Cours de mathématiques
Partie IV – Probabilités
MPSI 4

Alain TROESCH

Version du:

9 mai 2019

Table des matières

29

Combinatoire et dénombrement

"What's one and one and one and one and one and one and one and one and one and one ?"
"I don't know," said Alice. "I lost count."
"She can't do Addition."

(Lewis Carroll)

Le dénombrement est à l'origine des mathématiques : compter.

Dénombrer un ensemble d'objets, c'est déterminer le cardinal de cet ensemble. Le dénombrement se base souvent sur la combinatoire, qui est l'étude des configurations possibles d'un ensemble d'objets, donc l'étude de la structure d'un ensemble, passant éventuellement par une description décomposée des objets (ce qui correspond à une construction par choix successifs), ou un tri des objets suivant certains critères.

La clé du dénombrement combinatoire est la notion de bijection : si je peux construire une correspondance un-à-un entre les objets d'un ensemble E et les objets d'un ensemble F (c'est-à-dire une bijection de E dans F), alors E a autant d'objets que F.
C'est d'ailleurs, comme nous l'avons vu, comme cela qu'on définit le cardinal d'un ensemble fini, par comparaison à un ensemble de type $[\![1, n]\!]$, correspondant à une énumération des éléments de E (l'action de compter)

Par la suite, il est souvent un peu maladroit de revenir systématiquement à la définition en comparant à un ensemble $[\![1, n]\!]$, et on cherche plutôt la mise en bijection avec des ensembles de référence bien connus et étudiés. Une telle méthode nécessite souvent une étude de la structure de l'ensemble étudié, et de la façon de construire les objets de cet ensemble : c'est une méthode relevant de la combinatoire.

Les dénombrements sont la base d'un grand nombre de calculs de probabilités, notamment dans une situation d'équiprobabilité : dans ce cas, en effet, de façon assez intuitive, la probabilité d'un événement est le rapport entre le nombre d'issues favorables et le nombre total d'issues possibles. Même dans des situations plus complexes, les dénombrements élémentaires gardent une place centrale.
Pour cette raison, le dénombrement semble trouver sa place naturelle au début d'un cours de probabilité, même si ses applications sont beaucoup plus diverses dans l'ensemble des mathématiques.

I Combinatoire des ensembles finis

Dans cette section, on rappelle la définition et les propriétés suivantes, vues dans un chapitre antérieur :

Définition 29.1.1 (Cardinal d'un ensemble)

E est de cardinal $n \in \mathbb{N}$, si et seulement s'il existe une bijection $\varphi : [\![1, n]\!] \to E$. On note $\text{Card}(E) = n$, ou $|E| = n$, ou encore $\sharp E = n$.

Proposition 29.1.2 (Cardinal d'une union disjointe)
Soit A, B, $A_1, \dots, A_n$ des ensembles finis.

1. *Si $A \cap B = \varnothing$, alors $|A \cup B| = |A| + |B|$.*
2. *Plus généralement, si pour tout $(i,j) \in [\![1,n]\!]^2$ tel que $i \neq j$, $A_i \cap A_j = \varnothing$, alors*

$$|A_1 \cup \cdots \cup A_n| = |A_1| + \cdots + |A_n|.$$

Proposition 29.1.3 (Cardinal d'un complémentaire)
Si $A \subset B$, alors $\left|\complement_B A\right| = |B| - |A|$.

Proposition 29.1.4 (Cardinal d'un sous-ensemble)
Si $A \subset B$, alors $|A| \leqslant |B|$, avec égalité si et seulement si $A = B$.

Proposition 29.1.5 (Cardinal d'une union quelconque)
Soit A et B des ensembles finis. On a :

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Théorème 29.1.6 (Formule du crible de Poincaré, ou formule d'inclusion-exclusion)
Soit $A_1, \dots, A_n$ des ensembles finis. Alors :

$$|A_1 \cup \cdots \cup A_n| = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} |A_{i_1} \cap \cdots \cap A_{i_k}| = \sum_{\substack{I \subset [\![1,n]\!] \\ I \neq \varnothing}} (-1)^{|I|-1} \left| \bigcap_{i \in I} A_i \right|.$$

Proposition 29.1.7 (Cardinal d'un produit cartésien)

1. *Soit A et B deux ensembles finis. Alors $|A \times B| = |A| \times |B|$.*
2. *Plus généralement, soit $A_1, \dots, A_n$ des ensembles finis ; Alors*

$$|A_1 \times \cdots \times A_n| = \prod_{i=1}^{n} |A_i|.$$

II Combinatoire des ensembles d'applications

Nous dénombrons dans cette partie certains ensembles de fonctions, qui serviront de modèles par la suite pour des situations probabilistes liées à des expériences de tirages. Certains de ces résultats ont déjà été démontrés ; il s'agit ici plutôt de grouper et classifier un peu ces résultats rencontrés de façon éparse.

II.1 Applications quelconques ; p-listes

Proposition 29.2.1 (Cardinal de l'ensemble des applications)
Soit E et F deux ensembles finis. On rappelle qu'on note F^E l'ensemble des applications de E vers F. Alors $|F^E| = |F|^{|E|}$.

◁ **Éléments de preuve.**
une bijection $[\![1, n]\!] \to E$ définit une énumération $x_1, \dots, x_n$ des éléments de E. Remarquer que F^E est alors en bijection avec F^n, en associant à f le n-uplet $(f(x_1), \dots, f(x_n))$. ▷

Définition 29.2.2 (p-listes)
Une *p-liste* d'éléments de F (ou p-uplet) est un élément $(x_1, \dots, x_p)$ de F^p.

Une p-liste peut être vue indifféremment comme un élément d'un produit cartésien $F \times \cdots \times F$, ou de l'ensemble $F^{[\![1,p]\!]}$ des fonctions de $[\![1, p]\!]$ dans F, associant x_i à i. Ce second point de vue aura l'avantage de mieux faire comprendre certaines propriétés imposées sur une liste. Ainsi, les listes d'éléments distincts correspondent aux fonctions injectives.
Évidemment, les deux points de vue amènent de façon immédiate le dénombrement suivant :

Proposition 29.2.3 (Nombre de p-listes)
Le nombre de p-listes d'éléments de F est $|F|^p$.

Enfin, étant donné E un ensemble fini, L'ensemble $\mathcal{P}(E)$ peut être mis en bijection avec l'ensemble des applications de E vers $\{0, 1\}$ *via* les fonctions caractéristiques. On obtient donc :

Proposition 29.2.4 (Cardinal de l'ensemble des parties)
$|\mathcal{P}(E)| = 2^{|E|}$.

II.2 Lemme du berger

Le résultat suivant permet de donner de la rigueur à tous les arguments combinatoires reposant sur des choix successifs :

Lemme 29.2.5 (Lemme du berger)
Soit $f : E \to F$ une application surjective. On suppose qu'il existe un entier $k \in \mathbb{N}^*$ tel que pour tout $y \in F$, $|f^{-1}(y)| = k$ (tous les éléments de F ont le même nombre k d'antécédents). Alors $|E| = k \cdot |F|$.

◁ **Éléments de preuve.**
Considérer la partition de E définie par cette surjection : le nombre de parts est $|F|$ et chaque part est de cardinal k. ▷

Remarque 29.2.6
- Ce lemme dit essentiellement que si tout mouton a quatre pattes, il y a quatre fois plus de pattes que de moutons (la fonction f étant ici la fonction associant à une patte donnée le mouton auquel elle est rattachée).

- Le lemme du berger permet de formaliser la notion de « choix successifs ». Il est souvent utilisé de façon implicite dans les raisonnements. Il faut essentiellement en retenir que lorsqu'on fait des choix successifs, et qu'à chaque étape, le nombre de possibilité ne dépend pas de la situation dans laquelle on se trouve (c'est-à-dire des choix précédents), alors le nombre total de possibilités s'obtient en faisant le produit du nombre de possibilités à chaque étape.
- La formalisation d'un tel choix par le lemme du berger se fait en considérant la fonction qui à un choix possible associe la situation à l'étape précédente permettant ce choix. Le calcul du nombre d'injections est un exemple typique d'utilisation du lemme du berger.
- À la longue, on ne formalisera plus complètement ce type d'arguments, et on se contentera de l'approche intuitive des choix successifs. Mais il faut bien être conscient à tout moment que cette approche intuitive peut être formalisée rigoureusement. C'est seulement lorsqu'on est pleinement conscient de la possibilité de faire cette formalisation de la sorte qu'on peut s'en dispenser.

II.3 Injections ; p-listes d'éléments distincts

Nous avons déjà démontré par des arguments intuitifs (choix successifs) le résultat suivant. Nous en donnons maintenant une formalisation reposant sur le lemme du berger.

Théorème 29.2.7 (Dénombrement des injections)

Soit A et B deux ensembles de cardinaux respectifs p et n. Alors, si $p \leqslant n$, le nombre d'injections de A vers B est $A_n^p = \frac{n!}{(n-p)!}$. Si $p > n$, il n'existe pas d'injection de A vers B.

◁ **Éléments de preuve.**

Se ramener au cas où $A = [\![1, p]\!]$, par commodité.

L'idée intuitive consiste à dire qu'on a n façons de choisir une valeur pour $f(1)$, puis plus que $n-1$ façons de choisir une valeur pour $f(2)$ (pour préserver l'injectivité) etc. : à chaque étape, on diminue le nombre de choix possibles.

Il s'agit donc de choix successifs : chaque étape peut se formaliser par le lemme du berger. Faire une récurrence sur p permet alors de n'avoir qu'une étape à rédiger. Appliquer le lemme du berger à l'application qui à une injection $f : [\![1, p+1]\!] \to B$ associe sa restriction à $[\![1, p]\!]$. Comment gérer le cas $p > n$? ▷

Une transcription en terme de listes donne alors :

Proposition 29.2.8 (Dénombrement des p-arrangements)

Soit F de cardinal n et $p \leqslant n$. Le nombre de p-listes d'éléments distincts de F (ou p-arrangements de F) est $A_n^p = \frac{n!}{(n-p)!}$.

En particulier, comme on le sait déjà :

Corollaire 29.2.9 (Nombre de permutations d'un ensemble)

1. *Soit E un ensemble fini. Alors $|\mathfrak{S}E| = |E|!$*
2. *En particulier, $|\mathfrak{S}_n| = n!$*

Définition 29.2.10 (Permutation de n éléments)

Une permutation de n éléments distincts $x_1, \ldots, x_n$ est une façon d'ordonner entre eux ces éléments. Il s'agit donc d'un n-uplet $(y_1, \ldots, y_n)$ tel que chaque x_i soit égal à un et un seul des y_i. On note $\mathrm{Perm}(x_1, \ldots, x_n)$ l'ensemble des permutations des éléments $x_1, \ldots, x_n$.

Proposition 29.2.11
$\mathrm{Perm}(x_1, \dots, x_n) \simeq \mathfrak{S}_n$. *Ainsi,* $|\mathrm{Perm}(x_1, \dots, x_n)| = n!$.

II.4 Surjections

Dénombrer les surjections est un problème plus dur (lié à ce qu'on appelle les nombres de Stirling). L'exemple ci-dessous est simple, mais introduit les problèmes qu'on peut rencontrer lorsqu'on cherche à dénombrer des surjections dans des cas plus généraux.

Exemple 29.2.12
Le nombre de surjections de $[\![1, n]\!]$ dans $[\![1, n-1]\!]$ est $(n-1)!\binom{n}{2}$.

III Combinatoire des sous-ensembles

Nous avons déjà défini algébriquement le coefficient binomial, et nous en avons déduit les différentes propriétés importantes, ainsi que la formule du binôme. Nous donnons dans ce chapitre un autre point de vue, partant d'une definition purement combinatoire du coefficient binomial.

Définition 29.3.1 (Coefficient binomial)
Le *coefficient binomial* $\binom{n}{k}$ est le nombre de parties à k éléments de $[\![1, n]\!]$.

Proposition 29.3.2
Le coefficient binomial $\binom{n}{k}$ *est plus généralement le nombre de sous-ensembles de cardinal* k *de n'importe quel ensemble* E *de cardinal* n.

◁ **Éléments de preuve.**
Montrer qu'une bijection $E \to F$ induit une bijection $\mathcal{P}_k(E) \to \mathcal{P}_k(F)$, et appliquer cela à $E \to [\![1, n]\!]$. ▷

Un peu de combinatoire nous permet de retrouver la définition algébrique donnée en début d'année :

Proposition 29.3.3 (Expression factorielle du coefficient binomial)
Pour $k \in [\![0, n]\!]$, $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$.

◁ **Éléments de preuve.**
Appliquer le lemme du berger à l'application qui a un k-arrangement $(x_1, \dots, x_k)$ d'éléments de $[\![1, n]\!]$ associe le sous-ensemble $\{x_1, \dots, x_k\}$. ▷

Nous avons déjà eu l'occasion de mentionner le fait que la définition combinatoire même du coefficient binomial fournit diverses inerprétations possibles en terme de dénombrement, les plus importantes étant les suivantes :

- $\binom{n}{p}$ est le nombre de mots de longueur n, constitué de p lettres a et $n-p$ lettres b.

- $\binom{n}{p}$ est le nombre de chemins de longueur n consitués de p pas vers le haut et $n-p$ pas vers la droite. Ainsi, $\binom{a+b}{a}$ est le nombre de chemins constitués de pas à droite et vers le haut, reliant $(0,0)$ à (a,b).

La définition combinatoire du coefficient binomial permet d'obtenir assez élégamment les formules classiques associées :

Proposition 29.3.4 (Propriétés du coefficient binomial)
Soit $(n,k) \in \mathbb{Z}^2$. On a :

(i) $\binom{n}{k} = \binom{n}{n-k}$ *(symétrie)*

(ii) $k\binom{n}{k} = n\binom{n-1}{k-1}$ *(formule comité-président)*

(iii) si $(n,k) \neq (0,0)$, $\binom{n}{k} + binomnk + 1 = \binom{n+1}{k+1}$ *(formule de Pascal)*

◁ **Éléments de preuve.**

(i) La complémentation définit une bijection

(ii) Compter les couples (F,x) tels que $F \in \mathcal{P}_k(n)$ et $x \in F$.

(iii) Trier les sous-ensembles de $[\![1,n]\!]$ suivant qu'ils contiennent ou non l'élément n.

▷

On rappelle que cette dernière formule est à la base du triangle de Pascal. On termine en donnant une démonstration combinatoire de la formule du binôme.

Théorème 29.3.5 (Formule du binôme)
Pour $n \in \mathbb{N}$, $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.

◁ **Éléments de preuve.**
Compter les facteurs du type $a^k b^{n-k}$ apparaissant dans le développement. Il y en a autant que de choix de k des facteurs $a+b$, fournissant un terme a dans le développement. ▷

IV Bijection, Déesse de la Combinatoire

La bijection étant au coeur-même de la définition du cardinal, elle tient un rôle central dans un grand nombre de problèmes de dénombrement. On obtient en particulier le principe fondamental suivant :

Méthode 29.4.1 (Principe fondamental du dénombrement)
Pour montrer que deux ensembles ont même cardinal, il suffit de construire une bijection entre eux. Ainsi, pour déterminer le cardinal d'un ensemble, on le met souvent en bijection avec un ensemble « de référence » dont on connaît le cardinal.

Nous en avons déjà vu des exemples d'utilisation, par exemple la preuve de la symétrie des coefficients binomiaux. Voici quelques autres exemples, qui représentent des situations très classiques, à bien connaître :

Exemples 29.4.2

1. Dénombrer les p-listes $(k_1, \dots, k_p)$ d'entiers strictement positifs tels que $k_1 + \cdots + k_p = n$.
2. Dénombrer les p-listes $(k_1, \dots, k_p)$ d'entiers *positifs ou nuls* tels que $k_1 + \cdots + k_p = n$.
3. Dénombrer les p-listes strictement croissantes d'éléments de $[\![1, n]\!]$.
4. Dénombrer les p-listes croissantes d'éléments de $[\![1, n]\!]$.

V Preuves combinatoires d'identités ?

Idée : dénombrer de deux manières différentes le même ensemble, de manière à obtenir des relations, dont certaines ne sont pas toujours évidentes à démontrer analytiquement.

Méthode 29.5.1 (Démonstration combinatoire d'une formule)

1. Trouver un modèle adapté à la formule, autrement dit un ensemble d'objets dont le dénombrement fournira un des membres de l'égalité. Pour cela, il est préférable de s'aider du membre le plus simple de l'égalité.
2. Dénombrer cet ensemble de deux façons différentes. Souvent, on procède d'une part à un dénombrement direct, et d'autre part à un dénombrement après avoir effectué un tri (de façon formelle, cela revient à définir une partition de l'ensemble). Le résultat d'un dénombrement par tri se traduit par une somme.
3. Évidemment, cette méthode n'est adaptée qu'à des formules portant sur des nombres entiers, si possible positifs. Il est parfois possible de se ramener à cette situation par un prétraitement de la formule à démontrer.

Proposition 29.5.2 (Quelques formules se démontrant combinatoirement)

1. *Formule de Pascal :* $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$.
2. $\displaystyle\sum_{k=0}^{n} \binom{n}{k} = 2^n$.
3. *Formule de Vandermonde :* $\displaystyle\sum_{k=0}^{n} \binom{N}{k}\binom{M}{n-k} = \binom{N+M}{n}$.
4. *Formule de sommation :* $\displaystyle\sum_{k=0}^{p} \binom{n+k}{n} = \binom{n+p+1}{n+1}$
5. *C'est un cas particulier d'une formule plus générale, de type Vandermonde :*

$$\sum_{k=0}^{n} \binom{k}{N}\binom{n-k}{M} = \binom{n+1}{M+N+1}$$

◁ **Éléments de preuve.**

1. Déjà vu
2. Trier les sous-ensembles de $[\![1, n]\!]$ suivant le cardinal
3. Composer des bouquets de n fleurs, disposant de 2 types de fleurs.
4. Trier les sous-ensembles à $n+1$ éléments de $[\![1, n+p+1]\!]$ suivant la valeur de leur élément maximum.
5. Trier les sous-ensembles à $n+1$ éléments de $[\![1, M+N+1]\!]$ suivant la valeur de leur ..... élément.

▷

Ces formules peuvent intervenir notamment dans certains calculs d'espérance ou de variance.

Méthode 29.5.3

Remarquez qu'un signe $(-1)^k$ associé à un coefficient binomial correspond souvent à une comparaison de parités de cardinaux. On peut passer d'un cardinal pair à un cardinal impair, et vice-versa, en « allumant ou éteignant » un élément fixé à l'avance, suivant qu'il est déjà ou non dans notre ensemble (plus précisément, il s'agit de l'opération $X \mapsto X \triangle \{x\}$).
Nous appelerons cela le principe de l'interrupteur.

Exemple 29.5.4

Démonstration combinatoire de la formule $\displaystyle\sum_{k=0}^{n}(-1)^k\binom{n}{k} = \delta_{n,0}$, pour tout $n \in \mathbb{N}$.

30 Espaces probabilisés

« Ce n'est qu'en essayant continuellement que l'on finit par réussir, donc plus ça rate, plus on a de chances que ça marche »
« La fusée interplanétaire des Shadoks n'était pas très au point, mais ils avaient calculé qu'elle avait quand même une chance sur un million de marcher. Et ils se dépêchaient de bien rater les 999999 premiers essais pour être sûrs que le millionnième marche. »

(Jacques Rouxel)

Il s'agit dans ce chapitre de donner un cadre rigoureux au calcul des probabilités, et à établir un certain nombre de formules calculatoires essentielles, permettant de mener à bien et de façon rigoureuse la plupart des calculs de probabilités.

