Chapitre 17
Probabilités

Table des matières

Vidal AGNIEL

Introduction

Vous trouverez au fil de ce chapitre les notations mathématiques :

Ω	Un ensemble
ω	Les éléments de Ω
$\mathcal{P}(\Omega)$	L'ensemble des parties de Ω
A	Un élément de $\mathcal{P}(\Omega)$. A est aussi appelé un événement.
$\mathbb{P}$	Une mesure de probabilité sur (Ω)
$\mathbb{P}(A)$	La probabilité d'un événement A
X	Une variable aléatoire sur $(\Omega, \mathbb{P})$
$\mathbb{P}(X \in A)$	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 X On a $\mathbb{E}(X) = \sum_{x\in\mathbb{R}} x\mathbb{P}(X = x) = \sum_{x\in X(\Omega)} x\mathbb{P}(X^{-1}(x))$.
$Var(X)$	La variance d'une v.a. réelle X.
σ_X	L'écart-type d'une v.a. réelle X. On a $\sigma_X^2 = Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

1 Dénombrement - Rappels

1.1 L'ensemble des parties de Ω, $\mathcal{P}(\Omega)$

On revoit rapidement dans cette section les éléments de théorie des ensembles et de dénombrement que nous utiliserons par la suite.
Ici, les ensembles E, Ω sont finis.
Un **sous-ensemble** ou une **partie** de Ω est un ensemble dont tous les éléments sont dans Ω.

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 $\omega = \{\alpha \in \Omega \mid \alpha = \omega\}$.*

EXEMPLE 2 —

1. *Si $\Omega = \{1, 2, 3\}$, alors*

$$\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

2. *Si $\Omega = \emptyset$, alors $\mathcal{P}(\emptyset) = \{\emptyset\}$ et donc a un élément. Et $\mathcal{P}(\{\emptyset\}) = \{\emptyset\{\emptyset\}\}$.*
   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 !

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

REMARQUE 4 — *On peut décomposer un ensemble Ω avec des singletons :*

$$\Omega = \bigcup_{\omega\in\Omega} \{w\}.$$

DÉFINITION 5
Soient $\Omega_1, \ldots, \Omega_n$ une famille d'ensembles finis.
Le ***produit cartésien*** *de la famille $(\Omega_i)_{1\leq i\leq n}$ est l'ensemble $\prod_{i=1}^{n} \Omega_i$ des familles $(\omega_1, \ldots, \omega_n)$, avec $\omega_i \in \Omega_i$.*

Pour Ω un ensemble fini, nous allons définir des fonctions particulières sur $\mathcal{P}(\Omega)$, les mesures de probabilité. Avec cela nous pourrons définir et étudier les variables aléatoires (qui sont des fonctions sur Ω).
Avant de définir ces objets, commençons par un exemple fondamental : les fonctions indicatrices. Les fonctions indicatrices permettent de définir des opérations ensemblistes (union, intersection,...) en termes algébriques (produit, somme...)

DÉFINITION 6 (**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 7
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 8 — *Écrire les fonctions indicatrices de $A \setminus B$ et de $A\Delta B$ en fonction de celles de A et B.*

1.2 Cardinal d'un ensemble

Définition

PROPOSITION 9
Deux ensembles finis Ω et Ω' ont même cardinal si et seulement s'il existe une bijection de Ω dans Ω'.
On a $Card(\Omega) = n$ si et seulement si Ω est en bijection avec $\{1, \ldots, n\}$.
Par convention, $Card\,(\emptyset) = 0$.
Un ensemble de cardinal infini est un ensemble qui n'est pas de cardinal fini. On écrira $Card(\Omega) = \infty$.

PROPOSITION 10
Soit Ω un ensemble fini de cardinal n.
Si $F \subset \Omega$, alors F est fini et $card\,(F) \leq \text{Card}\,(\Omega)$.
De plus, $\text{Card}\,(\Omega) = \text{Card}\,(F)$ si et seulement si $F = \Omega$.

REMARQUE 11 — *Pour Ω un ensemble fini, en posant $n = Card(\Omega)$, on pourra écrire $\Omega = \{\omega_1, \ldots, \omega_n\}$ (on numérote les éléments de Ω).*

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

PROPOSITION 12
Soit Ω un ensemble et A une partie finie de Ω. Alors :

$$\text{Card}\,A = \sum_{x\in\Omega} \mathbb{1}_A(x)$$

Proposition 13
Si A et B sont deux sous-ensembles finis de Ω, alors $A \cup B$ est fini et

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

Proposition 14
Soient Ω et Ω' deux ensembles finis. Alors $\Omega \times \Omega'$ est fini et

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

Proposition 15
Soit Ω un ensemble fini de cardinal n.
Alors $\mathcal{P}(\Omega)$ est fini, de cardinal 2^n.

1.3 Coefficients binomiaux

Définition 16
Soient $n \geq 0$, $p \in \mathbb{Z}$.
On note $\binom{n}{p}$ (ou C_n^p) le nombre de parties à p éléments d'un ensemble à n éléments.
*Ce nombre est appelé **coefficient binomial** (ou p **parmi** n).*
On convient que si $p < 0$ ou si $p > n$, alors $\binom{n}{p} = 0$.

Proposition 17 (**Propriété du triangle de Pascal**)
Soient $n, p \in \mathbb{N}$. On a

$$\binom{n+1}{p} = \binom{n}{p-1} + \binom{n}{p}.$$

Proposition 18
Soient $n, p \in \mathbb{N}$. Alors on a

$$\binom{n}{p} = \binom{n}{n-p}.$$

Proposition 19
Soient $n, p \in \mathbb{N}$ avec $0 \leq p \leq n$. Alors on a

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

Remarque 20 — *On a ainsi $\binom{n}{0} = \binom{n}{n} = 1$, $\binom{n}{1} = \binom{n}{n-1} = 1$, $\binom{n}{2} = \frac{n(n-1)}{2}$.*

Nombres d'arrangements

Définition 21 (**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.*

Un arrangement à p éléments nécessite p éléments parmi n, et l'ordre de ces éléments compte. (par ex $(2, 3)$ et $(3, 2)$ sont deux arrangements differents à 2 él. de $\{1, 2, 3\}$)

Proposition 22
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* $A_p^n = 0 = p!\binom{n}{p}$.

1.4 Tirages

En probabilités, lorsque l'on tire des éléments (des cartes, des boules dans une urne,...) on rencontre 4 méthodes principales pour faire le tirage.
Le tirage avec ou sans remise, et avec ou sans ordre.
Dans un tirage avec remise, on peut tirer plusieurs fois le même élément.
Dans un tirage sans remise, tous les éléments tirés doivent être différents.
Dans un tirage avec ordre, l'ordre des éléments tirés compte (on regardera une liste).
Dans un tirage sans ordre, l'ordre des éléments tirés ne compte pas (on regardera un ensemble).
Prenons $E = \{1, \ldots, n\}$ pour ensemble de départ, et tirons k éléments ($0 \leq k$) parmi les n éléments de E.

