

Probabilités


Résumé de cours et guide de lecture

Charles SUQUET

<http://math.univ-lille1.fr/~suquet/>

Documentation du cours

- [ICP] Ch. Suquet, *Introduction au calcul des probabilités*, polycopié de L2, version 2010–2011, distribué en version imprimée.
<http://math.univ-lille1.fr/~suquet/Polys/ICP.pdf>
- [IFP] Ch. Suquet, *Intégration, Analyse de Fourier, Probabilités*, polycopié de Licence (L3) 2003–2004, réédition du 2 septembre 2010, distribué en version imprimée.
<http://math.univ-lille1.fr/~suquet/Polys/IFP.pdf>
- [PVIR] Ch. Suquet, *Probabilités via l'intégrale de Riemann*, Ellipses, collection Références Sciences, 2013, ISBN 978-2-7298-77613, disponible à la B.U. et partiellement accessible sur le Web à l'URL provisoire :
<http://math.univ-lille1.fr/~suquet/Polys/Agrint13.pdf>

Ce résumé de cours contient les résultats qu'il faut connaître, sans les démonstrations ni la plupart des exemples évoqués en cours. Il vous donne aussi des indications de lecture dans les trois documents référencés ci-dessus. Ces indications sont signalées dans le texte par le symbole .

Chapitre 1

Le modèle probabiliste

1.1 Rappels de théorie de la mesure

Soit Ω un ensemble, on note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω .

Définition 1.1 (tribu). Une famille \mathcal{F} de parties de Ω est appelée tribu (ou σ -algèbre) sur Ω si elle

- a) possède l'ensemble vide : $\emptyset \in \mathcal{F}$;
- b) est stable par passage au complémentaire : $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$;
- c) est stable par union dénombrable : $(\forall i \in \mathbb{N}^*, A_i \in \mathcal{F}) \Rightarrow \bigcup_{i \in \mathbb{N}^*} A_i \in \mathcal{F}$.

Le couple (Ω, \mathcal{F}) est appelé espace mesurable.

On vérifie à partir de cette définition qu'une tribu est stable par unions finies (prendre tous les A_i vides à partir d'un certain rang), intersections finies et par intersections dénombrables (combiner b) et c)).

Définition 1.2 (mesure). Soit \mathcal{F} une tribu sur Ω . On appelle mesure positive sur (Ω, \mathcal{F}) une application

$$m : \mathcal{F} \longrightarrow [0, +\infty],$$

vérifiant

- a) $m(\emptyset) = 0$;
- b) m est σ -additive : pour toute suite $(A_i)_{i \in \mathbb{N}^*}$ d'éléments de \mathcal{F} deux à deux disjoints,

$$m\left(\bigcup_{i \in \mathbb{N}^*} A_i\right) = \sum_{i=1}^{+\infty} m(A_i). \quad (1.1)$$

Attention, cette définition de la mesure n'est pas celle qui vous a été donnée dans le cours du S5 sur l'intégration au sens de Lebesgue (mesure extérieure). Il importe de noter que l'on définit souvent une mesure sur une tribu \mathcal{F} plus petite que $\mathcal{P}(\Omega)$.

Les trois exemples les plus simples de tribu sont les suivants.

- La tribu triviale sur Ω est $\mathcal{F} = \{\Omega, \emptyset\}$.
- $\mathcal{P}(\Omega)$ est une tribu.
- Si A est une partie de Ω , alors $\mathcal{F} := \{\Omega, \emptyset, A, A^c\}$ est une tribu. C'est la *plus petite* tribu possédant A comme élément, au sens où toute tribu \mathcal{G} telle que $A \in \mathcal{G}$ contient \mathcal{F} . On dit que \mathcal{F} est la tribu *engendrée* par A .

Cette notion de tribu engendrée se généralise en remarquant que si $(\mathcal{G}_i)_{i \in I}$ est une famille quelconque de tribus sur Ω , $\mathcal{G} := \bigcap_{i \in I} \mathcal{G}_i$ est une tribu sur Ω (vérification immédiate à partir de la définition 1.1).

Définition 1.3 (tribu engendrée). Soit \mathcal{C} une famille de parties d'un ensemble Ω . On appelle tribu engendrée par \mathcal{C} , et on note $\sigma(\mathcal{C})$, la plus petite tribu contenant \mathcal{C} . C'est l'intersection de toutes les tribus sur Ω contenant \mathcal{C} .

Définition 1.4 (tribu borélienne). On appelle tribu borélienne sur \mathbb{R}^d la tribu engendrée par la famille \mathcal{O} des ensembles ouverts¹ de \mathbb{R}^d . On la notera $\text{Bor}(\mathbb{R}^d)$. Ainsi $\text{Bor}(\mathbb{R}^d) = \sigma(\mathcal{O})$. Les sous-ensembles de \mathbb{R}^d qui sont éléments de sa tribu borélienne sont appelés boréliens de \mathbb{R}^d ou boréliens tout court quand il n'y a pas d'ambiguïté.

Remarque 1.5. On peut démontrer que $\text{Bor}(\mathbb{R})$ est aussi engendrée par les fermés de \mathbb{R} , ou par les intervalles ouverts, ou les intervalles fermés, ou les semi-ouverts, ou les intervalles (ouverts ou fermés) à extrémités rationnelles, ou les intervalles $] - \infty, a]$, ou les intervalles $[a, +\infty[$. De même, $\text{Bor}(\mathbb{R}^d)$ est engendrée par les pavés ouverts ou par les pavés de la forme $\prod_{k=1}^d]a_k, b_k]$.

Voyons maintenant des exemples importants de mesures.

Exemple 1.6 (masse de Dirac). Soit x_0 un élément fixé de Ω . On appelle *masse de Dirac* au point x_0 ou *mesure de Dirac* au point x_0 , la mesure δ_{x_0} sur $(\Omega, \mathcal{P}(\Omega))$ définie par

$$\forall A \in \mathcal{P}(\Omega), \quad \delta_{x_0}(A) := \begin{cases} 1 & \text{si } x_0 \in A, \\ 0 & \text{si } x_0 \notin A. \end{cases}$$

Par restriction, δ_{x_0} est aussi une mesure sur (Ω, \mathcal{F}) pour toute tribu \mathcal{F} sur Ω .

Exemple 1.7 (série de mesures finies). Si les $(\mu_k)_{k \in \mathbb{N}}$ sont des mesures finies² sur (Ω, \mathcal{F}) et si $(a_k)_{k \in \mathbb{N}}$ est une suite de réels positifs, la fonction d'ensembles $\mu = \sum_{k \in \mathbb{N}} a_k \mu_k$ définie sur \mathcal{F} par

$$\mu : \mathcal{F} \rightarrow \overline{\mathbb{R}}_+, \quad A \mapsto \mu(A) := \sum_{k \in \mathbb{N}} a_k \mu_k(A)$$

est une mesure sur (Ω, \mathcal{F}) . Le résultat se généralise à $\mu = \sum_{i \in I} a_i \mu_i$, avec I au plus dénombrable.

Exemple 1.8 (mesure ponctuelle). Soient I un ensemble au plus dénombrable, $(x_i)_{i \in I}$ une famille dans Ω et $(a_i)_{i \in I}$ une famille de réels positifs. Alors $\mu := \sum_{i \in I} a_i \delta_{x_i}$ est une mesure sur $(\Omega, \mathcal{P}(\Omega))$, donc aussi par restriction sur tout espace mesurable (Ω, \mathcal{F}) . C'est un cas particulier de l'exemple précédent. Les mesures de ce type sont appelées *mesures ponctuelles*.

Remarque 1.9. Si $\Omega = \{\omega_i; i \in I\}$ est au plus dénombrable, toute mesure sur $(\Omega, \mathcal{P}(\Omega))$ qui est *finie sur les singletons* est une mesure ponctuelle.

Exemple 1.10. Soit μ une mesure sur (Ω, \mathcal{F}) et $B \in \mathcal{F}$. La fonction d'ensembles $\nu = \mu(\cdot \cap B)$ définie sur \mathcal{F} par

$$\forall A \in \mathcal{F}, \quad \nu(A) := \mu(A \cap B)$$

1. Un ensemble ouvert de \mathbb{R}^d est une réunion (quelconque) de *pavés* ouverts $\prod_{k=1}^d]a_k, b_k[$. Un *fermé* est le complémentaire d'un ouvert.

2. La mesure μ_k est *finie* si $\mu_k(A) < +\infty$ pour tout $A \in \mathcal{F}$. Elle est donc aussi *bornée* puisque $\mu_k(A) \leq \mu_k(\Omega) < +\infty$. L'hypothèse « μ_k finie » nous évite ici la gestion du conflit entre $a_k = 0$ et $\mu_k(A) = +\infty$.

est une mesure sur (Ω, \mathcal{F}) . Si de plus $0 < \mu(B) < +\infty$, la fonction d'ensembles μ_B définie sur \mathcal{F} par

$$\forall A \in \mathcal{F}, \quad \mu_B(A) := \frac{\mu(A \cap B)}{\mu(B)}$$

est une mesure sur (Ω, \mathcal{F}) , vérifiant $\mu_B(\Omega) = 1$, autrement dit une *probabilité*³. Quand $\mu = P$ est déjà une probabilité sur (Ω, \mathcal{F}) , la mesure P_B est appelée *probabilité conditionnelle*; notation : $P_B(A) =: P(A | B)$.

Exemple 1.11 (mesure de Lebesgue λ_d). Nous admettons qu'il existe une unique mesure μ sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$ telle que pour tout pavé $\prod_{i=1}^d]a_i, b_i]$ non vide⁴,

$$\mu \left(\prod_{i=1}^d]a_i, b_i] \right) = \prod_{i=1}^d (b_i - a_i).$$

On l'appelle *mesure de Lebesgue* sur \mathbb{R}^d et on la note λ_d . Pour $d = 1$, on étend ainsi la notion de longueur des intervalles à tous les ensembles membres de $\text{Bor}(\mathbb{R})$, pour $d = 2$ on étend de même la notion d'aire des rectangles à tous les ensembles boréliens du plan \mathbb{R}^2 , pour $d = 3$, on étend la notion de volume. Pour $d > 3$, on continuera à parler de volume ou d'hypervolume. Voici les principales propriétés de la mesure de Lebesgue.

Proposition 1.12. *La mesure de Lebesgue λ_d sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$ a les propriétés suivantes.*

- i) λ_d est invariante par translations : si $h : x \mapsto x + v$ est une translation de \mathbb{R}^d , alors pour tout $B \in \text{Bor}(\mathbb{R}^d)$, $h(B) \in \text{Bor}(\mathbb{R}^d)$ et $\lambda_d(B) = \lambda_d(h(B))$.
- ii) λ_d est invariante par toute isométrie euclidienne de \mathbb{R}^d : symétrie, rotation, etc.
- iii) Si h est l'homothétie $x \mapsto cx$ dans \mathbb{R}^d , pour tout borélien B , $\lambda_d(h(B)) = |c|^d \lambda_d(B)$.
- iv) λ_d ne charge pas les points : $\lambda_d(\{x\}) = 0$ pour tout $x \in \mathbb{R}^d$. Si $A \subset \mathbb{R}^d$ est fini ou dénombrable, $\lambda_d(A) = 0$.
- v) $\lambda_d \left(\prod_{i=1}^d]a_i, b_i] \right) = \lambda_d \left(\prod_{i=1}^d]a_i, b_i[\right)$ et cette égalité implique, bien sûr, l'égalité des mesures des 4^d pavés obtenus en jouant sur l'ouverture ou la fermeture des extrémités a_i, b_i des intervalles.
- vi) Si E est un sous-espace affine de \mathbb{R}^d et $E \neq \mathbb{R}^d$, $\lambda_d(E) = 0$.



IFP-Prop. 1.3.9.

1.2 Évènements

La théorie moderne des probabilités utilise le langage des ensembles pour modéliser une expérience aléatoire. Nous noterons Ω un ensemble dont les éléments représentent tous les résultats possibles ou *évènements élémentaires* d'une expérience aléatoire donnée. Les *évènements* (ou évènements composés) seront représentés par des parties (sous-ensembles) de Ω .

Il n'est pas toujours facile de trouver un ensemble Ω permettant de modéliser l'expérience aléatoire. Voici une règle pratique pour y arriver : les évènements élémentaires sont

3. En anticipant un peu sur la suite du chapitre.

4. Ce qui suppose implicitement que $a_i < b_i$ pour chaque $i \in \llbracket 1, d \rrbracket$.

ceux qui contiennent *l'information maximale* qu'il est possible d'obtenir de l'expérience. Par exemple si on jette un dé, l'évènement A : « obtention d'un chiffre pair » n'est pas élémentaire. Il est composé des trois évènements élémentaires 2, 4, 6 : $A = \{2, 4, 6\}$. Ici $\Omega = \{1, 2, 3, 4, 5, 6\}$. De même si on lance trois fois une pièce de monnaie, les évènements élémentaires sont des triplets comme (p, f, p) indiquant le résultat précis de chacun des trois lancers. Ici $\Omega = \{f, p\}^3$. L'évènement B « obtention de pile au deuxième des trois lancers » est composé : $B = \{(f, p, f); (f, p, p); (p, p, f); (p, p, p)\}$.

Notations	Vocabulaire ensembliste	Vocabulaire probabiliste
\emptyset	ensemble vide	évènement impossible
Ω	ensemble plein	évènement certain
ω	élément de Ω	évènement élémentaire
A	sous-ensemble de Ω	évènement
$\omega \in A$	ω appartient à A	Le résultat ω est une des réalisations possibles de A
$A \subset B$	A inclus dans B	A implique B
$A \cup B$	réunion de A et B	A ou B
$A \cap B$	intersection de A et B	A et B
A^c	complémentaire de A dans Ω	évènement contraire de A
$A \cap B = \emptyset$	A et B sont disjoints	A et B sont incompatibles

TABLE 1.1 – Langage ensembliste - langage probabiliste

Avec ce mode de représentation, les opérations logiques sur les évènements : « et », « ou », « négation » se traduisent par des opérations ensemblistes : intersection, réunion, passage au complémentaire. Le tableau 1.1 présente la correspondance entre les deux langages.

Les opérations logiques sur les évènements peuvent bien sûr faire intervenir plus de deux évènements. Ainsi, si A_1, \dots, A_n sont des évènements,

$$\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n$$

est l'ensemble des ω qui sont dans l'un au moins des A_i . C'est donc l'évènement « réalisation de l'un au moins des A_i ($1 \leq i \leq n$) ». De même :

$$\bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap \dots \cap A_n$$

est l'ensemble des ω qui sont dans tous les A_i . C'est donc l'évènement « réalisation de chacun des A_i ($1 \leq i \leq n$) ». Ceci s'étend aux réunions et intersections d'une suite infinie d'évènements :

$$\begin{aligned} \bigcup_{i \in \mathbb{N}^*} A_i &= \{\text{réalisation de l'un au moins des } A_i, i \in \mathbb{N}^*\}, \\ \bigcap_{i \in \mathbb{N}^*} A_i &= \{\text{réalisation de tous les } A_i, i \in \mathbb{N}^*\}. \end{aligned}$$

Ces opérations logiques sur des suites d'événements sont très utiles pour analyser des événements complexes à l'aide d'événements plus simples et, comme nous le verrons plus tard, calculer ainsi des probabilités.



Une question de dés, PVIR-5.2.3, pp. 163–166, document distribué en amphi.

1.3 La probabilité comme mesure

La *probabilité* P est une fonction qui à un événement, associe un nombre compris entre 0 et 1 et censé mesurer les chances de réalisation de cet événement. Il n'est pas toujours possible d'attribuer ainsi de manière cohérente une probabilité à chaque partie de Ω . En d'autres termes, P ne peut pas être considérée comme une *application* de l'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω dans $[0, 1]$ mais comme une *fonction* ayant pour domaine de définition une tribu \mathcal{F} généralement plus petite que $\mathcal{P}(\Omega)$. La tribu \mathcal{F} est aussi appelée famille des événements observables⁵.