I Espaces probabilisables

I.1 Notion d'expérience aléatoire

Une *expérience aléatoire* est une donnée intuitive : c'est une expérience dont le résultat ne dépend que du hasard. Plus explicitement, cela signifie que s'il est possible de répéter l'expérience plusieurs fois dans les mêmes conditions et de façon identique, les différents résultats ne seront pas nécessairement les mêmes. Ainsi, le résultat d'une expérience aléatoire n'est en général pas prévisible. L'exemple typique est le lancer d'une pièce, qui va retomber sur Pile ou Face, ou bien le lancer d'un dé : les lois physiques régissant la mécanique d'une pièce lancée en l'air ou d'un dé roulant sur une table sont trop complexes et trop sensibles à des variations des conditions initiales, et aux éventuelles perturbations extérieures pour espérer obtenir un résultat déterministe, même avec un lancer très précis.
Une expérience, malgré son nom, peut être purement théorique, dans le sens ou elle ne peut pas être réalisée de façon effective. Par exemple l'expérience consistant à lancer une infinité de fois une pièce de monnaie. Il peut aussi s'agir d'une simple observation d'un phénomène naturel ne nécessitant pas d'intervention de l'observateur. C'est le cas en physique (notamment en physique des particules).

Définition 30.1.1 (Univers)

- Un *résultat*, ou une *issue* de l'expérience est une donnée issue de l'expérience aléatoire ; une même expérience peut fournir différents résultats, suivant ce qu'on veut étudier de l'expérience.
- L'*univers des possibles* Ω (ou plus simplement l'univers) est l'ensemble des issues possibles d'une expérience.
  Une même expérience peut fournir plusieurs univers différents suivant ce qu'on veut en tirer.

Intuitivement, un événement correspond à un groupement d'issues possibles vérifiant une certaine propriété. Ainsi, il s'agit d'un ensemble d'issues, donc d'un sous-ensemble de l'univers.

Définition 30.1.2 (Évenement, définition intuitive)

Un événement est un sous-ensemble de Ω. Pour certaines raisons techniques, lorsque Ω n'est pas fini, on est parfois amené à se restreindre et à ne pas considérer tous les sous-ensembles comme des événements, ce que nous formaliserons plus loin avec la notion de tribu.

La plupart des notions ensemblistes ont une traduction dans le langage des probabilités, afin de mieux traduire leur interprétation intuitive.

Terminologie 30.1.3 (Vocabulaire probabiliste)

- *Événement élémentaire*, ou *épreuve* : un singleton $\{\omega\} \subset \Omega$.
- *Événement certain* : l'événement Ω.
- *Événement impossible* : l'événement $\varnothing$.
- *Événement contraire de l'événement* A : l'événement $\overline{A} = \complement_\Omega A$.
- A *entraîne* B *si* $A \subset B$
- *Événement « A et B »* : $A \cap B$ (réalisation simultanée)
- *Événement « A ou B »* : $A \cup B$
- A et B sont dits *incompatibles* si $A \cap B = \varnothing$.
- une famille $\{A_i, i \in I\}$ est dite *constituée d'événements deux à deux incompatibles* si pour tout $(i,j) \in I^2$ tel que $i \neq j$, A_i et A_j sont incompatibles.

Définition 30.1.4 (Système complet d'événements)

Un système complet d'événements est une famille $\{A_i, i \in I\}$ formant une partition de Ω. Autrement dit :

- Les événements A_i sont non vides ;
- La famille est constituée d'événements deux à deux incompatibles ;
- $\bigcup_{i \in I} A_i = \Omega$.

I.2 σ-algèbres d'événements (ou tribus)

Souvent, la définition de la mesure de probabilité sur tous les sous-ensembles de Ω, n'est pas pertinente, ou peut poser des problèmes techniques. Pour cette raison, il peut être intéressant de pouvoir ne définir la mesure de probabilité que sur une classe particulière de sous-ensembles de Ω. L'objet de ce paragraphe et du suivant est de définir un type satisfaisant de classe de sous-ensembles à laquelle se restreindre.
Voici nos exigences pour la construction d'une telle classe :

- L'événement certain Ω et l'événement impossible ont une probabilité, donc appartiennent à la classe ;
- Pour tout événement A dont on sait calculer la probabilité, on voudrait pouvoir calculer la probabilité de l'événement contraire $\overline{A}$;
- Si on dispose d'une probabilité sur A et B, on voudrait disposer d'une probabilité de $A \cup B$.
- Plus généralement, on souhaite pouvoir calculer la probabilité d'une union infinie dénombrable d'événements.

Remarque 30.1.5

Le dernier point est motivé par l'étude d'un tirage infini à Pile ou Face : quelle est la probabilité qu'un tirage donné (par exemple n'obtenir que des Piles) ait lieu ? Pour ce calcul, on peut encore se dispenser des unions (ou intersections par passage au complémentaire) infinies, en effectuant des majorations, si on définit correctement la probabilité ; mais en revanche, si on se demande maintenant quelle est

la probabilité qu'un résultat soit stationnaire, on a besoin de regrouper une infinité de cas possibles (stationnaires à partir d'un certain rang).

Cela nous amène à la définition suivante :

Définition 30.1.6 (σ-algèbre, ou tribu, Spé)

Soit Ω un univers (fini ou non). Une *σ-algèbre* $\mathcal{A}$ d'événements sur Ω (ou tribu) est un sous-ensemble $\mathcal{T}$ de $\mathcal{P}(\Omega)$ telle que :

1. $\Omega \in \mathcal{A}$;
2. $\forall A \in \mathcal{P}(\Omega),\ A \in \mathcal{A} \Longrightarrow \overline{A} \in \mathcal{A}$ (stabilité par complémentation) ;
3. Pour tout famille dénombrable $(A_i)_{i\in I}$ d'éléments de $\mathcal{T}$, $\bigcup_{i\in I} A_i$ est dans $\mathcal{T}$ (stabilité par union dénombrable)

Nous complétons les propriétés imposées par la définition par les suivantes :

Proposition 30.1.7 (Propriétés des tribus, Spé)

Soit $\mathcal{T}$ une tribu d'événements sur Ω.

1. $\varnothing \in \mathcal{T}$;
2. $\forall (A,B) \in \mathcal{P}(\Omega),\ (A,B) \in \mathcal{T}^2 \Longrightarrow A \cup B \in \mathcal{T}$ *(stabilité par union finie) ;*
3. *Pour tout famille dénombrable $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{T}$, $\bigcap_{n=0}^{+\infty} A_n \in \mathcal{T}$.*
4. $\forall (A,B) \in \mathcal{P}(\Omega),\ (A,B) \in \mathcal{T}^2 \Longrightarrow A \cap B \in \mathcal{T}$ *(stabilité par intersection finie) ;*

◁ **Éléments de preuve.**
Le point 2 se montre en complétant la famille (A, B) en une famille dénombrable, en ajoutant l'ensemble vide, ce qui ne change pas l'union. Les autres s'obtiennent par passage au complémentaire. ▷

Par une récurrence immédiate, les points 2 et 4 se généralisent facilement aux unions ou intersections d'un nombre fini quelconque d'événements.

Remarque 30.1.8

En n'imposant que la stabilité par union finie, on retrouve la définition d'une algèbre de Boole, déjà rencontrée au cours de l'année. Ainsi, les σ-algèbres sont des algèbres de Boole, vérifiant une propriété de stabilité un peu plus forte.

Définition 30.1.9 (Espace probabilisable et événements, Spé)

- Un *espace probabilisable* est un couple $(\Omega, \mathcal{T})$, où Ω est un ensemble quelconque, appelé univers, et $\mathcal{T}$ une σ-algèbre d'événements sur Ω.
- Les éléments de $\mathcal{T}$ sont appelés *événements* (de l'espace probabilisable $(\Omega, \mathcal{T})$).

Lemme 30.1.10 (Intersection de tribus, Spé)

Soit $(\mathcal{T}_i)_{i\in I}$ une famille de σ-algèbres sur Ω. Alors $\bigcap_{i\in I} \mathcal{T}_i$ est une σ-algèbre d'événements.

◁ **Éléments de preuve.**
Vérification immédiate de la préservation des trois points de la définition par intersection. ▷

Définition 30.1.11 (Tribu engendrée par une famille, Spé)
Soit $\mathcal{C} = \{A_i,\ i \in I\}$ une famille quelconque d'événements dans $\mathcal{P}(\Omega)$. La σ-algèbre $\mathcal{T}$ (notée parfois $\mathcal{T}(\mathcal{C})$) engendrée par la famille $\mathcal{C}$ est la plus petite σ-algèbre contenant $\mathcal{C}$. Elle vérifie donc :

1. $\mathcal{C} \subset \mathcal{T}$;
2. pour toute σ-algèbre $\mathcal{T}'$, $\mathcal{C} \subset \mathcal{T}' \Longrightarrow \mathcal{T} \subset \mathcal{T}'$.

Proposition 30.1.12 (Existence de la tribu engendrée)
Cette σ-algèbre existe.

◁ **Éléments de preuve.**
C'est l'intersection de toutes les tribus $\mathcal{T}'$ vérifiant $\mathcal{C} \subset \mathcal{T}'$. ▷

Remarque 30.1.13
La σ-algèbre engendrée par une famille contient au moins tous les complémentaires des événements de cette famille, ainsi que toutes les unions et intersections finies ou dénombrables des événements de la famille et de leurs complémentaires.

Proposition 30.1.14 (Tribu engendrée par les singletons)
Soit Ω un univers dénombrable. La σ-algèbre engendrée par l'ensemble des événements élémentaires $\{\{\omega\}, \omega \in \Omega\}$ est $\mathcal{P}(\Omega)$.

◁ **Éléments de preuve.**
Tout sous-ensemble de Ω est intersection au plus dénombrable de ses singletons. ▷

Remarque 30.1.15
Dans le cas où Ω est dénombrable, si on veut pouvoir considérer les événements élémentaires (réalisations ponctuelles) comme des événements, la seule tribu possible sur Ω est $\mathcal{P}(\Omega)$, Ainsi, lorsque Ω est fini ou dénombrable, l'espace probabilisable que l'on considére généralement est $(\Omega, \mathcal{P}(\Omega))$.

Enfin, voici une tribu d'une grande importance dans l'étude des variables aléatoires réelles :

Définition 30.1.16 (tribu des boréliens, HP)
La *tribu borélienne* sur $\mathbb{R}$ est la tribu $\mathcal{B}^1$, souvent notée plus simplement $\mathcal{B}$, engendrée par tous les intervalles $]-\infty, a]$. Les ensembles de cette tribu sont appelés *ensembles boréliens*, ou tout simplement *boréliens*.

Proposition 30.1.17 (laissé en exercice)
On peut montrer que la tribu des boréliens contient tous les intervalles. Elle est aussi engendrée par les intervalles $]-\infty, a[$, ou encore par les $[a, b]$, ou encore par les $[a, b[$ etc.

◁ **Éléments de preuve.**
Par exemple, pour $]-\infty, a[$, montrer que ces ensembles sont dans $\mathcal{B}^1$ (les écrire comme union de $]-\infty, a_n]$). Réciproquement, les $]-\infty, a]$ sont dans la tribu engendrée par les $]-\infty, b[$ (écrire cette fois $]-\infty, a]$ comme intersection de $]-\infty, a_n[$). Les autres se démontrent sur le même principe. ▷

Définition 30.1.18 (Tribu des boréliens sur $\mathbb{R}^n$, HP)
La tribu $\mathcal{B}^n$ des boréliens de $\mathbb{R}^n$ est la tribu sur $\mathbb{R}^n$ engendrée par les pavés $I_1 \times \cdots \times I_n$, les I_k étant des intervalles (dont on peut imposer le type, à condition qu'ils soient de longueur non nulle)

II Espaces probabilisés

Les notions de ce paragraphe ne sont au programme de Sup que dans le cas où Ω est fini et où l'espace probabilisable considéré est $(\Omega, \mathcal{P}(\Omega))$. Vous pouvez vous limiter à ce cas, mais nous généralisons à un espace probabilisable $(\Omega, \mathcal{T})$ quelconque (programme de Spé).

II.1 Mesures de probabilité

Intuitivement, une probabilité mesure la possibilité qu'un événement donné se produise. Il s'agit donc d'un réel p indiquant la fréquence probable de réalisation de l'événement lorsqu'on répète un grand nombre de fois l'expérience dans des situations similaires, c'est-à-dire le rapport moyen entre le nombre de réalisations et le nombre d'expériences.
Ainsi, une probabilité est un réel positif et inférieur à 1. De plus, toutes les expériences amenant une issue de Ω, la probabilité de l'événement Ω doit être 1. Enfin, si A et B sont deux événements ne pouvant pas être réalisés simultanément, le nombre d'expériences amenant A ou B est la somme du nombre d'expériences amenant A et du nombre d'expériences amenant B (puisqu'une expérience ne peut amener à la fois A et B). De l'interprétation ci-dessus, il doit alors venir que la probabilité de l'union disjointe $A \sqcup B$ doit être la somme des probabilités de A et de B. Cela motive la définition suivante :

Définition 30.2.1 (Mesure de probabilité)
Soit $(\Omega, \mathcal{T})$ un espace probabilisable. Une mesure de probabilité sur $(\Omega, \mathcal{T})$ est une application $\mathbb{P} : \mathcal{T} \to \mathbb{R}$ telle que :

1. $\forall A \in \mathcal{T}$, $0 \leqslant \mathbb{P}(A) \leqslant 1$;
2. $\mathbb{P}(\Omega) = 1$
3. (propriété de σ-additivité) Pour tout famille dénombrable $(A_n)_{n\in\mathbb{N}}$ d'événements deux à deux incompatibles, $\sum \mathbb{P}(A_n)$ converge, et on a : $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$.

Remarque 30.2.2
Les probabilités étant positive, la convergence de $\sum \mathbb{P}(A_n)$ est absolue, et les propriétés de convergence, ainsi que la valeur de la somme sont donc indépendantes de l'ordre de sommation. Cela est cohérent avec le fait que le terme de gauche (l'union des A_i) est indépendant de l'ordre de numérotation des A_i.

Proposition 30.2.3 (Propriétés d'une mesure de probabilité)
Soit $\mathbb{P}$ une mesure de probabilités sur $(\Omega, \mathcal{T})$. Alors :

1. $\mathbb{P}(\varnothing) = 0$;
2. (Additivité) *Pour tout couple (A, B) d'événements incompatibles,* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$;

3. $\forall A \in \mathcal{T}$, $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$;
4. $\forall (A,B) \in \mathcal{T}$, $A \subset B \Longrightarrow \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$;
5. $\forall (A,B) \in \mathcal{T}$, $A \subset B \Longrightarrow \mathbb{P}(A) \leqslant \mathbb{P}(B)$;
6. $\forall (A,B) \in \mathcal{T}$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;
7. (Formule du crible de Poincaré, HP) *Plus généralement, soit* $(A_1, \dots, A_n) \in \mathcal{T}^n$. *Alors :*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{I \subset [\![1,n]\!]} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}).$$

◁ **Éléments de preuve.**
$\mathbb{P}(\varnothing)$ s'obtient en étudiant la convergence de $\sum \mathbb{P}(\varnothing)$, convergence assurée par la σ-additivité. L'additivité en découle en complétant la famille finie (A, B) en une famille dénombrable par ajout de $\varnothing$. Les points 3,4,5,6 sont évident, en décomposant le plus grand ensemble en union dijointe d'ensembles plus petits. La formule du crible s'obtient par récurrence, exactement comme dans le cas des cardinaux. ▷

Remarque 30.2.4
Dans le cas où Ω est fini (cadre du programme), la σ-additivité n'est pas pertinente, car une famille infinie $(A_n)_{n\in\mathbb{N}}$ ne peut être constituée d'éléments 2 à 2 disjoints que si seul un nombre fini d'événements A_i ne sont pas vides. Ainsi, dans le cas où Ω est fini, on peut se contenter de définir une mesure de probabilité en remplaçant la propriété de σ-additivité par la propriété d'additivité.

Le théorème suivant découle directement de la σ-additivité. Il n'a d'intérêt que dans le cas où Ω n'est pas fini (sinon, les suites considérées sont stationnaires). Pour cette raison, il sort du cadre du programme de Sup (mais pas de Spé).

Théorème 30.2.5 (Propriété de limite monotone, ou continuité monotone de la mesure $\mathbb{P}$)

1. *Soit* $(A_n)_{n\in\mathbb{N}}$ *une suite croissante d'événements (pour l'inclusion). Alors :*

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

2. *Soit* $(A_n)_{n\in\mathbb{N}}$ *une suite décroissante d'événements. Alors :*

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

◁ **Éléments de preuve.**
Écrire $\bigcup_{n\in\mathbb{N}} A_n = \bigsqcup_{n\in\mathbb{N}} (A_n \setminus A_{n-1})$, avec $A_{-1} = \varnothing$, puis σ additivité. Passer au complémentaire pour l'intersection. ▷

En définissant pour une suite croissante (A_n), $\bigcup_{n\in\mathbb{N}} A_n = \lim \uparrow A_n$ et de même pour une suite décroissante, $\bigcap_{n\in\mathbb{N}} A_n = \lim \downarrow A_n$, la propriété ci-dessus se réécrit :

$$\mathbb{P}(\lim \uparrow A_n) = \lim \mathbb{P}(A_n) \qquad \text{et} \qquad \mathbb{P}(\lim \downarrow A_n) = \lim \mathbb{P}(A_n).$$

Il s'agit donc d'une propriété de continuité séquentielle, d'où le nom donné au théorème.
Les unions et intersections ci-dessus dans le cas de suites monotones sont un cas particulier d'une construction plus générale :

Définition 30.2.6 (Limite supérieure et limite inférieure, HP)

Soit (A_n) une suite d'événements. On définit

- la limite supérieure $\limsup A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{k=n}^{\infty} A_k$: réalisation d'une infinité de A_k ;
- la limite inférieure $\liminf A_n = \bigcup_{n\in\mathbb{N}} \bigcap_{k=n}^{\infty} A_k$: réalisation de presque tous les A_k (tous sauf un nombre fini, c'est-à-dire tous à partir d'un certain rang).

Corollaire 30.2.7

Soit (A_n) une suite d'événements.

- *La probabilité qu'un nombre infini de A_k se réalisent est*
$$\mathbb{P}(\limsup A_n) = \lim \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right).$$
- *La probabilité que presque tous les A_k se réalisent est*
$$\mathbb{P}(\liminf A_n) = \lim \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k\right).$$

Remarque 30.2.8

Si (A_n) est décroissante, $\bigcup_{k=n}^{+\infty} A_k = A_n$ et $\limsup A_n = \lim \downarrow A_n$; de même si (A_n) est croissante, $\liminf A_n = \lim \uparrow A_n$. Le corollaire précédent englobe donc le théorème de continuité monotone.

Nous avons maintenant tout ce qu'il faut pour nous définir un cadre rigoureux pour nos probabilités.

Définition 30.2.9 (Espace probabilisé, ou modèle probabiliste de Kolmogorov)

Un *espace probabilisé* est un triplet $(\Omega, \mathcal{T}, \mathbb{P})$ où $(\Omega, \mathcal{T})$ est un espace probabilisable, et $\mathbb{P}$ une mesure de probabilité sur $(\Omega, \mathcal{T})$.

Voici un résultat commode pour décrire de façon rapide une mesure de probabilité.

Proposition 30.2.10 (Détermination d'une probabilité par l'image des singletons)

Soit Ω un ensemble fini ou dénombrable. Alors, une mesure de probabilité sur $(\Omega, \mathcal{P}(\Omega))$ est définie de manière unique par la donnée d'une application $p : \Omega \to [0,1]$ telle que $\sum_{\omega\in\Omega} p(\omega) = 1$ et telle que $\mathbb{P}(\{\omega\}) = p(\omega)$.