— **Tirage sans remise, sans ordre**
  L'élément tiré est $\{a_1, \ldots, a_k\}$, avec $a_1, \ldots, a_k \in E$ t.q. $\forall 1 \leq i < j \leq k$, on a $a_i \neq a_j$.
  Il y a $\binom{n}{k}$ tirages possibles.
  Exemple : Tirage au loto (p boules avec n valeurs possibles).
— **Tirage sans remise, avec ordre**
  L'élément tiré est $(a_1, \ldots, a_k)$, avec $a_1, \ldots, a_k \in E$ t.q. $\forall 1 \leq i < j \leq k$, $a_i \neq a_j$.
  Il y a $A_k^n = k!\binom{n}{k}$ tirages possibles.
  Exemple : Tiercé (p premiers chevaux à la fin de la course, n chevaux possibles)
— **Tirage avec remise, avec ordre**
  L'élément tiré est $(a_1, \ldots, a_k,\ a_i)$, avec $a_i \in E$.
  Il y a n^k tirages possibles.
  Exemple : Combinaison de cadenas (p numéros à choisir, chacun avec n valeurs possibles)
— **Tirage avec remise, sans ordre**
  L'élément tiré est $\{\{a_1, \ldots, a_k\}$, avec $a_i \in E$ (un ensemble où on autorise les répétitions).
  Il y a $\binom{n+k-1}{n-1}$ tirages possibles. (plus difficile)
  Exemple : Coupe de boules de glace (p boules de glace, avec n parfums possibles)

2 Probabilités - Le langage des probabilités

2.1 Expériences aléatoires, événements

DÉFINITION 23
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 24
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 25 —

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 26 — *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 des probabilités.*
On rappelle que nous n'étudierons que les cas d'ensembles finis (ce qui permet déjà de modéliser beaucoup de choses).

REMARQUE 27 — *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. Mais pas avec la température de la pièce par exemple.

DÉFINITION 28
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 29 —

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" :*

$$A = \{\omega \in \Omega, \sum_{i=1}^{n} \omega_i \geq \frac{n}{2}\}.$$

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).

2.2 Opérations ensemblistes sur les 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).

Dans le cas des ensembles finis, l'ensemble de tous les événements possibles dans l'expérience aléatoire est exactement $\mathcal{P}(\Omega)$, l'ensemble de toutes les parties de Ω.

Probabilités : approche intuitive

Remarque 30 — *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.

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.

3 Mesures de probabilité, Probabilité

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 deux objets fondamentaux : les **mesures de probabilités**, et les **variables aléatoires**.
Dans le cas des ensembles infinis, il faut aussi ajouter une troisième notion (les σ-algèbres), mais cela est totalement hors du programme.
Dans cette section, nous allons définir les mesures de probabilités. Les variables aléatoires seront définies par la suite.
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, et avec les propriétés des mesures de probabilités nous utiliserons les techniques de dénombrement pour calculer des probabilités.

3.1 Définitions

Définition 31
Soient Ω un ensemble et $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$ une fonction.
*On dit que $\mathbb{P}$ est une **(mesure de) probabilité** sur Ω si elle vérifie les propriétés suivantes :*

1. $$\mathbb{P}(\Omega) = 1. \quad (\textit{mesure totale de l'ensemble})(*)$$
2. *Pour toute famille $(A_1, \ldots, A_n)$ de parties de Ω deux-à-deux disjointes, on a*
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i). \quad (\textit{sigma-additivité})(**)$$

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

Remarque 32 — *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.
Dans le programme, il faut dire que $\mathbb{P}$ est une probabilité et que $\mathbb{P}(A)$ est une probabilité.

Exemple 33 — *Soient Ω un ensemble, et $\omega_0 \in \Omega$.*
On définit la fonction $\delta_{\omega_0} : A \in \mathcal{P}(\Omega) \mapsto \begin{cases} 1 \text{ si } \omega_0 \in A \\ 0 \text{ sinon} \end{cases}$.
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.

Mesure de probabilité uniforme

Proposition-Définition 34
Soit Ω un ensemble fini.
*On appelle **mesure de probabilité uniforme** sur Ω la mesure de probabilité telle que*
$$\forall \omega \in \Omega, \, \mathbb{P}(\{\omega\}) = \frac{1}{\text{Card}\,(\Omega)}.$$

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

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

Démonstration — On vérifie la définition de mesure de probabilité.

Remarque 35 —

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 de la première section).*
   La difficulté est de bien décrire et dénombrer l'ensemble total Ω et la partie A qui nous intéresse.

Il existe beaucoup d'autres mesures de probabilités sur un ensemble fini. Donnons d'abord les propriétés de ces fonctions.

3.2 Propriétés d'une mesure de probabilité

La définition d'une mesure de probabilité est courte, mais la propriété d'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).

Proposition 36 (**Propriétés élémentaires**)
Soient Ω un ensemble, et $\mathcal{P}(\Omega)$ une tribu sur Ω.
Soit $\mathbb{P}$ une mesure de probabilité. Soient $A, B \in \mathcal{P}(\Omega)$. 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. $\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cup B)$

Démonstration — Sur feuille. On utilise les propriétés d'une mesure de probabilité.

Si les parties $A_1, \ldots, A_n$ de Ω ne sont pas deux-à-deux disjointes, on ne peut pas appliquer la propriété $(**)$.
Par contre, nous avons la majoration suivante, très utile dans la pratique.

Proposition 37 (**Probabilité d'une réunion**)
Soient Ω un ensemble, et $\mathbb{P}$ une mesure de probabilité sur Ω.
Soit $(A_i)_{i \in I}$ une famille finie d'élements de $\mathcal{P}(\Omega)$ (une famille d'événements). On a alors :

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

Démonstration —Sur feuille.

Et voilà, nous avons énoncé ici toutes les propriétés les plus fondamentales d'une mesure de probabilité.
On regarde en général le couple suivant :

DÉFINITION 38
Soient Ω un ensemble et $\mathbb{P}$ une mesure de probabilité sur (Ω).
*Le couple $(\Omega, \mathbb{P})$ 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.

REMARQUE 39 — *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.

Nous allons maintenant décrire toutes les mesures de probabilité possibles sur un ensemble Ω fini.

3.3 Probabilités sur un ensemble fini - Calcul combinatoire

La définition d'une mesure de probabilité ne dépend pas de la forme des éléments de Ω, juste de leur nombre. Donc, pour Ω de cardinal n, on pourra se ramener à considérer des mesures de probabilités sur $\{1, \ldots, n\}$.

PROPOSITION 40 (**Caractérisation des mesures de probabilités**)
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 liste
$$(\mathbb{P}(\{\omega_1\}), \ldots, \mathbb{P}(\{\omega_n\})).$$

(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 pour tout $\omega_i \in \Omega$, on ait $p_i = P(\{\omega_i\})$.
On a : $\mathbb{P}(A) = \sum_{i=1}^n p_i \cdot \chi_A(\omega_i) = \sum_{i=1}^n p_i \delta_{\omega_i}(A)$.

Démonstration —Sur feuille.

REMARQUE 41 —
- *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 42 — *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 $\dfrac{1}{2}$.*
On retrouve alors une mesure de probabilité uniforme.

