $aX + b$, au sens des moindres carrés, est la v.a. :

$$\frac{Cov(X, Y)}{Var(X)}(X - \mathbb{E}(X)) + \mathbb{E}(Y) = \rho(X, Y)\frac{\sigma_Y}{\sigma_X}\big(X - \mathbb{E}(X)\big) + \mathbb{E}(Y).$$

Définition 247
*Avec les notations précédenes, on appelle **droite de régression linéaire** (ou la droite des moindres carrés) de Y en fonction de X la droite d'équation*

$$\big(y - \mathbb{E}(Y)\big) - \rho(X, Y)\frac{\sigma_Y}{\sigma_X}\big(x - \mathbb{E}(X)\big) = 0,\ (x, y) \in \mathbb{R}^2.$$

Remarque 248 — *Avec les résultats du cours sur l'orthogonalité, on en déduit aussi que l'écart entre Y et son approximation linéaire en X, au sens des moindres carrés, vaut :*

$$\begin{aligned}\mathbb{E}\left[\Big(\big(Y-\mathbb{E}(Y)\big)-a\big(X-\mathbb{E}(X)\big)\Big)^2\right] &= \mathbb{E}[(Y-\alpha X''-\beta)^2] \\ &= \mathbb{E}(Y^2)-\alpha^2-\beta^2=\mathbb{E}(Y^2)-\frac{Cov(X,Y)^2}{\sigma_X^2}-\mathbb{E}(Y)^2 \\ &= Var(Y)-\frac{Cov(X,Y)^2}{Var(X)}=Var(Y)(1-\rho(X,Y)^2)\end{aligned}$$

Ceci montre que plus $|\rho(X,Y)|$ est proche de 1 (plus $|Cov(X,Y)|$ est proche de $Var(X)Var(Y)$), plus l'approximation est bonne.
A l'inverse, si $Cov(X,Y)=0$, alors la distance est maximale et vaut $Var(Y)$.

3.8 Lois conditionnelles

Soit $(\Omega,\mathcal{A},\mathbb{P})$ un espace probabilisé, et $X:\Omega\to F$, $Y:\Omega\to G$, deux v.a. discrètes.
Nous avons vu dans le paragraphe précédent que si X et Y sont réelles, alors la covariance permet de trouver des v.a. de la forme $f(X)$ qui approximent Y de façon optimale (qui minimise une certaine quantité).
On suppose ici que F et G sont dénombrables.
Pour comprendre comment X et Y sont liées l'une par rapport à l'autre, nous allons utiliser les probabilités conditionnelles.
On considère le couple $Z=(X,Y):\Omega\to F\times G$. C'est encore une v.a. discrète à valeurs dans un ensemble dénombrable.
Alors, la loi de probabilité de $\mathbb{Z}$ est caractérisée par la famille $(\mathbb{P}(Z=(x,y)))_{(x,y)\in F\times G}$.
On rappelle que l'on a $\mathbb{P}(Z=(x,y))=\mathbb{P}(X=x \text{ et } Y=y)$.

Définition 249
Soit $(\Omega,\mathcal{A},\mathbb{P})$ un espace probabilisé, et $X:\Omega\to F$, $Y:\Omega\to G$, deux v.a. discrètes.
*Les mesures $\mathbb{P}_X$ et $\mathbb{P}_Y$ sont appellées les **lois marginales** de la v.a. $Z=(X,Y)$.*

Exemple 250 — *Un joueur lance en même temps un dé rouge et un dé bleu.*
Soient X le résultat du dé rouge et Y le résultat de la somme des deux dés.
Il est clair que la connaissance de la valeur de X va influer sur les valeurs possibles que peut prendre Y et sur sa loi.
Par exemple, si $X=3$, alors Y ne pourra prendre que des valeurs supérieures ou égales à 4, ce qui n'est pas le cas si $X=1$. Il est donc naturel de s'intéresser, pour chaque valeur fixée x_i de X, à la loi de Y avec l'information a priori que $X=x_i$.

Remarque 251 — *Plus généralement, quand on étudie un phénomène aléatoire, on obtient une série de mesures qui chacune donne une information partielle sur le résultat.*
Chacune de ces mesures correspond à une variable aléatoire.
Une bonne compréhension du phénomène correspond à étudier les liens entre ces valeurs.

Définition 252
Soient $X:\Omega\to F$, $Y:\Omega\to G$, deux v.a. discrètes, avec F,G dénombrables.
Soit $x\in F$ tel que $\mathbb{P}(X=x)>0$.
*On appelle **loi de probabilité conditionnelle de Y sachant** $X=x$ la mesure de probabilité sur G associée à la famille $(\mathbb{P}(Y=y|X=x))_{y\in G}$.*
On la note $\mathbb{P}_{Y|X=x}$.

Remarque 253 — *Ces lois conditionnelles de Y sachant $X=x$ sont a priori différentes pour chaque valeur de x, et différentes de la mesure de probas $\mathbb{P}_Y$.*
Cela vient des propriétés des probabilités conditionnelles que nous avons vues dans le chapitre précédent.

Nous avons en fait les relations suivantes.

Proposition 254
Soient $X : \Omega \to F$, $Y : \Omega \to G$, deux v.a. discrètes, avec F, G dénombrables. On a alors

1. $\mathbb{P}(X = x) = \sum_{y \in G} \mathbb{P}\big(Z = (x, y)\big)$, $\forall x \in F$;
2. $\mathbb{P}_{Y|X=x}(\{y\}) = \mathbb{P}(Y = y|X = x) = \dfrac{\mathbb{P}(Z = (x, y))}{\mathbb{P}(X = x)}$, *si* $\mathbb{P}(X = x) > 0$.
3. $\mathbb{P}\big(Z = (x, y)\big) = \begin{cases} \mathbb{P}(Y = y|X = x)\mathbb{P}(X = x) \text{ si } \mathbb{P}(X = x) > 0, \\ 0 \text{ sinon} \end{cases}$

Ainsi, la mesure de probas $\mathbb{P}_Z$ est caractérisée par $\mathbb{P}_X$ et par les $\mathbb{P}Y|X = x$ ($x \in F$), et la réciproque est vraie.

Preuve — Pour montrer le 1/ : l'ensemble $\{X = x_i\}$ est la réunion (finie ou dénombrable) des ensembles deux-à- deux disjoints $\{X = x_i, Y = y_j\} = \{Z = (x_i, y_j)\}$ pour $y_j \in G$. Par σ-additivité, on en déduit le résultat.
L'égalité 2/ est en fait la définition de la probabilité conditionnelle.
Enfin, si $p_i^X = P(X = x_i) > 0$, l'égalité vient de 2/ et sinon $\mathbb{P}(X = x_i, Y = y_j) = 0$ d'après 1/. □

Proposition 255
Avec les notations précédentes et en supposant Y intégrable, on a

$$\mathbb{E}(Y) = \sum_{i \geq 0} \mathbb{E}(Y|X = x_i)\mathbb{P}(X = x_i).$$

Preuve — Par définition $\mathbb{E}(Y) = \sum_{y_j \in G} y_j p_j^Y$. L'égalité 1/ de la proposition 254 nous dit que

$$p_j^Y = \sum_{x_i \in F} \mathbb{P}(Y = y_j, X = x_i).$$

On a donc

$$\mathbb{E}(Y) = \sum_{j \geq 0} \left(\sum_{i \geq 0} y_j \, \mathbb{P}(Y = y_j, X = x_i) \right)$$

La série étant absolument convergente, on peut intervertir les sommes

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{i \geq 0} \left(\sum_{j \geq 0} y_j \, \mathbb{P}(Y = y_j, X = x_i) \right) \\ &= \sum_{i \geq 0} \left(\sum_{j \geq 0} y_j \, \mathbb{P}(Y = y_j|X = x_i)\mathbb{P}(X = x_i) \right) \quad \text{Prop 254 3/} \\ &= \sum_{i \geq 0} \left(\mathbb{P}(X = x_i) \sum_{j \geq 0} y_j \, \mathbb{P}(Y = y_j|X = x_i) \right) \\ &= \sum_{i \geq 0} \mathbb{E}(Y|X = x_i)\mathbb{P}(X = x_i) \end{aligned}$$

On fait abstraction des indices i tels que $\mathbb{P}(X = x_i) = 0$, mais alors, $\mathbb{P}(X = x_i, Y = y_j) = 0$. □

Remarque 256 — *Ce résultat permet de calculer $\mathbb{E}(Y)$ en conditionnant par une variable auxiliaire X. Il généralise la formule des probabilités totales, qui correspond ici à $Y = \mathbb{1}_A$, et $B_i = \{X = x_i\}$:*

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap B_i) = \sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

où les B_i forment un système complet.

Exemple 257 — *Le nombre N de voitures passant devant une station d'essence en un jour suit la loi de Poisson de paramètre $\lambda > 0$. Chaque voiture décide de s'arrêter à la station avec probabilité p*

indépendamment des autres. On note K le nombre de voitures qui s'arrêtent à la station. Cherchons $\mathbb{E}(K)$. Par hypothèse,

$$p_n^N = e^{-\lambda}\frac{\lambda^n}{n!},\ p_k^{K|N=n} = \binom{n}{k}p^k(1-p)^{n-k}$$

D'où

$$\mathbb{E}(K|N=n) = \sum_{k=0}^{n} k p_k^{K|N=n} = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = np,$$

espérance d'une loi binomiale de paramètre n et p.
D'après la proposition précédente

$$\mathbb{E}(K) = \sum_{n\geq 0}\mathbb{E}(K|N=n)\mathbb{P}(N=n) = p\sum_{n\geq 0} n\,\mathbb{P}(N=n) = p\lambda$$

puisque $\lambda = \mathbb{E}(N)$.

Chapitre 4 Résultats asymptotiques

Table des matières du chapitre

Les variables aléatoires sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ sont des fonctions. On est ainsi amené à étudier des suites de variables aléatoires (des suites de fonctions), par exemple une suite de lancers de Pile ou Face (ou une suite de tirage de cartes, une suite de jets de dés,...).
Avec une suite de fonctions, on cherche à étudier le comportement de la suite (bornée, non-bornée, normes, convergences, limite,...).
En probabilités, on s'intéresse énormément à faire la même chose avec des v.a.

4.1 Loi faible des grands nombres, Inégalité de Markov

Nous commençons par une inégalité "simple" mais très souvent utile.

Proposition 258 (**Inégalité de Markov**)
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.
Pour tout $\varepsilon > 0$, on a

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X|)}{\varepsilon}.$$

Preuve — On rappelle que comme X est intégrable, alors par définition $|X|$ l'es aussi et $\mathbb{E}(|X|)$ existe.
Soit $D = \{\omega \in \Omega \text{ tq } |X(\omega)| \geq \varepsilon\}$.
On a alors l'inégalité de fonctions : $|X| \geq \epsilon \mathbb{1}_D(.)$. (Cela est vrai sur D et sur $\bar{D}$)
Les propriétés de l'espérance (qui est une forme d'intégrale) donnent :

$$\mathbb{E}(|X|) \geq \mathbb{E}(\epsilon \mathbb{1}_D) = \epsilon \mathbb{E}(\mathbb{1}_D) = \epsilon \mathbb{P}(D) = \epsilon \mathbb{P}(|X| \geq \varepsilon).$$

□

Remarque 259 — *On peut montrer de la même façon que pour tout $p \geq 1$, si $X^p \in L^1$, alors pour tout $\varepsilon > 0$ on a*

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X^p|)}{\varepsilon^p}.$$

Cela vient de l'inégalité de fonctions :

$$|X|^p \geq \varepsilon^p \mathbb{1}_{[\varepsilon;+\infty[}(|X|),$$

que l'on combine aux propriétés de l'espérance.
En particulier, en prenant $p = 2$ et $Y = X - \mathbb{E}(X)$, on obtient la proposition suivante

Proposition 260 (**Inégalité de Bienaymé-Tchebychev**)
Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$.
Alors, pour tout $\varepsilon > 0$ on a

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2} = \frac{\sigma_X^2}{\varepsilon^2}.$$

Remarque 261 — *L'inégalité de Bienaymé-Tchebychev majore la probabilité que la v.a. X s'éloigne d'au moins ε de sa moyenne.*
Cette inégalité va nous servir à démontrer le prochain théorème, la loi faible des grands nombres.