◁ **Éléments de preuve.**

Unicité car les singletons engendrent $\mathcal{P}(\Omega)$. Expliciter p ainsi obtenue en fonction des probabilités des singletons, et vérifier qu'il s'agit bien d'une mesure de probabilité. ▷

Ainsi, une mesure de probabilité sur un univers discret est entièrement déterminée par la probabilité des événements élémentaires.

II.2 Probabilités uniformes sur un univers fini

Soit Ω un univers fini. On considère l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$.
D'après la proposition 30.2.10, pour définir une mesure de probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{P}(\Omega))$, il suffit de définir $\mathbb{P}$ sur les singletons de sorte que la somme soit égale à 1. Voici un cas particulier très important :

Définition 30.2.11 (Probabilité uniforme sur un univers fini)
La *probabilité uniforme* (ou *équirépartition*) sur $(\Omega, \mathcal{P}(\Omega))$ est la probabilité définie par :

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

On retrouve alors, à partir du cadre formel qu'on s'est fixé, le résultat que vous utilisez intuitivement depuis longtemps déjà :

Théorème 30.2.12 (Formule de Laplace)
Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ l'espace probabilisé uniforme sur l'univers fini Ω. Alors, pour tout $A \in \mathcal{P}(\Omega)$:

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\textit{nombre de cas favorables}}{\textit{nombre de cas possibles}}.$$

◁ **Éléments de preuve.**
Écrire A comme union de ses singletons, plus utiliser l'additivité. ▷

II.3 Ensembles négligeables

Définition 30.2.13 (Ensembles négligeables, HP)
Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, et soit $A \in \mathcal{T}$.

1. On dit que A est *négligeable*, ou *presque-impossible*, ou *quasi-impossible* si $\mathbb{P}(A) = 0$.
2. On dit que A est *presque-certain* ou *quasi-certain* si $\mathbb{P}(A) = 1$.
3. Une propriété est vraie *presque sûrement* si l'événement « la propriété est satisfaite » est presque-certain.

Exemple 30.2.14
L'obtention d'une suite infinie de Pile dans une succession infinie de tirages d'une pièce équilibrée est un événement presque-impossible, mais pas impossible.

III Conditionnement et indépendance

III.1 Probabilités conditionnelles

On cherche maintenant à voir comment la connaissance de certaines informations sur le résultat d'une expérience modifie les probabilités. Pour cela, on définit la probabilité conditionnelle d'un événement A sachant que l'événement B est réalisé : la connaissance de la réalisation de B peut influer sur les chances que A se réalise aussi. Par exemple si on tire un dé à 6 faces équilibré, la probabilité d'obtenir 1 est $\frac{1}{6}$. Mais si on sait qu'on a obtenu une valeur impaire (événement B), on a une chance sur 3 pour que ce soit 1.

Définition 30.3.1 (Probabilités conditionnelles)

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, A et B deux événements de $\mathcal{T}$ tels que B ne soit pas presque-impossible. Alors, la *probabilité conditionnelle de A en B* (ou la *probabilité de A sachant B*) est :

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

On trouve également souvent la notation $\mathbb{P}(A \mid B)$.

Cette définition se comprend bien pour la mesure uniforme : il s'agit de la proportion des éléments de B qui satisfont A (c'est-à-dire la proportion des éléments de B qui sont dans $A \cap B$).

Proposition 30.3.2

Soit B un événement non presque-impossible. Alors $\mathbb{P}_B$ définit une mesure de probabilité sur l'espace probabilisable $(\Omega, \mathcal{T})$.

◁ **Éléments de preuve.**

Vérifications sans difficulté. ▷

Remarque 30.3.3

On utilise souvent la relation définissant les probabilités conditionnelles sous la forme suivante : $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$.
En effet, on se sert souvent des probabilités conditionnelles pour calculer les probabilités d'intersections, et non l'inverse, car les probabilités conditionnelles sont souvent plus simples à calculer. On verra un peu plus loin une généralisation de cette formule à une intersection de n événements (formule des probabilités composées, théorème 30.4.7)

III.2 Indépendance

Les probabilités conditionnelles mesurent l'impact de la réalisation d'un événement sur un autre. Si cet impact est nul (c'est-à-dire si la connaissance de la réalisation d'un événement n'influe pas sur la probabilité de l'autre), on dira que ces deux événements sont indépendants.

Ainsi, intuitivement, deux événements A et B sont indépendants si la connaissance de l'un ne modifie pas la probabilité de l'autre, donc si $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ et $\mathbb{P}(B \mid A) = \mathbb{P}(B)$. L'inconvénient de cette définition réside dans le fait qu'elle n'est valable que sous une hypothèse de non quasi-impossibilité, afin de pouvoir considérer les probabilités conditionnelles. On constate cependant que dans le cas de non quasi-impossibilité, les deux égalités sont équivalentes à l'égalité : $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Cette égalité pouvant être considérée sans hypothèse de quasi-impossibilité, c'est elle que nous prenons comme définition de l'indépendance. Elle a en outre l'avantage d'être symétrique.

Définition 30.3.4 (Événements indépendants)

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, et $(A, B) \in \mathcal{T}^2$. Les événements A et B sont *indépendants* si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

En faisant le cheminement inverse de celui précédant la définition, il vient :

Proposition 30.3.5 (Probabilités conditionnelles et indépendance)
Soit A et B deux événements indépendants tel que B ne soit pas quasi-impossible. Alors

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

On a évidemment aussi $\mathbb{P}(B) = \mathbb{P}(B \mid A)$ si A n'est pas quasi-impossible.

Définition 30.3.6 (Mutuelle indépendance pour une famille finie)
Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, et $(A_1, \dots, A_n) \in \mathcal{T}^n$. Les événements $A_1, \dots, A_n$ sont *mutuellement indépendants* si :

$$\forall J \subset [\![1, n]\!], \quad \mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

En particulier, en prenant $J = [\![1, n]\!]$:

Proposition 30.3.7 (Probabilité d'une intersection d'événements mut. indépendants)
Soit $A_1, \dots, A_n$ des événements mutuellement indépendants. Alors :

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

Avertissement 30.3.8
L'égalité ne suffit pas à avoir l'indépendance mutuelle.

Exemple 30.3.9
On fait des tirages à pile ou face
- A est réalisé si et seulement si le premier tirage est Pile.
- B est réalisé si et seulement si lors des 3 premiers tirages, il y a au plus un Pile.
- $C = B$

On a $\mathbb{P}(A \cap B \cap C) = \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$, mais les événements ne sont clairement pas mutuellement indépendants !

Proposition 30.3.10 (Stabilité de la mutuelle indépendance par restriction)
Soit $(A_i)_{i \in [\![1,n]\!]}$ une famille d'événements mutuellement indépendants. Alors toute sous-famille de $(A_i)_{i \in [\![1,n]\!]}$ est encore composée d'événements mutuellement indépendants.

En particulier, en prenant toutes les sous-familles de 2 événements, la mutuelle indépendance implique l'indépendance 2 à 2 des événements.

Avertissement 30.3.11
La réciproque est fausse : l'indépendance 2 à 2 d'une famille n'implique pas l'indépendance mutuelle.

Exemple 30.3.12

On procède à trois tirages à Pile ou Face. Soit A l'événement consistant à obtenir exactement 1 Pile lors des tirages 2 et 3, B de même lors des tirages 1 et 3, C de même lors des tirages 1 et 2. Les événements A, B et C sont 2 à 2 indépendants, mais pas mutuellement.

On verra dans le paragraphe suivant comment calculer la probabilité d'une intersection lorsque les événements ne sont pas mutuellement indépendants (formule des probabilités composées, théorème 30.4.7).

Définition 30.3.13 (Mutuelle indépendance pour une famille quelconque, HP)

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, et $(A_i)_{i\in I}$ une famille quelconque d'événements. Les événements $A_i, i \in I$ sont *mutuellement indépendants* si pour toute sous-ensemble fini J de I, les événements $A_i,\ i \in J$ sont mutuellement indépendants, donc si pour tout sous-ensemble *fini* J de I,

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j).$$

Proposition 30.3.14 (Indépendance et complémentation)

Si A et B sont indépendants, alors A et $\overline{B}$ sont indépendants.

◁ **Éléments de preuve.**
Calculer $\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A \setminus (A \cap B))$. ▷

Proposition 30.3.15 (Mutuelle indépendance et complémentation)

Si $A_1, \ldots, A_n$ sont mutuellement indépendants, il en est de même de $\overline{A_1}, A_2, \ldots, A_n$

◁ **Éléments de preuve.**
Considérer l'intersection d'une sous-famille. Seul cas à voir : $\overline{A_1}$ est dans la sous-famille. Se ramener au cas précédent en justifiant que A_1 est indépendant de l'intersection des autres termes de la sous-famille. ▷

En appliquant plusieurs fois de suite cette propriété, on peut complémenter un nombre quelconque des événements A_i. Par ailleurs, il est facile de voir que cela reste valable pour des familles quelconques.
Cette propriété pourrait se généraliser en terme de tribu, mais cela dépasse largement le cadre du programme.

IV Les trois théorèmes fondamentaux du calcul des probabilités

IV.1 Formule des probabilités totales

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé.

Définition 30.4.1 (Système quasi-complet d'événements, HP)

Soit $\mathcal{C} \subset \mathcal{T}$. On dit que $\mathcal{C}$ est un système quasi-complet d'événements si les événements de $\mathcal{C}$ sont non impossibles, deux à deux incompatibles, et si $\sum_{A\in\mathcal{C}} \mathbb{P}(A) = 1$.

Ainsi, la différence avec un système complet réside uniquement dans le fait que l'union des événements n'est égale à Ω qu'à un ensemble négligeable près (l'union est presque-certaine et non certaine).

Un système complet est bien entendu aussi quasi-complet, la réciproque étant fausse.

Théorème 30.4.2 (Formule des probabilités totales)
Soit $(A_i)_{i\in I}$ un système quasi-complet fini ou dénombrable, tels que les événements A_i ne sont pas presque-impossibles, et soit $B \in \mathcal{T}$. Alors

$$\mathbb{P}(B) = \sum_{i\in I} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i),$$

cette somme étant (absolument) convergente.

◁ **Éléments de preuve.**
Écrire $B = \left(\bigsqcup_{i\in I} B \cap A_i\right) \sqcup (B \cap C)$, où $\mathbb{P}(C) = 0$. ▷

Remarque 30.4.3
La formule des probabilités totales permet de retrouver une probabilité $\mathbb{P}(B)$ connaissant les probabilités conditionnelles $\mathbb{P}(B \mid A_i)$. Ainsi, on peut calculer $\mathbb{P}(B)$ en « distinguant suivant un certain nombre de cas » (suivant que A_i est réalisé), si ces cas partagent l'ensemble (ou presque) des cas possibles. Ainsi, cette formule est adapté au cas où le résultat d'une expérience dépend de résultats antérieurs, donc si on est amené à faire une discussion pour décrire le résultat de l'expérience.

Remarque 30.4.4
La somme intervenant dans la formule des probabilités totales étant absolument convergente (la convergence absolue découlant ici de la convergence, puisque la somme est à termes positifs), elle est correctement définie, de façon indépendante de l'ordre de sommation. Ce point est crucial ici, puisque nous ne nous sommes pas donné d'ordre sur l'ensemble I des indices.

Voici un cas particulier très important.

Corollaire 30.4.5 (Formule des probabilités totales pour le système complet $(A, \overline{A})$)
Soit A un événement non quasi-impossible et non quasi-certain. Alors, pour tout événement B,

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

En anticipant un peu sur les variables aléatoires, nous donnons un autre cas important. Vous pouvez passer ce cas à première lecture, et y revenir un peu plus tard.

Étant donné une variable aléatoire réelle discrète (donc une fonction de Ω dans $\mathbb{R}$ telle que son image $X(\Omega)$ soit au plus dénombrable) définissons les événements

$$[X = x] = X^{-1}(x) = \{\omega \in \Omega \mid X(\omega) = x\}$$

Les événements $[X = x],\ x \in X(\omega)$ forment un système complet d'événements.

Corollaire 30.4.6 (Formule des probabilités totales associée à une v.a.r.d.)
Soit X une variable aléatoire réelle discrète. Quitte à enlever quelques parts négligeables de sorte à se ramener à un système quasi-complet, on peut supposer que pour tout $x \in X(\Omega)$, $\mathbb{P}([X = x]) > 0$. On a alors, pour tout événement B :

$$\mathbb{P}(B) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x)\mathbb{P}(B \mid X = x).$$

IV.2 Formule des probabilités composées

La définition des probabilités conditionnelles permet d'exprimer la probabilité d'une intersection à l'aide de probabilités conditionnelles :

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

Cette formule est à comprendre de la façon suivante : en considérant les événements A et B comme successifs, pour obtenir $A \cap B$, il faut d'abord obtenir A, puis, A étant obtenu (ce qui donne la condition de la deuxième probabilité), il faut obtenir B. On pourrait bien sur continuer ainsi : si C est un troisième événement, consécutif à A et B, une fois réalisé $A \cap B$ (ce qui donne la condition de la troisième probabilité), il faut réaliser C. Ainsi :

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

De façon plus générale, on obtient :

Théorème 30.4.7 (Formule des probabilités composées)
Soit $n \geqslant 2$, et $(A_1, \ldots, A_n) \in \mathcal{T}^n$ tel que $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) \neq 0$ (donc $A_1 \cap \cdots \cap A_{n-1}$ n'est pas presque-impossible). Alors :

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) &= \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2)\cdots\mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}) \\ &= \mathbb{P}(A_1)\prod_{i=2}^{n}\mathbb{P}\left(A_i \;\middle|\; \bigcap_{j=1}^{i-1} A_j\right).\end{aligned}$$

◁ **Éléments de preuve.**
Récurrence sur n. ▷

Remarque 30.4.8
L'intérêt de cette formule est de donner une formule pour le calcul des probabilités d'intersections d'événements, notamment dans le cas d'une succession d'expérience. Cette formule dévoile toute son utilité lorsque les événements considérés ne sont pas mutuellement indépendants. En cas d'indépendance mutuelle, elle ne dit rien de plus que la proposition 30.3.7.

IV.3 Formules de Bayes

La troisième formule permet d'inverser des conditions. Sa version simple découle de façon directe de la définition d'une probabilité conditionnelle.

Théorème 30.4.9 (Formule de Bayes simple)

Soit A et B deux événements tels que $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$. Alors

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

◁ **Éléments de preuve.**

Calculer $\mathbb{P}(A \cap B)$ en conditionnant de deux manières. ▷

De façon plus générale, la connaissance des probabilités d'un système complet, et des probabilités d'un événement B conditionnées au système complet permet de retourner une à une les probabilités conditionnelles :

Théorème 30.4.10 (Formule de Bayes sur un système complet)

Soit $(A_i)_{i\in I}$ un système quasi-complet fini ou dénombrable, et B un événement non quasi-impossible. tel que pour tout $i \in I$, $\mathbb{P}(A_i) \neq 0$. Soit $j \in I$. Alors :

$$\mathbb{P}(A_j \mid B) = \frac{\mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}{\sum\limits_{i\in I} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}.$$

◁ **Éléments de preuve.**

Exprimer $\mathbb{P}(B)$ avec la formule des probabilités totales dans la formule précédente. ▷

Remarque 30.4.11 (Intérêt de la formule de Bayes)

Le système (A_i) représente souvent une liste de causes pouvant amener l'événement B lors d'une étape suivante de l'expérience. Il est alors généralement facile de déterminer la probabilité qu'une certaine conséquence B ait lieu, sachant que la cause A_i a eu lieu, c'est-à-dire la probabilité conditionnelle $\mathbb{P}(B \mid A_i)$ (on respecte dans ce cas l'ordre temporel). Ces données permettent, grâce à la formule de Bayes, de remonter le temps, en déterminant la probabilité qu'une certaine cause A_i ait eu lieu sachant la conséquence B. Pour cette raison, cette formule est aussi souvent appelée *formule de probabilité des causes.*

Exemple 30.4.12

Dans une population composée d'autant d'hommes que de femmes, 5 % des hommes et 0,25 % des femmes sont daltoniens. Quelle est la probabilité pour qu'une personne daltonienne choisie au hasard soit une femme ?

Voici d'autres exemples de situations de la vie courante dans lesquelles la formule de Bayes peut trouver des utilisations pratiques

Exemple 30.4.13

1. Diagnostics médicaux (retrouver la cause des symptômes)
2. Anti-SPAM

V Principes généraux du calcul des probabilités

Nous exposons ici le principe général du calcul (et de la rédaction de ce calcul) d'une probabilité. Évidemment, dans la pratique, les situations sont très variées, et on peut être amené à s'écarter légèrement de la voie tracée ci-dessous.

Méthode 30.5.1 (Comment aborder un calcul de probabilité)

1. Si on est dans une situation d'équiprobabilité, on peut se diriger vers un argument combinatoire (et utiliser la formule de Laplace). Dans les autres cas, la démarche générale est indiquée ci-dessous.
2. Définir des événements simples issus de l'expérience décrite, en n'oubliant pas la numérotation de ces événements en cas d'expérience répétée. Ces événements simples auront pour vocation de décrire ceux qui nous intéressent.

   Il ne faut pas hésiter à introduire des événements par vous-même, à donner des notations. Ce travail n'est pas nécessairement fait pour vous dans le sujet.
3. Décrire à l'aide d'une **phrase** l'événement dont on recherche la probabilité, sous forme une condition nécessaire et suffisante de réalisation :

   « L'événement A est réalisé si et seulement si ... »

   Cette condition nécessaire et suffisante de réalisation de l'événement A doit s'exprimer en fonction de la réalisation ou non des événements simples définis dans la première étape. Il s'agit donc de « décomposer » l'événement A en événements simples.
4. Traduire cette description dans le langage ensembliste (c'est-à-dire sous forme d'union, intersection ou complémentaires d'événements élémentaires). Les deux étapes sont nécessaires : la première (description verbale) fournit une explication facilement lisible, et éclaire la formule, la seconde (description ensembliste) donne de la rigueur à l'argument, et permet ensuite de s'orienter vers la bonne méthode de calcul.
5. Calcul de la probabilité $\mathbb{P}(A)$, en se servant de la décomposition ensembliste précédente, et des règles de calcul des probabilités d'unions, d'intersections ou de complémentaires. Nous rappelons l'essentiel de ces règles ci-dessous.

Ainsi, il est important de savoir s'orienter rapidement vers la bonne technique de calcul (probabilité d'une union, d'une intersection), suivant la situation rencontée. Voici les différents cas rencontrés les plus fréquemment :

Méthode 30.5.2 (Calcul de la probabilité d'un complémentaire)

Une seule formule à connaître, issue de la définition d'une mesure de probabilité :

$$\forall A \in \mathcal{T}, \quad \mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A).$$

Méthode 30.5.3 (Calcul de la probabilité d'une intersection)

Suivant les cas :

- Intersection dénombrable d'événements décroissants pour l'inclusion :

  On utilise le théorème de la limite monotone.
- Intersection d'un nombre fini d'événements mutuellement indépendants :

  On utilise la relation issue de la définition de l'indépendance mutuelle :

$$\mathbb{P}\left(A_1 \cap A_2 \cap \cdots \cap A_n\right) = \mathbb{P}(A_1)\mathbb{P}(A_2)\ldots\mathbb{P}(A_n).$$

- Intersection d'un nombre fini d'événements non nécessairement indépendants :

  On utilise la formule des probabilités composées.

  Cette formule est surtout utile lorsqu'il y a un enchainement temporel des événements (A_1 étant le premier à advenir, A_n le dernier), chaque résultat influant sur les suivants. C'est le cas par exemple dans le cadre de tirages successifs dans une urne évolutive, l'évolution se faisant différemment suivant la boule tirée à un tirage donné.