Définition 1.13. Soit Ω un ensemble et \mathcal{F} une tribu sur Ω . On appelle *probabilité* sur (Ω, \mathcal{F}) toute application P de \mathcal{F} dans $[0, 1]$ vérifiant :

- (i) $P(\Omega) = 1$.
- (ii) Pour toute suite $(A_j)_{j \geq 1}$ d'événements de \mathcal{F} deux à deux disjoints (incompatibles) :

$$P\left(\bigcup_{j \in \mathbb{N}^*} A_j\right) = \sum_{j=1}^{+\infty} P(A_j).$$

Le triplet (Ω, \mathcal{F}, P) s'appelle *espace probabilisé*.

Définir une probabilité sur (Ω, \mathcal{F}) c'est en quelque sorte attribuer une « masse » à chaque événement observable, avec par convention une masse totale égale à 1 pour l'événement certain Ω . Une formulation équivalente à la définition 1.13 est : P est une *mesure* sur (Ω, \mathcal{F}) telle que $P(\Omega) = 1$.

Proposition 1.14 (propriétés générales d'une probabilité).

Toute probabilité P sur (Ω, \mathcal{F}) vérifie les propriétés suivantes :

1. $P(\emptyset) = 0$.
2. *Additivité.*
 - a) Si $A \in \mathcal{F}$ et $B \in \mathcal{F}$ sont incompatibles ($A \cap B = \emptyset$), $P(A \cup B) = P(A) + P(B)$.
 - b) Si les $A_i \in \mathcal{F}$ ($1 \leq i \leq n$) sont deux à deux disjoints : $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.
3. $\forall A \in \mathcal{F}, P(A^c) = 1 - P(A)$.
4. $\forall A \in \mathcal{F}, \forall B \in \mathcal{F}, A \subset B \Rightarrow P(A) \leq P(B)$.
5. $\forall A \in \mathcal{F}, \forall B \in \mathcal{F}, P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
6. *Continuité monotone séquentielle.*
 - a) Si $(B_n)_{n \geq 0}$ est une suite croissante d'événements de \mathcal{F} convergente⁶ vers $B \in \mathcal{F}$, alors $P(B) = \lim_{n \rightarrow +\infty} P(B_n)$, en notation abrégée :

$$B_n \uparrow B \Rightarrow P(B_n) \uparrow P(B) \quad (n \rightarrow +\infty).$$

5. La définition générale d'une tribu \mathcal{F} ne suppose pas que tous les singletons $\{\omega\}$ soient des éléments de \mathcal{F} . Donc un « événement élémentaire » n'est pas toujours un événement observable. Néanmoins dans la plupart des exemples que nous étudierons, la tribu possèdera les singletons.


6. Ce qui signifie : $\forall n \geq 0, B_n \subset B_{n+1}$ et $B = \bigcup_{n \geq 0} B_n$.

- b) Si $(C_n)_{n \geq 0}$ est une suite décroissante d'évènements de \mathcal{F} convergente⁷ vers $C \in \mathcal{F}$, alors $P(C) = \lim_{n \rightarrow +\infty} P(C_n)$ en notation abrégée :

$$C_n \downarrow C \Rightarrow P(C_n) \downarrow P(C) \quad (n \rightarrow +\infty).$$

7. Sous-additivité et sous- σ -additivité.

- a) $\forall A \in \mathcal{F}, \forall B \in \mathcal{F}, \quad P(A \cup B) \leq P(A) + P(B).$
 b) $\forall A_1, \dots, A_n \in \mathcal{F}, \quad P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$
 c) $\forall A_1, \dots, A_n, \dots \in \mathcal{F}, \quad P\left(\bigcup_{i \in \mathbb{N}^*} A_i\right) \leq \sum_{i=1}^{+\infty} P(A_i).$

 ICP-Prop. 1.2, pp. 7–11.

Le calcul de probabilités de réunions ou d'intersections est une question cruciale. La propriété 5 montre qu'en général on ne peut pas calculer $P(A \cup B)$ à partir de la *seule connaissance* de $P(A)$ et $P(B)$ et qu'on se heurte à la même difficulté pour $P(A \cap B)$. Le calcul des probabilités d'intersections sera discuté plus tard, à propos du conditionnement. Pour les probabilités de réunions, on peut se demander comment se généralise la propriété 5 lorsqu'on réunit plus de deux évènements. Il est facile de vérifier (faites-le!) que :

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$


Le cas général est donné par la formule de Poincaré qui exprime $P(A_1 \cup \dots \cup A_n)$ à l'aide des probabilités de *toutes les intersections* des A_i : 2 à 2, 3 à 3, etc.

Proposition 1.15 (formule de Poincaré).

Pour tout entier $n \geq 2$ et tous évènements A_1, \dots, A_n :

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) + \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \quad (1.2)$$

$$= \sum_{\emptyset \neq I \subset \llbracket 1, n \rrbracket} (-1)^{1 + \text{card } I} P\left(\bigcap_{i \in I} A_i\right) \quad (1.3)$$

 ICP-Prop. 1.3, pp. 11–13.

1.4 Exemples de probabilités

Nous examinons maintenant quelques exemples d'espaces probabilisés et de calcul de probabilités d'évènements.

Exemple 1.16. On effectue une partie de pile ou face en trois coups. Quelle est la probabilité d'obtenir pile aux premier et troisième lancers ?

On peut modéliser cette expérience en prenant $\Omega = \{\mathbf{f}, \mathbf{p}\}^3$ et pour famille d'évènements observables $\mathcal{F} = \mathcal{P}(\Omega)$ l'ensemble de toutes les parties⁸ de Ω . La pièce étant supposée symétrique, nous n'avons *a priori* pas de raison de supposer que l'un des 8 triplets de résultats possibles soit favorisé ou défavorisé par rapport aux autres. Nous

7. Ce qui signifie : $\forall n \geq 0, C_{n+1} \subset C_n$ et $C = \bigcap_{n \geq 0} C_n$.

8. Lorsque Ω est fini, il est toujours possible de faire ce choix.

choisirons donc P de sorte que tous les événements élémentaires aient même probabilité (hypothèse d'équiprobabilité), soit :

$$\forall \omega \in \Omega, \quad P(\{\omega\}) = \frac{1}{\text{Card } \Omega} = \frac{1}{2^3}.$$

L'évènement B dont on veut calculer la probabilité s'écrit : $B = \{(\text{p}, \text{f}, \text{p}); (\text{p}, \text{p}, \text{p})\}$. D'où $P(B) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$.

Exemple 1.17. On fait remplir un questionnaire comportant 20 questions binaires. Quelle est la probabilité qu'un candidat répondant au hasard obtienne au moins 16 bonnes réponses ?

On choisit ici :

$$\Omega = \{\text{oui}, \text{non}\}^{20}, \quad \mathcal{F} = \mathcal{P}(\Omega).$$

Si le candidat répond complètement au hasard, on peut considérer que chacune des 2^{20} grilles de réponses possibles a la même probabilité d'apparaître (hypothèse d'équiprobabilité sur Ω). Pour tout $B \subset \Omega$, on a alors :

$$P(B) = \frac{\text{Card } B}{\text{Card } \Omega}.$$

En particulier pour $B = \{\text{obtention d'au moins 16 bonnes réponses}\}$,

$$P(B) = \frac{C_{20}^{16} + C_{20}^{17} + C_{20}^{18} + C_{20}^{19} + C_{20}^{20}}{2^{20}} = \frac{6196}{2^{20}} \simeq 0,006.$$

Exemple 1.18 (contrôle de production). On prélève au hasard un échantillon de k pièces dans une production totale de N pièces comprenant en tout n pièces défectueuses. Le prélèvement est *sans remise* (donc $k \leq N$). Cherchons la probabilité de :

$$A_j = \{\text{il y a exactement } j \text{ pièces défectueuses dans l'échantillon}\}.$$

On prend pour Ω l'ensemble de toutes les parties à k éléments d'un ensemble à N éléments (ensemble de tous les échantillons possibles de taille k), $\mathcal{F} = \mathcal{P}(\Omega)$ et P l'équiprobabilité sur Ω . Il suffit alors de dénombrer tous les échantillons ayant exactement j pièces défectueuses. Un tel échantillon se construit en prenant j pièces dans le sous-ensemble des défectueuses (C_n^j choix possibles) et en complétant par $k - j$ pièces prises dans le sous-ensemble des non-défectueuses (C_{N-n}^{k-j} choix possibles). On en déduit :

$$P(A_j) = \frac{C_n^j C_{N-n}^{k-j}}{C_N^k} \quad \text{si} \quad \begin{cases} 0 \leq j \leq n, \\ 0 \leq j \leq k, \\ k - j \leq N - n. \end{cases}$$

Si l'une de ces trois conditions n'est pas vérifiée, $P(A_j) = 0$.

Remarque 1.19. Lorsque Ω est fini, la façon la plus simple de construire une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ est de choisir $P(\{\omega\}) = 1/\text{card } \Omega$. On parle alors d'*équiprobabilité* ou de *probabilité uniforme* sur $(\Omega, \mathcal{P}(\Omega))$. C'est la modélisation qui s'impose naturellement lorsqu'on n'a pas de raison de penser *a priori* qu'un résultat élémentaire de l'expérience soit favorisé ou défavorisé par rapport aux autres. La situation est radicalement différente lorsque Ω est infini *dénombrable*. Sur un tel ensemble, *il ne peut pas y avoir d'équiprobabilité*. Imaginons que l'on veuille tirer une boule *au hasard* dans une urne contenant

une infinité de boules numérotées de manière bijective par les entiers naturels. Soit $\{\omega_i\}$ l'évènement *tirage de la boule numérotée i* ($i \in \mathbb{N}$) et p_i sa probabilité. Par σ -additivité, les p_i vérifient nécessairement :

$$\sum_{i \in \mathbb{N}} p_i = 1.$$

Mais si les p_i sont égaux, tous les termes de la série ci-dessus valent p_0 . Sa somme est alors $+\infty$ si $p_0 > 0$ ou 0 si $p_0 = 0$, il y a donc une contradiction.

Voici une caractérisation des probabilités sur les espaces au plus dénombrables.


Proposition 1.20. *Soit $\Omega = \{\omega_i ; i \in I\}$ un ensemble au plus dénombrable. La donnée d'une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ équivaut à la donnée d'une famille $(p_i)_{i \in I}$ dans \mathbb{R}_+ telle que :*

$$\sum_{i \in I} p_i = 1$$

et des égalités

$$P(\{\omega_i\}) = p_i, \quad i \in I.$$

La probabilité P s'écrit alors $P = \sum_{i \in I} p_i \delta_{\omega_i}$, où δ_ω désigne la masse de Dirac (ou mesure de Dirac) au point ω , définie sur $\mathcal{P}(\Omega)$ par $\delta_\omega(A) = 1$ si $\omega \in A$ et $\delta_\omega(A) = 0$ si $\omega \notin A$.

 PVIR-Prop. 5.23, pp. 174–175.

Exemple 1.21 (une probabilité définie sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$).

Soit a un réel strictement positif fixé. On pose :

$$\forall k \in \mathbb{N}, \quad p_k = \frac{e^{-a} a^k}{k!}.$$

On remarque que p_k est le terme général positif d'une série convergente :

$$\sum_{k=0}^{+\infty} \frac{e^{-a} a^k}{k!} = e^{-a} \sum_{k=0}^{+\infty} \frac{a^k}{k!} = e^{-a} e^a = 1.$$

Pour tout $A \subset \mathbb{N}$, on définit :

$$P(A) = \sum_{k \in A} p_k = \sum_{k \in \mathbb{N}} p_k \delta_k(A).$$

D'après la proposition 1.20, P est une probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. On l'appelle *loi de Poisson de paramètre a* . Calculons par exemple $P(2\mathbb{N})$ où $2\mathbb{N}$ désigne l'ensemble des entiers pairs.

$$P(2\mathbb{N}) = \sum_{k \in 2\mathbb{N}} p_k = \sum_{l=0}^{+\infty} \frac{e^{-a} a^{2l}}{(2l)!} = e^{-a} \operatorname{ch} a = \frac{1}{2}(1 + e^{-2a}).$$

Une conséquence de ce résultat est : si l'on tire un nombre entier au hasard suivant une loi de Poisson, la probabilité qu'il soit pair est strictement supérieure à $\frac{1}{2}$.

Exemple 1.22 (loi uniforme sur un segment). Prenons $\Omega = \mathbb{R}$, $\mathcal{F} = \operatorname{Bor}(\mathbb{R})$ et rappelons que λ_1 désigne la mesure de Lebesgue sur \mathbb{R} (cf. exemple 1.11). Soit $[a, b]$ un segment fixé de \mathbb{R} . On définit une probabilité P sur $(\mathbb{R}, \operatorname{Bor}(\mathbb{R}))$ en posant :

$$\forall A \in \operatorname{Bor}(\mathbb{R}), \quad P(A) = \frac{\lambda_1(A \cap [a, b])}{\lambda_1([a, b])} = \frac{\lambda_1(A \cap [a, b])}{b - a}. \quad (1.4)$$

D'après l'exemple 1.10, P est une mesure et comme $P(\mathbb{R}) = 1$, c'est une probabilité. On l'appelle *loi uniforme sur $[a, b]$* . Remarquons que pour cette probabilité, tout singleton est de probabilité nulle ($\forall x \in \mathbb{R}, P(\{x\}) = 0$), ce qui résulte de la propriété analogue de λ_1 . On voit sur cet exemple que la probabilité d'une union infinie *non dénombrable* d'événements deux à deux disjoints n'est pas forcément égale à la somme de la famille correspondante de probabilités d'événements. En effet,

$$1 = P([a, b]) = P\left(\bigcup_{x \in [a, b]} \{x\}\right) \neq \sum_{x \in [a, b]} P(\{x\}) = 0.$$

Exemple 1.23 (lois uniformes dans \mathbb{R}^d). On peut généraliser l'exemple précédent en prenant au lieu d'un segment un borélien B de \mathbb{R}^d , i.e. $B \in \text{Bor}(\mathbb{R}^d)$, tel que $0 < \lambda_d(B) < +\infty$. On définit alors une probabilité P sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$, en posant :

$$\forall A \in \text{Bor}(\mathbb{R}^d), \quad P(A) = \frac{\lambda_d(A \cap B)}{\lambda_d(B)}. \quad (1.5)$$

Cette probabilité est appelée *loi uniforme sur B* .

Nous allons donner maintenant une caractérisation de *toutes les probabilités P* sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$. Comme le montre l'exemple de la loi uniforme sur un segment, il est illusoire d'espérer caractériser ces probabilités par les $P(\{x\})$. La situation est donc radicalement différente du cas d'un espace Ω au plus dénombrable. Au lieu des $P(\{x\})$, nous allons utiliser les $P([a, b])$, ou les $P([-\infty, x])$. Nous aurons besoin pour cela de la notion de fonction de répartition.

Définition 1.24 (fonction de répartition). *Soit P une probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$. On appelle fonction de répartition de P , l'application*

$$F : \mathbb{R} \rightarrow [0, 1], \quad x \mapsto F(x) := P([-\infty, x]).$$

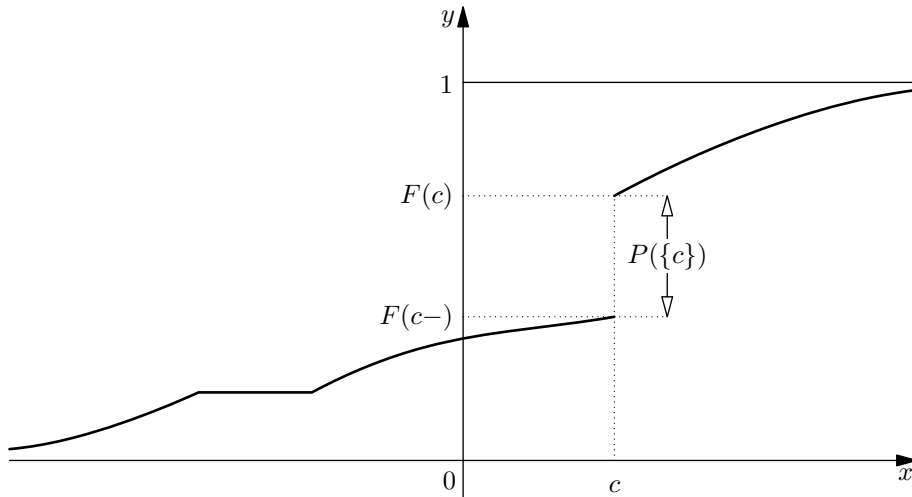


FIGURE 1.1 – Une fonction de répartition avec une discontinuité au point $x = c$

Voici les propriétés générales des fonctions de répartition.


