

Cours de probabilités

Louis Gass

Table des matières

1	Un peu de dénombrement	2
1.1	Les règles de base	2
1.2	Permutation	4
1.3	Arrangement	5
1.4	Combinaison	5
2	Espace probabilisé	8
2.1	Un peu d'heuristique	8
2.2	Quelques définitions et exemples	9
2.3	Probabilité conditionnelle et indépendance	11
2.4	Quelques théorèmes limites	15
3	Variable aléatoire	17
3.1	Quelques définitions	17
3.2	Loi d'une variable aléatoire	19
3.3	Moyennes et dispersion d'une variable aléatoire réelle	25
3.3.1	Espérance	25
3.3.2	Médiane	29
3.3.3	Variance	30
3.4	Quelques loi usuelles	31
3.4.1	Lois discrètes	31
3.4.2	Loi à densité	32
3.5	Couple de variables aléatoires	33
3.6	Variables aléatoires indépendantes	34
4	Théorèmes limites en probabilité	37
4.1	Modes de convergence de variables aléatoires	37
4.2	Loi des grands nombres	38
4.3	Théorème central limite	39

1 Un peu de dénombrement

1.1 Les règles de base

Dénombrer, c'est l'art de savoir compter le nombre d'éléments d'un ensemble. Sa maîtrise permet souvent de montrer de jolies formules parfois sans efforts. Elle demande un peu d'intuition car il peut être facile de se tromper sur son comptage.

Nous allons voir quelques règles dites "de base" pour apprendre à compter. Sans toutes les énoncer car celles-ci sont "évidentes", elles permettent déjà de dénombrer pas mal d'ensembles.

Définition 1.1. Le **cardinal** d'un ensemble A est le nombre d'éléments qui composent A .

A notre niveau, on distinguera trois possibilités pour le cardinal d'un ensemble.

Un ensemble est soit de cardinal

- **dénombrable**, que l'on sépare en deux catégories :
 - **fini** : il y a un nombre fini d'éléments dans l'ensemble.
 - **infini dénombrable** : il y a un nombre infini d'éléments que l'on peut lister dans une suite indexée par les entiers naturels ; le *premier*, le *deuxième*, etc. C'est le cas des entiers naturels, des fractions rationnelles, etc.
- **indénombrable** : il y a un nombre infini d'éléments et on ne peut pas les lister dans une suite indexée par les entiers. C'est le cas de l'ensemble des nombres réels, de l'ensemble des parties de \mathbb{N} , etc.

On décrit quelques règles de base sur les cardinaux.

Si A et B sont deux ensembles en bijections, alors :

$$\text{Card}(A) = \text{Card}(B)$$

Ce principe permet de ramener le dénombrement de A à un dénombrement plus simple, celui de B . Par exemple, considérons un marcheur qui se déplace sur une grille soit vers le haut de 1 case, soit vers la droite de 1 case. Il part du point $(0,0)$ et arrive au point $(2,2)$. Combien y a-t-il de chemin possible ? Il y en a autant que d'anagramme du mot $RURU$, c'est-à-dire 6.

Soient A et B deux ensembles finis. On note $A \times B$ l'ensemble des couples :

$$A \times B = \{(a, b) \mid a \in A, b \in B\}$$

Alors :

$$\text{Card}(A \times B) = \text{Card}(A) \times \text{Card}(B)$$

Si je lance un dé et une pièce de monnaie, combien y a-t-il d'issues possibles ? Réponse : $6 \times 2 = 12$. En probabilité, cela permet de calculer les probabilités lorsqu'on effectue deux expériences indépendantes successives. Dans une urne, il y a une boule blanche et deux boules noires. Je tire une boule au hasard et je la repose, puis j'en tire une deuxième au hasard. Quelle est la probabilité de tirer deux boules noires de suite ? Réponse :

$$\frac{2 \times 2}{3 \times 3} = \frac{4}{9}$$

Soit A et B deux sous ensemble d'un ensemble fini Ω . Si A et B sont disjoints alors :

$$\text{Card}(A \cup B) = \text{Card}(A) + \text{Card}(B)$$

En particulier si A et B forment une partition de Ω ¹, alors $\text{Card}(B) = \text{Card}(\Omega) - \text{Card}(A)$. On en déduit par suite la formule suivante pour A et B non nécessairement disjoints :

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

Si je lance deux dés au hasard, quelle est la probabilité que l'un de mes dés soit 6 ? Il y a 36 issues équiprobables possibles notées (n, m) ou n (resp. m) est l'issue du premier (resp. deuxième) dé, avec $n, m \in \{1, \dots, 6\}$. Alors la probabilité de faire un 6 vaut :

$$\frac{6 + 6 - 1}{36} = \frac{11}{36}$$

Soit Ω un ensemble et A un sous ensemble de Ω . On note $\mathbb{1}_A$ la **fonction indicatrice** du sous-ensemble A , c'est-à-dire la fonction

$$\begin{aligned} \mathbb{1}_A : \Omega &\longrightarrow \{0, 1\} \\ a &\longrightarrow \begin{cases} 1 \text{ si } a \in A \\ 0 \text{ si } a \notin A \end{cases} \end{aligned}$$

Cette fonction a une utilité théorique importante, il sera primordial de connaître sa définition. Par exemple, la fonction

$$\mathbb{1}_{[a, b[} : \mathbb{R} \rightarrow \mathbb{R}$$

est la fonction qui vaut 1 sur l'intervalle semi-ouvert $[a, b[$ et 0 en dehors. Autre exemple,

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

Soit Ω un ensemble. On note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω , c'est-à-dire l'ensemble :

$$\mathcal{P}(\Omega) = \{A \mid A \text{ est un sous-ensemble de } \Omega\}$$

Alors :

$$\text{Card}(\mathcal{P}(\Omega)) = 2^{\text{Card}(\Omega)}$$

En effet, notons $\omega_1, \dots, \omega_n$ les éléments de Ω . Alors $\mathcal{P}(\Omega)$ peut être mis en bijection avec l'ensemble produits de n termes :

$$\{0, 1\} \times \{0, 1\} \times \dots \times \{0, 1\}$$

Via l'application :

$$\mathcal{P}(\Omega) \longrightarrow \{0, 1\} \times \{0, 1\} \times \dots \times \{0, 1\} = \{0, 1\}^{\text{Card}(\Omega)}$$

$$A \longrightarrow (\mathbb{1}_A(\omega_1), \mathbb{1}_A(\omega_2), \dots, \mathbb{1}_A(\omega_n))$$

Par exemple si $\Omega = \{1, 2, 3, 4\}$ alors le sous ensemble $A = \{2, 4\}$ est associé à $(0, 1, 0, 1)$. Et bien sur, on a :

$$\text{Card}(\{0, 1\} \times \{0, 1\} \times \dots \times \{0, 1\}) = 2^n = 2^{\text{Card}(\Omega)}$$

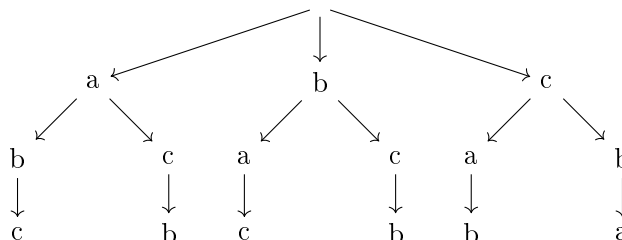
1. Les ensembles A et B forment une partition de Ω si A et B disjoints et $A \cup B = \Omega$. Cette définition se généralise à une famille de sous-ensembles (A_1, \dots, A_n) de Ω . Cette famille forme une partition de Ω si les ensembles A_1, \dots, A_n sont deux-à-deux disjoints et leur union est Ω . Autrement dit, on a découpé Ω en n morceaux A_1, \dots, A_n .

1.2 Permutation

Soit Ω un ensemble de n éléments. Par défaut, l'ensemble Ω n'est pas *ordonné* et sans répétition : on peut aussi bien écrire $\Omega = \{a, b, c\}$ que $\Omega = \{b, c, a\}$ ou encore $\Omega = \{a, b, c, a, b, b\}$. Quand on écrit "Soit $\omega_1, \dots, \omega_n$ les éléments de Ω ." c'est en fait une manière d'ordonner l'ensemble Ω : le premier est ω_1 , le deuxième est ω_2 , etc. Bien évidemment il n'y a pas qu'une manière d'ordonner Ω .

Définition 1.2. Une **permutation** de l'ensemble Ω est une façon de ranger les éléments de Ω de manière ordonnée. De manière équivalente, c'est une bijection de Ω sur l'ensemble $\{1, \dots, n\}$ (lorsque Ω n'est pas l'ensemble vide).

Par exemple, il y a six manières de ranger l'ensemble $\Omega = \{a, b, c\}$. On choisit le premier élément : il y a 3 possibilités parmi $\{a, b, c\}$. On choisit ensuite le deuxième : il y a deux possibilités parmi les deux éléments à placer restants. Enfin le dernier élément est automatiquement en dernière place : Donc on a $3 \times 2 \times 1 = 6$ possibilités.



Théorème 1.1. Il existe exactement $n!$ permutations de l'ensemble Ω , où $n!$ (prononcez "factorielle n ") est le nombre défini par :

$$n! = 1 \times 2 \times \dots \times n$$

Par convention, $0! = 1$. Il existe une seule manière d'ordonner un ensemble à 0 éléments.

Démonstration. Le raisonnement est analogue au cas précédent : on a n possibilités pour choisir le premier élément, puis $n - 1$ pour le deuxième, etc. Au total il y a $n \times (n - 1) \times \dots \times 1 = n!$ possibilités d'ordonner l'ensemble Ω . \square

Combien existe-t-il d'anagramme du mot *PROBA*? Toutes les lettres sont différentes, donc chaque permutation donne un mot différent. Par suite il existe $5! = 120$ anagrammes du mot *PROBA*.

Même question avec le mot *ECOLE*. Cette fois il faut faire attention car si l'on échange les deux "E" cela redonne le même mot. Notons E_1 et E_2 les deux "E". Comme précédemment, il existe 120 permutations pour le mot E_1COLE_2 . En revanche, chaque mot est compté en double puisque pour un anagramme du mot *ECOLE* (par exemple *CEOEL*) on peut construire CE_2OE_1L et CE_1OE_2L . Il existe donc $120/2 = 60$ possibilités.

Même question avec le mot *MISSISSIPPI*. Il y a 11 lettres mais beaucoup d'entre elles se répètent : $1 \times M + 4 \times I + 4 \times S + 2 \times P$. On écrit alors $MI_1S_1S_2I_2S_3S_4I_3P_1P_2I_4$. Ce mot a bien 11! permutations distinctes. Essayons de trouver le nombre de permutations de $MI_1SSI_2SSI_3P_1P_2I_4$ (on considère maintenant que tous les S sont identiques). Étant donné une permutation de ce mot, on peut en construire exactement 4! pour le mot $MI_1S_1S_2I_2S_3S_4I_3P_1P_2I_4$. Par exemple à partir de $SSSSMI_1I_2I_3I_4P_1P_2$ on peut construire :

- $S_1 S_2 S_3 S_4 M I_1 I_2 I_3 I_4 P_1 P_2$
- $S_2 S_1 S_4 S_3 M I_1 I_2 I_3 I_4 P_1 P_2$
- ...

En fait on peut permuter tous les S entre eux, cela me donnera exactement 24 permutations distinctes du mot $M I_1 S_1 S_2 I_2 S_3 S_4 I_3 P_1 P_2 I_4$. Donc il y a exactement $\frac{11!}{4!}$ permutations du mot $M I_1 S S I_2 S S I_3 P_1 P_2 I_4$. On peut réitérer avec les lettres restantes. Au final, le mot *MISSISSIPPI* aura :

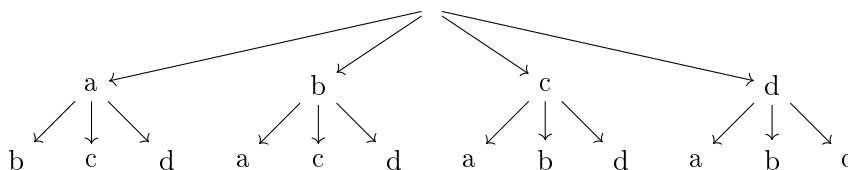
$$\frac{11!}{1!4!4!2!} = 34\,650$$

anagrammes.

1.3 Arrangement

Définition 1.3. Un **arrangement** de k éléments pris parmi n éléments d'un ensemble Ω est une suite ordonnée de k éléments distincts de Ω .

Par exemple (a, c, f) et (c, f, a) et (b, a, f) sont des arrangements distincts de 3 éléments parmi les 6 éléments $\{a, b, c, d, e, f\}$. Un arrangement de n éléments parmi un ensemble Ω à n éléments est simplement une permutation de Ω . Par exemple il y a 2 possibilité d'arranger 2 éléments parmi 4 : on a 4 possibilité pour choisir le premier, puis 3 pour le deuxième.