  Cette formule permet de formaliser le calcul de la probabilité associée à une branche de l'arbre des possibilités.

Méthode 30.5.4 (Calcul de la probabilité d'une union)

- Union dénombrable d'événements croissants pour l'inclusion :

  On utilise ici aussi le théorème de la limite monotone.
- Union d'un nombre fini ou dénombrable d'événements 2 à 2 incompatibles :

  On utilise l'additivité (cas fini) ou la σ-additivité (cas dénombrable) de la mesure de probabilité.
- Union d'un nombre fini d'événements mutuellement indépendants :

  On a intérêt dans ce cas à considérer l'événement complémentaire. On est alors ramené au calcul de la probabilité d'une intersection d'événements mutuellement indépendants.
- Union d'un nombre fini d'événements lorsqu'on a des informations sur les intersections :

  On utilise la formule du crible de Poincaré. Les cas $n = 2$ et $n = 3$ sont les plus utiles :
  * $n = 2 : \mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$
  * $n = 3 : \mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap A_2 \cap A_3)$

Méthode 30.5.5 (Cas d'une expérience dont l'issue dépend de résultats antérieurs)

Dans certaines situations, par exemple lorsque le mode opératoire d'une expérience (par exemple la composition d'une urne) dépend du résultat d'une expérience précédente, il peut être nécessaire de discuter suivant le résultat obtenu lors de l'expérience antérieure : autrement dit, dans ces situations, il est aisé de calculer des probabilités conditionnellement à chacun des résultats possibles de la première expérience. Il faut ensuite récupérer la probabilité globale à partir de toutes ces probabilités conditionnelle. On utilise pour cela la formule des probabilités totales.

A retenir : discussion à faire $\longrightarrow$ formule des probabilités totales

Enfin, vu qu'elles interviennent dans de nombreuses formules, nous faisons quelques remarques concernant le calcul des probabilités conditionnelles.

Méthode 30.5.6 (Calcul de probabilités conditionnelles)

Puisque la probabilité conditionnelle $\mathbb{P}_B$ sachant un événement B donné est une mesure de probabilité, tous les résultats, et toutes les règles vues ci-dessus pour le calcul des probabilités s'appliquent au calcul des probabilités conditionnelles. Lorsque ces formules font elles-même intervenir des probabilités conditionnelles, les conditions s'ajoutent à celle déjà présente (sous forme d'une intersection) :

$$(\mathbb{P}_B)_C = \mathbb{P}_{B \cap C}.$$

À titre d'exemple, voici la formule des probabilités totales pour une probabilité conditionnelle : $(A_i)_{i \in I}$ étant un système complet au plus dénombrable tel que pour tout $i \in I$ $\mathbb{P}_C(A_i) \neq 0$, on a :

$$\mathbb{P}(B \mid C) = \sum_{i \in I} \mathbb{P}(A_i \mid C)\mathbb{P}(B \mid C \cap A_i).$$

Avertissement 30.5.7

Attention, on ne peut pas définir de notion d'événement conditionnel. Ce n'est pas l'événement qui est conditionnel, mais sa réalisation. Ainsi, si vous adoptez la démarche générale de rédaction :

- ne décrivez pas l'événement conditionnel (cela n'a pas de sens), mais donnez bien une condition nécessaire et suffisante de réalisation, conditionnellement à un certain événement :

  « Sachant que B est réalisé, A est réalisé si et seulement si ... »
- Gardez-vous bien de transcrire cette phrase de manière ensembliste ; cela vous amènerait inévitablement à considérer des « événements conditionnels ». Sautez cette étape et passez directement aux probabilités. Vous pouvez transcrire votre phrase de façon ensembliste à l'intérieur des probabilités, la condition étant alors précisée non pas au niveau des ensembles, mais au niveau de la probabilité.

Avertissement 30.5.8

L'indépendance est une notion dépendant de la mesure de probabilité. Ce n'est pas parce que A et B sont indépendants (lorsqu'on dit cela, on sous-entend l'indépendance pour la mesure de probabilité totale $\mathbb{P}$) qu'ils sont indépendants pour une mesure de probabilité conditionnelle $\mathbb{P}_C$.

Exemple 30.5.9

On effectue 3 tirages à Pile ou Face indépendants. L'événement A est réalisé si et seulement si on obtient exactement 1 Pile lors des tirages 1 et 2, B de même avec les tirages 2 et 3. et C de même avec les tirages 1 et 3.
Les événements A et B sont indépendants pour $\mathbb{P}$, mais pas pour $\mathbb{P}_C$.

31 Variables aléatoires

Alea jacta est.

(Jules, un vieux pote d'Asterix)

Dans ce chapitre, sauf mention contraire, $(\Omega, \mathcal{T}, \mathbb{P})$ désigne un espace probabilisé quelconque.

Le programme se limite au cas où Ω est fini. Nous aborderons le cas un peu plus général d'une variable aléatoire discrète. La difficulté supplémentaire dans ce cadre réside souvent dans des problèmes de convergence. Ces problèmes de convergence n'ont pas lieu dans le cadre d'un univers fini, puisque toutes les sommes considérées sont alors finies. Nous rappelons pour cela le résultat fondamental suivant :

> *Si $\sum a_n$ est absolument convergente alors ni la convergence ni la valeur de la somme ne dépendent de l'ordre de sommation, ce qui permet de considérer dans ce cas une somme sur des ensembles I dénombrables non ordonnés (il suffit d'avoir convergence absolue pour une énumération possible des éléments de I pour que la somme sur I ait un sens bien défini). On dira dans ce cas que la famille $(a_i)_{i\in I}$ est sommable.*

Nous énonçons les résultats dans le contexte général, mais nous nous contenterons de donner les preuves dans le cas fini, conformément au programme, afin d'éviter les justifications un peu douloureuses de convergence et de commutation de sommes. Les justifications précises de ces points seront vues en Spé.

La situation la plus fréquente est celle d'une variable aléatoire à valeurs réelles (variable aléatoire réelle), ou à valeurs dans $\mathbb{R}^n$ (on parle de vecteur aléatoire, c'est équivalent à la donnée de n variables aléatoires réelles) mais on peut définir une variable aléatoire à valeurs dans n'importe quel ensemble muni d'une tribu.

I Variables aléatoires

I.1 Applications mesurables

Définition 31.1.1 (application mesurable)

Soit $(E, \mathcal{S})$ et $(F, \mathcal{T})$ deux espaces munis d'une tribu (on parlera d'espaces mesurables). Soit $f : E \to F$. On dit que f est mesurable si pour tout $B \in \mathcal{T}$, $f^{-1}(B) \in \mathcal{S}$.

Proposition 31.1.2 (Composition)

Soit $(E, \mathcal{S})$, $(F, \mathcal{T})$ et $(G, \mathcal{U})$ trois espaces mesurables, et $f : E \to F$ et $g : F \to G$ deux applications mesurables. La composée $g \circ f$ est encore mesurable.

Proposition 31.1.3 (Caractérisation des applications mesurables)
Avec les notations précédentes, soit $f : E \to F$, et $\mathcal{C}$ tel que $\mathcal{T} = \sigma(\mathcal{C})$. Alors f est mesurable si et seulement si pour tout $B \in \mathcal{C}$, $f^{-1}(B) \in \mathcal{S}$.

◁ **Éléments de preuve.**
Montrer que l'ensemble des B de $\mathcal{T}$ tels que $f^{-1}(B) \in \mathcal{S}$ est une tribu contenant $\mathcal{C}$. ▷

On rappelle que la tribu $\mathcal{B}^n$ des boréliens sur $\mathbb{R}^n$ est engendrée par les ensembles $I_1 \times I_2 \times \cdots \times I_n$, où les I_k sont des intervalles fermés $[a_k, b_k]$.

Lemme 31.1.4 (des éléments de $\mathcal{B}^n$)
Les produits $I_1 \times \cdots \times I_n$, où les I_k sont des intervalles quelconques, sont des éléments de $\mathcal{B}^n$.

◁ **Éléments de preuve.**
On l'avait déjà évoqué : récupérer des bornes ouvertes par intersection dénombrable, des bornes infinies par union dénombrable. ▷

Lemme 31.1.5 (D'autres systèmes générateurs de $\mathcal{B}^1$)
La tribu $\mathcal{B}^1$ sur $\mathbb{R}$ est aussi engendrée par les intervalles $]-\infty, a]$.

◁ **Éléments de preuve.**
Retrouver les $[a, b]$ par intersection, union. ▷

On montrerait sur le même principe de $\mathcal{B}^1$ est engendrée par les intervalles $]-\infty, a[$, ou par les intervalles $[a, +\infty[$, ou par les intervalles $]a, +\infty[$, ou encore par les intervalles $]a, b[$, ou encore $[a, b[$, ou $]a, b]$. Ces différents points sont laissés en exercice.
Un exemple important de fonctions mesurables est donné par les fonctions continues de $\mathbb{R}^n$ dans $\mathbb{R}^p$

Lemme 31.1.6 (Description des ouverts de $\mathbb{R}^n$)
Tout ouvert de $\mathbb{R}^n$ s'écrit comme union dénombrable de pavés ouverts $I_1 \times \cdots \times I_n$ où les I_k sont des intervalles ouverts. (non nécessairement deux à deux disjoints).

◁ **Éléments de preuve.**
Inclure tout singleton dans une boule elle-même incluse dans l'ouvert. Trouver un pavé ouvert contenant ce singleton et inclus dans cette boule. S'arranger ensuite pour en trouver un à coins rationnels. ▷

Lemme 31.1.7 (Caractérisation topologique de la continuité)
Soit $f : \mathbb{R}^n \to \mathbb{R}^p$ une application. Alors f est continue si et seulement si pour tout ouvert U de $\mathbb{R}$, $f^{-1}(U)$ est ouvert.

◁ **Éléments de preuve.**
- Sens direct : si $x \in f^{-1}(U)$, il existe un voisinage V de x tel que $f(V) \subset U$ (caractérisation topologique utilisée avec U, voisinage de $f(x)$). Alors $V \subset f^{-1}(U)$.
- Sens réciproque : Soit $x \in \mathbb{R}^n$ et W un voisinage de $f(x)$. Considérer un ouvert U tel que $x \in U \subset V$, et considérer son image réciproque.

▷

Ce lemme est bien sûr plus général : il est valide dès lors qu'on dispose d'espaces métriques (espaces munis d'une distance), la topologie (les ouverts) ainsi que la continuité étant définis à l'aide de cette métrique.

Théorème 31.1.8 (Mesurabilité des applications continues)

On munit $\mathbb{R}^n$ et $\mathbb{R}^p$ de la norme euclidienne canonique. Soit $f : \mathbb{R}^n \to \mathbb{R}^p$ une application continue. Alors si on munit $\mathbb{R}^n$ et $\mathbb{R}^p$ des tribus $\mathcal{B}^n$ et $\mathcal{B}^p$, f est mesurable.

◁ **Éléments de preuve.**

Il suffit de vérifier que l'image réciproque d'un pavé borné ouvert est un borélien, ce qui provient de la description des ouverts de $\mathbb{R}^n$. ▷

Corollaire 31.1.9 (Caractérisation par les coordonnées)

Soit $(E, \mathcal{T})$ un espace mesurable, et $f : E \to \mathbb{R}^n$ une application. On note $f_1, \dots, f_n$ les coordonnées de E. Ainsi, pour tout $x \in E$, $f(x) = (f_1(x), \dots, f_n(x))$. L'ensemble $\mathbb{R}^n$ étant muni de la tribu $\mathcal{B}^n$ et $\mathbb{R}$ de la tribu $\mathcal{B}^1$, l'application f est mesurable si et seulement si $f_1, \dots, f_n$ le sont.

◁ **Éléments de preuve.**

- Sens direct : composer par la projection qui est continue.
- Sens réciproque : L'image réciproque d'un pavé par f est une intersection d'images réciproques par les f_i.

▷

Une autre propriété importante, issue de la mesurabilité de la projection est :

Théorème 31.1.10 (Produit cartésien de boréliens)

Si $A_1, \dots, A_n$ sont des éléments de $\mathcal{B}_1$, alors $A_1 \times \cdots \times A_n$ est un élément de $\mathcal{B}_n$. Autrement dit, le produit cartésien de boréliens est un borélien.

◁ **Éléments de preuve.**

Remarquer que $A_1 \times \cdots \times A_n = \bigcap_{k=1}^{n} p_k^{-1}(A_k)$. ▷

I.2 Variables aléatoires

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, E un ensemble et $\mathcal{T}'$ une tribu sur E.

Définition 31.1.11 (Variable aléatoire)

- Une variable aléatoire à valeurs dans E est une application mesurable $X : \omega \mapsto X(\omega)$ de Ω dans $(E, \mathcal{T}')$.
- Une variable aléatoire réelle (v.a.r.) est une variable aléatoire à valeurs dans $(\mathbb{R}, \mathcal{B}^1)$
- Un vecteur aléatoire réel X est une variable aléatoire à valeurs dans $(\mathbb{R}^n, \mathcal{B}^n)$. D'après ce qui précède, la donnée d'un vecteur aléatoire équivaut à la donnée de n variables aléatoires réelles $X_1, \dots, X_n$.

Définition 31.1.12 (variable aléatoire réelle finie ou discrète)

Soit X une variable aléatoire réelle, ou un vecteur aléatoire réel. On dit que X est finie (resp. discrète) si $X(\Omega)$ est un sous-ensemble fini (resp. au plus dénombrable) de $\mathbb{R}$.

Remarque 31.1.13

- Un vecteur aléatoire réel est fini (resp. discret) si et seulement si toutes ses coordonnées sont des variables aléatoires réelles finies (resp. discrètes).
- Par définition, une variable finie est discrète.
- Dans le cadre du programme de Sup, Ω est fini et $\mathcal{T} = \mathcal{P}(\Omega)$. Dans ce cadre, toute fonction est mesurable !
- Cette remarque reste valide si Ω est dénombrable, la tribu choisie étant aussi $\mathcal{P}(\Omega)$.

Comme nous l'avons vu dans le paragraphe précédent, la mesurabilité peut se justifier en se restreignant à l'étude des images réciproques d'un système générateur de $\mathcal{T}'$. En appliquant ce résultat à la tribu $\mathcal{B}^1$ des boréliens, on obtient :

Proposition 31.1.14 (Caractérisation des variables aléatoires réelles)

Soit $X : (\Omega, \mathcal{T}) \to \mathbb{R}$ une application. X est une variable aléatoire si et seulement si pour tout $a \in \mathbb{R}$, $X^{-1}(]-\infty, a]) \in \mathcal{T}$.

On admet parfois la définition suivante, un peu plus large, et utile dans certaines expériences infinies :

Définition 31.1.15 (Variable aléatoire, notion étendue)

Une *variable aléatoire* X sur Ω à valeurs dans E est une application mesurable définie sur $\Omega' \in \mathcal{T}$ telle que $\mathbb{P}(\Omega') = 1$, donc définie presque partout.

Exemple 31.1.16

On lance une infinité de fois une pièce, et X correspond au rang de lancer du premier Pile obtenu. En adoptant la convention précédente, X est une v.a.r., prenant les valeurs $X(\Omega) = \mathbb{N}^*$. Remarquez que l'événement « n'obtenir que des faces », qui est quasi-impossible, mais pas impossible, n'admet pas d'image par X.

Remarque 31.1.17

Cette définition étendue se ramène à la définition générale :

- soit en considérant X comme une variable aléatoire sur Ω', muni de la tribu $\{A \cap \Omega', A \in \mathcal{T}\} = \{A \in \mathcal{T} \mid A \subset \Omega'\}$,
- soit en considérant X comme une variable aléatoire à valeurs dans $(\overline{\mathbb{R}}, \overline{\mathcal{B}^1})$, où $\overline{\mathcal{B}^1}$ est la tribu $\sigma(\mathcal{B}^1, \{-\infty\}, \{+\infty\})$, et où les éléments de $\overline{\Omega'}$ sont envoyés par X sur $+\infty$.

Notation 31.1.18 (Événements liés à une variable aléatoire)

Soit X une variable aléatoire à valeurs dans $(E, \mathcal{T}')$.

- Soit $A \in \mathcal{T}'$ (en particulier $A \subset E$). L'événement $X^{-1}(A) \in \mathcal{T}$ est noté $[X \in A]$ ou $\{X \in A\}$ ou encore $(X \in A)$
- Soit $x \in A$. Si $\{x\} \in \mathcal{T}'$, l'événement $X^{-1}(\{x\}) = [X \in \{x\}]$ sera simplement noté $[X = x]$, ou $(X = x)$, ou $\{X = x\}$.

Remarque 31.1.19

Si $\mathcal{T} = \mathcal{P}(\Omega)$ (cadre du programme), on peut toujours considérer sur E la tribu $\mathcal{P}(E)$. Dans ce cas, l'événement $[X \in A]$ est bien défini pour tout $A \subset E$.

Ainsi, l'ensemble des événements qu'on peut décrire à l'aide de X sont tous les événements $[X \in A]$, c'est-à-dire $X^{-1}(A)$, pour $A \in \mathcal{T}$. Cet ensemble admet naturellement une structure de tribu :

Proposition/Définition 31.1.20 (Tribu associée à une v.a.)

Soit $X : (\Omega, \mathcal{T}) \to (E, \mathcal{T}')$ une variable aléatoire. L'ensemble $\{X^{-1}(A), A \in \mathcal{T}'\} \subset \mathcal{T}$ est une tribu, notée $\mathcal{T}_X$, appelée tribu associée à la variable X.

Intuitivement, $\mathcal{T}_X$ est la plus petite sous-tribu de $\mathcal{T}$ rendant X mesurable.
Dans le cas d'une variable à valeurs réels, on note de la façon suivante les événements associés aux images réciproques des intervalles :

Notation 31.1.21 (Événements liés à une v.a.r.)

Dans le cas d'une variable aléatoire réelle, $E = \mathbb{R}$, muni de la tribu des boréliens. On utilise les notations suivantes pour les images réciproques des différents intervalles possibles (les intervalles appartiennent tous à la tribu des boréliens) :

- L'événement $\{\omega \mid X(\omega) < a\}$ est noté $[X < a]$ ou $(X < a)$ ou... ;
- L'événement $\{\omega \mid X(\omega) \leqslant a\}$ est noté $[X \leqslant a]$ ou... ;
- L'événement $\{\omega \mid a < X(\omega) < b\}$ est noté $[a < X < b]$,
- L'événement $\{\omega \mid a < X(\omega) \leqslant b\}$ est noté $[a < X \leqslant b]$,
- etc.

Par ailleurs, les singletons étant dans la tribu des boréliens, on peut définir pour tout $x \in \mathbb{R}$ l'événement $[X = x] = X^{-1}(\{x\})$, qui peut éventuellement être vide.

I.3 Loi d'une variable aléatoire

Définition 31.1.22 (Loi d'une variable aléatoire)

Soit X une variable aléatoire (quelconque) sur $(\Omega, \mathcal{T}, \mathbb{P})$, à valeurs dans $(E, \mathcal{T}')$. La loi de probabilité de X est la fonction $\mathbb{P}_X$ définie de $\mathcal{T}'$ dans $[0, 1]$ par :

$$\forall A \in \mathcal{T}', \quad \mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)).$$

Proposition 31.1.23 (La mesure $\mathbb{P}_X$)

La fonction $\mathbb{P}_X$ est une mesure de probabilité sur l'espace probabilisable $(E, \mathcal{T}')$.