Proposition 1.25. *La fonction de répartition F d'une probabilité P sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ a les propriétés suivantes.*

- a) F est croissante sur \mathbb{R} .
- b) F a pour limite 0 en $-\infty$ et 1 en $+\infty$.
- c) F est continue à droite sur \mathbb{R} et a une limite à gauche en tout point $x \in \mathbb{R}$.
- d) En notant $F(x-) := \lim_{\varepsilon \downarrow 0} F(x - \varepsilon)$ la limite à gauche⁹ au point x , on a

$$\forall x \in \mathbb{R}, \quad P(\{x\}) = F(x) - F(x-). \quad (1.6)$$

De plus l'ensemble des $x \in \mathbb{R}$ tels que $F(x) \neq F(x-)$ est au plus dénombrable.

- e) Si deux probabilités P_1 et P_2 sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ ont même fonction de répartition F , elles sont égales, autrement dit $P_1(B) = P_2(B)$ pour tout $B \in \text{Bor}(\mathbb{R})$.


 Pour a) – d), PVIR Prop. 1.28, pp. 176–178. Pour e), IFP, Th. 1.34 et Cor. 1.3.5.

Le théorème suivant permet de construire toutes les probabilités sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$.

Théorème 1.26. Soit F une fonction croissante et continue à droite sur \mathbb{R} , ayant pour limite 0 en $-\infty$ et 1 en $+\infty$. Il existe une unique probabilité P sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ telle que

$$\forall a, b \in \mathbb{R}, \text{ avec } a \leq b, \quad P([a, b]) = F(b) - F(a).$$

Alors F est la fonction de répartition de P .

 IFP Th. 1.36., pp. 45–47.

1.5 Probabilités conditionnelles

1.5.1 Introduction

Comment doit-on modifier la probabilité que l'on attribue à un évènement lorsque l'on dispose d'une *information supplémentaire*? Le concept de probabilité conditionnelle permet de répondre à cette question.

Définition 1.27 (probabilité conditionnelle). Soit H un évènement de probabilité non nulle. Pour tout évènement A , on définit :

$$P(A \mid H) = \frac{P(A \cap H)}{P(H)},$$

appelée probabilité conditionnelle de l'évènement A sous l'hypothèse H .

Remarquons que pour l'instant, il ne s'agit que d'un jeu d'écriture. On a simplement défini un réel $P(A \mid H)$ pour que :

$$P(A \cap H) = P(A \mid H)P(H).$$

Ce qui fait l'intérêt du concept de probabilité conditionnelle, c'est qu'il est souvent bien plus facile d'attribuer *directement* une valeur à $P(A \mid H)$ en tenant compte des conditions expérimentales (liées à l'*information* H) et d'en déduire ensuite la valeur de $P(A \cap H)$. Le raisonnement implicite alors utilisé est : tout espace probabilisé modélisant correctement la réalité expérimentale devrait fournir cette valeur pour $P(A \mid H)$.

9. Cette notation est un peu abusive, puisqu'il ne s'agit pas forcément d'une valeur prise par la fonction F . Attention à ne pas confondre le « $x-$ » dans $F(x-)$ avec le « x^- » partie négative de x .

Exemple 1.28. Une urne contient r boules rouges et v boules vertes. On en tire deux l'une après l'autre, sans remise. Quelle est la probabilité d'obtenir deux rouges ?

Notons H et A les évènements :

$$H = \{\text{rouge au 1}^{\text{er}} \text{ tirage}\}, \quad A = \{\text{rouge au 2}^{\text{e}} \text{ tirage}\}.$$

Un espace probabilisé (Ω, \mathcal{F}, P) modélisant correctement cette expérience devrait vérifier :

$$P(H) = \frac{r}{r+v}, \quad P(A | H) = \frac{r-1}{r+v-1}.$$

En effet, si H est réalisé, le deuxième tirage a lieu dans une urne contenant $r+v-1$ boules dont $r-1$ rouges. On en déduit :

$$P(\text{deux rouges}) = P(A \cap H) = P(A | H)P(H) = \frac{r-1}{r+v-1} \times \frac{r}{r+v}.$$

On aurait pu arriver au même résultat en prenant pour Ω l'ensemble des arrangements de deux boules parmi $r+v$, muni de l'équiprobabilité et en faisant du dénombrement :

$$\text{card } \Omega = A_{r+v}^2 = (r+v)(r+v-1), \quad \text{card}(A \cap H) = A_r^2 = r(r-1).$$

d'où :

$$P(A \cap H) = \frac{r(r-1)}{(r+v)(r+v-1)}.$$

Notons d'ailleurs que $\text{card } H = r(r+v-1)$ d'où

$$P(H) = \frac{\text{card } H}{\text{card } \Omega} = \frac{r(r+v-1)}{(r+v)(r+v-1)} = \frac{r}{r+v}.$$

En appliquant la définition formelle de $P(A | H)$ on retrouve :

$$P(A | H) = \frac{P(A \cap H)}{P(H)} = \frac{r(r-1)}{(r+v)(r+v-1)} \times \frac{r+v}{r} = \frac{r-1}{r+v-1},$$

ce qui est bien la valeur que nous avons attribuée *a priori* en analysant les conditions expérimentales.

Remarque 1.29. Il importe de bien comprendre que l'écriture « $A | H$ » ne désigne pas un nouvel évènement¹⁰ différent de A . Quand on écrit $P(A | H)$, ce que l'on a modifié, ce n'est pas l'évènement A , mais la valeur numérique qui lui était attribuée par la fonction d'ensembles P . Il serait donc en fait plus correct d'écrire $P_H(A)$ que $P(A | H)$. On conservera néanmoins cette dernière notation essentiellement pour des raisons typographiques : $P(A | H_1 \cap H_2 \cap H_3)$ est plus lisible que $P_{H_1 \cap H_2 \cap H_3}(A)$.

1.5.2 Propriétés

Proposition 1.30. Soit (Ω, \mathcal{F}, P) un espace probabilisé et H un évènement fixé tel que $P(H) \neq 0$. Alors la fonction d'ensembles $P(. | H)$ définie par :

$$P(. | H) : \mathcal{F} \rightarrow [0, 1] \quad B \mapsto P(B | H)$$

est une nouvelle probabilité sur (Ω, \mathcal{F}) .

¹⁰. En fait cette écriture prise isolément (sans le P) n'a pas de sens et ne devrait jamais être utilisée. Le symbole $|$ ne représente pas une opération sur les évènements qui l'entourent.

Une conséquence immédiate est que la fonction d'ensembles $P(\cdot \mid H)$ vérifie toutes les propriétés de la proposition 1.14.

Corollaire 1.31. *La fonction d'ensembles $P(\cdot \mid H)$ vérifie :*

1. $P(\emptyset \mid H) = 0$, $P(\Omega \mid H) = 1$ et si $A \supset H$, $P(A \mid H) = 1$.
2. Si les A_i sont deux à deux disjoints :

$$P\left(\bigcup_{i=1}^n A_i \mid H\right) = \sum_{i=1}^n P(A_i \mid H).$$

3. Pour tout $B \in \mathcal{F}$, $P(B^c \mid H) = 1 - P(B \mid H)$.
4. Pour tous $A \in \mathcal{F}$ et $B \in \mathcal{F}$, si $A \subset B$, $P(A \mid H) \leq P(B \mid H)$.
5. Pour tous $A \in \mathcal{F}$ et $B \in \mathcal{F}$,

$$P(A \cup B \mid H) = P(A \mid H) + P(B \mid H) - P(A \cap B \mid H).$$

6. Pour toute suite $(A_i)_{i \geq 1}$ d'évènements :

$$P\left(\bigcup_{i \in \mathbb{N}^*} A_i \mid H\right) \leq \sum_{i=1}^{+\infty} P(A_i \mid H).$$

7. Si $B_n \uparrow B$, $P(B_n \mid H) \uparrow P(B \mid H)$, $(n \rightarrow +\infty)$.
8. Si $C_n \downarrow C$, $P(C_n \mid H) \downarrow P(C \mid H)$, $(n \rightarrow +\infty)$.

Nous n'avons vu jusqu'ici aucune formule permettant de calculer la probabilité d'une intersection d'évènements à l'aide des probabilités de ces évènements. Une telle formule *n'existe pas dans le cas général*. Les probabilités conditionnelles fournissent une méthode générale tout à fait naturelle pour calculer une probabilité d'intersection.

Proposition 1.32 (règle des conditionnements successifs).

Si A_1, \dots, A_n sont n évènements tels que $P(A_1 \cap \dots \cap A_{n-1}) \neq 0$, on a :

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) &= P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \times \dots \\ &\quad \dots \times P(A_n \mid A_1 \cap \dots \cap A_{n-1}). \end{aligned}$$

Les probabilités conditionnelles permettent aussi de calculer la probabilité d'un évènement en conditionnant par *tous les cas possibles*. Du point de vue ensembliste, ces *cas possibles* correspondent à une *partition* de Ω .

Définition 1.33 (partition). *On dit qu'une famille $(H_i)_{i \in I}$ de parties de Ω est une partition de Ω si elle vérifie les trois conditions :*

- $\forall i \in I$, $H_i \neq \emptyset$.
- $\Omega = \bigcup_{i \in I} H_i$.
- Les H_i sont deux à deux disjoints ($i \neq j \Rightarrow H_i \cap H_j = \emptyset$).

Proposition 1.34 (conditionnement par les cas possibles¹¹).

- (i) Si H est tel que $P(H) \neq 0$ et $P(H^c) \neq 0$, on a

$$\forall A \in \mathcal{F}, \quad P(A) = P(A \mid H)P(H) + P(A \mid H^c)P(H^c).$$

11. Ou formule des probabilités totales.

(ii) Si H_1, \dots, H_n est une partition finie de Ω en évènements de probabilité non nulle,

$$\forall A \in \mathcal{F}, \quad P(A) = \sum_{i=1}^n P(A \mid H_i)P(H_i).$$

(iii) Si $(H_i)_{i \in \mathbb{N}}$ est une partition de Ω telle qu'aucun $P(H_i)$ ne soit nul,

$$\forall A \in \mathcal{F}, \quad P(A) = \sum_{i=0}^{+\infty} P(A \mid H_i)P(H_i).$$

Lorsqu'on a une partition de Ω en n hypothèses ou cas possibles H_i et que l'on sait calculer les $P(H_i)$ et les $P(A \mid H_i)$, on peut se poser le problème inverse : calculer $P(H_j \mid A)$ à l'aide des quantités précédentes. La solution est donnée par la formule suivante quelquefois appelée (abusivement) formule des probabilités des causes.

Proposition 1.35 (formule de Bayes).

Soit A un évènement de probabilité non nulle. Si les évènements H_i ($1 \leq i \leq n$) forment une partition de Ω et si aucun $P(H_i)$ n'est nul, on a pour tout $j \in \llbracket 1, n \rrbracket$:

$$P(H_j \mid A) = \frac{P(A \mid H_j)P(H_j)}{\sum_{i=1}^n P(A \mid H_i)P(H_i)}.$$

La même formule se généralise au cas d'une partition dénombrable. Ces formules sont plus faciles à retrouver qu'à mémoriser.



Pour toute cette sous-section, cf. ICP pp. 32–36.

Exemple 1.36. Revenons sur le problème de lancers répétés de deux dés en lecture suggérée à la page 5, à savoir l'attribution d'une probabilité à l'évènement E première obtention de la somme 9 avant celle de la somme 7. Les probabilités conditionnelles permettent d'en proposer une solution simple, sans calcul de série. En rappelant que F_1 , G_1 , H_1 désignent respectivement l'évènement *obtention au premier lancer d'une somme 9* (resp. 7, resp. ni 7 ni 9), la formule des probabilités totales nous donne :

$$P(E) = P(E \mid F_1)P(F_1) + P(E \mid G_1)P(G_1) + P(E \mid H_1)P(H_1). \quad (1.7)$$

On a clairement $P(E \mid F_1) = 1$ et $P(E \mid G_1) = 0$. Pour attribuer une valeur à $P(E \mid H_1)$, on considère que l'obtention d'une somme autre que 7 ou 9 au premier lancer ne devrait pas influencer sur l'apparition ultérieure du 7 ou du 9. Pour traduire cette idée, on pose $P(E \mid H_1) = P(E)$. En reportant ces valeurs dans (1.7), il vient

$$P(E) = P(F_1) + P(E)P(H_1) = \frac{1}{9} + \frac{13}{18}P(E), \quad \text{d'où} \quad \left(1 - \frac{13}{18}\right)P(E) = \frac{1}{9},$$

ce qui se résout en

$$P(E) = \frac{1}{9} \times \frac{18}{5} = \frac{2}{5}.$$

On retrouve ainsi la valeur obtenue dans PVIR p. 165, lorsque nous avons esquissé la construction d'un espace (Ω, \mathcal{F}, P) modélisant cette expérience aléatoire. Le point clé dans le raisonnement ci-dessus est l'égalité $P(E \mid H_1) = P(E)$, qui exprime l'indépendance des évènements H_1 et E , une notion que nous allons étudier maintenant.

1.6 Indépendance

1.6.1 Indépendance de deux évènements

Soient A et B deux évènements de probabilité non nulle. Il arrive que la connaissance de la réalisation de A ne modifie pas notre information sur celle de B , autrement dit que $P(B | A) = P(B)$. C'est le cas par exemple lorsque l'on fait deux tirages avec remise et que la réalisation de A ne dépend que du résultat du premier tirage, celle de B que du deuxième. Symétriquement on aura dans cet exemple $P(A | B) = P(A)$. Cette remarque se généralise :

Proposition 1.37. *Si A et B sont des évènements de probabilité non nulle, les trois égalités suivantes sont équivalentes :*

- (i) $P(B | A) = P(B)$,
- (ii) $P(A | B) = P(A)$,
- (iii) $P(A \cap B) = P(A)P(B)$.

D'autre part la relation (iii) est toujours vérifiée dans le cas dégénéré où $P(A) = 0$ ou $P(B) = 0$. En effet, on a alors à la fois $P(A)P(B) = 0$ et $0 \leq P(A \cap B) \leq \min(P(A), P(B)) = 0$ d'où $P(A \cap B) = 0$. Ainsi la relation (iii) est un peu plus générale que (i) et (ii). Elle a aussi sur les deux autres l'avantage de la symétrie d'écriture. C'est elle que l'on retient pour définir l'indépendance.

Définition 1.38. *Soit (Ω, \mathcal{F}, P) un espace probabilisé. Deux évènements A et B de cet espace sont dits indépendants lorsque :*

$$P(A \cap B) = P(A)P(B).$$

Exemple 1.39. On jette deux fois le même dé. Les évènements

$$\begin{aligned} A &= \{\text{obtention d'un chiffre pair au premier lancer}\}, \\ B &= \{\text{obtention du 1 au deuxième lancer}\}, \end{aligned}$$

sont indépendants.

En effet, en prenant $\Omega = \llbracket 1, 6 \rrbracket^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ et P l'équiprobabilité, on vérifie que :

$$\begin{aligned} P(A) &= \frac{3 \times 6}{36} = \frac{1}{2}, & P(B) &= \frac{6 \times 1}{36} = \frac{1}{6}, \\ P(A \cap B) &= \frac{3 \times 1}{36} = \frac{1}{12}, & P(A)P(B) &= \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}. \end{aligned}$$

Remarques 1.40.