Théorème 1.2. Il y a exactement :

$$A_n^k = n(n-1)(n-2)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

manières d'arranger k éléments parmi n .

Démonstration. Le raisonnement est analogue au cas précédent : on a n possibilités pour choisir le premier élément, puis $n-1$ pour le deuxième, etc. Au total il y a $n \times (n-1) \times \dots \times (n-k+1)$ possibilités d'arranger k éléments parmi n . \square

Par exemple, sur une course de 20 participants, il y a :

$$20 \times 19 \times 18 = 6840$$

podiums distincts possibles.

1.4 Combinaison

Définition 1.4. Une **combinaison** de k éléments parmi un ensemble Ω à n éléments est un sous-ensemble A de Ω contenant k éléments.

La différence entre une combinaison et un arrangement est que la combinaison est un ensemble, donc *non ordonnée*. Dans un problème de dénombrement où l'ordre importe (combinaison d'un coffre fort, podium, etc) on utilisera plutôt la notion d'arrangement. Dans un problème où l'ordre n'importe pas (nombre de boules blanches tirées lors d'un tirage avec remise, probabilité d'avoir 2 filles sachant qu'on a 5 enfants, etc) c'est la notion de combinaison qui sera importante.

Le nombre de combinaisons de k éléments parmi un ensemble à n éléments (on appelle ce nombre " k parmi n ") est noté $\binom{n}{k}$.

Théorème 1.3. On a :

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

Démonstration. Étant donné une combinaison de k éléments, on peut former exactement $k!$ arrangements distincts en permutant les k éléments choisis. On obtient ainsi l'égalité :

$$k! \binom{n}{k} = A_n^k = \frac{n!}{(n-k)!}$$

□

Si $k > n$ on stipule que :

$$\binom{n}{k} = 0$$

En général, toutes les formules de dénombrement ont deux preuves : une preuve calculatoire, et une preuve utilisant uniquement des outils de dénombrement. Il est bon de savoir jongler entre les deux : la preuve calculatoire est généralement (mais pas tout le temps) plus longue, tandis que la preuve par dénombrement est souvent plus jolie et intuitive (mais gare aux erreurs de raisonnement!).

Théorème 1.4. Soit $x, y \in \mathbb{R}$ et $n \in \mathbb{N}$. Alors :

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

Démonstration. On peut faire une preuve par récurrence sur n mais procédons plutôt par une preuve combinatoire. On développe :

$$\begin{aligned} (x + y)^n &= (x + y) \times (x + y) \times \dots \times (x + y) \\ &= \sum_{k=0}^n a_k x^k y^{n-k} \end{aligned}$$

Lorsqu'on développe on se retrouve avec des monômes de la forme $x^k y^{n-k}$ pour k variant entre 0 et n . Pour obtenir un tel monôme, on choisit k facteurs parmi les n et on développe selon " x ", et pour les $(n - k)$ facteurs restant on développe selon " y ". Il ne restera que le monôme $x^k y^{n-k}$. Le nombre de façon de choisir k facteurs parmi n est exactement le coefficient binomial k parmi n , d'où :

$$a_k = \binom{n}{k}$$

□

En particulier :

$$2^n = \sum_{k=0}^n \binom{n}{k}$$

2 Espace probabilisé

2.1 Un peu d'heuristique

La théorie des probabilités a pour but de quantifier la notion d'*incertitude*. La notion d'incertitude reflète un manque d'*information* qui nous empêche de prédire l'issue d'une expérience avec une absolue certitude.

Prenons l'exemple du lancer de pièce. Intuitivement, on tombera sur pile environ une fois sur 2. Mais qu'en est-il vraiment ? Le lancer de pièce, en tant que système physique, peut être considéré comme déterministe. Si je pouvais décrire avec une absolue précision l'action exercée par ma main sur la pièce, je pourrais en déduire sa trajectoire exacte et prédire sur quelle face la pièce tomberait. Supposons que la trajectoire de la pièce dépend d'un paramètre $\theta \in [0, 1]$. Puisque l'on a aucune information sur le paramètre θ , on suppose que celui-ci est choisi *au hasard* selon une certaine loi (par exemple une loi uniforme). Dans ce cas, on observerait que pour environ la moitié des valeurs de θ la pièce tombera sur pile, et pour l'autre moitié elle tombera sur face.

Prenons l'exemple d'un joueur de poker. Étant donnée son jeu et les cartes sur la table, il doit décider si oui ou non il décide de continuer à jouer. Comme il ne connaît pas la carte de ses adversaires, il fait l'hypothèse que le paquet de carte qui sert au jeu a été choisi *au hasard* uniformément parmi tous les paquets de cartes possibles (52!). De là, il est possible de calculer la proportion de paquets de cartes qui lui sont favorables.

La **théorie des probabilités** permet de quantifier un *manque d'information* sur les paramètres d'une expérience considérée. Ce manque d'information se répercute en une *incertitude* sur l'issue de l'expérience. Se donner un cadre pour effectuer des probabilités, c'est se donner une *modélisation de l'incertitude* à laquelle on fait face.

Lorsque l'on joue au Uno, les règles du jeu font que l'empilement des cartes à la fin d'une partie présente beaucoup de cartes de même couleur qui se suivent. En le mélangeant à la main, certaines de ces enchaînements ne sont pas cassés et le paquet de carte mélangé présente généralement une proportion de cartes de même valeur qui se suivent plus importante que si l'on avait effectué un mélange "parfait". Sans connaissance préalable sur le paquet mis à part que celui ci vient d'être mélangé suite à une précédente partie, un joueur expérimenté modélisera le paquet de carte comme un paquet choisi *au hasard* parmi les 52! paquets possibles, mais ce choix au hasard ne sera pas uniforme sur tous les paquets. Ceux présentant des cartes de même valeur qui se suivent seront considérés avec une plus grande probabilité d'apparition.

D'après nos exemples précédents nous avons besoin de trois ingrédients :

Il faut se donner l'*univers* des configurations possibles pour l'expérience ou le jeu considéré.

Pour le lancer de dés, cela peut être la trajectoire de ma main. Pour un jeu de carte, c'est l'ensemble des jeux de cartes possibles. Généralement c'est un espace gros et compliqué qu'il est difficile voire impossible de décrire précisément.

Il faut définir la notion d'*information*. C'est l'ensemble des événements que l'on souhaite considérer lors d'une expérience.

Par exemple, on fait un sondage au près d'une population. On demande à chaque personne de préciser sa catégorie d'âge : entre 0 et 20 ans, entre 20 et 50 ans, plus de 50 ans. Ici l'univers est l'ensemble des personnes sondées. Notre expérience nous permet de quantifier la probabilité d'apparition d'événements comme : "La personne a moins de 50 ans". En revanche nous n'avons

pas accès à des quantités comme "La personne a entre 40 et 60 ans" et encore moins à "La personne est un homme". Cette notion d'*information* semble un peu superflue car il paraît toujours possible de "grossir" l'expérience afin d'accéder à une information dont on ne dispose pas. Il y a au moins trois raisons à ne pas procéder ainsi :

- Cela dénature/complique la modélisation. Dans le cas du lancer de dé, on se fiche pas mal de savoir si le dé a fait trois tours sur lui-même avant de toucher le sol. L'information essentielle est seulement la valeur finale du dé.
- Lorsqu'on modélise un jeu, ou plus généralement une quantité qui évolue au cours du temps (le cours de la bourse, une particule quantique, chaîne de Markov, etc) cette notion d'information prend toute son importance pour la compréhension du jeu (les règles sont-elles en ma faveur ?) et l'élaboration d'une stratégie (dois-je piocher une carte, vendre mes actions ?). Le cadre général est celui de la théorie des *processus stochastiques*.
- Il y a certaines obstructions mathématiques qui empêchent de considérer la probabilité de réalisation de certains ensembles dans le cas où l'univers Ω est indénombrable, comme le montre le célèbre paradoxe de Banach-Tarski.

Il faut définir une notion de *mesure de probabilité*. C'est elle qui quantifie la probabilité de réalisation d'un événement donné.

Dans le cas d'un paquet de carte bien mélangé, on peut par exemple choisir une probabilité uniforme sur tous les paquets de cartes, de sorte que chaque paquet de carte a une probabilité égale à $1/(52!)$, mais s'il est mal mélangé, la probabilité peut être choisie différemment. La modélisation d'une expérience impose donc de choisir avec quelle probabilité apparaît chaque événement. Cette mesure est censé refléter aux mieux la réalité physique de l'expérience ce qui peut s'avérer compliqué. Un test statistique permet de vérifier la concordance entre un modèle probabiliste théorique et une expérience réelle répétée un grand nombre de fois.

2.2 Quelques définitions et exemples

D'après la discussion précédente, un espace probabilisé est donc constitué de trois ingrédients.

Définition 2.1. Un espace probabilisé est un triplet $(\Omega, \mathcal{F}, \mathbb{P})$, où :

- Ω est un ensemble
- \mathcal{F} est une tribu sur Ω .
- \mathbb{P} est une mesure de probabilité

Détaillons plus précisément les trois points de la définition :

- Ω est l'**univers**, c'est-à-dire l'ensemble des configurations possibles de mon expérience. Mathématiquement c'est simplement un ensemble.
- \mathcal{F} représente l'information que l'on peut acquérir au cours de l'expérience. Un élément de \mathcal{F} est appelé un **événement**, c'est un sous-ensemble de l'univers Ω . L'ensemble \mathcal{F} est donc un ensemble d'événements. Moralement, il est possible d'appliquer certaines opérations entre événements : union, intersection, différence, complémentaire, etc. Autrement dit l'ensemble \mathcal{F} est stable par un certain nombre d'opérations.

Définition 2.2. Une **tribu** \mathcal{F} sur un espace Ω est un sous-ensemble de $\mathcal{P}(\Omega)$ (l'ensemble des parties de Ω) tel que :

- $\Omega \in \mathcal{F}$ (l'univers est un événement)
- Si $A \in \mathcal{F}$ alors $\bar{A} \in \mathcal{F}$ (stabilité par passage au complémentaire)
- Si $(A_n)_{n \geq 0} \in \mathcal{F}$ alors $\bigcup_{n \geq 0} A_n \in \mathcal{F}$ (stabilité par union dénombrable)

La paire (Ω, \mathcal{F}) est un **espace mesurable**.

- \mathbb{P} permet de quantifier la probabilité de réalisation d'un événement donné. À un événement A donné elle associe un nombre entre 0 et 1, noté $\mathbb{P}(A)$ qui reflète la probabilité de réalisation de l'événement A . Pour être cohérent avec une notion intuitive de *mesure*, \mathbb{P} doit vérifier quelques propriétés :

Définition 2.3. Une **mesure de probabilité** \mathbb{P} sur un espace mesurable (Ω, \mathcal{F}) est une application :

$$\begin{aligned} \mathbb{P} : \mathcal{F} &\longrightarrow [0, 1] \\ A &\longrightarrow \mathbb{P}(A) \end{aligned}$$

Telle que :

- $\mathbb{P}(\Omega) = 1$ (l'univers est un événement de probabilité 1)
- Si $(A_n)_{n \geq 0}$ est une famille dénombrable d'événements deux à deux disjoints alors :

$$\mathbb{P} \left(\bigcup_{n=0}^{+\infty} A_n \right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$$

Si A est un événement d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ tel que $\mathbb{P}(A) = 1$, on dit que A est réalisé *presque-surement* (on écrira parfois p.s.). Inversement, si A est un événement d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ tel que $\mathbb{P}(A) = 0$, on dira que A est un événement *négligeable*. Cette terminologie est importante et sera utilisée de nombreuses fois dans la suite du cours.

Voyons ensemble quelques exemples fondamentaux de tribus.

Définition 2.4. La tribu $\{\emptyset, \Omega\}$ est appelée la **tribu triviale**. C'est la plus petite tribu sur Ω que l'on puisse considérer. La tribu $\mathcal{P}(\Omega)$ est appelée **tribu discrète**. C'est la plus grosse tribu sur Ω que l'on puisse considérer.

Lorsque Ω est dénombrable, on considère généralement la tribu discrète sur Ω .

Définition 2.5. Soit \mathcal{C} un ensemble de parties Ω . On note $\sigma(\mathcal{C})$ la plus petite tribu contenant \mathcal{C} . C'est l'intersection de toutes les tribus contenant \mathcal{C} .

Par exemple, si Ω est dénombrable et \mathcal{C} est l'ensemble des singletons, il est facile de voir que $\sigma(\mathcal{C}) = \mathcal{P}(\Omega)$.

Soit $\Omega = [0, 1]$. Si I est un intervalle d'extrémité a et b (pas forcément ouvert ni fermé), on pose $\mu(I) = b - a$. Cette définition est "cohérente" avec la notion d'espace probabilisé :

$$\mu(\Omega) = 1$$

I_1 et I_2 sont des intervalles disjointes telles que $I_1 \cup I_2$ est un intervalle (par exemple $I_1 = [a, b]$ et $I_2 =]b, c]$) alors :

$$\mu(I_1 \cup I_2) = \mu(I_1) + \mu(I_2)$$

Définition 2.6. On note $\mathcal{B}([0, 1])$ la tribu engendrée par les intervalles inclus dans $[0, 1]$. On l'appelle la **tribu borélienne** sur $[0, 1]$. De même, on note $\mathcal{B}(\mathbb{R})$ la tribu engendrée par les intervalles dans \mathbb{R} .