◁ **Éléments de preuve.**

Vérification simple, en se servant des propriétés de l'image réciproque. ▷

Dans le cas d'une variable réelle (donc à valeurs dans $\mathbb{R}$ muni de la tribu borélienne), il n'est pas nécessaire de déterminer $\mathbb{P}_X$ sur tous les éléments de $\mathcal{T}$. Des techniques fines sur les espaces mesurables permettent en effet de montrer qu'une mesure de probabilité est entièrement déterminée par ses valeurs sur un π-système $\mathcal{C}$ engendrant la tribu $\mathcal{T}'$, un π-système étant un sous-ensemble de $\mathcal{T}'$ stable par intersections finies. Ce résultat est une conséquence du « lemme de classe monotone ». On en déduit :

Proposition 31.1.24 (Détermination de la loi d'une v.a.r., admis)

Soit X une variable aléatoire réelle. Alors la loi $\mathbb{P}_X$ de E est entièrement déterminée par les probabilités $\mathbb{P}(X \leqslant x)$, $x \in \mathbb{R}$.

Définition 31.1.25 (Fonction de répartition)

Soit $X : \Omega \to \mathbb{R}$ une v.a.r.. La fonction de répartition de X est la fonction, notée généralement F_X, définie pour tout $x \in \mathbb{R}$ par $F_X(x) = \mathbb{P}(X \leqslant x)$.

Ainsi, d'après ce qui précède, la fonction de répartition détermine entièrement la loi $\mathbb{P}_X$.

Théorème 31.1.26 (Propriétés caractéristiques des fonctions de répartition)

Soit X une v.a.r., et F_X sa fonction de répartition. Alors :

1. *F_X est une fonction croissante sur $\mathbb{R}$;*
2. *F_X est continue à droite en tout point de $\mathbb{R}$;*
3. *F_X admet des limites en $\pm\infty$, et $\lim\limits_{x\to-\infty} F_X(x) = 0$, $\lim\limits_{x\to+\infty} F_X(x) = 1$.*

◁ **Éléments de preuve.**

1. C'est la croissance de la probabilité (pour l'inclusion)
2. La croissance donne déjà l'existence de la limite à droite. Le théorème de limite monotone montre que $F_X(a) = \lim\limits_{n\to+\infty} F_X\left(a + \dfrac{1}{n}\right)$.
3. Considérer une intersection d'ensembles de la forme $]-\infty, a]$ donnant $\varnothing$, une union donnant $\mathbb{R}$.

▷

Réciproquement, on peut montrer (on l'admettra) que ces propriétés caractérisent les fonctions de répartition, c'est-à-dire que toute fonction vérifiant cela peut être vue comme la fonction de répartition d'une certaine variable aléatoire réelle.

Proposition 31.1.27 (Expression de $\mathbb{P}(X \in I)$ à l'aide de la fonction de répartition)

Soit X une v.a.r., et $a \leqslant b$ dans $\overline{\mathbb{R}}$. On a :

- $\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a)$
- $\mathbb{P}(a < X < b) = \lim\limits_{y\to b^-} F_X(y) - F_X(a)$, $a < b$
- $\mathbb{P}(a \leqslant X \leqslant b) = F_X(b) - \lim\limits_{x\to a^-} F_X(x)$
- $\mathbb{P}(a \leqslant X < b) = \lim\limits_{y\to b^-} F_X(y) - \lim\limits_{x\to a^-} F_X(x)$.

◁ **Éléments de preuve.**

Même principe en décrivant tous ces ensembles comme unions, intersections, complémentations d'ensembles de la forme $]-\infty, c]$. ▷

En particulier, on a, pour tout $x \in \mathbb{R}$:

$$\mathbb{P}(X = x) = F(x) - \lim_{y\to x^-} F(y).$$

Les probabilités ponctuelles correspondent donc aux sauts de discontinuité de F_X.

I.4 La variable $f(X)$

La propriété de composition des fonctions mesurables amène directement :

Proposition/Définition 31.1.28 (Image d'une variable aléatoire par une fonction)

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire de Ω sur $(E, \mathcal{T}_E)$. Soit E' un sous-ensemble de E contenant $X(\Omega)$, et f une application de E' dans un ensemble F, muni d'une tribu $\mathcal{T}_F$. Si f est mesurable, $f \circ X$ est une variable aléatoire à valeurs dans $(F, \mathcal{T}_F)$, et est généralement notée $f(X)$. Plus explicitement :

$$\forall \omega \in \Omega, \quad f(X)(\omega) = f(X(\omega)).$$

Remarque 31.1.29

- Si $\mathcal{T} = \mathcal{P}(\Omega)$, toute fonction $\Omega \to E$ étant mesurable, on peut élargir cette définition au cas où f n'est pas nécessairement mesurable, puisque toute application $X' : \Omega \to F$ est mesurable.
- Si $X : \Omega \to \mathbb{R}^n$ est un vecteur aléatoire réel, et $f : \mathbb{R}^n \to \mathbb{R}^p$ une application continue, alors $f(X)$ est une variable aléatoire.

Corollaire 31.1.30 (Somme de variables aléatoires réelles)

Si $X_1, \dots, X_n$ sont des variables aléatoires réelles, alors $X_1 + \cdots + X_n$ est une variable aléatoire réelle.

◁ **Éléments de preuve.**

La somme est continue (somme des projections) ▷

Proposition 31.1.31 (Tribu associée à $f(X)$)

Avec les notations et conditions de la proposition/définition ci-dessus, $\mathcal{T}_{f(X)} \subset \mathcal{T}_X$. Plus précisément,

$$\mathcal{T}_{f(X)} = \{[X \in f^{-1}(A)],\ A \in \mathcal{T}_F\}.$$

◁ **Éléments de preuve.**

C'est juste une réexpression de $[f(X) \in A]$. ▷

Proposition 31.1.32 (Loi de $f(X)$)

Soit X une variable aléatoire sur $(\Omega, \mathcal{T}, \mathbb{P})$, à valeurs dans $(E, \mathcal{T}_E)$ et $f : (E, \mathcal{T}_E) \to (F, \mathcal{T}_F)$ mesurable. Soit $\widehat{f^{-1}}$ la fonction image réciproque, de $\mathcal{P}(F)$ dans $\mathcal{P}(E)$, se restreignant en une fonction de $\mathcal{T}_F$ dans $\mathcal{T}_E$. Alors

$$\mathbb{P}_{f(X)} = \mathbb{P}_X \circ \widehat{f^{-1}}.$$

Autrement dit, pour tout $A \in \mathcal{T}_F$,

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

◁ **Éléments de preuve.**

Comme ci-dessous, cela découle de $[X \in f^{-1}(A)] = [f(X) \in A]$. ▷

I.5 Variables aléatoires discrètes

Les définitions diffèrent un peu suivant les contextes et les besoins. La définition suivante est la moins contraignante, et est amplement suffisante pour toute la théorie.

Définition 31.1.33 (Variable aléatoire discrète)

- Une *variable aléatoire discrète* X est une variable aléatoire telle que $X(\Omega)$ soit un ensemble fini ou dénombrable.
- Si X est une variable aléatoire réelle, on abrègera « variable aléatoire réelle discrète » en v.a.r.d.

D'après les règles de cardinalité concernant les ensembles finis ou dénombrables, on obtient immédiatement :

Proposition 31.1.34 (Algèbre des v.a.r.d.)

Le sous-ensemble de $\mathbb{R}^\Omega$ constitué des variables aléatoires réelles discrètes est une sous-algèbre de $\mathbb{R}^\Omega$. Ainsi, si X et Y deux v.a.r.d. sur Ω, λ un réel quelconque, les v.a.r. définies de la manière suivante sont encore discrètes :

1. $X+Y \ : \omega \mapsto X(\omega)+Y(\omega)$;
2. $\lambda X \ : \ \omega \mapsto \lambda X(\omega)$;
3. $XY \ : \ \omega \mapsto X(\omega)Y(\omega)$;

◁ **Éléments de preuve.**
Pour l'union, $X+Y(\Omega) \subset \bigcup_{x\in X(\Omega)} x+Y(\Omega)$. De même pour le produit. ▷

Proposition 31.1.35 (Composition d'une variable aléatoire discrète)

Soit X une variable aléatoire discrète sur $(\Omega,\mathcal{T},\mathbb{P})$, à valeurs dans $(E,\mathcal{S})$, et et f une application mesurable de $X(\Omega)$ dans un ensemble F muni d'une tribu $\mathcal{S}'$. Alors $f(X)$ est une variable aléatoire discrète.

◁ **Éléments de preuve.**
$f : X(\Omega) \to f(X)(\Omega)$ est une surjection, et $X(\Omega)$ est au plus dénombrable. ▷

Proposition 31.1.36 (Système complet associé à une variable aléatoire)

Soit X une variable (ou un vecteur) aléatoire réelle discrète sur $(\Omega,\mathcal{T})$. Alors $\{[X=x],\ x\in X(\Omega)\}$ est un système complet d'événements.

◁ **Éléments de preuve.**
C'est la partition associée à une surjection. Le seul point à justifier est que ce sont bien des éléments de la tribu. Cela provient du fait que les singletons sont des boréliens. ▷

De façon plus générale, ceci reste valable dès lors que X est à valeurs dans $(E,\mathcal{T}')$ et que pour tout $x\in X(\Omega)$, $\{x\}\in\mathcal{T}'$.

Remarque 31.1.37

En utilisant la définition étendue d'une variable aléatoire, il ne s'agit que d'un système quasi-complet, mais cela permet tout de même d'utiliser les formules classiques.

I.6 Loi de probabilité d'une variable aléatoire discrète

On suppose dans ce paragraphe que la tribu sur E contient tous les singletons de E, ainsi que la triibu sur F lorsqu'on considère $f(X)$, où $f : E \to F$.

Proposition 31.1.38

Soit X une variable aléatoire discrète sur $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ à valeurs dans E. Alors la famille $(\mathbb{P}(X = x))_{x \in X(\Omega)}$ est sommable, et :

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

◁ **Éléments de preuve.**
C'est la σ-additivité utilisée sur le SCE précédent. ▷

La sommabilité (issue de la convergence absolue, ici automatique en cas de convergence, la série étant à termes positifs) assure que cette somme est bien définie, indépendamment de l'ordre de sommation.

Proposition 31.1.39 (Détermination de la loi d'une v.a. discrète)

La loi d'une v.a. discrète est déterminée par les probabilités ponctuelles $\mathbb{P}(X = x)$ pour $x \in X(\Omega)$. Plus précisément, pour tout $A \subset E$,

$$\mathbb{P}(X \in A) = \sum_{x \in A \cap X(\Omega)} \mathbb{P}(X = x),$$

cette somme étant bien définie.
Plus précisément, une famille $(p_x)_{x \in Y}$ définit la loi de probabilité d'une certaine variable X (i.e. il existe une variable aléatoire X telle que $X(\Omega) \subset Y$, $p_x = 0$ si $x \in Y \setminus X(\Omega)$ et $\mathbb{P}(X = x) = p_x$ si $x \in X(\Omega)$) si et seulement si $\sum_{x \in Y} p_x = 1$.

◁ **Éléments de preuve.**
Écrire A comme union de singletons. ▷

La réciproque de ce dernier point est admise.

Proposition 31.1.40 (Loi de $f(X)$)

Soit X une variable aléatoire discrète sur $(\Omega, \mathcal{T}, \mathbb{P})$, à valeurs dans E. La loi de $f(X)$ est alors entièrement déterminée par les probabilités ponctuelles :

$$\forall x \in f(X(\Omega)), \quad \mathbb{P}(f(X) = x) = \sum_{\substack{y | f(y) = x \\ y \in X(\Omega)}} \mathbb{P}(X = y).$$

◁ **Éléments de preuve.**
En effet $[f(X) = x] = [X \in f^{-1}(x)]$. Appliquer alors le résultat précédent. ▷

Exemples 31.1.41 (Loi d'une somme, loi du minimum)

Voici deux exemples particulièrement important, à bien connaître, dans le cas où X et Y sont des variables aléatoires réelles discrètes.

1. $\forall z \in \mathbb{R}, \quad \mathbb{P}(X + Y = z) = \sum_{\substack{(x,y) \in X(\Omega) \times Y(\Omega) \\ x+y=z}} \mathbb{P}(X = x, Y = y).$

2. $\forall z \in X(\Omega) \cup Y(\Omega),\ \mathbb{P}(\min(X,Y)=z) = \sum_{\substack{(x,y)\in X(\Omega)\times Y(\Omega)\\ y\geqslant z}} \mathbb{P}(X=z,Y=y) + \sum_{\substack{(x,y)\in X(\Omega)\times Y(\Omega)\\ x>z}} \mathbb{P}(X=x,Y=z).$

   Exemple : tirage de deux dés.

I.7 Loi d'un vecteur aléatoire

Soit $(X_1,\dots,X_n)$ un vecteur aléatoire, à valeurs dans $(\mathbb{R}^n,\mathcal{B}_n)$, ou plus généralement, dans un produit cartésien d'espaces mesurés. Pour commencer, introduisons une notation simplifiée pour désigner les événements associés au vecteur $(X_1,\dots,X_n)$ (c'est-à-dire les éléments de $\mathcal{T}_{(X_1,\dots,X_n)}$.

Notation 31.1.42 ($\mathbb{P}(X_1 \in A_1,\dots,X_n \in A_n)$)

Pour alléger les notations, on note $\mathbb{P}(X_1\in A_1,\dots,X_n\in A_n)$ au lieu de $\mathbb{P}([X_1\in A_1]\cap\dots\cap[X_n\in A])$. En particulier, on utilisera la notation $\mathbb{P}(X_1=x_1,\dots,X_n=x_n)$.

Définition 31.1.43 (Loi conjointe, lois marginales)

Soit $V=(X_1,\dots,X_n)$ un vecteur aléatoire, à valeurs dans $\mathbb{R}^n$.

- On appelle loi conjointe de V la loi de $(X_1,\dots,X_n)$. Il s'agit donc d'une mesure de probabilité $\mathbb{P}_{(X_1,\dots,X_n)}$ définie sur l'espace probabilisable $(\mathbb{R}^n,\mathcal{B}^n)$
- On appelle k-ième loi marginale la loi de X_k.

De façon évidente :

Proposition 31.1.44 (Expression d'une loi marginale)

Soit A un borélien de $\mathbb{R}$

1. *L'événement $(X_k\in A)$ est égal à $[X_1\in\mathbb{R},\dots,X_k\in A,\dots,X_n\in\mathbb{R}]$.*
2. *La k-ième loi marginale s'obtient de la loi conjointe par l'expression :*

$$\mathbb{P}(X_k\in A)=\mathbb{P}(X_1\in\mathbb{R},\cdots,X_k\in A,\cdots,X_n\in\mathbb{R})=\mathbb{P}_V(\mathbb{R}\times\cdots\times A\times\cdots\times\mathbb{R}).$$

Remarque 31.1.45

Ce résultat reste vrai dans un cadre plus général où V est à valeurs dans un produit cartésien $E_1\times\cdots\times E_n$, en remplaçant $X_i\in\mathbb{R}$ par $X_i\in E_i$, et $A\in\mathbb{R}$ par $A\in\mathcal{T}_i$, où $\mathcal{T}_i$ est la tribu sur E_i.

Corollaire 31.1.46

Pour tout $k\in[\![1,n]\!]$, $\mathcal{T}_{X_k}\subset\mathcal{T}_{(X_1,\dots,X_n)}$.

◁ **Éléments de preuve.**

En effet, si $A\in\mathcal{B}_1$, $\mathbb{R}\times\cdots\times A\times\cdots\times\mathbb{R}\in\mathcal{B}_n$. L'image réciproque du premier par X_k est l'image réciproque du second par V. ▷

Proposition 31.1.47 (Détermination de la loi de V dans le cas discret)

La donnée de la loi conjointe du vecteur aléatoire réel discret V équivaut à la donnée de l'application

$$\varphi_V : X_1(\Omega)\times\cdots\times X_n(\Omega)\to\mathbb{R}$$

qui à $(x_1,\dots,x_n)$ associe $\varphi_V(x_1,\dots,x_n)=\mathbb{P}(X_1=x_1,\dots,X_n=x_n)$.

◁ **Éléments de preuve.**
On a alors, pour tout $A \in \mathcal{B}^n$,

$$\mathbb{P}(V \in A) = \sum_{(x_1,\dots,x_n)\in A\cap V(\Omega)} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n).$$

▷

Proposition 31.1.48
On a : $\displaystyle\sum_{(x_1,\dots,x_n)\in X_1(\Omega)\times\cdots\times X_n(\Omega)} \varphi_V(x_1,\dots,x_n) = 1.$

◁ **Éléments de preuve.**
On peut d'abord réduire la somme à $V(\Omega)$ (les autres termes étant nuls). D'après la propriété précédente, il s'agit alors de $\mathbb{P}(V \in \mathbb{R}^n)$. ▷

Proposition 31.1.49 (Expression des lois marginales par la loi conjointe)
Les lois maginales se déduisent de la loi conjointe : pour tout $x_k \in X_k(\Omega)$ *:*

$$\mathbb{P}(X_k = x_k) = \sum_{(x_1,\dots,x_{k-1},x_{k+1},\dots,x_n)\in X_1(\Omega)\times\cdots\times X_{k-1}(\Omega)\times X_{k+1}(\Omega)\times\cdots\times X_n(\Omega)} \varphi_V(x_1,\dots,x_n)$$

◁ **Éléments de preuve.**
Toujours d'après ce qui précède, cette somme est $\mathbb{P}(V \in \mathbb{R}\times\cdots\times\mathbb{R}\times\{x_k\}\times\mathbb{R}\times\cdots\times\mathbb{R})$. ▷

Par exemple, pour un couple (X, Y),

$$\mathbb{P}(X = x) = \sum_{y\in Y(\Omega)} \mathbb{P}(X = x, Y = y).$$

Avertissement 31.1.50
Les lois marginales sont déterminées par la loi conjointe, mais la réciproque est fausse : les lois marginales ne déterminent pas la loi conjointe !

Exemple 31.1.51
Soit une urne contenant 1 boule blanche, 1 boule noire. On effectue un tirage, avec équiprobabilité. On note X_1, Y_1 et X_2 les trois variables (égales l'une à l'autre) égales à 1 si on tire la boule noire, et 0 sinon. On note Y_2 la variable égale à 1 si on tire la boule blanche, et 0 sinon. Alors, les lois marginales de (X_1, Y_1) et de (X_2, Y_2) sont les mêmes, pourtant les lois conjointes sont distinctes.

La loi conjointe de (X, Y) est souvent plus facile à déterminer que les lois de X et Y. On se sert donc du calcul de cette loi conjointe pour déterminer ensuite les lois de X et Y.

Exemple 31.1.52
On lance deux dés équilibrés, on note X le maximum, et Y le minimum. Déterminer la loi conjointe de (X, Y), puis les lois marginales.

II Espérance mathématique

Dans le cas de variables aléatoires réelles, ou plus généralement de variables à valeurs dans une $\mathbb{R}$-algèbre, on peut considérer des sommes, produits et moyennes de valeurs de X. Ainsi, on peut envisager de définir une valeur moyenne prise par X. On peut par exemple calculer une moyenne empirique, en répétant un grand nombre de fois l'expérience associée à X, et en faisant la moyenne sur ces expériences des valeurs obtenues pour X. On peut aussi définir une moyenne théorique (c'est l'espérance). Des propriétés de convergence (qui ne sont pas au programme de cette année) permettent de justifier que la moyenne empirique tend vers la moyenne théorique (l'espérance) lorsque le nombre d'expériences tend vers $+\infty$. C'est la loi faible des grands nombres.
Nous étudions dans ce paragraphe ces notions de moyennes, et les mesures d'écarts associés.