DÉFINITION 43
Soit $\Omega = \{\omega_1, \dots, \omega_n\}$ un ensemble fini, et $\mathbb{P}$ une mesure de probabilité sur Ω.
Le n-uplet $(\mathbb{P}(\{\omega_1\}), \dots, \mathbb{P}(\{\omega_n\})),\ \omega_i \in \Omega\}$ est appelé ***distribution (ou loi) de probabilité*** *de $\mathbb{P}$.*

On parle aussi de **distribution de probabilité** de $\mathbb{P}$.
Nous avons vu que la loi de probabilité de $\mathbb{P}$ caractérise la fonction $\mathbb{P}$.
Cela est une façon bien plus pratique de définir et de manipuler une mesure de probabilités.

Différentes méthodes de tirages

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 uniforme 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

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

Vous connaissez trois grandes façons de tirer les boules au hasard : tirage avec remise, tirage sans remise, tirage simultané. Pour chaque tirage, l'ensemble Ω des résultats est différent.
En fonction de la situation, il faudra choisir le tirage qui correspond, sinon les calculs de probabilités ne seront pas bons.

Tirage simultané

Nous tirons 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 $\binom{N}{n}$.
La probabilité recherchée est (après calculs) :

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

Dans le cas de deux couleurs, on a :

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

Exemple 44 — *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 uniforme 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 sans remise

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

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

Cependant, les événements que l'on regarde ne tiennent pas compte de l'ordre (on veut uniquement un certain nombre de boules de chaque couleur). Le calcul des probabilités donne :

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

On obtient ainsi la même mesure de probabilité que celle du cas de tirage simultané.
Proposition : Le tirage sans remise et le tirage simultané sont deux façons de "tirer au sort" qui sont équivalentes, si l'on ne s'intéresse pas à l'ordre des éléments tirés.

Tirage avec remise

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 $Card(\Omega) = N^n$.
On obtient après calculs une probabilité de :

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

Dans le cas particulier où $k = 2$, on pose $p = \dfrac{N_1}{N} = p_1$. La probabilité vaut alors :

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

Exemples

Exemple 45 — *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 46 — *On tire au hasard uniforme 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 uniforme" 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 $\frac{\text{nb de cas favorables}}{\text{nb de cas possibles}} = \dfrac{\binom{4}{2}\binom{48}{2}}{\binom{52}{4}}$.

Exemple 47 — *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.
Après un peu de comptage, on obtient que l'ensemble des (a_1, a_2, a_3) tels que $a_1 + a_2 + a_3 > 10$ contient $\frac{6^3}{2}$ éléments.
Donc, l'événement A est de probabilité $\frac{1}{2}$, ce qui implique que $\overline{A}$ est aussi de probabilité $\frac{1}{2}$.

Remarque 48 — *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 49 — *(Le problème du chevalier de Méré) Ce personnage de la cour de Louis XIV 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 permet de fournir à long terme des gains assurés.
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 fait avoir par un faux 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.

3.4 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 50 — *Lançons de deux dés équilibrés. Cherchons la probabilité de l'événement "la somme est supérieure ou égale à 10". Elle vaut :*
- $\frac{1}{6}$ *sans information supplémentaire.*
- $\frac{1}{2}$ *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 51 (**Probabilité conditionnelle**)
Soit $(\Omega, \mathbb{P})$ *un espace probabilisé.*
Soit $A, B \in \mathcal{P}(\Omega)$ *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 52
Soit $(\Omega, \mathbb{P})$ *un espace probabilisé.*
Soit $B \in \mathcal{P}(\Omega)$ *une partie telle que* $\mathbb{P}(B) > 0$. *Alors,*

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

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

Démonstration — On vérifie la définition de mesure de probabilité.

Proposition 53 (**Formule des probabilités composées**)
Soit $(\Omega, \mathbb{P})$ *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 :*

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

Démonstration — On utilise la définition d'une probabilité conditionnelle, et un produit télescopique.

Pour $A, B \in \mathcal{P}(\Omega)$ deux événements 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 54 (**Partition**)
Soient Ω *un ensemble,* $n \in \mathbb{N}^*$, *et* $A_1, \ldots, A_n$ *des parties de* Ω.
On dit que la famille $(A_1, \ldots, A_n)$ *est une* ***partition*** *de* Ω *si :*

1. *Les* A_i *sont deux à deux disjoints :* $\forall i, j \in \{1, \ldots, n\}$ *t.q.* $i \neq j$, $A_i \cap A_j = \emptyset$
2. *La réunion des* A_i *vaut* Ω *:* $\cup_{i=1}^n A_i = \Omega$.

Définition 55 (**Système complet d'événements**)
Soit $(\Omega, \mathbb{P})$ *un espace probabilisé.*
Soit $(B_1, \ldots, B_n)$ *une partition de* Ω, *telle que* $\mathbb{P}(B_i) > 0$ *pour chaque* i.
Une telle partition est appelée ***système complet d'événements***.

Proposition 56 (**Formule des probabilités totales**)
Soit $(\Omega, \mathbb{P})$ *un espace probabilisé, et soit* $(B_i)_{i\in I}$ *un système complet d'événements de* Ω.
Alors, pour tout $A \in \mathcal{P}(\Omega)$, *on a*

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

Démonstration — On utilise les propriétés de la mesure de probas $\mathbb{P}$.

Théorème 57 (**Formule de Bayes**)
Soit $(\Omega, \mathbb{P})$ *un espace probabilisé, et soit* $(B_i)_{i\in I}$ *un système complet d'événements de* Ω.
Soit $A \in \mathcal{P}(\Omega)$ *tel que* $\mathbb{P}(A) > 0$. *Alors, pour tout* $i \in I$, *on a*

$$\mathbb{P}(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j\in I} \mathbb{P}(A|B_j)P(B_j)}.$$

Démonstration — On utilise les résultats précédents.

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

Exemple 58 — *Un individu est tiré au hasard uniforme 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 59 — *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 uniforme un gérant de portefeuille, il y a une chance sur 10 que celui-ci soit un gérant bien informé.
Un client choisit au hasard uniforme 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

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

3.5 Evénements indépendants

La notion d'indépendance est un outil absolument fondamental en probabilités.
Intuitivement, deux événements A et B sont indépendants si le fait d'avoir une information sur A ne donne aucune information sur B, et réciproquement.
L'indépendance est modélisée mathématiquement par cette définition.

