

Introduction aux probabilités

Ioan Manolescu

basé sur un cours de Yvan Velenik

November 17, 2017

Ce cours est basé sur le cours du Prof. Yvan Velenik "*Probabilités et Statistique*", donné à l'université de Genève.

Table des matières

1	Introduction: espaces de probabilités	4
1.1	Espaces de probabilités	4
1.2	Construction des espaces de probabilité	7
1.2.1	Univers finis: combinatoire	7
1.2.2	Univers dénombrable	12
1.2.3	Univers continus	15
1.3	Conditionnement	19
1.4	Indépendance	22
1.5	Espaces produit	24
1.5.1	Produits finis	24
1.5.2	Produits infinis	26
2	Variables aléatoires	28
2.1	Variables aléatoires et leur lois	28
2.2	Exemples	29
2.2.1	Lois discrètes	29
2.2.2	Lois continues (à densité)	31
2.2.3	Fonctions d'une variable aléatoire	35
2.3	Indépendance des variables aléatoires	36
2.4	Loi jointe; vecteurs aléatoires	40
2.5	Espérance	47
2.5.1	Espérance des variables aléatoires discrètes	47
2.5.2	Espérance des variables aléatoires quelconques	50
2.5.3	Espérance des variables aléatoires à densité	51
2.5.4	Espérance d'une fonction d'une variable aléatoire	53
2.6	Variance et moments	57
2.6.1	Inégalité de Markov, Tchebychev	61
2.7	Espérance conditionnelle; loi conditionnelle	62
2.7.1	Conditionnement par rapport à un événement	62
2.7.2	Conditionnement par rapport à une variable discrète	63
2.7.3	Espérance conditionnelle par rapport à une variable générale	64
2.7.4	Espérance conditionnelle par rapport à une variable à densité	65

3 Fonctions génératrices et applications	67
3.1 Définitions et propriétés	67
3.2 Application: vecteurs gaussiens	71
3.3 Application: processus de branchement	72
4 Théorèmes limites	75
4.1 Convergence des variables aléatoires	75
4.2 Loi des grands nombres	80
4.3 Théorème central limite	81

Chapter 1

Introduction: espaces de probabilités

1.1 Espaces de probabilités

Supposons qu'on veut modéliser une expérience dont le résultat est aléatoire. Une façon de faire cela est de faire une liste de tous les résultats possibles et d'associer à chacun une probabilité. Ceci est le début du formalisme de la théorie des probabilités, qu'on décrit par la suite.

Fixons un ensemble Ω qu'on appelle un univers. Les éléments $\omega \in \Omega$, qu'on va appeler "scénarios" ou "issues", représente les différentes issues possible de notre expérience. Un ensemble de issues est généralement appelé un *événement*.

Pour une expérience donnée, les issues peuvent être complexes, et notre information dessus peut être limitée. Ainsi, on va se limiter à étudier des événements qu'on peut observer.

Définition 1.1. Une algèbre sur Ω est une collection $\mathcal{E} \subset \mathcal{P}(\Omega)$ de sous-ensembles de Ω telle que

- (i) $\emptyset \in \mathcal{E}$,
- (ii) si $A \in \mathcal{E}$, alors $A^c := \Omega \setminus A \in \mathcal{E}$,
- (iii) si $A, B \in \mathcal{E}$, alors $A \cup B \in \mathcal{E}$.

On dit que \mathcal{E} est une σ -algèbre (ou une tribu) si en plus des propriétés ci-dessus, on a

$$\text{pour tous } A_1, A_2, \dots \in \mathcal{E}, \quad \bigcup_{n \geq 1} A_n \in \mathcal{E}.$$

Exemple: Supposons qu'on dispose d'une urne avec 4 billes, deux rouges, deux noires. De plus, les billes de chaque couleur ont sont numérotées 1 et 2. On en tire une au hasard.

L'univers qui apparait naturellement est $\Omega = \{(R, 1); (R, 2); (N, 1); (N, 2)\}$. Une σ -algèbre naturelle sur Ω est $\mathcal{P}(\Omega) = \{A : A \subset \Omega\}$. Cette σ -algèbre nous permet de distinguer toutes les billes entre elles, car pour toute bille b_1 , l'événement

$\{b_1\} \in \mathcal{P}(\Omega)$. On peut donc parler de l'événement "la bille b_1 qui est tirée." Supposons toutefois que, après avoir tiré la bille, on observe uniquement sa couleur, alors les issues $(R, 1)$ et $(R, 2)$ sont, de notre point de vue, indistinguibles. Ainsi, les événements qu'on peut étudier sont uniquement $\mathbf{R} := \{(R, 1); (R, 2)\}$, $\mathbf{N} = \{(N, 1); (N, 2)\}$, $\Omega = \{(R, 1); (R, 2); (N, 1); (N, 2)\}$ et \emptyset . Ils correspondent à: "une bille rouge est tirée", "une bille noire est tirée" et "une bille rouge ou une bille noire est tirée", respectivement. L'ensemble vide ne correspond à aucun scénario. Cette collection d'événements, à savoir $\{\emptyset, \mathbf{R}, \mathbf{N}, \Omega\}$, est aussi une σ -algèbre sur Ω .

Un Corollaire directe de la définition de σ -algèbre est le suivant.

Corollaire 1.2. *Soit \mathcal{E} une σ -algèbre sur Ω . Alors,*

- (i) $A, B \in \mathcal{E} \Rightarrow A \cap B \in \mathcal{E}$;
- (ii) $A, B \in \mathcal{E} \Rightarrow A \setminus B \in \mathcal{E}$;
- (iii) $A_1, A_2, \dots \in \mathcal{E} \Rightarrow \bigcap_{n \geq 1} A_n \in \mathcal{E}$.

Preuve: La preuve est laissée en exercice. □

Les issues de notre expérience arrive avec une certaine probabilité. Plutôt que d'associer à chaque issue une probabilité, on va associer aux différents événements les probabilités qu'ils se produisent. Cela offre un cadre plus flexible, donc plus général; dans les cas les plus simples (ceux des univers discrets) les probabilités peuvent être vues comme étant associées à chaque issue.

Définition 1.3. *Soient Ω un ensemble et \mathcal{E} une tribu sur Ω . Une probabilité \mathbb{P} sur (Ω, \mathcal{E}) est une fonction $\mathbb{P} : \mathcal{E} \rightarrow [0, 1]$ telle que*

- (i) $\mathbb{P}(\emptyset) = 0$,
- (ii) si $(A_n)_{n \in \mathbb{N}}$ est une famille **dénombrable** d'ensembles de \mathcal{E} deux à deux disjoints, alors

$$\mathbb{P}\left(\bigsqcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n). \quad (1.1)$$

- (iii) $\mathbb{P}(\Omega) = 1$.

On appelle le triplet $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilité.

Attention! La théorie des probabilités ne se soucie pas du choix de $(\Omega, \mathcal{E}, \mathbb{P})$. En effet, on part toujours du principe qu'on modélise une expérience décrite précisément, ou le choix de l'espace de probabilité (surtout celui de \mathbb{P}) va de soi.

La recherche du bon espace de probabilité à partir d'observations est l'objet de la statistique.

Exemple: Dans l'exemple de l'urne contenant 4 billes, on a

$$\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\mathbf{N}) = \mathbb{P}(\mathbf{R}) = \frac{1}{2}, \quad \mathbb{P}(\Omega) = 1.$$

Si par contre, l'urne contenait 2 billes noires et 3 rouges (en adaptant les notations), on aurait

$$\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\mathbf{N}) = \frac{2}{5}, \quad \mathbb{P}(\mathbf{R}) = \frac{3}{5}, \quad \mathbb{P}(\Omega) = 1.$$

Les propriétés suivantes ressortent de la définition d'un espace de probabilité.

Proposition 1.4. *Soit $(E, \mathcal{E}, \mathbb{P})$ un espace de probabilité. Alors*

(i) *pour $A, B \in \mathcal{E}$ avec $A \cap B = \emptyset$,*

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

(ii) *pour $A, B \in \mathcal{E}$ avec $A \subset B$, $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.*

(iii) *pour $A, B \in \mathcal{E}$,*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

(iv) *pour $A_1, A_2, \dots \in \mathcal{E}$ avec $A_1 \subset A_2 \subset \dots$,*

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} \uparrow \mathbb{P}(A_n).$$

(v) *pour $A_1, A_2, \dots \in \mathcal{E}$ avec $A_1 \supset A_2 \supset \dots$,*

$$\mathbb{P}\left(\bigcap_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} \downarrow \mathbb{P}(A_n).$$

(vi) *pour $A_1, A_2, \dots \in \mathcal{E}$,*

(σ - sous additivité)

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 1} \mathbb{P}(A_n).$$

(vii) *pour $A_1, A_2, \dots, A_n \in \mathcal{E}$,*

(inclusion-exclusion)

$$\mathbb{P}\left(\bigcup_{j=1}^n A_j\right) = \sum_{\substack{I \subset \{1, \dots, n\} \\ I \neq \emptyset}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right).$$

Les opérations d'union et intersection d'événements ont une interprétation intuitive. En général on définit les événements comme les ensembles d'issues satisfaisant certaines condi-

tions. Posons ainsi

$$A = \{\omega \in \Omega : \omega \text{ satisfait condition (a)}\} \quad \text{et} \quad B = \{\omega \in \Omega : \omega \text{ satisfait condition (b)}\},$$

ou (a) et (b) sont certaines conditions sur les issus. Alors

$$\begin{aligned} \text{intersection:} \quad & A \cap B = \{\omega \in \Omega : \omega \text{ satisfait les conditions (a) et (b)}\}, \\ \text{union:} \quad & A \cup B = \{\omega \in \Omega : \omega \text{ satisfait au moins une des conditions (a) ou (b)}\}, \\ & A \setminus B = \{\omega \in \Omega : \omega \text{ satisfait (a) mais pas (b)}\}, \\ & A^c = \{\omega \in \Omega : \omega \text{ ne satisfait pas (a)}\}. \end{aligned}$$

1.2 Construction des espaces de probabilité

1.2.1 Univers finis: combinatoire

Dans la plus part des exemples qui nous viennent à l'esprit, il n'existe que un nombre fini d'issues possible. Ainsi, on peut commencer par étudier des univers finis.

Pour un univers fini Ω , il est naturel de poser $\mathcal{E} = \mathcal{P}(\Omega)$ et d'associer à chaque issue une probabilité.

Définition 1.5. Soit Ω un ensemble non-vide. Une fonction de masse sur Ω est une fonction $f : \Omega \rightarrow [0, 1]$ avec

$$\sum_{\omega \in \Omega} f(\omega) = 1.$$

A une telle fonction de masse, on associe la probabilité \mathbb{P}_f définie par

$$\mathbb{P}_f(A) = \sum_{\omega \in A} f(\omega), \quad \forall A \in \mathcal{P}(\Omega).$$

Proposition 1.6. Soit Ω un ensemble fini, non-vide. Pour toute fonction de masse f , \mathbb{P}_f est bien une probabilité. Inversement, toute probabilité \mathbb{P} sur $(\Omega, \mathcal{P}(\Omega))$, il existe une unique fonction de masse f telle que $\mathbb{P}_f(A) = \mathbb{P}(A)$ pour $A \in \mathcal{E}$.

Preuve: Voir le preuve de la proposition plus générale 1.9. □

L'exemple le plus simple d'espace de probabilité fini est celui muni d'une probabilité uniforme. La *probabilité uniforme* \mathbb{P} sur Ω est celle donnée par la fonction de masse f avec

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

Alors,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}, \quad \forall A \in \mathcal{P}(\Omega).$$

Ainsi, sous la probabilité uniforme, la probabilité d'un événement est donné par le nombre de cas favorables divisé par le nombre de cas possibles.

Mentionnons qu'on rencontre aussi des univers finis munis d'une probabilité qui n'est pas uniforme. Supposons par exemple qu'on lance une pièce de monnaie n fois, mais que celle-ci est biaisée et qu'elle tombe sur pile avec probabilité $p \in (0, 1)$. Alors $\Omega = \{P, F\}^n$ et la fonction de masse est

$$f(\omega) = p^{\text{nombre de } P \text{ dans } \omega} (1-p)^{\text{nombre de } F \text{ dans } \omega}, \quad \text{pour tout } \omega \in \Omega.$$

Combinatoire. Soit E un ensemble fini avec $|E| = n$. On va être amenés dans des exemples à étudier certains types de sous-ensembles, ou sous-familles de E . Imaginons E comme l'ensemble des billes dans une urne.

Avant de commencer, rappelons la notation *factoriel*:

$$n! = n(n-1) \dots 1, \quad \forall n \geq 1.$$

De plus, on pose par convention $0! = 1$.

Tirages ordonnés avec remise. Supposons qu'on tire k billes de l'urne, mais qu'après chaque tirage, on remet la bille dans l'urne. De plus, on va regarder les résultats du tirage, dans l'ordre. L'univers qui s'impose alors est

$$E^k = \underbrace{E \times E \times \dots \times E}_{k \text{ fois}} := \{(x_1, \dots, x_k) : x_1, \dots, x_k \in E\}.$$

On a alors,

$$|E^k| = |E|^k = n^k, \quad \text{pour tout } k \geq 1. \quad (1.2)$$

Preuve: Fixons k . Alors, pour choisir $(x_1, \dots, x_k) \in E^k$ on a n -choix pour x_1 , n choix pour x_2 etc. Ainsi, on a en tout n^k choix pour la famille (x_1, \dots, x_k) , d'où $|E^k| = n^k$. \square

Tirages ordonnés sans remise. Supposons maintenant qu'on tire k billes, sans les remettre dans l'urne, et que on regarde les résultats du tirage dans l'ordre. L'univers approprié est alors

$$\mathcal{T}_n^k := \{(x_1, \dots, x_k) : x_1, \dots, x_k \in E, x_i \neq x_j \forall i \neq j\}. \quad (1.3)$$

Cet univers est évidemment vide pour $k > n$. Pour $1 \leq k \leq n$, on a

$$|\mathcal{T}_n^k| = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}.$$

Un cas spécial apparaît quand $k = n$. Dans ce cas la, l'univers est l'ensemble de façons d'ordonner les n billes de E . Il s'agit donc de l'ensemble des *permutations* \mathcal{S}_n :

$$\mathcal{S}_n = \{f : \{1, \dots, n\} \rightarrow \{1, \dots, n\} : f \text{ bijective}\}.$$

On a alors

$$|\mathcal{S}_n| = n!$$

Preuve: Fixons $k > 0$. Si $k > n$ il est évident que \mathcal{T}_n^k est vide. Supposons donc que $k \leq n$. Alors, pour choisir $(x_1, \dots, x_k) \in \mathcal{T}_n^k$, on a n -choix pour x_1 , $n-1$ choix pour x_2 (car il doit être différent de x_1), $n-2$ choix pour x_3 (car il doit être différent de x_1 et de x_2), etc. Pour x_k on a donc $n-k+1$ choix. Ainsi, on a en tout $n(n-1)\dots(n-k+1)$ choix pour la famille $(x_1, \dots, x_k) \in \mathcal{T}_n^k$. \square

Exemple: Dans un groupe de k personnes, quelle est la probabilité d'en avoir deux avec le même anniversaire?

On modélise de façon simplifiée cette question en supposant que chaque personne du groupe est née avec probabilité égale un des 365 jours de l'année (on ignore les ans bissextiles). Posons $A = \{1, \dots, 365\}$, les jours de l'année. Ainsi, notre univers est

$$\Omega = A^k,$$

et il y a 365^k possibilités en tout pour les dates de naissance des personnes du groupe. Toutes les possibilités étant équiprobables, elles ont chacune une probabilité de 365^{-k} .

L'événement B qui consiste à ce que tous les membres du groupe aient des anniversaires différentes est alors

$$\{(a_1, \dots, a_k) : a_1, \dots, a_k \in A, a_i \neq a_j \forall i \neq j\}.$$

Ainsi $|B| = \frac{365!}{(365-k)!}$. Il s'en suit que la probabilité de B est de

$$\mathbb{P}(B) = \frac{365!}{(365-k)! 365^k} = \frac{365 \cdot \dots \cdot (365-k+1)}{365^k}.$$

Ainsi, le fait que au moins deux personnes aient le même anniversaire a une probabilité

$$\mathbb{P}(B^c) = 1 - \mathbb{P}(B) = 1 - \frac{365!}{(365-k)! 365^k}.$$

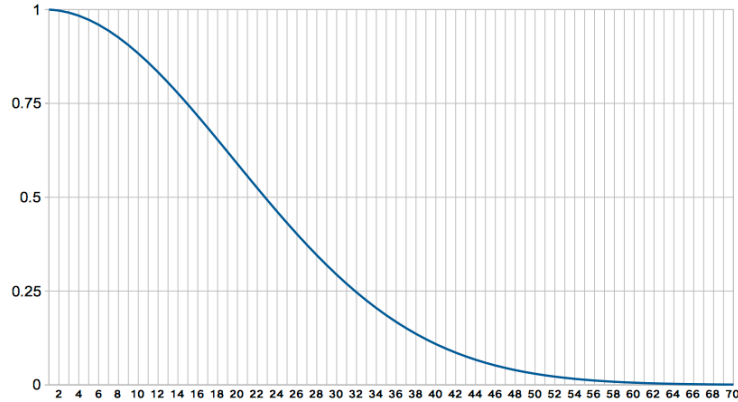


Figure 1.1: La probabilité que toutes les personnes d'un groupe aient des anniversaires différentes, en fonction de la taille du groupe.

On peut voir que cette probabilité est la plus proche de $1/2$ pour $k = 23$. De plus, la probabilité de B décroît assez vite vers 0 (voir le graphe fig. 1.1).

Tirages non-ordonnés sans remise: coefficients binomiaux Supposons maintenant qu'on tire k billes, sans les remettre dans l'urne. De plus on regarde les résultats du tirage sans se soucier de l'ordre. L'univers approprié est alors

$$\mathcal{P}_k(E) = \{T \subset E : |T| = k\}.$$

Cet univers est évidemment vide pour $k > n$. Pour $0 \leq k \leq n$, on a

$$|\mathcal{P}_k(E)| = \frac{n!}{k!(n-k)!} = \binom{n}{k}. \quad (1.4)$$

Preuve: Chaque ensemble de $\mathcal{P}_k(E)$ peut être ordonné de $k!$ façons, et chaque telle façon, correspond exactement à un élément de \mathcal{T}_n^k . Ainsi

$$|\mathcal{P}_k(E)| \cdot k! = |\mathcal{T}_n^k| = \frac{n!}{(n-k)!}.$$

La conclusion (1.4) s'en suit. □

Proposition 1.7 (Binôme de Newton). *Soient $a, b \in \mathbb{R}$ et $n \geq 1$. Alors*

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}.$$

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

Preuve: Ecrivons

$$(a + b)^n = \underbrace{(a + b) \dots (a + b)}_{n \text{ fois}}.$$

On peut développer ce produit. Chaque terme du développement correspond à un choix de a ou b dans chaque parenthèse. Les termes du type $a^k b^{n-k}$ pour un certain $k \in \{0, \dots, n\}$ fixé, correspondent au choix de k parenthèses dans les-quelles on prend le terme a . Ainsi, il y a $\binom{n}{k}$ tels terms. En sommant sur toutes les possibilités pour k (à savoir $0, 1, \dots, n$) on obtient la conclusion. \square

Exercice 1.

Soient $n \geq 1$ et $0 \leq k \leq n$. On observe de la définition que $\binom{n}{k} = \binom{n}{n-k}$, et on conclut que

$$|\mathcal{P}_k(\{1, \dots, n\})| = |\mathcal{P}_{n-k}(\{1, \dots, n\})|.$$

Donner une raison combinatoire pour cette relation.

Partitions. Supposons qu'on veut placer les n billes de E dans j urnes (numérotées de 1 à j) avec k_i billes dans la i^{eme} urne. Ainsi, on va se donner des nombres k_1, \dots, k_j avec $k_1 + \dots + k_j = n$. Alors, notre univers est

$$\left\{ (U_1, \dots, U_j) : U_1, \dots, U_j \subset E \text{ avec } \bigsqcup_{i=1}^j U_i = E \text{ et } |U_i| = k_i \forall 1 \leq i \leq j \right\}.$$

Le nombre d'éléments de cet univers est

$$\frac{n!}{k_1! k_2! \dots k_j!}.$$

Equivalent de Stirling En combinatoire, beaucoup de quantités s'expriment à l'aide des factoriels. Pour estimer des telles quantités, il est utile d'avoir un équivalent pour $n!$.

Proposition 1.8. *Quand $n \rightarrow \infty$*

$$n! \left(n^n e^{-n} \sqrt{2\pi n} \right)^{-1} \rightarrow 1,$$

ce qui s'écrit aussi

$$n! = \sqrt{2\pi n} \frac{n^n}{e^n} (1 + o(1)).$$

Preuve: La preuve va être faite en séance d'exercices. □

1.2.2 Univers dénombrable

On considère maintenant le cas des univers dénombrables. On dit que Ω est un ensemble dénombrable s'il existe une fonction bijective $f : \Omega \rightarrow \mathbb{N}$.

Ce contexte est similaire à celui des univers finis. En particulier, on va considérer par défaut $\mathcal{E} = \mathcal{P}(\Omega)$. De plus, comme dans le cas fini, les mesures de probabilité sur Ω sont données par des fonctions de masse. La définition de fonction de masse 1.5 est valable également pour les univers dénombrables, tout comme celle de la probabilité générée par une fonction de masse f :

$$\mathbb{P}_f(A) := \sum_{\omega \in A} f(\omega), \quad \forall A \in \mathcal{P}(\Omega). \quad (1.5)$$

Un petit mot sur les sommes infinies : Une somme $\sum_{\omega \in A} f(\omega)$, ou f prend valeurs positives, est définie comme suit:

- si $A = \emptyset$, on pose par convention $\sum_{\omega \in \emptyset} f(\omega) = 0$;
- si $A = \{\omega_1, \dots, \omega_n\}$ est fini, alors

$$\sum_{\omega \in A} f(\omega) := f(\omega_1) + \dots + f(\omega_n).$$

Il s'agit donc d'une simple somme dans \mathbb{R} ;

- si A est infini, alors on pose

$$\sum_{\omega \in A} f(\omega) := \sup \left\{ \sum_{\omega \in B} f(\omega) : B \subset A, B \text{ fini} \right\}. \quad (1.6)$$

Cette définition est valable car ce qu'on somme (notamment $f(\omega)$) est positif ⁽ⁱ⁾.

Mentionnons que, pour une fonction positive f quelconque, si A est infini, $\sum_{\omega \in A} f(\omega)$ peut prendre la valeur $+\infty$.

On utilisera la notation $\sum_{\omega \in A} f(\omega)$ surtout pour des ensembles A dénombrables, ou pour A indénombrable, mais tel qu'il existe $\tilde{A} \subset A$ dénombrable avec $f(\omega) = 0$ si $\omega \notin \tilde{A}$.

Observons que la formule (1.6) ne fait pas intervenir un ordre sur les éléments de A . En effet, si $A = \{\omega_1, \omega_2, \dots\}$ admet un ordre naturel, on peut se dire que $\sum_{\omega \in A} f(\omega)$ peut être

⁽ⁱ⁾Quand f prend des valeurs positives et négatives, la somme $\sum_{\omega \in A} f(\omega)$ n'est pas toujours définie. Pour pouvoir lui donner un sens cohérent, il faut que $\sum_{\omega \in A} |f(\omega)| < \infty$; à ce moment là, on la définit comme $\sum_{\omega \in A} \max\{f(\omega), 0\} - \sum_{\omega \in A} \max\{-f(\omega), 0\}$. Observons que les trois sommes infinies de la phrase précédente font intervenir que des termes positifs, elles sont donc définies par (1.6); de plus $\sum_{\omega \in A} |f(\omega)| < \infty$ assure que la différence qui définit $\sum_{\omega \in A} f(\omega)$ est entre deux termes finis.

défini par

$$\sum_{\omega \in A} f(\omega) = \lim_{k \rightarrow \infty} \sum_{k=1}^n f(\omega_k).$$

Cette relation est vraie, mais on préfère ne pas l'utiliser pour définir $\sum_{\omega \in A} f(\omega)$, car elle fait intervenir un ordre sur A qui n'est pas pertinent.

En fin, mentionnons que cette définition de somme infinie a certaines propriétés intuitives et très convenables: si A_1, A_2, \dots sont des ensembles deux à deux disjoints et on pose $A = \bigsqcup_{n \geq 1} A_n$, alors

$$\sum_{n \geq 1} \sum_{\omega \in A_n} f(\omega) = \sum_{\omega \in A} f(\omega). \quad (1.7)$$

La preuve est laissée en exercice (voir exercice 3).

La proposition 1.6 s'adapte aussi aux univers dénombrables.

Proposition 1.9. *Soit Ω un ensemble dénombrable, non-vidé. Pour toute fonction de masse f , \mathbb{P}_f est bien une probabilité. Inversement, pour toute probabilité \mathbb{P} sur $(\Omega, \mathcal{P}(\Omega))$, il existe une unique fonction de masse f telle que $\mathbb{P}_f(A) = \mathbb{P}(A)$ pour $A \in \mathcal{P}(\Omega)$.*

Preuve: Supposons pour commencer que f est une fonction de masse. Alors $\mathbb{P}_f(\emptyset) = 0$ (par le premier point de la définition de \sum) et \mathbb{P}_f est σ -additive par (1.7).

Inversement, supposons que \mathbb{P} est une probabilité sur un univers dénombrable Ω . Posons $f(\omega) = \mathbb{P}(\{\omega\})$. Pour $A \subset \Omega$, on peut écrire $A = \bigsqcup_{\omega \in A} \{\omega\}$, ou l'union est au plus dénombrable, car Ω , donc A aussi, le sont. Ainsi, par la σ -additivité de \mathbb{P} ,

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

□

Exemple: Supposons qu'on lance une pièce de monnaie jusqu'à obtenir une fois pile. On peut poser dans ce cas la

$$\Omega = \{P, HP, HHP, HHHP, \dots\},$$

et

$$f(\underbrace{H \dots H}_n P) = \mathbb{P}_f(\underbrace{\{H \dots H\}_n P}) = 2^{n+1}, \quad \text{pour } n \geq 0.$$

Attention! Vu les similarités entre le cas fini et celui dénombrable, on appelle les deux des cadres *discrets*.

Observons que pour un univers infini et dénombrable, il n'y a pas de mesure uniforme.

Exercice 2.

Montrer qu'il n'existe pas de fonction $f : \mathbb{N} \rightarrow [0, 1]$ avec $f(x) = f(y)$ pour tout $x, y \in \mathbb{N}$ et $\sum_{n \in \mathbb{N}} f(n) = 1$.

Exercice 3.

Soit Ω un ensemble quelconque et $f : \Omega \rightarrow [0, +\infty)$ une fonction.