Théorème 262 (**Loi faible des grands nombres**)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. discrètes, de carré intégrable, indépendantes, et de même loi de probabilité.
Soit $S_n = \sum_{k=1}^{n} X_k$.
Alors pour tout $\varepsilon > 0$, on a

$$\lim_{n\to+\infty} \mathbb{P}\left(|\frac{S_n}{n} - \mathbb{E}(X_1)| \geq \varepsilon\right) = 0.$$

Preuve — On va utiliser les résultats précédents sur la v.a. $\frac{X_1+\ldots+X_n}{n}$.
Comme les v.a. X_n ont la même loi de probabilité, on a en particulier que $\mathbb{E}(X_n) = \mathbb{E}(X_1)$ et $Var(X_n) = Var(X_1)$, $\forall n \geq 1$.
Avec la linéarité de l'espérance, on obtient donc

$$\mathbb{E}\left(\frac{S_n}{n}\right) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}(X_k) = \mathbb{E}(X_1).$$

Avec l'indépendance des X_k, le calcul de variance donne

$$Var(\frac{S_n}{n}) = \frac{1}{n^2}\sum_{k=1}^{n} Var(X_k) = \frac{Var(X_1)}{n}.$$

On applique alors l'inégalité de Bienaymé-Tchebychev à la v.a. $\frac{S_n}{n}$:

$$0 \leq \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{Var(X_1)}{n\varepsilon^2}.$$

Le terme de droite tend vers 0 quand $n \to +\infty$, ce qui conclut. □

Remarque 263 — *La loi faible des grands nombrs nous dit que la suite de v.a. $(\frac{1}{n}S_n)_{n\geq 1}$ converge (en un certain sens) vers la v.a. constante $\mathbb{E}(X_1)$.*
• *Cette suite de v.a. $(X_n)_n$ représente une suite de réalisations de la même expérience (on relance la même pièce autant de fois que l'on veut, le même dé autant de fois que l'on veut) et cela de façon indépendante (chaque relance ne tient pas compte des résultats précédents). Ce théorème nous dit alors que le résultat moyen de n réalisations de la même expérience ($\frac{S_n}{n}$) a une probabilité 1 de converger vers l'espérance $\mathbb{E}(X_1)$.*
• *Elle permet de mieux comprendre l'approche des probabilités basée sur la notion de fréquence, celle que nous avons utilisée en début de ce cours.*
• *Attention, l'hypothèse d'indépendance est indispensable : Si on considère une variable aléatoire X non constante et $(X_n)_{n\geq 1}$ avec $X_n = X$. Alors on a $\frac{1}{n}S_n = X$, et $\mathbb{P}\left(\left|\frac{1}{n}S_n - \mathbb{E}(X)\right| \geq \varepsilon\right)\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \neq 0$.*

Exemple 264 — *Preuve du théorème d'approximation de Weierstrass par les probabilités.*
Soit $f : [0,1] \to \mathbb{R}$ une fonction continue. Alors il existe une suite de polynômes $(B_n)_{n\geq 1}$ qui converge uniformément vers f sur $[0,1]$.
Preuve : *On pose*

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k(1-x)^{n-k},$$

avec la convention $0^0 = 1$.
Le polynôme B_n s'appelle polynôme de Bernstein de degré au plus n associé à f.
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Pour $x \in]0,1[$ fixé, on considère une suite $(X_n)_{n\geq 1}$ de v.a. de loi de probas de Bernouilli de paramètre x, et indépendantes.
On pose $S_n = \sum_{i=1}^{n} X_i$. On va pouvoir appliquer des résultats précédents à la v.a. S_n.

1. *La loi de S_n est binomiale : pour tout $k \in \{0, \cdots, n\}$, $\mathbb{P}(S_n = k) = \binom{n}{k} x^k(1-x)^{n-k}$.*
2. *Le théorème de transfert nous dit que*

$$\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k(1-x)^{n-k}.$$

3. Pour tout $\varepsilon > 0$, on pose

$$\delta(\varepsilon) = \sup\left\{|f(x) - f(y)|,\ x, y \in [0,1] \text{ et } |x-y| \leq \varepsilon\right\}.$$

Le théorème de Heine nous dit que f est uniformément continue et donc $\delta(\varepsilon)$ tend vers 0 quand ε tend vers 0.
Alors, pour tout $x \in]0,1[$, on a

$$|B_n(x) - f(x)| = \left|\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] - f(x)\right|,$$

et donc

$$\begin{aligned} |B_n(x) - f(x)| &\leq \mathbb{E}\left[\mathbb{1}_{|\frac{S_n}{n} - x| < \varepsilon}\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right] \\ &+ \mathbb{E}\left[\mathbb{1}_{|\frac{S_n}{n} - x| \geq \varepsilon}\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right] \\ &\leq \delta(\varepsilon) + 2\|f\|_\infty \mathbb{P}\left(\left|\frac{S_n}{n} - x\right| \geq \varepsilon\right) \end{aligned}$$

Mais d'après l'inégalité de Bienaymé-Tchebychev, on a

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\sigma^2_{\frac{S_n}{n}}.$$

Or

$$\mathbb{E}\left(\frac{S_n}{n}\right) = x$$

et les variables X_n étant indépendantes, on a

$$\sigma^2_{\frac{S_n}{n}} = \frac{1}{n^2}\sigma^2_{S_n} = \frac{1}{n}\sigma^2_{X_1}.$$

On sait enfin que $\sigma^2_{X_1} = x(1-x) \leq 1$ et donc

$$|B_n(x) - f(x)| \leq \delta(\varepsilon) + 2\|f\|_\infty \frac{1}{n\varepsilon^2}.$$

Comme $\lim\limits_{\varepsilon \to 0} \delta(\varepsilon) = 0$, pour tout $\varepsilon' > 0$, il existe $\delta' > 0$, tel que $\varepsilon < \delta'$ implique $\delta(\varepsilon) \leq \frac{\varepsilon'}{2}$.
Et il existe $N \in \mathbb{N}$, tel que $n \geq N$ implique $2\|f\|_\infty \frac{1}{n\varepsilon^2} \leq \frac{\varepsilon'}{2}$. Et finalement, on a montré que pour $n \geq N$,

$$|B_n(x) - f(x)| \leq \delta(\varepsilon) + 2\|f\|_\infty \frac{1}{n\varepsilon^2} \leq \varepsilon',$$

la majoration ne dépendant pas de x, la convergence uniforme est donc bien prouvée.

4.2 Approximation d'une loi de Poisson

Théorème 265 (**Théorème de Poisson**)
Soit $\lambda > 0$. Soit $(p_n)_{n\in\mathbb{N}}$ une suite de réels dans $]0,1[$ telle que $\lim\limits_{n\to+\infty} np_n = \lambda$.
Soit $(X_n)_{n\geq 1}$ une suite de v.a., telle que X_n est de loi binomiale $B(n, p_n)$.
Alors, pour tout entier $k \geq 0$, la suite de probabilités $(\mathbb{P}(X_n = k))_n$ converge. On a

$$\lim_{n\to+\infty} \mathbb{P}(X_n = k) = \exp(-\lambda)\frac{\lambda^k}{k!}.$$

Preuve — Soit $k \geq 0$ fixé.
Comme X_n est une v.a. de loi binomiale de paramètres n et p_n, on a

$$\mathbb{P}(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}.$$

Par hypothèse, on a

$$p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right).$$

On obtient alors :

$$\mathbb{P}(X_n = k) = \binom{n}{k} \left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^k \left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^{n-k}.$$

Quand n tend vers $+\infty$, le de gauche est équivalent à

$$\binom{n}{k} \left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^k = \frac{n(n-1)\cdots(n-k+1)}{k! n^k} [\lambda + o(1)]^k \sim \frac{\lambda^k}{k!}.$$

Le terme de droite est équivalent à

$$\left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^{n-k} \sim \exp(-\lambda).$$

Avec les propriétés des équivalents, on obtient donc

$$\mathbb{P}(X_n = k) \sim_{n\to+\infty} \exp(-\lambda)\frac{\lambda^k}{k!}.$$

□

Remarque 266 —

1. *Pour approximer correctement une loi de probabilités avec une suite de v.a., nous avons déjà remarqué que l'espérance et l'écart-type doivent coïncider.*
   L'espérance d'une variable aléatoire de loi $B(n, p_n)$ est np_n qui tend bien vers λ, l'espérance de la loi de Poisson correspondante. L'écart-type est $np_n(1-p_n)$ qui tend bien vers λ puisque (p_n) tend vers 0.
2. *On peut améliorer le résultat en obtenant une information sur la vitesse de convergence des $(\mathbb{P}(X_n = k))_n$:*

$$\sum_{k=0}^{+\infty} \left|\mathbb{P}(X_n = k) - \exp(-\lambda)\frac{\lambda^k}{k!}\right| \leq \frac{2\lambda}{n} \min(2, \lambda).$$

   La preuve est cependant plus technique et sort du cadre de ce cours.

4.3 Convergences de v.a.

Nous allons définir trois notions de convergence pour les suites de v.a. (pour les v.a. réelles et discrètes).

Définition 267
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient X et $(X_n)_{n\geq 0}$ des v.a.r. discrètes.
On dit que

1. *$(X_n)_n$ converge **en probabilité** vers X si*

$$\lim_{n\to+\infty} \mathbb{P}(|X - X_n| \geq \varepsilon) = 0, \forall \varepsilon > 0$$

2. *$(X_n)_n$ converge **presque sûrement** vers X si*

$$\mathbb{P}(\{\omega \in \Omega \text{ tq } \lim_{n\to+\infty} X_n(\omega) \text{ existe et vaut } X(\omega)\}) = 1.$$

   On notera parfois $\mathbb{P}(\lim_{n\to+\infty} X_n = X) = 1$.
3. *$(X_n)_n$ converge **dans** L^2 vers X si ces v.a. sont de carré intégrable et si $\mathbb{E}\left[(X - X_n)^2\right]$ converge vers 0.*

Exemple 268 — *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(A_n)_{n\in\mathbb{N}}$ des éléments de $\mathcal{A}$.*

1. *La suite de v.a.* $(\mathbb{1}_{A_n})_{n\geq 0}$ *converge en probabilités vers* 0 *si et seulement si* $\lim_{n\to+\infty} \mathbb{P}(A_n) = 0$.
   En effet, pour tout $\varepsilon > 0$ *avec* $\varepsilon < 1$, *on a*
$$\mathbb{P}(|\mathbb{1}_{A_n} - 0| > \varepsilon) = \mathbb{P}(A_n).$$
   Ainsi, le terme de gauche tend vers 0 *ssi celui de droite tend vers* 0.
2. *La suite de v.a.* $(\mathbb{1}_{A_n})_{n\geq 0}$ *converge dans* L^2 *vers* 0 *si et seulement si* $\mathbb{P}(A_n)$ *tend vers* 0.
   En effet, on a $\mathbb{E}((\mathbb{1}_{A_n} - 0)^2) = \mathbb{E}(\mathbb{1}_{A_n}) = \mathbb{P}(A_n)$.
   Dans cet exemple, la cv en probabilités est équivalente à la cv L^2.
3. *On suppose que la suite* $(\mathbb{1}_{A_n})_{n\geq 0}$ *converge presque sûrement vers* 0.
   Soit $\omega \in \Omega$ *tel que* $\mathbb{1}_{A_n}(\omega)$ *tend vers* 0. *Alors la suite* $(\mathbb{1}_{A_n}(\omega))_n$ *est stationnaire à partir d'un certain rang et vaut* 0.
   Donc ω *n'appartient pas à un nombre infini de* A_n. *Par hypothèse de CV presque sûre, on en déduit que*
$$\mathbb{P}(\limsup A_n) = \mathbb{P}\left(\bigcap_{n\in\mathbb{N}} \bigcup_{p\geq n} Ap\right) = 0.$$
   Par propriété de limite décroissante, on en déduit que $\lim_{n\to+\infty} \mathbb{P}\left(\cup_{p\geq n} A_p\right) = 0$. *Comme* $A_n \subset \cup_{p\geq n} A_p$, *on obtient alors* $\lim_{n\to+\infty} \mathbb{P}(A_n) = 0$. *Dans cet exemple, la convergence presque sûre implique la convergence en probabilités (ou la cv* L^2*).*
4. *On prend maintenant*
$$\begin{array}{llll} A_1 = [0,1] & & & \\ A_2 = [0,1/2] & A_3 =]1/2,1] & & \\ A_4 = [0,1/3] & A_5 =]1/3,2/3] & A_6 =]2/3,1] & \\ A_7 = [0,1/4] & A_8 =]1/4,1/2] & A_9 =]1/2,3/4] & A_{10} =]3/4,1] \\ \quad \cdots & & & \end{array}$$
   On a alors $\lim \mathbb{P}(A_n) = 0$ *et* $\limsup A_n = [0,1]$. *Donc* $\mathbb{P}\big(\limsup A_n\big) = 1$.
   D'après les points précédents, la suite $(\mathbb{1}_{A_n})_{n\geq 0}$ *cv en probabilités (et dans* L^2*), mais elle ne converge pas presque sûrement.*

Remarque 269 — *La loi faible des grands nombres nous dit que pour* $(X_n)_n$ *une suite de v.a. indépendantes et de même loi de probas, la suite des moyennes* $S_n = \dfrac{X_1 + \cdots + X_n}{n}$ *converge en probabilité vers la v.a. constante* $\mathbb{E}(X_1)$.