II.1 Espérance d'une variable aléatoire réelle discrète

L'espérance mathématique correspond à la moyenne théorique.
Par exemple, si on dispose d'un dé à 6 faces, dont 5 faces sont notées 1 et une face est numérotée 2, le 2 sortira en moyenne 1 fois sur 6. et le 1 apparaîtra 5 fois sur 6. Ainsi, sur 6 tirage, on peut « espérer » avoir en moyenne 5 fois 1 et 1 fois 2. La moyenne de ces 6 tirages est alors

$$\frac{1}{6}(5\times 1+1\times 2)=1\times\mathbb{P}(X=1)+2\times\mathbb{P}(X=2),$$

où X désigne le résultat d'un tirage.
Cela motive la définition suivante de la moyenne théorique, appelée plus généralement espérance mathématique.

Définition 31.2.1 (espérance mathématique)
Soit X une v.a.r.d.. On dit que X admet une espérance si la famille $(x\mathbb{P}(X=x))_{x\in X(\Omega)}$ est sommable, et dans ce cas :

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

Remarque 31.2.2
L'espérance peut ne pas exister :

- série divergente : $X(\Omega)=\mathbb{N}^*$ et $\forall n\in\mathbb{N}^*$, $\mathbb{P}(X=n)=\dfrac{6}{\pi^2n^2}$.
- série semi-convergente : $X(\Omega)=\{(-1)^n n,\ n\in\mathbb{N}^*\}$, et : $\forall n\in\mathbb{N}^*,\ \ \mathbb{P}(X=(-1)^n n)=\dfrac{6}{\pi^2n^2}$.

En effet, la semi-convergence n'est pas suffisante pour définir correctement l'espérance, car, $X(\Omega)$ n'étant pas muni d'un ordre naturel, il faut s'assurer de l'indépendance du résultat par rapport à l'ordre de sommation.
Dans le premier cas cependant, on peut donner un sens à l'espérance dans $\overline{\mathbb{R}}$, la série définissant l'espérance étant alors de somme infinie, quel que soit l'ordre de sommation choisi.

Définition 31.2.3 (Espérance infinie)
Soit X une variable aléatoire **positive**, telle que $\sum_{x\in X(\omega)} x\mathbb{P}(X=x)$ diverge. On pourra alors dire que $\mathbb{E}(X)=+\infty$

Ainsi, lorsque X est positive, l'espérance existe toujours dans $\overline{\mathbb{R}}$, et dire qu'elle existe dans $\mathbb{R}$ équivaut à dire que $\mathbb{E}(X)<+\infty$.

Remarque 31.2.4

Dans le cadre du programme de Sup, seul le cas de variables aléatoires finies (i.e. $X(\Omega)$ fini) est à considérer. Dans ce cas, la somme définissant l'espérance est finie, et il n'y a pas de convergence à justifier. Nous résumons cela dans la proposition suivante.

Proposition 31.2.5 (Espérance d'une variable finie)

Toute variable finie admet une espérance.

◁ **Éléments de preuve.**

La somme définissant $\mathbb{E}(X)$ est alors finie ! ▷

Plus généralement :

Proposition 31.2.6 (Espérance d'une variable bornée)

Soit X une v.a.r.d. bornée. Alors X admet une espérance.

◁ **Éléments de preuve.**

Par majoration, $|x\mathbb{P}(X=x)| \leqslant M\mathbb{P}(X=x)$, et cette famille est sommable. ▷

Du fait même de l'interprétation en terme de moyenne pondérée, on obtient :

Proposition 31.2.7 (Réexpression de l'espérance)

Soit X une variable aléatoire réelle discrète sur $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, Ω étant dénombrable. Alors X admet une espérance si et seulement la famille $(\mathbb{P}(\{\omega\})X(\omega))_{\omega\in\Omega}$ est sommable, et dans ce cas :

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

◁ **Éléments de preuve.**

Groupements de terme (propriété d'associativité pour les familles sommables) sur les $[X=x]$ pour passer de l'une à l'autre. ▷

II.2 Théorème de transfert et propriétés de l'espérance

Théorème 31.2.8 (Théorème de transfert)

Soit X un vecteur aléatoire réel discret à valeurs dans $E \subset \mathbb{R}^n$, et $f : E \to \mathbb{R}$ une fonction mesurable, alors $f(X)$ admet une espérance si et seulement si la famille $(f(x)\mathbb{P}(X=x))_{x\in X(\Omega)}$ est sommable et dans ce cas :

$$\mathbb{E}(f(X)) = \sum_{x\in X(\Omega)} f(x)\mathbb{P}(X=x).$$

◁ **Éléments de preuve.**

Groupement de terme (propriété d'associativité) sur les $f^{-1}(\{y\})$ pour passer de l'une à l'autre. ▷

Ainsi, l'espérance de $f(X)$ est entièrement déterminée par la loi de X et la fonction f.

Ce théorème est fréquemment utilisé pour $n = 1$ (X est une variable aléatoire réelle discrète). Par exemple, en cas de sommabilité, on peut calculer les moments :

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

On peut aussi remarquer que la condition de convergence absolue imposée dans la définition de $\mathbb{E}(X)$ et le théorème de transfert nous donnnent immédiatement la propriété suivante :

Proposition 31.2.9 (Condition d'existence de $\mathbb{E}(X)$)

X admet une espérance si et seulement $|X|$ admet une espérance, donc si et seulement si $\mathbb{E}(|X|) < +\infty$

Mais on peut aussi utiliser le théorème de transfert pour des variables vectorielles. Par exemple, si (X, Y) est un couple de vecteurs aléatoires, et $f : \mathbb{R}^2 \to \mathbb{R}$ une fonction mesurable, la formule se traduit par :

$$\mathbb{E}(f(X,Y)) = \sum_{(x,y) \in V(\Omega)} f(x,y)\mathbb{P}(V = (x,y)) = \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} f(x,y)\mathbb{P}(X = x, Y = y),$$

où $V = (X, Y)$.
Dans ce cadre, une application importante du théorème de transfert est la suivante :

Théorème 31.2.10 (Linéarité de l'espérance)

Soit X et Y deux variables aléatoires réelles discrètes sur $(\Omega, \mathcal{T}, \mathbb{P})$ et $\lambda \in \mathbb{R}$. Alors, si X et Y admettent une espérance, il en est de même de $\lambda X + Y$, et

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

Autrement dit, l'ensemble des v.a.r.d. admettant une espérance est un espace vectoriel, et $\mathbb{E}$ est une forme linéaire sur cet espace.

◁ **Éléments de preuve.**

Théorème de transfert, appliqué à $f : (x, y) \mapsto \lambda x + y$, puis manipulations simples de somme et expression des lois marginales en fonction de la loi conjointe. ▷

Proposition 31.2.11 (Positivité et croissance de l'espérance)

Soit X et Y deux variables aléatoires réelles discrètes admettant une espérance

- *Si $X \geqslant 0$, (c'est-à-dire pour tout $\omega \in \Omega$, $X(\omega) \geqslant 0$), alors $\mathbb{E}(X) \geqslant 0$, avec égalité ssi $X = 0$ presque sûrement.*
- *Si $X \geqslant Y$ (c'est-à-dire, pour tout $\omega \in \Omega$, $X(\omega) \geqslant Y(\omega)$, alors $\mathbb{E}(X) \geqslant \mathbb{E}(Y)$.*

◁ **Éléments de preuve.**

Le premier point est évident, le cas d'égalité nécessitant $\mathbb{P}(X = x) = 0$ dès que $x > 0$. Le deuxième s'en déduit par linéarité de l'espérance. ▷

Remarquons qu'encore une fois, le théorème de transfert est valide dans $\overline{\mathbb{R}}$ avec des hypothèses de positivité supplémentaire, puisque si les séries considérées sont positives et divergentes, alors leur somme vaut $+\infty$. Ainsi, si f est une fonction à valeur dans $\mathbb{R}_+$, $f(X)$ admet toujours une espérance dans $\overline{\mathbb{R}}$, égale à $\sum_{x \in X(\Omega)} f(x)\mathbb{P}(X = x)$.

Corollaire 31.2.12 (Linéarité de l'espérance dans $\overline{\mathbb{R}}$)

Soit X et Y deux v.a.r. positives, et $\lambda = \mathbb{R}_+$. Alors, dans $\overline{\mathbb{R}}$, on dispose de l'égalité

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

◁ **Éléments de preuve.**

Évacuer le cas $\lambda = 0$. Seul cas à étudier : $\mathbb{E}(X)$ ou $\mathbb{E}(Y)$ infini. Par comparaison à partir du théorème de transfert, on obtient la divergence de la somme donnant $\mathbb{E}(\lambda X + Y)$. ▷

On en déduit en particulier la propriété suivante :

Corollaire 31.2.13 (existence de $\mathbb{E}$ par majoration)

Soit X et Y deux v.a.r. Si $0 \leqslant |X| \leqslant |Y|$ et si Y admet une espérance (dans $\mathbb{R}$), alors X aussi.

◁ **Éléments de preuve.**

Utiliser l'égalité précédente dans $\overline{\mathbb{R}}$, avec l'égalité $|X| + Z = |Y|$, où $Z \geqslant 0$. ▷

II.3 Théorème de l'espérance totale

Nous donnons dans ce paragraphe un théorème très utile, formalisant un certain nombre de raisonnements intuitifs de calculs sur les moyennes. Ce résultat, malgré sa simplicité, n'est malheureusement pas au programme.

Définition 31.2.14 (Espérance conditionnelle, HP)

Soit A un événement non presque-impossible et X une variable aléatoire réelle discrète. Alors, on dit que X admet une espérance conditionnelle sachant A (ou espérance conditionnée à A) si X admet une espérance condionnelle pour la mesure de probabilité $\mathbb{P}_A$, donc si la famille $(x\mathbb{P}_A([X = x]))_{x \in X(\Omega)}$ est sommable. Dans ce cas, on définit :

$$\mathbb{E}(X \mid A) = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x \mid A).$$

Théorème 31.2.15 (Formule de l'espérance totale, HP)

Soit $(A_n)_{n \in I}$ un système quasi-complet fini ou dénombrable, constitué d'événements non quasi-impossibles, et X une v.a.r.d.. Alors X admet une espérance si et seulement si les deux conditions ci-dessous sont satisfaites :

(i) pour tout $n \in I$, X admet une espérance conditionnelle sachant A_n, (donc $|X|$ aussi) ;

(ii) la série $\sum_{n \in I} \mathbb{E}(|X| \mid A_n)\mathbb{P}(A_n)$ est convergente

Dans ce cas :

$$\mathbb{E}(X) = \sum_{n \in I} \mathbb{E}(X \mid A_n)\mathbb{P}(A_n)$$

◁ **Éléments de preuve.**

Injecter la formule des probabilités totales dans la formule de l'espérance, utiliser la propriété d'associativité pour obtenir la sommabilité de la somme double, et pouvoir utiliser Fubini. ▷

Ce théorème est une formalisation de raisonnements en moyennes que l'on est amené à faire de façon assez intuitive. Par exemple, si, lorsqu'on choisit un restaurant, on choisit le restaurant « À la bonne fourchette » avec une probabilité de $\frac{3}{4}$ et le restaurant « Ah ! la chère fourchette ! » avec une probabilté de $\frac{1}{4}$, et qu'en moyenne, on dépense 20 euros en allant dans le premier et 80 euros en allant dans le second, une soirée au restaurant nous coûte en moyenne $\frac{3}{4} \times 20 + \frac{1}{4} \times 80 = 35$ euros.

III Variance, dispersion, moments

Connaître la valeur moyenne prise par une variable est une chose, mais n'a pas forcément beaucoup de pertinence pour avoir un ordre de grandeur du résultat d'une expérience. En effet, l'espérance ne mesure pas la dispersion de la variable aléatoire : les valeurs prises peuvent tout aussi bien rester toujours très proches de $\mathbb{E}(X)$, ou alors s'en éloigner beaucoup, de part et d'autre, de façon à ce qu'en moyenne cela se compense. Il est donc intéressant de définir une quantité mesurant cette dispersion. La dispersion correspond à la moyenne de l'écart à $\mathbb{E}(X)$, c'est-à-dire de $|X - \mathbb{E}(X)|$. La moyenne arithmétique correspondrait donc à l'espérance de $|X - \mathbb{E}(X)|$. Pour des raisons techniques, ce choix n'est pas bon (les valeurs absolues rendent le calcul de cette quantité difficile). D'ailleurs, s'agissant d'une distance, le choix euclidien usuel est plutôt quadratique. Pour cette raison, on a plutôt adopté la moyenne quadratique, c'est-à-dire la racine carrée de la moyenne de $(X - \mathbb{E}(X))^2$. Ce choix est surtout motivé par la facilité de calcul qui découle de ce choix, notamment grâce à la formule de König-Huygens.

III.1 Moments d'ordre k

Définition 31.3.1 (Moments d'ordre k)

1. Le moment d'ordre k de X est $\mathbb{E}(X^k)$ (si ce moment existe)
2. Si X admet une espérance, le moment centré d'ordre k de X est $\mathbb{E}((X - \mathbb{E}(X))^k)$, donc le moment d'ordre k de la variable centrée $X - \mathbb{E}(X)$.

Le théorème de transfert nous donne alors

Proposition 31.3.2 (Expression des moments)

Soit X une variable aléatoire réelle discrète.

1. *X admet (dans $\mathbb{R}$) un moment d'ordre k si et seulement si $(x^k \mathbb{P}(X = x))_{x \in X(\Omega)}$ est absolument sommable, et dans ce cas*
$$\mathbb{E}(X^k) = \sum_{x \in X(\omega)} x^k \mathbb{P}(X = x).$$
2. *X admet (dans $\mathbb{R}$) un moment centré d'ordre k si et seulement $((x - \mathbb{E}(X))^k \mathbb{P}(X = x))_{x \in X(\Omega)}$ est absolument sommable, et dans ce cas*
$$\mathbb{E}((X - \mathbb{E}(X))^k) = \sum_{x \in X(\omega)} (x - \mathbb{E}(X))^k \mathbb{P}(X = x).$$

Par exemple, le moment d'ordre 1 est égal à l'espérance, tandis que le moment centré d'ordre 1 est nul.

Théorème 31.3.3

Soit X une variable aléatoire réelle discrète. Si X admet un moment d'ordre k, alors X admet des moments de tous ordres $\ell \in [\![1, k]\!]$.

◁ **Éléments de preuve.**
$\mathbb{E}(|X|^\ell \mid |X| \leqslant 1) < +\infty$ (variable bornée ps pour la mesure de probabilité conditionnelle) et $\mathbb{E}(|X|^\ell \mid |X| \geqslant 1) < +\infty$ (comparaison à X^k), puis théorème de l'espérance totale. ▷

Proposition 31.3.4 (Positivité des moments d'ordre pair)
Soit X une variable aléatoire admettant un moment d'ordre pair $2k$. Alors :
- $\mathbb{E}(X^{2k}) \geqslant 0$*, avec égalité si et seulement si X est nulle presque sûrement.*
- $\mathbb{E}((X - \mathbb{E}(X))^{2k}) \geqslant 0$*, avec égalité si et seulement si X est constante presque sûrement.*

◁ **Éléments de preuve.**
Le cas d'égalité découle du cas d'égalité dans 31.2.11. ▷

III.2 Variance

Un cas particulier important est le suivant :

Définition 31.3.5 (Variance d'une variable aléatoire)
Soit X une v.a.r.d. admettant une espérance $\mathbb{E}(X)$. Alors, sous réserve d'existence (c'est-à-dire sous reserve de sommabilité), la variance de X est :

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

Ainsi, $\mathbb{V}$ est le moment centré d'ordre 2 de X.

Comme cas particulier de 31.3.4, on a toujours $\mathbb{V}(X) \geqslant 0$ avec égalité si et seulement si X est constante ps.
On peut alors définir l'écart quadratique moyen à la moyenne :

Définition 31.3.6 (Écart-type)
L'écart-type est l'écart quadratique moyen à la moyenne, à savoir :

$$\sigma(X) = \sqrt{\mathbb{V}(X)}.$$

Remarque 31.3.7
La variance peut ne pas exister même si l'espérance existe. Pouvez-vous trouver un exemple ?

Et pour rester dans le cadre du programme :

Théorème 31.3.8 (variance d'une variable bornée)
Soit X une variable bornée (par exemple finie) sur $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Alors X admet une variance et un écart-type.

◁ **Éléments de preuve.**
$(X - \mathbb{E}(X))^2$ est aussi bornée. ▷

Dans la pratique, la variance se calcule le plus souvent avec le :

Théorème 31.3.9 (Koenig-Huyghens)

Soit X une v.a.r.d. Alors X admet une variance si et seulement si $\mathbb{E}(X^2)$ existe (on dit que X admet un moment d'ordre 2) et dans ce cas :

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

◁ **Éléments de preuve.**

Développer $(X - \mathbb{E}(X))^2$. La relation obtenue montre que si 2 des 3 variables X, X^2 et $(X - \mathbb{E}(X))^2$ admettent une espérance, la troisième aussi. Or, si X admet une variance, par définition, elle admet une espérance, et si X admet un moment d'ordre 2, par 31.3.3, elle admet une espérance. La formule de K-H est alors immédiate par linéarité de l'espérance. ▷

La définition-même de la variance amène :

Proposition 31.3.10 (propriétés de la variance)

Soit a et b deux réels et X une variable aléatoire admettant une variance. Alors $aX + b$ aussi, et

$$\mathbb{V}(aX + b) = a^2\mathbb{V}(X).$$

Avertissement 31.3.11

En général, on n'a pas $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$. On verra dans la suite du chapitre qu'une condition suffisante pour que ce soit vrai est que X et Y soient indépendantes, dans un sens qu'on définira dans ce chapitre. Dans le cas général, la variance d'une somme s'exprime à l'aide de la covariance, ce que nous verrons plus loin.

III.3 Variables centrées réduites

Définition 31.3.12 (variable centrée, réduite)

Une variable X est dite :

- centrée si elle admet une espérance, et $\mathbb{E}(X) = 0$
- réduite si elle admet une variance, et $\mathbb{V}(X) = 1$
- centrée réduite si elle est à la fois centrée et réduite.

Proposition 31.3.13 (Variable centrée réduite associée à X)

Soit X une variable aléatoire réelle discrète admettant une variance non nulle. Il existe d'uniques réels $a > 0$ et b tels que la variable X^* définie par $X^* = aX + b$ soit centrée réduite. Plus précisément,

$$X^* = \frac{X - \mathbb{E}(X)}{\sigma(X)}.$$

On dit que X^* est la variable aléatoire centrée réduite associée à X.

◁ **Éléments de preuve.**

Analyse synthèse par linéarité de $\mathbb{E}$ et quadraticité de $\mathbb{V}$. ▷

IV Indépendance de variables aléatoires

IV.1 Couples de variables aléatoires indépendantes

Définition 31.4.1 (Variables indépendantes)

Soit X et Y deux variables aléatoires définies sur $(\Omega, \mathcal{T}, \mathbb{P})$, à valeurs dans $(E_1, \mathcal{T}_1)$ et $(E_2, \mathcal{T}_2)$ respectivement. Les variables X et Y sont indépendantes si et seulement si pour tout $(A, B) \in \mathcal{T}_1 \times \mathcal{T}_2$, les événements $[X \in A]$ et $[Y \in B]$ sont indépendants, c'est-à-dire :

$$\forall (A, B) \in \mathcal{T}_1 \times \mathcal{T}_2, \quad \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Autrement dit, X et Y sont indépendantes si et seulement les tribus $\mathcal{T}_X$ et $\mathcal{T}_Y$ sont indépendantes (dans le sens où un prenant un événement quelconque dans chacune de ces tribus, ces deux événements sont indépendants).