- (a) Soient $A_1, \dots, A_N \subset \Omega$ des ensembles finis, deux à deux disjoints. Posons $A = \bigsqcup_{n=1}^N A_n$. Montrer que

$$\sum_{n=1}^N \sum_{\omega \in A_n} f(\omega) = \sum_{\omega \in A} f(\omega).$$

- (b) Soient $A_1, A_2, \dots \subset \Omega$ des ensembles deux à deux disjoints. Posons $A = \bigsqcup_{n=1}^N A_n$. Supposons que $\sum_{\omega \in A} f(\omega) < \infty$. Fixons $\epsilon > 0$. Argumenter l'existence de $B \subset A$, fini tel que

$$\sum_{\omega \in B} f(\omega) \leq \sum_{\omega \in A} f(\omega) - \epsilon.$$

Argumenter l'existence de N tel que $B \cap A_n = \emptyset$ pour tout $n > N$.

Poser $B_n = A_n \cap B$ pour $n = 1, \dots, N$ et utiliser le premier point pour conclure que

$$\sum_{n \geq 1} \sum_{\omega \in A_n} f(\omega) \geq \sum_{\omega \in A} f(\omega) - \epsilon.$$

Conclure que

$$\sum_{n \geq 1} \sum_{\omega \in A_n} f(\omega) \geq \sum_{\omega \in A} f(\omega).$$

- (c) Soient $A_1, A_2, \dots \subset \Omega$ des ensembles deux à deux disjoints. Posons $A = \bigsqcup_{n=1}^N A_n$. Supposons que $\sum_{n \geq 1} \sum_{\omega \in A_n} f(\omega) < \infty$ (ce qui entraîne $\sum_{\omega \in A_n} f(\omega) < \infty$ pour tout n). Fixons $\epsilon > 0$. Argumenter l'existence de $N \geq 1$ tel que

$$\sum_{n=1}^N \sum_{\omega \in A_n} f(\omega) \geq \sum_{n \geq 1} \sum_{\omega \in A_n} f(\omega) - \epsilon.$$

Pour chaque $n \leq N$, argumenter l'existence de $B_n \subset A_n$, fini, tel que

$$\sum_{\omega \in B_n} f(\omega) \geq \sum_{\omega \in A_n} f(\omega) - \frac{\epsilon}{N}.$$

En utilisant le premier point pour les ensembles $(B_n)_{n=1,\dots,N}$, montrer que

$$\sum_{\omega \in A} f(\omega) \geq \sum_{n \geq 1} \sum_{\omega \in A_n} f(\omega) - 2\epsilon.$$

Conclure que

$$\sum_{n \geq 1} \sum_{\omega \in A_n} f(\omega) \leq \sum_{\omega \in A} f(\omega).$$

- (d) Refaire la preuve de (b) et (c) dans le cas où $\sum_{\omega \in A} f(\omega) = \infty$ et $\sum_{n \geq 1} \sum_{\omega \in A_n} f(\omega) = \infty$, respectivement.
- (e) Conclure que (1.7) est toujours vrai.

1.2.3 Univers continus

Commençons par quelques exemples d'expériences qui ne sont pas modélisées par des univers finis ou dénombrables. On va voir que dans ces cas l'utilisation d'une fonction de masse n'est pas adaptée.

Exemple: Supposons qu'on lance une pièce de monnaie une infinité de fois. Alors l'univers Ω à considérer est l'ensemble des suites infinies de P ou F . On le note

$$\Omega = \{P, F\}^{\mathbb{N}} := \{(x_n)_{n \in \mathbb{N}} : x_1, x_2, \dots \in \{P, F\}\}.$$

Cet ensemble est infini et indénombrable; en effet, il peut être mis en bijection avec $[0, 1]$, qui est indénombrable.

On voit bien que la probabilité de toute séquence *infinie* $\omega \in \{P, F\}^{\mathbb{N}}$ est 0. Pour formaliser cela, observons par exemple que la probabilité que notre lancé correspondent à ω pour les n premiers lancers est de 2^{-n} . Ainsi $\mathbb{P}(\{\omega\}) \leq 2^{-n}$ pour tout n . La fonction de masse associée à \mathbb{P} devrait donc être nulle, ce qui contredit sa définition.

On va revenir à cet exemple et à la façon dont on définit \mathbb{P} et la σ -algèbre \mathcal{E} associée dans la partie 1.5

Exemple: Supposons qu'on veut prendre un **nombre réel entre 0 et 1 de manière uniforme**. Le choix de Ω est évident: $\Omega = [0, 1]$. À nouveau, Ω est indénombrable et pour tout $\omega \in [0, 1]$, la probabilité de choisir *exactement* ω est $\mathbb{P}(\{\omega\}) = 0$.

Cet exemple est un cas particulier des probabilités sur \mathbb{R} , dont on parle plus bas dans cette partie.

La grande difficulté quand on parle d'univers indénombrable est de définir la probabilité \mathbb{P} . Cela est impossible si on veut le faire avec $\mathcal{E} = \mathcal{P}(\Omega)$; on doit réduire la σ -algèbre des événements pour pouvoir définir \mathbb{P} . On commence par un lemme qui donne lieu à une définition.

Lemme 1.10. Soient Ω un ensemble non-vide et $(\mathcal{E}_i)_{i \in I}$ une famille quelconque de σ -algèbres sur Ω . Alors

$$\bigcap_{i \in I} \mathcal{E}_i$$

est une σ -algèbre sur Ω .

Preuve: Il s'agit de vérifier simplement les conditions de la définition de σ -algèbre. □

Vu ce lemme, on peut définir une σ -algèbre engendrée.

Définition 1.11. Soient Ω un ensemble et $\mathcal{A} \subset \mathcal{P}(\Omega)$. La σ -algèbre engendrée par \mathcal{A} est la σ -algèbre

$$\sigma(\mathcal{A}) = \bigcap_{\mathcal{E} \text{ } \sigma\text{-algèbre: } \mathcal{A} \subset \mathcal{E}} \mathcal{E}.$$

L'exemple d'univers indénombrable qui va nous intéresser le plus est celui de \mathbb{R} . Pour \mathbb{R} on va considérer la σ -algèbre borélienne.

Définition 1.12. La σ -algèbre borélienne $\mathcal{B}(\mathbb{R})$ est la σ -algèbre sur \mathbb{R} engendrée par l'ensemble des intervalles semi-infinis:

$$\mathcal{B}(\mathbb{R}) := \sigma(\{(-\infty, a] : a \in \mathbb{R}\}).$$

Désormais, quand on parle de l'univers \mathbb{R} (ou des sous-parties de \mathbb{R}), on considère toujours la σ -algèbre borélienne $\mathcal{B}(\mathbb{R})$. Tous les sous-ensembles de \mathbb{R} qu'on rencontre dans ce cours font partie de $\mathcal{B}(\mathbb{R})$ et vont donc être mesurables par les mesures dont on va munir \mathbb{R} .

Définition 1.13. Une fonction de répartition sur \mathbb{R} est une fonction $F : \mathbb{R} \rightarrow [0, 1]$ telle que

- (i) F est croissante,
- (ii) $\lim_{a \rightarrow -\infty} F(a) = 0$ et $\lim_{a \rightarrow +\infty} F(a) = 1$,
- (iii) pour tout $a \in \mathbb{R}$, $F(a) = \lim_{x \searrow a} F(x)$ (continuité à droite).

Les fonctions de répartition sont utilisées pour définir et décrire des probabilités sur \mathbb{R} .

Théorème 1.14. Soit $F : \mathbb{R} \rightarrow [0, 1]$ une fonction de répartition. Alors il existe une unique probabilité \mathbb{P}_F sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que

$$\mathbb{P}_F((-\infty, a]) = F(a), \quad \forall a \in \mathbb{R}. \quad (1.8)$$

Inversement, si \mathbb{P} est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, l'équation (1.8) définit une fonction de répartition.

Ce théorème est admis; il est un de principaux théorèmes de la théorie de la mesure.

Exemple: Pour l'expérience qui consiste à choisir un nombre uniforme sur $[0, 1]$, on prend la probabilité sur \mathbb{R} associée à la fonction de répartition:

$$F(x) = \begin{cases} 0 & \text{si } x < 0; \\ x & \text{si } 0 \leq x < 1; \\ 1 & \text{si } x \geq 1. \end{cases}$$

Par contre, la probabilité $\tilde{\mathbb{P}}$ uniforme sur $\{1, \dots, n\}$, la fonction de répartition est donnée par

$$\tilde{F}(x) = \begin{cases} 0 & \text{si } x < 1; \\ \frac{k}{n} & \text{si } k \leq x < k + 1, \text{ pour } k = 1, \dots, n - 1; \\ 1 & \text{si } x \geq n. \end{cases}$$

Théorème 1.15. Soit $f : \mathbb{R} \rightarrow [0, +\infty)$ une fonction continue par morceaux avec

$$\int_{-\infty}^{+\infty} f(x) dx = 1.$$

Alors il existe une probabilité \mathbb{P}_f sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que

$$\mathbb{P}_f(]a, b]) = \int_a^b f(x) dx, \quad \forall -\infty \leq a \leq b \leq \infty. \quad (1.9)$$

On dit que f est la densité de \mathbb{P}_f .

Attention! Pas toutes les probabilités sur \mathbb{R} sont données par une densité.

Pour une probabilité \mathbb{P} (de fonction de répartition F) qui admet une densité,

l'unique densité f qui soit continue à droite est donnée par

$$f(x) = \lim_{h \searrow 0} \frac{\mathbb{P}_f(]x, x+h])}{h} = \lim_{h \searrow 0} \frac{F(x+h) - F(x)}{h}, \quad x \in \mathbb{R}.$$

En particulier, la fonction de répartition F de \mathbb{P} admet dans ce cas la une dérivée à droite en tout point et cette dérivée est la densité f .

Ainsi $f(x)$ mesure la probabilité d'être "aux point x " (ou plus précisément "proche du point x "); pourtant, à cause de la division par h , ce n'est pas une probabilité en sois. Ainsi, on ne devrait pas penser qu'une densité prend des valeurs uniquement dans $[0, 1]$!

Preuve: Existence. Soit $f : \mathbb{R} \rightarrow [0, +\infty)$ comme dans le théorème. Alors la fonction F définie par

$$F(a) = \int_{-\infty}^a f(x) dx, \quad \forall a \in \mathbb{R}$$

est une fonction de répartition. Elle est en effet croissante, car f est positive; elle est continue car f est bornée (à vérifier...); $\lim_{a \rightarrow -\infty} F(a) = 0$ par la définition de l'intégrale et $\lim_{a \rightarrow +\infty} F(a) = \int_{-\infty}^{\infty} f(x) dx = 1$.

Notons \mathbb{P} la probabilité de fonction de répartition F . Alors, pour $-\infty \leq a \leq b \leq \infty$,

$$\mathbb{P}_f(]a, b]) = \mathbb{P}_f((-\infty, b]) - \mathbb{P}_f((-\infty, a]) = \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx = \int_a^b f(x) dx.$$

Unicité: Supposons que \mathbb{P} et $\tilde{\mathbb{P}}$ sont deux probabilités satisfaisant (1.9). Alors $\mathbb{P}((-\infty, a]) = \tilde{\mathbb{P}}((-\infty, a])$ pour tout $a \in \mathbb{R}$. Donc les deux probabilités ont la même fonction de répartition, et par le théorème 1.14, elles sont égales. \square

Exemple: La densité d'une probabilité uniforme sur $[0, 1]$ est

$$f(x) = \begin{cases} 0 & \text{si } x < 0; \\ 1 & \text{si } 0 \leq x < 1; \\ 0 & \text{si } x \geq 1. \end{cases}$$

Exercice 4.

Soit \mathbb{P} une probabilité sur \mathbb{R} avec une densité f . Montrer que pour tout $x \in \mathbb{R}$, $\mathbb{P}(\{x\}) = 0$. Donner un exemple de probabilité sur \mathbb{R} qui n'admet pas de densité.

Exercice 5.

Donner un exemple de deux fonctions de densité différentes f_1, f_2 donnant la même probabilité \mathbb{P} sur \mathbb{R} .

Montrer que si on impose que les densités soient continues à droite, alors une probabilité sur \mathbb{R} admet au plus une densité. *Indication:* utiliser la remarque sous le théorème 1.15.

Exercice 6.

Soit F une fonction de répartition. Le but de cet exercice est de montrer que F admet un nombre au plus dénombrable de points de discontinuité.

- (a) Pour $x \in \mathbb{R}$ soit $m_x = F(x) - \lim_{y \nearrow x} F(y)$. Montrer que $m_x \geq 0$ pour tout $x \in \mathbb{R}$ et que $m_x > 0$ si et seulement si F est discontinue en x .
- (b) Pour tout $N \geq 1$ montrer qu'il existe au plus N points $x \in \mathbb{R}$ avec $m_x \geq \frac{1}{N}$.
- (c) Conclure qu'il existe au plus un nombre dénombrable de points $x \in \mathbb{R}$ avec $m_x > 0$.

1.3 Conditionnement

Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilité, fini, dénombrable ou indénombrable. De plus, soit $B \in \mathcal{E}$ avec $\mathbb{P}(B) > 0$.

Définition 1.16. La probabilité conditionnelle d'un événement $A \in \mathcal{E}$ sachant B est

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

On insiste sur le fait que la probabilité de A sachant B est définie uniquement si $\mathbb{P}(B) > 0$.

Les probabilités conditionnelles apparaissent très naturellement: si on joue au poker texas hold'em, nos prévisions changent au fur et à mesure que cartes sont retournées!

Exemple: Des fois, la probabilité conditionnelle est facile à calculer: quelle est la probabilité d'avoir un score de au plus 7 sur un lancé de deux dés? Et si le premier dé tombe sur 5? Un calcul direct montre que les deux probabilités sont de $7/12$ et $1/3$, respectivement.

Dans d'autres cas on peut avoir des surprises qui vont à l'encontre de notre intuition: On choisie deux cartes dans un paquet de 52. Quelle est la probabilité d'avoir deux as? Et si on sait qu'on en a au moins un? Et si on sait qu'on a l'as de pique?

Si on note C l'ensemble des 52 cartes, l'univers à considérer est $\mathcal{P}_2(C)$, l'ensemble de toutes les paires de cartes distinctes de C . Notons A , A^\spadesuit et A^2 les événements d'avoir aux moins un as, l'as de pique et deux as, respectivement. Alors on a

$$\mathbb{P}(A^2) = \frac{6}{\binom{52}{2}}, \quad \mathbb{P}(A) = \frac{6 + 4 \cdot 48}{\binom{52}{2}}, \quad \mathbb{P}(A^\spadesuit) = \frac{51}{\binom{52}{2}}, \quad \mathbb{P}(A^\spadesuit \cap A^2) = \frac{3}{\binom{52}{2}},$$

d'où

$$\mathbb{P}(A^2|A) = \frac{6}{6 + 4 \cdot 48} = \frac{1}{33} \quad \text{et} \quad \mathbb{P}(A^2|A^\spadesuit) = \frac{3}{51} = \frac{1}{17}.$$

Proposition 1.17. *La fonction $\mathbb{P}(\cdot|B) : \mathcal{E} \rightarrow \mathbb{R}$ est une probabilité sur (Ω, \mathcal{E}) .*

Preuve: Observons que, pour tout $A \in \mathcal{E}$, $0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(B)$, ce qui implique que $\mathbb{P}(A|B) \in [0, 1]$.

De plus

$$\mathbb{P}(\emptyset|B) = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(B)} = 0 \quad \text{et} \quad \mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

En fin, soient A_1, A_2, \dots des événements deux à deux disjoints. Alors

$$\mathbb{P}\left(\bigsqcup_{n \geq 1} A_n \mid B\right) = \frac{\mathbb{P}\left(\bigsqcup_{n \geq 1} A_n \cap B\right)}{\mathbb{P}(B)} = \frac{\sum_{n \geq 1} \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n \geq 1} \mathbb{P}(A_n|B).$$

□

Proposition 1.18 (formule de Bayes). *Soient $A, B \in \mathcal{E}$ avec $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$. Alors*

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

Preuve: Par définition on a

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

□

On dit qu'une famille $(B_i)_{i \in I}$ est une *partition* de Ω si les ensembles $(B_i)_{i \in I}$ sont dans \mathcal{E} , ils sont deux à deux disjoints et

$$\bigsqcup_{i \in I} B_i = \Omega.$$

Proposition 1.19 (loi de probabilité totale). *Soient $A \in \mathcal{E}$ et $(B_i)_{i \in I}$ est une partition de Ω avec I fini ou dénombrable et $\mathbb{P}(B_i) > 0$ pour tout $i \in I$. Alors*

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

Preuve: Du fait que $(B_i)_{i \in I}$ est une partition de Ω , on a que

$$\bigsqcup_{i \in I} (A \cap B_i) = A,$$

ou l'union est disjointe. Par la σ -additivité de \mathbb{P} ,

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

□

La formule de Bayes et la loi de probabilité totale sont surtout utiles quand on veut calculer la probabilité d'une cause, sachant l'effet. On rencontre des cas où la probabilité d'un résultat est affecté par une certaine cause. Si on connaît la cause, on estime facilement la probabilité (conditionnelle) du résultat. Mais si on observe le résultat, peut-on estimer la cause?

Si A est un événement qui porte sur le résultat et $(B_i)_{i \in I}$ sont différentes causes possibles qui produisent A avec différentes probabilités, alors on utilise la formule de Bayes et la loi de probabilité totale ensemble sous la forme

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

Exemple: On dispose de deux urnes, la première contenant 2 boules blanches et une noire, la deuxième contenant 1 boule blanche et 5 noires. On extrait au hasard une boule de la première urne et on la met dans la deuxième. On tire ensuite une boule de la deuxième urne; on observe qu'elle est blanche. Quelle est la probabilité que la boule tirée de la première urne était blanche?

Soit A l'événement que la première boule tirée soit blanche, et B que la deuxième soit blanche. On veut calculer $\mathbb{P}(A|B)$. Pourtant, ce qu'on peut estimer facilement est

$$\mathbb{P}(A) = \frac{2}{3}, \quad \mathbb{P}(B|A) = \frac{2}{7}, \quad \mathbb{P}(B|A^c) = \frac{1}{7}.$$

De plus, A et A^c forment une partition de l'espace de probabilité. Ainsi on peut écrire:

$$\mathbb{P}(A|B) = \mathbb{P}(B|A) \frac{\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} = \frac{\frac{2}{7} \cdot \frac{2}{3}}{\frac{2}{7} \cdot \frac{2}{3} + \frac{1}{7} \cdot \frac{1}{3}} = \frac{4}{5}.$$

Dans cet exemple, le résultat est la boule finale (donc l'événement B qui s'y appuie) et la cause est la boule transférée de la première urne vers la deuxième (décrite par A et A^c).

Exemple: Les chiffres admissions à l'université de Berkley en 1973 en fonction du sexe étaient les suivantes ⁽ⁱⁱ⁾

	Applicants	Admis	
Hommes	2691	44%	(1.10)
Femmes	1835	32%	

Ainsi on peut croire que l'université est biaisée en faveur des hommes. Pourtant, si on regarde les chiffres d'admissions par département on obtient une image différente:

Département	Hommes		Femmes	
	Applicants	Admis	Applicants	Admis
Engineering	825	61%	108	82%
Physics	560	63%	25	68%
Social sciences	325	37%	593	40%
Mathematics	417	33%	375	35%
English	191	28%	393	24%
Philosophy	373	6%	341	7%

Notons que ces chiffres sont cohérentes avec le tableau (1.10) (appliquer la loi de probabilité totale). Ainsi chaque département a admis en plus grande proportion les femmes que les hommes. Pourtant, sur l'intégralité de l'université, la proportion d'admission est plus grande parmi les hommes.

L'explication se trouve dans le fait que les femmes ont candidaté surtout dans des départements avec des taux d'admissions bas, alors que les hommes se sont concentré sur ceux avec des taux d'admissions élevés.

1.4 Indépendance

On fixe un espace de probabilité $(\Omega, \mathcal{E}, \mathbb{P})$ fini, dénombrable ou indénombrable.

Définition 1.20. Soient $A, B \in \mathcal{E}$ deux événement. On dit que A et B sont indépendants, et on l'écrit $A \perp\!\!\!\perp B$ si

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

⁽ⁱⁱ⁾Le cas du biais des admissions à Berkley en 1973 est réel; les chiffres présentés ici ont été légèrement modifiés

Une famille d'événements $(A_i)_{i \in I}$ est dite indépendante si, pour tout $J \subset I$ fini,

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

Lemme 1.21. Soient A, B deux événements avec $\mathbb{P}(B) > 0$. Alors A et B sont indépendants si et seulement si

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

L'indépendance apparaît surtout quand on modélise plusieurs expériences faites séparément, sans interaction entre elles. Néanmoins, il existe des situations où des expériences fortement liées produisent des résultats indépendants.

Exemple: On lance une pièce de monnaie n fois. La pièce est biaisée et tombe sur pile avec probabilité $p \in [0, 1]$. Notons A l'événement qu'on obtient pile au premier lancé et B_k l'événement qu'on obtient k fois pile pendant les n lancers. Alors

$$\mathbb{P}(A) = p, \quad \mathbb{P}(B_k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \mathbb{P}(A \cap B_k) = p \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}.$$

Ainsi, si $p = \frac{k}{n}$, les événements A et B_k sont indépendants.

On dit que des événements $(A_i)_{i \in I}$ sont deux à deux indépendants, si pour tout $i, j \in I$ avec $i \neq j$, $A_i \perp\!\!\!\perp A_j$.

Exercice 7.

(a) Donner un exemple de trois événements A_1, A_2, A_3 qui sont deux à deux indépendants, mais pas tous indépendants.

(b) Donner un exemple de trois événements A_1, A_2, A_3 avec $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$, mais qui ne sont pas indépendants.

Exercice 8.

Que peut-on dire d'un événement A qui est indépendant de lui-même?

Exercice 9.

Soient A et B deux événements. Montrer que si $\mathbb{P}(A) \in \{0, 1\}$ alors $A \perp\!\!\!\perp B$.

Lemme 1.22. Soient A, B deux événements indépendants. Alors A^c et B sont aussi indépendants.

Un lemme similaire s'applique aux familles d'événements indépendants.

Dans certaines situations on peut avoir des événements qui ne sont pas indépendants, mais pour lesquels on peut identifier la cause de leur dépendance. Ainsi on introduit le concept d'*indépendance conditionnelle*.

Définition 1.23. Soient $(A_i)_{i \in I}$ et B des événements avec $\mathbb{P}(B) > 0$. On dit que la famille $(A_i)_{i \in I}$ est indépendante conditionnellement à B si elle est indépendante pour la probabilité $\mathbb{P}(\cdot|B)$. Cela revient à

$$\mathbb{P}\left(\bigcap_{i \in J} A_i \mid B\right) = \prod_{i \in J} \mathbb{P}(A_i|B), \quad \forall J \subset I, J \text{ fini.}$$

Exemple: On dispose de deux pièces de monnaie. Une est biaisée pour tomber sur pile avec probabilité p (un nombre entre 0 et 1, différent de $1/2$), l'autre tombe sur chaque face avec probabilité $1/2$. On choisie une des pièces au hasard et on la lance deux fois. Notons A et B les événements d'obtenir pile au premier et deuxième lancé, respectivement. Alors A et B ne sont pas indépendants:

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}\left(p + \frac{1}{2}\right), \quad \mathbb{P}(A \cap B) = \frac{1}{2}\left(p^2 + \frac{1}{4}\right).$$

Par contre, si on défini C comme l'événement de choisir la pièce biaisée, alors A et B sont indépendants conditionnellement à C et C^c .

$$\begin{aligned} \mathbb{P}(A|C) = \mathbb{P}(B|C) &= p, & \mathbb{P}(A \cap B|C) &= p^2 & \text{et} \\ \mathbb{P}(A|C^c) = \mathbb{P}(B|C^c) &= \frac{1}{2}, & \mathbb{P}(A \cap B|C^c) &= \frac{1}{4}. \end{aligned}$$

Attention! Il n'y a pas d'implication générale entre l'indépendance et l'indépendance conditionnelle.

1.5 Espaces produit

Supposons qu'on fait deux (ou plusieurs) expériences, indépendamment, chaque décrite par un espace de probabilité. Il est alors utile de construire un espace de probabilité total, permettant de décrire simultanément les deux expériences, ainsi que la relation d'indépendance entre elles.

1.5.1 Produits finis

Soient $(\Omega_1, \mathcal{E}_1, \mathcal{P}_1)$ et $(\Omega_2, \mathcal{E}_2, \mathcal{P}_2)$ deux espaces de probabilités. On pose alors

$$\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}.$$

Sur $\Omega_1 \times \Omega_2$ on considère la σ -algèbre $\mathcal{E}_1 \otimes \mathcal{E}_2$ engendrée par les *rectangles*, c.-à-d. par les ensembles de la forme $A \times B$ ou $A \in \mathcal{E}_1$ et $B \in \mathcal{E}_2$:

$$\mathcal{E}_1 \otimes \mathcal{E}_2 = \sigma(\{A \times B : A \in \mathcal{E}_1, B \in \mathcal{E}_2\}).$$

En fin, on mentionne un résultat qu'on va admettre ici et qui nous permet de construire une probabilité sur $(\Omega_1 \times \Omega_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$

Théorème 1.24. *Il existe une unique probabilité \mathbb{P} sur $(\Omega_1 \times \Omega_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ telle que, pour tout $A \in \mathcal{E}_1$ et $B \in \mathcal{E}_2$*

$$\mathbb{P}(A \times B) = \mathbb{P}_1(A) \mathbb{P}_2(B).$$

On appelle l'espace de probabilité $(\Omega_1 \times \Omega_2, \mathcal{E}_1 \otimes \mathcal{E}_2, \mathbb{P})$ l'espace produit de $(\Omega_1, \mathcal{E}_1, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{E}_2, \mathbb{P}_2)$.

Observons que, pour tout $A \in \mathcal{E}_1$ et $B \in \mathcal{E}_2$, les événements correspondants dans l'espace produit, à savoir $A \times \Omega_2$ et $\Omega_1 \times B$, respectivement, sont indépendants. Ainsi l'espace produit décrit bien deux expériences indépendantes.

En utilisant plusieurs fois cette construction, on peut créer des produits de tout nombre *fini* d'espaces de probabilités.