La convergence en probabilité ou la cv dans L^2 n'impliquent pas la convergence presque sure !

Remarque 270 — *Ces trois notions de convergence sont distinctes dans le cas général.*
La convergence L^2 *implique la convergence en probabilités.*
La convergence presque sûre implique la convergence en probabilités.
Toutes les autres implications sont fausses en général.

Chapitre 5 Chaîne de Markov

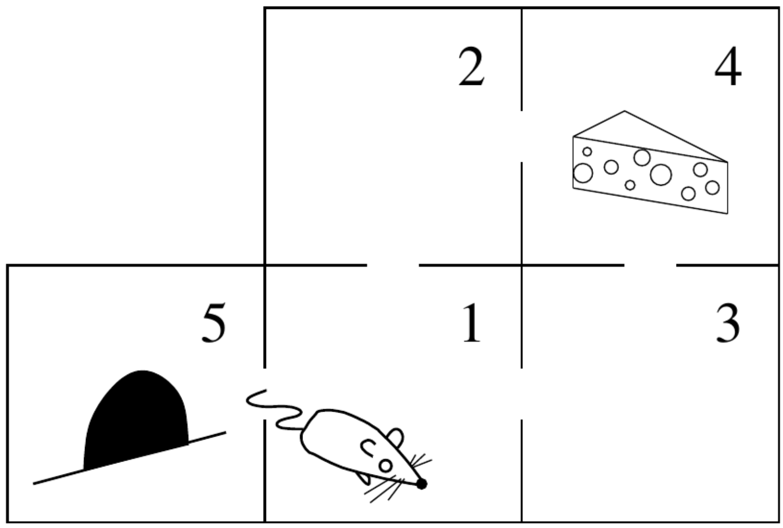
Table des matières du chapitre

5.1 Exemples pour introduire les chaînes de Markov

Les chaînes de Markov sont intuitivement très simples à définir.
Une chaîne de Markov est une suite de variables aléatoires. (une suite $(X_n)_n$)
Cela représente l'évolution d'un système au cours du temps, de façon discrète.
Et, on veut qu'à chaque instant n, la v.a. suivante X_{n+1} ne dépende que de l'état de la v.a. actuelle X_n (et du hasard), mais pas de l'état des v.a. précédentes $(X_{n-1}, X_{n-2}, \ldots)$.

Voyons quelques exemples qui montrent ce principe.

Exemple 271 (La souris dans le labyrinthe) — *Une souris se déplace dans le labyrinthe de la figure ci-dessous*



Initialement, elle se trouve dans la case 1.
A chaque minute, elle change de case en choisissant, de manière équiprobable, l'une des cases adjacentes.
Dès qu'elle atteint soit la nourriture (case 4), soit sa tanière (case 5), elle y reste.

Le mouvement de la souris au temps $n+1$ dépend ainsi uniquement de sa position au temps n (et du hasard), mais pas des positions précédentes (au temps $n-1, n-2, \ldots$).
On se pose alors les questions suivantes :

1. *Avec quelle probabilité la souris atteint-elle la nourriture plutôt que sa tanière ?*
2. *Au bout de combien de temps atteint-elle sa tanière ou la nourriture ?*

On peut essayer de répondre à ces questions en construisant un arbre décrivant les chemins possibles. Par exemple, il est clair que la souris se retrouve dans sa tanière au bout d'une minute avec probabilité 1/3. Sinon, elle passe soit dans la case 2, soit dans la case 3, et depuis chacune de ces cases elle a une chance sur deux de trouver la nourriture.
Il y a donc une probabilité de 1/6 que la souris trouve la nourriture au bout de deux minutes.
Dans les autres cas, elle se retrouve dans la case de départ, ce qui permet d'établir une formule de récurrence pour les probabilités cherchées.

Exemple 272 (Le joueur au casino) — *Un joueur vient au casino avec 1000 yuan.*
Toutes les minutes, il joue 200 yuan à la roulette, en misant sur le rouge (proba de $\frac{18}{38}$ de doubler sa mise, et $\frac{20}{38}$ de perdre).
Si le joueur tombe à 0 yuan ou atteint les 2000 yuan, il s'arrête de jouer (il conserve l'argent qui lui reste).
On peut se poser les mêmes questions que pour la souris

1. *Avec quelle probabilité le joueur repart-il avec 2000 yuan ?*
2. *Au bout de combien de temps en moyenne arrête-t-il de jouer ?*

5.2 Chaînes de Markov sur un ensemble fini ou dénombrable

Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé.
On prend un ensemble E fini ou dénombrable, et on considère une suite de v.a. $X_n : \Omega \to E$.
La famille de v.a. $(X_n)_n$ va modéliser un système (souris dans un labyrinthe, joueur au casino, ...). Ce système prend des valeurs dans un certain ensemble E, appelé ensemble d'états, t il évolue au cours du temps de façon discrète.

Pour $x \in E$, le système est ainsi dans l'état x à l'instant $n \geq 0$ avec une probabilité de $\mathbb{P}(X_n = x)$.
Pour que cette suite de v.a. modélise de façon intéressante nos exemples, il faut ajouter une condition supplémentaire.

Définition 273
Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé. Soit E un ensemble fini ou dénombrable. Soit, $(X_n)_{n\geq 0}$ une famille de v.a. de Ω dans E.

La suite $(X_n)_{n\in\mathbb{N}}$ est appelée ***chaîne de Markov*** *si, pour tout $n \in \mathbb{N}$ et tous $a_0, \cdots, a_{n+1} \in E$ tels que $\mathbb{P}((X_n = a_n) \cap \cdots \cap (X_0 = a_0)) > 0$, on a*

$$P((X_{n+1} = a_{n+1}) \mid (X_n = a_n) \cap \cdots \cap (X_0 = a_0)) = P((X_{n+1} = a_{n+1}) \mid (X_n = a_n)).$$

On dit que la chaîne de Markov $(X_n)_n$ est ***finie*** *si l'ensemble E est fini. Elle est infinie sinon.*

La loi de probabilité de X_0 est appelée la ***distribution initiale*** *de la chaîne.*

À chaque instant n, l'état futur X_{n+1} dépend de l'état présent X_n mais pas des états passés $X_0, \cdots, X_{n-1}$.
Une chaîne de Markov modélise donc une évolution aléatoire *sans mémoire*. (par ex. la souris se déplace seulement en fonction de la case où elle se trouve, sans penser au passé)
Pour une chaîne de Markov, comme pour une variable aléatoire en général, on ne cherche pas à décrire proprement l'ensemble Ω, ni les valeurs $X_n(\omega)$ pour chaque ω. On veut étudier de façon probabiliste cette suite de v.a. : probabilité d'arriver à un état x, temps moyen pour arriver à l'état x, nombre moyen de passages par l'état x, ...

Exemple 274 (La souris dans le labyrinthe 2) — *Le déplacement de la souris dans le labyrinthe (voir exemple 271) se modélise par une chaîne de Markov finie $(X_n)_{n\geq 0}$.*
L'ensemble des états sera $E = \{1, 2, 3, 4, 5\}$.
L'énoncé nous donne $X_0 = 1$.
Et, pour tout $n \geq 0$, on a les probabilités conditionnelles suivantes :

$$\begin{gathered}
\mathbb{P}(X_{n+1} = 2|X_n = 1) = \tfrac{1}{3}, \mathbb{P}(X_{n+1} = 3|X_n = 1) = \tfrac{1}{3}, \mathbb{P}(X_{n+1} = 5|X_n = 1) = \tfrac{1}{3} \\
\mathbb{P}(X_{n+1} = 4|X_n = 2) = \tfrac{1}{2}, \mathbb{P}(X_{n+1} = 4|X_n = 2) = \tfrac{1}{2} \\
\mathbb{P}(X_{n+1} = 1|X_n = 3) = \tfrac{1}{2}, \mathbb{P}(X_{n+1} = 4|X_n = 3) = \tfrac{1}{2} \\
\mathbb{P}(X_{n+1} = 4|X_n = 4) = 1 \\
\mathbb{P}(X_{n+1} = 5|X_n = 5) = 1.
\end{gathered}$$

Toutes les autres probabilités conditionnelles de la même forme sont nulles ou n'existent pas (ex : $\mathbb{P}(X_{n+1} = 4|X_n = 1) = 0$, $\mathbb{P}(X_1 = 1|X_0 = 3)$ n'existe pas).
On verra plus loin quelles sont les conditions qui impliquent l'existence de la chaîne de Markov $(X_n)_n$. (On peut faire la construction et la vérification, mais cela est long et moins intéressant.)
On peut alors retrouver par des calculs faciles de probabilités conditionnelles que $\mathbb{P}(X_1 = 5) = \frac{1}{3}$ ou que $\mathbb{P}(X_2 = 4) = \frac{1}{6}$.
On peut utiliser les outils de probabilités des chapitres précédents pour répondre aux questions posées dans les exemples précédents (pour calculer des probabilités et des espérances).

Exemple 275 — *Soit E un ensemble fini ou dénombrable. Soit $(X_n)_n$ une suite de v.a. à valeurs dans E, telle que X_n est indépendante de $X_0, X_1, \ldots, X_{n-1}$.*
Alors la suite $(X_n)_n$ est une chaîne de Markov.
En effet, par indépendance des variables aléatoires, si $\mathbb{P}((X_n = a_n) \cap \cdots \cap (X_0 = a_0)) > 0$, on a

$$\mathbb{P}((X_{n+1} = a_{n+1}) \mid (X_n = a_n) \cap \cdots \cap (X_0 = a_0)) = \mathbb{P}(X_{n+1} = a_{n+1}) = \mathbb{P}((X_{n+1} = a_{n+1}) \mid (X_n = a_n)).$$

Les suites de v.a. indépendantes sont donc des chaînes de Markov.
Cependant, elles ne sont pas intéressantes à étudier en tant que chaînes de Markov. (l'état de X_{n+1} ne dépend pas de $X_0, \ldots, X_{n-1}$, ni de X_n)

Une chaîne de Markov $(X_n)_n$ est en quelque sorte définie par les valeurs $\mathbb{P}(X_{n+1} = y|X_n = x)$, pour tous $x, y \in E$ et tout $n \geq 0$, et par le choix de X_0.
Cela laisse énormément de choix, qui donnent des chaînes de Markov très variées.
Nous allons nous intéresser à une famille un peu plus petite : les chaînes de Markov homogènes finies.

5.3 Chaînes de Markov homogènes finies

5.3.1 Définition, probabilité de transition

Définition 276
Soit $(X_n)_{n\geq 0}$ une chaîne de Markov sur un ensemble E.
La chaîne de Markov $(X_n)_n$ est ***homogène*** *si pour tous $i, j \in E$ on a*

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_n = i), \forall n \geq 0.$$

Les probabilités conditionnelles $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ ne dépendent pas de n.