Définition 60 (**Evénements indépendants**)
Soient $(\Omega, \mathbb{P})$ un espace probabilisé, et $A, B \in \mathcal{P}(\Omega)$ 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 61 —

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 62 —

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* $A = \{\textit{faire Pile}\}$ *et* $B = \{\textit{faire Face}\}$ *dans un jeu de Pile ou Face équilibré)*
3. *On tire une carte au hasard uniforme 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 $\mathbb{P}(A) = \dfrac{4}{52}$*,* $\mathbb{P}(B) = \dfrac{13}{52}$*, 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 mesure 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
$$\mathbb{Q}(A \cap B) = \frac{1}{2} \neq \mathbb{Q}(A)\mathbb{Q}(B) = (\frac{1}{2} + \frac{3}{102}) \times (\frac{1}{2} + \frac{12}{102}).$$
Les événements A *et* B *ne sont pas indépendants pour la mesure de probas* $\mathbb{Q}$*.*

Proposition 63
Soient $(\Omega, \mathbb{P})$ *un espace probabilisé, et* $A, B \in \mathcal{P}(\Omega)$ *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 64 (**Indépendance d'une famille d'événements**)
Soient $(\Omega, \mathbb{P})$ *un espace probabilisé, et* $(A_1, \ldots, A_n)$ *une famille d'événements.*
On dit que cette famille est ***indépendante*** *si pour tous entiers* $i_1, \cdots, i_k$*, avec* $1 \leq i_1 < i_2 < \ldots < i_k \leq n$*, on a*

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

REMARQUE 65 — *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 $\mathbb{P}(A) = \dfrac{1}{2}$*,* $\mathbb{P}(B) = \dfrac{1}{2}$*,* $\mathbb{P}(C) = \dfrac{2}{3}$*.*
   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.*

4 Variables aléatoires

4.1 Définition

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

Définition 66
Soient $(\Omega, \mathbb{P})$ un espace probabilisé fini et E un ensemble.
Soit $X : \Omega \to E$ une fonction.
*On dit alors que X est une **variable aléatoire** de $(\Omega, \mathbb{P})$ dans E.*
*Si E est un sous-ensemble de $\mathbb{R}$, on dit alors que X est une **variable aléatoire réelle**.*

Remarque 67 — *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.**.*
Nous n'étudions ici que les variables aléatoires finies (les v.a. définies sur un ensemble Ω fini).

Faisons un exemple.

Exemple 68 — *Etudions un lancer de deux dés équilibrés.*
Dans ce cas, l'ensemble des états est $\Omega = \{(i, j) : 1 \leq i \leq 6; 1 \leq j \leq 6\}$.
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".

4.2 Loi d'une variable aléatoire

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

Proposition-Définition 69 (**Loi d'une variable aléatoire**)
Soient $(\Omega, \mathbb{P})$ un espace probabilisé fini et E un ensemble. Soit $X : \Omega \to E$ une variable aléatoire.
On définit la fonction $\mathbb{P}_X : \mathcal{P}(X(\Omega)) \to [0, 1]$ par

$$\mathbb{P}_X(B) \underset{def}{=} \mathbb{P}(\{\omega \in \Omega \ t.q. \ X(\omega) \in B\}).$$

Alors, la fonction $\mathbb{P}_X$ est une (mesure de) probabilité, sur $X(\Omega)$, l'image de Ω par X.
*Cette mesure de probabilité est appelée **loi de probabilité de** X.*
Dans le langage probabiliste, pour $B \subset X(\Omega)$ on note $(X \in B) = \{\omega \in \Omega \ t.q. \ X(\omega) \in B\}$, de sorte que $\mathbb{P}_X(B) =_{def} \mathbb{P}(X \in B)$.
Pour $X(\Omega) = \{x_1, \ldots, x_r\}$, la famille $(\mathbb{P}(X = x_1), \ldots, \mathbb{P}(X = x_r))$ est la distribution de X (et de $\mathbb{P}_X$).

Démonstration — On montre que cette fonction vérifie la définition d'une mesure de probas.

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 $\mathbb{P}$ donne énormément d'informations sur $\mathbb{P}$.

Exemple 70 — *Reprenons l'exemple précédent du lancer de deux dés équilibrés. Pour X la variable aléatoire "somme des deux faces obtenues", on a $X : \{1,\dots,6\}^2 \to \{2,\dots,12\}$.*
Ainsi la loi de probabilité de X, $\mathbb{P}_X$, est une mesure de probas sur l'ensemble $\{2,\dots,12\}$.
On a par exemple :

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

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 71 —

1. *Le nombre de 6 obtenus dans un lancer de 3 dés équilibrés.*
   Les espaces sont $\Omega = \{1,\dots,6\}^3$, $\mathcal{P}(\Omega) = \mathcal{P}(\Omega)$, $\mathbb{P}$ la mesure uniforme sur Ω, $F = \{0,1,2,3\}$.
   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 72 —
Pour Ω fini, toute fonction $X : \Omega \to F$ est une variable aléatoire.
De plus, $X(\Omega)$ est un ensemble fini, on peut numéroter ses éléments. On regarde ainsi une partie finie de F. (pour tout A disjoint de $X(\Omega)$ on a $\mathbb{P}(X \in A) = 0$).
Ainsi, $\mathbb{P}_X$ est une mesure de probas sur $X(\Omega)$.
Or, une mesure de probas est caractérisée par sa loi.
Donc, pour $X(\Omega) = \{x_1,\dots,x_m\}$, la loi de probas de X est caractérisée par la liste $(\mathbb{P}(X = x_1), \mathbb{P}(X = x_2),\dots,\mathbb{P}(X = x_r))$.
Si l'on veut calculer une probabilité de la forme $\mathbb{P}(X = B)$, on pourra faire le calcul à partir des $\mathbb{P}(X = x_i)$.

Exemple 73 — *(**Fonction indicatrice**) Soit $(\Omega,\mathbb{P})$ un espace probabilisé. Soit $A \in \mathcal{P}(\Omega)$.*
*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 $(\Omega,\mathbb{P})$.
Ces variables aléatoires sont les v.a. les plus simples que l'on puisse utiliser (avec les v.a. constantes).
Elles sont extrêmement utiles dans les calculs. (pour des sommes, produits, découpages en partition)
On a par exemple que $\mathbb{P}_{\mathbb{1}_A}(\{1\}) = \mathbb{P}(\mathbb{1}_A^{-1}(\{1\})) = \mathbb{P}(A)$.

Proposition 74
Soient $(\Omega,\mathbb{P})$ un espace probabilisé et $X : \Omega \to F$ une variable aléatoire. Pour tout $y \in F$, on a :

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

On peut calculer toutes les probabilités de la forme $\mathbb{P}(X = B)$ en utilisant la probabilité de tous les singletons $\{\omega\}$.

Exemple 75 — *Une variable aléatoire $X : \Omega \to \{1, \dots, n\}$ de loi uniforme a pour loi de probas la famille* $(\frac{1}{n})_{1\leq k\leq n}$.

5 Espérance et variance d'une variable aléatoire réelle

Dans la majorité des exemples que nous avons vus, les variables aléatoires étaient réelles. Etudions-les.

Espérance d'une v.a. - Approche intuitive

Motivation : Considérons X une variable aléatoire réelle, définie sur un ensemble Ω fini.
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

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

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 76 (**Espérance**)
Soient $(\Omega, \mathbb{P})$ un espace probabilisé et $X : \Omega \to F$ une variable aléatoire réelle.
*On appelle **espérance de la v.a.** X le nombre* $\mathbb{E}(X) = \sum_{y\in X(\Omega)} y\mathbb{P}(X = y)$.

Remarque 77 —

1. *On peut remarquer que le nombre réel $\mathbb{E}(X)$ ne dépend que de la loi de X (de la famille $(\mathbb{P}_X(\{y\}))_{y\in X(\Omega)}$.*
2. *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)*

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 $f : \mathbb{R} \to \mathbb{R}$).
Toutes ces opérations définissent encore des v.a. réelles.
On peut alors étudier ce qui se passe par rapport à l'espérance.