Produit d'espaces discrets. Supposons que $(\Omega_1, \mathcal{E}_1, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{E}_2, \mathbb{P}_2)$ sont deux espaces de probabilités discrets. Rappelons que la probabilité \mathbb{P}_i ($i = 1, 2$) est donnée par la fonction de masse

$$f_i(\omega) = \mathbb{P}_i(\{\omega\}), \quad \forall \omega \in \Omega_i.$$

Dans ce contexte, la construction de l'espace produit est explicite: l'univers $\Omega_1 \times \Omega_2$ est discret et $\mathcal{E}_1 \otimes \mathcal{E}_2 = \mathcal{P}(\Omega_1 \times \Omega_2)$. De plus, la probabilité produit \mathbb{P} est donnée par la fonction de masse

$$f(\omega_1, \omega_2) = f_1(\omega_1)f_2(\omega_2), \quad \forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2.$$

Exemple: On fait l'expérience suivante: on lance un dé et, séparément, une pièce de monnaie biaisée, avec probabilité p de tomber sur pile. Le premier lancé est décrit par l'univers

$$\Omega_1 = \{1, 2, \dots, 6\} \quad \text{avec } \mathbb{P}_1 \text{ uniforme.}$$

Le deuxième lancé est décrit par l'univers

$$\Omega_2 = \{P, F\} \quad \text{avec } \mathbb{P}_2 \text{ donnée par } \mathbb{P}_2(\{P\}) = p \text{ et } \mathbb{P}_2(\{F\}) = 1 - p.$$

L'expérience globale est alors décrite par

$$\Omega = \{1P, 1F, 2P, 2F, \dots, 6P, 6F\}$$

et une probabilité \mathbb{P} avec fonction de masse:

$$\mathbb{P}(\{iP\}) = \frac{p}{6} \quad \text{et} \quad \mathbb{P}(\{iF\}) = \frac{1-p}{6}, \quad \forall i \in \{1, \dots, 6\}.$$

Produit d'espaces avec mesures à densité. Supposons que $(\Omega_1, \mathcal{E}_1, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{E}_2, \mathbb{P}_2)$ sont deux espaces de probabilités continus, avec $\Omega_1 = \Omega_2 = \mathbb{R}$, $\mathcal{E}_1 = \mathcal{E}_2 = \mathcal{B}(\mathbb{R}^2)$ et les probabilités \mathbb{P}_1 et \mathbb{P}_2 ayant des densités f_1 et f_2 , respectivement. Alors, la probabilité \mathbb{P} sur l'espace produit \mathbb{R}^2 est aussi donnée par une densité, à savoir $f(x, y) = f_1(x)f_2(y)$, $\forall x, y \in \mathbb{R}$. Ainsi, pour tout ensemble $A \in \mathcal{B}(\mathbb{R}^2)$,

$$\mathbb{P}(A) = \int_A f(x, y) dx dy. \quad (1.11)$$

Dans toutes les situations qu'on va étudier, l'ensemble A va avoir une structure simple, et on va pouvoir calculer l'intégrale au-dessus en intégrant coordonnée par coordonnée.

Plus généralement, pour décrire des probabilités \mathbb{P} sur \mathbb{R}^d avec $d \geq 2$, on peut utiliser la fonction de répartition multidimensionnelle

$$F(a_1, \dots, a_d) = \mathbb{P}((-\infty, a_1] \times \dots \times (-\infty, a_d]).$$

Exemple: Prenons un nombre uniformément entre 0 et 1 et un autre, indépendamment, entre 0 et 2. Quelle est la probabilité que le premier soit plus petit que le deuxième?

Les espaces de probabilité pour les deux expériences individuelles sont $\Omega_1 = [0, 1]$ et $\Omega_2 = [0, 2]$ avec les probabilités \mathbb{P}_1 et \mathbb{P}_2 de densités $f_1 = \mathbf{1}_{[0,1]}$ et $f_2 = \frac{1}{2}\mathbf{1}_{[0,2]}$, respectivement.

L'espace décrivant les deux expériences est l'espace produit, à savoir $\Omega = [0, 1] \times [0, 2]$ avec la probabilité \mathbb{P} de densité $f(x, y) = \frac{1}{2}\mathbf{1}_{(x,y) \in [0,1] \times [0,2]}$. On est intéressé par la probabilité de l'événement $A = \{(x, y) \in \Omega : x \leq y\}$. En vertu de (1.11), on a

$$\mathbb{P}(A) = \int_{x=0}^1 \int_{y=x}^2 \frac{1}{2} dx dy = \int_{x=0}^1 \left(1 - \frac{x}{2}\right) dx = 1 - \left[\frac{x^2}{4}\right]_0^1 = \frac{3}{4}.$$

1.5.2 Produits infinis

Dans certaines situations, il est utile de considérer des produits infinis d'espaces de probabilités. Un exemple typique est celui d'une suite infinie de lancers de pile ou face. Pour décrire un lancer on utilise l'univers fini $\Omega = \{P, F\}$ avec la probabilité uniforme. Pour décrire la suite de lancers l'univers naturel est

$$\Omega^{\mathbb{N}} = \{(\omega_n)_{n \geq 1} : \omega_n \in \Omega \forall n \geq 1\}.$$

Cet univers est indénombrable, et on doit donc prendre soin à bien définir une σ -algèbre dessus.

Soit $(\Omega_n, \mathcal{E}_n, \mathbb{P}_n)_{n \geq 1}$ une suite d'espaces de probabilités. Posons

$$\Omega = \Omega_1 \times \Omega_2 \times \dots = \{(\omega_n)_{n \geq 1} : \omega_n \in \Omega_n \forall n \geq 1\}.$$

Sur Ω , considérons la σ -algèbre \mathcal{E} engendrée par les événements du type

$$A_1 \times A_2 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots, \quad \text{avec } A_1 \in \mathcal{E}_1, \dots, A_n \in \mathcal{E}_n,$$

qu'on appelle des *cylindres*. Alors, il existe une unique probabilité \mathbb{P} sur (Ω, \mathcal{E}) telle que, pour tout cylindre comme au-dessus,

$$\mathbb{P}(A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n).$$

L'existence de la probabilité \mathbb{P} découle du théorème d'extension de Kolmogorov, un des théorèmes fondateur de la théorie des probabilités.

Chapter 2

Variables aléatoires

2.1 Variables aléatoires et leur lois

Fixons un espace de probabilité $(\Omega, \mathcal{E}, \mathbb{P})$.

Définition 2.1. Une variable aléatoire sur $(\Omega, \mathcal{E}, \mathbb{P})$ est une fonction $X : \Omega \rightarrow \mathbb{R}$ telle que, pour tout $a \in \mathbb{R}$,

$$X^{-1}((-\infty, a]) = \{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{E}. \quad (2.1)$$

Grace à la condition (2.1), on dit que X est une fonction mesurable. Il s'en suit que, pour tout $A \in \mathcal{B}(\mathbb{R})$, $X^{-1}(A) \in \mathcal{E}$, et on peut parler de la probabilité de cet ensemble. La fonction

$$\begin{aligned} \mu_X : \mathcal{B}(\mathbb{R}) &\rightarrow [0, 1] \\ A &\mapsto \mathbb{P}(X^{-1}(A)) \end{aligned}$$

définit une probabilité sur \mathbb{R} . On appelle cette probabilité la *loi* de la variable aléatoire X .

Attention! Dire que μ_X est une probabilité sur \mathbb{R} signifie que μ_X satisfait les conditions de la définition 1.3, avec $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ jouant le rôle de (Ω, \mathcal{E}) . La différence de notation entre \mathbb{P} et μ_X vient du fait que \mathbb{P} est définie sur l'univers abstrait Ω (et ne dépend pas de X), alors que μ_X est définie sur l'espace \mathbb{R} où X prend ses valeurs.

Comme vu dans le théorème 1.14, la loi μ_X d'une variable aléatoire X est décrite par sa fonction de répartition

$$F_X(t) = \mu_X((-\infty, t]) = \mathbb{P}(X \leq t).$$

Un mot sur les notations. Les variables aléatoires sont souvent, mais pas exclusivement, notées par des majuscules latines (X, Y, Z, \dots). Pour $A \in \mathcal{B}(\mathbb{R})$, on utilise la notation

$$\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}).$$

L'espace de probabilité $(\Omega, \mathcal{E}, \mathbb{P})$ n'est pas en général explicite; les quantités qu'on va étudier vont dépendre uniquement des lois des variables aléatoires qu'on considère, pas du choix précis de l'espace de probabilité. Retenons, qu'on peut définir des variables aléatoires de même loi sur des espaces de probabilité différents.

En fin, mentionnons que pour toute probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, il existe un espace de probabilité et une variable aléatoire définie dessus, dont la loi est μ .

Preuve: du fait que μ_X est une probabilité sur \mathbb{R} : Fixons une variable aléatoire X définie sur un espace de probabilité $(\Omega, \mathcal{E}, \mathbb{P})$. On vérifie facilement que la collection de sous-ensembles de \mathbb{R}

$$\{A \subset \mathbb{R} : X^{-1}(A) \in \mathcal{E}\}$$

est une σ -algèbre. Par (2.1) elle contient tout intervalle de la forme $(-\infty, a]$, et par conséquent la tribu que ceux-ci engendrent, à savoir $\mathcal{B}(\mathbb{R})$. Ainsi la fonction μ_X est bien définie.

De plus, $\mu_X(\emptyset) = \mathbb{P}(X \in \emptyset) = \mathbb{P}(\emptyset) = 0$ et $\mu_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1$. En fin, pour $A_1, A_2, \dots \in \mathcal{B}(\mathbb{R})$ deux à deux disjoints,

$$\mu_X\left(\bigsqcup_{n \geq 1} A_n\right) = \mathbb{P}\left(X \in \bigsqcup_{n \geq 1} A_n\right) = \mathbb{P}\left(\bigsqcup_{n \geq 1} X^{-1}(A_n)\right) = \sum_{n \geq 1} \mathbb{P}(X^{-1}(A_n)) = \sum_{n \geq 1} \mu_X(A_n).$$

On conclut que μ_X est bien une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. □

2.2 Exemples

Certaines loi (et les variables aléatoires associées) décrivent des situations naturelles, et donc portent des noms spécifiques. On en liste ici les plus courantes.

2.2.1 Lois discrètes

On dit qu'une variable aléatoire X (ou de manière équivalente, une loi μ_X) est discrète s'il existe un ensemble fini ou dénombrable $A \subset \mathbb{R}$ tel que

$$\mathbb{P}(X \in A) = 1 \quad (\text{ou de manière équivalente } \mu_X(A) = 1).$$

Il s'en suit que $\mathbb{P}(X \notin A) = 0$.

Une telle variable est décrite par la fonction de masse f_X :

$$f_X(t) = \mathbb{P}(X = t) \in [0, 1], \quad \forall t \in \mathbb{R}.$$

Observons que

$$\sum_{t \in \mathbb{R}} f_X(t) = \sum_{t \in A} f_X(t) = \mathbb{P}(X \in A) = 1.$$

On appelle le support de X l'ensemble $\{t \in \mathbb{R} : \mathbb{P}(X = t) > 0\}$; on dit que X prend des valeurs dans cet ensemble.

Attention! Cette définition de support s'applique uniquement aux variables discrètes.

Loi uniforme Soit $A \subset \mathbb{R}$ un ensemble fini, non vide. Une variable aléatoire uniforme sur A est une variable X telle que

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

Il s'en suit que $\mathbb{P}(X \in A) = 1$.

Loi de Bernoulli Une variable de Bernoulli de paramètre $p \in [0, 1]$ est une variable à valeurs dans $\{0, 1\}$, avec

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

Cette loi est utilisée pour décrire le résultat d'un lancé de pile ou face d'une pièce biaisée. Ainsi, si la pièce a probabilité p de tomber sur pile et on définit X comme 1 si on obtient pile et 0 si on obtient face, alors X suit une loi de Bernoulli de paramètre p .

Loi binomiale Une variable X est dite binomiale de paramètres $p \in [0, 1]$ et $n \geq 1$ si elle prend valeurs dans $\{0, \dots, n\}$, avec

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

Supposons qu'on lance la pièce du point précédent n fois, et qu'on note X le nombre total de piles obtenus. Alors X suit une loi binomiale de paramètres n, p .

Loi géométrique Une variable X est dite géométrique de paramètre $p \in [0, 1]$ si elle prend valeurs dans $\mathbb{N}^* = \{1, 2, \dots\}$, avec

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

Lançons la même pièce jusqu'à obtenir pile une première fois. Le nombre de lancers effectués suit alors une loi géométrique de paramètre p .

Fixons une variable géométrique X de paramètre p (on écrit cela $X \sim \text{Geo}(p)$). Alors, pour $k \in \mathbb{N}$ on a

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

Il s'en suit que, pour $k, \ell \geq 0$,

$$\mathbb{P}(X > k + \ell \mid X > \ell) = \frac{(1-p)^{k+\ell}}{(1-p)^\ell} = \mathbb{P}(X > k).$$

Cette propriété s'appelle *l'absence de mémoire* des variables géométriques.

Exercice 10.

Montrer que les variables géométriques sont les seules variables aléatoires X à valeurs dans \mathbb{N}^* qui satisfont:

$$\mathbb{P}(X > k + \ell \mid X > \ell) = \mathbb{P}(X > k), \quad \forall k, \ell \geq 0, \text{ entiers.}$$

Loi de Poisson Une variable X est dite de Poisson de paramètre $\lambda > 0$ si elle prend valeurs dans $\mathbb{N} = \{0, 1, 2, \dots\}$, avec

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

La loi de Poisson décrit les occurrences d'événements rares, comme le nombre de fois que la loterie est gagnée. Fixons $\lambda > 0$. Pour $n \geq 1/\lambda$, soit Y_n une variable de Bernoulli de paramètres $n, \lambda/n$. Alors, pour $k \geq 0$ fixé,

$$\begin{aligned} \mathbb{P}(Y_n = k) &= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \cdot \frac{n(n-1)\dots(n-k+1)}{(n-\lambda)^k} \xrightarrow{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^k}{k!}. \end{aligned}$$

Ainsi, on dit que la suite $(Y_n)_n$ admet comme loi limite la loi de Poisson de paramètre λ .

Loi de Dirac En fin, mentionnons qu'on peut considérer une variable aléatoire qui prend une unique valeur avec probabilité 1. Pour $t \in \mathbb{R}$, une telle variable X est telle que $\mathbb{P}(X = t) = 1$ et $\mathbb{P}(X \neq t) = 0$. Sa loi est la loi de Dirac en t , notée δ_t . On dit que X est déterministe pour accentuer le manque d'aléa.

2.2.2 Lois continues (à densité)

On peut imaginer des variables aléatoires (et leur lois) qui prennent des valeurs dans un ensemble "continu", à savoir dans un ensemble non dénombrable. Il y a une très grande variété de telles lois, et leur étude nécessite des outils avancés. On va se placer dans le cadre le plus souvent utilisé, notamment celui des variables à densité.

Définition 2.2. Une variable aléatoire X (ou sa loi μ_X) admet une densité, s'il existe une

fonction continue par morceaux (ou plus généralement intégrable) $f : \mathbb{R} \rightarrow [0, \infty)$ telle que

$$\mathbb{P}(X \in [a, b]) = \mu_X([a, b]) = \int_a^b f(x) dx.$$

On dit que f est la densité de μ ⁽ⁱ⁾.

Pour une variable X admettant la densité f_X , on a

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^t f_X(s) ds.$$

En particulier

$$\int_{-\infty}^{\infty} f_X(s) ds = 1.$$

On peut montrer que, plus généralement, pour tout $A \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(X \in A) = \int_A f_X(s) ds.$$

Cette formule est très utile quand on travaille avec des variables à densité.

Attention! Pas toutes les variables admettent une densité; il se peut même que F_X soit continue, sans que X admette une densité.

Exercice 11.

Montrer que si X est une variable aléatoire à densité, alors

$$\mathbb{P}(X = t) = 0, \quad \forall t \in \mathbb{R}.$$

Loi uniforme Soient $a < b$. Alors la loi uniforme sur l'intervalle $[a, b]$ est donnée par la densité

$$f_U(t) = \frac{1}{b-a} \mathbf{1}_{a \leq t \leq b}.$$

Loi exponentielle On dit qu'une variable aléatoire X a une loi exponentielle de paramètre $\lambda > 0$ si sa loi est donnée par la densité

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{x \geq 0}.$$

Fixons X une variable exponentielle de paramètre $\lambda > 0$ (on écrit cela $X \sim \text{Exp}(\lambda)$). Alors, pour $t \geq 0$,

$$\mathbb{P}(X > t) = \int_t^{\infty} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_t^{\infty} = e^{-\lambda t}.$$

⁽ⁱ⁾La densité est essentiellement unique. On peut demander que f soit continue à droite et admettant des limites à gauche en tout point, pour la rendre unique.

(Observons en particulier que cela montre bien que $\int_{-\infty}^{\infty} \lambda e^{-\lambda x} dx = 1$). Ainsi, la fonction de répartition de X est

$$F_X(t) = 1 - e^{-\lambda t}.$$

De plus, on observe que, pour $s, t \geq 0$,

$$\mathbb{P}(X > t + s \mid X > t) = \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > t)} = e^{-\lambda(t+s)} e^{\lambda t} = e^{-\lambda s} = \mathbb{P}(X > s). \quad (2.2)$$

Cette propriété s'appelle *l'absence de mémoire* des variables exponentielles.

Exercice 12.

Montrer que les variables exponentielles sont les seules variables aléatoires X à valeurs dans $[0, \infty)$ à satisfaire

$$\mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s), \quad \forall s, t \geq 0.$$

Loi normale On dit qu'une variable aléatoire N a une loi normale de moyenne $\mu \in \mathbb{R}$ et variance σ^2 (ou $\sigma > 0$) si elle admet la densité

$$f_N(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \quad (2.3)$$

On écrit $N \sim \mathcal{N}(\mu, \sigma^2)$.

La loi normale apparaît naturellement comme limite de différents phénomènes naturelles; c'est une des lois les plus importantes des probabilités et statistiques. Mentionnons que sa fonction de répartition

$$\Phi_{\mu, \sigma^2}(t) = \mathbb{P}(N \leq t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^t \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx,$$

n'admet pas de formule explicite.

La loi normale de moyenne 0 et variance 1 est appelé la loi *normale standard*. On va voir par la suite, que si $X \sim \mathcal{N}(0, 1)$ alors $\mu + \sigma X \sim \mathcal{N}(\mu, \sigma^2)$.

Exercice 13 (Attention! utilise le changement de variables multidimensionnel).

(a) Posons $I(0, 1) = \int_{-\infty}^{\infty} e^{-x^2/2} dx$. Montrer que

$$I(0, 1)^2 = \int \int_{(x,y) \in \mathbb{R}^2} \exp\left(-\frac{\|(x,y)\|^2}{2}\right) dx,$$

où $\|\cdot\|$ désigne la norme euclidienne. Conclure par un changement de variable que $I(0, 1)^2 = 2\pi$.

(b) En appliquant un changement de variables, montrer que pour $\mu \in \mathbb{R}$ et $\sigma > 0$,

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sigma I(0, 1).$$

(c) Conclure que (2.3) définit bien une densité.

Loi gamma Un variable aléatoire X a une loi Gamma de paramètres $\lambda > 0$ et $t > 0$ si elle admet la densité

$$f_X(x) = \frac{1}{\Gamma(t)} \lambda^t x^{t-1} e^{-\lambda x} \mathbf{1}_{x \geq 0},$$

ou

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx.$$

Lorsque $\lambda = 1/2$ et $t = \frac{1}{2}d$ on dit que X suit une loi du χ^2 à d degrés de liberté. Cette loi apparaît souvent en statistique.

Exercice 14.

Fixons $t > 0$. Observons que, si $\lambda = 1$, alors $\int_0^{+\infty} f_X(x) dx = 1$. Appliquer un changement de variables pour déduire cela pour toute valeur de $\lambda > 0$.

Exercice 15.

Utiliser une intégration par parties pour montrer que, pour $t > 0$,

$$\Gamma(t+1) = t\Gamma(t).$$

Calculer $\Gamma(1)$, ensuite utiliser la formule précédente pour déduire $\Gamma(t)$ pour tout entier $t \geq 1$.

Exercice 16 (pour plus tard).

Soient X_1, \dots, X_n des v.a. i.i.d. exponentielles de paramètre $\lambda > 0$. Montrer que $S_n = X_1 + \dots + X_n$ suit une loi Gamma de paramètres λ et n .

Indication: procéder par récurrence sur n .

Loi de Cauchy Un variable aléatoire X a une loi de Cauchy si elle admet la densité

$$f_X(x) = \frac{1}{\pi(1+x^2)}.$$

Notons que la primitive de f_X peut être calculé explicitement:

$$\int \frac{1}{\pi(1+x^2)} = \frac{1}{\pi} \arctan(x) + c.$$

Ainsi on trouve

$$\int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)} = \frac{1}{\pi} [\arctan(\infty) - \arctan(-\infty)] = 1.$$

2.2.3 Fonctions d'une variable aléatoire

Une fonction $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ est dite mesurable si

$$\varphi^{-1}(A) \in \mathcal{B}(\mathbb{R}) \quad \text{pour tout } A \in \mathcal{B}(\mathbb{R}). \quad (2.4)$$

De manière équivalente, on peut demander que $\varphi^{-1}(A) \in \mathcal{B}(\mathbb{R})$ pour A dans une classe \mathcal{A} qui engendre $\mathcal{B}(\mathbb{R})$. Considérons par exemple $\mathcal{A} = \{(-\infty, a] : a \in \mathbb{R}\}$. Si on suppose de plus que φ est continue, alors $\varphi^{-1}(A)$ est fermé pour tout $A \in \mathcal{A}$, donc $\varphi^{-1}(A) \in \mathcal{B}(\mathbb{R})$. On conclut que toutes les fonctions continues sont mesurables.

Plus généralement, toutes les fonctions rencontrées en pratique sont mesurables.

Lemme 2.3. *Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable et X une variable aléatoire. Alors $\varphi(X)$ est aussi une variable aléatoire. De plus, si X est discrète, alors $\varphi(X)$ l'est aussi.*

Attention! Si X est une variable continue (par exemple à densité), $\varphi(X)$ n'est pas nécessairement continue.

Preuve: Soit $A \in \mathcal{B}(\mathbb{R})$. Alors

$$\{\omega \in \Omega : \varphi(X(\omega)) \in A\} = \{\omega \in \Omega : X(\omega) \in \varphi^{-1}(A)\} \in \mathcal{E},$$

car $\varphi^{-1}(A) \in \mathcal{B}(\mathbb{R})$ et X est une variable aléatoire

Supposons que X est discrète prenant les valeurs x_1, x_2, \dots avec probabilités positives. Posons $A = \{\varphi(x_i) : i > 0\}$. Alors, pour tout $y \in A$,

$$\mathbb{P}(\varphi(X) = y) = \sum_{\substack{i \geq 1 \text{ t.q.} \\ \varphi(x_i) = y}} \mathbb{P}(X = x_i).$$

De plus, $\mathbb{P}(\varphi(X) \in A) = \mathbb{P}(X \in \{x_1, x_2, \dots\}) = 1$. □

Proposition 2.4 (Changement de variables unidimensionnel). *Soit X une variable aléatoire admettant une densité f_X et $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction bijective et de classe \mathcal{C}^1 (donc mesurable). Alors la variable aléatoire $\varphi(X)$ admet la densité*

$$f_{\varphi(X)}(t) = \frac{1}{|\varphi'(\varphi^{-1}(t))|} f_X(\varphi^{-1}(t)).$$

Preuve: Supposons pour commencer que φ est croissante. Pour $t \in \mathbb{R}$ et $\epsilon > 0$ on a

$$\begin{aligned} \frac{1}{\epsilon} \mathbb{P}(\varphi(X) \in [t, t + \epsilon]) &= \frac{1}{\epsilon} \mathbb{P}(X \in [\varphi^{-1}(t), \varphi^{-1}(t + \epsilon)]) \\ &= \frac{1}{\epsilon} \int_{\varphi^{-1}(t)}^{\varphi^{-1}(t+\epsilon)} f_X(s) ds \xrightarrow{\epsilon \rightarrow 0} (\varphi^{-1})'(\varphi^{-1}(t)) f_X(\varphi^{-1}(t)) \\ &= \frac{1}{\varphi'(\varphi^{-1}(t))} f_X(\varphi^{-1}(t)). \end{aligned}$$

Si on suppose que φ est décroissante, les bornes de l'intégrale sont inversées et on trouve

$$\begin{aligned} \frac{1}{\epsilon} \mathbb{P}(\varphi(X) \in [t, t + \epsilon]) &= \frac{1}{\epsilon} \mathbb{P}(X \in [\varphi^{-1}(t + \epsilon), \varphi^{-1}(t)]) \\ &= \frac{1}{\epsilon} \int_{\varphi^{-1}(t+\epsilon)}^{\varphi^{-1}(t)} f_X(s) ds \xrightarrow{\epsilon \rightarrow 0} -(\varphi^{-1})'(\varphi^{-1}(t)) f_X(\varphi^{-1}(t)) \\ &= -\frac{1}{\varphi'(\varphi^{-1}(t))} f_X(\varphi^{-1}(t)). \end{aligned}$$

□

Une façon alternative de prouver cette proposition est de dériver en t la fonction $\mathbb{P}[\varphi(X) < t] = \mathbb{P}[X < \varphi^{-1}(t)]$ pour φ croissante.

Remarque 2.5. Souvent on utilise cette formule pour φ une fonction affine: $\varphi : x \mapsto ax + b$ avec $a \neq 0$. Alors on obtient:

$$f_{aX+b}(t) = \frac{1}{|a|} f_X\left(\frac{t-b}{a}\right).$$

Exercice 17.

Soient X une variable de loi normale standard et $\mu, \sigma \in \mathbb{R}$ avec $\sigma \neq 0$. Montrer que $\mu + \sigma X \sim \mathcal{N}(\mu, \sigma^2)$.

Exercice 18.

Soit X une variable aléatoire de loi exponentielle de paramètre $\lambda > 0$. Quelle est la loi de $\lfloor X \rfloor + 1$ (ou $\lfloor \cdot \rfloor$ désigne la partie entière)?

2.3 Indépendance des variables aléatoires

Souvent on décrit des expériences avec plusieurs quantités observées, donc plusieurs variables aléatoires. Si par exemple on lance deux dés et on note X et Y les résultats, on obtient deux variables aléatoires définies sur un même espace de probabilité. Comme pour les événements, on aimerai dire que ces variables sont indépendantes.

Fixons un espace de probabilité $(\Omega, \mathcal{E}, \mathbb{P})$; toutes les variables aléatoires dans cette partie sont définies implicitement sur cet espace.

Définition 2.6. Soit $(X_i)_{i \in I}$ une famille de variables aléatoires. On dit qu'elles sont indépendantes si, pour tout $J \subset I$ fini et toute collection $\{[a_j, b_j] : j \in J\}$ d'intervalles de \mathbb{R} , on a

$$\mathbb{P}\left(\bigcap_{j \in J} \{X_j \in [a_j, b_j]\}\right) = \prod_{j \in J} \mathbb{P}(X_j \in [a_j, b_j]). \quad (2.5)$$

Lemme 2.7. Soit $(X_i)_{i \in I}$ une famille de variables aléatoires. Elles sont indépendantes si et seulement si, pour toute famille $(A_i)_{i \in I}$ de boréliens de \mathbb{R} , les événements

$$\{X_i \in A_i\} : i \in I$$

sont indépendants.

Observons que la définition d'indépendance des variables aléatoires donne l'indépendance des événements $\{X_i \in A_i\} : i \in I$ si les A_i sont des intervalles. Ainsi, une implication du lemme est immédiate. On admet l'implication inverse.