Théorème 2.1. (*Extension de Carathéodory*) il existe une unique mesure de probabilité (encore notée μ) sur $(\Omega, \mathcal{B}([0, 1]))$ telle que :

- $([0, 1], \mathcal{B}([0, 1]), \mu)$ soit un espace probabilisé.
- $\mu(I) = b - a$ pour tout intervalle I d'extrémité a et b .

Démonstration. Admis. □

Autrement dit, si l'on décide que la mesure d'un intervalle est exactement sa longueur, alors on peut étendre cette mesure à toute la tribu engendrée par les intervalles : les unions dénombrables d'intervalles, de singletons, ... Et bien plus encore !

En revanche il existe des sous-ensembles de Ω qui ne sont pas dans $\mathcal{B}([0, 1])$ (un contre-exemple est compliqué à exhiber). En fait on peut montrer qu'il est impossible d'étendre de manière cohérente cette mesure à la tribu $\mathcal{P}([0, 1])$ tout entier (contre-exemple de Vitali). C'est pour cela que l'on doit se restreindre à la tribu borélienne.

2.3 Probabilité conditionnelle et indépendance

Lorsque l'on souhaite modéliser deux quantités distinctes par un modèle probabiliste, il arrive souvent que ces deux quantités soient *corrélées*. Par exemple, la recette d'un marchand de glace sur une journée est fortement corrélée à la température. Connaître finement cette dépendance permet au marchand de glace de mieux ajuster ses stocks pour ne pas être en rupture dans les jours de canicule et avoir des pertes lors de journées plus fraîches. Considérons les événements suivants :

- A : le vendeur vend plus de 100 glaces au cours de la journée
- B : la température de la journée a dépassé les 20 degrés.

Le vendeur modélise sa situation de la façon suivante. Il estime que :

- $\mathbb{P}(A) = 0.5$
- $\mathbb{P}(B) = 0.4$
- $\mathbb{P}(A \cap B) = 0.3$

Le soir, le journal météo du lendemain annonce des températures supérieures à 20 degrés. Comment quantifier la probabilité que le vendeur vende plus de 100 glaces le lendemain ? Notons $\mathbb{P}(A|B)$ cette probabilité. Alors on a :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = 0.75$$

L'univers des possibles pour le lendemain n'est plus Ω tout entier mais B , l'ensemble des configurations pour lesquelles la température dépasse 20 degrés au cours de la journée. Nous avons donc en quelque sorte changé l'espace probabilisé :

$$(\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (B, \mathcal{F}_B, \mathbb{P}(\cdot|B))$$

Où :

$$\mathcal{F}_B = \{A \cap B \mid A \in \mathcal{F}\}$$

Dans cet exemple, la notion de dépendance s'explique par une notion de *causalité* :

Il fait beau \Rightarrow les gens achètent des glaces \Rightarrow Le vendeur fait des recettes

Mais ce n'est pas toujours le cas. Par exemple, on fait une enquête sur des gens habitant près de pylônes électriques². On note :

- A : "la personne tombe malade plus de 5 fois au cours de l'année"
- B : "la personne habite à moins de 200m d'un pylône électrique"

Sur l'ensemble d'une ville, on remarque que :

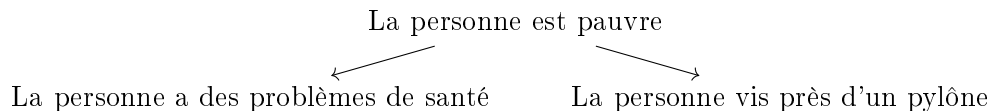
- $\mathbb{P}(A) = 0.2$
- $\mathbb{P}(B) = 0.1$
- $\mathbb{P}(A \cap B) = 0.04$

Alors :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = 0.4$$

Autrement dit, une personne a deux fois plus de chance de tomber souvent malade si elle habite près d'un pylône électrique. Faut-il en conclure que les pylônes électriques sont dangereux pour la santé ?

Non, car en réalité la catégorie de population vivant près de pylônes électriques est plus pauvre en moyenne, et une personne pauvre a moins accès aux soins. Ici, les événements A et B sont corrélés entre eux mais sans forcément de lien de causalité. En fait, ces deux événements sont corrélés à un troisième événement C : "la personne est en dessous du seuil de pauvreté". Et on a plutôt les relations de cause-à-effet suivantes entre A et B



Exemple plus flagrant : 100% des gens qui boivent de l'eau finissent par décéder... Ne buvez pas d'eau !

Définition 2.7. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et B un événement de mesure non-nulle. Alors pour A un événement quelconque on appelle **probabilité conditionnelle** de A sachant B la quantité :

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

Elle indique la probabilité qu'un événement A se réalise sachant que l'événement B est réalisé.

2. Adapté d'une vraie enquête.

Définition 2.8. On dit qu'une famille $\{B_1, \dots, B_n\}$ est un **système complet d'événements** si :

- $\forall i \in \{1, \dots, n\}, \mathbb{P}(B_i) \neq 0$
- $\forall i, j \in \{1, \dots, n\}, \mathbb{P}(B_i \cap B_j) = 0$
- $\mathbb{P}(\bigcup_{i=1}^n B_i) = 1$ ou de manière équivalente $\sum_{i=1}^n \mathbb{P}(B_i) = 1$

Autrement dit la famille $\{B_1, \dots, B_n\}$ forme une partition *probabiliste* de l'univers Ω . Il est possible d'adapter la définition pour une famille dénombrable d'événements.

Par exemple, si $0 < \mathbb{P}(A) < 1$ alors la famille $\{A, \bar{A}\}$ forme un système complet d'événements. De même, sous réserve que les événements suivants soient de probabilités non-nulles, la famille suivante :

$$\{A \cap B, \bar{A} \cap B, A \cap \bar{B}, \bar{A} \cap \bar{B}\}$$

est un système complet d'événements.

Théorème 2.1 (Formule des probabilités totales). *Soit $\{B_1, \dots, B_n\}$ un système complet d'événements, et A un événement quelconque. Alors :*

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(B_i) \mathbb{P}(A|B_i)$$

Théorème 2.2 (Formule de Bayes). *Soit A et B deux événements de probabilité non-nulles. Alors on a l'identité :*

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

Si la famille $\{A_1, \dots, A_n\}$ est un système complet d'événements, alors pour $1 \leq i \leq n$ on a :

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i) \mathbb{P}(B|A_i)}{\sum_{i=1}^n \mathbb{P}(A_i) \mathbb{P}(B|A_i)}$$

La formule de Bayes permet, en connaissant les *effets*, de remonter à une probabilité sur les *causes*. Elle est à la base de *l'inférence bayésienne*.

La notion de probabilité conditionnelle nous mène naturellement à la notion d'*indépendance*. Soient B un événement de probabilité non-nulle et A un événement quelconque. L'événement A est indépendant de l'événement B si la probabilité qu'il soit réalisé ne dépend pas du fait que B soit réalisé ou non. Autrement dit :

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

Si on lance deux dés, la connaissance de la valeur du premier dé n'influe en rien sur la valeur du deuxième dé. Les événements "le premier dé est un 6" et "le deuxième est pair" sont indépendants.

Définition 2.9. Deux événements A et B sont dits **indépendants** si :

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

On note alors $A \perp B$. De manière générale, deux tribus \mathcal{F} et \mathcal{G} sont dites indépendantes si :

$$\forall A \in \mathcal{F}, \forall B \in \mathcal{G}, \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

et l'on note :

$$\mathcal{F} \perp \mathcal{G}$$

On peut également parler d'indépendance pour une famille d'événements.

Définition 2.10. Les événements A_1, \dots, A_n sont dits mutuellement indépendants si : pour toute sous-partie I de $\{1, \dots, n\}$ on a :

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

De même, les tribus $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont mutuellement indépendantes si pour tout événements A_1, \dots, A_n telle que :

$$A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$$

Les événements A_1, \dots, A_n sont indépendants.

On remarque facilement que les événements A_1, \dots, A_n sont indépendants si et seulement si les tribus $\sigma(A_1), \dots, \sigma(A_n)$ sont indépendantes. En effet, pour un événement A on a :

$$\sigma(\{A\}) = \{\emptyset, A, \bar{A}, \Omega\}$$

Par exemple, si A et B sont des événements indépendants alors

$$\begin{aligned} \mathbb{P}(\bar{A} \cap B) &= \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(\bar{A})\mathbb{P}(B) \end{aligned}$$

Théorème 2.3 (Lemme de coalition). Soit $\mathcal{F}_1, \dots, \mathcal{F}_n, \mathcal{G}_1, \dots, \mathcal{G}_m$ des tribus mutuellement indépendantes. Alors :

$$\sigma\left(\bigcup_{i=1}^n \mathcal{F}_i\right) \perp \sigma\left(\bigcup_{j=1}^m \mathcal{G}_j\right)$$

Démonstration. Admis. □

Par exemple, soient A_1, A_2, A_3, A_4, A_5 des événements mutuellement indépendants. Alors :

$$(A_1 \cup A_3) \perp (A_2 \cap (\bar{A}_4 \cup A_5))$$

De manière générale, si l'on sépare en deux groupes une famille d'événements mutuellement indépendants, alors un événement construit à partir du premier groupe et des opérations \cap, \cup et $-$ sera indépendant d'un événement construit à partir du deuxième groupe et des mêmes opérations. Il faut comprendre l'indépendance mutuelle en terme d'indépendance d'*informations*.

L'information apportée par la connaissance de A_1 et A_3 est indépendante de l'information engendrée par la connaissance de A_2, A_4 et A_5 .

On peut aussi définir l'indépendance d'une famille dénombrable d'événements.

Définition 2.11. Une famille $(A_n)_{n \in \mathbb{N}}$ d'événements est dite mutuellement indépendante si pour toute sous-partie *finie* I de \mathbb{N} on a :

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

De même, les tribus $(\mathcal{F}_n)_{n \in \mathbb{N}}$ sont mutuellement indépendantes si pour toute famille d'événements $(A_n)_{n \in \mathbb{N}}$ telle que :

$$\forall n \in \mathbb{N}, A_n \in \mathcal{F}_n$$

les événements $(A_n)_{n \in \mathbb{N}}$ sont mutuellement indépendants.

2.4 Quelques théorèmes limites

Le but de cette section est de s'intéresser à certaines probabilités "limites". Nous rappelons la définition des limites supérieures et inférieures d'une suite d'ensembles.

Définition 2.1. Soit $(A_n)_{n \geq 0}$ une suite d'ensembles. On définit la limite supérieure de la suite comme

$$\limsup_{n \geq 0} A_n = \bigcap_{n \geq 0} \bigcup_{k \geq n} A_k,$$

et la limite inférieure de la suite comme

$$\liminf_{n \geq 0} A_n = \bigcup_{n \geq 0} \bigcap_{k \geq n} A_k.$$

Les interprétations de ces deux quantités sont les suivantes :

- Un élément $\omega \in \Omega$ appartient à l'ensemble $\limsup_{n \geq 0} A_n$ si et seulement si ω appartient à une infinité d'événements de la suite $(A_n)_{n \geq 0}$.
- Un élément $\omega \in \Omega$ appartient à l'ensemble $\liminf_{n \geq 0} A_n$ si et seulement si ω appartient à tous les événements de la suite $(A_n)_{n \geq 0}$ à partir d'un certain rang.

Nous commençons avec le théorème fondamental suivant.

Théorème 2.2 (Borel–Cantelli). Soit $(A_n)_{n \geq 0}$ une suite d'événements d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Si

$$\sum_{n \geq 0} \mathbb{P}(A_n) < +\infty,$$

alors

$$\mathbb{P}(\limsup_{n \geq 0} A_n) = 0.$$

Si les événements $(A_n)_{n \geq 0}$ sont mutuellement indépendants et que

$$\sum_{n \geq 0} \mathbb{P}(A_n) = +\infty,$$

alors

$$\mathbb{P}(\limsup_{n \geq 0} A_n) = 1.$$

Démonstration. La preuve n'est pas difficile mais nous l'admettrons pour ce cours. \square

Autrement dit, si la série de terme $\mathbb{P}(A_n)$ converge alors les événements de la suite $(A_n)_{n \geq 0}$ ne se réalisent presque sûrement qu'un nombre fini de fois. Inversement, si l'on ajoute l'hypothèse d'indépendance de la suite $(A_n)_{n \geq 0}$ et que la série de terme $\mathbb{P}(A_n)$ diverge alors les événements de la suite $(A_n)_{n \geq 0}$ se réalisent infiniment souvent presque sûrement. Ce théorème a une importance fondamentale pour prouver des convergences presque-sûres de variables aléatoires, tel que la loi forte des grands nombres.