- Si A est un évènement tel que $P(A) = 0$ ou $P(A) = 1$, alors il est indépendant de tout évènement, *y compris de lui-même* (c'est le cas en particulier pour Ω et \emptyset).
- Deux évènements *incompatibles* A et B avec $P(A) > 0$ et $P(B) > 0$ ne sont *jamais indépendants*. En effet $A \cap B = \emptyset$ implique $P(A \cap B) = 0$ or $P(A)P(B) \neq 0$.
- L'indépendance de deux évènements A et B n'est pas une propriété intrinsèque aux évènements, elle est toujours relative au modèle (Ω, \mathcal{F}, P) que l'on a choisi.



Pour la dernière remarque, étudier l'exemple 2.6 p. 38 dans ICP.

Proposition 1.41. *Si A et B sont indépendants, il en est de même pour les paires d'évènements A et B^c , A^c et B , A^c et B^c .*

1.6.2 Indépendance mutuelle

On se propose de généraliser la notion d'indépendance à plus de deux évènements. Examinons d'abord la situation suivante.

Exemple 1.42. Une urne contient quatre jetons : un bleu, un blanc, un rouge et un bleu-blanc-rouge. On en tire un au hasard. Considérons les trois évènements

$$\begin{aligned} A &= \{\text{le jeton tiré contient du bleu}\}, \\ B &= \{\text{le jeton tiré contient du blanc}\}, \\ C &= \{\text{le jeton tiré contient du rouge}\}. \end{aligned}$$

Il est clair que $P(A) = P(B) = P(C) = 2/4 = 1/2$. D'autre part :

$$P(A \cap B) = P(\text{tricolore}) = \frac{1}{4} = P(A)P(B)$$

et de même $P(B \cap C) = 1/4 = P(B)P(C)$, $P(C \cap A) = 1/4 = P(C)P(A)$. Ainsi les évènements A, B, C sont *deux à deux indépendants*.

D'autre part $P(A | B \cap C) = 1$ car $B \cap C = \{\text{tricolore}\}$. Donc la connaissance de la réalisation *simultanée* de B et C modifie notre information sur A . La notion d'indépendance deux à deux n'est donc pas suffisante pour traduire l'idée intuitive d'indépendance de plusieurs évènements. Ceci motive la définition suivante.

Définition 1.43. *Trois évènements A, B, C sont dits mutuellement indépendants lorsqu'ils vérifient les quatre conditions :*

$$\begin{aligned} P(A \cap B) &= P(A)P(B), \\ P(B \cap C) &= P(B)P(C), \\ P(C \cap A) &= P(C)P(A), \\ P(A \cap B \cap C) &= P(A)P(B)P(C). \end{aligned}$$

Avec cette définition de l'indépendance des évènements A, B et C on a bien¹² $P(A | B) = P(A)$, $P(A | B \cap C) = P(A)$, ainsi que toutes les égalités qui s'en déduisent par permutation sur les lettres A, B, C . On peut généraliser cette définition comme suit.

Définition 1.44. *Les n évènements A_1, \dots, A_n sont dits mutuellement indépendants si pour toute sous-famille A_{i_1}, \dots, A_{i_k} avec $1 \leq i_1 < \dots < i_k \leq n$,*

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \times \dots \times P(A_{i_k}). \quad (1.8)$$

L'indépendance mutuelle implique évidemment l'indépendance deux à deux et la réciproque est fausse comme le montre l'exemple 1.42. Dans toute la suite, lorsque nous parlerons d'une famille de plusieurs évènements indépendants sans autre précision, nous sous-entendrons systématiquement *mutuellement* indépendants.

Proposition 1.45. *Si $\{A_1, \dots, A_n\}$ est une famille de n évènements indépendants, toute famille obtenue en remplaçant certains des A_i par leur complémentaire est encore indépendante.*

12. Lorsque les probabilités conditionnelles existent.

Définition 1.46 (indépendance d'une suite d'évènements). *Une suite infinie d'évènements est dite indépendante si toute sous-suite finie est formée d'évènements mutuellement indépendants.*

Remarque 1.47. Compte-tenu de la proposition 1.45, on voit immédiatement que si $(A_i)_{i \in \mathbb{N}}$ est une suite indépendante d'évènements, toute suite formée en remplaçant certains des A_i (éventuellement tous) par leur complémentaire est encore indépendante.

1.6.3 Épreuves répétées

Considérons une suite d'épreuves réalisées dans les mêmes conditions expérimentales, par exemple tirages avec remise dans la même urne, lancers successifs d'un dé, etc. Il est alors raisonnable de supposer que les résultats de tout sous-ensemble fini d'épreuves n'ont aucune influence sur ceux des autres épreuves.

Définition 1.48. *On dit que les épreuves sont indépendantes si toute suite $(A_i)_{i \geq 1}$ telle que la réalisation de chaque A_i est déterminée uniquement par le résultat de la i^e épreuve est une suite indépendante d'évènements.*

Exemple 1.49. On réalise une suite d'épreuves indépendantes. Chaque épreuve résulte en un succès avec probabilité $p \in]0, 1[$ ou en un échec avec probabilité $q = 1 - p$. Quelle est la probabilité des évènements suivants ?

- a) $A = \{\text{Au moins un succès au cours des } n \text{ premières épreuves}\}.$
- b) $B = \{\text{Exactement } k \text{ succès au cours des } n \text{ premières épreuves}\}.$
- c) $C = \{\text{Toutes les épreuves donnent un succès}\}.$

Notons pour tout $i \geq 1$: $R_i = \{\text{succès à la } i^e \text{ épreuve}\}$, R_i^c est alors l'échec à la i^e épreuve.

- a) $A = \bigcup_{i=1}^n R_i$, d'où $A^c = \bigcap_{i=1}^n R_i^c$. Les R_i^c étant indépendants, on en déduit :

$$P(A^c) = \prod_{i=1}^n P(R_i^c) = (1 - p)^n = q^n,$$

d'où $P(A) = 1 - q^n$.

- b) Traitons d'abord le cas $0 < k < n$. L'évènement B est la réunion disjointe de tous les évènements du type :

$$B_I = \left(\bigcap_{i \in I} R_i \right) \cap \left(\bigcap_{j \in J} R_j^c \right),$$

où I est une partie de cardinal k de $\llbracket 1, n \rrbracket$ et J son complémentaire dans $\llbracket 1, n \rrbracket$. L'ensemble d'indices I représente un choix possible des k épreuves donnant un succès, les autres épreuves indexées par J donnant alors un échec. En considérant tous les choix possibles de l'ensemble I (il y en a C_n^k), on obtient une partition de B par les B_I . Par indépendance des épreuves, pour tout I on a :

$$P(B_I) = \prod_{i \in I} P(R_i) \times \prod_{j \in J} P(R_j^c) = p^k q^{n-k}.$$

On voit ainsi que $P(B_I)$ ne dépend pas de I . On en déduit :

$$P(B) = \sum_{\substack{I \subset \llbracket 1, n \rrbracket \\ \text{card } I = k}} P(B_I) = C_n^k p^k q^{n-k}.$$

La vérification de la validité de la formule $P(B) = C_n^k p^k q^{n-k}$ dans les cas $k = 0$ et $k = n$ est laissée au lecteur.

c) Pour $n \geq 1$, soit $C_n = \{\text{succès aux } n \text{ premières épreuves}\}$. Clairement C est inclus dans C_n donc $0 \leq P(C) \leq P(C_n)$. En utilisant l'indépendance des R_i on obtient :

$$P(C_n) = P\left(\bigcap_{i=1}^n R_i\right) = \prod_{i=1}^n P(R_i) = p^n.$$

donc pour tout $n \geq 1$, $0 \leq P(C) \leq p^n$. En faisant tendre n vers $+\infty$, on en déduit $P(C) = 0$.

1.6.4 L'algorithme du rejet

Nous allons voir maintenant comment utiliser des épreuves répétées pour « fabriquer » des probabilités conditionnelles. La méthode exposée ci-dessous est le coeur de l'algorithme du rejet utilisé pour la simulation de variables ou de vecteurs aléatoires. Pour en comprendre le principe, il n'est pas nécessaire de connaître en détail ces notions qui seront étudiées dans les chapitres suivants.

Considérons une suite d'épreuves répétées où la i^{e} épreuve consiste à générer un point aléatoire M_i du plan¹³. Formellement, on a une suite $(M_i)_{i \geq 1}$ de points aléatoires telle que pour tout borélien A , les événements $\{M_i \in A\} = \{\omega \in \Omega; M_i(\omega) \in A\}$ sont indépendants et de même probabilité ne dépendant que de A . On fixe un borélien B de \mathbb{R}^2 tel que $P(M_1 \in B) > 0$ et on définit l'indice $T(\omega)$ comme le premier indice i tel que $M_i(\omega)$ soit dans B , voir figure 1.2. Si $M_i(\omega)$ n'appartient à B pour aucun $i \in \mathbb{N}^*$, on pose $T(\omega) = +\infty$. On note M_T le point aléatoire ainsi obtenu (c'est le premier M_i à tomber dans B , mais attention son numéro est lui même aléatoire et peut changer avec ω), en prenant par exemple $M_T(\omega) = (0, 0)$ dans le cas particulier où $T(\omega) = +\infty$. On se propose de calculer $P(M_T \in A)$ pour A borélien quelconque de \mathbb{R}^2 .

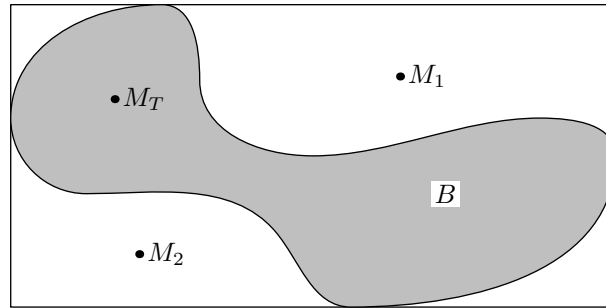


FIGURE 1.2 – Simulation par rejet, ici $T(\omega) = 3$

Nous supposons implicitement ici que les écritures $\{M_i \in A\}$, $\{M_T \in A\}$, $\{T = k\}$ représentent effectivement des événements¹⁴.

Comme les événements $\{T = k\}$, $k \in \mathbb{N}^*$ réalisent une partition dénombrable de Ω , on a la décomposition en union dénombrable disjointe :

$$\forall A \in \text{Bor}(\mathbb{R}^2), \quad \{M_T \in A\} = \bigcup_{k \in \mathbb{N}^*} \{M_T \in A \text{ et } T = k\}. \quad (1.9)$$

13. Tout ce qui suit est valable aussi avec des points aléatoires de \mathbb{R}^d au lieu de \mathbb{R}^2 .

14. On peut justifier ces suppositions avec la notion de mesurabilité qui sera vue lors de l'étude des variables aléatoires ou des vecteurs aléatoires.

Pour tout $k \in \mathbb{N}^*$, l'évènement $\{M_T \in A \text{ et } T = k\}$ peut s'écrire comme suit :

$$\{M_T \in A \text{ et } T = k\} = \{M_k \in A \text{ et } T = k\} = \underbrace{\bigcap_{i < k} \{M_i \notin B\} \cap \{M_k \in B\}}_{\{T=k\}} \cap \{M_k \in A\}$$

Ceci nous permet d'exploiter l'indépendance des épreuves répétées pour obtenir :

$$P(M_T \in A \text{ et } T = k) = (1 - P(M_1 \in B))^{k-1} P(M_k \in A \cap B).$$

En notant $p = P(M_1 \in B)$, $q = 1 - p$ et en appliquant cette formule avec $A = \mathbb{R}^2$, on obtient :

$$P(M_T \in \mathbb{R}^2 \text{ et } T = k) = P(T = k) = q^{k-1} p.$$

En rappelant que p est strictement positif, d'où $0 < q < 1$ et en sommant pour tous les $k \in \mathbb{N}^*$, il vient :

$$P(T \in \mathbb{N}^*) = \sum_{k \in \mathbb{N}^*} P(T = k) = \sum_{k \in \mathbb{N}^*} q^{k-1} p = \frac{p}{1 - q} = 1.$$

Par conséquent, $P(T = +\infty)$ est nulle, tout comme $P(M_T \in A \text{ et } T = +\infty)$ qu'elle majore. Donc pour calculer $P(M_T \in A)$ par σ -additivité à partir de la décomposition (1.9), on ne perd rien en laissant tomber $P(M_T \in A \text{ et } T = +\infty)$. Ainsi :

$$\begin{aligned} P(M_T \in A) &= \sum_{k \in \mathbb{N}^*} P(M_T \in A \text{ et } T = k) = \sum_{k \in \mathbb{N}^*} q^{k-1} P(M_1 \in A \cap B) \\ &= \frac{P(M_1 \in A \cap B)}{1 - q} \\ &= \frac{P(M_1 \in A \cap B)}{P(M_1 \in B)}. \end{aligned}$$

Finalement, pour tout ensemble borélien A du plan \mathbb{R}^2 :

$$P(M_T \in A) = P(M_1 \in A \mid M_1 \in B).$$

Voici une première application. Supposons que les M_i soient choisis au hasard suivant la loi uniforme sur un rectangle C (c'est très facile à simuler informatiquement) contenant B , comme celui de la figure 1.2, autrement dit que pour tout entier i ,

$$P(M_i \in A) = \frac{\lambda_2(A \cap C)}{\lambda_2(C)}.$$

Alors pour tout borélien A ,

$$P(M_T \in A) = \frac{\frac{\lambda_2(A \cap B \cap C)}{\lambda_2(C)}}{\frac{\lambda_2(B \cap C)}{\lambda_2(C)}} = \frac{\lambda_2(A \cap B)}{\lambda_2(B)},$$

ce qui montre que le point aléatoire M_T suit la loi uniforme sur B .

Chapitre 2

Variables aléatoires

D'un point de vue formel, une variable aléatoire X sur un espace probabilisable (Ω, \mathcal{F}) est simplement une application mesurable $X : \Omega \rightarrow \mathbb{R}$. La mesurabilité de X permet lorsque (Ω, \mathcal{F}) est muni d'une mesure de probabilité P de définir la loi de X comme la mesure image $P \circ X^{-1}$. Cette mesure est une nouvelle probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$. L'intérêt de cette notion est qu'elle permet dans de nombreux problèmes de remplacer l'espace plus ou moins abstrait et compliqué Ω par \mathbb{R} . On revoit d'abord ces notions de mesurabilité et de mesure image.

2.1 Mesurabilité

Définition 2.1 (mesurabilité). Soit $h : \Omega_1 \rightarrow \Omega_2$, où Ω_1 et Ω_2 sont munis respectivement des tribus \mathcal{F}_1 et \mathcal{F}_2 . On dit que h est mesurable¹ \mathcal{F}_1 - \mathcal{F}_2 si pour tout $B \in \mathcal{F}_2$, $h^{-1}(B) \in \mathcal{F}_1$.

Définition 2.2 (application borélienne). Soit $h : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$ et $\mathcal{F}_i = \text{Bor}(\mathbb{R}^{d_i})$ pour $i = 1, 2$. On dit que h est une application borélienne si elle est mesurable \mathcal{F}_1 - \mathcal{F}_2 .

On démontre en théorie de la mesure le résultat suivant.

Proposition 2.3. Soient Ω_1 et Ω_2 deux ensembles munis respectivement des tribus \mathcal{F}_1 et \mathcal{F}_2 et \mathcal{C} une famille de parties de Ω_2 engendrant \mathcal{F}_2 ($\sigma(\mathcal{C}) = \mathcal{F}_2$). L'application $h : \Omega_1 \rightarrow \Omega_2$ est mesurable \mathcal{F}_1 - \mathcal{F}_2 si pour tout $B \in \mathcal{C}$, $h^{-1}(B) \in \mathcal{F}_1$.