Lemme 2.8. Soient X, Y deux variables aléatoires discrètes. Alors X et Y sont indépendantes (noté $X \perp\!\!\!\perp Y$); si et seulement si

$$\mathbb{P}(X = x \text{ et } Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \quad \text{pour tout } x, y \in \mathbb{R}. \quad (2.6)$$

Preuve: Si $X \perp\!\!\!\perp Y$, alors on peut appliquer (2.5) avec des intervalles $[x, x]$ et $[y, y]$ pour obtenir (2.6). Inversement, supposons (2.6). Soient A, B deux boréliens. Alors

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

Ainsi $X \perp\!\!\!\perp Y$. □

Lemme 2.9. Soient X, Y deux variables aléatoires et $f, g : \mathbb{R} \rightarrow \mathbb{R}$ des fonctions mesurables. Alors $f(X) \perp\!\!\!\perp g(Y)$.

Preuve: Soient A, B deux boréliens. Alors

$$\begin{aligned} \mathbb{P}[f(X) \in A \text{ et } g(Y) \in B] &= \mathbb{P}[X \in f^{-1}(A) \text{ et } Y \in g^{-1}(B)] \\ &= \mathbb{P}[X \in f^{-1}(A)] \mathbb{P}[Y \in g^{-1}(B)] = \mathbb{P}[f(X) \in A] \mathbb{P}[g(Y) \in B]. \end{aligned}$$

□

Définition 2.10. Une famille de variables aléatoires $(X_i)_{i \in I}$ est dite indépendante et identiquement distribuée (abrégé **i.i.d.**) si les variables X_i sont indépendantes et ont toutes la même loi.

Grace au théorème d'extension de Kolmogorov, pour toute loi μ , il existe un espace de probabilité $(\Omega, \mathcal{E}, \mathbb{P})$ et des variables aléatoires $(X_i)_{i \in \mathbb{N}}$ qui sont i.i.d. de loi μ .

Exercice 19.

Soient $(X_i)_{1 \leq i \leq n}$ des variables i.i.d. de Bernoulli de paramètre $p \in (0, 1)$. Quelle est la loi de $Y = \sum_{i=1}^n X_i$?

Complément: Processus de Poisson ponctuel On commence par deux lemmes essentiels.

Lemme 2.11. Soient X, Y deux variables de Poissons de paramètres λ et μ , respectivement. Supposons de plus que $X \perp\!\!\!\perp Y$. Alors $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$.

Preuve: Pour $n \geq 0$ on peut écrire

$$\begin{aligned} \mathbb{P}(X + Y = n) &= \mathbb{P}\left(\bigsqcup_{k=0}^n \{X = k, Y = n - k\}\right) \\ &= \sum_{k=0}^n \mathbb{P}(X = k \text{ et } Y = n - k) = \sum_{k=0}^n \mathbb{P}(X = k)\mathbb{P}(Y = n - k) \\ &= e^{-(\lambda+\mu)} \sum_{k=0}^n \frac{\lambda^k}{k!} \frac{\mu^{n-k}}{(n-k)!} = e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} \\ &= e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^n}{n!}. \end{aligned}$$

□

Lemme 2.12. Soient X une variable de Poissons de paramètres λ . De plus, soient $(Y_n)_{n \in \mathbb{N}}$ des variables i.i.d. de Bernoulli de paramètre $p \in (0, 1)$, indépendantes de X . Posons

$$\tilde{X} = \sum_{n=1}^X Y_n.$$

Alors \tilde{X} suit une loi de Poisson de paramètre $p\lambda$. De plus, $X - \tilde{X}$ suit une loi de Poisson de paramètre $(1 - p)\lambda$ et est indépendante de \tilde{X} .

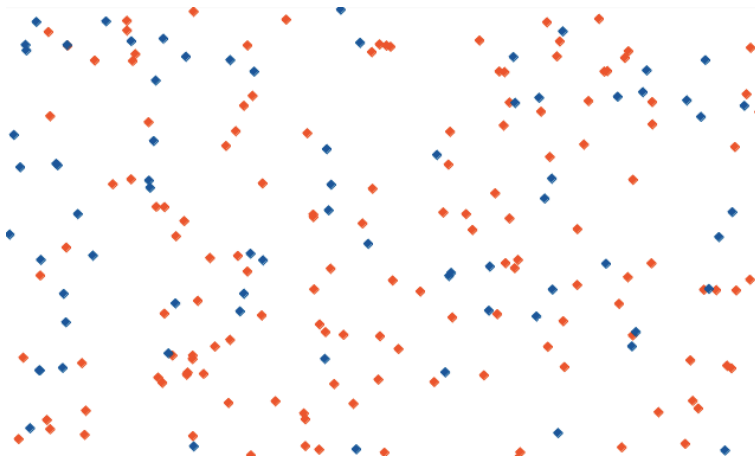


Figure 2.1: Deux processus de Poisson ponctuels dans le plan, un d'intensité 1 (en orange) et un autre d'intensité 1/2 (bleu). Ensemble ils forment un processus de Poisson ponctuel d'intensité 3/2.

Preuve: Pour $n, m \geq 0$ on a

$$\begin{aligned} \mathbb{P}(\tilde{X} = n \text{ et } X - \tilde{X} = m) &= \mathbb{P}\left(X = n + m \text{ et } \sum_{k=1}^{n+m} Y_k = n\right) = \mathbb{P}(X = n + m) \mathbb{P}\left(\sum_{k=1}^{n+m} Y_k = n\right) \\ &= e^{-\lambda} \frac{\lambda^{(n+m)}}{(n+m)!} \binom{n+m}{n} p^n (1-p)^m = e^{-\lambda} \frac{(p\lambda)^n ((1-p)\lambda)^m}{n! m!}. \end{aligned}$$

En sommant sur $m \geq 0$, on trouve

$$\begin{aligned} \mathbb{P}(\tilde{X} = n) &= \sum_{m \geq 0} \mathbb{P}(\tilde{X} = n \text{ et } X - \tilde{X} = m) \\ &= e^{-\lambda} \frac{(p\lambda)^n}{n!} \sum_{m \geq 0} \frac{((1-p)\lambda)^m}{m!} = e^{-\lambda} \frac{(p\lambda)^n}{n!} e^{(1-p)\lambda} = e^{-p\lambda} \frac{(p\lambda)^n}{n!}. \end{aligned}$$

Ainsi \tilde{X} est une variable de Poisson de paramètre $p\lambda$. De même, on peut montrer que $X - \tilde{X}$ suit une loi de Poisson de paramètre $(1-p)\lambda$. En fin, le calcul initial montre que

$$\mathbb{P}(\tilde{X} = n \text{ et } X - \tilde{X} = m) = \mathbb{P}(\tilde{X} = n) \mathbb{P}(X - \tilde{X} = m),$$

donc que $\tilde{X} \perp\!\!\!\perp X - \tilde{X}$. □

À l'aide de ces résultats, on peut définir un emplacement aléatoire de points dans le plan \mathbb{R}^2 , tel que pour chaque rectangle $R = [a, b] \times [c, d]$ le nombre de points N_R dans ce rectangle suit une loi de Poisson de paramètre $\lambda \text{Aire}(R) := \lambda(b-a)(d-c)$ et tels que les nombres de points dans des rectangles disjoints R_1, R_2, \dots sont indépendants.

Soit $(N_{k,\ell})_{k,\ell \in \mathbb{Z}}$ une famille de variables i.i.d. de Poisson de paramètre λ . Pour chaque carré $C_{k,\ell} = [k, k+1] \times [\ell, \ell+1]$, plaçons $N_{k,\ell}$ points dedans, uniformément. Cela définit un emplacement de points dans le plan, montrons qu'il a les propriétés désirées.

- Exercice 20.** (a) Pour un rectangle R contenu dans un carré $C_{k,\ell}$, montrer que le nombre N_R de points dans R suit une loi de Poisson de paramètre $\lambda \text{Aire}(R)$.
- (b) Pour un rectangle général R , en décomposant $R = \bigsqcup_{k,\ell} R \cap C_{k,\ell}$, montrer que N_R suit une loi de Poisson de paramètre $\lambda \text{Aire}(R)$.
- (c) Montrer que si R_1, \dots, R_n sont des rectangles disjoints contenus dans un même carré $C_{k,\ell}$ alors les variables N_{R_1}, \dots, N_{R_n} sont indépendantes.
- (d) En utilisant la même décomposition que au point (b), montrer le résultat du point (c) pour tous rectangles disjoints R_1, \dots, R_n .
- (e) Que obtient on si on superpose deux processus de Poissons indépendants de paramètres λ et μ ?

2.4 Loi jointe; vecteurs aléatoires

Définition 2.13. *Un vecteur aléatoire est une collection $\mathbf{X} = (X_1, \dots, X_n)$ de variables aléatoires définies sur un même espace de probabilité.*

Pour un vecteur aléatoire, la fonction $\mu_{\mathbf{X}} : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$ définie par

$$\mu_{\mathbf{X}}(A) = \mathbb{P}(\mathbf{X} \in A)$$

est bien définie et induit une probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ qu'on appelle la loi jointe de X_1, \dots, X_n , ou la loi du vecteur aléatoire \mathbf{X} .

On a vu précédemment que la loi d'une variable aléatoire est caractérisée par sa fonction de répartition. Pour un vecteur aléatoire on parle de la fonction de répartition conjointe.

Définition 2.14. *Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire. La fonction répartition conjointe de \mathbf{X} et la fonction $F_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$ définie par*

$$F_{\mathbf{X}}(t_1, \dots, t_n) = \mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n).$$

La fonction répartition conjointe a des propriétés similaires aux fonctions de répartition déjà étudiées.

Proposition 2.15. *Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire. Alors*

- (i) *La fonction $F_{\mathbf{X}}$ détermine la loi de \mathbf{X} .*
- (ii) *La fonction $F_{\mathbf{X}}$ est croissante et continue à droite en chaque coordonnée.*

(iii) On a

$$\lim_{t_1 \rightarrow -\infty, \dots, t_n \rightarrow -\infty} F(t_1, \dots, t_n) = 0 \quad \text{et} \quad \lim_{t_1 \rightarrow \infty, \dots, t_n \rightarrow \infty} F(t_1, \dots, t_n) = 1.$$

(iv) Pour tout $1 \leq i \leq n$, la fonction de répartition de $\tilde{\mathbf{X}} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ est

$$F_{\tilde{\mathbf{X}}}(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n) = \lim_{t_i \rightarrow \infty} F_{\mathbf{X}}(t_1, \dots, t_n).$$

En particulier $F_{X_i}(t_i) = \lim_{t_j \rightarrow \infty, j \neq i} F_{\mathbf{X}} = (t_1, \dots, t_n)$.

(v) Les variables X_1, \dots, X_n sont indépendantes si et seulement si

$$F_{\tilde{\mathbf{X}}}(t_1, \dots, t_n) = F_{X_1}(t_1) \dots F_{X_n}(t_n), \quad \forall t_1, \dots, t_n \in \mathbb{R}. \quad (2.7)$$

Preuve: Le point a) est similaire au théorème 1.14; sa preuve nécessite des concepts de la théorie de la mesure, on va donc l'admettre ici.

b) On va étudier uniquement la première coordonnée. Pour $t_2, \dots, t_n \in \mathbb{R}$, on a

$$F_{\mathbf{X}}(t, t_2, \dots, t_n) = \mathbb{P}(X_1 \leq t, X_2 \leq t_2, \dots, X_n \leq t_n).$$

L'événement dont on calcule la probabilité est croissant en t , il en est donc de même de $F(\cdot, t_2, \dots, t_n)$. Comme pour la fonction de répartition d'une variable aléatoire, on peut montrer que $F(\cdot, t_2, \dots, t_n)$ est continue à droite.

c) Les limites en question existent par monotonie de $F_{\mathbf{X}}$. Ainsi

$$\begin{aligned} \lim_{t_1 \rightarrow \infty, \dots, t_n \rightarrow \infty} F(t_1, \dots, t_n) &= \lim_{m \rightarrow \infty} F(m, \dots, m) \\ &= \lim_{m \rightarrow \infty} \mathbb{P}(X_1 \leq m, \dots, X_n \leq m) = \mathbb{P}(X_1 < \infty, \dots, X_n < \infty) = 1. \end{aligned}$$

Le même raisonnement s'applique pour prouver la limite en $-\infty$.

d) On a

$$\begin{aligned} F_{\tilde{\mathbf{X}}}(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n) &= \mathbb{P}(X_1 \leq t_1, \dots, X_{i-1} \leq t_{i-1}, X_{i+1} \leq t_{i+1}, \dots, X_n \leq t_n) \\ &= \mathbb{P}(X_1 \leq t_1, \dots, X_{i-1} \leq t_{i-1}, X_i < \infty, X_{i+1} \leq t_{i+1}, \dots, X_n \leq t_n) \\ &= \lim_{t_i \rightarrow \infty} \mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n). \end{aligned}$$

Le deuxième point suit en appliquant ce qu'on vient de prouver successivement.

e) Observons que (2.7) revient à

$$\mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n) = \mathbb{P}(X_1 \leq t_1) \dots \mathbb{P}(X_n \leq t_n), \quad \forall t_1, \dots, t_n \in \mathbb{R}.$$

Si les variables X_1, \dots, X_n sont indépendantes, cette égalité suit de la définition.

Inversement, en appliquant des intersections et unions de tels événements, on peut en

déduire que

$$\mathbb{P}(s_1 \leq X_1 \leq t_1, \dots, s_n \leq X_n \leq t_n) = \mathbb{P}(s_1 \leq X_1 \leq t_1) \dots \mathbb{P}(s_n \leq X_n \leq t_n),$$

pour tous $s_1, t_1, \dots, s_n, t_n \in \mathbb{R}$. □

Vecteurs aléatoires discrets Un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$ est dit discret si chaque composante X_i est une variable aléatoire discrète. De manière équivalente \mathbf{X} est discret s'il existe un ensemble $A \subset \mathbb{R}^n$ fini ou dénombrable tel que

$$\mathbb{P}(\mathbf{X} \in A) = 1.$$

Tout comme les variables aléatoires discrètes, les vecteurs aléatoires discrets (ou plus précisément leur lois) sont décrits par des fonctions de masse:

$$f_{\mathbf{X}}(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n).$$

Une telle fonction de masse prends des valeurs dans $[0, 1]$, est non-nulle que pour un nombre au plus dénombrable de n -uplets (x_1, \dots, x_n) (à savoir ceux dans A) et est telle que

$$\sum_{(x_1, \dots, x_n) \in \mathbb{R}^n} f_{\mathbf{X}}(x_1, \dots, x_n) = \sum_{(x_1, \dots, x_n) \in A} f_{\mathbf{X}}(x_1, \dots, x_n) = 1.$$

Ainsi, pour tout $B \subset \mathbb{R}^n$,

$$\mathbb{P}(\mathbf{X} \in B) = \sum_{(x_1, \dots, x_n) \in B} f_{\mathbf{X}}(x_1, \dots, x_n).$$

En particulier les fonctions de masses marginales sont données par

$$f_{X_i}(x_i) = \sum_{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in \mathbb{R}^{n-1}} f_{\mathbf{X}}(x_1, \dots, x_n).$$

Mentionnons en fin que les variables X_1, \dots, X_n sont indépendantes si et seulement si

$$f_{\mathbf{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n), \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

Exercice 21.

Soit (X, Y) un vecteur aléatoire discret. Montrer que $X + Y$ est une variable discrète et que sa fonction de masse est donnée par

$$f_{X+Y}(t) = \sum_{x \in \mathbb{R}} f_{(X,Y)}(x, t - x). \tag{2.8}$$

Vecteurs aléatoires à densité Un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$ est dit à densité s'il existe une fonction $f_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, \infty)$ telle que

$$F_{\mathbf{X}}(t_1, \dots, t_n) = \int_{x_1=-\infty}^{t_1} \cdots \int_{x_n=-\infty}^{t_n} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n,$$

pour tous $t_1, \dots, t_n \in \mathbb{R}$. Il s'en suit alors que, pour tout $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}(\mathbf{X} \in A) = \iint_{\mathbb{R}^n} \mathbf{1}_A(x_1, \dots, x_n) f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

La fonction $f_{\mathbf{X}}$ est appelée la densité jointe de (X_1, \dots, X_n) .

Lemme 2.16. *Si \mathbf{X} est à densité, il en est de même de ses composantes X_1, \dots, X_n . Leur densités respectives sont appelées les densités marginales de \mathbf{X} et sont données par*

$$f_{X_i}(t_i) = \iint_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(x_1, \dots, x_{i-1}, t_i, x_{i+1}, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n.$$

Preuve: On peut voir cela en observant que si on définit f_{X_i} comme au-dessus,

$$\begin{aligned} F_{X_i}(t_i) &= \iint_{\mathbb{R}^n} \mathbf{1}_{x_i \leq t_i} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_{x_i=-\infty}^{t_i} \iint_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n dx_i \\ &= \int_{x_i=-\infty}^{t_i} f_{X_i}(x_i) dx_i. \end{aligned}$$

□

Corollaire 2.17. *Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire à densité. Alors ses composantes sont indépendantes si et seulement si*

$$f_{\mathbf{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n) \quad \forall x_1, \dots, x_n \in \mathbb{R}.^{(ii)} \quad (2.9)$$

Observons que pour montrer (2.9) il suffit d'observer que $f_{\mathbf{X}}$ s'écrit comme un produit. En effet, si on peut écrire

$$f_{\mathbf{X}}(x_1, \dots, x_n) = g_1(x_1) \dots g_n(x_n) \quad \forall x_1, \dots, x_n \in \mathbb{R}$$

ou g_1, \dots, g_n sont des fonctions positives, on déduit que

$$\prod_{k=1}^n \int g_k(x_k) dx_k = \iint g_1(x_1) \dots g_n(x_n) dx_1 \dots dx_n = 1,$$

⁽ⁱⁱ⁾Par là on veut dire, si et seulement si $f_{X_1}(\cdot) \dots f_{X_n}(\cdot)$ est une densité pour \mathbf{X}

et donc que, pour tout k ,

$$f_{X_k}(t) = \frac{1}{\int g_k(x_k) dx_k} g_k(t), \quad \forall t \in \mathbb{R}.$$

Cela implique (2.9).

Preuve: Supposons que X_1, \dots, X_n sont indépendantes. Alors, pour $t_1, \dots, t_n \in \mathbb{R}$,

$$\begin{aligned} F_{\mathbf{X}}(t_1, \dots, t_n) &= \prod_{i=1}^n F_{X_i}(t_i) = \prod_{i=1}^n \int_{x_i=-\infty}^{t_i} f_{X_i}(x_i) dx_i \\ &= \int_{x_1=-\infty}^{t_1} \dots \int_{x_n=-\infty}^{t_n} f_{X_1}(x_1) \dots f_{X_n}(x_n) dx_1 \dots dx_n. \end{aligned}$$

Inversement, si (2.9) est vraie, alors le calcul précédent montre que $F_{\mathbf{X}}(t_1, \dots, t_n) = \prod_{i=1}^n F_{X_i}(t_i)$ pour tous $t_1, \dots, t_n \in \mathbb{R}$, donc que X_1, \dots, X_n sont indépendantes. \square

Une fonction $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est dite de classe \mathcal{C}^1 si sa différentielle (aussi appelée la Jacobienne)

$$D\Phi_{\mathbf{x}} := \left(\frac{\partial \Phi_i(\mathbf{x})}{\partial x_j} \right)_{1 \leq i, j \leq n} \in \mathcal{M}_n(\mathbb{R})$$

existe en tout point $\mathbf{x} = (x_1, \dots, x_n)$ de \mathbb{R}^n et si elle est continue en \mathbf{x} . On appelle le Jacobien le déterminant de cette matrice:

$$J_{\Phi}(\mathbf{x}) = \det D\Phi_{\mathbf{x}}$$

On rappelle le changement de variables utilisé pour le calcul des intégrales. Si $g : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction intégrable (au sens de Riemann), alors

$$\iint_{\mathbb{R}^n} g(\Phi(x_1, \dots, x_n)) |J_{\Phi}(x_1, \dots, x_n)| dx_1 \dots dx_n = \iint_{\mathbb{R}^n} g(y_1, \dots, y_n) dy_1 \dots dy_n. \quad (2.10)$$

On va admettre cette formule, et en déduire le changement de variables multidimensionnel pour les densités des vecteurs aléatoires.

Proposition 2.18 (Changement de variables multidimensionnel). *Soient \mathbf{X} un vecteur aléatoire à densité et $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction bijective, de classe \mathcal{C}^1 , avec $J_{\mathbf{x}}(\Phi) \neq 0$ pour tout $\mathbf{x} \in \mathbb{R}^n$. Alors $\Phi(\mathbf{X})$ est un vecteur aléatoire à densité, admettant la densité*

$$f_{\Phi(\mathbf{X})}(\mathbf{y}) = f_{\mathbf{X}}(\Phi^{-1}(\mathbf{y})) |J_{\Phi^{-1}}(\mathbf{y})| = f_{\mathbf{X}}(\Phi^{-1}(\mathbf{y})) \frac{1}{|J_{\Phi}(\Phi^{-1}(\mathbf{y}))|}.$$

Mentionnons qu'on peut appliquer ce même changement de variable ((2.10) ainsi que Prop. 2.18) aussi si Φ est une fonction bijective entre des ouverts U et V de \mathbb{R}^n et si $\mathbb{P}(\mathbf{X} \in U) = 1$.

Preuve: Pour $A \subset \mathbb{R}^n$ ouvert, appliquons (2.10) à la fonction $f_{\mathbf{X}}$ et la transformation $\Phi^{-1} : A \rightarrow \Phi^{-1}(A)$ (qui est en effet une bijection de classe \mathcal{C}^1). On obtient ainsi

$$\begin{aligned} \mathbb{P}(\Phi(\mathbf{X}) \in A) &= \mathbb{P}(\mathbf{X} \in \Phi^{-1}(A)) = \iint_{\Phi^{-1}(A)} f_{\mathbf{X}}(\mathbf{y}) dx_1 \dots dx_n \\ &= \iint_A f_{\mathbf{X}}(\Phi^{-1}(\mathbf{y})) |J_{\Phi^{-1}}(\Phi^{-1}(\mathbf{y}))| dx_1 \dots dx_n. \end{aligned}$$

Cette égalité se généralise aux boréliens $A \in \mathcal{B}(\mathbb{R}^n)$ par des outils standard de la théorie de la mesure, et on déduit que $f_{\mathbf{X}}(\Phi^{-1}(\cdot)) |J_{\Phi^{-1}}(\Phi^{-1}(\cdot))|$ agit comme une densité pour $\Phi(\mathbf{X})$. La dernière égalité suit du simple fait que $J_{\Phi^{-1}}(\mathbf{y}) = J_{\Phi}(\Phi^{-1}(\mathbf{y}))^{-1}$ pour tout $\mathbf{y} \in \mathbb{R}^n$. \square

Corollaire 2.19. *Soient X, Y deux variables aléatoires de densité jointe $f_{(X,Y)}$. Alors la variable $X + Y$ admet la densité*

$$f_{X+Y}(s) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, s-x) dx.$$

En particulier, si $X \perp\!\!\!\perp Y$,

$$f_{X+Y}(s) = \int_{-\infty}^{+\infty} f_X(x) f_Y(s-x) dx.$$

Preuve: Soit $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ défini par $\Phi(x, y) = (x, x + y)$. Cette fonction satisfait les hypothèses de la Proposition 2.18 et son Jacobien est constant égal à 1. Ainsi

$$f_{(X, X+Y)}(x, s) = f_{(X,Y)}(x, s-x), \quad \forall s, y \in \mathbb{R}.$$

Il s'en suit que

$$f_{X+Y}(s) = \int_{-\infty}^{+\infty} f_{(X, X+Y)}(x, s) dx = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, s-x) dx.$$

Si en plus $X \perp\!\!\!\perp Y$, alors $f_{(X,Y)}(x, s-x) = f_X(x) f_Y(s-x)$. \square

Exemple: Soit $\mu = (\mu_1, \dots, \mu_n) \in \mathcal{M}_{1,n}(\mathbb{R})$ un vecteur ligne et $\Sigma \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique définie positive ⁽ⁱⁱⁱ⁾. On dit que $\mathbf{Y} = (Y_1, \dots, Y_n)$ est un vecteur Gaussien de moyennes μ et matrice de covariance Σ s'il admet la densité

$$f_{\mathbf{Y}}(y_1, \dots, y_n) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp \left[-\frac{1}{2} (\mathbf{y} - \mu) \Sigma^{-1} (\mathbf{y} - \mu)^T \right],$$

⁽ⁱⁱⁱ⁾c.-à-d. telle que $U \Sigma U^T > 0$ pour tout vecteur ligne $U \in \mathcal{M}_{1,n}(\mathbb{R})$ non-nul. Si Σ a cette propriété, il en est de même de Σ^{-1} .

ou on écrit $\mathbf{y} = (y_1, \dots, y_n) \in \mathcal{M}_{1,n}$.

L'exemple le plus simple est celui de n gaussiennes standard indépendantes $\mathbf{X} = (X_1, \dots, X_n)$. Alors

$$\begin{aligned} f_{\mathbf{X}}(x_1, \dots, x_n) &= \prod_{k=1}^n f_{X_k}(x_k) = \frac{1}{\sqrt{(2\pi)^n}} \exp \left[-\frac{1}{2}(x_1^2 + \dots + x_n^2) \right] \\ &= \frac{1}{\sqrt{(2\pi)^n}} \exp \left(-\frac{1}{2} \mathbf{x} \mathbf{x}^T \right). \end{aligned}$$

Soient $\mathbf{Y} = (Y_1, \dots, Y_n)$ un vecteur gaussien général (avec matrice de covariance Σ et moyennes μ), $A \in \mathcal{M}_n(\mathbb{R})$ une matrice inversible et $B \in \mathcal{M}_{1,n}(\mathbb{R})$ un vecteur ligne. Posons alors $\mathbf{Z} = \mathbf{Y}A + B \in \mathcal{M}_{1,n}(\mathbb{R})$. Alors, en utilisant le changement de variable $\mathbf{y} \mapsto \mathbf{y}A + B$, on peut calculer la densité de \mathbf{Z} :

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= |\det(A)|^{-1} f_{\mathbf{Y}}[(\mathbf{z} - B)A^{-1}] \\ &= \frac{1}{\sqrt{(2\pi)^n \det(\Sigma) |\det(A)|}} \exp \left[-\frac{1}{2} (\mathbf{z} - B - \mu A) A^{-1} \Sigma^{-1} (A^{-1})^T (\mathbf{z} - B - \mu A)^T \right]. \end{aligned}$$

On reconnaît la densité d'un vecteur gaussien de moyenne $B + \mu A$ et matrice de covariance $A^T \Sigma A$ (cette dernière est encore symétrique et définie positive).