Dans la suite, $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé et $(\mathcal{F}_n)_{n \geq 0}$ une suite de sous-tribus de \mathcal{F} mutuellement indépendantes.

Définition 2.2 (Tribu asymptotique). On définit la tribu asymptotique

$$\mathcal{F}_\infty = \bigcap_{n \geq 0} \sigma \left(\bigcup_{k \geq n} \mathcal{F}_k \right).$$

La tribu asymptotique regroupe les événements qui ne dépendent que de la "queue" de la suite de tribus $(\mathcal{F}_n)_{n \geq 0}$, c'est à dire qui ne dépend pas des premiers éléments de la suite de tribus. Un événement de \mathcal{F}_∞ est appelé un événement *asymptotique*.

Théorème 2.3 (Loi du 0–1 de Kolmogorov). Soit A un événement de la tribu asymptotique \mathcal{F}_∞ . Alors

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

Démonstration. La preuve n'est pas difficile mais nous l'admettrons pour ce cours. Elle repose sur le fait que la tribu \mathcal{F}_∞ est indépendante d'elle-même. \square

Autrement dit, un événement de la tribu asymptotique se réalise presque sûrement, ou bien presque jamais. Il n'y a pas de possibilité d'entre-deux. En revanche il peut être difficile de savoir dans quel cas on se trouve. Des exemples typique d'événements asymptotiques sont les suivants : "Est ce qu'une suite de variables aléatoires indépendantes est bornée ? converge ?." Nous verrons quelques applications de ces deux théorèmes en cours et dans les exercices.

3 Variable aléatoire

3.1 Quelques définitions

Une variable aléatoire X est une application à qui une configuration, une issue donnée (donc un élément de Ω) associe un objet, un nombre, une couleur, etc.

Une variable aléatoire modélise les différentes valeurs que peut prendre le résultat d'une expérience au caractère aléatoire.

Par exemple on peut considérer les variables aléatoires associées aux expériences suivantes.

- Le lancer de dé. On considère la variable aléatoire $X : \Omega \rightarrow \{1, 2, 3, 4, 5, 6\}$ l'application à qui une configuration donnée associe la valeur du dé.
- Le pile ou face. On considère la variable aléatoire $X : \Omega \rightarrow \{\text{Pile}, \text{Face}\}$ l'application à qui une configuration donnée associe la valeur de la pièce.
- Le jeu de fléchette. On considère la variable aléatoire $X : \Omega \rightarrow \mathbb{R}_+$ l'application à qui une configuration donnée associe la distance de la fléchette au centre de la cible.
- Le lancé de deux dés. On considère la variable aléatoire $X : \Omega \rightarrow \{2, \dots, 12\}$ l'application à qui une configuration donnée associe la somme des deux dés.

Pour obtenir des informations sur la probabilité qu'une variable aléatoire appartienne à un certain ensemble de valeurs, elle doit satisfaire certaines contraintes. Par exemple pour le pile ou face, il faut s'assurer que le sous-ensemble

$$\{X = \text{"Pile"}\} = \{\omega \in \Omega \mid X(\omega) = \text{"Pile"}\}$$

soit bien un événement, c'est-à-dire un élément de la tribu \mathcal{F} . En effet, la proposition "la probabilité que la pièce tombe sur pile est de $1/2$ " s'écrit mathématiquement :

$$\mathbb{P}(X = \text{"Pile"}) = 1/2.$$

Dans le cas où X est à valeur dans un espace dénombrable E , il n'y a pas de difficultés. On exigera que pour tout $x \in E$ l'ensemble :

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

soit bien un événement (donc appartienne à la tribu \mathcal{F}). Puisque tout sous-ensemble U de E peut s'écrire comme une union dénombrable de singletons, alors :

$$\{X \in U\} = \bigcup_{x \in U} \{X = x\}$$

Et donc :

$$\forall U \subset E, \{X \in U\} \subset \mathcal{F}$$

Si X est à valeur réelles (ou dans un espace indénombrable), les choses se compliquent un peu. Si on demande seulement que les ensembles $\{X = x\}$ soient des événements, on ne peut pas affirmer que l'ensemble :

$$\{X \in [0, 1]\} = \bigcup_{x \in [0, 1]} \{X = x\}$$

soit un événement puisqu'une tribu est seulement stable par union dénombrable, et l'ensemble $[0, 1]$ est indénombrable. A l'inverse, on ne peut pas (pour des raisons mathématiques³) exiger que pour tout sous-ensemble U de \mathbb{R} , l'ensemble :

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

3. Cette définition serait trop forte. Par exemple on pourrait montrer qu'avec cette définition, il n'existe pas de loi uniforme sur $[0, 1]$.

soit un événement. Il faut restreindre la collection de sous-ensembles que l'on s'autorise à regarder. Le minimum que l'on puisse exiger est de pouvoir donner un sens à des propositions comme "la probabilité que ma fléchette arrive à moins de 20cm de la cible est de 1/2". Pour cela il faut ici que l'ensemble :

$$\{X \leq 20\} = \{\omega \in \Omega \mid X(\omega) \leq 20\}$$

soit un événement. De manière générale, on exigera que pour tout $x \in \mathbb{R}$, l'ensemble :

$$\{X \leq x\} = \{X \in]-\infty, x]\}$$

soit un événement. A partir de là on montre facilement⁴ que les sous-ensembles suivants de Ω sont des événements :

$$\begin{aligned} \{X < x\} &= \bigcup_{n \geq 0} \left\{ X \leq x - \frac{1}{n} \right\} \\ \{X > x\} &= \overline{\{X \leq x\}} \\ \{X \geq x\} &= \overline{\{X < x\}} \\ \{x \leq X \leq y\} &= \{X \in [x, y]\} = \{X \geq x\} \cap \{X \leq y\} \\ \{X = x\} &= \{x \leq X \leq x\} \end{aligned}$$

En particulier, l'ensemble :

$$\{X \in U\}$$

est un événement lorsque U est un intervalle, mais aussi une union dénombrable d'intervalle, de singletons, ... Et bien plus encore ! En fait, l'ensemble des U pour lesquelles cet ensemble est un événement contient toute la tribu engendrée par les intervalles, que l'on appelle la tribu borélienne sur \mathbb{R} , notée $\mathcal{B}(\mathbb{R})$ (on a déjà vu ce concept au chapitre précédent).

Essayons de généraliser. Soit X une variable aléatoire à valeur dans un espace quelconque E . Il faut décider quels sont les sous-ensembles U de E pour lesquelles l'ensemble $\{X \in U\}$ soit un événement (donc un élément de la tribu \mathcal{F}). Cela revient à choisir une ensemble de parties de E que l'on notera \mathcal{G} , et telle que :

$$\forall U \in \mathcal{G}, \{X \in U\} \in \mathcal{F}$$

Si $U \in \mathcal{G}$ et $(U_n)_{n \in \mathbb{N}}$ est une famille d'éléments de la collection \mathcal{G} alors :

$$\begin{aligned} \{X \in E\} &= \Omega \in \mathcal{F}, \\ \{X \in \overline{U}\} &= \overline{\{X \in U\}} \in \mathcal{F}, \\ \left\{ X \in \bigcup_{n=0}^{+\infty} U_n \right\} &= \bigcup_{n=0}^{+\infty} \{X \in U_n\} \in \mathcal{F}, \end{aligned}$$

De sorte qu'il est naturel d'exiger que \mathcal{G} soit une tribu sur E .

Définition 3.1. Soit X une application d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeur dans un espace mesurable (E, \mathcal{G}) . On dit que X est une **variable aléatoire** à valeur dans (E, \mathcal{G}) si :

$$\forall U \in \mathcal{G}, \{\omega \in \Omega \mid X(\omega) \in U\} \in \mathcal{F}$$

Le théorème suivant détaille le cas où tribu \mathcal{G} est engendrée par une ensemble \mathcal{C} de parties de E stable par intersection.

4. L'ensemble des événements forment une tribu. Donc l'intersection, l'union et le complémentaire d'événements sont encore des événements.

Théorème 3.1 (Lemme des classes monotones). Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (E, \mathcal{G})$ une application. Supposons que la tribu \mathcal{G} soit engendrée par un ensemble \mathcal{C} de parties de E stable par intersection. C'est-à-dire :

$$\mathcal{G} = \sigma(\mathcal{C})$$

$$\forall U, V \in \mathcal{C}, \quad U \cap V \in \mathcal{C}$$

Alors X est une variable aléatoire si et seulement si :

$$\forall U \in \mathcal{C}, \quad \{\omega \in \Omega \mid X(\omega) \in U\} \in \mathcal{F}$$

Démonstration. Admis. □

Autrement dit, il suffit de vérifier que $\{X \in U\}$ est un événement pour tout $U \in \mathcal{C}$ pour en déduire que c'est un événement pour tout $A \in \mathcal{G}$. Cela nous permet de préciser la définition d'une variable aléatoire dans les deux cas qui nous intéressent les variables : aléatoires réelles et les variables aléatoires discrètes.

Définition 3.1 (Cas discret). Soit X une application d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans un espace dénombrable E , muni de la tribu discrète. C'est à dire

$$X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (E, \mathcal{P}(E)).$$

On dit que X est une **variable aléatoire discrète** à valeurs dans E , si pour tout $e \in E$, l'ensemble $\{X = e\}$ est un événement de la tribu \mathcal{F} .

Définition 3.2 (Cas réel). Soit X une application d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans \mathbb{R} , muni de la tribu borelienne $\mathcal{B}(\mathbb{R})$, c'est à dire

$$X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

On dit que X est une **variable aléatoire réelle** si pour tout $x \in \mathbb{R}$, l'ensemble $\{X \leq x\}$ est un événement de la tribu \mathcal{F} .

Dans la pratique, on a rarement à montrer qu'une application est une variable aléatoire car l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ est justement construit de sorte que les quantités qui nous intéressent (le chiffre du dé, la distance à la cible, etc) soient des variables aléatoires. Concrètement, un problème pourra commencer par :

"Soit X une variable aléatoire qui modélise le résultat d'un lancer de dé..."

Le but étant de s'éloigner au maximum de cet espace probabilisé dont on ne connaît pas grand chose. En revanche, il pourra être demandé de montrer que d'autres applications dépendants de X soient bien des variables aléatoires.

3.2 Loi d'une variable aléatoire

Lorsqu'on se donne une variable aléatoire X , il est important de connaître sa loi, c'est-à-dire la probabilité que X appartienne à un certain ensemble de valeurs.

Définition 3.2. La **loi** d'une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{G})$ est la donnée pour tout $U \in \mathcal{G}$ des quantités :

$$\mathbb{P}(X \in U) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in U\})$$

Théorème 3.2 (Lemme des classes monotones). *Si la tribu \mathcal{G} est engendrée par un ensemble \mathcal{C} de parties de E stable par intersection, alors la loi d'une variable aléatoire X est entièrement caractérisée par la donnée pour tout $U \in \mathcal{C}$ des quantités :*

$$\mathbb{P}(X \in U)$$

Démonstration. Admis. □

Il y a deux exemples fondamentaux à retenir.

Définition 3.3 (Cas discret). Soit X une variable aléatoire discrète, à valeurs dans un espace au plus dénombrable E . Dans ce cas, la loi de X est entièrement caractérisée par la donnée des quantités :

$$\mathbb{P}(X = e)$$

Pour $e \in E$.

Par exemple, si la variable aléatoire X modélise le nombre de personnes dans une file d'attente, alors la loi de X est caractérisée par la donnée pour $k \in \mathbb{N}$ des quantités :

$$\mathbb{P}(X = k)$$

Définition 3.4. Soit X une variable aléatoire réelle. La loi de X est entièrement caractérisée par la donnée des quantités :

$$\mathbb{P}(X \leq x)$$

Pour $x \in \mathbb{R}$.