En particulier en prenant $\Omega_1 = \Omega$, $\mathcal{F}_1 = \mathcal{F}$, $\Omega_2 = \mathbb{R}$, $\mathcal{F}_2 = \text{Bor}(\mathbb{R})$ et \mathcal{C} la famille des intervalles $]a, b]$, on voit que pour que $h : \Omega \rightarrow \mathbb{R}$ soit mesurable \mathcal{F} - $\text{Bor}(\mathbb{R})$, il suffit que

$$\forall a, b \in \mathbb{R}, \quad h^{-1}(]a, b]) \in \mathcal{F}.$$

La mesurabilité \mathcal{F} - $\text{Bor}(\mathbb{R})$ est conservée par toutes les opérations usuelles de l'algèbre (somme, produit, combinaisons linéaires, composition, ...) et de l'analyse, pourvu que les familles d'applications concernées soient au plus dénombrables (inf, sup, lim inf, lim sup, limite si elle existe d'une suite de fonctions, série de fonctions).



IFP, pp. 61–68.

1. Ce langage est trompeur : la mesurabilité ne fait intervenir aucune mesure, elle concerne seulement h et les tribus.

2.2 Généralités sur les variables aléatoires

2.2.1 Variables aléatoires et lois

Définition 2.4 (variable aléatoire réelle). *Soit (Ω, \mathcal{F}, P) un espace probabilisé. On appelle variable aléatoire réelle sur (Ω, \mathcal{F}) , ou plus simplement variable aléatoire, toute application X :*

$$X : \Omega \longrightarrow \mathbb{R} \quad \omega \mapsto X(\omega),$$

mesurable \mathcal{F} - $\text{Bor}(\mathbb{R})$, c'est-à-dire vérifiant :

$$\forall B \in \text{Bor}(\mathbb{R}), \quad X^{-1}(B) \in \mathcal{F}.$$

En raison de la proposition 2.3, il suffit que la condition ci-dessus soit vérifiée pour $B =]a, b]$ avec a et b réels quelconques, pour que X soit une variable aléatoire.

Remarque 2.5. Il importe de noter que la mesure de probabilité P sur (Ω, \mathcal{F}) ne joue *aucun rôle* dans la définition de la notion de variable aléatoire. C'est pour cela que nous parlons de « variable aléatoire sur (Ω, \mathcal{F}) » plutôt que sur (Ω, \mathcal{F}, P) .

Définition 2.6 (variable aléatoire discrète). *On appelle variable aléatoire discrète sur (Ω, \mathcal{F}) , toute application $X : \Omega \rightarrow \mathbb{R}$ vérifiant les deux conditions suivantes.*

- (i) *L'ensemble des images $X(\Omega) = \{X(\omega), \omega \in \Omega\}$ est une partie au plus dénombrable de \mathbb{R} . On peut donc numéroter ses éléments par des indices entiers²*

$$X(\Omega) = \{x_0, x_1, \dots, x_k, \dots\}.$$

- (ii) *X est mesurable \mathcal{F} - $\mathcal{P}(\mathbb{R})$, ce qui équivaut ici à*

$$\forall x \in X(\Omega), \quad X^{-1}(\{x\}) \in \mathcal{F}. \quad (2.1)$$

Remarquons que si X est mesurable \mathcal{F} - $\mathcal{P}(\mathbb{R})$, elle est *a fortiori* mesurable \mathcal{F} - $\text{Bor}(\mathbb{R})$ puisque la condition $X^{-1}(B) \in \mathcal{F}$ doit être satisfaite dans ce cas seulement pour les B appartenant à la tribu $\text{Bor}(\mathbb{R})$ qui est une sous-famille de $\mathcal{P}(\mathbb{R})$. Par conséquent, toute variable aléatoire discrète est aussi une variable aléatoire réelle.

Proposition 2.7. *Soit X une variable aléatoire sur (Ω, \mathcal{F}) et P une probabilité sur (Ω, \mathcal{F}) . La fonction d'ensembles $P_X = P \circ X^{-1}$ définie sur $\text{Bor}(\mathbb{R})$ par*

$$\forall B \in \text{Bor}(\mathbb{R}), \quad P_X(B) := P(X^{-1}(B)) = P(X \in B) \quad (2.2)$$

est une probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$.

 IFP, prop. 2.25, p. 74.

Définition 2.8 (loi d'une variable aléatoire). *Soient (Ω, \mathcal{F}, P) un espace probabilisé et X une variable aléatoire sur (Ω, \mathcal{F}) . On appelle loi de X sous P , ou plus simplement loi de X , la probabilité P_X sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ définie par (2.2).*

2. Pour tous les exemples classiques que nous rencontrerons, il est possible de les numéroter de manière croissante : $x_0 < x_1 < x_2 \dots$. Mais ce n'est pas toujours le cas, car $X(\Omega)$ peut être par exemple, l'ensemble des décimaux (ou des rationnels) de $[0, 1]$, cf. *Une loi discrète pathologique* (ICP exercice 3.15 p. 78, corrigé dans PVIR pp. 455–462).

Si μ est une mesure de probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$, on dit que X « suit » la loi μ si $P_X = P \circ X^{-1} = \mu$, autrement dit, si la loi de X sous P est la mesure μ .

Remarque 2.9. Dans les problèmes usuels de probabilités, on travaille souvent avec un seul (Ω, \mathcal{F}, P) et on se contente alors de l'appellation *loi de X* . Il n'en va pas de même en statistique où l'on met généralement en concurrence *plusieurs* modèles $(\Omega, \mathcal{F}, P_\theta)$, où θ est un paramètre inconnu et où on se propose de choisir un de ces modèles au vu des valeurs $X(\omega)$ observées. C'est là que l'appellation loi de X sous P_θ s'impose. Pour donner un exemple simple, considérons le problème du sondage d'un échantillon de 500 personnes avant le second tour d'une élection présidentielle opposant le candidat A au candidat B . Ici θ est la proportion *inconnue* d'électeurs votant A dans la population totale. Si X est le nombre de personnes interrogées favorables à A , la loi de X sous P_θ est la loi binomiale³ $\text{Bin}(500, \theta)$.



IFP 2.5.5, pp. 81–83 : un premier exemple de modèle statistique.

Une autre situation où il est naturel de considérer plusieurs lois pour une même variable aléatoire est celle du *conditionnement*. Rappelons que si (Ω, \mathcal{F}, P) est un espace probabilisé et $H \in \mathcal{F}$ un évènement tel que $P(H) > 0$, on peut définir sur \mathcal{F} une nouvelle *mesure* de probabilité $P_H = P(\cdot | H)$ par

$$\forall A \in \mathcal{F}, \quad P_H(A) := P(A | H) = \frac{P(A \cap H)}{P(H)}.$$

Définition 2.10 (loi conditionnelle). Soient (Ω, \mathcal{F}, P) un espace probabilisé, $H \in \mathcal{F}$ tel que $P(H) > 0$, X une variable aléatoire sur (Ω, \mathcal{F}) . On appelle loi conditionnelle de X sachant H , la loi de X sous P_H . En la notant $P_{X|H}$, on a donc

$$\forall B \in \text{Bor}(\mathbb{R}), \quad P_{X|H}(B) = P_H(X^{-1}(B)) = P(X \in B | H).$$

Il importe de ne pas se laisser induire en erreur par la notation $P_{X|H}$, elle ne concerne pas une nouvelle variable aléatoire « $X | H$ » mais bien toujours *la même* variable aléatoire X . Ce qui a changé, c'est la probabilité dont on munit (Ω, \mathcal{F}) et sous laquelle on considère la loi de X .

Dans le cas d'une variable aléatoire discrète X , il est facile de donner une formule explicite pour la loi P_X .

Proposition 2.11. Soient (Ω, \mathcal{F}, P) un espace probabilisé et X une variable aléatoire discrète sur (Ω, \mathcal{F}) . La loi de X sous P est la probabilité

$$P_X = \sum_{x \in X(\Omega)} P(X = x) \delta_x, \tag{2.3}$$

que l'on peut considérer comme probabilité sur $(X(\Omega), \mathcal{P}(X(\Omega)))$, ou sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ ou sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$.



IFP, prop. 2.27, p. 75.

3. En fait c'est une loi hypergéométrique (tirages sans remise), mais en raison du théorème 2.38, on peut la remplacer en pratique par une binomiale.

Définition 2.12 (loi discrète sur \mathbb{R}). *On appelle loi discrète sur \mathbb{R} , toute mesure ponctuelle μ sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ qui est aussi une probabilité. Une telle loi admet donc une représentation sous la forme*

$$\mu = \sum_{i \in I} p_i \delta_{x_i},$$

où I est un ensemble d'indices au plus dénombrable, $(x_i)_{i \in I}$ une famille de nombres réels et $(p_i)_{i \in I}$ une famille sommable de réels positifs, de somme $\sum_{i \in I} p_i = 1$.

Il est clair que par restriction, μ est aussi une probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$.

Remarque 2.13. Deux variables aléatoires peuvent avoir même loi sans être égales. Par exemple considérons le jet de deux dés, l'un bleu et l'autre rouge. Notons X le nombre de points indiqué par le dé bleu et Y celui du rouge. Les variables aléatoires X et Y sont définies sur le même espace probabilisé $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ muni de l'équiprobabilité. On a $X(\Omega) = Y(\Omega) = \llbracket 1, 6 \rrbracket$ et :

$$\forall k \in \llbracket 1, 6 \rrbracket, \quad P(X = k) = \frac{1}{6}, \quad P(Y = k) = \frac{1}{6}.$$

Donc X et Y ont même loi : $P_X = P_Y = \sum_{k=1}^6 \frac{1}{6} \delta_k$. Pour autant, on n'a pas l'égalité des variables aléatoires X et Y qui signifierait $X(\omega) = Y(\omega)$ pour tout $\omega \in \Omega$ (égalité de deux applications). Autrement dit, en lançant deux dés on obtiendrait à coup sûr un double. En revanche nous pouvons considérer l'évènement $\{X = Y\}$ dont la réalisation n'est pas certaine et calculer sa probabilité :

$$P(X = Y) = P\left(\bigcup_{k=1}^6 \{(X, Y) = (k, k)\}\right) = \frac{6}{36} = \frac{1}{6}.$$

On en déduit : $P(X \neq Y) = 5/6$.

Remarque 2.14. Deux variables aléatoires peuvent avoir même loi en étant définies sur des espaces probabilisés différents (Ω, \mathcal{F}, P) et $(\Omega', \mathcal{F}', P')$. Prenons par exemple pour X les points du dé bleu comme ci-dessus et posons $Z = 0$ si X est pair, $Z = 1$ si X est impair. La variable aléatoire Z est définie sur $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ muni de la tribu $\mathcal{P}(\Omega)$ et de l'équiprobabilité P sur Ω . Sa loi est $P_Z = \frac{1}{2}(\delta_0 + \delta_1)$. Prenons maintenant $\Omega' = \{-1, 1\}$, muni de la tribu $\mathcal{P}(\Omega')$ et de l'équiprobabilité P' sur Ω' et posons pour $\omega' \in \Omega'$, $Z'(\omega') := (1 + \omega')/2$. Alors la loi de Z' est aussi $P'_{Z'} = \frac{1}{2}(\delta_0 + \delta_1) = P_Z$. Remarquons que si X et Y sont définies sur des espaces probabilisés différents, il n'y a pas d'évènement $\{X = Y\}$, pas plus que de variable aléatoire « $X + Y$ ». Essayez d'en écrire la définition explicite pour vous en convaincre.

Remarque 2.15. Si X est une v.a. discrète sur (Ω, \mathcal{F}) , alors pour toute probabilité P sur (Ω, \mathcal{F}) , la loi de X sous P est une loi discrète au sens de la définition 2.12. C'est une conséquence de la proposition 2.11. Mais il peut aussi exister sur (Ω, \mathcal{F}) une variable aléatoire réelle Y et une probabilité P telles que Y ne soit pas discrète ($Y(\Omega)$ étant un ensemble infini non dénombrable) mais que la loi de Y sous P soit une loi discrète. Voici un exemple simple. On prend $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{P}(\mathbb{R})$ et $Y : \omega \mapsto \omega$ l'application identité sur \mathbb{R} . Alors Y n'est pas une v.a. discrète puisque $Y(\Omega) = \mathbb{R}$. Notons au passage que Y est mesurable relativement à n'importe quelle tribu sur l'ensemble d'arrivée puisque l'ensemble de départ Ω est muni ici de la tribu $\mathcal{P}(\Omega)$. En particulier, Y est bien une variable aléatoire réelle. Munissons maintenant $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{P}(\mathbb{R}))$ de la mesure de probabilité $P = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ et cherchons la loi de Y sous P . Comme Y est l'identité sur \mathbb{R} , on a

l'égalité $Y^{-1}(B) = B$ pour tout borélien B . Par conséquent $P_Y(B) = P(Y^{-1}(B)) = P(B)$ pour tout $B \in \text{Bor}(\mathbb{R})$. On a donc $P_Y = P = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$, ce qui montre que la loi de Y sous P est la loi discrète $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Bien sûr on a un résultat analogue en remplaçant P par n'importe quelle loi discrète Q sur \mathbb{R} , au sens de la définition 2.12.

Remarque 2.16. Pour toute probabilité Q sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$, il existe au moins un espace probabilisé (Ω, \mathcal{F}, P) et une variable aléatoire réelle X sur (Ω, \mathcal{F}) dont la loi sous P soit égale à Q . Il suffit de prendre $\Omega = \mathbb{R}$, $\mathcal{F} = \text{Bor}(\mathbb{R})$ et pour X l'application identité de $\mathbb{R} \rightarrow \mathbb{R}$. En prenant $P = Q$, on a clairement $P_X = Q$. Bien entendu, il y a une infinité d'autres solutions à ce problème.

Il y a donc identité entre les mesures de probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ et les lois des variables aléatoires réelles. Comme nous savons caractériser une probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ par sa fonction de répartition (théorème 1.26), ceci va nous permettre de classer les lois des variables aléatoires réelles.

2.2.2 Fonction de répartition

Définition 2.17 (f.d.r. d'une variable aléatoire). Soient (Ω, \mathcal{F}, P) un espace probabilisé et X une variable aléatoire sur (Ω, \mathcal{F}) . On appelle fonction de répartition (f.d.r.) de X , la fonction F_X définie sur \mathbb{R} par :

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

La fonction F_X est la fonction de répartition de la probabilité P_X , au sens de la définition 1.24. Elle ne dépend donc que de la loi⁴ de X . Deux variables aléatoires de même loi ont même fonction de répartition. La proposition suivante donnant les propriétés générales des fonctions de répartition des variables aléatoires n'est qu'une simple réécriture de la proposition 1.25.

Proposition 2.18. La fonction de répartition F_X d'une variable aléatoire X est croissante sur \mathbb{R} , avec limite 0 en $-\infty$ et 1 en $+\infty$. Elle est continue à droite et limitée à gauche en tout point de \mathbb{R} et vérifie :

$$\forall x \in \mathbb{R}, \quad P(X = x) = F(x) - F(x-). \quad (2.4)$$

La fonction de répartition d'une variable aléatoire caractérise sa loi, autrement dit : $F_X = F_Y$ si et seulement si les variables aléatoires X et Y ont même loi.

On vérifie facilement les formules suivantes de calcul à l'aide de F_X des $P(X \in I)$ pour I intervalle de \mathbb{R} :

$$P(a < X \leq b) = F_X(b) - F_X(a), \quad (2.5)$$

$$P(a \leq X \leq b) = F_X(b) - F_X(a-), \quad (2.6)$$

$$P(a \leq X < b) = F_X(b-) - F_X(a-), \quad (2.7)$$

$$P(a < X < b) = F_X(b-) - F_X(a), \quad (2.8)$$

$$P(X \leq a) = F_X(a), \quad (2.9)$$

$$P(X < a) = F_X(a-), \quad (2.10)$$

$$P(X > b) = 1 - F_X(b), \quad (2.11)$$

$$P(X \geq b) = 1 - F_X(b-). \quad (2.12)$$

4. Il serait plus correct, mais plus long, de parler de f.d.r. de la loi de X ou même de f.d.r. de la loi de X sous P .

Si a et b sont des points de continuité de F_X , ces formules se simplifient considérablement.

Dans le cas particulier des variables aléatoires discrètes, on peut donner une formule explicite de calcul de la fonction de répartition.