Sachant que Σ , comme toute matrice symétrique et définie positive, s'écrit $\Sigma = A^T A$ pour une certaine matrice inversible A , on déduit que $\mathbf{X}A + \mu$ admet la même densité que \mathbf{Y} , donc qu'ils sont de même loi^(iv). La conclusion de ce calcul est le lemme suivant.

Lemme 2.20. *Tout vecteur gaussien \mathbf{Y} peut s'écrire $\mathbf{Y} = \mathbf{X}A + \mu$ ou μ est le vecteur des moyennes de \mathbf{Y} , A est une matrice carrée inversible et \mathbf{X} est un vecteur de gaussiennes indépendantes.*

Exercice 22 (Important!).

Soit $\mathbf{Y} = (Y_1, \dots, Y_n)$ un vecteur gaussien comme dans l'exemple.

- Calculer la densité de Y_1 et déduire que c'est une gaussienne de moyenne et variance à déterminer.
- Montrer que pour tout $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, non-tous nuls,

$$\tilde{Y} = \lambda_1 Y_1 + \dots + \lambda_n Y_n$$

est une gaussienne de moyenne et variance à déterminer.

Indication: En supposant que $\lambda_1 \neq 0$, utiliser le changement de variable

$$(y_1, \dots, y_n) \mapsto (\lambda_1 y_1 + \dots + \lambda_n y_n, y_2, \dots, y_n),$$

^(iv)Mentionnons que cela prouve implicitement que la loi de \mathbf{Y} est bien définie, c.-à-d. que $\iint_{\mathbb{R}^n} f_{\mathbf{Y}}(y_1, \dots, y_n) dy_1 \dots dy_n = 1$.

pour déduire que $(\tilde{Y}, Y_2, \dots, Y_n)$ est un vecteur gaussien.

2.5 Espérance

L'espérance d'une variable aléatoire est la "valeur moyenne" que la variable prend. Ainsi, on peut intuitivement dire que l'espérance du résultat d'un lancé d'un dé est de 3.5 et celle d'une variable uniforme sur $[0, 1]$ est de $1/2$. Pour définir précisément l'espérance des variables aléatoires générales, on va commencer par celles discrètes.

2.5.1 Espérance des variables aléatoires discrètes

Définition 2.21. Soit X une variable aléatoire discrète prenant uniquement des valeurs positives, qu'on note x_1, x_2, \dots . Alors l'espérance de X est donnée par

$$\mathbb{E}(X) = \sum_{k \geq 1} x_k \mathbb{P}(X = x_k) = \sum_{x \geq 0} x \mathbb{P}(X = x).$$

Observons que l'espérance est définie comme une somme possiblement infinie de nombres positifs. Elle est ainsi toujours définie et peut prendre des valeurs entre 0 et $+\infty$, inclus.

Exemple: Une X à valeurs dans $\{1, 3, 9, 27, \dots\}$ avec $\mathbb{P}(X = 3^n) = \frac{3}{4}4^{-n}$ pour tout $n \geq 0$ admet une espérance

$$\mathbb{E}(X) = \frac{3}{4} \sum_{n \geq 0} \left(\frac{3}{4}\right)^n = \frac{3}{4} \cdot \frac{1}{1 - 3/4} = 3.$$

Par contre si on prend Y prenant les mêmes valeurs mais avec $\mathbb{P}(Y = 3^n) = 2^{-(n+1)}$ pour tout $n \geq 0$ on trouve

$$\mathbb{E}(X) = \frac{1}{2} \sum_{n \geq 0} \left(\frac{3}{2}\right)^n = \infty.$$

Observons toutefois que si X prend que un nombre fini de valeurs, alors son espérance est toujours finie.

Passons au cas des variables discrète à valeurs réelles générales.

Définition 2.22. Soit X une variable aléatoire discrète prenant les valeurs x_1, x_2, \dots . On dit que X est intégrable (ou qu'elle admet une espérance) si $\mathbb{E}(|X|) < \infty$. Alors l'espérance

de X est donnée par

$$\mathbb{E}(X) = \sum_{k \geq 1} \mathbb{P}(X = x_k) x_k = \mathbb{E}(\max\{X, 0\}) - \mathbb{E}(\max\{-X, 0\}).$$

Observons que $|X|$ est une variable aléatoire discrète positive, on peut donc parler de son espérance grâce à la définition 2.21. La condition $\mathbb{E}(|X|) < \infty$ peut alors se réécrire

$$\sum_{k \geq 1} \mathbb{P}(X = x_k) |x_k| < \infty.$$

La définition de $\mathbb{E}(X)$ a un sens que sous cette condition. La dernière égalité de la définition se démontre facilement.

Lemme 2.23. *Soient X, Y deux variables aléatoires discrètes et intégrables ou positives.*

(i) *Alors, pour tout $\lambda \in \mathbb{R}$,*

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

(ii) *Si $\mathbb{P}(X \geq 0) = 1$, alors $\mathbb{E}(X) \geq 0$. Si de plus $\mathbb{E}(X) = 0$, alors $\mathbb{P}(X = 0) = 1$.*

(iii) *Si $\mathbb{P}(X \leq Y) = 1$ avec probabilité 1, alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.*

(iv) *$|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.*

(v) *Si $\mathbb{P}(X = \lambda) = 1$, alors $\mathbb{E}(X) = \lambda$.*

Preuve: Commençons par la première propriété. On a

$$\begin{aligned} \mathbb{E}(\lambda X + Y) &= \sum_{z \in \mathbb{R}} z \mathbb{P}(\lambda X + Y = z) \\ &= \sum_{z \in \mathbb{R}} \sum_{x, y \in \mathbb{R}} z \mathbf{1}_{\lambda x + y = z} \mathbb{P}(X = x \text{ et } Y = y) \\ &= \sum_{x, y \in \mathbb{R}} (\lambda x + y) \mathbb{P}(X = x \text{ et } Y = y) \\ &= \sum_{x \in \mathbb{R}} \left(\lambda x \sum_y \mathbb{P}(X = x \text{ et } Y = y) \right) + \sum_{y \in \mathbb{R}} \left(y \sum_x \mathbb{P}(X = x \text{ et } Y = y) \right) \\ &= \lambda \sum_{x \in \mathbb{R}} x \mathbb{P}(X = x) + \sum_y y \mathbb{P}(Y = y) = \lambda \mathbb{E}(X) + \mathbb{E}(Y). \end{aligned}$$

La propriété (ii) suit directement de la définition.

Pour (iii) écrivons $Y = X + (Y - X)$ avec $\mathbb{P}(Y - X \geq 0) = 1$ grâce à l'hypothèse. Ainsi, par les points (i) et (ii), $\mathbb{E}(Y) = \mathbb{E}(X) + \mathbb{E}(Y - X) \geq \mathbb{E}(X)$.

Le point (iv) suit en observant que $\mathbb{P}(X \leq |X|) = \mathbb{P}(-X \leq |X|) = 1$, donc $\mathbb{E}(X) \leq \mathbb{E}(|X|)$ et $\mathbb{E}(X) \geq -\mathbb{E}(|X|)$

En fin (v) est une conséquence directe de la définition. \square

Remarquons que l'espérance d'une variable aléatoire X dépend uniquement de sa loi (à travers sa fonction de masse dans le cas présent). Ainsi, on peut parler par exemple de l'espérance d'une variable de Poisson de paramètre λ , sans mentionner l'univers Ω sur lequel cette variable est définie.

Exemples: espérance des lois usuelles Calculons les espérance des lois déjà rencontrées:

Loi de Bernoulli : Si X suit une loi de Bernoulli de paramètre p alors

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

Loi de binomiale : Si Y suit une loi de binomiale de paramètres n, p alors on peut écrire $Y = X_1 + \dots + X_n$ ou X_1, \dots, X_n sont i.i.d. de Bernoulli de paramètre p . Ainsi

$$\mathbb{E}(Y) = \sum_{k=1}^n \mathbb{E}(X_k) = np.$$

Loi géométrique : Soit X une variable géométrique de paramètre $p \in (0, 1)$. Posons, pour $z < 1/p$

$$G(z) = \sum_{k \geq 1} z^k \mathbb{P}(X = k) = \sum_{k \geq 1} z^k p(1 - p)^{k-1} = \frac{zp}{1 - z(1 - p)}.$$

On peut dériver sous le signe somme dans l'expression précédente, car la série est absolument convergente. Ainsi on trouve

$$G'(z) = \sum_{k \geq 1} k z^{k-1} \mathbb{P}(X = k) = \frac{p(1 - z(1 - p)) + zp(1 - p)}{(1 - z(1 - p))^2} = \frac{p}{(1 - z(1 - p))^2}.$$

Posons $z = 1$ pour trouver

$$\mathbb{E}(X) = \sum_{k \geq 1} k \mathbb{P}(X = k) = \frac{1}{p}.$$

Loi de Poisson : Soit X une variable de Poisson de paramètre $\lambda > 0$. Alors

$$\mathbb{E}(X) = \sum_{k \geq 0} k \mathbb{P}(X = k) = \sum_{k \geq 1} e^{-\lambda} k \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{\lambda \geq 1} \frac{\lambda^{k-1}}{(k-1)!} = \lambda.$$

Exercice 23.

Soit X une variable aléatoire géométrique de paramètre p et posons $Y = (-1)^X X$. Calculer l'espérance de Y .

Exercice 24 (Important!).

Soit X une variable aléatoire à valeurs dans \mathbb{N} . Montrer que

$$\mathbb{E}(X) = \sum_{n \geq 0} \mathbb{P}(X > n) = \sum_{n \geq 1} \mathbb{P}(X \geq n).$$

2.5.2 Espérance des variables aléatoires quelconques

Pour une variable aléatoire continue X , on ne peut pas répliquer la construction du cas direct, car il se peut que $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$.

Définition 2.24. Soit X une variable aléatoire positive (c.-à-d. avec $\mathbb{P}(X \geq 0) = 1$). Pour $n \geq 1$, posons

$$X_n = 2^{-n} \lfloor 2^n X \rfloor = \sum_{k=0}^{\infty} k 2^{-n} \mathbf{1}_{k 2^{-n} \leq X < (k+1) 2^{-n}}.$$

Alors on pose

$$\mathbb{E}(X) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n).$$

Les variables X_n sont des approximations discrètes de X . Le choix de X_n est tel que $X_n(\omega)$ est une suite croissante vers $X(\omega)$ pour tout ω . Ainsi $\mathbb{E}(X_n)$ est croissante, et la limite existe toujours dans $[0, +\infty]$.

Mentionnons aussi que, si X est discrète, alors les définitions 2.21 et 2.24 donne la même valeur pour $\mathbb{E}(X)$.

Définition 2.25. Soit X une variable aléatoire. On dit que X est intégrable (ou qu'elle admet une espérance) si $\mathbb{E}(|X|) < \infty$. Alors l'espérance de X est donnée par

$$\mathbb{E}(X) = \mathbb{E}(\max\{X, 0\}) - \mathbb{E}(\max\{-X, 0\}).$$

Vu que $|X| \geq \max\{X, 0\}$ et $|X| \geq \max\{-X, 0\}$, si on suppose que $\mathbb{E}(|X|) < \infty$, alors il en est de même de $\mathbb{E}(\max\{X, 0\})$ et $\mathbb{E}(\max\{-X, 0\})$. Ainsi $\mathbb{E}(X)$ est bien définie.

Remarque 2.26. Observons que toute variable aléatoire X bornée (c.-à-d. telle qu'il existe $c > 0$ avec $\mathbb{P}(|X| < c) = 1$) est intégrable. L'inverse n'est pas vrai, comme on a déjà vu dans les exemples des variables aléatoires discrètes de Poisson et géométriques.

Proposition 2.27. Soient X, Y deux variables aléatoires et intégrables ou positives.

- Alors, pour tout $\lambda \in \mathbb{R}$,

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

- Si $X \geq 0$, alors $\mathbb{E}(X) \geq 0$. Si de plus $\mathbb{E}(X) = 0$, alors $\mathbb{P}(X = 0) = 1$.

- Si $X \leq Y$ avec probabilité 1, alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.
- $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

Cette proposition est obtenue par passage à la limite des résultats du lemme 2.23.

Preuve: Traitons uniquement le cas X et Y positives et montrons que $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$. Les autres propriétés de montre de la même façon.

On rappelle les notations X_n, Y_n et $(X + Y)_n$ pour les approximations de X, Y et $X + Y$, respectivement. Alors on vérifie simplement que, pour tout $\omega \in \Omega$ et $n \geq 1$,

$$X_n + Y_n \leq (X + Y)_n \leq X_n + Y_n + 2^{n-1}$$

Cette inégalité se traduit pour l'espérance en

$$\mathbb{E}(X_n) + \mathbb{E}(Y_n) \leq \mathbb{E}[(X + Y)_n] \leq \mathbb{E}(X_n) + \mathbb{E}(Y_n) + 2^{n-1}.$$

En passant à la limite, on obtient l'égalité recherchée. □

2.5.3 Espérance des variables aléatoires à densité

Proposition 2.28. Soit X une variable à densité, d'une densité f_X . Alors

$$\mathbb{E}(|X|) = \int_{\mathbb{R}} |x| f_X(x) dx.$$

De plus, si cette quantité est finie, alors

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

Preuve: Supposons que X est positive; la preuve pour X de signe quelconque suit le même raisonnement. Posons

$$g_n(x) = 2^{-n} \lfloor x 2^n \rfloor f_X(x).$$

Cette fonction est telle que

$$\int_{k2^{-n}}^{(k+1)2^{-n}} g_n(x) dx = k 2^{-n} \int_{k2^{-n}}^{(k+1)2^{-n}} f_X(x) dx = k 2^{-n} \mathbb{P}(X \in [k 2^{-n}, (k+1) 2^{-n}]).$$

Ainsi

$$\begin{aligned} \mathbb{E}(X_n) &= \sum_{k=0}^{\infty} k 2^{-n} \mathbb{P}(X \in [k 2^{-n}, (k+1) 2^{-n}]) = \sum_{k=0}^{\infty} \int_{k 2^{-n}}^{(k+1) 2^{-n}} g_n(x) dx \\ &= \int_0^{\infty} g_n(x) dx \leq \int_0^{\infty} x f_X(x) dx. \end{aligned}$$

Il s'en suit que

$$\mathbb{E}(X) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n) \leq \int_0^{\infty} x f_X(x) dx.$$

Il nous reste à montrer l'inégalité inverse. Observons que

$$\int_0^{\infty} x f(x) dx - \mathbb{E}(X_n) = \int_0^{\infty} (x - 2^{-n} \lfloor x 2^n \rfloor) f_X(x) dx \leq \int_0^{\infty} 2^{-n} f_X(x) dx \leq 2^{-n} \quad (2.11)$$

En prenant $n \rightarrow \infty$, on trouve donc $\int_0^{\infty} x f(x) dx - \mathbb{E}(X_n) \leq 0$, ce qui achève la preuve. \square

Exemples: espérance des lois usuelles

Loi uniforme Si U est uniforme sur un intervalle $[a, b]$, alors sa densité est $\frac{1}{b-a} \mathbf{1}_{[a,b]}$ donc son espérance peut être calculé comme

$$\mathbb{E}(U) = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Loi exponentielle. Si X a une loi exponentielle de paramètre $\lambda > 0$, alors, en intégrant par parties, on trouve

$$\mathbb{E}(X) = \int_0^{\infty} x \lambda e^{-\lambda x} = [-x e^{-\lambda x}]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} = \frac{1}{\lambda}.$$

Loi normale. Soit N suit une loi normale de moyenne 0 et variance σ^2 . Comme N prend des valeurs positives et négative, on doit commencer par s'assurer que $\mathbb{E}(|N|) < \infty$. On a

$$E(|N|) = \int_0^{\infty} 2x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx < \infty.$$

Ainsi N est intégrable et

$$E(N) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx = \int_0^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx + \int_{-\infty}^0 x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx = 0,$$

car les deux dernières intégrales sont égales en valeur absolue mais de signes opposés. Ce n'est pas surprenant que $\mathbb{E}(N) = 0$, car N est distribuée symétriquement par rapport à 0.

En fin, si \tilde{N} suit une loi normale de moyenne μ et variance σ^2 , alors $\tilde{N} - \mu$ suit une loi normale de moyenne 0 et variance σ^2 . Ainsi \tilde{N} est intégrable et

$$\mathbb{E}(\tilde{N}) = \mathbb{E}(N) + \mu = \mu.$$

Loi gamma. Si X suit loi Gamma de paramètres $\lambda > 0$ et $t > 0$, alors

$$\mathbb{E}(X) = \frac{1}{\Gamma(t)} \int_0^{\infty} x \lambda^t x^{t-1} e^{-\lambda x} dx = \frac{1}{\Gamma(t)} [x^t \lambda^t e^{-\lambda x}]_0^{\infty} + \frac{1}{\Gamma(t)} \int_0^{\infty} \frac{t}{\lambda} \lambda^t x^{t-1} e^{-\lambda x} dx = \frac{t}{\lambda}.$$

Loi de Cauchy. En fin, soit X une variable de loi de Cauchy. On pourrait croire que $\mathbb{E}(X) = 0$ par symétrie, mais on doit pour commencer s'assurer que X est intégrable. On a

$$\mathbb{E}(|X|) = \int_0^\infty 2x \frac{1}{\pi(1+x^2)} = \infty.$$

Ainsi X n'est pas intégrable et donc n'admet pas d'espérance.

Exercice 25.

Soit X une variable aléatoire à valeurs dans $[0, +\infty)$. Montrer que

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > x) dx = \int_0^\infty (1 - F_X(x)) dx.$$

Commencer par le cas où X admet une densité f_X .

2.5.4 Espérance d'une fonction d'une variable aléatoire

Commençons par une observation qui, malgré sa simplicité, peut être très utile. Soit X une variable aléatoire et $A \subset \mathbb{R}$ un borélien. Alors on peut écrire:

$$\mathbb{P}(X \in A) = \mathbb{E}[\mathbf{1}_A(X)].$$

L'avantage de l'écriture de droite est qu'elle nous permet d'utiliser la linéarité de l'espérance. De plus, si $E \subset \Omega$ est un événement, alors $\mathbf{1}_E$ est une variable aléatoire et

$$\mathbb{P}(E) = \mathbb{E}(\mathbf{1}_E).$$

Exemple: Si A, B sont deux événements d'un espace de probabilité. Alors

$$\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{E}(\mathbf{1}_A) + \mathbb{E}(\mathbf{1}_B) = \mathbb{E}(\mathbf{1}_A + \mathbf{1}_B) = \mathbb{E}(\mathbf{1}_{A \cup B} + \mathbf{1}_{A \cap B}) = \mathbb{P}(A \cup B) + \mathbb{P}(A \cap B).$$

On déduit ainsi la formule $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Lemme 2.29. Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire et $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction mesurable. Alors, si ϕ est à valeurs positives ou si $\phi(X_1, \dots, X_n)$ est intégrable,

- si \mathbf{X} est discret

$$\mathbb{E}[\phi(\mathbf{X})] = \sum_{x_1, \dots, x_n} \phi(x_1, \dots, x_n) \mathbb{P}(X_1 = x_1, \dots, X_n = x_n),$$

ou la somme est sur toutes les valeurs $(x_1, \dots, x_n) \in \mathbb{R}^n$ que \mathbf{X} prend avec probabilité positive;

- si \mathbf{X} admet une densité $f_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, \infty)$, alors

$$\mathbb{E}[\phi(\mathbf{X})] = \iint_{\mathbb{R}^n} \phi(x_1, \dots, x_n) f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Remarque 2.30. Cette formule s'applique évidemment aussi à une variable aléatoire X :

$$\mathbb{E}[\phi(X)] = \begin{cases} \sum_{x \in \mathbb{R}} \phi(x) \mathbb{P}(X = x), & \text{si } X \text{ est discrète,} \\ \int_{-\infty}^{+\infty} \phi(x) f_X(x) dx & \text{si } X \text{ admet la densité } f_X. \end{cases}$$

Preuve: Supposons que \mathbf{X} est un vecteur discret. Alors

$$\begin{aligned} \mathbb{E}[\phi(\mathbf{X})] &= \sum_y y \mathbb{P}[\phi(\mathbf{X}) = y] \\ &= \sum_y \sum_{x_1, \dots, x_n} \mathbf{1}_{\{\phi(x_1, \dots, x_n) = y\}} \mathbb{P}[\mathbf{X} = (x_1, \dots, x_n)] \\ &= \sum_{x_1, \dots, x_n} \mathbb{P}[\mathbf{X} = (x_1, \dots, x_n)] \left(\sum_y \mathbf{1}_{\{\phi(x_1, \dots, x_n) = y\}} \right) \\ &= \sum_{x_1, \dots, x_n} \phi(x_1, \dots, x_n) \mathbb{P}[\mathbf{X} = (x_1, \dots, x_n)]. \end{aligned}$$

Pour les variables à densité, il faut utiliser l'approximation par des variables discrètes. Supposons que ϕ est à valeurs dans $[0, +\infty)$ pour simplifier les calculs. Pour $k \geq 1$, soit $\phi_k(x_1, \dots, x_n) = 2^{-k} \lfloor \phi(x_1, \dots, x_n) 2^k \rfloor$. Alors

$$\begin{aligned} \mathbb{E}[\phi_k(\mathbf{X})] &= \sum_{j \geq 0} j 2^{-k} \mathbb{P}[\phi_k(\mathbf{X}) = j 2^{-k}] \\ &= \sum_{j \geq 0} j 2^{-k} \iint_{\phi^{-1}\left(\left[\frac{j}{2^k}, \frac{j+1}{2^k}\right]\right)} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \iint_{\mathbb{R}^n} \sum_{j \geq 0} j 2^{-k} \mathbf{1}_{\phi^{-1}\left(\left[\frac{j}{2^k}, \frac{j+1}{2^k}\right]\right)}(x_1, \dots, x_n) f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \iint_{\mathbb{R}^n} \phi_k(x_1, \dots, x_n) f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

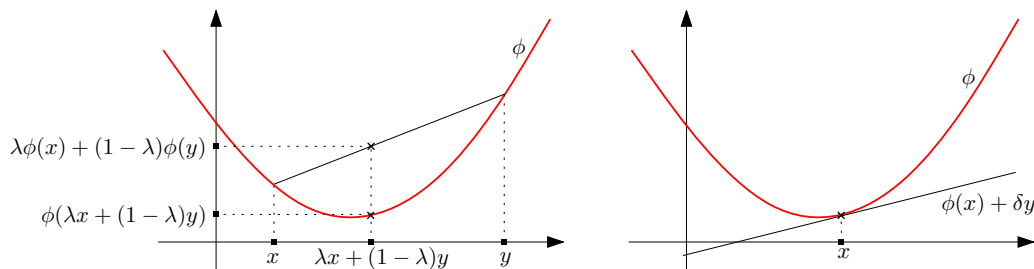
En fin, comme dans (2.11), quand $k \rightarrow \infty$, le terme de gauche tend vers $\mathbb{E}[\phi_k(\mathbf{X})]$ alors que le terme de droite tend vers $\iint_{\mathbb{R}^n} \phi(x_1, \dots, x_n) f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n$. \square

Convexité & Jensen Un outil qui nous permet d'obtenir des nombreuses inégalités utiles est la convexité. Une fonction $\phi : \mathbb{R} \rightarrow \mathbb{R}$ est dite *convexe* si, pour tout $x, y \in \mathbb{R}$ et $\lambda \in (0, 1)$,

$$\phi(\lambda x + (1 - \lambda)y) \leq \lambda \phi(x) + (1 - \lambda)\phi(y). \quad (2.12)$$

Elle est dite *strictement convexe* si l'inégalité au-dessus est stricte (pour tout $x \neq y$).

Rappelons qu'une fonction dérivable est convexe si et seulement si sa dérivée est croissante. Si la fonction est \mathcal{C}^2 , on peut vérifier sa convexité en montrant que sa deuxième dérivée est positive.


 Figure 2.2: Deux façons de voir qu'une fonction ϕ est convexe.

Cela implique qu'une fonction convexe dérivable reste au-dessus de ses tangentes. Ceci peut être généralisé comme suit.

Lemme 2.31. *Une fonction $\phi : \mathbb{R} \rightarrow \mathbb{R}$ est convexe si et seulement si pour tout $x \in \mathbb{R}$ il existe $\delta \in \mathbb{R}$ tel que, pour tout $y \in \mathbb{R}$,*

$$\phi(x + y) \geq \phi(x) + \delta y.$$

De plus ϕ est strictement convexe si et seulement s'il existe δ tel que $\phi(x + y) > \phi(x) + \delta y$ pour tout x et $y \neq 0$.

On laisse la preuve de ce lemme en exercice. Mentionnons uniquement qu'on peut choisir $\delta = \liminf_{y \searrow x} \frac{\phi(y) - \phi(x)}{y - x}$.

Mentionnons que les deux cotés de (2.12) peuvent être perçues comme des moyennes pondérées. Alors (2.12) se lit: " ϕ d'une moyenne est plus petit que la moyenne de ϕ ". Le théorème suivant en est une généralisation naturelle.

Théorème 2.32 (Inégalité de Jensen). *Soit X une variable aléatoire et $\phi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe. Si les deux espérances en-dessous existent, alors*

$$\mathbb{E}[\phi(X)] \geq \phi[\mathbb{E}(X)].$$

De plus, si ϕ est strictement convexe, alors on a égalité au-dessus si et seulement si $\mathbb{P}(X = \mathbb{E}(X)) = 1$.

Preuve: En appliquant le lemme 2.31 avec $x = \mathbb{E}(X)$ et $y = X - \mathbb{E}(X)$, on trouve

$$\mathbb{E}[\phi(X)] = \mathbb{E}\left[\phi\left(\underbrace{X - \mathbb{E}(X)}_y + \underbrace{\mathbb{E}(X)}_x\right)\right] \geq \mathbb{E}[\delta(X - \mathbb{E}(X)) + \phi(\mathbb{E}(X))] = \phi(\mathbb{E}(X)).$$

Si on suppose de plus que $\mathbb{E}[\phi(X)] = \phi[\mathbb{E}(X)]$ et que ϕ est strictement convexe, alors on a égalité dans la ligne précédente, donc

$$\mathbb{P}[\phi(X) - \phi(\mathbb{E}(X)) = \delta(X - \mathbb{E}(X))] = 1.$$

Par la deuxième partie du lemme 2.31, cela implique $\mathbb{P}(X = \mathbb{E}(X)) = 1$. \square

Loi et espérance

Proposition 2.33. *Soient X, Y deux variables aléatoires. Si*

$$\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$$

pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ bornée et continue, alors X et Y ont la même loi.

Preuve: Observons pour commencer que, si f est bornée, $f(X)$ et $f(Y)$ sont des variables aléatoires bornées. Ainsi elles sont intégrables.