5.1 Propriétés de l'espérance

Proposition 78

Soient $(\Omega, \mathbb{P})$ *un espace probabilisé et* $X, Y : \Omega \to \mathbb{R}$ *deux v.a. réelles. Alors :*

1. *L'espérance* $\mathbb{E} : L^1(\Omega, \mathbb{P}) \to \mathbb{R}$ *est une application linéaire.*
   On a : $\forall a, b \in \mathbb{R}$, $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.
2. *L'espérance est positive : Si pour tous* $X \geq 0$ *(i.e.* $\forall \omega \in \Omega$, $X(\omega) \geq 0$*), alors on a* $\mathbb{E}(X) \geq 0$.
3. *Si* $X \leq Y$*, alors on a* $\mathbb{E}(X) \leq \mathbb{E}(Y)$.
4. *Si* X *est une v.a. constante (*$X = a$ *pour un* $a \in \mathbb{R}$*), alors* $\mathbb{E}(X) = a$.

Démonstration —Les points 2 et 4 découlent de la définition de l'espérance. Les points 1 et 3 nécessitent le lemme de transfert pour être démontrés.

Exemple 79 — *Soient* $(\Omega, \mathbb{P})$ *un espace probabilisé et* $A \in \mathcal{P}(\Omega)$. *Pour* $\mathbb{1}_A$ *la fonction indicatrice de* A*, 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 finie car la v.a. X prend un nombre fini de valeurs.
Mais, on peut décomposer chaque terme $\mathbb{P}(X = y)$ en une somme sur des parties de Ω, et exprimer l'espérance comme une somme sur chaque $\omega \in \Omega$.

5.2 Lemme de transfert, Théorème de transfert

Théorème 80 (**Lemme de transfert**)

Soit $(\Omega, \mathbb{P})$ *un espace probabilisé fini. Soit* $X : \Omega \to F$ *une v.a. réelle.*
On a alors la formule fondamentale suivante :

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

Démonstration — On utilise le fait que $\{X = y\} = \{\omega \in \Omega, \text{ t.q. } X(\omega) = y\}$, ainsi qu'une somme double.

Exemple 81 — *Dans le cas du lancer de 3 dés à 6 faces équilibrés, on a* $\Omega = \{1, \dots, 6\}^3$ *et* $\mathbb{P}$ *est la mesure uniforme sur* Ω. *On regarde la somme des 3 faces obtenues. On pose donc* $X : \Omega \to \{3, \dots, 18\}$ *la v.a. définie par* $X(a_1, a_2, a_3) = a_1 + a_2 + a_3$.
D'après le lemme de transfert, on a ainsi :

$$\begin{array}{rcl} \mathbb{E}(X) & = & \sum_{(a_1,a_2,a_3)\in\Omega}(a_1 + a_2 + a_3)\mathbb{P}(\{(a_1, a_2, a_3)\}) = \sum_{i,j,k=1}^{6}(i + j + k)\frac{1}{6^3} \\ & = & \frac{1}{6^3}(\sum_{i=1}^{6}\sum_{j=1}^{6}\sum_{k=1}^{6}(i + j + k) = \frac{1}{6^3}(6^2\sum_{i=1}^{6} i + 6^2\sum_{j=1}^{6} j + 6^2\sum_{k=1}^{6} k) \\ & = & \frac{6^2.3}{6^3}\frac{6.7}{2} = \frac{21}{2}. \end{array}$$

Exemple 82 — *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.*
On choisit une séquence de questions à poser. Puis, pour chaque valeur de m *on définit* $N(m)$ *le nombre de questions nécessaires pour deviner* m. *Nous regardons la fonction* N *comme une v.a.*
Calculons l'espérance de N *pour les séquences de questions suivantes :*

1. *Premier cas : La question numéro i est "Est-ce que $m = i$?".*
   Avec ce choix de questions, on obtient :
$$\mathbb{P}(N = k) = \mathbb{P}(\textit{le nombre } k \textit{ a été choisi}) = \frac{1}{10}.$$
   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.

Soit maintenant $f : F \to \mathbb{R}$ une fonction. La fonction $f \circ X$ est alors une autre v.a. réelle. Par définition, on a $\mathbb{E}(f(X)) = \sum_{\omega \in \Omega} f(X(\omega))P(\{\omega\})$.
Pour $f \circ X(\Omega) = \{y_1, \ldots, y_x\}$, la formule de transfert donne $\mathbb{E}(f(X)) = \sum_{i=1}^{s} y_i P(f \circ X = y_i)$.
Ce résultat manque de commodité en général car il implique de déterminer l'image de $f \circ X$ et de calculer la probabilité associée à chaque élément de l'image.
On peut cependant exprimer $\mathbb{E}(f \circ X)$ d'une autre façon, plus adaptée aux informations que l'on connaît en général sur X.

Théorème 83 (**Théorème de transfert**)
Soit $(\Omega, \mathbb{P})$ un espace probabilisé fini . Soient $X : \Omega \to F$ une v.a. réelle, et $f : F \to \mathbb{R}$.
On a alors :
$$\mathbb{E}(f(X)) = \sum_{x \in X(\omega)} f(x)\mathbb{P}(X = x).$$

On peut ainsi calculer facilement $\mathbb{E}(f(X))$, car le calcul de la somme est basé sur les quantités $\mathbb{P}(X = x)$ (sur la loi de la loi de probas de X).

5.3 Variance et écart-type

Définition 84 (**Variance**)
Soit $(\Omega, \mathbb{P})$ un espace probabilisé fini. Soit $X : \Omega \to F$ une v.a. réelle.
On définit la ***variance*** *de X par :*
$$\mathrm{Var}\,(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \sum_{x_i \in F} (x_i - \mathbb{E}(X))^2\mathbb{P}(X = x_i).$$

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

Proposition 85
Soit $(\Omega, \mathbb{P})$ un espace probabilisé fini. Soit $X : \Omega \to F$ une v.a. réelle.

1. *En développant le carré $(X - \mathbb{E}(X))^2$ on obtient :*
$$\sigma_X^2 = Var(X) = \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$
Ainsi, on a $\mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0$.
2. *Pour $a, b \in \mathbb{R}$, on a :*
$$Var(aX + b) = a^2 Var(X).$$

Remarque 86 — *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 $d(x, y) = \sqrt{\langle x - y, x - y\rangle}$ dans le plan ou dans l'espace) Elle mesure, dans un sens, à quel point la v.a. X s'écarte en moyenne de $\mathbb{E}(X)$.*

Exemple 87 — *(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 $g(n)$ 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. 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).

6 Variables aléatoires usuelles

6.1 Variable aléatoire uniforme

La loi de variable aléatoire que l'on rencontre le plus souvent est la loi uniforme. C'est celle qui correspond au tirage uniforme, c'est-à-dire à un tirage parmi m éléments, où chaque élément a une probabilité $\frac{1}{m}$ d'être tirée.

Définition 88
Soit $(\Omega, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{R}$ une v.a. Soit $n \geq 1$.
*Si on a $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \ldots = \mathbb{P}(X = n) = \frac{1}{n+1}$, on dit que X est une **variable aléatoire uniforme** sur $\{0, \ldots, n\}$.*

On fera attention, l'ensemble $\{0, \ldots, n\}$ a $n+1$ éléments.