Proposition 2.19 (f.d.r. d'une variable aléatoire discrète). *Soient (Ω, \mathcal{F}, P) un espace probabilisé et X une variable aléatoire discrète sur (Ω, \mathcal{F}) . Fixons une numérotation de l'ensemble au plus dénombrable $X(\Omega)$ par les entiers : $X(\Omega) = \{x_0, x_1, \dots, x_k, \dots\}$ et notons $p_k := P(X = x_k)$. La fonction de répartition F_X vérifie alors :*


$$\forall x \in \mathbb{R}, \quad F_X(x) = \sum_{x_k \in X(\Omega)} p_k \mathbf{1}_{[x_k, +\infty[}(x), \quad (2.13)$$

ce qui s'écrit aussi

$$\forall x \in \mathbb{R}, \quad F_X(x) = \sum_{\substack{x_k \in X(\Omega) \\ x_k \leq x}} P(X = x_k). \quad (2.14)$$

De plus, si on peut numéroter les éléments de $X(\Omega)$ de manière croissante ($x_k < x_{k+1}$ pour tout k), la fonction F_X est constante sur chaque intervalle $[x_n, x_{n+1}[$ et vaut sur cet intervalle $F_X(x) = \sum_{k \leq n} p_k$.

Cette proposition se généralise au cas des variables aléatoires de loi discrète⁵, en remplaçant $X(\Omega)$ par $X_P(\Omega) := \{x \in \mathbb{R} ; P(X = x) > 0\}$.

 PVIR, prop. 6.20, pp. 212–213.

Si l'on veut esquisser une classification sommaire des lois des variables aléatoires, on peut commencer par les partager entre les lois à f.d.r. continue sur \mathbb{R} et les lois à f.d.r. non continue⁶ sur \mathbb{R} . On parle plus simplement de *lois continues* ou encore *lois diffuses* dans le premier cas et de lois non continues ou non diffuses dans le deuxième. Dans la famille des lois non continues, nous connaissons déjà la sous-famille des lois discrètes. Dans la famille des lois continues, une importante sous-famille est celle des *lois à densité* que nous allons examiner maintenant.

2.2.3 Lois à densité

Définition 2.20 (densité de probabilité). *On appelle densité de probabilité sur \mathbb{R} toute fonction borélienne positive f vérifiant*

$$\int_{\mathbb{R}} f \, d\lambda = 1,$$

où λ désigne la mesure de Lebesgue sur \mathbb{R} .

Si f est une fonction borélienne positive définie seulement sur un intervalle I de \mathbb{R} et telle que $\int_I f \, d\lambda = 1$, on peut en faire une densité en la prolongeant à tout \mathbb{R} en posant $f(t) := 0$ pour $t \notin I$. Voici quatre exemples simples de densités :

$$\begin{aligned} f_1(t) &:= \frac{1}{b-a} \mathbf{1}_{[a,b]}(t); & f_2(t) &:= \frac{1}{2\sqrt{t}} \mathbf{1}_{]0,1]}(t); \\ f_3(t) &:= e^{-t} \mathbf{1}_{[0,+\infty[}(t); & f_4(t) &:= \frac{1}{\pi(1+t^2)}. \end{aligned}$$

⁵. Voir la remarque 2.15.

⁶. Rappelons que l'ensemble des points de discontinuité d'une f.d.r. quelconque est au plus dénombrable.

Remarque 2.21 (usage des indicatrices dans les formules explicites). La définition de f_2 repose sur un *abus d'écriture* d'usage courant. En effet il y a en toute rigueur un problème pour calculer $f_2(t)$ lorsque $t \leq 0$, puisqu'alors il nous faut former le produit de l'expression $\frac{1}{2\sqrt{t}}$ non définie (du moins en tant que nombre réel) par 0. La convention adoptée est que si la formule de calcul d'une fonction contient le produit d'une indicatrice par une expression non définie lorsque cette indicatrice est nulle, le produit vaut 0 dans ce cas. Ceci permet de considérer que la « définition » de f_2 comme ci-dessus est un raccourci d'écriture commode pour :

$$f_2(t) := \begin{cases} \frac{1}{2\sqrt{t}} & \text{si } t \in]0, 1], \\ 0 & \text{si } t \notin]0, 1]. \end{cases}$$

Définition 2.22. Soient (Ω, \mathcal{F}, P) un espace probabilisé et X une variable aléatoire réelle sur (Ω, \mathcal{F}) . La loi de X sous P a pour densité f si :

$$\forall B \in \text{Bor}(\mathbb{R}), \quad P(X \in B) = \int_B f \, d\lambda. \quad (2.15)$$

On dit aussi par abus de langage que X a pour densité f (lorsqu'il n'y a pas ambiguïté sur P).

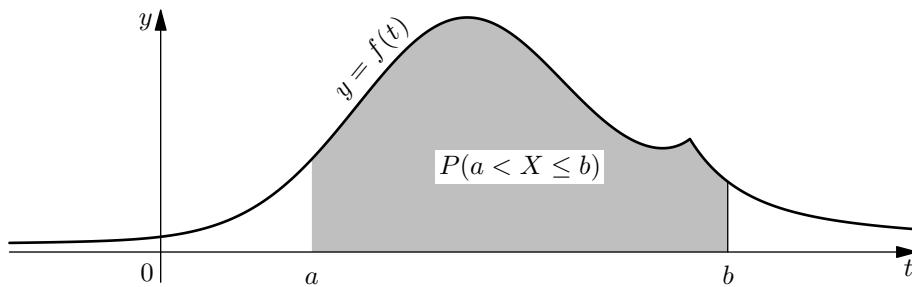


FIGURE 2.1 – $P(a < X \leq b) = \int_a^b f(t) \, dt$ pour X de densité f

Remarque 2.23. La notation $\int_B f \, d\lambda$ est commode pour intégrer sur un borélien B quelconque. En pratique B sera souvent un intervalle I d'extrémités a et b dans \mathbb{R} et nous utiliserons la notation

$$\int_I f \, d\lambda = \int_a^b f(t) \, dt,$$

comme s'il s'agissait d'une intégrale de Riemann (éventuellement généralisée), ce qui d'ailleurs sera vrai dans la plupart des cas.



Pour la comparaison entre intégrale de Riemann et de Lebesgue : IFP pp. 132-135.

Remarque 2.24. Il est clair d'après la définition 2.22 que si Y est une autre variable aléatoire ayant même loi que X (donc mêmes probabilités d'appartenance aux boréliens), elle a aussi la densité f . D'autre part, il n'y a pas unicité de la densité d'une variable aléatoire. Par exemple $g_1 = \mathbf{1}_{[0,1]}$ et $g_2 = \mathbf{1}_{]0,1]}$ sont deux densités de probabilité qui donnent les mêmes intégrales : $\int_B g_1 \, d\lambda = \int_B g_2 \, d\lambda$ pour tout borélien B . Ces deux fonctions peuvent chacune être prise comme densité de la loi uniforme sur $[0, 1]$.

L'exemple ci-dessus montre qu'il ne suffit pas de vérifier que deux variables aléatoires ont des densités qui diffèrent en un point pour en déduire qu'elles n'ont pas même loi. Le lemme suivant donne une condition suffisante pratique pour que deux variables à densité n'aient pas même loi.

Lemme 2.25. *Soient X et Y deux variables aléatoires admettant respectivement pour densité les fonctions f et g . On suppose qu'il existe un réel t_0 tel que $f(t_0) \neq g(t_0)$ et que de plus, f et g sont toutes deux continues au point t_0 . Alors X et Y n'ont pas même loi.*



P VIR lemme 6.25, pp. 215–216.

Examinons maintenant les relations entre densité (lorsqu'elle existe) et fonction de répartition (qui elle, existe toujours) d'une variable aléatoire.

Proposition 2.26. *Si la variable aléatoire X a pour densité f , sa fonction de répartition F vérifie :*

- a) $\forall x \in \mathbb{R}, F(x) = \int_{-\infty}^x f(t) dt$;
- b) F est continue sur \mathbb{R} ;
- c) si f est continue au point x_0 , alors F est dérivable en ce point et $F'(x_0) = f(x_0)$.



P VIR pp. 216–217 ou ICP pp. 189–190.

Corollaire 2.27. *Si la variable aléatoire X a pour densité f , les égalités suivantes sont vérifiées pour tous $a, b \in \mathbb{R}$ tels que $a < b$:*

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f(t) dt,$$

$$P(X < a) = P(X \leq a) = \int_{-\infty}^a f(t) dt,$$

$$P(X > b) = P(X \geq b) = \int_b^{+\infty} f(t) dt.$$

Remarques 2.28.

1. Pour toute densité f (au sens de la définition 2.20), il existe une variable aléatoire X ayant f pour densité : il suffit d'appliquer le théorème 1.26 pour obtenir l'existence d'une mesure de probabilité μ sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ ayant pour f.d.r F définie par a). La remarque 2.16 nous assure de l'existence d'une variable aléatoire X de loi μ , donc de f.d.r. F . La preuve du b) ci-dessus nous montre que F est continue sur \mathbb{R} . En particulier pour toute paire de réels $a \leq b$, on a $P(a < X \leq b) = F(b) - F(a) = \int_{-\infty}^b f(t) dt - \int_{-\infty}^a f(t) dt = \int_a^b f(t) dt$.
2. D'après b) toute variable aléatoire à densité a une fonction de répartition continue. La réciproque est fausse : il existe des lois à fonction de répartition continue sans densité.
3. Par ailleurs si X a une densité, sa fonction de répartition n'est pas forcément dérivable en tout point. Par exemple la densité f_2 ci-dessus a pour fonction de répartition associée $F_2(x) = \sqrt{x} \mathbf{1}_{]0,1]}(x) + \mathbf{1}_{]1,+\infty[}(x)$ (cette écriture condensée signifie que $F_2(x)$ est nul sur \mathbb{R}^- , vaut \sqrt{x} entre 0 et 1 et reste constant égal à 1 sur $]1, +\infty[$). F_2 est dérivable en tout point sauf en 0 et en 1.

♣ Pour un exemple de loi continue sans densité, PVIR pp. 219–220 et 239–243.

La proposition suivante donne une règle pratique permettant de trouver la densité (lorsqu'elle existe !) à partir de la fonction de répartition dans les cas les plus courants.

Proposition 2.29. *On suppose que la fonction de répartition F de X est C^1 par morceaux au sens suivant : F est continue sur \mathbb{R} et dérivable sur \mathbb{R} privé (éventuellement) d'un ensemble fini de points $a_1 < \dots < a_n$. Sur chacun des intervalles ouverts $] -\infty, a_1[$, $]a_i, a_{i+1}[$ ($1 \leq i < n$), $]a_n, +\infty[$, la dérivée f de F est continue. Alors X a pour densité f .*

♣ PVIR p. 218, ICP p. 191 ou IFP p. 137.

L'idée sous-jacente à la proposition 2.29 est que si la f.d.r. F d'une variable aléatoire X est suffisamment régulière, alors la loi de X a une densité. On a vu d'autre part, proposition 2.26 b), que si la loi de X est à densité, sa f.d.r. F est continue. La question qui surgit alors naturellement est : peut-on caractériser l'existence d'une densité par la régularité de la fonction de répartition ? La réponse est oui et la notion de régularité de la f.d.r. qui équivaut à l'existence d'une densité est celle d'absolue continuité. La preuve de cette équivalence est hors programme. Nous allons néanmoins introduire cette notion et montrer que toute loi à densité a une f.d.r. absolument continue. Cela nous sera utile pour montrer l'existence de lois à f.d.r. continue mais sans densité.

Définition 2.30. *La fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ est dite absolument continue sur \mathbb{R} , si pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que pour tout $n \in \mathbb{N}^*$ et toute famille finie d'intervalles $[a_k, b_k] \subset \mathbb{R}$, $1 \leq k \leq n$ tels que les $]a_k, b_k[$ soient 2 à 2 disjoints,*

$$\sum_{k=1}^n (b_k - a_k) < \delta \implies \sum_{k=1}^n |F(b_k) - F(a_k)| < \varepsilon. \quad (2.16)$$

Notons que si F est croissante, les valeurs absolues sont superflues dans la dernière inégalité ci-dessus. La continuité absolue sur \mathbb{R} implique évidemment la continuité uniforme sur \mathbb{R} . Par ailleurs toute f.d.r. continue sur \mathbb{R} est *uniformément* continue sur \mathbb{R} (exercice). Nous verrons ultérieurement qu'il existe des f.d.r. continues sur \mathbb{R} mais non absolument continues, donc que la continuité absolue est une propriété strictement plus forte que la continuité uniforme.

Proposition 2.31. *La fonction de répartition F d'une loi à densité f est absolument continue sur \mathbb{R} .*

La preuve de cette proposition est grandement simplifiée par le lemme suivant de théorie de la mesure.

Lemme 2.32. *Soit (E, \mathcal{F}, μ) un espace mesuré et $f : E \rightarrow \overline{\mathbb{R}}_+$ μ intégrable sur E . On définit sur \mathcal{F} une fonction d'ensembles ν en posant :*

$$\forall A \in \mathcal{F}, \quad \nu(A) := \int_A f \, d\mu.$$

Alors

$$\forall \varepsilon > 0, \exists \delta > 0, \quad \mu(A) < \delta \implies \nu(A) < \varepsilon.$$

Il se trouve que ν ainsi définie est une mesure (cf. IFP th. 3.17 pp. 95–96), mais nous n'avons pas besoin de cette propriété pour prouver le lemme.

Preuve du lemme. De l'hypothèse $\int_E f \, d\mu < +\infty$, on déduit que

$$\lim_{n \rightarrow \infty} \int_{\{f > n\}} f \, d\mu = \int_{\{f = +\infty\}} f \, d\mu = 0. \quad (2.17)$$

Pour justifier la première égalité, on peut utiliser le théorème de convergence dominée ou le théorème de convergence décroissante (avec premier terme intégrable). Pour justifier la deuxième égalité, on note que $\mu(\{f = +\infty\}) = 0$, cf. IFP prop. 4.19 p. 116.

Fixons $\varepsilon > 0$ arbitraire. D'après (2.17), il existe un n_0 tel que

$$\int_{\{f > n_0\}} f \, d\mu < \frac{\varepsilon}{2}.$$

Prenons ensuite $\delta = \varepsilon/(2n_0)$. Alors pour tout $A \in \mathcal{F}$ tel que $\mu(A) < \delta$,

$$\begin{aligned} \nu(A) &= \int_A f \, d\mu = \int_{A \cap \{f \leq n_0\}} f \, d\mu + \int_{A \cap \{f > n_0\}} f \, d\mu \\ &\leq n_0 \mu(A \cap \{f \leq n_0\}) + \int_{\{f > n_0\}} f \, d\mu \\ &\leq n_0 \mu(A) + \frac{\varepsilon}{2} \leq \varepsilon. \end{aligned}$$

Comme $\varepsilon > 0$ était arbitraire, le lemme est prouvé. \square

Preuve de la proposition 2.31. On applique le lemme 2.32 avec $E = \mathbb{R}$, $\mathcal{F} = \text{Bor}(\mathbb{R})$, $\mu = \lambda$ mesure de Lebesgue et $\nu = P_X$ loi de X à densité f . On sait déjà que la f.d.r. F de X est continue sur \mathbb{R} donc en particulier que pour tout k , $P_X(\{a_k\}) = P_X(\{b_k\}) = 0$ et $P_X([a_k, b_k]) = P_X(]a_k, b_k]) = F(b_k) - F(a_k)$.