Pour $n \in \mathbb{N}$ et $t \in \mathbb{R}$, soit $f_{n,t} : \mathbb{R} \rightarrow \mathbb{R}$ définie par

$$f_{n,t}(x) = \begin{cases} 1 & \text{si } x < t, \\ 1 - 2^n(x - t) & \text{si } t \leq x \leq t + 2^{-n}, \\ 0 & \text{si } x > t + 2^{-n}. \end{cases}$$

Ces fonctions sont continues, donc $\mathbb{E}[f_{n,t}(X)] = \mathbb{E}[f_{n,t}(Y)]$ pour tout $n \geq 1$ et $t \in \mathbb{R}$. D'autre part $\mathbf{1}_{(-\infty, t]} \leq f_{n,t} \leq \mathbf{1}_{(-\infty, t+2^{-n}]}$, d'où

$$F_X(t) \leq \mathbb{E}[f_{n,t}(X)] \leq F_X(t + 2^{-n}).$$

En appliquant cela à Y aussi et en prenant $n \rightarrow \infty$, on trouve $F_X(t) = F_Y(t)$ pour tout t . \square

Indépendance et espérance

Proposition 2.34. *Soient X, Y deux variables aléatoires. Alors X et Y sont indépendantes si et seulement si*

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)] \tag{2.13}$$

pour tous fonctions $f, g : \mathbb{R} \rightarrow \mathbb{R}$ bornées et continues (ou simplement bornées mesurables).

Preuve: Supposons pour commencer que X et Y sont indépendantes. Si de plus X et Y sont discrètes, alors $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ suit directement du Lemme 2.29. Pour des variables aléatoires indépendantes quelconques, on utilise des approximations comme dans la définition de l'espérance pour arriver à (2.13).

Supposons maintenant qu'on a (2.13) pour deux variables X et Y . Avec les mêmes notations que dans le preuve de la proposition 2.33 on trouve, pour $t, s \in \mathbb{R}$

$$\mathbb{E}[f_{n,s}(X)f_{n,t}(Y)] = \mathbb{E}[f_{n,s}(X)]\mathbb{E}[f_{n,t}(Y)].$$

En prenant $n \rightarrow \infty$, on trouve

$$\mathbb{P}(X \leq s \text{ et } Y \leq t) = \mathbb{P}(X \leq s)\mathbb{P}(Y \leq t),$$

ce qui implique que X et Y sont indépendantes par (2.7). \square

2.6 Variance et moments

Définition 2.35. On dit qu'une variable aléatoire X admet un moment d'ordre $p \geq 1$ si

$$\mathbb{E}(|X|^p) < \infty.$$

On appelle alors $\mathbb{E}(X^p)$ le moment d'ordre p de X . L'ensemble des variables aléatoires (sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$) admettant un moment d'ordre p est noté L^p .

On pose aussi

$$L^\infty = \{X : \text{variable aléatoire t.q. il existe } C > 0 \text{ avec } \mathbb{P}(|X| \leq C) = 1\}.$$

Les ensembles L^p ($p \geq 1$) sont des objets très étudiés en analyse, comme en probabilité. De leurs nombreuses propriétés on en mentionne ici une seule.

Lemme 2.36. Si $X, Y \in L^p$ pour un $p \geq 1$, alors $X + Y \in L^p$.

Si $X \in L^p$ pour un $p \geq 1$, alors X admet des moments de tout ordre $0 < \ell \leq p$.

En d'autres termes, ce lemme nous dit que chaque L^p ($p \geq 1$) est un espace vectoriel et que ces espaces sont décroissants en p .

Preuve: Supposons que $X, Y \in L^p$ pour un certain $p \geq 1$. Alors, comme la fonction $x \mapsto |x|^p$ est convexe,

$$\frac{1}{2^p} \mathbb{E}(|X + Y|^p) = \mathbb{E}\left(\left|\frac{X + Y}{2}\right|^p\right) \leq \mathbb{E}\left[\frac{1}{2}(|X|^p + |Y|^p)\right] = \frac{1}{2}[\mathbb{E}(|X|^p) + \mathbb{E}(|Y|^p)] < \infty.$$

Soit $0 < \ell < p$. Alors, comme la fonction $x \mapsto x^{p/\ell}$ est convexe

$$\mathbb{E}[|X|^\ell]^{p/\ell} \leq \mathbb{E}[(|X|^\ell)^{p/\ell}] = \mathbb{E}(|X|^p) < \infty.$$

En fin, on observe facilement que L^∞ est aussi un espace vectoriel et que $L^\infty \subset L^p$ pour tout $p < \infty$. \square

Définition 2.37. Soit X une variable aléatoire admettant un moment d'ordre 2. Alors la variance de X est donnée par

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

L'écart-type de X est alors $\sigma(X) = \sqrt{\text{Var}(X)}$.

L'écart-type (et la variance) d'une variable X mesure à quel point X dévie (typiquement) autour de sa moyenne. L'avantage d'utiliser l'écart-type est qu'il "est mesuré dans les mêmes unités que X et $\mathbb{E}(X)$ ". Pour mesurer la déviation typique de X autour de $\mathbb{E}(X)$, il est peut être plus naturel d'utiliser une mesure du type

$$\mathbb{E}[|X - \mathbb{E}(X)|].$$

L'avantage de l'écart-type est qu'il s'exprime à l'aide de fonctions polynomiales simples. Cela résulte dans des formules plus simple pour le manipuler. De plus, est deux quantités sont comparables (voir exercice 28).

Définition 2.38. Soient X, Y deux variables aléatoires admettant un moment d'ordre 2. Alors la covariance de X et Y est donnée par

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

La corrélation de X et Y est alors^(v)

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Les deux expression données pour la variance et la covariance sont égales du à un calcul direct:

$$\begin{aligned} \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] &= \mathbb{E}(XY) - \mathbb{E}[Y\mathbb{E}(X)] - \mathbb{E}[X\mathbb{E}(Y)] + \mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \end{aligned}$$

Le fait que la covariance est une quantité bien définie suit du théorème suivant.

Théorème 2.39 (Inégalité de Cauchy-Schwarz). Soient X, Y deux variables aléatoires admettant un moment d'ordre 2. Alors

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

De plus, on a égalité si et seulement s'il existe $\lambda \neq 0$ tel que $X = \lambda Y$ ou $Y = \lambda X$.

^(v)La corrélation est définie uniquement pour deux variables de variance strictement positive chacune.

Observons que, si $X, Y \in L^2$, en appliquant l'inégalité de Cauchy-Schwarz à $|X|$ et $|Y|$, on observe que $\mathbb{E}(|XY|) < \infty$, donc que XY est intégrable. Ainsi, les deux termes de l'inégalité sont toujours bien définis.

Preuve: Quitte à permuter les variables et à éliminer le cas trivial $\mathbb{P}(X = 0) = \mathbb{P}(Y = 0) = 1$, on peut supposer $\mathbb{E}(X^2) \neq 0$. Observons alors que, pour $\lambda \in \mathbb{R}$,

$$0 \leq \mathbb{E}[(\lambda X + Y)^2] = \lambda^2 \mathbb{E}(X^2) + 2\lambda \mathbb{E}(XY) + \mathbb{E}(Y^2).$$

On reconnaît à droite un polynôme d'ordre 2 en λ . Du fait qu'il n'admet pas de racines réelles, on déduit que son discriminant est négatif:

$$4\mathbb{E}(XY)^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0.$$

De plus, le discriminant est nul si et seulement si le polynôme admet une racine réelle, ce qui revient à $\mathbb{P}(Y = \lambda X) = 1$ pour un certain $\lambda \in \mathbb{R}$. \square

Corollaire 2.40. *Soient $X, Y \in L^2$ deux variables aléatoires de variance strictement positive. Alors*

$$\rho(X, Y) \in [-1, 1]. \tag{2.14}$$

De plus $\rho = 1$, (respectivement -1) si et seulement s'il existe $\lambda > 0$ (respectivement $\lambda < 0$) telle que $\mathbb{P}(X = \lambda Y) = 1$.

Preuve: La preuve suit directement de l'inégalité de Cauchy-Schwarz. En effet

$$\begin{aligned} \text{Cov}(X, Y)^2 &= \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]^2 \\ &\leq \mathbb{E}[(X - \mathbb{E}(X))^2] \mathbb{E}[(Y - \mathbb{E}(Y))^2] \quad \text{par Cauchy-Schwarz} \\ &= \text{Var}(X)\text{Var}(Y). \end{aligned}$$

Après division par $\text{Var}(X)\text{Var}(Y)$ et application de la racine carrée, on obtient le résultat. (Pour cette étape, il faut avoir $\text{Var}(X)\text{Var}(Y) > 0$.)

Si on a égalité dans (2.14), il faut avoir égalité dans l'inégalité de Cauchy-Schwarz au-dessus. Cela implique que $X = \lambda Y$ pour un certain $\lambda \in \mathbb{R}$. Alors, il suffit d'observer que

$$\rho(\lambda Y, Y) = \frac{\text{Cov}(\lambda Y, Y)}{\sqrt{\text{Var}(\lambda Y)\text{Var}(Y)}} = \frac{\lambda \text{Var}(Y)}{\sqrt{\lambda^2 \text{Var}(Y)}} = \frac{\lambda}{|\lambda|}.$$

\square

Proposition 2.41. *Soient $X, Y, Z \in L^2$ des variables aléatoires. Alors*

- (i) $\text{Cov}(X, X) = \text{Var}(X)$;
- (ii) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- (iii) $\text{Cov}(\lambda X + \mu Y, Z) = \lambda \text{Cov}(X, Z) + \mu \text{Cov}(Y, Z)$;
- (iv) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.
- (v) $\text{Var}(X) = 0$ si et seulement si $\mathbb{P}(X = \mathbb{E}(X)) = 1$. Dans le cas échéant, $\text{Cov}(X, Y) = 0$ pour toute variable Y admettant un moment d'ordre 2.

Preuve: Les deux premiers points suivent directement de la définition. Le troisième suit de la linéarité de l'espérance.

Pour le quatrième, on écrit

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y - \mathbb{E}(X + Y))^2] = \mathbb{E}[(X - \mathbb{E}(X) + Y - \mathbb{E}(Y))^2] \\ &= \mathbb{E}[(X - \mathbb{E}(X))^2] + \mathbb{E}[(Y - \mathbb{E}(Y))^2] + \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y). \end{aligned}$$

Pour le dernier point, observons que $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = 0$ si et seulement si $\mathbb{P}[(X - \mathbb{E}(X))^2 = 0] = 1$, donc si et seulement si $\mathbb{P}(X = \mathbb{E}(X)) = 1$. Dans ce cas, $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}[0(Y - \mathbb{E}(Y))] = 0$. \square

Pour un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$ dont chaque composante admet un moment d'ordre 2, on définit sa matrice de covariance $\Sigma \in \mathcal{M}_n(\mathbb{R})$ par

$$\Sigma_{ij} = \text{Cov}(X_i, X_j).$$

Alors, pour tout $\Lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$,

$$\text{Var}(\lambda_1 X_1 + \dots + \lambda_n X_n) = \Lambda \Sigma \Lambda^T.$$

En particulier on déduit que la matrice symétrique Σ est positive; même définie positive si les variables X_1, \dots, X_n forment une famille libre.

Indépendance et non-corrélation

Proposition 2.42. *Si X, Y sont des variables aléatoires indépendantes admettant un moment d'ordre 2, alors*

$$\text{Cov}(X, Y) = 0.$$

L'inverse n'est pas vrai!

Preuve: Si $X, Y \in L^2$ sont indépendantes, alors

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(X - \mathbb{E}(X)) \mathbb{E}(Y - \mathbb{E}(Y)) = 0.$$

Pour le second point, voir l'exercice 26 □

Exercice 26.

Soient X, Y deux variables indépendantes prenant les valeurs ± 1 avec probabilité $1/2$. Posons

$$Z = \begin{cases} Y & \text{si } X = -1 \\ 2Y & \text{si } X = 1 \end{cases}$$

Montrer que $\rho(X, Z) = 0$ mais que X et Z ne sont pas indépendantes.

En vous inspirant de cet exemple, construire un couple de variables gaussiennes de covariance nulle mais qui ne sont pas indépendantes. Pourquoi cette construction ne contredit pas l'exercice suivant?

Exercice 27 (Important!).

Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur Gaussien de moyennes μ et matrice de covariance Σ . Calculer la covariance de X_i et X_j pour $1 \leq i, j \leq n$.

Conclure que les variables X_1, \dots, X_n sont indépendantes si et seulement si $\text{Cov}(X_i, X_j) = 0$ pour tous $i \neq j$.

Exercice 28.

Soit X une variable aléatoire admettant un moment d'ordre 2. En utilisant l'inégalité de Cauchy-Schwarz, montrer que

$$\mathbb{E}[|X - \mathbb{E}(X)|] \leq \sigma(X).$$

Supposons qu'il existe une constante $C > 0$ telle que $\mathbb{P}(|X - \mathbb{E}(X)| < C) = 1$. Montrer alors qu'on peut trouver $C' > 0$ telle que

$$\mathbb{E}[|X - \mathbb{E}(X)|] \geq C'\sigma(X).$$

Toutefois, montrer que si X n'est pas borné, on ne peut pas obtenir une inégalité comme au-dessus. Pour cela, calculer $\sigma(X)$ et $\mathbb{E}[|X - \mathbb{E}(X)|]$ pour une variable aléatoire X avec

$$\mathbb{P}(X = x) = \begin{cases} \frac{1}{2n} & \text{pour } x = n \text{ et } x = -n, \\ 1 - \frac{1}{n} & \text{pour } x = 0. \end{cases}$$

2.6.1 Inégalité de Markov, Tchebychev



Théorème 2.43. *Soit X une variable aléatoire positive. Alors, pour tout $\lambda > 0$,*

$$\mathbb{P}(X \geq \lambda) \leq \frac{1}{\lambda} \mathbb{E}(X).$$

Preuve: Soit $\lambda > 0$. Alors

$$\mathbb{P}(X \geq \lambda) = \frac{1}{\lambda} \mathbb{E}(\lambda \mathbf{1}_{X \geq \lambda}) \leq \frac{1}{\lambda} \mathbb{E}(X \mathbf{1}_{X \geq \lambda}) \leq \frac{1}{\lambda} \mathbb{E}(X).$$

□

Corollaire 2.44 (Inégalité de Tchebychev). *Soit X une variable aléatoire admettant un moment d'ordre 2. Alors, pour tout $\lambda > 0$,*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{1}{\lambda^2} \text{Var}(X).$$

Preuve: Il suffit d'appliquer l'inégalité de Markov à la variable aléatoire positive $(X - \mathbb{E}(X))^2$, dont l'espérance est $\text{Var}(X)$. □

2.7 Espérance conditionnelle; loi conditionnelle

2.7.1 Conditionnement par rapport à un événement

Soit X une variable aléatoire et A un événement avec $\mathbb{P}(A) > 0$. On peut alors parler de la loi de la variable X , conditionnellement à A . Elle est donnée par

$$\mu_{X|A}(B) = \mathbb{P}(X \in B | A), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Il s'agit simplement de la loi de X sous la probabilité $\mathbb{P}(\cdot|A)$. Ainsi, on peut parler de l'espérance de la variable X (pourvu que celle-ci existe) sous la probabilité $\mathbb{P}(\cdot|A)$. Elle est donnée par

$$\mathbb{E}(X|A) = \frac{\mathbb{E}(X \mathbf{1}_A)}{\mathbb{P}(A)}.$$

Evidemment, si X admet une espérance non-conditionnelle, alors elle en admet une conditionnelle aussi, car $\mathbb{E}(|X| | A) \leq \frac{\mathbb{E}(|X|)}{\mathbb{P}(A)}$.

Exemple: On peut prendre une personne ω en Suisse au hasard (uniformément) et noter X son salaire. Il est alors naturel de parler du salaire moyen en suisse (donnée par $\mathbb{E}(X)$), mais aussi du salaire moyen pour les hommes (donné par $\mathbb{E}(X | \omega \text{ homme})$).

On peut également parler de la distribution des salaires (la loi de X) et de la distribution des salaires pour les hommes (la loi de X conditionnellement à l'événement $\{\omega \text{ est un homme}\}$).

Attention! Il est essentiel dans cette partie d'avoir $\mathbb{P}(A) > 0$, pour pouvoir diviser par cette quantité!

Une variante de la loi de probabilité totale s'applique pour le calcul des espérances. Si X est une variable aléatoire intégrable et $(A_k)_{k \geq 1}$ est une famille (finie ou dénombrable) d'événements formant une partition de Ω , alors

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{k \geq 1} X \mathbf{1}_{A_k}\right) = \sum_{k \geq 1} \mathbb{E}(X|A_k) \mathbb{P}(A_k). \quad (2.15)$$

2.7.2 Conditionnement par rapport à une variable discrète

Continuons l'exemple du salaire en Suisse comme variable aléatoire X . Supposons maintenant qu'on veut parler de salaire moyen par tranche d'âge. On peut alors associer à chaque âge y la valeur $\phi(y) = \mathbb{E}(X | \omega \text{ a âge } y)$.

Si on considère que ω est une personne prise au hasard, on peut créer une variable aléatoire Y qui désigne son âge. La variable aléatoire $\phi(Y)$ se lit alors "le salaire moyen dans la tranche d'âge de la personne sélectionnée."

En probabilité, on appelle cette *variable aléatoire* l'espérance de X sachant Y , et on la note $\mathbb{E}(X|Y)$.

Définition 2.45. Soit $X \in L^1$ une variable aléatoire quelconque et Y une variable aléatoire discrète. Alors l'esperance conditionnelle de X par rapport à Y est la variable aléatoire $\mathbb{E}(X|Y) = \phi(Y)$ ou

$$\phi(y) = \mathbb{E}(X|Y = y)$$

pour tout y tel que $\mathbb{P}(Y = y) > 0$.

Proposition 2.46. Soit $X \in L^1$ une variable aléatoire quelconque et Y une variable aléatoire discrète. Alors, pour toute fonction borélienne bornée $f : \mathbb{R} \rightarrow \mathbb{R}$,

$$\mathbb{E}[f(Y)X] = \mathbb{E}[f(Y)\mathbb{E}(X|Y)]. \quad (2.16)$$

Si on applique 2.16 avec f la fonction constante égale à 1, on déduit

$$\mathbb{E}[\mathbb{E}(X|Y)] = \mathbb{E}[X].$$

Preuve: Notons $\mathcal{Y} = \{y \in \mathbb{R} : \mathbb{P}(Y = y) > 0\}$ et ϕ une fonction telle que $\mathbb{E}(X|Y) = \phi(Y)$.

Alors on peut écrire $f(z) = \sum_{y \in \mathcal{Y}} f(y) \mathbf{1}_{z=y}$ pour tout $z \in \mathcal{Y}$. Ainsi, par la linéarité de l'intégrale (et le théorème de convergence dominée),

$$\begin{aligned} \mathbb{E}[f(Y)X] &= \sum_{y \in \mathcal{Y}} \mathbb{E}[f(y) \mathbf{1}_{Y=y} X] = \sum_{y \in \mathcal{Y}} f(y) \mathbb{E}[\mathbf{1}_{Y=y} X] \\ &= \sum_{y \in \mathcal{Y}} f(y) \phi(y) \mathbb{P}(Y = y) = \mathbb{E}[f(Y)\phi(Y)] = \mathbb{E}[f(Y)\mathbb{E}(X|Y)]. \end{aligned}$$

□

2.7.3 Espérance conditionnelle par rapport à une variable générale

Si Y n'est pas une variable aléatoire discrète, on ne peut plus appliquer la stratégie de la partie précédente pour définir $\mathbb{E}(X|Y)$, car $\{Y = y\}$ peut avoir probabilité 0 pour tout $y \in \mathbb{R}$. Toutefois, l'espérance conditionnelle peut être définie.

La solution est de définir l'espérance conditionnelle à travers sa propriété 2.16.

Théorème 2.47. *Soit X une variable aléatoire dans L^1 et Y une variable aléatoire quelconque. Alors il existe une unique^(vi) variable aléatoire notée $E(X|Y)$ telle que*

- elle s'écrit $E(X|Y) = \phi(Y)$ pour une fonction mesurable ϕ ;
- pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ mesurable bornée, on a

$$\mathbb{E}[f(Y)X] = \mathbb{E}[f(Y)\mathbb{E}(X|Y)]. \quad (2.17)$$

La variable $E(X|Y)$ est l'espérance de X conditionnellement à Y .

Ce théorème va être vu dans le cours de théorie de la mesure; on l'admet pour l'instant.

Une variable qui s'écrit $f(Y)$ pour une certaine fonction mesurable f est dite *mesurable par rapport à Y* .

Si on pose $\mathcal{F}_Y = \{Y^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$ alors \mathcal{F}_Y est une σ -algèbre sur Ω . Il s'agit de la plus petite σ -algèbre sur Ω telle que $Y : (\Omega, \mathcal{F}_Y) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ soit mesurable. Les événements de \mathcal{F}_Y s'identifie facilement quand Y est discrète. Dans ce cas, $A \in \mathcal{F}_Y$ si et seulement si, pour tout $\omega, \tilde{\omega} \in A$ avec $Y(\omega) = Y(\tilde{\omega})$

$$\omega \in A \Leftrightarrow \tilde{\omega} \in A.$$

Une variable aléatoire $Z : \Omega \rightarrow \mathbb{R}$ s'écrit $Z = f(Y)$ pour une fonction mesurable $f : \mathbb{R} \rightarrow \mathbb{R}$ si et seulement si $Z : (\Omega, \mathcal{F}_Y) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable. Dans le cas échéant on dit que Z est \mathcal{F}_Y -mesurable. Ainsi, l'espérance conditionnelle $\mathbb{E}(X|Y)$ est l'unique variable aléatoire \mathcal{F}_Y -mesurable telle que, pour toute variable aléatoire bornée \mathcal{F}_Y -mesurable Z ,

$$\mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}(X|Y)].$$

^(vi)La variable est unique au sens ou, si Z et \tilde{Z} satisfont toutes les deux les propriétés de $\mathbb{E}(X|Y)$, alors $\mathbb{P}(Z = \tilde{Z}) = 1$.

On identifie deux cas extrêmes de dépendance entre X et Y . Il se peut que X soit indépendant de Y (écrit $X \perp\!\!\!\perp Y$) ou que X soit complètement déterminé par Y (ce qui veut dire que $X = g(Y)$ pour une fonction mesurable g , ou alors que X est \mathcal{F}_Y -mesurable). Dans ces deux cas, l'espérance conditionnelle est facilement déterminée.

- Si $X \perp\!\!\!\perp Y$ alors $\mathbb{E}(X|Y) = \mathbb{E}(X)$ avec probabilité 1. On dit que $\mathbb{E}(X|Y)$ est une variable déterministe.
- Si X s'écrit $X = g(Y)$ pour une fonction mesurable g , alors $\mathbb{E}(X|Y) = X$.

Pour prouver ces deux affirmations, il suffit d'observer que l'expression donnée pour la variable $\mathbb{E}(X|Y)$ satisfait les conditions du théorème 2.47.

2.7.4 Espérance conditionnelle par rapport à une variable à densité

Si Y n'est pas une variable aléatoire discrète, on ne peut plus appliquer la stratégie de la partie précédente pour définir $\mathbb{E}(X|Y)$, car $\{Y = y\}$ peut avoir probabilité 0 pour tout $y \in \mathbb{R}$. Toutefois, l'espérance conditionnelle peut être définie.

La solution est de définir l'espérance conditionnelle à travers sa propriété 2.16.

Théorème 2.48. *Soit X, Y des variables aléatoires admettant une densité jointe continue $f_{X,Y} : \mathbb{R} \rightarrow [0, +\infty)$ avec $X \in L^1$. Alors, l'espérance conditionnelle de X par rapport à Y s'écrit $\mathbb{E}(X|Y) = \phi(Y)$ ou*

$$\phi(y) = \frac{\int_{-\infty}^{\infty} x f_{X,Y}(x, y) dx}{\int_{-\infty}^{\infty} f_{X,Y}(x, y) dx}. \quad (2.18)$$

Plus généralement, pour $y \in \mathbb{R}$ on définit la densité de X sachant que $Y = y$ par ^(vii)

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{\int_{-\infty}^{\infty} f_{X,Y}(t, y) dt}. \quad (2.19)$$

Alors, pour chaque $y \in \mathbb{R}$, $f_{X|Y=y}$ est une densité; on appelle la loi avec cette densité la loi de X sachant $Y = y$. L'espérance conditionnelle définie dans (2.18) est l'espérance sous cette loi.

Pour y fixé et $dy > 0$ petit, on peut approximer $\mathbb{P}(Y \in [y, y + dy])$ par $f_Y(y)dy = dy \int_{-\infty}^{\infty} f_{X,Y}(t, y) dt$. De plus, pour tout x fixé, on peut approximer $\int_y^{y+dy} f_{X,Y}(x, s) ds$ par $f_{X,Y}(x, y) dy$. Ainsi,

$$\mathbb{P}(X < u | Y \in [y, y + dy]) = \frac{\int_{-\infty}^u dx \int_y^{y+dy} f_{X,Y}(x, s) ds}{\mathbb{P}(Y \in [y, y + dy])} \sim \int_{-\infty}^u \frac{f_{X,Y}(x, y)}{\int_{-\infty}^{\infty} f_{X,Y}(t, y) dt} dx.$$

^(vii) On se limite au cas où $\int_{-\infty}^{\infty} f_{X,Y}(t, y) dt > 0$.

Ce-ci montre que $\frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(t,y)dt}$ joue bien le rôle de la densité de X conditionné à $\{Y = y\}$. Cette heuristique peut être rendue rigoureuse quand $f_{X,Y}$ est continue.

Remarque 2.49. Les formules du théorème précédent se généralise à des situations plus complexes. Si $\mathbf{X} = (X_1, \dots, X_n)$ est un vecteur aléatoire de densité jointe $f_{\mathbf{X}}$ et $\Psi : \mathbb{R}^k \rightarrow \mathbb{R}$ est une fonction mesurable, alors $\mathbb{E}[\Psi(X_1, \dots, X_k) | X_{k+1}, \dots, X_n] = \phi(X_{k+1}, \dots, X_n)$ ou

$$\phi(x_{k+1}, \dots, x_n) = \frac{\int_{\mathbb{R}^k} \Psi(x_1, \dots, x_k) f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_k}{\int_{\mathbb{R}^k} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_k}.$$

Preuve: Commençons par prouver (2.18). On doit montrer que, si on définit ϕ comme dans (2.18), alors, pour toute fonction mesurable bornée g ,

$$\mathbb{E}[Xg(Y)] = \mathbb{E}[g(Y)\mathbb{E}(X|Y)] = \int_{-\infty}^{\infty} g(y)\phi(y)f_Y(y)dy.$$

Pour cela, calculons le terme de droite comme expliqué dans le Lemme 2.29:

$$\begin{aligned} \mathbb{E}[Xg(Y)] &= \iint_{\mathbb{R}^2} xg(y)f_{X,Y}(x,y)dx dy \\ &= \int_{y=-\infty}^{\infty} \left(dyg(y) \int_{x=-\infty}^{\infty} x f_{X,Y}(x,y)dx \right) \\ &= \int_{y=-\infty}^{\infty} \left(dyg(y)\phi(y) \int_{x=-\infty}^{\infty} f_{X,Y}(x,y)dx \right) = \int_{y=-\infty}^{\infty} g(y)\phi(y)f_Y(y) dy, \end{aligned}$$

ce qui montre que ϕ est bien la densité de $\mathbb{E}(X|Y)$.