En particulier, le lemme de classe monotone, qu'on a déjà évoqué sans l'énoncer, amène, pour les variables aléatoires réelles :

Proposition 31.4.2 (v.a.r. indépendantes, admis)

Soit X et Y deux variables aléatoires réelles. Alors X et Y sont indépendantes si et seulement si pour tout $(x, y) \in \mathbb{R}^2$, les événements $[X \leqslant x]$ et $[Y \leqslant y]$ sont indépendants, c'est-à-dire :

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

Pour les variables aléatoires réelles discrètes, on obtient une description ponctuelle :

Proposition 31.4.3 (v.a.r.d. indépendantes)

Soit X et Y deux variables aléatoires réelles discrètes. Alors X et Y sont indépendantes si et seulement si pour tout $x \in X(\Omega)$ et pour tout $y \in Y(\Omega)$, les événements $[X = x]$ et $[Y = y]$ sont indépendants, c'est-à-dire :

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

◁ **Éléments de preuve.**
Exprimer A et B comme union de singletons. ▷

Remarque 31.4.4

Ce dernier résultat affirme qu'en cas d'indépendance, les lois marginales déterminent la loi conjointe d'un couple de variables aléatoires réelles discrètes. Ce fait reste vrai dans une situation plus générale (variables non discrètes), mais repose sur des résultats un peu plus délicat sur les tribus (notamment le lemme de classe monotone).

En effet, dans le cas où X et Y sont deux variables aléatoires réelles indépendantes de X et Y, la loi du couple (X, Y) coïncide avec le produit des lois de X et de Y sur les boréliens de $\mathbb{R}^2$ de la forme $A \times B$, où A et B sont des boréliens de $\mathbb{R}$. Or, les boréliens du type $A \times B$ engendrent (par définition) $\mathcal{B}_2$ et forment une classe stable par intersection finie. Or, d'après le lemme λ-π de Dynkin vu en DM, il existe une unique mesure de probabilité sur $\mathbb{R}^2$ prenant des valeurs déterminées sur les $A \times B$. Ainsi, $\mathbb{P}_X$ et $\mathbb{P}_Y$ déterminent $\mathbb{P}_{(X,Y)}$.
Ce raisonnement reste valable plus généralement pour des variables quelconques à valeurs respectivement dans $(E_1, \mathcal{T}_1)$ et $(E_2, \mathcal{T}_2)$, en remarquant que le couple (X, Y) définit dans ce cas une variable aléatoire à valeur dans le produit $E_1 \times E_2$, muni de la tribu produit, engendrée par les produits $A \times B$, pour $A \in \mathcal{T}_1$ et $B \in \mathcal{T}_2$.

IV.2 Familles de v.a.r. indépendantes

On généralise ici la notion d'indépendance au cas d'un nombre plus grand de variables aléatoires. Comme dans le cas des événements, la notion d'indépendance deux à deux est souvent insuffisante pour traduire complètement les situations d'indépendance rencontrées. Nous devons ici aussi distinguer l'indépendance deux à deux de l'indépendance mutuelle définie ci-dessous.
Soit $(X_k)_{k\in K}$ une famille quelconque de variables aléatoires, X_k étant à valeurs dans $(E_k, \mathcal{T}_k)$

Définition 31.4.5 (Famille de variables mutuellement indépendantes)

On dit que les variables aléatoires X_k, pour $k \in K$, sont mutuellement indépendantes si pour toute famille $(A_k)_{k\in K}$ telle que pour tout $k \in K$, $A_k \in \mathcal{T}_k$, les événements $[X_k \in A_k]$, pour $k \in K$, sont mutuellement indépendants, donc si et seulement si pour tout sous-ensemble fini J de K,

$$\mathbb{P}\left(\bigcap_{j\in J}[X_j \in A_j]\right) = \prod_{j\in J}\mathbb{P}(X_j \in A_k).$$

De même que pour 2 variables, cela se traduit par le fait que les tribus $\mathcal{T}_{X_j}$ forment une famille de tribus mutuellement indépendantes (dans le sens où en prenant un événement quelconque de chaque tribu, on obtient toujours une famille mutuellement indépendante).
Dans le cas de variables aléatoires réelles discrètes, on a la définition équivalente ci-dessous.

Proposition 31.4.6 (Caractérisation ponctuelle de la mutuelle indépendance)

Si les X_k sont réelles discrètes, elles sont mutuellement indépendantes si et seulement si pour toute famille $(x_k)_{k\in K}$ telle que pour tout $k \in K$, $x_k \in X_k(\Omega)$, les événements $[X_k = x_k]$, pour $k \in K$, sont mutuellement indépendants, donc si et seulement si pour tout sous-ensemble fini J de K,

$$\mathbb{P}\left(\bigcap_{j\in J}[X_j = x_j]\right) = \prod_{j\in J}\mathbb{P}(X_j = x_j).$$

◁ **Éléments de preuve.**
On peut se contenter de montrer l'équivalence dans le cas $|J| = 2$ (récurrence). Décomposer les $[X_k \in A_k]$ en union finie ou dénombrable de $[X_k = x_{k,i}]$. ▷

IV.3 Fonctions de variables indépendantes

Proposition 31.4.7 (Fonctions de variables indépendantes)

Soit X et Y deux variables aléatoires independantes, et f et g deux fonctions mesurables. Alors $f(X)$ et $g(Y)$ sont indépendantes.

Ceci est un cas particulier du résultat plus général ci-dessous :

Proposition 31.4.8 (Fonctions de variables mutuellement indépendantes)

Soit $(X_n)_{n\in K}$ une famille de variables aléatoires mutuellement indépendantes, et $(f_n)_{n\in K}$ des fonctions mesurables. Alors les $f_n(X_n)$ sont mutuellement indépendantes.

◁ **Éléments de preuve.**
Vérification facile. ▷

V Covariance

V.1 Espérance d'un produit

On a vu que l'espérance est linéaire. On peut se demander si elle est multiplicative. La réponse est non en général. On peut cependant obtenir :

Théorème 31.5.1 (Espérance d'un produit de variables indépendantes)
Si X et Y sont indépendantes et admettent une espérance, alors XY admet une espérance, et

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

◁ **Éléments de preuve.**
Théorème de transfert avec la fonction $f : (x, y) \mapsto xy$. ▷

Il est important de noter que ce résultat entre en défaut lorsque X et Y ne sont pas indépendants :

Exemple 31.5.2
Soit X_1, X_2, et X_3, indépendantes, suivant des lois de Bernoulli de paramètre p. Soit $Y_1 = X_1X_2$ et $Y_2 = X_2X_3$. Alors $\mathbb{E}(Y_1Y_2) \neq \mathbb{E}(Y_1)\mathbb{E}(Y_2)$.

On donne tout de même un résultat d'existence :

Lemme 31.5.3 (CS d'existence de $\mathbb{E}(XY)$)
Si X et Y admettent des moments d'ordre 2, alors XY admet une espérance.

◁ **Éléments de preuve.**
Majoration $|xy| \leqslant \frac{1}{2}(x^2 + y^2)$ dans la somme déterminant XY par le théorème de transfert. ▷

On en déduit en particulier :

Proposition 31.5.4 (L'espace des v.a.r.d. admettant un moment d'ordre 2)
L'ensemble des variables aléatoires discrètes sur $(\Omega, \mathcal{T}, \mathbb{P})$ admettant un moment d'ordre 2 est un sous-espace vectoriel de l'espace de toutes les v.a.r.d..

◁ **Éléments de preuve.**
La linéarité nécessite d'étudier $(\lambda X + Y)^2$, et fait donc intervenir XY. ▷

V.2 Covariance

Nous introduisons maintenant une quantité permettant de comparer les variations de X et de Y. Si X et Y sont indépendantes, ces variations sont indépendantes ce qui se traduira par le fait que cette quantité est nulle. Cette propriété sera le reflet de l'égalité que l'on vient d'obtenir $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Ainsi, notre mesure du défaut d'indépendance se définira comme la différence $\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. Afin de mieux comprendre la signification de cette différence, nous l'intruisons sous une forme légèrement différente.

Définition 31.5.5 (Covariance)
Soit X et Y deux variables aléatoires réelles discrètes. Sous réserve d'existence de ces espérances, on définit la *covariance* de X et Y par :

$$\operatorname{cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big).$$

Remarque 31.5.6 (Interprétation de la covariance)

Si $\mathrm{cov}(X,Y) > 0$, $X - \mathbb{E}(X)$ et $Y - \mathbb{E}(Y)$ ont tendance à être de même signe, donc X et Y ont tendance à se situer du même côté de leur espérance. Ainsi :

- $\mathrm{cov}(X,Y) > 0$ signifie que X et Y ont tendance à évoluer parallèlement
- $\mathrm{cov}(X,Y) < 0$ signifie que X et Y ont tendance à avoir une évolution opposée.

Définition 31.5.7 (Variables décorrélées)

Les variables X et Y sont dites décorrélées lorsque $\mathrm{cov}(X,Y) = 0$.

Proposition 31.5.8 (CS d'existence de $\mathrm{cov}(X,Y)$)

Si X et Y admettent des moments d'ordre 2, alors $\mathrm{cov}(X,Y)$ existe.

◁ **Éléments de preuve.**

Elles admettent alors aussi des moments d'ordre 1. Développer. ▷

Proposition 31.5.9 (Propriétés de la covariance)

Soit X et Y des v.a.r.d. admettant des moments d'ordre 2.

1. $\mathrm{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$;
2. $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$ *(symétrie)*
3. cov *est bilinéaire sur l'espace des v.a.r.d. sur $(\Omega, \mathcal{T}, \mathbb{P})$ admettant un moment d'ordre 2.*
4. $\mathrm{cov}(X,X) = \mathbb{V}(X)$.
5. $\mathrm{cov}(1,X) = 0$.
6. *Si X et Y sont indépendantes,* $\mathrm{cov}(X,Y) = 0$.

◁ **Éléments de preuve.**

Assez immédiat par linéarité de l'espérance. ▷

Avertissement 31.5.10

La réciproque de la dernière propriété est fausse : il existe des variables décorrélées, mais non indépendantes. Voir les exercices pour des exemples.

Les propriétés 2, 3 et 4 permettent d'affirmer que cov est une forme bilinéaire symétrique positive, de forme quadratique associée égale à la variance. En particulier, on dispose dans cette situation de l'inégalité de Cauchy-Schwarz :

Théorème 31.5.11 (Inégalité de Cauchy-Schwarz)

Soit X et Y des variables aléatoire réelles discrètes. Alors

$$|\mathrm{cov}(X,Y)| \leqslant \sigma(X)\sigma(Y),$$

avec égalité si et seulement s'il existe une relation affine presque sûrement entre X et Y (c'est-à-dire une relation non triviale $aX + bY + c = 0$)

◁ **Éléments de preuve.**

Si cela n'a pas encore été fait à ce moment, on renvoie au cours d'algèbre bilinéaire, pour la preuve de cette inégalité, plus généralement pour des formes bilinéaires symétriques positives. ▷

Définition 31.5.12 (Coefficient de corrélation)

Le coefficient de corrélation $\rho(X,Y)$ est défini, pour des variables non quasi-certaines, par :

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

D'après le résultat précédent, $\rho(X,Y) \in [-1,1]$; $\rho(X,Y) = 0$ indique une décorrélation, alors que $|\rho(X,Y)| = 1$ indique une corrélation très forte (les variables aléatoires X et Y sont liées par la relation de dépendance la plus forte qu'on puisse imaginer).

Remarque 31.5.13

$(X,Y) \mapsto \mathbb{E}(XY)$ est également bilinéaire symétrique positive, donc on a aussi :

$$|E(XY)| \leqslant \sqrt{E(X^2)E(Y^2)}$$

Cela donne au passage une nouvelle preuve de la CS d'existence de $E(XY)$.

V.3 Variance d'une somme

Une autre formule issue directement de l'algèbre bilinéaire (et du lien entre une forme bilinéaire symétrique et la forme quadratique associée) est la formule de polarisation :

Proposition 31.5.14 (Formule de polarisation)

Soit X et Y deux variables aléatoires. Alors :

$$\mathrm{cov}(X,Y) = \frac{1}{2}(\mathbb{V}(X+Y) - \mathbb{V}(X) - \mathbb{V}(Y)).$$

Cette formule est dans notre contexte surtout utilisée pour calculer la variance d'une somme à l'aide des covariances. Nous la reexprimons donc de la façon suivante :

Proposition 31.5.15 (Variance d'une somme)

Si X et Y admettent une variance, alors $X+Y$ également, et :

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\cdot\mathrm{cov}(X,Y).$$

On en déduit alors de façon immédiate :

Théorème 31.5.16 (Variance d'une somme de variables indépendantes)

Si X et Y sont indépendantes, et admettent une variance, alors $X+Y$ aussi, et $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.

Remarque 31.5.17

La multiplication des réels est aussi une forme bilinéaire symétrique. A quoi correspond la formule de polarisation dans ce contexte ?

En utilisant la propriété de bilinéarité généralisée, on obtient de même l'expression de la variance d'une somme de n terme :

Proposition 31.5.18 (Variance de $X_1 + \cdots + X_n$)

Soit $X_1, \ldots, X_n$ des v.a.r.d. ayant un moment d'ordre 2 ; $X_1 + \cdots + X_n$ admet une variance et :

$$\mathbb{V}(X_1 + \cdots + X_n) = \mathbb{V}(X_1) + \cdots + \mathbb{V}(X_n) + 2 \cdot \sum_{1 \leqslant i < j \leqslant n} \operatorname{cov}(X_i, X_j).$$

Théorème 31.5.19 (Variance d'une somme de variables 2 à 2 indépendantes)

Sous les mêmes hypothèses, si $X_1, \ldots, X_n$ sont deux à deux *indépendantes :*

$$\mathbb{V}(X_1 + \cdots + X_n) = \mathbb{V}(X_1) + \cdots + \mathbb{V}(X_n).$$

Remarque 31.5.20

Il est remarquable qu'il n'y ait besoin que d'une hypothèse d'indépendance 2 à 2 ici, et non d'une indépendance mutuelle. Une hypothèse de non corrélation 2 à 2 serait même suffisante.

V.4 Matrice des variances-covariances

Nous terminons l'étude de la covariance en exposant certaines techniques matricielles pour le calcul des covariances et des variances. Ces techniques sont issues de techniques plus générales liées aux formes bilinéaires.
On se donne dans ce paragraphe une famille $(X_1, \ldots, X_n)$ de variables aléatoires réelles discrètes admettant un moment d'ordre 2.

Définition 31.5.21 (Matrice des variances-covariances)

On définit la matrice des variances-covariances d'un vecteur aléatoire discret $X = (X_1, \ldots, X_n)$ par :

$$\mathbb{V}(X) = (\operatorname{cov}(X_i, X_j))_{1 \leqslant i,j \leqslant n}$$

Ainsi, cette matrice est symétrique, et sa diagonale est formée des variances des X_i. Par exemple,

$$\mathbb{V}((X, Y)) = \begin{pmatrix} \mathbb{V}(X) & \operatorname{cov}(X, Y) \\ \operatorname{cov}(X, Y) & \mathbb{V}(Y). \end{pmatrix}$$

On peut alors calculer matriciellement les covariances de toutes combinaison linéaires des X_i :

Proposition 31.5.22 (Expression matricielle de la covariance de CL)

Pour tout $(\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_n) \in \mathbb{R}^{2n}$:

$$\operatorname{cov}(\lambda_1 X_1 + \cdots + \lambda_n X_n, \mu_1 X_1 + \cdots + \mu_n X_n) = \begin{pmatrix} \lambda_1 & \cdots & \lambda_n \end{pmatrix} \mathbb{V}(X_1, \ldots, X_n) \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

En particulier, on obtient une expression matricielle de la variance de la somme, permettant souvent de mener les calculs de façon plus ordonnée que l'utilisation directe de la formule précédente, même si *stricto sensu* il s'agit de la même formule !

Proposition 31.5.23 (Expression matricielle de la variance d'une somme)

Soit $C = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$. *Alors :* $\mathbb{V}(X_1 + \cdots + X_n) = {}^tC\mathbb{V}(X_1, \ldots, X_n)C$.

VI Lois discrètes classiques

Un certain nombre de lois interviennent fréquemment. De nombreuses situations peuvent s'y ramener. Il est donc intéressant de les connaître afin de s'économiser un certain nombre de calculs d'espérance ou de variance. Il est crucial de bien connaître l'expérience-type associée à chacune de ces lois, car c'est *via* la description de cette expérience que ces lois interviennent dans des situations diverses.

VI.1 Loi uniforme

L'expérience consiste à tirer une boule dans une urne contenant n boules numérotées de 1 à n. La variable X est le numéro de la boule tirée. Ainsi, on définit :

Définition 31.6.1 (Loi uniforme)

Soit $n \in \mathbb{N}^*$. Une v.a.r. X suit la loi uniforme de paramètre n si $X(\Omega) = [\![1, n]\!]$ et si :

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

On note $X \hookrightarrow \mathcal{U}(n)$.

Proposition 31.6.2 (Espérance et variance d'une loi uniforme)

Soit $X \hookrightarrow \mathcal{U}(n)$. *Alors* $\mathbb{E}(X) = \dfrac{n+1}{2}$ *et* $\mathbb{V}(X) = \dfrac{n^2-1}{12}$.

◁ **Éléments de preuve.**

Utiliser KH. On est alors ramené à des sommes d'entiers et de carrés d'entiers. ▷

Pour tout ensemble fini E, on note plus généralement $X \hookrightarrow \mathcal{U}(E)$ pour une v.a.r. suivant la loi uniforme sur X, c'est-à-dire une v.a.r. telle que :

$$\forall x \in E, \quad \mathbb{P}(X = x) = \frac{1}{|E|}.$$

Proposition 31.6.3 (Expérience-type associée à une loi uniforme)

Soit U une urne contenant n boules numérotées de 1 à n. On effectue un tirage avec équiprobabilité, et on note X le numéro de la boule obtenu. Alors $X \hookrightarrow \mathcal{U}(n)$.

Plus généralement, si X est une variable suivant une loi uniforme sur l'ensemble $[\![a, b]\!]$, elle vérifie :

$$\mathbb{E}(X) = \frac{a+b}{2} \qquad \text{et} \qquad \mathbb{V}(X) = \frac{(b-a+1)^2 - 1}{12}.$$

VI.2 Loi de Bernoulli

L'expérience consiste à tirer dans une urne contenant une proportion p de boules blanches. On note X la variable aléatoire égale à 1 si on tire une boule blanche, et 0 sinon.

On peut aussi définir X à l'aide de tirages à pile ou face avec une pièce déséquilibrée, dont la probabilité de pile est p ; X prend alors la valeur 1 si on tire pile, et 0 si on tire face.
Ainsi, $[X = 1]$ représente le succès dans une expérience ayant une probabilité p de succès.

De façon plus générale, une situation similaire se produit dès lors qu'on a une expérience à deux issues : succès et échec. L'événement $[X = 1]$ représente alors le succès, et $[X = 0]$ représente l'échec.
On définit donc :

Définition 31.6.4 (Loi de Bernoulli)
Soit X une v.a.r. et $p \in]0,1[$. On dit que X suit une loi de Bernoulli de paramètre p (et on note $X \hookrightarrow \mathcal{B}(1,p)$ ou $X \hookrightarrow \mathcal{B}(p)$) si :

$$X(\Omega) = \{0,1\} \qquad \text{et} \qquad \begin{cases} \mathbb{P}(X=0) = 1-p \\ \mathbb{P}(X=1) = p \end{cases}$$

Désormais, on se donne un réel $p \in]0,1[$, et on note $q = 1-p$.

Proposition 31.6.5 (Espérance et variance d'une loi de Bernoulli)
Soit $X \hookrightarrow \mathcal{B}(p)$. Alors $\mathbb{E}(X) = p$ et $\mathbb{V}(X) = pq$.