Par exemple, si la variable aléatoire X modélise la distance d'une fléchette au centre de la cible, alors sa loi est caractérisée par la donnée pour $x \in \mathbb{R}_+$ des quantités :

$$\mathbb{P}(X \leq x)$$

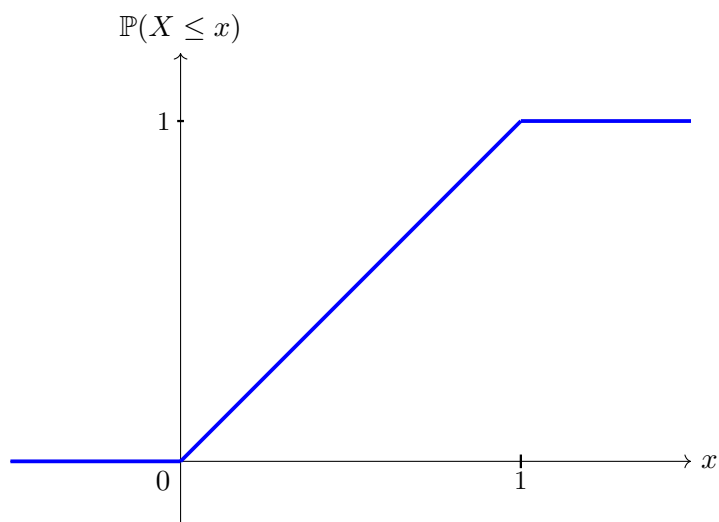
Définition 3.5. Soit X une variable aléatoire réelle. La fonction F définie par :

$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow [0, 1] \\ x &\longrightarrow \mathbb{P}(X \leq x) \end{aligned}$$

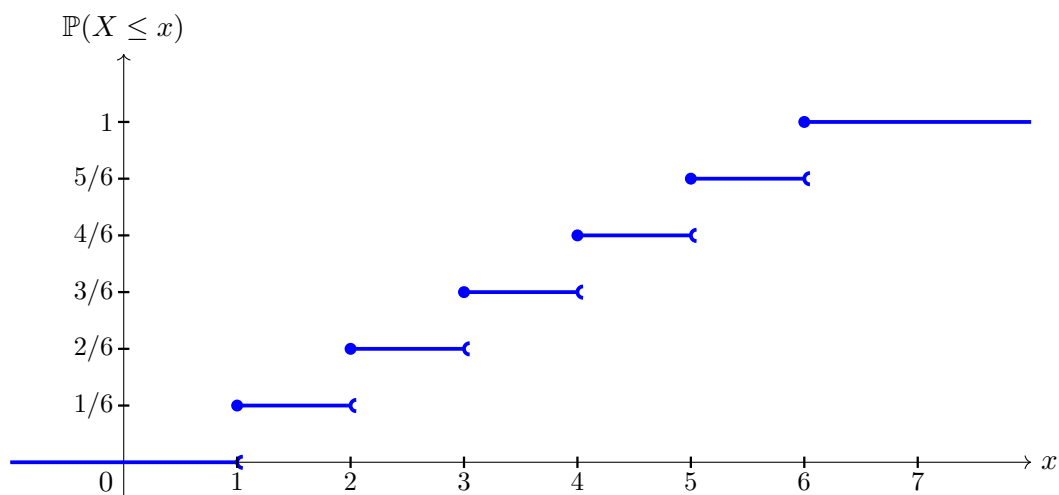
est appelée **fonction de répartition** de la variable X .

Par le théorème précédent, la fonction de répartition caractérise la loi de X . Cette fonction est centrale dans l'étude des variables aléatoires réelles.

Soit X une variable aléatoire qui modélise le tirage uniforme d'un nombre réel entre 0 et 1. Sa fonction de répartition est la suivante :



Soit X une variable aléatoire qui modélise un lancer de dé équilibré. C'est une variable aléatoire réelle puisque $\{1, 2, 3, 4, 5, 6\} \subset \mathbb{R}$. Sa fonction de répartition est la suivante :



De manière générale, lorsque X est une variable aléatoire discrète, sa fonction de répartition est constante par morceaux, et les points de discontinuité sont exactement les valeurs atteintes par X avec une probabilité non-nulle.

Théorème 3.1. *La fonction de répartition F_X d'une variable aléatoire vérifie les propriétés suivantes :*

- F_X est croissante et :

$$\lim_{x \rightarrow -\infty} F_X(x) = 0$$

$$\lim_{x \rightarrow +\infty} F_X(x) = 1$$

- F_X est continue à droite : si $(x_n)_{n \geq 0}$ est une suite décroissante qui converge vers un réel x alors :

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x_n) = \mathbb{P}(X \leq x)$$

- F_X admet une limite à gauche en tout point : si $(y_n)_{n \geq 0}$ est une suite croissante qui converge vers un réel y alors :

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X \leq y_n) = \mathbb{P}(X < y)$$

Une fonction F vérifiant les deux derniers points est dite **càdlàg** (continue à droite, limite à gauche).

Démonstration. Pour le premier point, soit x et y des réels tels que $x < y$. Alors :

$$\{X \leq x\} \subset \{X \leq y\}$$

Donc :

$$F_x(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y)$$

Par ailleurs, soit $(x_n)_{n \in \mathbb{N}}$ une suite croissante vers l'infini. Les événements $(\{X \leq x_n\})_{n \in \mathbb{N}}$ forment une suite croissante d'événements dont l'union est \mathbb{R} tout entier. D'où :

$$\begin{aligned} \lim_{n \rightarrow +\infty} F_X(x_n) &= \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x_n) \\ &= \mathbb{P}\left(\bigcup_{n=0}^{\infty} \{X \leq x_n\}\right) \\ &= \mathbb{P}(X \in \mathbb{R}) \\ &= 1 \end{aligned}$$

Le cas en $-\infty$ est similaire. Soit $(x_n)_{n \in \mathbb{N}}$ une suite décroissante vers un réel x . Alors :

$$\begin{aligned} \lim_{n \rightarrow +\infty} F_X(x_n) &= \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x_n) \\ &= \mathbb{P}\left(\bigcap_{n=0}^{\infty} \{X \leq x_n\}\right) \\ &= \mathbb{P}(X \leq x) \\ &= F_X(x) \end{aligned}$$

Soit $(y_n)_{n \in \mathbb{N}}$ une suite croissante vers un réel y . Alors :

$$\begin{aligned} \lim_{n \rightarrow +\infty} F_X(y_n) &= \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq y_n) \\ &= \mathbb{P}\left(\bigcup_{n=0}^{\infty} \{X \leq y_n\}\right) \\ &= \mathbb{P}(X < y) \end{aligned}$$

□

Inversement, soit F une fonction vérifiant les trois points du Théorème 3.1. On définit la fonction quantile $F^{-1} : [0, 1] \rightarrow \mathbb{R}$ par

$$F^{-1}(u) = \inf \{x \in \mathbb{R} \mid F(x) \geq u\}.$$

Théorème 3.2. Soit U une loi uniforme sur $[0, 1]$. La fonction quantile F^{-1} vérifie les propriétés suivantes.

- Si F est continue et croissante alors F^{-1} est la bijection réciproque de F .
- La variable aléatoire $X = F^{-1}(U)$ a pour fonction de répartition la fonction F .

Démonstration. Admis. □

Dans le cas discret, il suffit de se donner une suite de nombre positifs $(\alpha_n)_{n \in \mathbb{N}}$ tels que :

$$\sum_{n \geq 0} \alpha_n = 1$$

Pour $(x_n)_{n \geq 0}$ une suite de réels, il existe une variable aléatoire X telle que :

$$\forall n \geq 0, \quad \mathbb{P}(X = x_n) = \alpha_n.$$

Définition 3.6. Soit $x \in \mathbb{R}$. On dit qu'une variable aléatoire réelle X a un **atome** au point x si :

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

Théorème 3.3. Soit X une variable aléatoire réelle. La fonction de répartition de X présente un saut de discontinuité au point x si et seulement si X a un atome au point x . Dans ce cas, le saut est de taille $\mathbb{P}(X = x)$.

On peut vérifier le théorème sur les précédents exemples et exercices.

Démonstration. La fonction de répartition de X est continue en un point x si et seulement si les limites de F à gauche et à droite du point x coïncident, ce qui s'écrit :

$$\mathbb{P}(X < x) = \mathbb{P}(X \leq x)$$

Or on a l'égalité :

$$\mathbb{P}(X = x) = \mathbb{P}(X \leq x) - \mathbb{P}(X < x)$$

Le membre de droite s'annule si et seulement si X est sans atome. Dans le cas contraire, la taille du saut est donné par la différence des deux limites qui est exactement $\mathbb{P}(X = x)$. □

Définition 3.7 (Variable à densité). Soit X une variable aléatoire réelle de fonction de répartition F_X . On dit que la variable aléatoire X est **à densité** s'il existe une fonction intégrable f_X telle que :

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(t) dt$$

La fonction f_X est appelée la densité de probabilité de la loi de X .

La fonction F_X est une primitive de la fonction f_X . On en déduit que F_X est continue, donc X est sans atome. On en déduit également le théorème suivant :

Théorème 3.4. Soit X une variable aléatoire réelle. Si sa fonction de répartition F_X est continue et dérivable par morceaux, alors X est une variable aléatoire de densité de probabilité f et :

$$f_X = (F_X)'$$

Démonstration. C'est le lien entre primitive et dérivée. □

Inversement, si X est une variable aléatoire à densité, sa fonction caractéristique est continue et donc X n'a pas d'atome. En revanche, il existe des variables aléatoires sans atome qui n'admettent pas de densité (contre-exemple de l'escalier de Cantor).

Soient a et b des réels avec $a < b$, et X une variable aléatoire à densité f_X . Alors :

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(t) dt$$

Par exemple, soit X une variable aléatoire uniforme sur l'ensemble $[0, 1]$. Pour $0 \leq a \leq b \leq 1$ on a :

$$\mathbb{P}(a \leq X \leq b) = b - a = \int_a^b 1 dt$$

Autrement dit, la densité de la loi uniforme est la fonction $\mathbb{1}_{[0,1]}$.

Si maintenant $I = [x, x + \varepsilon]$ est un petit intervalle, alors on a :

$$\mathbb{P}(x \leq X \leq x + \varepsilon) = \int_x^{x+\varepsilon} f_X(t) dt \simeq \varepsilon f_X(x)$$

Soit X une variable aléatoire à densité f_X . La probabilité que X appartienne à un petit intervalle autour d'un point $x \in \mathbb{R}$ est proportionnelle à la taille de cet intervalle. Le coefficient de proportionnalité est exactement $f_X(x)$.

Théorème 3.5. Soit X une variable aléatoire à densité f_X . Alors :

- f_X est positive
- f_X est de masse 1 :

$$\int_{-\infty}^{+\infty} f_X(t) dt = 1$$

Inversement, si f est une fonction vérifiant ces deux propriétés, alors il existe une variable aléatoire X de densité f .

Démonstration. Pour le premier point, on a pour ε un petit réel positif :

$$0 \leq \mathbb{P}(x \leq X \leq x + \varepsilon) \simeq \varepsilon f(x)$$

Et donc f est nécessairement positif. Pour le deuxième point, on a :

$$1 = \mathbb{P}(-\infty < X < +\infty) = \int_{-\infty}^{+\infty} f(t) dt$$

□

3.3 Moyennes et dispersion d'une variable aléatoire réelle

Dans cette section, on considère uniquement des variables aléatoires réelles. Nous allons définir les notions de *moyenne* et de *dispersion* d'une variable aléatoire réelle. Ces outils permettent d'avoir une meilleure compréhension de la variable aléatoire à laquelle on fait face.

3.3.1 Espérance

Comment donner un sens à la notion de *moyenne* d'une variable aléatoire ? Une moyenne est intuitivement une valeur réelle qui est censé refléter l'ensemble de la variable aléatoire. C'est donc un moyen de réduire toute l'*information* engendrée par la variable aléatoire en un seul nombre.

Considérons dans un premier temps une variable aléatoire réelle X définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, où Ω est un ensemble fini ou dénombrable, et \mathcal{F} est la tribu discrète. Par exemple $\Omega = \{a, b, c\}$, avec :

$$\mathbb{P}(\{a\}) = \mathbb{P}(\{b\}) = \mathbb{P}(\{c\}) = \frac{1}{3}$$

Une manière naturelle de définir sa moyenne $\mathbb{E}[X]$ est via la formule :

$$\mathbb{E}[X] = \frac{X(a) + X(b) + X(c)}{3}$$

De manière générale, soit $\Omega = \{\omega_1, \omega_2, \dots\}$, muni de la tribu discrète et d'une mesure de probabilité \mathbb{P} quelconque. Si la somme :

$$\mathbb{E}[X] = \sum_{n=1}^{+\infty} X(\omega_n) \mathbb{P}(\{\omega_n\})$$

converge, on dit que X est d'espérance finie et on note $\mathbb{E}[X]$ sa moyenne.

Supposons maintenant que l'on considère une variable aléatoire réelle définie sur l'espace probabilisé $([0, 1], \mathcal{B}([0, 1]), \mu)$. Une variable aléatoire X est alors une application de $[0, 1]$ dans \mathbb{R} (qui vérifie la condition du théorème 3.1). Si X est une fonction intégrable, on définit alors sa moyenne par :

$$\mathbb{E}[X] = \int_0^1 X(\omega) d\omega$$

Dans ces deux exemples, nous avons un outil permettant de moyenner :

- La somme dans le cas où Ω est fini ou dénombrable
- L'intégrale dans le cas où $\Omega = [0, 1]$.