On observe directement que la fonction $f_{X|Y=y}$ définie dans (2.19) est positive et s'intègre à 1 pour tout $y \in \mathbb{R}$. De plus, l'espérance d'une variable avec cette densité est

$$\int_{-\infty}^{\infty} x f_{X|Y=y} dx = \int_{-\infty}^{\infty} x \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(t,y)dt} dx = \frac{\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dx}{\int_{-\infty}^{\infty} f_{X,Y}(t,y)dt} = \phi(y).$$

□

Chapter 3

Fonctions génératrices et applications

3.1 Définitions et propriétés

Définition 3.1. Soit X une variable aléatoire à valeurs dans \mathbb{R} . La fonction génératrice des moments de X est donnée par

$$M_X(t) = \mathbb{E}(e^{tX}) \in [0, \infty] \quad \forall t \in \mathbb{R}.$$

Remarque 3.2. Si X est une variable à valeurs dans \mathbb{N} , on définit aussi la *fonction génératrice des probabilités* par

$$G_X(t) = \mathbb{E}(t^X) = \sum_{n \geq 0} t^n \mathbb{P}(X = n).$$

Il s'agit d'une série entière avec un rayon de convergence plus grand ou égale à 1. Les deux sont reliées par la relation simple $G_X(e^t) = M_X(t)$.

Pour une variable quelconque X à valeurs dans \mathbb{R} , on définit également la *fonction caractéristique* de X par

$$\Phi_X(t) = \mathbb{E}(e^{itX}).$$

La fonction caractéristique est l'espérance d'une variable aléatoire à valeurs dans les nombre complexes de module 1. Ainsi Φ_X prend valeurs dans $\{z \in \mathbb{C} : |z| \leq 1\}$.

Proposition 3.3. Soit X un variable aléatoire.

- (i) La fonction $t \mapsto M_X(t)$ est finie sur un intervalle I de \mathbb{R} avec $0 \in I$. De plus $M_X(0) = 1$.
- (ii) $\log M_X(\cdot)$ est une fonction convexe sur I .

(iii) Supposons que I contient un voisinage de 0. Alors X admet des moments de tout ordre, M_X est C^∞ sur l'intérieur de I et

$$\mathbb{E}(X^k) = M_X^{(k)}(0). \quad (3.1)$$

(iv) Soient X et Y deux variables aléatoires telles que

$$M_X(t) = M_Y(t) < \infty \quad \text{pour tout } t \text{ dans un voisinage de 0.}$$

Alors X et Y ont la même loi.

Preuve: Commençons par le point (ii). Soient $s < t$ et $\lambda \in [0, 1]$. On va traiter en détail le cas $\lambda = 1/2$. Par l'inégalité de Cauchy Swartz pour les variables $e^{sX/2}$ et $e^{tX/2}$ on a

$$\log \mathbb{E}[e^{(\lambda s + (1-\lambda)t)X}] = \log \mathbb{E}[e^{sX/2} e^{tX/2}] \leq \log \mathbb{E}[e^{sX}]^{\frac{1}{2}} \mathbb{E}[e^{tX}]^{\frac{1}{2}} = \frac{1}{2} (\log \mathbb{E}[e^{sX}] + \log \mathbb{E}[e^{tX}]).$$

Ainsi on obtient (2.12) pour $\lambda = 1/2$. Pour obtenir l'inégalité pour tout autre $\lambda \in (0, 1)$ on utilise une généralisation de l'inégalité de Cauchy-Schwarz appelée inégalité de Hölder. Celle ci affirme que, pour U et V des variables aléatoires positives et $p, q > 1$ avec $\frac{1}{p} + \frac{1}{q} = 1$, on a

$$\mathbb{E}(UV) \leq \mathbb{E}(U^p)^{\frac{1}{p}} \mathbb{E}(V^q)^{\frac{1}{q}}.$$

En appliquant cela à $p = 1/\lambda$, $q = 1/(1-\lambda)$, $U = e^{sX}$ et $V = e^{tX}$ on obtient (2.12).

Passons au point (i). La convexité de $\log M_X$ montre que, si $M_X(s) < \infty$ et $M_X(t) < \infty$ pour certains $s < t$, alors $M_X(r) < \infty$ pour tout $r \in [s, t]$. Ainsi M_X est finie sur un intervalle. Le fait que $M_X(0) = 1$ suit directement de la définition de M_X .

(iii) Supposons que $M_X(t) < \infty$ pour tout $|t| \leq \delta$ pour un certain $\delta > 0$. Alors $\mathbb{E}[e^{\delta X}] < \infty$ et $\mathbb{E}[e^{-\delta X}] < \infty$, donc

$$\mathbb{E}[e^{\delta|X|}] < \infty. \quad (3.2)$$

Observons que, pour tout $k \in \mathbb{N}$, il existe une constante $C = C(k) > 0$ telle que, pour tout $x \in \mathbb{R}$, $|x|^k \leq C e^{\delta|x|}$. Ainsi, on déduit de (3.2) que X^k est intégrable.

En fin, (3.2) nous permet de dériver sous le signe \mathbb{E} pour $|t| < \delta$ (on admet cela dans ce cours). Ainsi

$$M_X'(t) = \mathbb{E}[X e^{tX}] \quad \text{et plus généralement} \quad M_X^{(k)}(t) = \mathbb{E}[X^k e^{tX}] \quad \text{pour } k \geq 0.$$

En posant $t = 0$, on trouve $M_X^{(k)}(0) = \mathbb{E}[X^k]$.

(iv) Pour simplicité, on va se limiter au cas où X et Y sont à support compact, c.-à-d. sont telles qu'il existe un certain $M > 0$ tel que $\mathbb{P}(X \in [-M, M]) = \mathbb{P}(Y \in [-M, M]) = 1$. Si on suppose que $M_X(t) = M_Y(t)$ pour tout t dans un voisinage de 0, on déduit par le point (iii) que $\mathbb{E}(X^k) = \mathbb{E}(Y^k)$ pour tout $k \geq 0$. Par linéarité on trouve que, pour tout

polynôme $P \in \mathbb{R}[X]$,

$$\mathbb{E}[P(X)] = \mathbb{E}[P(Y)]. \quad (3.3)$$

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue. Alors, pour tout $\epsilon > 0$ il existe un polynôme P tel que $\sup\{|f(x) - P(x)| : x \in [-M, M]\} < \epsilon$ ⁽ⁱ⁾. Il s'en suit que

$$|\mathbb{E}[f(X)] - \mathbb{E}[P(X)]| \leq \mathbb{E}[|f(X) - P(X)|] \leq \epsilon.$$

En utilisant (3.3), on trouve $|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| < 2\epsilon$. Comme $\epsilon > 0$ est arbitraire, on a $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$. De la Proposition 2.33 on déduit que X et Y ont la même loi. \square

Le nom “fonction génératrice des moments” s’explique par (3.1). Pour une suite $(a_n)_{n \geq 0}$, la fonction génératrice de cette suite est la série entière

$$z \mapsto \sum_{n \geq 0} a_n z^n.$$

De (3.1), en écrivant le développement en série Taylor de M_X , on trouve

$$M_X(t) = \sum_{k \geq 0} t^k \frac{M_X^{(k)}(0)}{k!} = \sum_{k \geq 0} t^k \frac{1}{k!} \mathbb{E}(X^k).$$

Proposition 3.4. *Soient X, Y deux variables aléatoires indépendantes. Alors*

$$M_{X+Y}(t) = M_X(t)M_Y(t), \quad \forall t \in \mathbb{R}.$$

Preuve: Soit $t \in \mathbb{R}$. Alors,

$$M_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX} e^{tY}] = \mathbb{E}[e^{tX}] \mathbb{E}[e^{tY}] = M_X(t)M_Y(t).$$

\square

Remarque 3.5. Des résultats similaires existent pour les fonctions génératrices des probabilités et les fonctions caractéristiques.

Calculs des fonctions génératrices des moments pour les lois usuelles. Dans la suite, si on ne mentionne pas le domaine pour t (et s), c’est car la formule est vraie pour tout $t \in \mathbb{R}$ (respectivement, que le rayon de convergence de G est ∞).

Loi de Bernoulli: Soit X une variable aléatoire de loi de Bernoulli de paramètre $p \in [0, 1]$.

⁽ⁱ⁾Ceci est la conclusion du théorème de Stone-Weierstrass

Alors

$$M_X(t) = (1-p)e^0 + pe^t = 1-p + pe^t \quad \text{et} \quad G_X(s) = 1-p + sp.$$

Loi binomiale: Soit Y une variable binomiale de paramètres $p \in [0, 1]$ et $n \geq 1$. Alors $Y = X_1 + \dots + X_n$ ou X_1, \dots, X_n sont i.i.d. de Bernoulli de paramètre p . Ainsi

$$M_Y(t) = M_X(t)^n = (1-p + pe^t)^n \quad \text{et} \quad G_Y(s) = (1-p + sp)^n.$$

Loi géométrique: Soit X une variable géométrique de paramètre $p \in [0, 1]$. Alors

$$M_X(t) = \sum_{k \geq 1} p(1-p)^{k-1} e^{kt} = \frac{pe^t}{1-(1-p)e^t} \quad \text{et} \quad G_X(s) = \frac{ps}{1-(1-p)s},$$

pour $t \leq -\log(1-p)$ et $|s| < 1/(1-p)$.

Loi de Poisson: Soit X une variable de Poisson de paramètre $\lambda > 0$. Alors

$$M_X(t) = e^{-\lambda} \sum_{k \geq 0} e^{kt} \frac{\lambda^k}{k!} = \exp[\lambda(e^t - 1)] \quad \text{et} \quad G_X(s) = \exp[\lambda(s - 1)].$$

Loi uniforme: Soient $a < b$ et U de loi uniforme sur l'intervalle $[a, b]$. Alors

$$M_U(t) = \frac{1}{b-a} \int_a^b e^{xt} dx = \frac{e^{bt} - e^{at}}{t(b-a)}.$$

Loi exponentielle: Soit X une variable aléatoire de loi exponentielle de paramètre $\lambda > 0$. Alors

$$M_X(t) = \int_0^\infty \lambda e^{(t-\lambda)x} dx = \frac{\lambda}{\lambda-t}, \quad \forall t < \lambda.$$

Loi normale: Soit N une variable aléatoire de loi normale de moyenne $\mu \in \mathbb{R}$ et variance σ^2 . Alors

$$\begin{aligned} M_N(t) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^\infty \exp\left(tx - \frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^\infty \exp\left(-\frac{(x-\mu+\sigma^2 t)^2}{2\sigma^2} + t\mu + \frac{t^2\sigma^2}{2}\right) dx = \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right). \end{aligned}$$

Fonction génératrice des moments pour des vecteurs aléatoires

Définition 3.6. Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire. La fonction génératrice des moments de \mathbf{X} est la fonction $M_{\mathbf{X}} : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par

$$M_{\mathbf{X}}(t_1, \dots, t_n) = \mathbb{E}[e^{t_1 X_1 + \dots + t_n X_n}] = \mathbb{E}[e^{\langle \mathbf{t}, \mathbf{X} \rangle}].$$

Les fonctions génératrices des variables individuelles peuvent être déduites de $M_{\mathbf{X}}$:

$$M_{\mathbf{X}}(0, \dots, 0, t_i, 0, \dots, 0) = M_{X_i}(t_i).$$

De plus, comme dans (3.1), on peut obtenir une formule pour les moments (joints) de \mathbf{X} . En effet, pour $k_1, \dots, k_n \geq 0$,

$$\mathbb{E}[X_1^{k_1} \dots X_n^{k_n}] = \frac{\partial^{k_1 + \dots + k_n} M_{\mathbf{X}}}{\partial t_1^{k_1} \dots \partial t_n^{k_n}}(0, \dots, 0).$$

Proposition 3.7. *Soient \mathbf{X} et \mathbf{Y} deux vecteurs aléatoires. Supposons que*

$$M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{Y}}(\mathbf{t}) < \infty \quad \text{pour tout } \mathbf{t} \in \mathbb{R}^n \text{ dans un voisinage de } 0.$$

Alors \mathbf{X} et \mathbf{Y} ont la même loi.

La preuve de cette proposition est très similaire à celle de la proposition 3.3.

3.2 Application: vecteurs gaussiens

Proposition 3.8. *Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire avec la propriété que pour tout $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, $\lambda_1 X_1 + \dots + \lambda_n X_n$ suit une loi normale. Alors \mathbf{X} est un vecteur gaussien.*

Remarque 3.9. On rappelle que, si on suppose que \mathbf{X} est un vecteur gaussien, alors toute combinaison linéaire de X_1, \dots, X_n est une gaussienne (voir exercice 22).

Ainsi, le fait que la loi de toute combinaison linéaire des composantes d'un vecteur aléatoire est gaussienne est une définition équivalente des vecteurs gaussiens.

Preuve: Soit \mathbf{X} un vecteur avec la propriété que pour tout $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, $\lambda_1 X_1 + \dots + \lambda_n X_n$ suit une loi normale. Posons $\mu = (\mu_1, \dots, \mu_n)$ le vecteur des espérances de \mathbf{X} : $\mu_i = \mathbb{E}(X_i)$. De plus, soit Σ la matrice de covariance de \mathbf{X} :

$$\Sigma_{ij} = \text{Cov}(X_i, X_j) \quad \forall 1 \leq i, j \leq n. \quad (3.4)$$

Soit $\mathbf{Y} = (Y_1, \dots, Y_n)$ un vecteur gaussien de moyennes μ et matrice de covariance Σ .

Pour $\mathbf{t} = (t_1, \dots, t_n) \in \mathbb{R}^n$, les variables $t_1 X_1 + \dots + t_n X_n$ et $t_1 Y_1 + \dots + t_n Y_n$ sont les deux des variables gaussiennes de moyenne $t_1 \mu_1 + \dots + t_n \mu_n = \mathbf{t}\mu$ et variance $\mathbf{t}\Sigma\mathbf{t}^T$. Ainsi

$$M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{Y}}(\mathbf{t}) = \exp\left(\mathbf{t}\mu + \frac{1}{2}\mathbf{t}\Sigma\mathbf{t}^T\right) < \infty.$$

On conclut par la proposition 3.7 que \mathbf{X} et \mathbf{Y} ont la même loi, donc que \mathbf{X} est un vecteur gaussien. \square

3.3 Application: processus de branchement

Dans cette partie on va parler uniquement de variables aléatoires à valeurs dans \mathbb{N} (donc discrètes). Il va être plus commode d'utiliser la fonction génératrice des probabilités.

Soit μ une loi de probabilité sur \mathbb{N} . On rappelle qu'on peut voir μ comme une suite $(p_n)_{n \in \mathbb{N}}$ avec $p_n \in [0, 1]$ et $\sum_{n \geq 0} p_n = 1$. Le lien entre μ et $(p_n)_{n \in \mathbb{N}}$ est donné par $p_n = \mu(\{n\})$.

Si Y est une variable aléatoire de loi μ , on pose

$$G_\mu(t) = G_Y(t) = \sum_{n \geq 0} \mathbb{P}(Y = n)t^n = \sum_{n \geq 0} p_n t^n.$$

Cette série converge absolument sur $[-1, 1]$ et on a $G_\mu(1) = 1$.

On commence par une formule de sommation très utile.

Proposition 3.10. (i) Soient Y_1, Y_2, \dots des variables indépendantes à valeurs dans \mathbb{N} . Alors, pour tout $n \in \mathbb{N}$,

$$G_{Y_1 + \dots + Y_n}(t) = G_{Y_1}(t) \cdot \dots \cdot G_{Y_n}(t),$$

pour tout t dans le domaine de convergence des fonctions G_{Y_k} .

(ii) Supposons que les variables $(Y_k)_{k \geq 1}$ sont i.i.d de loi μ et soit N une variable à valeurs dans \mathbb{N} , indépendante de des variables $(Y_k)_{k \geq 1}$. Alors si on pose $S = \sum_{k=1}^N Y_k$,

$$G_S(t) = G_N(G_\mu(t)),$$

Preuve: Le point (i) suit des propriétés d'indépendance:

$$G_{Y_1 + \dots + Y_n}(t) = \mathbb{E}(t^{Y_1 + \dots + Y_n}) = \mathbb{E}(t^{Y_1}) \cdot \dots \cdot \mathbb{E}(t^{Y_n}) = G_{Y_1}(t) \cdot \dots \cdot G_{Y_n}(t).$$

Passons au point (ii). Ecrivons la loi de probabilité totale pour l'espérance (2.15):

$$\begin{aligned} G_S(t) &= \mathbb{E}(t^S) = \sum_{n=0}^{\infty} \mathbb{E}\left(t^{\sum_{k=1}^N Y_k} \mid N = n\right) \mathbb{P}(N = n) \\ &= \sum_{n=0}^{\infty} \mathbb{E}\left(t^{\sum_{k=1}^n Y_k}\right) \mathbb{P}(N = n) \\ &= \sum_{n=0}^{\infty} \prod_{k=1}^n \mathbb{E}(t^{Y_k}) \mathbb{P}(N = n) = \sum_{n=0}^{\infty} G_\mu(t)^n \mathbb{P}(N = n) = G_N(G_\mu(t)). \end{aligned}$$

\square

Un processus de branchement (ou de Galton-Watson) est censée décrire l'évolution d'une population avec des règles de reproductions simplifiées, mais aléatoires. Mathématiquement, un processus de branchement est une séquence de variables aléatoires $(X_n)_n$ définie de la manière suivante. Chaque variable X_n représente le nombre d'individus de la population à la génération n .

Au moment initial, on commence par un unique individu ($X_0 = 1$). En passant de la génération n à $n + 1$, chacun des X_n individus de la génération n donne naissance à un nombre aléatoire d'enfants de loi μ , puis meure. De plus, les nombres d'enfants des différents individus sont indépendants.

La suite de variables $(X_n)_{n \geq 0}$ ainsi définie est appelée un processus de branchement (ou de Galton-Watson) de loi de reproduction μ .

Pour définir cela formellement, soit $(Y_{n,k})_{n,k \in \mathbb{N}}$ une famille de variables i.i.d. de loi μ . Posons $X_0 = 1$ et, pour tout $n \geq 0$,

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k}. \quad (3.5)$$

Dans cette formalisation, $Y_{n,k}$ représente le nombre d'enfants du k ème individu de la génération n (si $X_n \geq k$; sinon $Y_{n,k}$ n'intervient pas dans la définition de X_{n+1}). Observons que l'indépendance des variables $(Y_{n,k})_{n,k \in \mathbb{N}}$ garantie que les progénitures de chaque individu de la génération n sont indépendantes entre elles et de X_1, \dots, X_n .

Si pour un certain n on a $X_n = 0$, alors les variables suivantes sont toutes nulles: $X_{n+1} = X_{n+2} = \dots = 0$. Quand cela arrive, on dit que la population s'éteint et on définit la probabilité d'extinction

$$\nu := \mathbb{P}(\exists n \text{ t.q. } X_n = 0) = \mathbb{P}\left(\bigcup_{n \geq 1} \{X_n = 0\}\right) = \lim_{n \rightarrow \infty} \mathbb{P}(X_n = 0).$$

Proposition 3.11. *Soit G_μ la fonction génératrice des variables μ . Alors la probabilité d'extinction ν est la plus petite solution positive de l'équation*

$$G_\mu(t) = t.$$

En particulier,

- si $\mathbb{E}(Y_{n,k}) = \sum_{k \geq 0} k\mu(k) < 1$, alors $\nu = 1$;
- si $\mathbb{E}(Y_{n,k}) > 1$, alors $\nu < 1$;
- si $\mathbb{E}(Y_{n,k}) = 1$ et $\mu(1) \neq 1$, alors $\nu = 1$;
- si $\mu(1) = 1$, alors $\nu = 0$.

Preuve: Soit G_n la fonction génératrice des probabilités de X_n . En appliquant la proposition 3.10 à la définition de X_n (3.5), on obtient

$$G_{n+1}(t) = G_n(G_\mu(t)) = (G_\mu \circ \dots \circ G_\mu)(t),$$

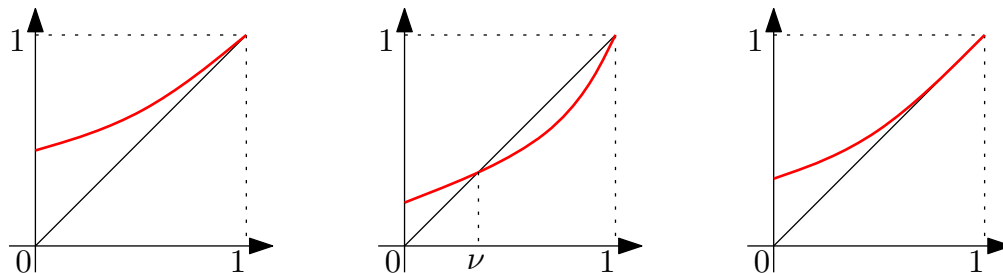


Figure 3.1: Les trois cas de figure possibles: $G'_\mu(1) < 1$, $G'_\mu(1) > 1$ et $G'_\mu(1) = 1$ (avec $\mu(1) \neq 1$). Dans le premier et troisième cas, le graphe de G_μ (en rouge) intersecte la diagonale $\{x = y\}$ uniquement au point 1, dans le deuxième cas, comme $G'_\mu(1) > 1$, il existe un deuxième point d'intersection $t_0 = \nu$ dans $[0, 1]$.

ou il y a $n + 1$ termes dans la composition. En particulier, si on pose $\nu_n = \mathbb{P}(X_n = 0) = G_n(0)$, on a

$$\nu_{n+1} = G_\mu(\nu_n), \quad \forall n \geq 0 \quad \text{et} \quad \nu_0 = 0. \quad (3.6)$$

La suite (ν_n) est croissante et $\nu = \lim_{n \rightarrow \infty} \nu_n$. Il s'en suit, en prenant $n \rightarrow \infty$ dans l'équation précédente (et en utilisant la continuité de G_μ sur $[0, 1]$), que ν est un point fixe de G_μ .

Soit t_0 la plus petite solution positive de $G_\mu(t) = t$. Observons que $G_\mu(1) = 1$, donc $t_0 \in [0, 1]$. On a $\nu_0 = 0 \leq t_0$. De plus G_μ (comme toute fonction génératrice des probabilités) est croissante sur $[0, 1]$. Ainsi, par récurrence, on trouve $\nu_n \leq t_0$ pour tout $n \geq 0$, donc aussi $\nu = \lim_n \nu_n \leq t_0$. Comme ν est un point fixe de G_μ , on a $\nu = t_0$.

Passons à l'étude des différents cas pour ν en fonction de $\mathbb{E}(Y_{n,k})$. Observons que, pour $t \in [0, 1]$,

$$G''_\mu(t) = \sum_{k \geq 2} k(k-1)t^{k-2}\mu(k) \geq 0.$$

Ainsi, la fonction G_μ est convexe, même strictement convexe si $\mu(\{2, 3, \dots\}) > 0$. On en déduit que G_μ admet au plus deux points fixes (c.-à-d. des intersections avec la droite $\{x = y\}$) dans $[0, 1]$. Un point fixe est 1; un deuxième peut exister, cela dépend de la tangente de G_μ au point 1:

- si $G'_\mu(1) < 1$, alors G_μ n'admet pas un autre point fixe dans $[0, 1]$, donc $\nu = 1$;
- si $G'_\mu(1) > 1$, alors G_μ admet un second point fixe dans $[0, 1]$ (car $G_\mu(0) \geq 0$), donc $\nu < 1$;
- si $G'_\mu(1) = 1$ et G_μ est strictement convexe, alors G_μ n'admet pas un autre point fixe dans $[0, 1]$, donc $\nu = 1$.

Ces trois cas sont illustrés dans la figure 3.1. En fin, rappelons que $G'_\mu(1) = \mathbb{E}(Y_{n,k})$. Ainsi les deux premiers points ci-dessus correspondent au deux premiers points de la proposition.

Si $G'_\mu(1) = \mathbb{E}(Y_{n,k}) = 1$, alors soit $\mu(1) = 1$, soit G_μ est strictement convexe. Dans le premier cas, on trouve $\mathbb{P}(X_0 = X_1 = \dots = 1) = 1$ et $\nu = 0$, dans le second on a $\nu = 1$ par la discussion précédente. \square

Chapter 4

Théorèmes limites

4.1 Convergence des variables aléatoires

Considérons une suite de variable aléatoires $(X_n)_{n \geq 1}$ définies sur un même espace $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 4.1. Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires et X une variable aléatoire.

- On dit que X_n converge vers X en probabilité, si pour tout $\epsilon > 0$,

$$\mathbb{P}(|X - X_n| \geq \epsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

On l'écrit $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$.

- On dit que X_n converge vers X presque sûrement, si

$$\mathbb{P}(X_n \xrightarrow[n \rightarrow \infty]{} X) = \mathbb{P}(\{\omega \in \Omega : X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega)\}) = 1.$$

On l'écrit $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$.

- On dit que X_n converge vers X en loi, si pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue et bornée,

$$\mathbb{E}[f(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[f(X)].$$

On l'écrit $X_n \xrightarrow[n \rightarrow \infty]{(d)} X$ ou $X_n \Rightarrow X$.

- On dit que X_n converge vers X dans L^p (pour $p \geq 1$) si

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

On l'écrit $X_n \xrightarrow[n \rightarrow \infty]{L^p} X$.

Proposition 4.2. *La convergence p.s. implique la convergence en probabilité, qui implique celle en loi.*

Les implications inverses sont fausses. De plus, aucune de ces convergences n'implique la convergence dans L^p pour $p \geq 1$. Inversement, la convergence dans L^p implique celle en probabilité et en loi, mais pas celle p.s..

Par la suite, les convergences qui vont nous intéresser le plus sont celle p.s. et celle en loi.

Preuve: Supposons que $X_n \xrightarrow{p.s.} X$ et montrons la convergence en probabilité. Soit $\epsilon > 0$ et notons

$$A_n = \{|X_n - X| \geq \epsilon\}.$$

Alors

$$\left\{ \limsup_{n \rightarrow \infty} |X_n - X| \geq \epsilon \right\} = \bigcap_{N \geq 1} \bigcup_{n \geq N} A_n.$$

Mais l'ensemble à gauche est de probabilité nulle par hypothèse. Il s'en suit que $\mathbb{P}(A_N) \leq \mathbb{P}(\bigcup_{n \geq N} A_n) \rightarrow 0$, ce qui montre la convergence en probabilité.

Supposons maintenant que $X_n \xrightarrow{\mathbb{P}} X$ et montrons la convergence en loi. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ continue et bornée (par $\|f\|_\infty := \sup\{|f(x)| : x \in \mathbb{R}\} < \infty$).