PROPOSITION 89
Soient $n \geq 1$ et X une v.a. uniforme sur $\{0, \ldots, n\}$.
Alors, on a

$$\begin{aligned} \mathbb{E}(X) &= \frac{n}{2} \\ \operatorname{Var}(X) &= \frac{n(n+2)}{12}. \end{aligned}$$

REMARQUE 90 — *Pour $x_1, \ldots, x_n$ des réels distincts, on peut définir les v.a. de loi uniforme sur $\{x_1, \ldots, x_n\}$ avec $\mathbb{P}(X = x_1) = \ldots = \mathbb{P}(X = x_n) = \frac{1}{n}$.*
Il n'y a pas besoin de se limiter aux nombres entiers pour les v.a. uniformes.
On trouvera alors que $\mathbb{E}(X) = \frac{1}{n} \sum_{k=1}^{n} x_k$. L'espérance de X est alors la moyenne des nombres $x_1, \ldots, x_n$.
Par exemple, lancer un dé à 6 faces équilibré revient à utiliser une v.a. de loi uniforme sur $\{1, \ldots, 6\}$.

Nous présentons deux autres v.a. réelles usuelles. Ces v.a. sont à valeurs dans $\mathbb{N}$.
On l'a dit, la v.a. X est totalement déterminée par sa mesure de probabilité associée $\mathbb{P}_X$.
Pour décrire une v.a., il n'est pas nécessaire de vraiment décrire l'espace probabilisé $(\Omega, \mathbb{P})$.

6.2 Variable aléatoire de Bernouilli

DÉFINITION 91
Soit $(\Omega, \mathbb{P})$ 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 92
Soient $p \in [0, 1]$ et X une v.a. de Bernouilli de paramètre p.
Alors, on a

$$\begin{aligned} \mathbb{E}(X) &= p \\ \operatorname{Var}(X) &= p(1-p). \end{aligned}$$

Démonstration — On calcule espérance et variance.

REMARQUE 93 — *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.*

6.3 Variable aléatoire binomiale

DÉFINITION 94
Soit $(\Omega, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{N}$ une v.a.
Soient $p \in [0, 1]$ et $n \geq 1$.
*Si on a $X(\Omega) = \{0, \ldots, n\}$ avec, pour tout $0 \leq k \leq n$, $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, 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 95 —

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 96
Soient $n \geq 1$, $p \in [0,1]$, et X une variable binomiale de loi $B(n,p)$.
Alors, on a

$$\begin{aligned} \mathbb{E}(X) &= np \\ \operatorname{Var}(X) &= np(1-p). \end{aligned}$$

Démonstration —(Plus tard).

EXEMPLE 97 — *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).

EXEMPLE 98 — *Dans une tombola, chaque participant a 1 chance sur 1000 de gagner le gros lot. Si 1000 candidats jouent, de façon indépendante les uns des autres, quelle est la probabilité qu'au moins l'un d'eux gagne le gros lot ?*
On peut modéliser cette situation par une v.a. X de loi binomiale, de paramètres $n = 1000$ et $p = \frac{1}{1000}$. La v.a. X compte le nombre de personnes qui auront gagné le gros lot.
La probabilité qui nous intéresse est $\mathbb{P}(X \geq 1)$.
Comme l'image de X est $\{0, \ldots, 1000\}$, on a $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X=0) = 1 - \times(\frac{999}{1000})^{1000}$.
Or, on a que $(\frac{n-1}{n})^n = (1-\frac{1}{n})^n = \exp(n\ln(1-\frac{1}{n}))$. Pour n grand ($n \to +\infty$), un DL_1 nous fournit $(\frac{n-1}{n})^n = \exp(n.(-\frac{1}{n} + o(\frac{1}{n})) = \exp(-1+o(1)) = \exp(-1)\exp(o(1)) = \exp(-1) + o(1)$.
Avec $n = 1000$ ici, on a donc $\frac{999}{1000} \simeq \frac{1}{e} \simeq 0.36$.
Donc, $\mathbb{P}(X \geq 1) \simeq 1 - \frac{1}{e} \simeq 0.63$.
Pour une expérience avec une probabilité de succès de $p = \frac{1}{1000}$, on a environ 63% de chances d'obtenir au moins un succès si l'on répète cette expérience (de façon indépendante) 1000 fois. Ce résultat reste vrai pour d'autres valeurs de n et de p, tant que $n.p = 1$ et que n est grand.

7 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.

Définition 99 (**V.a. indépendantes**)
Soit $(\Omega, \mathbb{P})$ *un espace probabilisé fini. Soient* $X : \Omega \to F$ *et* $Y : \Omega \to G$ *deux v.a..*
On dit que X *et* Y *sont des* ***variables aléatoires indépendantes*** *si on a*

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

Proposition 100
Soit $(\Omega, \mathbb{P})$ *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*

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

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. finies, il n'y a qu'un nombre fini d'éléments de A (resp. B) qui sont dans l'image de X (resp. Y). et

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

□

Remarque 101 — *Pour* X, Y *deux v.a.d. indépendantes, posons* $Z = (X, Y)$.
Alors Z *est une v.a., 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 102 (**V.a. (mutuellement) indépendantes**)
Soit $(\Omega, \mathbb{P})$ *un espace probabilisé. Soient* $X_i : \Omega \to F_i$, $1 \le i \le n$, n *v.a..*
On dit que $X_1, \ldots, X_n$ *sont des* ***variables aléatoires indépendantes*** *si*

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

L'indépendance est utile pour déterminer la loi de la v.a. $X_1 + \ldots + X_n$ dans certains cas. Tout comme pour les événements, si X_1, X_2, X_3 sont deux à deux indépendantes, cela n'implique pas qu'elles sont indépendantes.

Proposition 103 (**Somme de v.a. de Bernouilli de paramètre p, indépendantes**)
Soient $n \in \mathbb{N}^*$ *et* $p \in [0, 1]$. *Soient* X_1, ..., X_n *des v.a. de Bernoulli de paramètre* $p \in]0, 1[$ *qui sont indépendantes.*
Alors, la v.a. $X_1 + X_2 + \ldots + X_n$ *est une v.a. binomiale de paramètres* n *et* p.

Preuve — Posons $S = X_1 + \ldots + X_n$. Soit $k \in \{0, \ldots, n\}$. Il faut prouver que $\mathbb{P}(S = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.
Comme chaque v.a. X_i est à valeur dans $\{0, 1\}$, on a :
$\mathbb{P}(X_1 + \ldots + X_n = k) = \mathbb{P}(\bigcup_{(x_1,\ldots,x_n)\in\{0,1\}^n,\, x_1+\ldots+x_n=k} (X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n))$.