Nous allons définir la notion d'**intégrale** (ou d'**espérance**) sur un espace probabilisé quelconque $(\Omega, \mathcal{F}, \mathbb{P})$. Commençons par définir l'intégrale de fonctions "faciles". Soit A un événement. Il est naturel de poser :

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

On demande à ce que l'espérance soit *linéaire*. Si A_1, \dots, A_n sont des événements et $\lambda_1, \dots, \lambda_n$ des réels on pose :

$$\mathbb{E}[\lambda_1 \mathbb{1}_{A_1} + \dots + \lambda_n \mathbb{1}_{A_n}] = \lambda_1 \mathbb{P}(A_1) + \dots + \lambda_n \mathbb{P}(A_n)$$

Il est facile de voir que cette définition est cohérente. Supposons qu'il existe des événements B_1, \dots, B_m et des réels $\alpha_1, \dots, \alpha_m$ tels que :

$$\lambda_1 \mathbb{1}_{A_1} + \dots + \lambda_n \mathbb{1}_{A_n} = \alpha_1 \mathbb{1}_{B_1} + \dots + \alpha_m \mathbb{1}_{B_m}$$

Dans ce cas :

$$\lambda_1 \mathbb{P}(A_1) + \dots + \lambda_n \mathbb{P}(A_n) = \alpha_1 \mathbb{P}(B_1) + \dots + \alpha_m \mathbb{P}(B_m)$$

Par exemple :

$$\mathbb{1}_{[0,1/2]} + 2\mathbb{1}_{[1/2,1]} = \mathbb{1}_{[0,1]} + \mathbb{1}_{[1/2,1]}$$

Et :

$$\frac{1}{2} + 2 * \frac{1}{2} = 1 + \frac{1}{2}$$

Une variable aléatoire X s'écrivant sous la forme :

$$X = \lambda_1 \mathbb{1}_{A_1} + \dots + \lambda_n \mathbb{1}_{A_n}$$

est dite **étagée**. Ce sont les briques de bases pour la définition de notre intégrale (au même titre que les fonctions en escalier sont les briques de bases pour l'intégrale de Riemann).

Définition 3.3. Soit X une variable aléatoire réelle **positive** sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Si X est étagée avec :

$$X = \lambda_1 \mathbb{1}_{A_1} + \dots + \lambda_n \mathbb{1}_{A_n}$$

On pose :

$$\mathbb{E}[X] = \lambda_1 \mathbb{P}(A_1) + \dots + \lambda_n \mathbb{P}(A_n)$$

Dans le cas général, on pose :

$$\mathbb{E}[X] = \sup \{ \mathbb{E}[Y] \mid Y \text{ étagée et } Y \leq X \}$$

On dit que X est intégrable (ou d'espérance finie) si :

$$\mathbb{E}[X] < +\infty$$

Définition 3.4. Soit X une variable aléatoire réelle **quelconque** (pas nécessairement positive). On décompose X en :

$$X = X \mathbb{1}_{X \geq 0} + X \mathbb{1}_{X < 0}$$

Si les variables aléatoires (positives) $X \mathbb{1}_{X \geq 0}$ et $(-X \mathbb{1}_{X < 0})$ sont d'espérances finies, on dit que X est d'espérance finie et on définit sa moyenne par :

$$\mathbb{E}[X] = \mathbb{E}[X \mathbb{1}_{X \geq 0}] - \mathbb{E}[X \mathbb{1}_{X < 0}]$$

En théorie de la mesure, cette intégrale est appelée intégrale de Lebesgue. Dans le cas discret, elle coïncide avec la somme comme vu en introduction. Dans le cas réel elle coïncide avec l'intégrale de Riemann sur les fonctions continues par morceaux, mais elle permet d'intégrer encore plus de fonctions. On peut par exemple intégrer des fonctions très irrégulières comme $\mathbb{1}_{\mathbb{Q}}$. Les théorèmes de convergences dominées sont également plus puissants et plus simples à énoncer.

Il se peut que l'espérance d'une variable aléatoire positive soit infinie. Par exemple, dans le cas où $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \mu)$ la fonction :

$$X : \omega \mapsto \frac{1}{\omega}$$

est d'intégrale infinie.



FIGURE 1 – À gauche : approximation de l'intégrale de Riemann par des fonctions en escalier. À droite, approximation de l'intégrale de Lebesgue par des fonctions étagées.

Théorème 3.6. Soit X et Y deux variables aléatoires intégrables définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, et λ un réel. Alors

Linéarité : Pour $\lambda \in \mathbb{R}$ on a

$$\mathbb{E}[\lambda X + Y] = \lambda \mathbb{E}[X] + \mathbb{E}[Y]$$

— *Croissance :* Si pour presque tout $\omega \in \Omega$, on a $X(\omega) \leq Y(\omega)$ alors

$$\mathbb{E}[X] \leq \mathbb{E}[Y]$$

— Si X suit une loi de Bernoulli de paramètre p alors

$$\mathbb{E}[X] = p.$$

Démonstration. On a vu que l'espérance est linéaire sur l'ensemble des variables aléatoires étagées. On admettra que la propriété reste vraie en passant à la limite. La croissance de l'espérance se prouve de manière similaire. \square

Malheureusement, cette définition de moyenne n'est pas du tout pratique. En effet elle semble dépendre fortement de l'espace probabilisé Ω . Fort heureusement, on dispose du théorème fondamental suivant qui nous permet de contourner le problème.

Théorème 3.7 (théorème de transfert). Soit X une variable aléatoire défini sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, et g une fonction de \mathbb{R} dans \mathbb{R} . Alors la quantité $\mathbb{E}[g(X)]$ ne dépend que de la fonction g et de la loi de X .

Démonstration. On utilise un argument d'approximation. Si U est un sous ensemble mesurable de \mathbb{R} et $g = \mathbb{1}_U$ alors :

$$\mathbb{E}[g(X)] = \mathbb{E}[\mathbb{1}_{\{X \in U\}}] = \mathbb{P}(X \in U)$$

Cette quantité ne dépend que de la loi de X . C'est encore vrai lorsque g est une somme d'indicateur, par linéarité de l'espérance. On utilise un argument d'approximation par fonction étagées si la fonction g est quelconque. \square

Nous allons préciser ce théorème dans le cas où X est une variable aléatoire discrète ou à densité.

Théorème 3.8. Soit X une variable aléatoire discrète et notons $(x_n)_{n \in \mathbb{N}}$ les valeurs atteintes par X . Alors X est d'espérance finie si et seulement si :

$$\sum_{n=0}^{\infty} |x_n| \mathbb{P}(X = x_n) < +\infty$$

Dans ce cas :

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} x_n \mathbb{P}(X = x_n)$$

De plus, si g est une fonction de \mathbb{R} dans \mathbb{R} alors (sous réserve que la somme converge absolument) :

$$\mathbb{E}[g(X)] = \sum_{n=0}^{\infty} g(x_n) \mathbb{P}(X = x_n)$$

Démonstration. Il suffit de reprendre le début de preuve du théorème 3.7 en adaptant au cas où la variable aléatoire X n'est pas nécessairement positive. \square

Théorème 3.9. Soit X une variable aléatoire réelle de densité f_X . Alors X est d'espérance finie si et seulement si :

$$\int_{\mathbb{R}} |x| f_X(x) dx < +\infty$$

Dans ce cas :

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx$$

De plus, si g est une fonction de \mathbb{R} dans \mathbb{R} alors (sous réserve que l'intégrale converge absolument) :

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f(x) dx.$$

Démonstration. Si $g = \mathbb{1}_{[a,b]}$:

$$\mathbb{E}[g(X)] = \mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx = \int_{\mathbb{R}} g(x) f(x) dx$$

Par linéarité, la formule est encore vraie si g est une somme d'indicatrices. Par un résultat d'approximation on en déduit le cas général. \square

Un théorème important pour borner certaines probabilités est le suivant.

Théorème 3.10 (Inégalité de Markov). Soit X une variable aléatoire positive d'espérance finie, et α un réel strictement positif. Alors

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}.$$

Plus généralement, si g une fonction positive et strictement croissante, alors

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[g(X)]}{g(\alpha)}.$$

Démonstration. On note

$$A = \{X \geq \alpha\}.$$

On observe alors que

$$\forall \omega \in \Omega, \quad \alpha \mathbb{1}_A(\omega) \leq X(\omega).$$

En passant à l'espérance on obtient

$$\begin{aligned} \alpha \mathbb{E}[\mathbb{1}_A] &\leq \mathbb{E}[X] \\ \mathbb{P}(A) &\leq \frac{\mathbb{E}[X]}{\alpha} \\ \mathbb{P}(X \geq \alpha) &\leq \frac{\mathbb{E}[X]}{\alpha} \end{aligned}$$

Pour la deuxième inégalité, on observe par croissance de g que

$$A = \{X \geq \alpha\} = \{g(X) \geq g(\alpha)\}.$$

Il suffit alors d'appliquer l'inégalité à la variable aléatoire $g(X)$ et au réel $g(\alpha)$. □

3.3.2 Médiane

Un enquête INSEE a révélé que le salaire moyen d'un Français était de 2263 euros net par mois. Pourtant, la moitié de la population touche un salaire net mensuel inférieur à 1683 euros. Comment expliquer une telle différence (580 euros) ?

La moyenne d'un échantillon est sensible aux valeurs extrêmes prises par l'échantillon. C'est pourquoi elle n'est pas toujours adapté à l'étude d'une statistique.

Dans notre cas, une petite portion de personne ayant un très haut revenu fait "gonfler" la moyenne des salaires. Imaginons un village de 1000 habitants dont le salaire moyen est de 2000 euros par mois. Si Bill Gates (salaire mensuel : 500 000 euros par mois) décidait de s'installer dans ce village, le salaire moyen augmenterait alors de 500 euros, soit une augmentation de 25%.

C'est pourquoi il est plus raisonnable ici d'utiliser un estimateur robuste vis-à-vis des valeurs extrêmes.

Définition 3.8. Une médiane d'une variable aléatoire est un nombre m tel que :

$$\mathbb{P}(X \leq m) \geq 1/2 \quad \text{et} \quad \mathbb{P}(X \geq m) \geq 1/2$$

On montre facilement qu'il existe toujours au moins une médiane, mais celle-ci n'est pas nécessairement unique. Par exemple, si X suit une loi de Bernoulli, tout nombre compris entre 0 et 1 est une médiane.

Dans l'exemple précédent, le salaire médian est donc de 1683 euros net par mois. L'intérêt d'une médiane est qu'elle est insensible aux valeurs extrêmes. Si les salaires étaient plafonnés à 3000 euros net par mois, le salaire médian s'en verrait inchangé. Une inégalité marquante entre le salaire moyen et le salaire médian montre qu'une majorité de la population vit pauvrement tandis qu'une petite partie de la population accumule les richesses.

Dans le cas de la désintégration nucléaire, la médiane est aussi appelée demi-vie. La probabilité qu'un atome se désintègre avant sa demi-vie est exactement 1/2.

3.3.3 Variance

La variance est une mesure de la *dispersion* d'un variable aléatoire.

La dispersion quantifie à quel point une variable aléatoire (ou un échantillon) s'éloigne de sa valeur moyenne. Une variable aléatoire dont la dispersion est nulle est donc constante.

En finance, la dispersion, ou volatilité, est liée au risque. Une variance élevée indique une action dont la valeur a tendance à beaucoup fluctuer. Une variance faible indique une action dont la valeur est relativement stable.

Autre exemple, je regarde l'ensemble des notes d'un devoir que je viens de corriger. Si ma variance est faible, cela signifie que mes élèves sont dans un "mouchoir de poche". Ils ont tous ou presque répondu aux questions faciles, tandis qu'ils ont tous bloqué aux questions difficiles. Il est donc plus difficile de départager un classement des élèves. Inversement, une variance élevée indique le niveau du devoir était bien dosé, et il est plus facile de départager les candidats. C'est le but recherché lors d'un concours par exemple.

Définition 3.9. Soit X une variable aléatoire d'espérance finie. On note $\text{Var}(X)$ la **variance** de X définie par :

$$\text{Var}(X) = \mathbb{E}[(X - E[X])^2]$$

Si cette quantité est finie on dit que X est de variance finie. Dans ce cas on note :

$$\sigma(X) = \sqrt{\text{Var}(X)}$$

son écart-type.

La variance mesure donc l'écart à la moyenne. Ce n'est bien sûr pas seul moyen de mesurer la dispersion d'une variable aléatoire mais c'est de loin le plus courant.