Fixons $\epsilon > 0$. Soit $M \geq 0$ tel que $\mathbb{P}(|X| < M) \leq \epsilon/\|f\|_\infty$. Sur l'intervalle compact $[-M, M]$ la fonction f est uniformément continue. Soit $\delta > 0$ tel que, si $x, y \in [-M, M]$ sont tels que $|x - y| < \delta$, alors $|f(x) - f(y)| < \epsilon$. Alors on peut écrire

$$\begin{aligned} \mathbb{E}[|f(X_n) - f(X)|] &\leq \mathbb{E}[|f(X_n) - f(X)|\mathbf{1}_{|X| > M}] \\ &\quad + \mathbb{E}[|f(X_n) - f(X)|\mathbf{1}_{|X| \leq M}\mathbf{1}_{|X_n - X| \geq \delta}] \\ &\quad + \mathbb{E}[|f(X_n) - f(X)|\mathbf{1}_{|X| \leq M}\mathbf{1}_{|X_n - X| < \delta}] \\ &\leq 2\|f\|_\infty \mathbb{P}(|X| > M) + 2\|f\|_\infty \mathbb{P}(|X_n - X| \geq \delta) + \epsilon \\ &= 3\epsilon + 2\|f\|_\infty \mathbb{P}(|X_n - X| \geq \delta) \end{aligned}$$

Notre hypothèse nous dit que $\mathbb{P}(|X_n - X| \geq \delta) \rightarrow 0$ quand $n \rightarrow \infty$. Ainsi,

$$\limsup_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \limsup_{n \rightarrow \infty} \mathbb{E}[|f(X_n) - f(X)|] \leq 3\epsilon,$$

pour tout $\epsilon > 0$. On conclut que la limite est nulle, donc que $X_n \xrightarrow{(d)} X$. □

Théorème 4.3. *Soient $X, X_1, X_2 \dots$ des variables aléatoires. Les affirmations suivantes sont équivalentes*

- (i) X_n converge en loi vers X ;
- (ii) Pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ de classe C^∞ et dont toutes les dérivées sont bornées, $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ quand $n \rightarrow \infty$.

- (iii) Pour tout ouvert $U \subset \mathbb{R}$, $\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in U) \geq \mathbb{P}(X \in U)$.
- (iv) Pour tout fermé $F \subset \mathbb{R}$, $\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$.
- (v) Pour tout $x \in \mathbb{R}$ ou F_X est continue, $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$.
- (vi) $\Phi_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} \Phi_X(t)$ pour tout $t \in \mathbb{R}$.

Preuve: On va uniquement montrer l'équivalence entre (i),(ii) et (v). On parlera brièvement du liens avec les autres points à la fin.

(i) \Rightarrow (ii) est évident.

(ii) \Rightarrow (v): Soit $t \in \mathbb{R}$ tel que F_X est continue en t . Pour $\delta > 0$ soient $f, g : \mathbb{R} \rightarrow [0, 1]$ deux fonctions de classe \mathcal{C}^∞ telles que:

- $f(x) = 1$ si $x \leq t$ et $f(x) = 0$ si $x > t + \delta$,
- $g(x) = 1$ si $x < t - \delta$ et $g(x) = 0$ si $x \geq t$.

La construction de telles fonctions f et g peut se faire explicitement, voir la fin de la preuve. Toutes les dérivées de f et de g sont continues et nulles en dehors de $[t - \delta, t + \delta]$, elles sont donc bornées. Le point (ii) s'applique donc à f et g .

Par construction de f et g , on a

$$F_{X_n}(t) \leq \mathbb{E}[f(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[f(X)] \leq F_X(t + \delta),$$

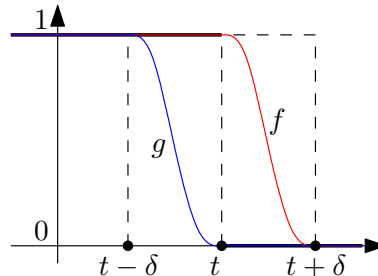
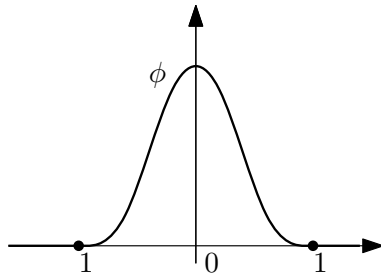
et

$$F_{X_n}(t) \geq \mathbb{E}[g(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[g(X)] \geq F_X(t - \delta).$$

Ainsi,

$$F_X(t - \delta) \leq \liminf_{n \rightarrow \infty} F_{X_n}(t) \leq \limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t + \delta).$$

En prenant $\delta \rightarrow 0$ dans cette équation, comme F_X est continue en t , on obtient $F_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} F_X(t)$.



Construction de f et g : Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ la fonction définie par

$$\phi(x) = \begin{cases} e^{-\frac{1}{1-x^2}} & \text{si } |x| \leq 1, \\ 0 & \text{si } |x| > 1. \end{cases}$$

Alors ϕ est une fonction de classe \mathcal{C}^∞ , positive, non-nulle seulement sur un intervalle de \mathbb{R} . Posons

$$h(x) = 1 - \frac{\int_{-\infty}^x \phi(u) du}{\int_{-\infty}^{+\infty} \phi(u) du}.$$

Ainsi h est de classe \mathcal{C}^∞ , elle prend valeurs dans $[0, 1]$ et vaut 1 (et 0 respectivement) sur $(-\infty, -1]$ (et $[1, +\infty)$). Pour t, δ comme avant, posons

$$f(x) = h\left(\frac{x-t}{2\delta} - 1\right) \quad \text{et} \quad g(x) = h\left(\frac{x-t}{2\delta} + 1\right).$$

Alors f, g sont des fonctions avec les propriétés recherchées.

(v) \Rightarrow (i): Soit A l'ensemble des points de discontinuité de F_X . Le fait que F_X est croissante implique que A est au plus dénombrable (voir exercice 6). Pour $a \notin A$, on a

$$\mathbb{E}[\mathbf{1}_{(-\infty, a]}(X_n)] = F_{X_n}(a) \rightarrow F_X(a) = \mathbb{E}[\mathbf{1}_{(-\infty, a]}(X)].$$

En appliquant cela pour $a < b$ avec $a, b \in A$ et faisant la différence, on trouve aussi

$$\mathbb{E}[\mathbf{1}_{(a,b)}(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[\mathbf{1}_{(a,b)}(X)]. \quad (4.1)$$

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue bornée. On désire montrer que $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$. Supposons pour commencer que f est égale à 0 en dehors d'un intervalle I de \mathbb{R} .

Soit $\epsilon > 0$. Il existe alors une fonction constante par morceaux $g : \mathbb{R} \rightarrow \mathbb{R}$ telle que, $|g(x) - f(x)| < \epsilon$ pour tout $x \in \mathbb{R}$. De plus la fonction g peut être choisie

$$g = \sum_1^m \lambda_i \mathbf{1}_{(a_i, b_i]}$$

avec $a_i, b_i \in A$ pour tout i . Alors, par (4.1), $\mathbb{E}[g(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[g(X)]$. Ainsi

$$\begin{aligned} & \limsup_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \\ & \leq \limsup_{n \rightarrow \infty} [|\mathbb{E}[f(X_n) - g(X_n)]| + |\mathbb{E}[g(X_n) - g(X)]| + |\mathbb{E}[f(X) - g(X)]|] \leq 2\epsilon \end{aligned}$$

Comme $\epsilon > 0$ est arbitraire, on trouve $\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)] \xrightarrow{n \rightarrow \infty} 0$.

En fin, considérons le cas où f est continue et bornée. Notons $\|f\|_\infty = \sup\{|f(x)| : x \in \mathbb{R}\} < \infty$. Soient $\epsilon > 0$ et $M > 0$ tel que $\mathbb{P}(|X| < M) < \epsilon/\|f\|_\infty$. De (4.1) on déduit que $\mathbb{P}(|X_n| < M) < \epsilon/\|f\|_\infty$ pour n assez grand. Alors

$$|\mathbb{E}[f(X_n) - f(X)]| \leq |\mathbb{E}[f(X_n) - \tilde{f}(X_n)]| + |\mathbb{E}[\tilde{f}(X_n) - \tilde{f}(X)]| + |\mathbb{E}[f(X) - \tilde{f}(X)]|.$$

Dans la partie droite, le premier et dernier terme sont bornés par $2\|f\|_\infty \mathbb{P}(|X_n| > M)$ et $2\|f\|_\infty \mathbb{P}(|X| > M)$, respectivement. Ainsi, pour n assez grand, leur contribution est bornée de au plus 4ϵ . Le terme $\mathbb{E}[\tilde{f}(X_n) - \tilde{f}(X)]$ converge vers 0 par le point précédent appliqué à \tilde{f} . Ainsi on conclut que $\limsup_{n \rightarrow \infty} |\mathbb{E}[f(X_n) - f(X)]| < 4\epsilon$, et comme $\epsilon > 0$ est arbitraire, $|\mathbb{E}[f(X_n) - f(X)]| \rightarrow 0$. Ceci montre (i).

Traisons quelques implications entre (iii), (iv), (vi) et les points (i),(ii) et (v).

(iii) \Leftrightarrow (iv): Il s'agit simplement de passer au complémentaire.

(iv) et (iii) \Rightarrow (v): En supposant (iv) on trouve $\limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t)$ pour tout $t \in \mathbb{R}$. Par (iii), on trouve $\liminf_{n \rightarrow \infty} \mathbb{P}(X_n < t) \geq \mathbb{P}(X < t)$ pour tout $t \in \mathbb{R}$. Fixons $t \in \mathbb{R}$ un point de continuité de F_X . En appliquant la deuxième égalité à $t - \epsilon$ pour $\epsilon > 0$ arbitraire, on trouve:

$$F_X(t) = \mathbb{P}(X < t) \leq \liminf_{n \rightarrow \infty} \mathbb{P}(X_n < t) \leq \liminf_{n \rightarrow \infty} F_{X_n}(t) \leq \limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t)$$

En prenant $\epsilon \rightarrow 0$ on trouve $F_{X_n}(t) \xrightarrow{t \rightarrow \infty} F_X(t)$.

(i) \Rightarrow (vi): C'est immédiat car, pour tout $t \in \mathbb{R}$, $x \mapsto \cos(tx)$ et $x \mapsto \sin(tx)$ sont des fonctions continues bornées.

Les implications (i) \Rightarrow (iii) et (vi) \Rightarrow (i) sont plus compliquées et on va les admettre. \square

Proposition 4.4. Soient $(X_n)_{n \geq 1}$ et X des variables aléatoires admettant toutes des moments exponentiels d'ordre $\delta > 0$. Alors, si $M_{X_n}(t) \xrightarrow{n \rightarrow \infty} M_X(t)$ pour tout $t \in [-\delta, \delta]$, X_n converge en loi vers X .

On va admettre cette proposition.

Lemme de Borel-Cantelli On a vu dans la preuve de la proposition 4.2 que si $(A_n)_{n \geq 1}$ est une famille d'événements avec

$$\mathbb{P}\left(\bigcup_{N \geq 1} \bigcap_{n \geq N} A_n\right) = 0, \quad (4.2)$$

alors $\mathbb{P}(A_n) \rightarrow 0$ quand $n \rightarrow \infty$.

Le lemme suivant, nous dit qu'une implication inverse est possible, sous des hypothèses plus fortes.

Lemme 4.5 (Borel Cantelli). Soient $(A_n)_{n \geq 1}$ une famille d'événements avec $\sum_{n \geq 1} \mathbb{P}(A_n) < \infty$. Alors

$$\mathbb{P}\left(\bigcup_{N \geq 1} \bigcap_{n \geq N} A_n\right) = 0.$$

Mentionnons que l'équation (4.2) peut s'interpréter comme "la probabilité qu'une infinité d'événements A_n se produisent est nulle". En effet

$$\bigcup_{N \geq 1} \bigcap_{n \geq N} A_n = \{\omega \in \Omega : \text{il existe un ensemble infini } I \subset \mathbb{N} \text{ t.q. } \omega \in A_n \text{ pour tout } n \in I\}.$$

Preuve: Soit $(A_n)_{n \geq 1}$ comme dans l'énoncé. Alors, pour $N \geq 1$,

$$\mathbb{P}\left(\bigcup_{n \geq N} A_n\right) \leq \sum_{n \geq N} \mathbb{P}(A_n) \xrightarrow{N \rightarrow \infty} 0.$$

La convergence vient du fait que le terme central est le reste d'une série convergente. Ainsi

$$\mathbb{P}\left(\bigcap_{N \geq 1} \bigcup_{n \geq N} A_n\right) = \lim_{N \rightarrow \infty} \mathbb{P}\left(\bigcup_{n \geq N} A_n\right) = 0.$$

□

4.2 Loi des grands nombres

Théorème 4.6 (Loi forte des grands nombres). *Soient X_1, X_2, \dots une suite de variables aléatoires i.i.d. dans L^1 avec $\mu = \mathbb{E}(X_i)$. Alors*

$$S_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{p.s.} \mu.$$

Preuve: On va montrer ce théorème sous l'hypothèse supplémentaire que $X_n \in L^4$.

Observons pour commencer que $\mathbb{E}(S_n) = \frac{1}{n} \sum_{k=1}^n \mathbb{E}(X_k) = \mu$. On va étudier par la suite $\mathbb{P}(|S_n - \mu| > \epsilon)$ en utilisant l'inégalité de Markov. Pour simplifier les notations, on va supposer $\mu = 0$; si ce n'est pas le cas, il suffit de poser $\tilde{X}_n = X_n - \mu$ pour se réduire au cas où l'espérance est nulle.

Observons que

$$\begin{aligned} \mathbb{E}[S_n^4] &= \frac{1}{n^4} \mathbb{E}\left[\sum_{k=1}^n X_k^4 + 2 \sum_{1 \leq k < \ell \leq n} X_k^2 X_\ell^2 + \sum_{k=1}^n X_k \left(\sum_{\ell \neq k} X_\ell\right)^3\right] \\ &= \frac{1}{n^3} \mathbb{E}(X_1^4) + \frac{n-1}{n^3} \mathbb{E}(X_1^2 X_2^2) < \frac{C}{n^2}, \end{aligned}$$

pour une constante C (on peut prendre $C = \mathbb{E}(X_1^4) + \mathbb{E}(X_1^2 X_2^2)$ par exemple). Ainsi, pour $\epsilon > 0$ fixé,

$$\mathbb{P}(|S_n| > \epsilon) = \mathbb{P}[S_n^4 > \epsilon^4] \leq \frac{1}{\epsilon^4} \mathbb{E}[S_n^4] \leq \frac{C}{\epsilon^4 n^2}.$$

On conclut donc que $\sum_{n \geq 1} \mathbb{P}(|S_n| > \epsilon) < \infty$.

On suit maintenant un argument standard pour en déduire la convergence p.s. de S_n vers 0. Pour $N \geq 1$, on a

$$\mathbb{P}(\exists n \geq N \text{ t.q. } |S_n| > \epsilon) = \mathbb{P}\left(\bigcup_{n \geq N} \{|S_n| > \epsilon\}\right) \leq \sum_{n \geq N} \mathbb{P}(|S_n| > \epsilon) \xrightarrow[N \rightarrow \infty]{} 0.$$

Ainsi, par passage au complémentaire,

$$\begin{aligned} \mathbb{P}(\exists N \geq 1 \text{ t.q. } \forall n \geq N, |S_n| \leq \epsilon) &= \mathbb{P}\left(\bigcup_{N \geq 1} \{\forall n \geq N, |S_n| \leq \epsilon\}\right) \\ &= \lim_{N \rightarrow \infty} \mathbb{P}(\forall n \geq N, |S_n| \leq \epsilon) = 1. \end{aligned}$$

On viens donc de montrer que $\mathbb{P}(\limsup_{n \rightarrow \infty} |S_n| \leq \epsilon) = 1$. En prenant l'intersection décroissante sur $\epsilon = 1/k$ pour $k \geq 1$, on conclut que $\mathbb{P}(\limsup_{n \rightarrow \infty} |S_n| = 0) = 1$, donc que $S_n \xrightarrow{p.s.} 0$. \square

Remarque 4.7. D'autres variantes de ce théorème existent. Certaines (connues comme lois faibles des grandes nombres) stipule une convergence en probabilité, plutôt que presque sure.

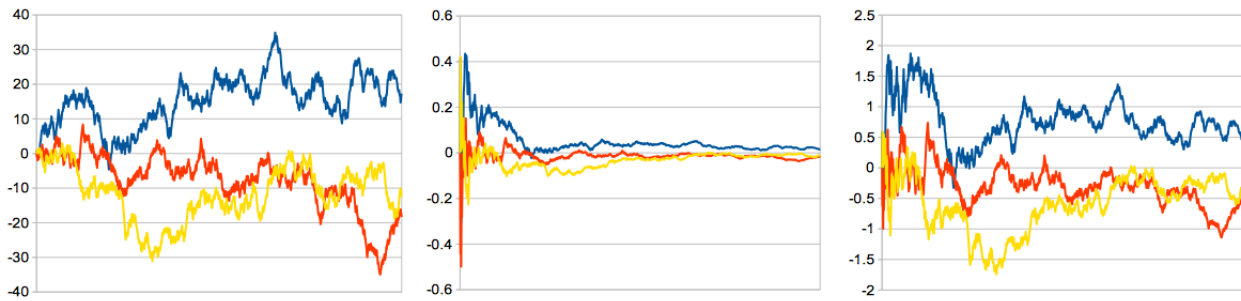


Figure 4.1: Trois instances de l'évolution de $nS_n = \sum_{k=1}^n X_k$, de S_n et de $\sqrt{n}S_n$, respectivement, pour $n = 1, \dots, 1000$. Les variables X_k sont d'espérance nulle et variance 1. On voit que S_n converge vers 0 pour les trois instances; que nS_n diverge (sa valeur absolue devient de plus en plus grande), mais que $\sqrt{n}S_n$ rester du même ordre, sans que les trajectoires individuelles convergent vers une valeur précise.

Exercice 29.

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d. de loi de Cauchy.

- (i) Montrer que, pour $\lambda \in (0, 1)$, la variable $\lambda X_1 + (1 - \lambda)X_2$ suit aussi une loi de Cauchy.
Indication: Ecrire la densité jointe de X_1, X_2 et utiliser (2.8).
- (ii) Dédire la loi de $S_n = \frac{1}{n} \sum_{k=1}^n X_k$. Pourquoi cela ne contredit pas la loi des grands nombres?

4.3 Théorème central limite

La loi des grandes nombres nous montre que, si $(X_n)_{n \geq 0}$ sont i.i.d. d'espérance nulle, alors $S_n = \frac{1}{n} \sum_{k=1}^n X_k \rightarrow 0$ presque sûrement. On peut alors se demander quelles sont les variations de S_n autour de 0. Le théorème central limite décrit ces variations. Elles sont d'ordre $1/\sqrt{n}$, ce qui revient à dire que $S_n \sqrt{n}$ est d'ordre constant. Ceci n'est pas une surprise car si X_k admet un moment d'ordre 2, alors $\text{Var}(S_n) = \text{Var}(X_1)/n$, donc $\text{Var}(\sqrt{n}S_n) = \text{Var}(X_1)$ pour tout n .

On peut donc se demander si $S_n\sqrt{n}$ converge quand $n \rightarrow \infty$. Le théorème central limite nous dit que c'est le cas, mais pas " ω par ω "; le comportement de la variable aléatoire $S_n\sqrt{n}$ s'approche d'une loi limite, notamment la loi normale, quand n augmente.

Théorème 4.8. Soit $(X_n)_{n \geq 0}$ une suite de variables i.i.d. de variance finie, avec $\mathbb{E}(X) = \mu$ et $\text{Var}(X) = \sigma^2$. Alors

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - \mu) \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{N}(0, \sigma^2). \quad (4.3)$$

ou $\mathcal{N}(0, \sigma^2)$ suit une loi gaussienne d'espérance 0 et variance σ^2 .

On va donner deux preuves différentes de ce théorème. Dans chaque une on aura besoin d'hypothèses supplémentaires. La première utilise la fonction génératrice définie dans le chapitre précédent.

Preuve: Quitte à remplacer les variables X_k par $X_k - \mu$, on peut supposer que $\mathbb{E}(X_k) = 0$. Supposons de plus que les variables X_k admettent un moment exponentiel, c.-à-d. que $M_{X_k}(t) < \infty$ pour $|t| < \delta$ pour un certain $\delta > 0$.

Notons $Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$. Ecrivons M_X pour la fonction des variables X_k . Alors, pour $|t| < \delta$,

$$\log M_{Y_n}(t) = \log \mathbb{E}[e^{t(X_1 + \dots + X_n)/\sqrt{n}}] = n \log M_X(t/\sqrt{n}).$$

Ecrivons la formule de Taylor pour la fonction $\log M_X(\cdot)$ autour de 0:

$$\log M_X(s) = \log M_X(0) + \frac{M'_X(0)}{M_X(0)}s + \frac{M''_X(0) - M'_X(0)^2}{2M_X(0)^2}s^2 + O(s^3).$$

De plus, $M_X(0) = 1$, $M'_X(0) = \mathbb{E}(X) = 0$ et $M''_X(0) = \mathbb{E}(X^2) = \text{Var}(X) = \sigma^2$. Ainsi

$$\log M_{Y_n}(t) = n \left[\frac{\sigma^2}{2} (t/\sqrt{n})^2 + O((t/\sqrt{n})^3) \right] \xrightarrow[n \rightarrow \infty]{} \frac{\sigma^2 t^2}{2}.$$

On trouve donc $M_{Y_n}(t) \xrightarrow[n \rightarrow \infty]{} \exp(\frac{\sigma^2 t^2}{2})$. On reconnaît à droite la fonction génératrice des moments d'une normale N centrée, de variance σ^2 . Ainsi, par la proposition 4.4, $Y_n \xrightarrow[n \rightarrow \infty]{(d)} N$. \square

L'hypothèse $\mathbb{E}(e^{\delta|X_i|}) < \infty$ utilisée dans la preuve est assez restrictive, mais nécessaire pour la preuve précédente. Pour s'en débarrasser, on peut utiliser la fonction caractéristique à la place de celle génératrice.

On donne une deuxième preuve, plus constructive. Remarquons que la première preuve est basée sur l'affirmation assez obscure que la convergence des fonctions génératrices implique la convergence en loi. Cette deuxième preuve est complètement explicite.

Preuve: Quitte à remplacer X_i par $\frac{X_i - \mu}{\sigma}$, on peut supposer que $\mathbb{E}(X_i) = 0$ et $\text{Var}(X_i) = 1$. On va faire la preuve que dans le cas où $\mathbb{E}[|X_n|^3] < \infty$.

Soit (N_i) une suite de variables i.i.d. gaussiennes d'espérance 0 et variance 1. Pour n fixé, $\mathcal{N} = \frac{1}{\sqrt{n}} \sum_{i=1}^n N_i$ est aussi une gaussienne d'espérance 0 et variance 1 (voir exercice 22). Posons

$$Z_k = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^k X_i + \sum_{i=k+1}^n N_i \right).$$

Ainsi $Z_0 = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ et $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n N_i = \mathcal{N}$. De plus, posons

$$U_k = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{k-1} X_i + \sum_{i=k+1}^n N_i \right),$$

de sorte que $Z_k = U_k + \frac{X_k}{\sqrt{n}}$ et $Z_{k-1} = U_k + \frac{N_k}{\sqrt{n}}$.

Soit f une fonction classe \mathcal{C}^∞ , nulle en dehors d'un intervalle (en particulier f et f''' sont bornées). On veut estimer

$$\mathbb{E} \left[f \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right) \right] - \mathbb{E}[f(\mathcal{N})] = \mathbb{E}[f(Z_0) - f(Z_n)],$$

plus précisément on veut montrer que la quantité au-dessus tend vers 0 quand $n \rightarrow \infty$. Cela suffit pour déduire la convergence en loi par le Théorème 4.3. En utilisant la formule de Taylor à l'ordre 3 autour des points U_i , on écrit:

$$\begin{aligned} & \mathbb{E}[f(Z_n) - f(Z_0)] \\ &= \sum_{i=1}^n \mathbb{E}[f(Z_i) - f(Z_{i-1})] \\ &= \sum_{i=1}^n \mathbb{E} \left[f \left(U_i + \frac{N_i}{\sqrt{n}} \right) - f \left(U_i + \frac{X_i}{\sqrt{n}} \right) \right] \\ &= \sum_{i=1}^n \mathbb{E} \left[f'(U_i) \frac{N_i - X_i}{\sqrt{n}} + f''(U_i) \frac{N_i^2 - X_i^2}{2n} + \frac{1}{6n\sqrt{n}} (f'''(Y_i) N_i^3 - f'''(\tilde{Y}_i) X_i^3) \right], \end{aligned}$$

où Y_i et \tilde{Y}_i sont des variables comprises entre U_i et $U_i + N_i$ et U_i et $U_i + X_i$, respectivement. Rappelons nous que X_i et N_i sont indépendantes de U_i et que les deux ont la même espérance et variance. Ainsi

$$\mathbb{E} \left[\frac{1}{\sqrt{n}} f'(U_i) (N_i - X_i) \right] = \mathbb{E} \left[\frac{1}{\sqrt{n}} f'(U_i) \right] \mathbb{E}[N_i - X_i] = 0$$

et

$$\mathbb{E} \left[\frac{1}{2n} f''(U_i) (N_i^2 - X_i^2) \right] = \mathbb{E} \left[\frac{1}{2n} f''(U_i) \right] \mathbb{E}[N_i^2 - X_i^2] = 0.$$

Soit $M = \sup\{|f'''(x)| : x \in \mathbb{R}\}$. Vu les deux dernières équations,

$$\begin{aligned} \left| \mathbb{E} \left[f \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right) - f(\mathcal{N}) \right] \right| &= \frac{1}{6n\sqrt{n}} \left| \sum_{i=1}^n \mathbb{E} \left[f'''(Y_i) N_i^3 - f'''(\tilde{Y}_i) X_i^3 \right] \right| \\ &\leq \frac{1}{6n\sqrt{n}} \sum_{i=1}^n \mathbb{E} \left[|f'''(Y_i) N_i^3| + |f'''(\tilde{Y}_i) X_i^3| \right] \\ &\leq \frac{M}{6\sqrt{n}} \mathbb{E} [|N_1^3| + |X_1^3|] \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Ainsi $\mathbb{E} \left[f \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right) \right] \rightarrow \mathbb{E} [f(\mathcal{N})]$, et donc $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{N}$. □

Références

- [1] G. Grimmett et D. Welsh, *Probability, an Introduction*, Oxford University Press (2014).
- [2] Y. Velenik, *Probabilités et Statistique*, notes du cours donné à l'Université de Genève, <http://www.unige.ch/math/folks/velenik/Cours/2016-2017/Probastat/probastat.pdf>