Pour chaque $(i, j) \in E^2$, on note

$$Q_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_0 = i).$$

Q_{ij} est appelée la ***probabilité de transition*** *de l'état i vers l'état j.*

Une chaîne de Markov $(X_n)_n$ est en quelque sorte définie par les valeurs $Q_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$, pour tous $i, j \in E$ (et par le choix de X_0).

Exemple 277 (Une chaîne de Markov homogène infinie : la marche aléatoire sur $\mathbb{Z}$) — *Un marcheur se déplace sur l'axe $\mathbb{Z}$. On choisit $p \in]0,1[$.*
À chaque pas, le marcheur va vers la droite avec la probabilité p ou vers la gauche avec la probabilité $q = 1 - p$.
Les états sont les positions du marcheur ($E = \mathbb{Z}$) et, pour chaque instant $n \in \mathbb{N}$, la variable aléatoire réelle X_n est la position avant le $(n+1)$-ième pas.
La suite $(X_n)_n$ définit alors une chaîne de Markov infinie (à valeurs dans $\mathbb{Z}$), qui est homogène.

On a

$$Q_{ij} = \begin{cases} p \text{ si } j = i+1 \\ q \text{ si } j = i-1 \\ 0 \text{ sinon} \end{cases}, \forall (i,j) \in \mathbb{Z}^2.$$

Les marches aléatoires (sur $\mathbb{Z}$, sur $\mathbb{Z}^2$, $\mathbb{Z}^3$,...) sont des chaînes de Markov qui ont beaucoup de propriétés très étonnantes.

Si la chaîne de Markov homogène $(X_n)_n$ est de plus finie, alors les Q_{ij} qui la définissent sont en nombre fini. Si $Card(E) = N$ on a N^2 probabilités de transition.
Cela permet de construire une matrice avec les probabilités de transition.

5.3.2 Matrices stochastiques

Définition 278
Soit $N > 0$ un entier. Soit $P = (p_{i,j})_{i,j} \in \mathscr{M}_N(\mathbb{R})$.
*On dit que P est une matrice **stochastique** si ses coefficients $p_{i,j}$ vérifient*

$$0 \leq p_{i,j} \leq 1, \forall 1 \leq i, j \leq n,$$

$$\sum_{j=1}^{N} p_{i,j} = 1, \forall 1 \leq i \leq n.$$

Remarque 279 —

1. *Une matrice stochastique est une matrice à coefficients dans $[0,1]$, et dont la somme des coeffs sur chaque ligne vaut 1.*
2. *La seconde condition est équivalente à : ${}^t(1, \cdots, 1)$ est un vecteur propre de P, pour la valeur propre 1.*
3. *On vérifie facilement que si P et Q sont deux matrices stochastiques, alors le produit PQ est encore une matrice stochastique.*
   En particulier, toutes les puissances P^n de P sont encore des matrices stochastiques.

On va utiliser les matrices stochasiques pour définir des chaînes de Markov homogènes finies.
Les coefficients $p_{i,j}$ seront les probabilités de transition de la chaîne de Markov (de l'état i vers l'état j).

5.3.3 Matrice de transition et matrice stochastique

Définition 280
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $P \in \mathscr{M}_N(\mathbb{R})$ une matrice stochastique.
Soit $(X_n)_{n \geq 0}$ une chaîne de Markov homogène sur E.
*On dit que $(X_n)_n$ a pour **matrice de transition** P si on a*

$$\mathbb{P}(X_{n+1} = j | X_n = i_n, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i_n) = p_{i_n, j},$$

pour tout $n \geq 0$ et tous $i_0, i_1, \cdots, i_n, j$ dans E.

La loi de probabilité de X_0 est appelée la ***distribution initiale*** *de la chaîne.*

Remarque 281 —
• *Pour vérifier que cette définition a du sens, il faut contrôler que les probabilités de transition Q_{ij} forment bien une matrice stochastique, c'est-à-dire que la somme sur tous les $j \in E$ des probabilités $\mathbb{P}(X_{n+1} = j|X_n = i)$ vaut 1.*
En effet, on a $\sum_{j\in\chi} \mathbb{P}(X_{n+1} = j|X_n = i) = \mathbb{P}(X_{n+1} \in E|X_n = i) = \mathbb{P}(\Omega|X_n = i) = 1$.
• *Toute chaîne de Markov homogène finie est donc associée à une matrice stochastique P.*
• *Une chaîne de Markov homogène finie ne dépend ainsi que de sa matrice stochastique P et de la loi de probabilité de X_0.*
• *Rappel : Comme X_0 est une v.a. discrète, sa loi de probabilité est la famille $(\mathbb{P}(X_0 = i))_{i\in E}$. C'est une famille de $[0,1]^N$ dont la somme vaut 1.*
On notera μ cette loi de probabilité (cette famille).

Exemple 282 —
• *Dans l'exemple 271, la trajectoire de la souris définit une chaîne de Markov homogène finie. Sa matrice de transition est*

$$P = \begin{pmatrix} 0 & 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

et sa distribution initiale (la loi de probas de X_0) est $\mu = (1, 0, 0, 0, 0)$.
• *Dans le cas de l'exemple 272 du joueur au casino, l'argent du joueur définit une chaîne de Markov homogène finie (à valeurs dans $E = \{0, 200, 400, \ldots, 1800, 2000\}$).*
Sa matrice de transition est

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{10}{19} & 0 & \frac{9}{19} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{10}{19} & 0 & \frac{9}{19} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{10}{19} & 0 & \frac{9}{19} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{10}{19} & 0 & \frac{9}{19} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{10}{19} & 0 & \frac{9}{19} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{10}{19} & 0 & \frac{9}{19} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{10}{19} & 0 & \frac{9}{19} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{10}{19} & 0 & \frac{9}{19} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{10}{19} & 0 & \frac{9}{19} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

et sa distribution initiale (la loi de probas de X_0) est $\mu = (0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0)$.
• *Dans une urne on place 5 boules, numérotées $1, 2, 3, 4, 4$. A chaque minute n on tire une boule, on regarde son numéro, et on la replace dans l'urne. Les tirages sont tous indépendants.*
Le numéro des boules tirées définit alors une suite de v.a. $(X_n)_n$, à valeurs dans $1, 2, 3, 4$, indépendantes, et toutes de loi de probabilité $(1/5, 1/5, 1/5, 2/5)$.
La suite $(X_n)_n$ est donc une chaîne de Markov, qui est homogène finie.
Sa matrice de transition est

$$P = \begin{pmatrix} 1/5 & 1/5 & 1/5 & 2/5 \\ 1/5 & 1/5 & 1/5 & 2/5 \\ 1/5 & 1/5 & 1/5 & 2/5 \\ 1/5 & 1/5 & 1/5 & 2/5 \end{pmatrix},$$

et sa distribution initiale (la loi de probas de X_0) est $\mu = (1/5, 1/5, 1/5, 2/5)$.

Donnons quelques théorèmes fondamentaux sur la construction de chaînes de Markov homogènes et sur leurs propriétés.

THÉORÈME 283 (Existence de chaînes de Markov homogènes finies)
Soit $N > 0$. Soit $E = \{a_1, \ldots, a_N\}$ un ensemble fini.
Soient $P \in \mathscr{M}_N(\mathbb{R})$ une matrice stochastique,et $\mu = (p_1, \ldots, p_N) \in [0,1]^N$ avec $p_1 + \ldots + p_N = 1$.
Alors, il existe $(X_n)_{n\geq 0}$ une chaîne de Markov homogène sur E, dont la loi de probas de X_0 et μ et dont la matrice de transition est P.

Ce théorème est admis.
Ainsi, il suffit de choisir une matrice stochastique et une loi de probabilité initiale pour construire une chaîne de Markov homogène finie.
Ce théorème justifie l'existence des chaînes de Markov qui modélisaient les exemples précédents (souris dans le labyrinthe, joueur au casino,...).

5.3.4 Matrices stochastiques et calculs de probabilités

THÉORÈME 284
Soit $N > 0$. Soit $E = \{a_1, \ldots, a_N\}$ un ensemble fini.
Soit $(X_n)_{n\geq 0}$ une chaîne de Markov homogène sur E, de distribution initiale μ et de matrice de transition P.
Alors, la loi de probabilité de X_n est donnée par le N-uplet μP^n.
C'est-à-dire,

$$(\mathbb{P}(X_n = a_1), \ldots, \mathbb{P}(X_n = a_N)) = \mu \cdot P^n = (\mathbb{P}(X_0 = a_1), \ldots, \mathbb{P}(X_0 = a_N)) \cdot P^n.$$

Autrement dit, pour tout $a_j \in E$, on a $\mathbb{P}(X_k = a_j) = \sum_{r=1}^N \mu_r (P^k)_{r,j}$.

Preuve — On démontre le résultat par récurrence sur $n \geq 0$.
Initialisation : Pour $n = 0$, on a bien ${}^t\mu = P^0\, {}^t\mu$.
Hérédité : Supposons le résultat vrai pour un $n \geq 0$.
Par hypothèse, on a donc

$$\mathbb{P}(X_n = a_i) = (\mu P^n)_i.$$

On pose $X = \mu P^{n+1} = \begin{pmatrix} x_1 & \ldots x_N \end{pmatrix}$.
Soit $1leqj \leq N$. D'une part, on a

$$x_i = (\mu P^{n+1})_i = \sum_{k=1}^N \mathbb{P}(X_0 = a_k) c_{k,i}.$$

D'autre part, on a

$$\begin{aligned} x_i &= (\mu P^{n+1})_i = ((\mu P^n).P)_i \\ &= \sum_{k=1}^N (\mu P^n)_k p_{k,i} = \sum_{k=1}^N \mathbb{P}(X_n = a_k) p_{k,i} \\ &= \sum_{k=1}^N \mathbb{P}(X_n = a_k)\mathbb{P}(X_{n+1} = a_i | X_n = a_k) \\ &= \sum_{k=1}^N \mathbb{P}(X_{n+1} = a_i, X_n = a_k) = \mathbb{P}(X_{n+1} = a_i), \end{aligned}$$

ce qui termine la récurrence. □

Ainsi, quand on connaît la matrice P et la loi de probas μ, il est très facile de calculer la loi de probas de n'importe quel terme X_n de la chaîne de Markov : un produit de matrices suffit. (calculer P^n, puis μP^n)
Ce résultat est très important en pratique, il simplifie grandement les calculs.
De plus, le calcul de P^n ne dépend pas de la loi de probas μ. On peut donc calculer la loi de probas de X^n pour différentes distributions initiales très facilement (on calcule P^n une seule fois).