Théorème 3.11. Soit X une variable aléatoire de variance finie, et λ un réel. On a :

- $\text{Var}(\lambda X) = \lambda^2 \text{Var}(X)$
- $\text{Var}(X + \lambda) = \text{Var}(X)$
- X est constante p.s. si et seulement si $\text{Var}(X) = 0$
- $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Démonstration. Pour le premier point :

$$\text{Var}(\lambda X) = \mathbb{E}[(\lambda X - E[\lambda X])^2] = \mathbb{E}[(\lambda(X - E[X]))^2] = \lambda^2 \text{Var}(X)$$

Pour le deuxième point :

$$\text{Var}(X + \lambda) = \mathbb{E}[(X + \lambda - (E[X] + E[\lambda]))^2] = \mathbb{E}[(X - E[X])^2] = \text{Var}(X)$$

Pour le troisième point, on a $X = c$ avec c une constante. Alors :

$$\text{Var}(X) = \mathbb{E}[(c - E[c])^2] = \mathbb{E}[(c - c)^2] = \mathbb{E}[0] = 0.$$

Inversement, une variable aléatoire positive est d'espérance nulle si et seulement si elle est presque sûrement nulle. Dans ce cas, on a $X = E[X]$ et est bien constante p.s.. Pour le dernier point on développe l'expression de la variance :

$$\text{Var}(X) = \mathbb{E}[(X - E[X])^2] = \mathbb{E}[X^2] - 2E[X]^2 + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

□

On retiendra également les formules suivante pour la variance dans les cas discrets et continus :

Théorème 3.12. *Si X est une variable aléatoire discrète à valeur dans $\{x_1, x_2, \dots\}$ de variance finie, alors :*

$$\text{Var}(X) = \left(\sum_{n=1}^{+\infty} x_n^2 \mathbb{P}(X = x_n) \right) - \left(\sum_{n=1}^{+\infty} x_n \mathbb{P}(X = x_n) \right)^2$$

Théorème 3.13. *Si X est une variable aléatoire de densité f_X à variance finie, alors :*

$$\text{Var}(X) = \left(\int_{\mathbb{R}} t^2 f_X(t) dt \right) - \left(\int_{\mathbb{R}} t f_X(t) dt \right)^2$$

On énonce également le théorème suivant qu'il sera important de connaître pour la suite.

Théorème 3.14 (Inégalité de Bienaymé-Tchebychev). *Soit X une variable aléatoire de variance finie et β un réel non nul. A l'aide de l'inégalité de Markov 3.10, montrer que :*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \beta) \leq \frac{\text{Var}(X)}{\beta^2}$$

Démonstration. On applique l'inégalité de Markov 3.10 à la variable aléatoire $|X - \mathbb{E}[X]|$ et la fonction $g(x) = x^2$. Alors

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}[X]| \geq \beta) &\leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{\beta^2} \\ &\leq \frac{\text{Var}(X)}{\beta^2}. \end{aligned}$$

□

3.4 Quelques lois usuelles

On fait un petit rappel des lois usuelles fréquemment rencontrées en probabilité.

3.4.1 Lois discrètes

- Loi de Bernoulli $\mathcal{B}(p)$: C'est la loi associée à une expérience ayant deux issues possibles : 0 ou 1. La variable aléatoire X suit une loi de Bernoulli de paramètre p si :

$$\mathbb{P}(X = 0) = 1 - p \quad \text{et} \quad \mathbb{P}(X = 1) = p$$

Si A est un événement, la variable aléatoire $\mathbb{1}_A$ suit une loi de Bernoulli de paramètre $\mathbb{P}(A)$.

$$\mathbb{E}[X] = p \quad \text{Var}(X) = p(1 - p)$$

- Loi binomiale $\mathcal{B}(n, p)$: C'est la loi associée à la répétition de n variable aléatoires indépendantes et identiquement distribuées suivant une loi de Bernoulli de paramètre p . La variable aléatoire X suit une loi binomiale de paramètres n et p si :

$$\forall k \in \{0, \dots, n\}, \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Si on lance n fois une pièce truquée (on note p la probabilité qu'elle tombe sur pile), alors la variable X qui compte le nombre de pile obtenu suit une loi binomiale de paramètre n et p .

$$\mathbb{E}[X] = np \quad \text{Var}(X) = np(1-p)$$

- Loi uniforme $\mathbb{U}(a, b)$: C'est la loi uniforme sur les entiers $a, a+1, \dots, b$. On note $n = b - a + 1$. La variable X suit une loi uniforme de paramètre a et b si :

$$\forall k \in \{0, \dots, n\}, \quad \mathbb{P}(X = k) = \frac{1}{n}$$

On a :

$$\mathbb{E}[X] = \frac{a+b}{2} \quad \text{Var}(X) = \frac{n^2-1}{12}$$

- Loi géométrique $\mathcal{G}(p)$: C'est la loi associée au premier instant de réalisation. La variable aléatoire X suit une loi géométrique de paramètre p si :

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(X = k) = (1-p)^{k-1}p$$

On a :

$$\mathbb{E}[X] = \frac{1}{p} \quad \text{Var}(X) = \frac{1-p}{p^2}$$

Si on lance une pièce truquée (on note p la probabilité qu'elle tombe sur pile), la variable X qui donne le premier instant où la pièce tombe sur pile suit une loi géométrique de paramètre p .

- Loi de Poisson $\mathcal{P}(\lambda)$: C'est la loi associée au nombre de personne dans une file d'attente. La variable aléatoire X suit une loi de Poisson de paramètre λ si :

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Dans une file d'attente, on suppose que le temps entre l'arrivée de deux clients suit une loi sans mémoire (loi exponentielle) de paramètre λ . Dans ce cas, la nombre de personnes dans la file d'attente à l'instant 1 suit une loi de Poisson de paramètre λ (à l'instant t , le paramètre est λt).

$$\mathbb{E}[X] = \lambda \quad \text{Var}(X) = \lambda$$

3.4.2 Loi à densité

- Loi uniforme $\mathcal{U}([a, b])$: C'est la loi uniforme sur l'intervalle $[a, b]$ borné. X suit une loi uniforme sur $[a, b]$ si :

$$\forall c, d \text{ tels que } a \leq c \leq d \leq b, \quad \mathbb{P}(c < X < d) = \frac{d-c}{b-a}.$$

La densité est donnée par

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}.$$

On peut considérer des variables aléatoires uniformes sur n'importe quel ensemble dont la mesure est finie (intervalle bornée, cercle, sphère, tore, etc).

$$\mathbb{E}[X] = \frac{b+a}{2} \quad \text{Var}(X) = \frac{(b-a)^2}{12}$$

- Loi exponentielle $\mathcal{E}(\lambda)$: C'est l'unique famille de lois sans mémoires. Sa densité est donnée par :

$$f_X : t \mapsto \lambda e^{-\lambda t}$$

On a :

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

- Loi normale $\mathcal{N}(m, \sigma^2)$: Cette loi apparaît naturellement dans le théorème central limite. Elle modélise souvent les fluctuations aléatoire d'un paramètre autour de sa valeur moyenne (température, bruit blanc, cours d'une action, etc). Sa densité est donnée par :

$$f_X : t \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-m)^2}{2\sigma^2}}$$

Si $m = 0$ et $\sigma = 1$ on parle de loi normale centrée réduite.

$$\mathbb{E}[X] = m \quad \text{Var}(X) = \sigma^2$$

3.5 Couple de variables aléatoires

On est régulièrement amené à regarder plusieurs variables aléatoires simultanément. On peut par exemple calculer leur somme, regarder si elles sont corrélées, etc. Nous donnons la définition d'un couple de variables aléatoires mais cette définition se généralise à un ensemble dénombrable de variables aléatoires. Soient X une variable aléatoire

$$X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (E, \mathcal{G})$$

et Y une autre variable aléatoire définie sur le même espace :

$$Y : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (F, \mathcal{H})$$

Alors il est possible de définir la variable aléatoire produit :

$$(X, Y) : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (E \times F, \mathcal{G} \otimes \mathcal{H})$$

Définition 3.5. Soient (E, \mathcal{G}) et (F, \mathcal{H}) deux espaces mesurables. L'espace produit :

$$E \times F = \{(x, y) \mid x \in E, y \in F\}$$

Peut naturellement être muni d'une tribu, appelée la tribu produit et notée $\mathcal{G} \otimes \mathcal{H}$. C'est la tribu engendrée par les pavés :

$$\mathcal{G} \otimes \mathcal{H} = \sigma(\{A \times B \mid A \in \mathcal{G}, B \in \mathcal{H}\})$$

La loi d'un couple (X, Y) de variables aléatoires défini comme précédemment est donc la donnée des probabilités :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(\{X \in A\} \cap \{Y \in B\})$$

Pour tout $A \in \mathcal{G}$ et $B \in \mathcal{H}$.

Lorsque l'on connaît la loi du couple (X, Y) , il est évidemment possible de connaître les lois de X et Y . On appelle cela les lois **marginales** du couple (X, Y) .

Théorème 3.3 (Loi marginale). Soit X une variable aléatoire :

$$X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (E, \mathcal{G})$$

et Y une autre variable aléatoire définie sur le même espace :

$$Y : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (F, \mathcal{H})$$

Si l'on connaît la loi du couple (X, Y) on en déduit celle de X par la formule :

$$\forall A \in \mathcal{G}, \quad \mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in F)$$

Démonstration. Trivial. □

Le lemme des classes monotones reste encore vrai. Détaillons les deux cas qui nous intéressent.

Théorème 3.15 (discret). Soit X et Y deux variables aléatoires discrètes définies sur le même espace probabilisé. La loi du couple (X, Y) est la donnée des quantités :

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

où x_k (resp. y_j) parcourt l'ensemble des valeurs prises par X (resp. Y).

Les lois marginales de X et Y sont alors données par

$$\mathbb{P}(X = x_k) = \sum_j \mathbb{P}(X = x_k, Y = y_j) \quad \text{et} \quad \mathbb{P}(Y = y_j) = \sum_k \mathbb{P}(X = x_k, Y = y_j).$$

Théorème 3.16 (réel). Soit X et Y deux variables aléatoires réelles définies sur le même espace probabilisé. La loi du couple (X, Y) est la donnée des quantités

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

où x et y parcourent \mathbb{R} .

Nous pouvons préciser le cas particulier où le couple (X, Y) admet une densité jointe.

Théorème 3.17. Soit X et Y deux variables aléatoires réelles définies sur le même espace probabilisé. On dit que le couple (X, Y) admet une densité jointe $f_{(X,Y)}$ si pour tout intervalle I et J de \mathbb{R} on a :

$$\mathbb{P}(X \in I, Y \in J) = \int_{I \times J} f_{(X,Y)}(x, y) dx dy$$

Dans ce cas, les variables aléatoires X et Y admettent des densités marginales données par

$$f_X : x \mapsto \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy \quad \text{et} \quad f_Y : y \mapsto \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx.$$

3.6 Variables aléatoires indépendantes

Soit X une variable aléatoire :

$$X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (E, \mathcal{G})$$

et Y une autre variable aléatoire définie sur le même espace :

$$Y : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (F, \mathcal{H})$$

On définit l'indépendance de deux variables aléatoires X et Y .

Définition 3.6. Deux variables aléatoires X et Y définies comme ci-dessus sont indépendantes si :

$$\forall A \in \mathcal{G}, \forall B \in \mathcal{H}, \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

Autrement dit, les événements $\{X \in A\}$ et $\{Y \in B\}$ sont indépendants.

Précisons les cas discrets et réels.

Définition 3.10 (discret). Soit X et Y deux variables aléatoires discrètes définies sur le même espace probabilisé. Les variables aléatoires X et Y sont indépendantes si et seulement si

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

où x_k (resp. y_j) parcourt l'ensemble des valeurs prises par X (resp. Y).

Définition 3.11 (réel). Soit X et Y deux variables aléatoires réelles définies sur le même espace probabilisé. Les variables aléatoires X et Y sont indépendantes si et seulement si :

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

Si X et Y sont deux variables aléatoires de densités respectives $f_X(x)$ et $f_Y(y)$, elles sont indépendantes si et seulement si le couple (X, Y) admet une densité jointe donnée par

$$f_{(X,Y)} : (x, y) \mapsto f_X(x)f_Y(y).$$

Ainsi, si deux variables aléatoires X et Y sont indépendantes alors on connaît la loi du couple (X, Y) , c'est simplement la loi donnée comme "loi produit" de X et de Y . On peut voir l'indépendance en terme d'*information*. Si X est une variable aléatoire à valeurs dans (E, \mathcal{G}) on peut définir la tribu $\sigma(X)$:

$$\sigma(X) = \{\{X \in A\} \mid A \in \mathcal{G}\}$$

C'est l'ensemble des événements engendrés par la variable X . Dans ce cas, les variables aléatoires X et Y sont indépendantes si et seulement si les tribus $\sigma(X)$ et $\sigma(Y)$ sont indépendantes.

De manière générale il est possible de parler d'une famille $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires mutuellement indépendantes. Elles sont indépendantes si et seulement si les tribus $(\sigma(X_n))_{n \in \mathbb{N}}$ sont mutuellement indépendantes. Par exemple, si X, Y et Z sont le résultat du lancer de trois dés, alors $X + Y$ est indépendant de Z .

Théorème 3.18. Soit X et Y deux variables aléatoires indépendantes d'espérances finies. Alors :

$$E[XY] = E[X]E[Y]$$

Et plus généralement, pour toute fonction f et g mesurable de \mathbb{R} dans \mathbb{R} :

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$

Si X et Y sont de variances finies :

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Démonstration. Pour le premier point, on le prouve facilement pour f et g des fonctions indicatrices : si $f = \mathbb{1}_A$ et $g = \mathbb{1}_B$ avec A et B des intervalles alors :

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[\mathbb{1}_{\{X \in A\}} \mathbb{1}_{\{Y \in B\}}] = \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$$

Par linéarité puis en passant à la limite le résultat général s'en déduit. Pour le deuxième point :

$$\text{Var}(X + Y) = \mathbb{E}[(X - \mathbb{E}[X]) + (Y - \mathbb{E}[Y])]^2 = \text{Var}(X) + \text{Var}(Y) + 2\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

On note :

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

la covariance de X et Y . Dans le cas où X et Y sont indépendants, la covariance des deux termes est nulle d'après la première partie du théorème, et on en déduit la proposition. \square

Théorème 3.19. Soient X_1, \dots, X_n des variables aléatoires mutuellement indépendantes. Alors

$$\mathbb{E}[X_1 \dots X_n] = \mathbb{E}[X_1] \dots \mathbb{E}[X_n],$$

et

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

Démonstration. Similaire à la preuve du théorème précédent. \square

Un cas d'application important de l'indépendance est la somme de variables aléatoires réelles indépendantes.