Fixons $\varepsilon > 0$ arbitraire et associons lui δ comme dans le lemme. Pour toute suite finie d'intervalles $[a_k, b_k]$, $1 \leq k \leq n$ deux à deux disjoints vérifiant $\sum_{k=1}^n (b_k - a_k) < \delta$, on pose $A = \bigcup_{k=1}^n [a_k, b_k]$ et on remarque que

$$\lambda(A) = \sum_{k=1}^n (b_k - a_k).$$

Ainsi $\lambda(A) < \delta$ et donc par le lemme $\nu(A) = P_X(A) < \varepsilon$. Mais

$$P_X(A) = P_X\left(\bigcup_{k=1}^n [a_k, b_k]\right) = P_X\left(\bigcup_{k=1}^n]a_k, b_k]\right) = \sum_{k=1}^n P_X(]a_k, b_k]) = \sum_{k=1}^n (F(b_k) - F(a_k)).$$

On en déduit que

$$\sum_{k=1}^n (F(b_k) - F(a_k)) < \varepsilon$$

et comme ε était arbitraire, ceci prouve l'absolue continuité de F . \square

2.3 Lois discrètes classiques

Dans toute la suite du chapitre, on fixe un espace probabilisé (Ω, \mathcal{F}, P) , on désigne par X une variable aléatoire sur (Ω, \mathcal{F}) et par P_X sa loi sous P . Cette clause sera implicite chaque fois que nous écrirons dans les définitions « La variable aléatoire X suit la loi ... si ... ». Pour X de loi discrète, nous utiliserons la notation :

$$X_P(\Omega) := \{x \in \mathbb{R}; P(X = x) > 0\}. \quad (2.18)$$

Bien sûr, $X_P(\Omega)$ est toujours inclus dans l'ensemble des valeurs « possibles » $X(\Omega)$ et dans la plupart des situations pratiques d'utilisation des variables aléatoires, ces deux ensembles sont égaux⁷.

2.3.1 Lois de Bernoulli

Définition 2.33. *La variable aléatoire X suit la loi de Bernoulli de paramètre p ($p \in]0, 1[$) si $X_P(\Omega) = \{0, 1\}$ avec :*

$$P(X = 1) = p, \quad P(X = 0) = 1 - p = q.$$

On notera $X \sim \text{Bern}(p)$.

Si A est un évènement de probabilité p , son indicatrice définie par :

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases} = \begin{cases} 1 & \text{si } A \text{ est réalisé} \\ 0 & \text{si } A \text{ n'est pas réalisé} \end{cases}$$

est une variable aléatoire suivant la loi de Bernoulli de paramètre p . Réciproquement, si X est une v.a. de Bernoulli, on peut toujours écrire que $X = \mathbf{1}_A$ *presque-sûrement*, c'est-à-dire $P(X = \mathbf{1}_A) = 1$, en définissant $A = \{\omega \in \Omega, X(\omega) = 1\}$.

2.3.2 Loi uniforme sur un ensemble fini de réels

Définition 2.34. *La variable aléatoire X suit la loi uniforme sur $\{x_1, \dots, x_n\} \subset \mathbb{R}$ si P_X est l'équiprobabilité sur cet ensemble. Notation : $X \sim \text{Unif}\{x_1, \dots, x_n\}$.*

Autrement dit, $X_P(\Omega) = \{x_1, \dots, x_n\}$ et

$$\forall k \in \llbracket 1, n \rrbracket, \quad P(X = x_k) = \frac{1}{n}.$$

D'où

$$P_X = \frac{1}{n} \sum_{k=1}^n \delta_{x_k}.$$

Par exemple, le nombre de points indiqué par un dé équilibré suit la loi uniforme sur $\llbracket 1, 6 \rrbracket$.

7. Pour comprendre l'utilité de ce distinguo entre $X(\Omega)$ et $X_P(\Omega)$, (re)lisez la remarque 2.15.

2.3.3 Loïs binomiales

Définition 2.35. La variable aléatoire X suit la loi binomiale de paramètres n et p , $n \in \mathbb{N}^*$ et $p \in [0, 1]$, notation $X \sim \text{Bin}(n, p)$, si $X_P(\Omega) = \llbracket 0, n \rrbracket$ et

$$\forall k \in \llbracket 0, n \rrbracket, \quad P(X = k) = C_n^k p^k (1-p)^{n-k}.$$

La formule ci-dessus définit bien une loi de probabilité puisque les $C_n^k p^k (1-p)^{n-k}$ sont positifs et :

$$\sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1,$$

en appliquant la formule du binôme de Newton (d'où le nom de la loi). La loi binomiale $\text{Bin}(n, p)$ est la loi du nombre de succès obtenus en une suite de n épreuves répétées indépendantes avec pour chaque épreuve une probabilité de succès p . Ceci a été démontré dans l'exemple 1.49.

De même, soit A_1, \dots, A_n une famille d'événements mutuellement indépendants tous de même probabilité p et notons X_i la variable de Bernoulli indicatrice de A_i :

$$X_i(\omega) = \begin{cases} 1 & \text{si } \omega \in A_i, \\ 0 & \text{si } \omega \in A_i^c. \end{cases}$$

Alors la variable aléatoire $S_n = \sum_{i=1}^n X_i$ suit la loi binomiale $\text{Bin}(n, p)$.

2.3.4 Loïs hypergéométriques

Alors que la loi binomiale intervient dans les tirages avec remise, la loi hypergéométrique correspond aux tirages sans remise.

Exemple 2.36. Dans une production totale de N objets dont M sont défectueux, on prélève au hasard un échantillon de n objets (tirage sans remise). Soit X le nombre aléatoire d'objets défectueux dans l'échantillon. Quelle est sa loi ?

On peut prendre comme espace Ω l'ensemble de tous les échantillons possibles (toutes les parties à n éléments d'un ensemble de cardinal N) muni de l'équiprobabilité. Chaque échantillon a ainsi une probabilité $1/C_N^n$ d'être choisi. Les échantillons (événements élémentaires) réalisant l'événement $\{X = k\}$ sont ceux qui contiennent k objets défectueux et $n-k$ objets non défectueux. Ceci n'est réalisable que si $0 \leq k \leq M$ et $0 \leq n-k \leq N-M$. Dénombrons ces échantillons. On les forme en choisissant k objets défectueux dans une sous-population de taille M et en complétant par $n-k$ objets non défectueux choisis dans une sous-population de taille $N-M$. Il y en a donc $C_M^k \times C_{N-M}^{n-k}$. Finalement :

$$P(X = k) = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} \quad \text{si} \quad \begin{cases} 0 \leq k \leq M, \\ 0 \leq n-k \leq N-M. \end{cases} \quad (2.19)$$

Définition 2.37. La loi définie par (2.19) s'appelle loi hypergéométrique de paramètres N , M et n . Notation : $X \sim \text{Hypg}(N, M, n)$. Le paramètre N est l'effectif de la population totale, M celui de la sous-population à laquelle on s'intéresse et n la taille de l'échantillon observé.

Pour une taille d'échantillon n fixée, plus N et M sont grands, moins les tirages sans remise diffèrent des tirages avec remise. Plus précisément, la loi hypergéométrique converge vers la loi binomiale au sens suivant.

Théorème 2.38 (convergence de l'hypergéométrie vers la binomiale). *On suppose que quand N tend vers $+\infty$, $M = M(N)$ tend vers $+\infty$ en vérifiant la condition :*

$$\lim_{N \rightarrow +\infty} \frac{M}{N} = p \quad \text{avec} \quad 0 < p < 1. \quad (2.20)$$

Alors, n restant fixé, la loi hypergéométrique $\text{Hypg}(N, M, n)$ converge vers la loi binomiale $\text{Bin}(n, p)$, ce qui signifie que si $(X_N)_{N \geq 1}$ est une suite de v.a. avec $X_N \sim \text{Hypg}(N, M, n)$ et Y est une v.a. de loi binomiale $\text{Bin}(n, p)$, alors :

$$\forall k \in \llbracket 0, n \rrbracket, \quad \lim_{N \rightarrow +\infty} P(X_N = k) = P(Y = k), \quad (2.21)$$

autrement dit :

$$\forall k \in \llbracket 0, n \rrbracket, \quad \lim_{N \rightarrow +\infty} \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}. \quad (2.22)$$



ICP pp. 60–61.

2.3.5 Lois géométriques

Exemple 2.39 (un problème de temps d'attente).

Considérons une suite infinie d'épreuves répétées indépendantes avec même probabilité de succès $p \in]0, 1[$. Soit X le numéro (aléatoire) de la première épreuve où l'on obtient un succès. Si l'on n'obtient jamais de succès, on conviendra que $X = +\infty$. Calculer $P(X = k)$ pour tout $k \in \mathbb{N}^*$. En déduire les valeurs de $P(X \in \mathbb{N}^*)$ et $P(X = +\infty)$.

En notant $R_i = \{\text{succès à la } i\text{-ème épreuve}\}$, on a :

$$\begin{aligned} \{X = k\} &= \{\text{échec aux } (k-1) \text{ premières et succès à la } k\text{-ième}\} \\ &= \left(\bigcap_{i=1}^{k-1} R_i^c \right) \cap R_k. \end{aligned}$$

D'où par indépendance des épreuves :

$$P(X = k) = \left(\prod_{i=1}^{k-1} P(R_i^c) \right) \times P(R_k) = (1-p)^{k-1} p.$$

Posons $q = 1 - p$ et notons que $q \in]0, 1[$. La décomposition de l'évènement $\{X \in \mathbb{N}^*\}$ en la réunion disjointe des $\{X = k\}$, $k \in \mathbb{N}^*$, nous donne par σ -additivité :

$$\begin{aligned} P(X \in \mathbb{N}^*) &= \sum_{k \in \mathbb{N}^*} P(X = k) = \sum_{k \in \mathbb{N}^*} q^{k-1} p \\ &= p \sum_{l \in \mathbb{N}} q^l \quad (l = k-1) \\ &= p \frac{1}{1-q} = 1. \end{aligned}$$

Ainsi avec probabilité 1, le premier succès a lieu au bout d'un nombre *fini* d'épreuves⁸. Remarquons qu'on aurait pu arriver au même résultat en montrant que $P(X = +\infty) = 0$ par la méthode utilisée à l'exemple 1.49 c) en échangeant les rôles de succès et échec.

8. Mais pas borné par un nombre fixé choisi avant le début des épreuves...

En toute rigueur, X n'est pas une variable aléatoire discrète au sens de la définition 2.6 puisque $X(\Omega)$ est une partie dénombrable de $\overline{\mathbb{R}}$ au lieu de \mathbb{R} . Néanmoins $X' := X\mathbf{1}_{\{X < +\infty\}}$ est une variable aléatoire discrète et ce qui précède montre que X' a même loi⁹ que X . Cette loi est celle du temps d'attente du premier succès dans une suite d'épreuves répétées indépendantes, on l'appelle *loi géométrique de paramètre p* .

Définition 2.40. Une variable aléatoire X suit la loi géométrique de paramètre $p \in]0, 1[$, si $X_P(\Omega) = \mathbb{N}^*$ et :

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

Notation : $X \sim \text{Geom}(p)$.

Lorsque X suit une loi géométrique, les probabilités $P(X > n)$ ont une expression particulièrement simple en fonction de $q = 1 - p$:

$$\forall n \in \mathbb{N}, \quad P(X > n) = q^n.$$

 ICP p. 62.

2.3.6 Lois de Poisson

Définition 2.41. On dit que la variable aléatoire discrète X suit la loi de Poisson de paramètre $\alpha > 0$ si $X_P(\Omega) = \mathbb{N}$ et

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

Notation : $X \sim \text{Pois}(\alpha)$.

On sait que la fonction exponentielle a un développement en série entière avec rayon de convergence infini. En particulier :

$$\forall \alpha > 0, \quad e^\alpha = \sum_{k=0}^{+\infty} \frac{\alpha^k}{k!}.$$

On a donc bien :

$$\sum_{k=0}^{+\infty} P(X = k) = e^{-\alpha} \sum_{k=0}^{+\infty} \frac{\alpha^k}{k!} = e^{-\alpha} e^\alpha = 1.$$

Une des raisons de l'importance de cette loi est le théorème de convergence de la loi binomiale vers la loi de Poisson.

Théorème 2.42. Si $(p_n)_{n \geq 1}$ est une suite de réels de $[0, 1]$ vérifiant

$$np_n \xrightarrow[n \rightarrow +\infty]{} \alpha \in]0, +\infty[, \quad (2.23)$$

alors pour tout $k \in \mathbb{N}$,

$$C_n^k p_n^k (1 - p_n)^{n-k} \xrightarrow[n \rightarrow +\infty]{} \frac{e^{-\alpha} \alpha^k}{k!}.$$

9. Pour être tout à fait rigoureux, il faudrait avoir défini les v.a. à valeurs dans $\overline{\mathbb{R}}$ et leurs lois, cf IFP pp. 57–60.

♣ ICP pp. 63–64.

Le théorème 2.42 sert de justification théorique à la règle pratique suivante : lorsque n est « grand » et np « petit », on peut remplacer la loi $\text{Bin}(n, p)$ par la loi $\text{Pois}(\alpha)$ où $\alpha = np$. En général on considère que n de l'ordre de quelques centaines et np de l'ordre de quelques unités donnent une bonne approximation. Sous cette forme, cette règle relève plus de la cuisine que des mathématiques. Il est possible par des techniques élémentaires de contrôler l'erreur commise en utilisant cette approximation.

♣ ICP exemple 3.3. p. 64 et exercice 3.19 p. 81.

Comparaison graphique. Les *diagrammes en bâtons* ci-dessous représentent la loi binomiale $\text{Bin}(n, p)$ et la loi de Poisson approximante $\text{Pois}(\alpha)$ avec $\alpha = np$. Les segments verticaux (les bâtons) du diagramme représentant la loi d'une variable discrète X (à valeurs dans \mathbb{N}) ont une hauteur égale à $P(X = k)$ avec une extrémité inférieure au point d'abscisse k de l'axe horizontal. Pour la lisibilité, on a légèrement décalé vers la gauche les bâtons de la loi de Poisson et vers la droite ceux de la loi binomiale. Bien que le diagramme en bâtons de la loi binomiale $\text{Bin}(n, p)$ soit constitué théoriquement de $n + 1$ bâtons (et que celui de la loi de Poisson en ait une infinité), seul un petit nombre de bâtons est visible sur les graphiques, les autres correspondant à des probabilités trop petites¹⁰. On constate que pour $n = 200$ (figure 2.5), la différence entre les deux diagrammes n'est pratiquement plus discernable *visuellement*.

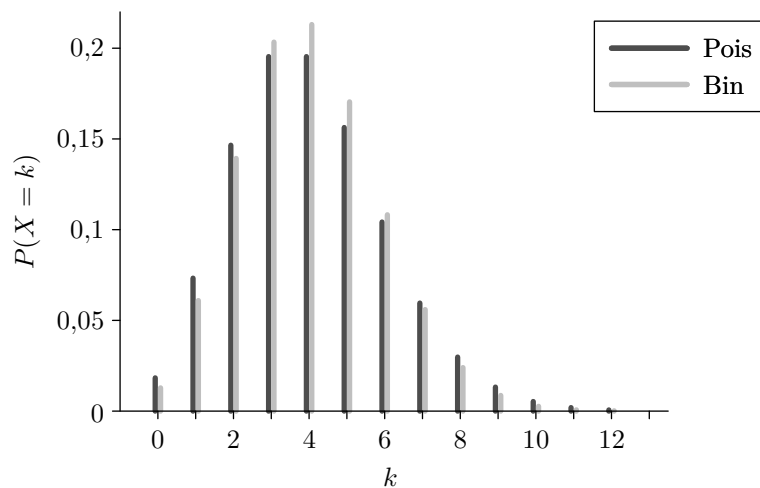


FIGURE 2.2 – Lois $\text{Bin}(25; 0.16)$ et $\text{Pois}(4)$

♣ ICP pp. 70–73, sur le caractère universel de la loi de Poisson.

¹⁰. En fait, on s'est contenté d'afficher les probabilités correspondant à k inférieur ou égal à la partie entière supérieure de $2\alpha + 4$. On peut vérifier que la somme des probabilités ainsi négligées est inférieure à 1%, pour chacune des deux lois.