EXEMPLE 285 — *Pour la matrice de transition de la souris dans le labyrinthe, on trouve*

$$P^2 = \begin{pmatrix} 1/3 & 0 & 0 & 1/3 & 1/3 \\ 0 & 1/6 & 1/6 & 1/2 & 1/6 \\ 0 & 1/6 & 1/6 & 1/2 & 1/6 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

La ième ligne de la matrice correspond aux probabilités d'être dans les différents états si la souris est partie de la case i.
• *Avec pour distribution initiale* $(1, 0, 0, 0, 0)$*, on voit que la loi de probas après deux mouvements est* $(1/3, 0, 0, 1/3, 1/3)$*. Elle trouvera donc sa tanière (en 5) avec une proba de* $1/3$ *(même chose pour sa nourriture (en 4)).*
• *Si la souris était partie de la case 2 ou de la case 3 (avec choix équiprobable), on aurait une distribution initiale de* $(0, 1/2, 1/2, 0, 0)$*.*
Donc, la loi de probas après deux mouvements serait $(0, 1/6, 1/6, 1/2, 1/6)$*. Dans cette situation, elle aurait une proba de* $1/2$ *de trouver sa tanière et une proba de* $1/6$ *de trouver sa nourriture en deux mouvements.*

Exemple 286 (Prévisions météo) —
Des études sur la météo de Pékin montrent que le climat du jour (soleil, nuages, pluie) varie de façon aléatoire en suivant une chaîne de Markov homogène.
Ainsi, en posant $E = \{soleil,\ nuages,\ pluie\ \}$ *l'ensemble des climats possibles, on peut modéliser l'évolution du climat chaque jour (à partir d'aujourd'hui) par une chaîne de Markov homogène finie* $(X_n)_n$*, dont la matrice de transition est la matrice* P*.*
Des estimations sur les années précédentes donnent :

$$P = \begin{pmatrix} 0.8 & 0.15 & 0.05 \\ 0.5 & 0.3 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}.$$

Ainsi, s'il fait soleil au jour n, on a 80% *de chances d'avoir du soleil le lendemain,* 15% *de chances d'avoir des nuages, et* 5% *de chances d'avoir de la pluie. De même,* $\mathbb{P}(X_{n+1} = pluie|X_n = pluie = 0.5)$*.*
Le 27 Février 2022, il fait soleil. On pose $X_0 = soleil$*. Quelle est la probabilité qu'il fasse soleil une semaine plus tard ?*
Avec les théorèmes précédents, on obtient facilement la réponse : On calcule P^7*. On prend* $\mu = (1, 0, 0)$ *la loi de probas de* X_0*. Puis, on calcule* μP^7*. Cela donnera la loi de probas de* X_7*, la météo à Pékin dans une semaine.*
Avec le calcul de P^7*, on peut facilement répondre à la question : "Il pleut aujourd'hui, quelle est la probabilité qu'il pleuve ou qu'il y ait des nuages une semaine plus tard ?"*
Il suffit de prendre pour distribution initiale $\mu = (0, 0, 1)$*, puis de calculer* μP^7 *pour obtenir* $\mathbb{P}(X_7 \in \{nuages,\ pluie\})$*.*
Il faut bien remarquer qu'avec un tel modèle, même si l'on connaît la matrice de transition P*, le choix de la distribution initiale (de la loi de* X_0*) est tout aussi important.*
Les prévisions météo 7 jours plus tard sont totalement différentes s'il fait beau aujourd'hui (X_0 *=soleil) ou s'il pleut (*X_0 *=pluie). Même chose s'il y a* 70% *de chances d'avoir des nuages et* 30% *de chances d'avoir du soleil aujourd'hui (*$\mu = (0.3, 0.7, 0)$*).*

Les chaînes de Markov homogènes ont une autre propriété parfois très utile.

Théorème 287
Soit $N > 0$*. Soit* $E = \{a_1, \ldots, a_N\}$ *un ensemble fini.*
Soit $(X_n)_{n\geq 0}$ *une chaîne de Markov homogène sur* E*, de distribution initiale* μ *et de matrice de transition* P*.*
Alors, pour tout $n \geq 0$*, pour tous* $i_0, \ldots, i_n \in E$*, on a*

$$\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mu_{i_0} p_{i_0,i_1} p_{i_1,i_2} \ldots p_{i_{n-1},i_n} = \mu_{i_0} \Pi_{k=0}^{n-1} p_{i_k,i_{k+1}}.$$

Ce théorème donne une expression simple de la probabilité $\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)$, c'est-à-dire de la probabilité que la chaîne de Markov $(X_m)_m$ commence par la trajectoire $i_0, i_1, \ldots, i_n$. Ce théorème est admis.

Proposition 288
Soit $N > 0$*. Soit* $E = \{a_1, \ldots, a_N\}$ *un ensemble fini.*
Soit $(X_n)_{n\geq 0}$ *une chaîne de Markov homogène sur* E*, de distribution initiale* μ *et de matrice de transition* P*.*
Soient $0 \leq n < m$*, et* $i_0, \ldots, i_m \in E$*.*

Si on a $\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$, *alors*

$$\mathbb{P}(X_m = i_m, \dots, X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_m = i_m, \dots, X_{n+1} = i_{n+1} | X_n = i_n).$$

Preuve — On utilise $\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B}{\mathbb{P}(B)}$, et on applique le théorème précédent pour transformer $\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)$ et $\mathbb{P}(X_m = i_m, X_{m-1} = i_{m-1}, \dots, X_0 = i_0)$.
Cela donne le résultat. □

Remarque 289 — *Cette proposition montre que pour une chaîne de Markov homogène, l'évolution de la chaîne sur les moments* $\{n+1, \dots, m\}$ *ne dépend que de* X_n *(l'état au temps n), et pas de la trajectoire passée (de* X_0 *à* X_{n-1}*).*

Exemple 290 — *Si on reprend l'exemple 286 de la météo à Pékin chaque jour, on peut utiliser les résultats précédents pour calculer la probabilité d'avoir* 5 *jours de soleil consécutifs (ou plus).*
Cela corresond au fait de commencer par une météo X_0 *ensoleillé, puis d'avoir les météos suivantes* $X_1, \dots, X_4$ *ensoleillées.*
Le choix du numérotage des jours ne compte pas pour une chaîne de Markov homogène, comme l'indique le théorème précédent. (que l'on parte du jour 0 ou du jour 314 *ne changera rien)*
On veut donc calculer $\mathbb{P}(X_4 = \textit{soleil}, \dots, X_0 = \textit{soleil})$.
Cette probabilité vaut $p_{1,1}^4 = 0.8^4 = 0.4096$. *Il y a donc* 41% *de chances que cela se produise.*
Une question que l'on va souvent poser avec cet exemple est : Quel est le nombre moyen de jours de soleil que l'on a en un an à Pékin ?
Il nous faudra d'autres résultats pour pouvoir y répondre (avoir une valeur exacte, ou avoir une valeur approchée).

Proposition 291
Soit $N > 0$. *Soit* $E = \{a_1, \dots, a_N\}$ *un ensemble fini.*
Soit $(X_n)_{n\geq 0}$ *une chaîne de Markov homogène sur* E*, de distribution initiale* μ *et de matrice de transition* P*.*
Soient $a_i, a_j \in E$ *et* $k, m > 0$. *On a*

1. $\mathbb{P}(X_k = a_j | X_0 = a_i) = (P^k)_{i,j}$,
2. $\mathbb{P}(X_{k+m} = a_j | X_m = a_i) = \mathbb{P}(X_k = a_j | X_0 = a_i) = (P^k)_{i,j}$,
3. *La suite* $(X_{nk})_{n\geq 0}$ *est une chaîne de Markov homogène, de matrice de transition* P^k.

Preuve —

1.
2. On décompose $\mathbb{P}(X_{k+m} = a_j | X_m = a_i)$ en une somme sur toutes les trajectoires de $X_{m+1}, \dots, X_{m+k-1}$ possibles, et on applique le théorème précédent. Cela donne :

$$\begin{aligned}\mathbb{P}(X_{k+m} = a_j | X_m = a_i) &= \sum_{a_{i(k-1)}\in E} \cdots \sum_{a_{i(1)}\in E} \mathbb{P}(X_{k+m} \\ &= a_j, X_{k+m-1} = a_{i(k-1)}, \dots, X_{m+1} = a_i(1) | X_m = a_i) \\ &= \sum_{a_{i(k-1)}\in E} \cdots \sum_{a_{i(1)}\in E} p_{i,i(1)} p_{i(2),i(3)} \cdots p_{i(k-1),j}.\end{aligned}$$

   Par définition du produit matriciel, la somme ci-dessus n'est autre que le coefficient (i, j) de la matrice de la matrice P^k.
   Comme le calcul ne dépend pas de la valeur de m, cela prouve les points (1) et (2).
3. Il faut vérifier que la suite $(X_{nk})_n$ satisfait la définition d'une chaîne de Markov. Cela est long à écrire, mais il n'y a pas de difficulté (on utilise le fait que $(X_n)_n$ est une c.M.).
   La propriété d'homogénéité et la valeur de la matrice de transition s'obtiennent avec le point (2).

□

5.3.5 Puissances de matrices stochastiques

Remarque 292 — *Pour une c.M. homogène de matrice de transition* P*, la loi de probas de* X_n *est* μP^n*.*
Si la suite de matrices P^n *convergeait quand n tend vers plus l'infini, alors la suite des lois de probas des* X_n *serait elle aussi convergente (en un certain sens).*

Est-ce que cela peut arriver ou arrive toujours ? Comment calculer la valeur de la limite (si elle existe) ? Nous allons nous intéresser à ces questions.
Comme cela concerne des matrices et de la convergence, les résultats des cours d'Algèbre 3 (valeurs propres, puissances de matrices) et d'Analyse 4 (normes de matrices, convergence) vont intervenir.

PROPOSITION 293
Soient $N > 0$ et $P \in \mathscr{M}_N(\mathbb{R})$ une matrice stochastique. On a

1. *1 est une valeur propre de P.*
2. *Pour $\|.\|_\infty$ la norme l^∞ sur $\mathbb{C}^N$, on a $|||P||| = 1$.*
3. *Pour $\lambda \in \mathbb{C}$ une valeur propre de P, on a $|\lambda| \leq 1$.*
4. *La suite $(P^n)_{n\geq 0}$ possède toujours une sous-suite convergente.*

Preuve —

1. Cela a déjà été démontré, c'est une conséquence immédiate de la condition de matrice stochastique.
2. Il faut faire calcul de norme de matrices, comme en Analyse 4. (montrer que la norme est ≤ 1, puis dire qu'avec le vecteur propre ${}^t(1,\ldots,1)$ la norme est d'au moins 1)
3. Pour λ une valeur propre de P, on a X non-nul tel que $PX = \lambda X$.
   En prenant la norme, on obtient $\|\lambda X\| = \|PX\| \leq |||P||| \|X\| \leq \|X\|$, d'où $|\lambda| \geq 1$.
4. La suite $(P^n)_n$ est bornée en norme. Elle est donc contenue dans une boule fermée de $\mathscr{M}_N(\mathbb{R})$.
   Comme cet ensemble est un $\mathbb{R}$-ev de dimension finie, ses boules fermées sont compactes.
   D'après le cours d'Analyse 4, toute suite d'un ensemble compact possède une sous-suite convergente.

□

REMARQUE 294 —

— *Les matrices stochastiques contiennent beaucoup de matrices très différentes. Il y a des matrices P telles que P^n ne converge pas du tout. Par exemple, toutes les matrices de permutations A_σ sont des matrices stochastiques. La matrice $A_{(123)} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ est une matrice stochastique, mais la suite des $A^n_{(123)}$ ne converge pas car elle est périodique de période 3.*

— *On peut montrer avec le théorème de Jordan que la suite P^n converge si et seulement si les valeurs propres de P sont dans un certain ensemble $(\mathbb{D} \cup \{1\})$.*
Mais, calculer les valeurs propres de P n'est pas simple. En probabilités on connaît en général les matrices P seulement par leurs coefficients, on ne leur connaît pas de relations algébriques (symétrique, polynôme annulateur, etc).
Par exemple, calculer les valeurs propres des matrices de transition dans l'exemple de la souris ou dans celui de la météo n'est pas simple.

— *On va étudier par la suite des graphes, définis à partir des matrices de transition, qui permettront d'avoir facilement certaines informations sur les valeurs propres de P.*

5.4 Chaînes de Markov absorbantes

DÉFINITION 295
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n\geq 0}$ une c.M. homogène sur E de matrice de transition P.
On définit la relation $\hookrightarrow$ sur E par : $a_i \hookrightarrow a_j$ s'il existe $n \geq 0$ tel que $(P^n)_{i,j} > 0$.
*On dit alors que a_j est **accessible** à partir de a_i.*
On a $a_i \hookrightarrow a_j$ si, en partant de i, on a une probabilité positive d'atteindre j en un nombre fini de pas.
On définit la relation $\sim$ sur E par : $a_i \sim a_j$ si $a_i \hookrightarrow a_j$ et $a_j \hookrightarrow a_i$.