◁ **Éléments de preuve.**
Un peu trivial... ▷

Exemple 31.6.6 (Fonction caractéristique d'un événement)
Étude de $X(\omega) = \mathbb{1}_A(\omega)$, $A \in \mathcal{P}(\Omega)$.

La motivation même de la définition de la loi de Bernoulli amène :

Proposition 31.6.7 (Expérience-type associée à une loi de Bernoulli)
Soit une expérience à 2 issue (par exemple P/F, ou tirage dans une urne contenant 2 types de boules), l'une représentant le succès et l'autre l'échec, la probabilité d'obtenir un succès étant p. Soit X prenant la valeur 1 en cas de succès et la valeur 0 en cas d'échec. Alors $X \hookrightarrow \mathcal{B}(p)$.

Une expérience telle que décrite dans la proposition précédente est usuellement appelée « expérience de Bernoulli ».

Remarque 31.6.8
Toute variable aléatoire prenant ses valeurs dans $\{0,1\}$ est une variable de Bernoulli, à condition d'élargir la définition en acceptant de considérer les cas dégénérés $p = 0$ et $p = 1$.

VI.3 Loi binomiale – nombre de succès

Soit $n \in \mathbb{N}^*$. On répète n fois une expérience de Bernoulli, et on note X le nombre de succès obtenus. Par exemple on compte le nombre de Pile obtenus dans une succession de n lancers (indépendants) avec une pièce déséquilibrée donnant Pile avec une probabilité de p.
Déterminons la loi de X. Tout d'abord, on peut avoir de 0 à n succès, ainsi $X(\Omega) = [\![0, n]\!]$. Soit donc $k \in [\![0, n]\!]$. L'événement $[X = k]$ est obtenu si on a eu k succès. Soit, pour tout $i \in [\![1, n]\!]$, S_i l'événement : obtenir un succès à la i-ième expérience. Alors :

$$[X = k] = \bigcup_{\substack{I \subset [\![1,n]\!] \\ |I|=k}} \left(\bigcap_{i \in I} S_i \cap \bigcap_{i \in [\![1,n]\!] \setminus I} \overline{S_i} \right)$$

Les événements S_i, $i \in [\![1, n]\!]$, sont mutuellement indépendants, et les événements de l'union sont deux à deux incompatibles. Ainsi,

$$\mathbb{P}(X = k) = \sum_{\substack{I \subset [\![1,n]\!] \\ |I|=k}} \left(\prod_{i \in I} p \cdot \prod_{i \in [\![1,n]\!] \setminus I} (1-p) \right) = \sum_{\substack{I \subset [\![1,n]\!] \\ |I|=k}} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Le coefficient binomial correspond à la position des k succès, p^r est la probabilité d'obtention de ces r succès et $(1-p)^r$ est la probabilité d'obtention des échecs aux autres tirages.
On définit donc :

Définition 31.6.9 (Loi binomiale)
Soit X une v.a.r.. On dit que X suit la loi binomiale de paramètres (n, p) (et on note $X \hookrightarrow \mathcal{B}(n, p)$) si :

$$X(\Omega) = [\![0, n]\!], \qquad \text{et} \qquad \forall k \in [\![0, n]\!],\ \mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}, \quad \text{où } q = 1 - p.$$

Ceci définit bien une loi de probabilités. En effet : $\displaystyle\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1.$

Le calcul introductif nous ayant servi de motivation à la définition de la loi binomiale amène directement :

Proposition 31.6.10 (Expérience-type associée à la loi binomiale)
Soit X le nombre de succès obtenus lors d'une répétition de n épreuves de Bernoulli indépendantes de même paramètre p. Alors $X \mapsto \mathcal{B}(n, p)$.

Proposition 31.6.11 (Espérance et variance d'une loi binomiale)
Soit $X \hookrightarrow \mathcal{B}(n, p)$. Alors $\mathbb{E}(X)$ et $\mathbb{V}(X)$ existent, et :

$$\mathbb{E}(X) = np, \qquad \text{et} \qquad \mathbb{V}(X) = npq.$$

◁ **Éléments de preuve.**
On est ramené à des sommes du binôme, éventuellement après un ou deux comité-président. Il peut être intéressant de remplacer le calcul de $\mathbb{E}(X^2)$ par le calcul de l'espérance d'un autre polynôme en X de degré 2, permettant directement l'utilisation répétée de la formule comité-président. ▷

On peut obtenir ce résultat soit à l'aide d'un calcul direct, soit en remarquant que X est une somme de n variables de Bernoulli indépendantes.

VI.4 Loi géométrique – temps d'attente du premier succès

On considère maintenant une succession infinie d'expériences de Bernoulli indépendantes de paramètre p. On considère X la variable aléatoire correspondant au rang du premier succès.
Par exemple X est le rang du premier pile obtenu dans une succession de lancers avec une probabilité d'obtention de pile égale à p pour chaque tirage.
Déterminons la loi de X. Les valeurs possibles de X sont toutes les valeurs entières strictement positives : $X(\Omega) = \mathbb{N}^*$. Soit $k \in \mathbb{N}^*$. Alors, en utilisant les mêmes notations que précédemment,

$$\mathbb{P}(X = k) = \mathbb{P}(\overline{S_1} \cap \cdots \cap \overline{S_{k-1}} \cap S_k) = \mathbb{P}(\overline{S_1}) \cdots \mathbb{P}(\overline{S_{k-1}})\mathbb{P}(S_k) = pq^{k-1}$$

On définit donc :

Définition 31.6.12 (Loi géométrique)
Soit X une v.a.r.. On dit que X suit la loi géométrique de paramètre p, et on note $X \hookrightarrow \mathcal{G}(p)$, si :

$$X(\Omega) = \mathbb{N}^*, \qquad \text{et} \qquad \forall k \in \mathbb{N}^*,\ \mathbb{P}(X = k) = pq^{k-1}.$$

Cela définit bien une loi de probabilité : $\displaystyle\sum_{k=1}^{+\infty} \mathbb{P}(X = k) = \sum_{k=1}^{+\infty} pq^{k-1} = \frac{p}{1-q} = 1.$

Et par notre motivation-même :

Proposition 31.6.13 (Expérience-type associée à une loi géométrique)
Soit X le temps d'attente du premier succès (c'est-à-dire le nombre d'expériences nécessaires pour obtenir le premier succès) lors d'une répétition infinie d'expériences de Bernoulli indépendantes de même paramètre p. Alors $X \hookrightarrow \mathcal{G}(p)$.

Proposition 31.6.14 (Espérance et variance d'une loi géométrique)
Soit $X \hookrightarrow \mathcal{G}(p)$. Alors X admet une espérance et une variance, et :

$$\mathbb{E}(X) = \frac{1}{p} \qquad \textit{et} \qquad \mathbb{V}(X) = \frac{q}{p^2}.$$

◁ **Éléments de preuve.**
On est ramené à des sommes géométriques dérivées. Comme plus haut, on peut avoir intérêt à ne pas considérer directement $\mathbb{E}(X^2)$. ▷

Remarque 31.6.15
Il est parfois commode de définir X à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$, et en définissant la probabilité $\mathbb{P}(X = +\infty) = 0$. Cela permet d'englober le cas du malchanceux éternel.

VI.5 Loi de Poisson

On définit une dernière loi, un peu à part puisqu'elle ne correspond pas à une expérience précise, mais qui apparaît souvent comme modélisation, ou comme loi limite (il s'agit d'une limite dans un certain sens de lois binomiales).

Définition 31.6.16 (Loi de Poisson)
Soit $\lambda \in \mathbb{R}_+$ et X une v.a.r.. On dit que X suit la loi de Poisson de paramètre λ, et on note $X \hookrightarrow \mathcal{P}(\lambda)$, si :

$$X(\Omega) = \mathbb{N} \qquad \text{et} \qquad \forall n \in \mathbb{N},\ \ \mathbb{P}(X = n) = \mathrm{e}^{-\lambda} \cdot \frac{\lambda^n}{n!}.$$

D'après les résultats sur les séries exponentielles, cela définit bien une loi de probabilité. De plus :

Proposition 31.6.17 (Espérance et variance d'une loi de Poisson)
Soit $X \hookrightarrow \mathcal{P}(\lambda)$. Alors X admet une espérance et une variance, et :

$$\mathbb{E}(X) = \lambda \qquad \textit{et} \qquad \mathbb{V}(X) = \lambda.$$

◁ **Éléments de preuve.**
Se ramener à des séries exponentielles, et toujours la même remarque. ▷

VI.6 Stabilité des lois classiques

Par stabilité d'une loi classique, on entend la chose suivante : si X et Y suit un certain type de loi et sont indépendante, alors $X + Y$ suit une loi de même type (avec des paramètres éventuellement différents). Nous énonçons ici deux propriétés de stabilité.

Proposition 31.6.18 (Stabilité des lois binomiales)
Soit $(m, n) \in (\mathbb{N}^*)^2$ et $p \in]0, 1[$. Si $X \hookrightarrow \mathcal{B}(n, p)$ et $Y \hookrightarrow \mathcal{B}(m, p)$ sont indépendantes, alors $X + Y \hookrightarrow \mathcal{B}(n + m, p)$.
En particulier si $X \hookrightarrow \mathcal{B}(n, p)$, X est la somme de n variables de Bernoulli de paramètre p mutuellement indépendantes.

◁ **Éléments de preuve.**
Se comprend bien par l'expérience. Le démontrer par le calcul en décomposant $[X + Y = k]$ en union de $[X = i] \cap [Y = k - i]$, et en calculant ces probabilités par indépendance. ▷

Corollaire 31.6.19 (Somme de variables de Bernoulli indépendantes)
Soit $X_1, \ldots, X_n$ des variables aléatoires mutuellement indépendantes suivant la loi $\mathcal{B}(p)$. Alors $X_1 + \cdots + X_n \hookrightarrow \mathcal{B}(n, p)$.

Proposition 31.6.20 (Stabilité des lois de Poisson)
Soit $(\lambda, \mu) \in (\mathbb{R}_+)^2$. Si $X \hookrightarrow \mathcal{P}(\lambda)$ et $Y \hookrightarrow \mathcal{P}(\mu)$ sont indépendantes, alors $X + Y \hookrightarrow \mathcal{P}(\lambda + \mu)$.

◁ **Éléments de preuve.**
Même principe de calcul. ▷

VI.7 Tableau récapitulatif

La table 31.1 récapitule les résultats obtenus pour les différentes lois classiques étudiées. Nous indiquons en gras les lois qui sont au programme de Sup et de Spé. Nous avons indiqué dans ce tableau quelques autres lois classiques, que vous pouvez étudier en exercice :

- Loi de Pascal $\mathcal{P}(r,p)$: temps d'attente du r-ième succès dans une répétition d'épreuves de Bernoulli indépendantes de même paramètre. Lorsque $r=1$, on retrouve une loi géométrique. On a une propriété de stabilité au sens du paragraphe précédent, qui se comprend bien de façon intuitive.
- Loi binomiale négative $\mathcal{J}(r,p)$: nombre d'échecs avant le r-ième succès. On peut remarquer que $X \hookrightarrow \mathcal{J}(r,p)$ ssi $X+r \hookrightarrow \mathcal{P}(r,p)$.
- Loi hypergéométrique $\mathcal{H}(N,n,p)$: nombre de boules blanches tirées lors de n tirages sans remise dans une urne contenant initialement Np boules blanches et $N(1-p)$ boules noires
- Les temps d'attente de la première boule blanche, ou de la r-ième, dans le même contexte.

VII Inégalités et convergences

Il est fréquent d'aboutir à l'étude d'une suite de variables aléatoires et de s'intéresser à la variable obtenue « par passage à la limite ». Il faut pour cela définir ce qu'on entend par passage à la limite sur des variables aléatoires et se donner quelques outils d'étude. Il existe plusieurs définitions de différents types de convergence. Les deux types les plus fréquemment utilisés sont la convergence en probabilités et la convergence en loi. Les deux étant hors-programme, nous nous contenterons d'étudier la convergence en probabilités. Attention cependant au fait que si les convergences sont hors-programme, les deux inégalités (Markov et Bienaymé-Tchebychev) sont elles bien au programme.

VII.1 Convergence en probabilités

Définition 31.7.1 (Convergence en probabilités, HP)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.r.d. sur $(\Omega,\mathcal{T},\mathbb{P})$, et X une autre *v.a.r.d.*. On dit que X_n *converge en probabilité* (ou *converge stochastiquement*) vers X si :

$$\forall \varepsilon>0,\quad \lim_{n\to+\infty}\mathbb{P}\big(|X_n-X|\geqslant\varepsilon\big)=0.$$

VII.2 Inégalités en probabilités

Théorème 31.7.2 (Inégalité de Markov)

Soit X une v.a.r.d. sur $(\Omega,\mathcal{T},\mathbb{P})$ telle que $X(\Omega)\subset\mathbb{R}_+$, admettant une espérance non nulle $\mathbb{E}(X)$. *Alors, pour tout $\lambda\in\mathbb{R}_+^*$:*

$$\mathbb{P}\big(X\geqslant\lambda\mathbb{E}(X)\big)\leqslant\frac{1}{\lambda}.$$

◁ **Éléments de preuve.**
Séparer la somme définissant $\mathbb{E}(X)$ en 2, suivant que $x<\lambda\mathbb{E}(X)$ ou $x\geqslant\lambda\mathbb{E}(X)$. Majorer puis simplifier par $\mathbb{E}(X)$. ▷

Voici deux formes souvent plus commodes à utiliser de l'inégalité de Markov :

Corollaire 31.7.3 (Réeexpression de l'inégalité de Markov)

Sous les mêmes hypothèses, pour tout $\varepsilon>0$,

$$\mathbb{P}(X\geqslant\varepsilon)\leqslant\frac{\mathbb{E}(X)}{\varepsilon}\qquad\text{et}\qquad\mathbb{P}(X>\varepsilon)\leqslant\frac{\mathbb{E}(X)}{\varepsilon}$$

◁ **Éléments de preuve.**
Avec $\lambda = \frac{\varepsilon}{\mathbb{E}(X)}$. ▷

Corollaire 31.7.4 (Inégalité de Markov quadratique)
Soit X une variable aléatoire (discrète ou à densité) admettant un moment d'ordre 2. Alors :

$$\mathbb{P}(|X| \geqslant \varepsilon) \leqslant \frac{\mathbb{E}(X^2)}{\varepsilon^2}.$$

◁ **Éléments de preuve.**
Si $\mathbb{E}(X^2) \neq 0$, appliquer ce qui précède à X^2. Et sinon ? ▷

Théorème 31.7.5 (Inégalité de Bienaymé-Tchebychev)
Soit Y une v.a.r.d. admettant une espérance m et une variance σ^2. Alors :

$$\forall \varepsilon > 0,\ \mathbb{P}\big(|Y - m| \geqslant \varepsilon\big) \leqslant \frac{\sigma^2}{\varepsilon^2}.$$

◁ **Éléments de preuve.**
Inégalité de markov quadratique appliquée à... ▷

VII.3 Loi faible des grands nombres

Un des exemples les plus importants de convergence en probabilité est celle qui résulte du calcul de la moyenne obtenue lors d'une répétition de variables aléatoires.

Théorème 31.7.6 (Loi faible des grands nombres, HP)
Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a.r.d. mutuellement indépendantes suivant une même loi (on dit= « indépendantes identiquement distribuées », ou, en abrégé : i.i.d.), ayant une espérance m et une variance σ^2. Soit $(Z_n)_{n\in\mathbb{N}^*}$ définie par :

$$\forall n \in \mathbb{N},\quad Z_n = \frac{X_1 + \cdots + X_n}{n}.$$

Alors $(Z_n)_{n\in\mathbb{N}^*}$ converge en probabilité vers la variable certaine égale à m. Plus précisément :

$$\forall \varepsilon > 0,\ \forall n \in \mathbb{N}^*,\quad \mathbb{P}\big(|Z_n - m| \geqslant \varepsilon\big) \leqslant \frac{\sigma^2}{n\varepsilon^2}.$$

◁ **Éléments de preuve.**
Inégalité de B-T appliquée à Z_n dont on sait facilement déterminer l'espérance et la variance. ▷

Théorème 31.7.7 (Théorème d'or de Bernoulli, HP)
Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables mutuellement indépendantes, suivant toutes une loi de Bernoulli de paramètre p. Soit :

$$\forall n \in \mathbb{N}^*,\ Z_n = \frac{X_1 + \cdots + X_n}{n}.$$

Alors $(Z_n)_{n\in\mathbb{N}^*}$ converge en probabilité vers la v.a.r. certaine égale à p. Plus précisément :

$$\forall \varepsilon > 0,\ \forall n \in \mathbb{N}^*,\quad \mathbb{P}\big(|Z_n - p| \geqslant \varepsilon\big) \leqslant \frac{1}{4n\varepsilon^2}.$$

◁ **Éléments de preuve.**

Conséquence immédiate de la loi faible des grands nombres, en utilisant une certaine inégalité archi-classique. ▷

Ainsi, la fréquence statistique d'un événement tend vers la probabilité de réalisation de l'événement.

Exemple 31.7.8

On tire 1000 fois à pile ou face avec une pièce déséquilibrée dont la probabilité d'obtention de Pile est p. On obtient 570 fois Pile. Donner un intervalle I tel que la probabilité que $p \in I$ soit supérieure à 0.9.

Nom	Paramètres	Notation	Valeurs	Loi	$\mathbb{E}(X)$	$\mathbb{V}(X)$
Uniforme	$n \in \mathbb{N}^*$	$\mathcal{U}(n)$	$[\![1, n]\!]$	$\mathbb{P}(X = k) = \dfrac{1}{n}$	$\dfrac{n+1}{2}$	$\dfrac{n^2-1}{12}$
Bernoulli	$p \in]0, 1[$	$\mathcal{B}(1, p)$	$\{0, 1\}$	$\mathbb{P}(X = 1) = p$	p	pq
Binomiale	(n, p)	$\mathcal{B}(n, p)$	$[\![0, n]\!]$	$\mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}$	np	npq
Géométrique	p	$\mathcal{G}(p)$	$\mathbb{N}^*$	$\mathbb{P}(X = k) = pq^{k-1}$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$
Pascal	(r, p)	$\mathcal{P}(r, p)$	$[\![r, +\infty[\![$	$\mathbb{P}(X = k) = \binom{k-1}{r-1} p^r q^{k-r}$	$\dfrac{r}{p}$	$\dfrac{rq}{p^2}$
Binomiale négative	(r, p)	$\mathcal{J}(r, p)$	$\mathbb{N}$	$\mathbb{P}(X = k) = \binom{k+r-1}{k} p^r q^k$	$\dfrac{rq}{p}$	$\dfrac{rq}{p^2}$
Hypergéométrique	(N, n, p)	$\mathcal{H}(N, n, p)$	$\subset [\![0, n]\!]$	$\mathbb{P}(X = k) = \dfrac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}}$	np	$npq\dfrac{N-n}{N-1}$
Attente du 1er succès (tirage sans remise)	(N, p)		$[\![1, Nq+1]\!]$	$\mathbb{P}(X = k) = \dfrac{Np}{k} \cdot \dfrac{\binom{Nq}{k-1}}{\binom{N}{k}}$		
Attente du r^{e} succès (tirage sans remise)	(r, N, p)		$[\![r, Nq+r]\!]$	$\mathbb{P}(X = k) = \dfrac{r}{k} \dfrac{\binom{Np}{r}\binom{Nq}{k-r}}{\binom{N}{k}}$		
Poisson	$\lambda \in \mathbb{R}_+$	$\mathcal{P}(\lambda)$	$\mathbb{N}$	$\mathbb{P}(X = k) = \mathrm{e}^{-\lambda} \cdot \dfrac{\lambda^k}{k!}$	λ	λ

Figure 31.1 – Table des lois discrètes classiques