Théorème 3.20. Soit X et Y deux variables aléatoires indépendantes à valeur dans \mathbb{N} , et $Z = X + Y$. Alors :

$$\mathbb{P}(Z = n) = \sum_{k=0}^n \mathbb{P}(X = k)\mathbb{P}(Y = n - k)$$

Soit X et Y deux variables aléatoires à densité et indépendantes, alors $Z = X + Y$ a une densité donnée par :

$$f_Z(z) = \int_{\mathbb{R}} f_X(u)f_Y(z - u)du$$

Démonstration. Dans le cas discret, c'est la formule des probabilités totales. Dans le cas à densité on a :

$$\mathbb{P}(X + Y \leq z) = \mathbb{P}((X, Y) \in A_z)$$

Avec A_z l'ensemble :

$$A_z = \{(x, y) \in \mathbb{R}^2 \mid x + y \leq z\}$$

D'où :

$$\begin{aligned} \mathbb{P}(X + Y \leq z) &= \int_{x+y \leq z} f_X(x)f_Y(y)dx dy \\ &= \int_{-\infty}^z \int_{\mathbb{R}} f_X(u)f_Y(v - u)dudv \end{aligned}$$

On a fait le changement de variable $(u, v) = (x, x + y)$. La densité s'obtient en dérivant par rapport à la variable z . \square

4 Théorèmes limites en probabilité

4.1 Modes de convergence de variables aléatoires

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires et X une variable aléatoire, toutes définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On cherche à donner un sens à la convergence de la suite $(X_n)_{n \geq 0}$ vers X . Pour cela on distingue 4 modes de convergence.

Définition 4.1 (Convergence presque-sûre). On dit que la suite de variables aléatoires $(X_n)_{n \geq 0}$ converge presque-sûrement vers X si pour presque tout $\omega \in \Omega$,

$$\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega).$$

On écrit alors

$$X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X.$$

C'est la convergence ponctuelle de variables aléatoires. Elle apparaît dans la loi des grands nombres. Intuitivement, si l'on simule la suite $X_n(\omega) - X(\omega)$ pour un ω fixé, alors cette suite converge vers 0 en temps long. On la retrouve dans la loi forte des grands nombre, et sa principale application est la methode de Monte-Carlo. Un cas simple de convergence presque sûre est lorsque l'on a affaire à une suite croissante et bornée de variables aléatoires. Alors elle converge presque-sûrement vers une variable aléatoire limite.

Définition 4.2 (Convergence en moyenne d'ordre p). On dit que $(X_n)_{n \geq 0}$ converge en moyenne d'ordre p si

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|X_n - X|^p] = 0.$$

On écrit alors

$$X_n \xrightarrow[n \rightarrow +\infty]{L^p} X.$$

Lorsque $p = 2$ on parle de convergence quadratique. C'est un mode de convergence utile pour certaines applications théoriques. Il permet en particulier de prouver la convergence de la moyenne, de la variance, etc.

Définition 4.3 (Convergence en probabilité). On dit que la suite de variables aléatoires $(X_n)_{n \geq 0}$ converge en probabilité vers X si pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

On écrit alors

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X.$$

Intuitivement, une suite de variables aléatoires converge en probabilité vers sa limite lorsque la probabilité que la suite prenne des valeurs éloignées de sa limite tend vers 0. Par exemple, si $(X_n)_{n \geq 0}$ est une suite de variables aléatoires suivant une loi Bernoulli de paramètre $1/n$, alors elle converge en probabilité vers la variable aléatoire constante égale à 0, car

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X_n - 0| > \varepsilon) \leq \mathbb{P}(X_n = 1) = \frac{1}{n} \xrightarrow[n \rightarrow +\infty]{} 0.$$

Définition 4.4 (Convergence en loi). On dit que la suite de variables aléatoires $(X_n)_{n \geq 0}$ (de fonctions de répartition $(F_n)_{n \geq 0}$) converge en loi vers X (de fonction de répartition F) si en tout point de continuité x de la fonction F on a

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x).$$

On écrit alors

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X.$$

La convergence en loi est particulière parmi les 4 modes de convergence. Elle dépend uniquement de la loi de la suite et non de l'espace probabilisé sur laquelle elle est définie. Par exemple, si X est une loi symétrique, la loi de la suite

$$(X, -X, X, -X, \dots)$$

est constante et converge donc en loi vers X , mais ne converge en général pas pour d'autres modes de convergence. Lorsqu'on simule une instance de la suite, on observe pas en général de convergence particulière. En revanche on l'observe en répétant l'expérience un grand nombre de fois et en faisant par exemple un histogramme. C'est un mode de convergence faible mais qui est facile à prouver car il ne dépend que de la suite des lois de la suite $(X_n)_{n \geq 0}$, pas de leur interdépendance. En place de X on pourra directement mettre une loi. On pourra par exemple écrire

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

On a les implications suivantes sur les convergences :

Convergence presque-sûre \Rightarrow Convergence en probabilités \Rightarrow Convergence en loi

Convergence L^p \Rightarrow Convergence en probabilités \Rightarrow Convergence en loi

Les réciproques sont en général fausses, hormis quelques cas particuliers.

4.2 Loi des grands nombres

La loi des grands nombre fait le lien entre la moyenne empirique et la moyenne théorique. Par exemple, lors d'un lancer de dé, j'ai une chance sur six de tomber sur 5. La loi des grands nombre affirme qu'en moyenne, au bout d'un grand nombre d'expérience, je serais tombé environ une fois sur six sur le chiffre 5.

La loi des grands nombres permet de valider (ou d'invalider) un modèle probabiliste. Par exemple, pour vérifier qu'une pièce est truquée ou non, il suffit de la lancer un grand nombre de fois. Si le nombre moyen de pile est significativement différent de 0.5 on peut en conclure que la pièce est truquée.

Dans la suite, $(X_n)_{n \geq 1}$ est une suite de variables aléatoires réelles, indépendantes et identiquement distribuées, c'est-à-dire qu'elle ont toutes la même loi. Notons X une variable aléatoire suivant cette loi. On abrège parfois en disant que $(X_n)_{n \geq 1}$ est une suite de variables aléatoires i.i.d de même loi que X . On définit la suite des *moyennes empiriques* $(\bar{X}_n)_{n \geq 1}$ par

$$\forall n \geq 0, \quad \bar{X}_n = \frac{X_1 + \dots + X_n}{n}.$$

Théorème 4.1 (Loi des grands nombres). *Si la loi de X est d'espérance finie alors*

$$\overline{X}_n \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}[X].$$

Démonstration. Admis. □

Par exemple, je lance une infinité de fois une pièce truquée, de sorte que la probabilité de faire "Pile" est p . On peut alors définir A_n l'événement "La pièce tombe sur "Pile" au n -ième lancer, et

$$X_n = \mathbb{1}_{A_n}.$$

La variable aléatoire X_n vaut 1 si la pièce tombe sur pile et 0 sinon. Ici, la suite $(X_n)_{n \geq 1}$ est bien une suite de variables aléatoires i.i.d suivant une loi de Bernoulli de paramètre p . Dans ce cas,

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n} = \frac{\text{nombre de "Pile" sur les } n \text{ premiers lancers}}{n}$$

correspond à la proportion de "Pile" sur les n premiers lancers. La loi des grands nombre affirme alors que cette proportion converge vers l'espérance de X_n , c'est-à-dire la probabilité de l'événement "La pièce tombe sur "Pile" " :

$$\overline{X}_n \xrightarrow[n \rightarrow +\infty]{p.s.} p.$$

Une application de la loi des grands nombres est la méthode de Monte-Carlo pour estimer la valeur d'une intégrale. Cette methode est particulièrement utilisée en dimension $d \geq 1$.

Théorème 4.2 (Monte-Carlo). *Soit $f : [0, 1]^d \rightarrow \mathbb{R}$ une fonction intégrable et $(U_n)_{n \geq 1}$ une suite i.i.d de loi uniforme sur $[0, 1]^d$. Alors*

$$\frac{f(U_1) + \dots + f(U_n)}{n} \xrightarrow[n \rightarrow +\infty]{} \int_{[0,1]^d} f(x) dx$$

Démonstration. On définit $X_n = f(U_n)$. La suite $(X_n)_{n \geq 1}$ est une suite de variables aléatoires réelles iid d'espérance

$$\mathbb{E}[f(U)] = \int_{[0,1]^d} f(x) dx,$$

d'après le théorème de transfert. La conclusion découle de la loi des grands nombres appliquée à la suite $(X_n)_{n \geq 1}$. □

4.3 Théorème central limite

Le théorème central limite (TCL) permet de préciser les fluctuations de la moyenne empirique autour de la moyenne théorique. Il permet de construire des intervalles de confiance lorsque l'on cherche à estimer un certain paramètre. Par exemple, si je lance 10000 fois une pièce et que je tombe 4500 fois sur pile, puis-je conclure que la pièce est équilibrée? C'est le but du TCL.

Comme précédemment, on définit $(X_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d de même loi que X , et

$$\forall n \geq 0, \quad \overline{X}_n = \frac{X_1 + \dots + X_n}{n}.$$

Pour rappel, une variable aléatoire réelle Z suit une loi normale $\mathcal{N}(m, \sigma^2)$ si elle admet une densité donnée par

$$f_Z(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Théorème 4.3. *On suppose que la loi de X est de variance finie. Alors*

$$\frac{\bar{X}_n - \mathbb{E}[\bar{X}_n]}{\sqrt{\text{Var}(\bar{X}_n)}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Démonstration. Admis. □

L'observation du TCL se fait à l'aide d'un histogramme Si l'on note $\sigma^2 = \text{Var}(X)$ alors

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}[X] \quad \text{et} \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Si Z suite une loi normale centrée réduite (c'est-à-dire $\mathcal{N}(0, 1)$) alors

$$\sigma Z + m \sim \mathcal{N}(m, \sigma^2).$$

Dans ce cas, on peut écrire le TCL sous différentes manières.

De manière équivalente le TCL peut s'écrire

$$\sqrt{n} (\bar{X}_n - \mathbb{E}[X]) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

$$\frac{(X_1 + \dots + X_n) - n\mathbb{E}[X]}{\sqrt{n}\sigma} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Ou encore de manière informelle

$$\bar{X}_n \simeq \mathcal{N}\left(\mathbb{E}[X], \frac{\sigma^2}{n}\right).$$

Le TCL affirme que l'erreur commise entre la moyenne théorique et empirique est de l'ordre de $\sqrt{\text{Var}(\bar{X}_n)} = \sigma/\sqrt{n}$. Cela permet de calculer des intervalle de confiance asymptotiques. Voyons un exemple pour estimer la moyenne d'une variable aléatoire à l'aide de l'observation d'un échantillon de taille n , dans le cas où la variance σ est connue.

Théorème 4.4 (Intervalle de confiance asymptotique au niveau α). *Soit Z une loi normale centrée réduite et $0 < \alpha < 1$. On définit le nombre $q_{\alpha/2}$ tel que*

$$\mathbb{P}(-q_{\alpha/2} \leq Z \leq q_{\alpha/2}) = 1 - \alpha.$$

Alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}q_{\alpha/2} \leq \mathbb{E}[X] \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}}q_{\alpha/2}\right) = 1 - \alpha.$$

Démonstration. D'après le TCL, on a

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(-q_{\alpha/2} \leq \sqrt{n} \frac{\bar{X}_n - \mathbb{E}[X]}{\sigma} \leq q_{\alpha/2}\right) = 1 - \alpha.$$

La conclusion s'ensuit. □

On a $\mathbb{P}(-2 \leq Z \leq 2) \simeq 0.95$ et lorsque la suite $(X_n)_{n \geq 0}$ est une suite i.i.d de loi de Bernoulli, on a l'égalité

$$\sigma \leq 1/2.$$

Dans ce cas, on utilise généralement l'intervalle de confiance suivant :

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\bar{X}_n - \frac{1}{\sqrt{n}} \leq \mathbb{E}[X] \leq \bar{X}_n + \frac{1}{\sqrt{n}} \right) \gtrsim 0.95.$$

C'est l'intervalle de confiance "classique" qui se résume parfois sous l'adage "L'erreur commise pour estimer la proportion d'une population est de l'ordre de l'inverse de la racine carré de la taille de l'échantillon."