Par σ-additivité, on a $\mathbb{P}(X_1+\ldots+X_n=k)=\sum_{(x_1,\ldots,x_n)\in\{0,1\}^n,\,x_1+\ldots+x_n=k}\mathbb{P}(X_1=x_1,X_2=x_2,\ldots,X_n=x_n)$.
Par indépendance des v.a. X_i on a

$$\mathbb{P}(X_1=x_1,X_2=x_2,\ldots,X_n=x_n)=\prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}=p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

On a donc $\mathbb{P}(S=k)=\sum_{(x_1,\ldots,x_n)\in\{0,1\}^n,\,x_1+\ldots+x_n=k}p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$
$=\sum_{(x_1,\ldots,x_n)\in\{0,1\}^n}p^k(1-p)^{n-k}=Card(\{(x_1,\ldots,x_n)\in\{0,1\}^n\}).p^k.(1-p)^{n-k}=\binom{n}{k}p^k(1-p)^{n-k}$. Cela prouve donc que S est une v.a. de loi $Bin(n,p)$. □

Proposition 104 (**Lemme des coalitions**)
Soient $n\in[\![2,+\infty[\![$ et $m\in[\![1,n-1[\![$. Soient $X_1,\ldots,X_n$ des v.a. réelles indépendantes. Soient $f:\mathbb{R}^m\to\mathbb{R}$ et $g:\mathbb{R}^{n-m}\to\mathbb{R}$.
Alors, les v.a. $f(X_1,\ldots,X_m)$ et $g(X_{m+1},\ldots,X_n)$ sont indépendantes.

7.1 Fonction de transfert et v.a. indépendantes

Proposition 105
Soient X,Y deux v.a. indépendantes. Soient $f:F\to\mathbb{R}$ et $g:G\to\mathbb{R}$.
Alors les v.a. $f(X)$ et $g(Y)$ sont indépendantes. De plus, on a :

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

Démonstration — Sur feuille. L'indépendance est nécessaire pour réaliser un découpage de somme double.

Corollaire 106
Soient X,Y deux v.a. réelles et indépendantes.
Alors, on a $\mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(Y)$ et $Var(X+Y)=Var(X)+Var(Y)$.

Démonstration — On utilise la proposition précédente.

Corollaire 107
Soit $(\Omega,\mathbb{P})$ 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 $Var(X_1+\cdots+X_n)=nVar(X_1)$.

Démonstration — On raisonne par récurrence sur n.

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.

Remarque 108 — *Nous pouvons maintenant revenir au calcul de $\mathbb{E}(X)$ et $Var(X)$ pour X une v.a. binomiale $Bin(n,p)$.*
En effet, si $X_1,\ldots,X_n$ sont des v.a. de Bernouilli de paramètre p et indépendantes, alors $X_1+\ldots+X_n$ est une v.a. binomiale de paramètres n et p (pour faire n Pile ou Face consécutifs il faut n fois un seul Pile ou Face, et que chaque Pile ou Face soit indépendant des autres).
On a alors $\mathbb{E}(X_1+\ldots+X_n)=\mathbb{E}(X_1)+\ldots+\mathbb{E}(X_n)=p+\ldots+p=np$, et $Var(X_1+\ldots+X_n)=Var(X_1)+\ldots+Var(X_n)=p(1-p)+\ldots+p(1-p)=np(1-p)$.

8 Lois conditionnelles

Soit $(\Omega,\mathbb{P})$ un espace probabilisé fini, et $X:\Omega\to F$, $Y:\Omega\to G$, deux v.a..
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.
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 109
Soit $(\Omega, \mathbb{P})$ un espace probabilisé fini, 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 110 — *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 111 — *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 112
Soient $X : \Omega \to F$, $Y : \Omega \to G$, deux v.a.
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 113 — *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 114
Soient $X : \Omega \to F$, $Y : \Omega \to G$, deux v.a. On a alors :

1. $\forall x \in F$, $\mathbb{P}(X = x) = \sum\limits_{y \in G} \mathbb{P}\big(Z = (x, y)\big)$;
2. *Si* $\mathbb{P}(X = x) > 0$, $\mathbb{P}_{Y|X=x}(\{y\}) = \mathbb{P}(Y = y|X = x) = \dfrac{\mathbb{P}(Z = (x, y))}{\mathbb{P}(X = x)}$.
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.

Démonstration —Admis.

REMARQUE 115 — *Attention ! Pour caractériser la loi de probas de $Z = (X, Y)$, il faut utiliser la loi de probas de X et les lois de probas de Y sachant $(X = x)$.*

Connaître la loi de probas de X et celle de Y ne suffisent pas. Cela nous dit comment sont X et Y, mais pas comment ces 2 v.a. interagissent.

PROPOSITION 116
Avec les notations précédentes, on a :

$$\mathbb{E}(Y) = \sum_{x_i \in F \ tq \ PP(X=x_i)>0} \mathbb{E}(Y|X = x_i)\mathbb{P}(X = x_i).$$

Démonstration —Admis.

REMARQUE 117 — *Ce résultat permet de calculer l'espérance $\mathbb{E}(Y)$ en conditionnant par une autre v.a. X.*
Il généralise la formule des probabilités totales, qui correspond ici à $Y = \mathbb{1}_A$, et $B_i = \{X = x_i\}$:

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

9 Inégalité de Markov, Inégalité de Bienaymé-Tchebychev

Nous commençons par une inégalité "simple" mais très souvent utile.

PROPOSITION 118 (**Inégalité de Markov**)
Soient $(\Omega, \mathbb{P})$ un espace probabilisé fini, et X une v.a. réelle.
Pour tout $\varepsilon > 0$, on a

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X|)}{\varepsilon}.$$

Preuve — 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 119 — *On peut montrer de la même façon que pour tout $p \geq 1$ et 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 120 (**Inégalité de Bienaymé-Tchebychev**)
Soient $(\Omega, \mathbb{P})$ un espace probabilisé fini, et X une v.a. réelle.
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 121 — *L'inégalité de Bienaymé-Tchebychev majore la probabilité que la v.a. X s'éloigne d'au moins ε de sa moyenne.*
Cette inégalité sert entre autres à démontrer le théorème suivant, la loi faible des grands nombres.

Théorème 122 (**Loi faible des grands nombres (HP)**)
Soit $(\Omega, \mathbb{P})$ *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 $S_n = \sum\limits_{k=1}^{n} X_k$.
Alors pour tout $\varepsilon > 0$, *on a*

$$\lim_{n\to+\infty} \mathbb{P}\left(|\frac{S_n}{n} - \mathbb{E}(X_1)| \geq \varepsilon \right) = 0.$$