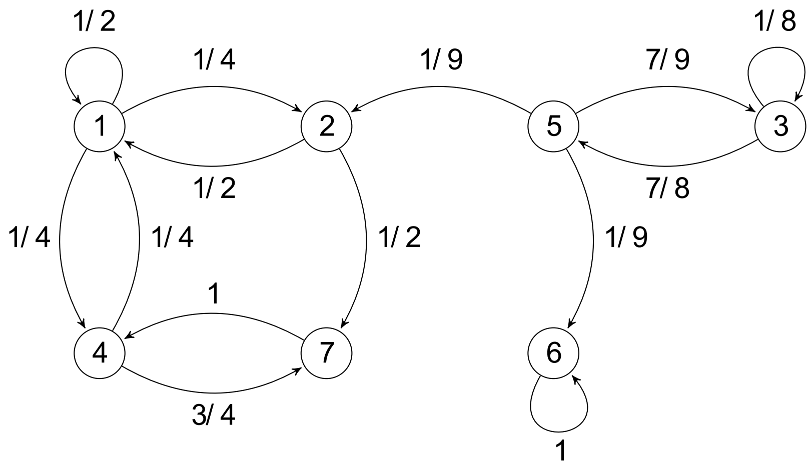
REMARQUE 296 —
• *La relation $\sim$ est une relation d'équivalence.*
*Une classe d'équivalence pour cette relation est appelée **classe d'état**.*
• *On peut associer une c.M. homogène à un graphe, en reliant le sommet a_i au sommet a_j si $p_{i,j} > 0$.*

La relation $\sim$ va permettre de découper ce graphe avec ses classes d'équivalence. On pourra alors réordonner les éléments de E en suivant les classes d'équivalence, et cela donnera des informations sur la matrice P.

Exemple 297 — *On considère la chaîne de Markov $(X_n)_{n\geq 0}$ de matrice de transition P donnée par*

$$P = \begin{pmatrix} 1/2 & 1/4 & 0 & 1/4 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1/8 & 0 & 7/8 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 0 & 0 & 3/4 \\ 0 & 1/9 & 7/9 & 0 & 0 & 1/9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Le graphe de la chaîne associée est



Trouver les classes d'état de cette chaîne de Markov.

Définition 298
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n\geq 0}$ une c.M. homogène sur E de matrice de transition P.
On dit que $a_i \in E$ est ***un état absorbant*** *si $p_{i,i} = 1$ (et donc $p_{i,j} = 0$ pour tout $j \neq i$).*
On dit que la c.M. $(X_n)_n$ est ***absorbante*** *si pour tout $a_i \in E$ il existe un état absorbant a_j tel que $a_i \hookrightarrow a_j$ (un état absorbant est accessible depuis a_i).*

Remarque 299 — *Les exemples de la souris et du jeu de Pile ou Face sont des exemples de chaînes absorbantes avec deux états absorbants.*
L'exemple de la météo n'a pas d'états absorbants.
L'exemple précédent possède 1 état absorbant.
Nous allons étudier les chaînes de Markov absorbantes avec r états absorbants.

Tout d'abord, commençons par une proposition sur les chaînes de Markov avec r états absorbants. Pour une expression plus simple, on choisit $a_{N-r+1}, \ldots, a_N$ comme états absorbants (quitte à réordonner les éléments de E).

Proposition 300
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n\geq 0}$ une c.M. homogène sur E. Soient $a_{N-r+1}, \ldots, a_N$ les r états absorbants de la c.M.
Alors, on a

1. *La matrice de transition P est de la forme*

$$P = \begin{pmatrix} Q & R \\ 0 & I_r \end{pmatrix},$$

avec $Q \in \mathscr{M}_{N-r}(\mathbb{R})$, R de taille $(N-r) \times r$.

2. $P^n = \begin{pmatrix} Q^n & [I + Q + \cdots + Q^{n-1}]R \\ 0 & I_r \end{pmatrix}$, $\forall n \geq 1$.

Preuve —

1. La forme de P découle du fait que chaque état $a_{N-r+1}, \ldots, a_N$ est absorbant.
2. Cette relation se démontre par récurrence sur $n \geq 1$.

□

PROPOSITION 301

Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n \geq 0}$ une c.M. homogène et absorbante sur E. Soient $a_{N-r+1}, \ldots, a_N$ les r états absorbants de la c.M., et soit P sa matrice de transition.
Alors, on a

1. $\lim_{n \to +\infty} Q^n = 0$.
2. *La matrice $I - Q$ est inversible, d'inverse* $(I-Q)^{-1} = \sum_{k=0}^{+\infty} Q^k$.
3. $$\lim_{n \to +\infty} P^n = \lim_{n \to +\infty} \begin{pmatrix} Q^n & [I + Q + \cdots + Q^{n-1}]R \\ 0 & I_r \end{pmatrix} = \begin{pmatrix} 0 & (I-Q)^{-1}R \\ 0 & I \end{pmatrix}.$$

Preuve —

1. Pour $i, j \leq q$, le coefficient $(Q^n)_{i,j}$ de Q^n est la probabilité de se trouver dans l'état non absorbant j, après n pas, partant de i.
   A contrario, pour $j > q$, le coefficient $\left([I + Q + \cdots + Q^{n-1}]R\right)_{i,j-q}$ représente la probabilité de se trouver à l'état absorbant j en au plus n pas.
   On en déduit que pour $i \leq q$ et $j > q$ fixé, la suite $\left([I + Q + \cdots + Q^{n-1}]R\right)_{i,j-q}$ est croissante. L'hypothèse P absorbante implique que pour $i \leq q$, il existe un état absorbant $j > q$ et un rang N_i tel que pour $n \geq N_i$, $\left([I + Q + \cdots + Q^{n-1}]R\right)_{i,j-q} > 0$.
   Soit $N = \max(N_i,\ i \in [\![1, q]\!])$. Par construction, $|||Q^N|||_\infty < 1$, où $|||A|||_\infty = \max_i \sum_{j=1}^{n} |a_{i,j}|$ est la norme subordonnée à la norme $\|(x_1, \cdots, x_n)\|_\infty = \max_i(|x_i|)$.
   Donc $(Q^N)^n$ tend vers 0, puisque $|||(Q^N)^n|||_\infty \leq |||Q^N|||_\infty^n$.
   Pour $i \in [\![1, q]\!]$, la somme des coefficients $(Q^n)_{i,j}$ formant une suite décroissante, on en déduit que $|||Q^n|||_\infty$ décroît et donc tend vers 0.
   De plus, pour $j > q$, le coefficient $\left(Q^{n-1}R\right)_{i,j-q}$ représente la probabilité d'arriver à l'état absorbant j en n pas en partant de i.
2. On a
   $$(I-Q)\sum_{k=0}^{n} Q^k = I - Q^{n+1}$$
   On en déduit que $(I-Q)\sum_{k=0}^{+\infty} Q^k = I$, ce qui montre que $I - Q$ est inversible d'inverse $\sum_{k=0}^{+\infty} Q^k$.
3. Cela découle des points précédents.

□

REMARQUE 302 — *La matrice $F = (I-Q)^{-1} = \sum_{k=0}^{+\infty} Q^k$ est appelée la* ***matrice fondamentale*** *de la c.M. absorbante.*

THÉORÈME 303

Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n \geq 0}$ une c.M. homogène et absorbante sur E. Soient $a_{N-r+1}, \ldots, a_N$ les r états absorbants de la c.M., P sa matrice de transition, et μ sa distribution initiale. On note $F = (f_{i,j}) = (I-Q)^{-1}$.

1. *Pour $1 \leq j \leq N - r$, on définit $Y_{n,j} = \mathbb{1}_{\{X_n = j\}}$.*

Alors, le nombre moyen de passages par l'état a_j avant d'atteindre un état absorbant vaut

$$\mathbb{E}(\sum_{n\geq 0} Y_{n,j}) = \sum_{i=1}^{N-r} \mu_i f_{i,j}.$$

2. *Soit $\tau = \inf\{nt.q.X_n > N-r\}$ la v.a. donnant le temps avant d'atteindre un état absorbant. Alors, le temps moyen avant d'atteindre un état absorbant vaut*

$$\mathbb{E}(\tau) = \sum_{i=1}^{N-r} \mu_i \left(\sum_{j=1}^{q} f_{i,j} \right).$$

3. *Soit $N-r+1 \leq j \leq N$. On pose $B = (b_{i,j}) = (I-Q)^{-1}R$. Alors, la probabilité de s'arrêter à l'état absorbant a_j vaut*

$$\mathbb{P}(X_\tau = a_j) = \mu_j + \sum_{k=1}^{N-r} \mu_k b_{k,j}.$$

Preuve —

1. Soit $1 \leq j \leq N-r$.
La quantité $\mathbb{E}(\sum_{n\geq 0} Y_{n,j})$ représente le nombre moyen de passages par l'état a_j avant d'atteindre un état absorbant. Comme les $Y_{n,j}$ sont des v.a. de Bernouilli, on a

$$\begin{aligned}\mathbb{E}(Y_{n,j}) = \mathbb{P}(X_n = a_j) &= \sum_{i=1}^{N} \mu_i (P^n)_{i,j} \\ &= \sum_{i=1}^{N-r} \mu_i (P^n)_{i,j} \\ &= \sum_{i=1}^{N-r} \mu_i (Q^n)_{i,j}\end{aligned}$$

D'autre part, comme $(I-Q)^{-1} = \sum_{n\geq 0} Q^n$, on a $f_{i,j} = \sum_{n\geq 0} (Q^n)_{i,j}$.

Ainsi,

$$\mathbb{E}(\sum_{n\geq 0} Y_{n,j}) = \sum_{n\geq 0} \sum_{i=1}^{N-r} \mu_i (Q^n)_{i,j} = \sum_{i=1}^{N-r} \mu_i f_{i,j}.$$

2. La v.a. $\tau = \inf\{nt.q.X_n > N-r\}$ la v.a. donne le temps avant d'atteindre un état absorbant. La quantité $\mathbb{E}(\tau)$ représente le temps moyen avant d'atteindre un état absorbant.
Un premier calcul donne

$$\mathbb{E}(\tau) = \sum_{n\geq 0} n\mathbb{P}(\tau = n) = \sum_{n\geq 0} \mathbb{P}(\tau > n).$$

Puis,

$$\begin{aligned}\mathbb{E}(\tau) = \sum_{n\geq 0} \mathbb{P}(\tau > n) &= \sum_{n\geq 0} \mathbb{E}(\mathbb{1}_{\{X_n \leq N-r\}}) \\ &= \mathbb{E}(\sum_{n\geq 0} \sum_{j=1}^{N-r} Y_{n,j}) \\ &= \sum_{j=1}^{q} f_{i,j}.\end{aligned}$$

3. Admis.

□

Exemple 304 (Jeu de Pile ou Face) —
Hugues et Alexandre jouent à la variante suivante de Pile ou Face. Ils jettent une pièce de monnaie (parfaitement équilibrée) de manière répétée. Hugues gagne dès que la pièce tombe trois fois de suite sur Face, alors que Alexandre gagne dès que la suite Pile-Face-Pile apparaît. On se pose les questions suivantes :

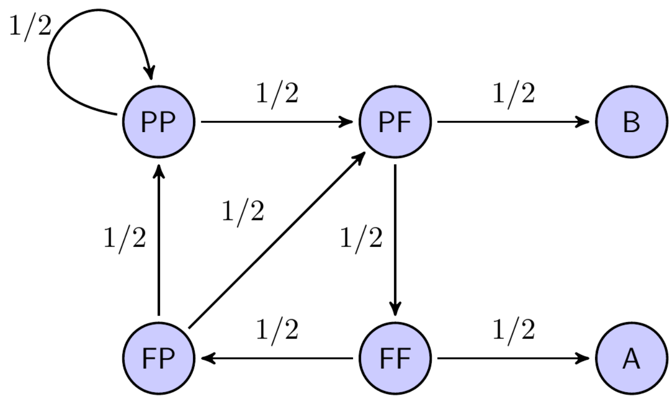
1. *Avec quelle probabilité est-ce Hugues qui gagne le jeu ?*

2. Au bout de combien de jets de la pièce l'un des deux joueurs gagne-t-il ?

Un peu de réflexion montre que si personne n'a gagné au bout de n jets de la pièce, la probabilité que l'un des deux joueurs gagne au coup suivant ne dépend que des deux derniers résultats.
On peut modéliser ce jeu par une chaîne de Markov sur l'ensemble

$$E = \{PP, PF, FP, FF, H \text{ gagne}, V \text{ gagne}\},$$

où PP signifie que la pièce est tombée sur Pile lors des deux derniers jets.
Cette chaîne de Markov est aussi homogène. On peut remarquer qu'elle est de même absorbante (ses deux derniers états sont absorbants, et tous les autres états peuvent mener à un état absorbant avec une probabilité non-nulle).
On détermine les probabilités de transition entre les six états. Cela donne