Exemple 123 — *On réalise* $n = 4000$ *Pile ou Face, indépendants entre eux, avec probabilité* $p = 0.15$ *d'obtenir Pile, et on compte le nombre de Pile obtenus.*
La fonction X *qui compte le nombre de Pile obtenus est ainsi une v.a. de loi binomiale de paramètres* n *et* p.
Une personne indique avoir obtenu 800 *Pile de cette façon.*
Dans cette situation, on a $\mathbb{E}(X) = 4000 \times 0.15 = 600$, *et* $Var(X) = 4000 \times 0.15 \times 0.85 = 510$.
Le résultat déclaré par la personne est significativement au-dessus de la valeur moyenne (l'espérance) de X. *Regardons la probabilité d'avoir été au moins aussi "chanceux" que la personne.*
On veut ainsi calculer $\mathbb{P}(X \geq 800)$.
L'image de X *est* $\{0, \ldots, 4000\}$, *donc* $\mathbb{P}(X \geq 800) = \sum_{k=600}^{4000} \mathbb{P}(X = k) = \sum_{k=800}^{4000} \binom{4000}{k} 0.15^k 0.85^{4000-k}$.
Mais, on ne peut pas simplifier davantage cette somme pour pouvoir la calculer (sauf à demander à Python une valeur approchée, mais cela demande du temps de calcul et de la manipulation de nombres réels relativement petits). Utilisons à la place l'inégalité de Bienaymé-Tchebychev.
On a $\mathbb{P}(X \geq 800) = \mathbb{P}(X - 600 \geq 200) \leq \mathbb{P}(|X - \mathbb{E}(X)| \geq 200) \leq \frac{Var(X)}{200^2} = 0.01275$.
Cette probabilité d'obtenir un résultat autant éloigné (ou plus) de $\mathbb{E}(X)$ *est donc inférieure à* 1.3%.
La valeur exacte de $\mathbb{P}(X \geq 800)$ *peut être bien plus faible que cela, mais cette majoration permet rapidement de dire que l'événement "obtenir 800 Pile ou mieux" a une faible probabilité d'arriver.*
L'inégalité de Markov fournit quant à elle $\mathbb{P}(X \geq 800) \leq \frac{\mathbb{E}(X)}{800} = \frac{600}{800} = \frac{3}{4}$, *ce qui est bien moins précis dans ce cas que l'inégalité de Bienaymé-Tchebychev.*

Bilan du contenu <u>nécessaire</u> à maîtriser :

— Rappels sur le dénombrement, calcul du cardinal d'un ensemble fini. Cardinal d'une union, d'un produit. Coefficients binomiaux, propriétés. Nombres de tirages de k éléments dans un ensemble à n éléments (avec ou sans ordre, avec ou sans remise).

— Fonction indicatrice d'une partie A, $\mathbb{1}_A$. $\mathbb{1}_A(x)$ vaut 1 si $x \in A$ et 0 sinon.

— Connaître le langage des probabilités (univers, événement, probabilité).

— Définition d'une (mesure de) probabilité $\mathbb{P}$ sur Ω. C'est une fonction de $\mathcal{P}(\Omega)$ vers $[0,1]$ qui est sigma-additive. Le nombre $\mathbb{P}(A)$ est la probabilité de la partie A.
Propriétés des probabilités (somme disjointe, somme, complémentaire).

— (Mesure de) probabilités uniforme. Si $\mathbb{P}$ est uniforme alors $\mathbb{P}(A) = \frac{Card(A)}{Card(\Omega)}$.

— Caractérisation des mesures de probabilités, sur un ensemble fini $\Omega = \{\omega_1, \ldots, \omega_n\}$. La distribution d'une mesure de probas est la famille $(\mathbb{P}(\{\omega_1\}), \ldots, \mathbb{P}(\{\omega_n\}))$.
Relations $\mathbb{P}(A) = \sum_{x\in A} \mathbb{P}(\{x\})$ et $\mathbb{P} = \sum_{i=1}^{n} \mathbb{P}(\{\omega_i\})\mathbb{1}_{\{\omega_i\}}$. Exemples classiques (jets de dés, tirages de cartes,...).

— Définition d'espace probabilisé $(\Omega, \mathbb{P})$.
La notion de "hasard" sur Ω dépend de la (mesure de) probas $\mathbb{P}$ que l'on choisit.

— Tirage dans une urne : Le tirage sans remise et le tirage simultané donnent des probabilités identiques. Le tirage avec remise est différent.

— Probabilités conditionnelles $\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$.
Formule des probabilités composées : $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \ldots \mathbb{P}(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$
Formule des probabilités totales : $\mathbb{P}(A) = \sum\limits_{i\in I} \mathbb{P}(A \cap B_i) = \sum\limits_{i\in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$.
Formule de Bayes : $\forall i \in I$, $\mathbb{P}(B_i|A) = \dfrac{P(A|B_i)P(B_i)}{\sum_{j\geq 0} \mathbb{P}(A|B_j)P(B_j)}$.
Savoir calculer des probabilités conditionnelles. Savoir calculer des probabilités à l'aide des probas conditionnelles.

— Evénements indépendants ($\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$), définition générale pour $A_1, \ldots, A_n$..

— Une variable aléatoire est une **fonction** $X : \Omega \to E$. Une v.a. réelle est $X : \Omega \to \mathbb{R}$.
Pour $A \subset \mathbb{R}$, on pose $(X \in A) = \{\omega \in \Omega, \text{ t.q. } X(\omega) \in A\}$ et $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$.
Pour $X(\Omega) = \{x_1, \ldots, x_r\}$, la loi de X est la fonction $\mathbb{P}_X$. Sa distribution est la famille $(\mathbb{P}(X = x_1), \ldots, \mathbb{P}(X = x_r))$.
Calcul de probabilités : $\mathbb{P}(X \in A) = \sum_{i=1}^{r} \mathbb{P}(X = x_i)\mathbb{1}_A(x_i) = \sum_{\omega\in(X\in A)} \mathbb{P}(\omega)$.

— On définit l'espérance de la v.a.r. X par $\mathbb{E}(X) = \sum_{i=1}^{r} x_i\mathbb{P}(X = x_i)$.
Formule de transfert : $\mathbb{E}(X) = \sum_{j=1}^{n} X(\omega_j)\mathbb{P}(\omega_j)$.
L'espérance $X \mapsto \mathbb{E}(X)$ est une application linéaire. Si $X \leq Y$ alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.
L'espérance a les mêmes propriétés que l'intégrale $f \mapsto \int_0^1 f(t)dt$.

— Théorème de transfert : $\mathbb{E}(f(X)) = \sum\limits_{x\in X(\omega)} f(x)\mathbb{P}(X = x)$.

— Varince de X : $Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$. Ecart-type : $\sigma_X = \sqrt{Var(X)}$.
Propriété : $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.
Intérêt de l'espérance et de la variance pour étudier la v.a. X.

— Lois de v.a. usuelles : Bernouilli $\mathcal{B}(p)$, binômiale $Bin(n,p)$ (distribution, espérance, variance), uniforme sur $\{x_1, \ldots, x_r\}$ ou $\{0, \ldots, n\}$ (distribution, espérance).

— Variables aléatoires indépendantes, mutuellement indépendantes.
Si X et Y sont indépendantes alors $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$.
Lemme des coalitions : Si $X_1, \ldots, X_n$ sont mutuellement indépendantes, alors $f(X_1, \ldots, X_k)$ et $g(X_{k+1}, \ldots, X_n)$ sont indépendantes.
Si $X_1, \ldots, X_n$ sont mutuellement indépendantes, alors $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$.

— Calcul d'espérance avec les probas. conditionnelles et la formule des probabilités totales.

— Inégalité de Markov : $\mathbb{P}(|X| \geq \varepsilon) \leq \dfrac{\mathbb{E}(|X|)}{\varepsilon}$.
Inégalité de Bienaymé-Tchebychev : $\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2} = \dfrac{\sigma_X^2}{\varepsilon^2}$.