Ainsi, les matrices P, Q et R sont données par

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \ Q = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \end{pmatrix}, \ R = \begin{pmatrix} 0 & 0 \\ 0 & 1/2 \\ 0 & 0 \\ 1/2 & 0 \end{pmatrix}$$

On calcule alors la matrice fondamentale

$$F = [I - Q]^{-1} = \begin{pmatrix} 7/3 & 4/3 & 1/3 & 2/3 \\ 1/3 & 4/3 & 1/3 & 2/3 \\ 4/3 & 4/3 & 4/3 & 2/3 \\ 2/3 & 2/3 & 2/3 & 4/3 \end{pmatrix}$$

et la matrice donnant les probabilités d'absorption

$$B = F(I - Q)^{-1}R = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \\ 1/3 & 2/3 \\ 2/3 & 1/3 \end{pmatrix}.$$

- *Quelles sont les chances pour Hugues et Valentin de gagner ? Ainsi, en partant de l'un des états PP, PF ou FP, Hugues gagne avec probabilité 1/3, et Alexandre gagne avec probabilité 2/3.*

En partant de l'état FF, c'est Alexandre qui gagne avec probabilité 1/3, et Hugues qui gagne avec probabilité 2/3. Comme la pièce est équilibrée, il y a la même probabilité d'avoir PP,FP,PF,FF en deux lancers. Donc la distribution initiale qui correspond au jeu est $\mu = (1/4, 1/4, 1/4, 1/4, 0, 0)$.
Par conséquent, Hugues gagne le jeu avec une probabilité

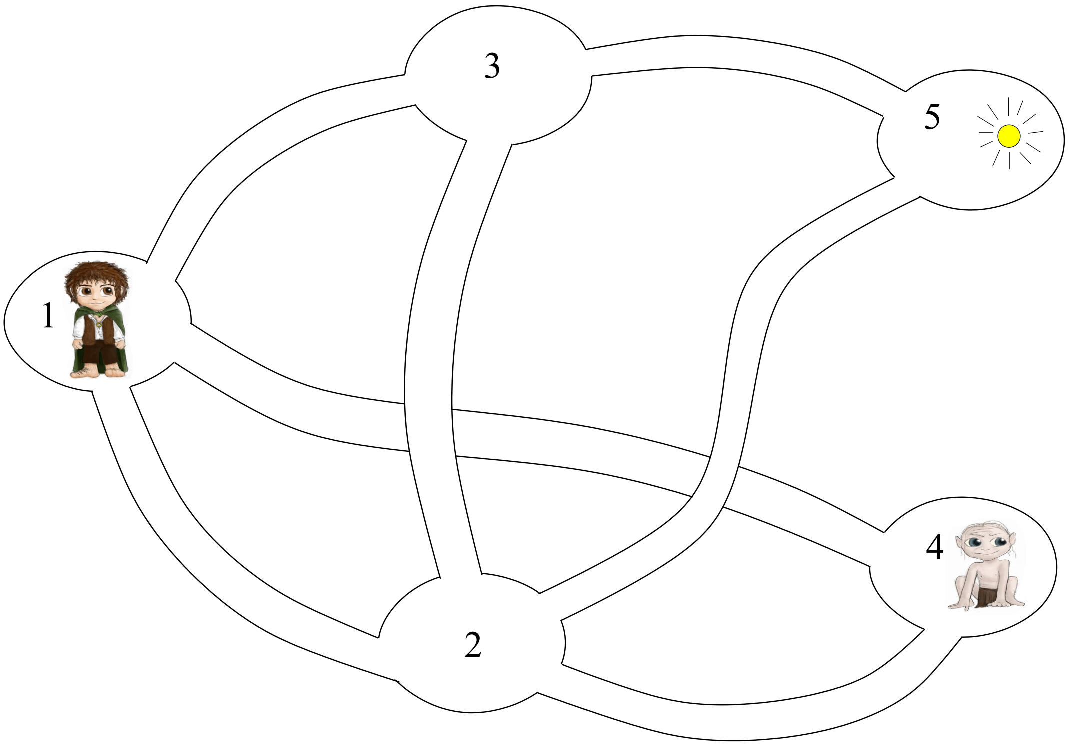
$$\mathbb{P}\{X_\tau = \text{“}H \text{ gagne”}\} = \sum_{i=1}^{4} \mu_i b_{i,1} = \frac{5}{12}.$$

Le jeu est donc déséquilibré entre Hugues et Valentin. La loi faible des grands nombres indique à Hugues qu'il ne vaut alors mieux pas jouer à ce jeu.

- *Quelle est la durée moyenne du jeu ?*

Calculons d'abord le temps moyen pour que la c.M. atteingne un état absorbant.
La somme des éléments de la ligne i de F donne l'espérance du temps d'absorption partant de i, donc par exemple $E_1(\tau) = 14/3$.
En moyennant avec la distribution initiale (utiliser la formule (2) du théorème), on trouve $E_\mu(\tau) = 23/6$. La durée moyenne du jeu est donc de $2 + 23/6 = 35/6$, soit un peu moins de 6 jets de pièce.

EXEMPLE 305 — *Bilbo le Hobbit est plongé dans le soir des cavernes de Gollum. Il ne peut rien voir et se déplace au hasard. Quelle est la probabilité qu'il trouve la sortie (le jeu est fini) ou sur le Gollum (censuré !) ?*



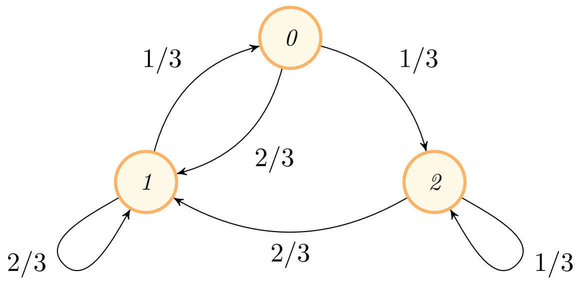
Il faut écrire les matrices P, Q, R F et B. Nous le ferons en TD.

Pour une chaîne de Markov absorbante, nous avons un théorème qui nous permet d'obtenir toutes les informations qui nous intéressent.
Nous allons définir une seconde famille de chaînes de Markov homogènes : les c.M. irréductibles.

5.5 Chaînes de Markov irréductibles, régulières

DÉFINITION 306
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n\geq 0}$ une c.M. homogène sur E de matrice de transition P.
On dit que la c.M. est ***irréductible*** *(ou ergodique si $i \sim j$, $\forall i, j \in E$.*
On dit que la c.M. est ***régulière*** *s'il existe $n \geq 1$ tel que tous les coefficients de P^n sont strictement positifs.*

EXERCICE 307 — *Le graphe suivant représente une chaîne de Markov :*



1. *Écrire la matrice de transition.*
2. *Cette chaîne de Markov est-elle régulière ? irréductible ?*

REMARQUE 308 —
• *Une chaîne de Markov régulière est irréductible, car tout état a_i est accessible depuis tout autre état a_j en n pas au plus.*
La réciproque n'est pas vraie.
Dans les énoncés, vous pouvez vous attendre à des matrices de transition dont tous les coefficients sont strictement positifs, ce qui donne donc des c.M. régulières.
• *Une chaîne de Markov irréductible n'a pas d'état absorbant, sauf si E ne contient qu'un élément.*
Les c.M. absorbantes et irréductibles sont donc deux familles de c.M. très différentes.

EXEMPLE 309 —
• *La chaîne de Markov dans l'exemple de la météo est irréductible et régulière.*
• *Une c.M. homogène de matrice de transition* $P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ *est irréductible, mais elle n'est pas régulière.*
En effet, à chaque pas de temps on se déplace avec probabilité 1 d'un état vers un autre. Il n'y a aucun moment où il y a une probabilité non-nulle d'aller d'un état a_i vers tous les états.
Un autre argument est de dire que la suite des P^n est périodique de période 3, et de voir que I_3, P, P^2 ont des coefficients nuls.

DÉFINITION 310
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n\geq 0}$ une c.M. homogène sur E. Soit $A \subset E$ un sous-ensemble de E.
On définit la variable aléatoire $\tau_A = \inf\{n \geq 0 : X_n \in A\}$.
Cette v.a. τ_A représente le ***temps du premier passage*** *dans A.*
Dans le cas où $A = \{i\}$ consiste en un seul point, nous écrirons aussi τ_i au lieu de $\tau_{\{i\}}$.

REMARQUE 311 — *Les c.M. absorbantes vont terminer avec probabilité 1 sur un état absorbant.*
A l'opposé, les c.M. irréductibles repassent toujours sur chacun de leurs états.

PROPOSITION 312
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n\geq 0}$ une c.M. homogène et irréductible sur E. Soit $A \subset E$ un sous-ensemble de E.
Alors, le temps de premier passage en tout sous-ensemble $A \subset \chi$ est fini presque sûrement :

$$\mathbb{P}(\tau_A < \infty) = \lim_{n\to\infty} \mathbb{P}(\tau_A \leq n) = 1.$$

Preuve — Considérons une autre chaîne de Markov de matrice de transition $\widehat{P}$, obtenue à partir de la chaîne de départ en rendant absorbants les états de A :

$$\widehat{p}_{i,j} = \begin{cases} \delta_{i,j} & \text{si } i \in A \\ p_{i,j} & \text{sinon.} \end{cases}$$

Les trajectoires de la chaîne initiale et de la chaîne modifiée coïncident jusqu'au temps τ_A. Il suffit donc de montrer la proposition pour la chaîne absorbante. Or dans ce cas, le résultat est une conséquence directe de la Proposition 301. En effet, la probabilité $\mathbb{P}_i\{\tau_A > n\}$ de ne pas avoir été absorbé jusqu'au temps n, partant de i, est donnée par la somme des $(Q^n)_{i,j}$ sur les $j \in A$, qui tend vers 0. □

REMARQUE 313 — *Il est important de remarquer que das le cadre plus général des c.M. infinies, ce résultat n'est plus forcément vrai.*

THÉORÈME 314
Soit $N > 0$. Soient $E = \{a_1, \dots, a_N\}$ un ensemble fini, et $(X_n)_{n\geq 0}$ une c.M. homogène et régulière sur E. Soit P sa matrice de transition.
Alors il existe des nombres $\pi_1, \cdots, \pi_N \in]0,1]$, dont la somme vaut 1, tels que

$$\lim_{n\to\infty} P^n = \Pi = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_N \\ \pi_1 & \pi_2 & \cdots & \pi_N \\ \vdots & & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_N \end{pmatrix}.$$

De plus, le vecteur ligne $\pi = (\pi_1, \pi_2, \cdots, \pi_N)$ satisfait $\pi P = \pi$.

Preuve — Si la chaîne n'a qu'un état, le résultat est immédiat, donc nous pouvons admettre que $N \geq 2$. Soit $d \geq 0$ le plus petit élément de P.
Alors $d \leq \dfrac{1}{2}$, puisque $Nd \leq 1$.
Soit $Y = (y_1, \cdots, y_N)^T$ tel que pour tout $i \in [\![1, N]\!]$, $y_i \geq 0$. On note

$$0 \leq m_0 = y_{i_0} = \min_{i\in[\![1,N]\!]} y_i \leq y_i \leq M_0 = y_{j_0} = \max_{i\in[\![1,N]\!]} y_i.$$

Soit $Z = PY$. On a

$$z_k = p_{j_0} y_{j_0} + \sum_{\substack{k=1\\k\neq j_0}}^{N} p_{i,k} y_k \geq dM_0 + (1-d)m_0$$

et

$$z_k = p_{i_0} y_{i_0} + \sum_{\substack{k=1\\k\neq i_0}}^{N} p_{i,k} y_k \leq dm_0 + (1-d)M_0.$$

On en déduit

$$m_0 \leq dM_0 + (1-d)m_0 \leq m_1 = \min_{i\in[\![1,N]\!]} z_i \leq \max_{j\in[\![1,N]\!]} z_j = M_1 \leq dm_0 + (1-d)M_0 \leq M_0.$$

De plus,

$$0 \leq M_1 - m_1 \leq (1-2d)(M_0 - m_0).$$

Par récurrence, on construit m_r (resp. M_r) le plus petit (resp. grand) coefficient de P^rY, la suite (m_r) est croissante, (M_r) décroissante.
On a

$$0 \leq M_r - m_r \leq (1-2d)^r(M_0 - m_0)$$

et la suite $(M_r - m_r)$ est décroissante. Pour montrer qu'elle tend vers 0, il suffit de montrer que pour une suite extraite, la suite tend vers 0.
Or par hypothèse, il existe $k > 0$, tel que tous les coefficients de la matrice P^k sont strictement positifs, et alors on applique ce que l'on vient de faire à P^k mais alors $0 < d \leq 1/2$ et la suite $(M_{kr} - m_{kr})$ tend vers 0.
Nous avons donc montré que la suite (P^rY) converge vers un vecteur de la forme $(u, u, \cdots, u)^T$ avec $u > 0$.
On applique ce résultat aux vecteurs colonnes de P, et on obtient que (P^n) converge vers une matrice de la forme Π et la somme des π_i vaut 1 car P^n est stochastique pour tout n.
On a $P\Pi = \Pi$ et comme Π et P commutent, on a aussi $\Pi P = \Pi$, ce qui implique que $\pi P = \pi$. □

REMARQUE 315 —

1. *Le théorème montre que pour une matrice stochastique régulière, la valeur propre 1 est une valeur propre simple, et toutes les autres valeurs propres sont strictement inférieures à 1 en module. C'est un résultat partiel d'un théorème plus général (théorème de Perron-Frobenius).*
2. *Cela montre que ${}^t\pi$ est l'unique vecteur propre de ${}^tP$ associé à la valeur propre 1 tel que la somme de ses coefficients vaut 1.*

3. Si μ est une distribution de probabilité (donc $\sum_i \mu_i = 1$), alors on vérifie que $\mu\Pi = \pi$.

On introduit alors la définition suivante.

Définition 316
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n\geq 0}$ une c.M. homogène sur E. Soit P sa matrice de transition.
On appelle ***distribution stationnaire*** *de la c.M. une distribution de probabilité $\pi = (\pi_1, \pi_2, \cdots, \pi_N)$ telle que $\pi P = \pi$.*

Remarque 317 — *Les distributions stationnaires pour une c.M. correspondent à des vecteurs propres de ${}^tP$, pour la valeur propre 1, qui sont à coefficients tous strictement positifs.*
On sait qu'une matrice stochastique P a toujours 1 comme valeur propre, donc ${}^tP$ aussi, mais on a en général aucune information sur les vecteurs propres de ${}^tP$.
Une matrice stochastique peut donc très bien n'avoir aucune distribution stationnaire, ou en avoir une infinité.
Nous avons montré que les c.M. régulières ont une distribution stationnaire, et la remarque précédente permet de montrer qu'elle est unique.

Théorème 318
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n\geq 0}$ une c.M. homogène et régulière sur E. Soit $\pi = (\pi_1, \pi_2, \cdots, \pi_N)$ la distribution stationnaire de la c.M.
Alors, pour tout $a_j \in E$, on a
$$\lim_{n\to\infty} \mathbb{P}\{X_n = j\} = \pi_j.$$

Preuve — La loi de X_n est donnée par le vecteur ligne μP^n. Comme la matrice ¶^{n} converge vers la matrice Π, on en déduit que
$$\lim_{n\to\infty} \mu P^n = \mu\Pi = \pi,$$
ce qui donne le résultat. □

Avec ce théorème, nous pouvons montrer qu'une c.M. irréductible admet elle aussi une unique distribution stationnaire.
Par contre, la matrice P^n peut ne pas converger.

Proposition 319
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n\geq 0}$ une c.M. homogène et irréductible sur E.
Alors, la c.M. possède une distribution stationnaire, qui est unique.

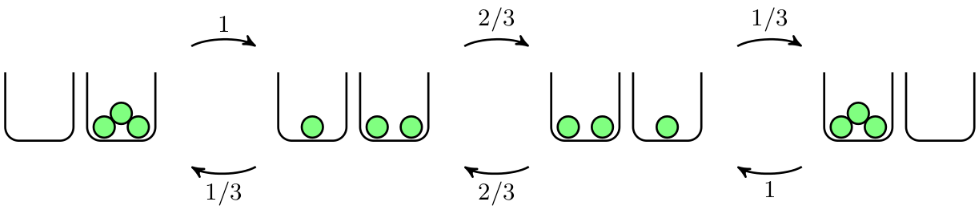
Preuve — Considérons la matrice stochastique $Q = \frac{1}{2}[P + I]$. Soit
$$m = \max_{i,j\in\mathcal{X}} \min\{n \geq 1 \mid p_{i,j}^{(n)} > 0\}.$$
Considérons la matrice
$$Q^m = \frac{1}{2^m}\left[I + \binom{m}{1}P + \binom{m}{2}P^2 + \cdots + \binom{m}{m-1}P^{m-1} + P^m\right].$$
Pour tout couple (i, j), il existe un terme de cette somme dont le coefficient (i, j) soit strictement positif. Comme tous les autres éléments de matrice sont non-négatifs, on conclut que $(Q^m)_{i,j} > 0$. Par conséquent, Q est la matrice de transition d'une chaîne régulière. Par le théorème précédent, il existe une unique distribution de probabilité π telle que $\pi Q = \pi$, ce qui implique $\frac{1}{2}[\pi + \pi P] = \pi$, donc $\pi P = \pi$. □

Exemple 320 (Modèle d'Ehrenfest) —
C'est un système motivé par la physique, qui a été utilisé pour modéliser de façon simple la répartition d'un gaz entre deux récipients.
N boules, numérotées de 1 à N , sont réparties sur deux urnes.
De manière répétée, on tire au hasard, de façon équiprobable, un numéro entre 1 et N , et on change d'urne la boule correspondante.

On voudrait savoir comment ce système se comporte asymptotiquement selon le temps (pour des durées aussi grandes que l'on veut) :

1. *Est-ce que la loi de probabilité du nombre de boules dans chaque urne approche une loi limite ?*
2. *Quelle est cette loi ?*
3. *Avec quelle fréquence toutes les boules se trouvent-elles toutes dans la même urne ?*

Ce système se modélise par une chaîne de Markov, sur 'ensemble $E = \{0, 1, \cdots, N\}$, où le numéro de l'état correspond au nombre de boules dans l'urne de gauche, par exemple :



Cette c.M. est homogène et irréductible.

Exemple 321 — *On vérifie facilement par calcul direct que la distribution stationnaire du modèle d'Ehrenfest est binomiale de paramètre $1/2$: $\mu_i = 2^{-N}\binom{N}{i}$.*
Quelle est l'interprétation de la distribution stationnaire ?
D'une part, nous savons déjà que si X_n suit la loi π à un temps n, alors X_m suivra la même loi π à tous les temps ultérieurs $m > n$.
En revanche, le Théorème précédent pour les c.M. régulières n'est pas vrai en général pour les c.M. irréductibles. Il suffit de considérer l'exemple du modèle d'Ehrenfest.
Mais, pour d'autres notions de convergence (en moyenne de Césaro), on peut encore obtenir des théorèmes de convergence vers la distribution stationnaire.

Théorème 322
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n \geq 0}$ une c.M. homogène et irréductible sur E. Soit π la distribution stationnaire de la c.M. Soit $a_j \in E$.
Alors, on a

$$\lim \frac{1}{n}\mathbb{E}_\mu\left(\sum_{m=0}^{n-1} \mathbb{1}_{\{X_m=j\}}\right) = \pi_j.$$

La fréquence moyenne de passage en tout état a_j converge vers π_j.

Preuve — Soit Π la matrice dont toutes les lignes sont égales à π. Alors on a $\Pi P = \Pi$ et $P\Pi = \Pi$. Il suit que

$$(I + P + \cdots + P^{n-1})(I - P + \Pi) = I - P^n + n\Pi.(*)$$

On commence par montrer que $I - P + \Pi$ est inversible : soit x un vecteur colonne tel que

$$(I - P + \Pi)x = 0.$$

Alors on a

$$0 = \pi(I - P + \Pi)x = \pi(I - P)x + \pi\Pi x = \pi x = 0$$

puisque $\pi(I - P) = \pi - \pi P = \pi - \pi = 0$ et $\pi\Pi = \pi$.
Il vient $\Pi x = 0$, et donc $(I - P)x = 0$. Comme P admet 1 comme valeur propre simple, avec vecteur propre à droite $\mathbf{1}$, ceci implique que $x = \lambda\mathbf{1}$, ce qui n'est possible que si $x = 0$ puisque $\pi x = 0$ et tous les π_i sont positifs. Le noyau de $I - P + \Pi$ est donc réduit à $\{0\}$, d'où $I - P + \Pi$ inversible.
Soit $Z = (I - P + \Pi)^{-1}$. Comme $\pi(I - P + \Pi) = \pi$, on a aussi $\pi = \pi Z$ et $\Pi = \Pi Z$. En multipliant $(*)$ à droite par Z, il vient

$$I + P + \cdots + P^{n-1} = (I - P^n)Z + n\Pi Z = (I - P^n)Z + n\Pi.$$

Or nous avons, pour tout état initial i,

$$\frac{1}{n}\mathbb{E}_\mu\left(\sum_{m=0}^{n-1} \mathbb{1}_{\{Xm=j\}}\right) = \frac{1}{n}\sum_{m=0}^{n-1} P_{i,j}^{(m)} = \frac{1}{n}\left((I - P^n)Z + n\Pi\right)_{i,j}$$

Comme les éléments de matrice de P^n sont uniformément bornés par 1, cette quantité converge vers $(\Pi)_{i,j} = \pi_j$ lorsque $n \to \infty$.
On a montré que

$$\Pi = \lim_{n\to+\infty} \frac{1}{n}(I_N + P + \cdots + P^{n-1}).$$

Mais pour calculer π, il est beaucoup plus simple de chercher la distribution stationnaire de $\frac{1}{2}(I_N + P)$.
Pour une distribution initiale quelconque μ, on obtient de la même manière la convergence vers $(\mu\Pi)_j = \pi_j$ puisque Π est une matrice de projection. □

Remarque 323 — *Le point de vue matriciel permet de mieux comprendre ce résultat : Soit la matrice $R = P - \Pi$, décrivant l'écart entre la matrice stochastique et le projecteur sur la distribution stationnaire. Il suit des égalités $\Pi P = P\Pi = \Pi^2 = \Pi$ que $\Pi R = R\Pi = 0$, et on en déduit*

$$P^n = \Pi + R^n$$

Dans le cas où P est régulière, R^n tend vers 0 lorsque $n \to \infty$.
Si P est irréductible mais pas régulière, la fonction $n \mapsto R^n$ est oscillante. La preuve montre cependant que R n'admet pas la valeur propre 1, et que la moyenne des R^n tend vers zéro, donc la moyenne des P^n tend vers Π.

Pour conclure, la distribution stationnaire π a également un lien intéressant avec l'espérance du temps de premier retour en un site j, appelé temps de récurrence moyen en j :

Théorème 324
Soit $N > 0$. Soient $E = \{a_1, \ldots, a_N\}$ un ensemble fini, et $(X_n)_{n\geq 0}$ une c.M. homogène et irréductible sur E. Soit π la distribution stationnaire de la c.M. Soit $a_j \in E$. On prend $X_0 = a_j$.
Alors, le temps de retour moyen à l'état a_j vaut

$$\mathbb{E}(\tau_j) = \frac{1}{\pi_j}.$$

Exemple 325 — *Dans le cas du modèle d'Ehrenfest avec N boules, le temps de récurrence moyen vers l'état à j boules est donné par*

$$\mathbb{E}(\tau_j) = \frac{1}{\pi_j} = 2^N \frac{i!(N-i)!}{N!}$$

En particulier, le temps moyen entre configurations où toutes les boules sont dans l'urne de gauche est de 2^N. Ce temps devient gigantesque pour des nombres de boules de l'ordre du nombre d'Avogadro, c'est-à-dire du nombre de molécules dans un échantillon d'une mole de gaz. Ce modèle simple peut donc justifier pourquoi, lorsque deux récipients contenant des gaz sont mis en contact, on n'observe jamais toutes les molécules dans le même récipient.