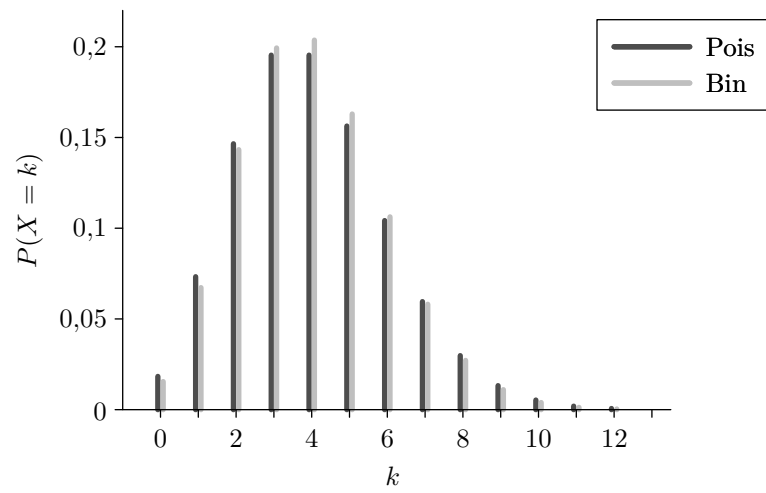


FIGURE 2.3 – Lois Bin(50; 0,08) et Pois(4)

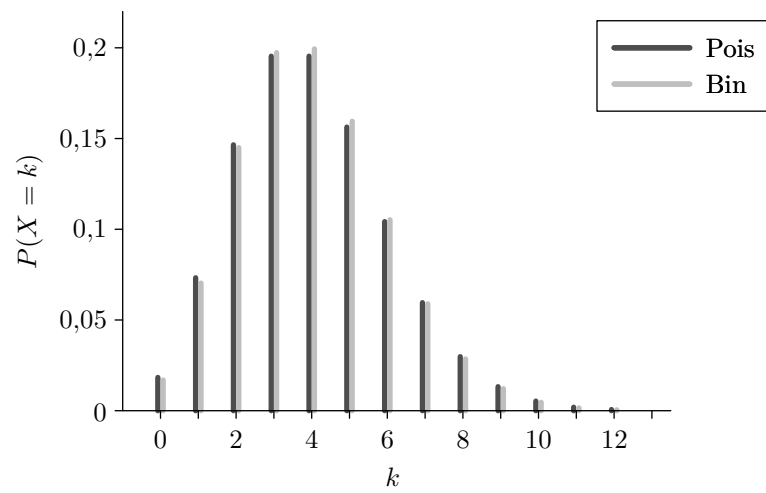


FIGURE 2.4 – Lois Bin(100; 0,04) et Pois(4)

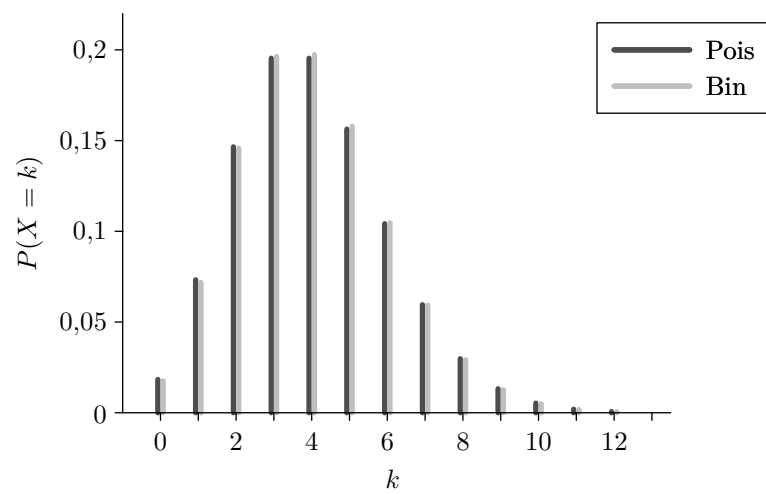


FIGURE 2.5 – Lois Bin(200; 0,02) et Pois(4)

2.4 Lois à densité classiques

2.4.1 Lois uniformes

Définition 2.43. La variable aléatoire réelle X suit la loi uniforme sur l'intervalle $[a, b]$ ($-\infty < a < b < +\infty$) si

$$\forall B \in \text{Bor}(\mathbb{R}), \quad P(X \in B) = P_X(B) = \frac{\lambda_1([a, b] \cap B)}{\lambda_1([a, b])}, \quad (2.24)$$

où λ_1 désigne la mesure de Lebesgue sur \mathbb{R} (en particulier $\lambda_1([a, b]) = b - a$). Notation : $X \sim \text{Unif}[a, b]$.

Calculons la fonction de répartition F en prenant $B =]-\infty, x]$ pour x quelconque dans (2.24).

$$F(x) = P_X(]-\infty, x]) = \frac{\lambda_1([a, b] \cap]-\infty, x])}{\lambda_1([a, b])} = \begin{cases} 0 & \text{si } -\infty < x < a; \\ \frac{x-a}{b-a} & \text{si } a \leq x < b; \\ 1 & \text{si } b \leq x < +\infty. \end{cases}$$

La fonction de répartition F est affine par morceaux, donc aussi C^1 par morceaux au sens de la proposition 2.29, avec dérivabilité sur $\mathbb{R} \setminus \{a, b\}$ (figure 2.6). La loi a donc une densité

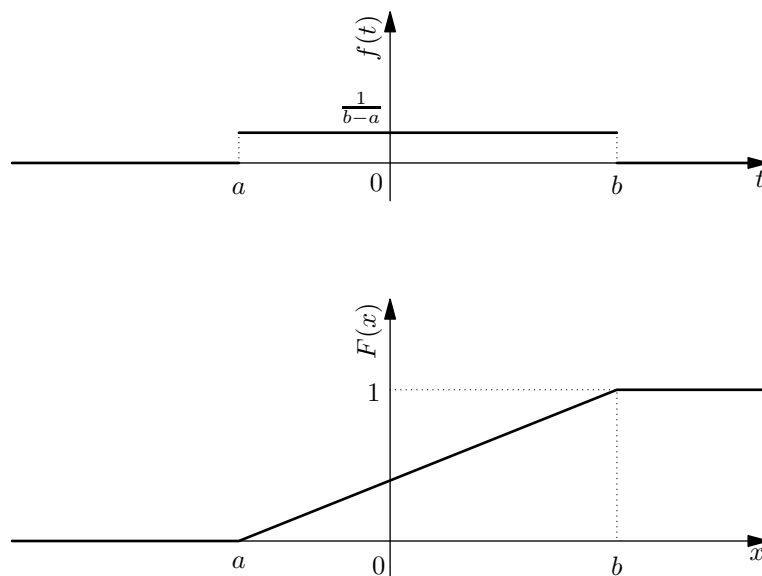


FIGURE 2.6 – Densité f et f.d.r. F de la loi $\text{Unif}[a, b]$

f qui s'obtient par dérivation de F , ce qui nous donne $f(t) = 0$ si $t < a$, $f(t) = \frac{1}{b-a}$ si $a < t < b$ et $f(t) = 0$ si $t > b$. On complète la définition de f en la prolongeant en a et b , par exemple en posant $f(a) = f(b) = \frac{1}{b-a}$. La loi uniforme sur $[a, b]$ admet donc pour densité

$$f = \frac{1}{b-a} \mathbf{1}_{[a, b]}.$$

Dans les calculs faisant intervenir la loi uniforme sur $[a, b]$, il est vivement conseillé d'utiliser chaque fois que c'est possible la formule (2.24) de préférence aux calculs d'intégrales de f .

Remarque 2.44. Comme $\lambda_1(\{a\}) = \lambda_1(\{b\}) = 0$, la loi uniforme sur $[a, b]$ est aussi la loi uniforme sur $]a, b]$, $[a, b[$ ou $]a, b[$.

Une des raisons de l'importance de la loi uniforme sur $[0, 1]$ est le théorème suivant.

Théorème 2.45. Si X est une variable aléatoire réelle de fonction de répartition continue strictement croissante F et si U est une variable aléatoire de loi uniforme sur $[0, 1]$, alors la variable aléatoire $Y := F^{-1}(U)$ a même loi que X .

Rappelons qu'avoir même loi que X ne signifie aucunement être égale à X . Ce théorème permet de réduire la simulation informatique de la loi de X à celle de U . Nous verrons ultérieurement que ce résultat s'étend à toutes les fonctions de répartition, sans hypothèse de continuité ni de croissance stricte, via une redéfinition de F^{-1} .

2.4.2 Lois exponentielles

Définition 2.46. Soit a un réel strictement positif. La variable aléatoire réelle X suit la loi exponentielle de paramètre a si elle admet pour densité

$$f(t) = ae^{-at}\mathbf{1}_{[0, +\infty[}(t).$$

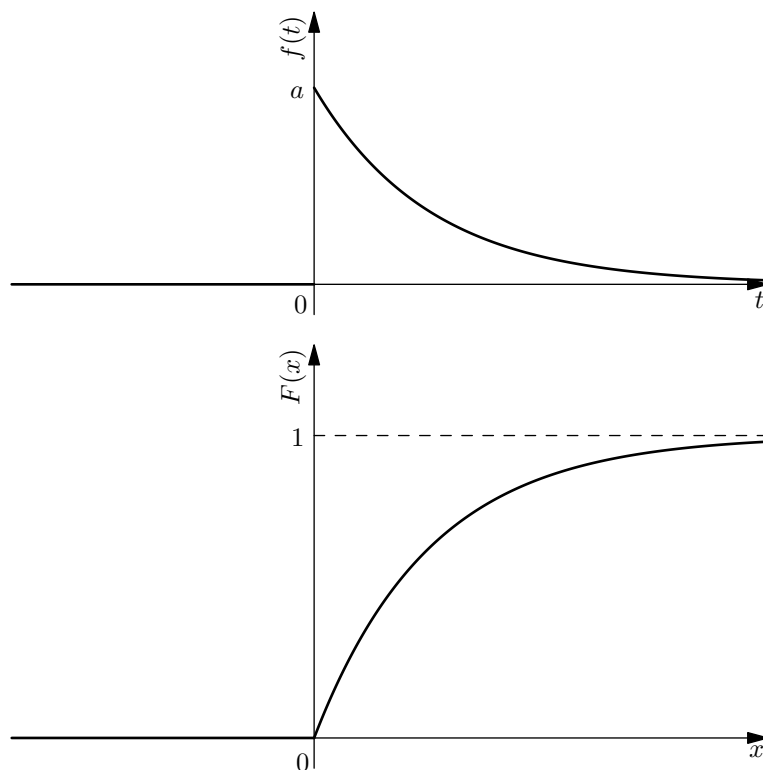


FIGURE 2.7 – Densité et f.d.r. de la loi $\text{Exp}(a)$

En pratique, plutôt que de travailler avec la fonction de répartition d'une loi exponentielle, il est plus commode d'utiliser la *fonction de survie* G :

$$G(x) = P(X > x) = 1 - F(x) = \begin{cases} 1 & \text{si } x \leq 0, \\ e^{-ax} & \text{si } x > 0. \end{cases}$$

Les lois exponentielles sont souvent choisies pour modéliser des temps d'attente : temps d'attente du prochain tremblement de terre, du prochain faux numéro sur une ligne téléphonique, de la prochaine désintégration d'un atome de radium, etc.

La raison de ce choix est la propriété *d'absence de mémoire* en temps continu qui caractérise la famille des lois exponentielles.

Théorème 2.47 (absence de mémoire).

i) Si la variable aléatoire X suit une loi exponentielle, elle vérifie la propriété d'absence de mémoire :

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad P(X > t + s \mid X > t) = P(X > s). \quad (2.25)$$

ii) Réciproquement si une variable aléatoire X vérifie (2.25), elle suit une loi exponentielle.



ICP pp. 196–198.

2.4.3 Lois gaussiennes

Ces lois jouent un rôle capital dans l'étude des lois limites de sommes de variables aléatoires indépendantes. Par exemple, selon le théorème de Moivre Laplace, si S_n suit la loi $\text{Bin}(n, p)$, alors pour tout $x \in \mathbb{R}$, $P(S_n - np \leq x\sqrt{np(1-p)})$ converge quand n tend vers l'infini vers $\Phi(x)$, où Φ est la f.d.r. de la loi gaussienne $\mathfrak{N}(0, 1)$.

Définition 2.48. On dit que la variable aléatoire X suit la loi gaussienne ou normale $\mathfrak{N}(m, \sigma)$ si elle a pour densité la fonction :

$$f_{m,\sigma} : \mathbb{R} \longrightarrow \mathbb{R}^+ \quad t \longmapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

La loi $\mathfrak{N}(0, 1)$ est appelée loi normale standard.

Tous les calculs de probabilités concernant une variable aléatoire de loi $\mathfrak{N}(m, \sigma)$ peuvent se ramener à des calculs sur une variable de loi normale standard.

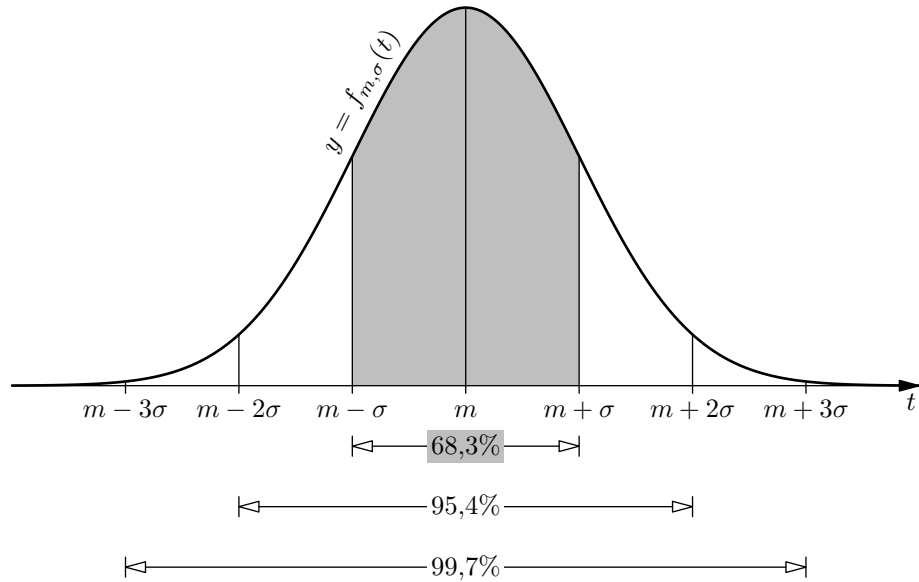
Proposition 2.49. Si la variable aléatoire X suit la loi $\mathfrak{N}(m, \sigma)$, $Y := (X - m)/\sigma$ suit la loi $\mathfrak{N}(0, 1)$. Autrement dit, toute v.a. gaussienne X de loi $\mathfrak{N}(m, \sigma)$ peut s'écrire $X = \sigma Y + m$ avec Y de loi $\mathfrak{N}(0, 1)$.

Remarque 2.50. On voit facilement que la famille des lois gaussiennes est *stable par transformations affines* : si X a pour loi $\mathfrak{N}(m, \sigma)$, alors pour tout $(\alpha, \beta) \in \mathbb{R}^* \times \mathbb{R}$, la v.a. $\alpha X + \beta$ est encore gaussienne, de loi $\mathfrak{N}(\alpha m + \beta, |\alpha|\sigma)$.

La figure 2.8 illustre la signification du paramètre de position m et du paramètre de dispersion σ pour la loi gaussienne $\mathfrak{N}(m, \sigma)$.

Cette concentration de pratiquement toute la probabilité dans l'intervalle $[m - 3\sigma, m + 3\sigma]$ permet l'utilisation des lois gaussiennes pour modéliser des grandeurs aléatoires qui *a priori* prennent leurs valeurs seulement dans un petit intervalle de \mathbb{R}^+ : taille, poids, ..., même si théoriquement une variable gaussienne peut prendre toute valeur entre $-\infty$ et $+\infty$.

Il n'existe pas d'expression d'une primitive de la densité gaussienne $f_{m,\sigma}$ à l'aide des fonctions usuelles. Les valeurs de la fonction de répartition Φ de $\mathfrak{N}(0, 1)$ sont tabulées. D'après la prop. 2.49, ceci suffit pour calculer numériquement n'importe quelle f.d.r. de loi gaussienne.

FIGURE 2.8 – Concentration de la loi $\mathfrak{N}(m, \sigma)$ autour de m

2.4.4 Lois de Cauchy

Définition 2.51. La variable aléatoire X suit la loi de Cauchy (ou loi de Cauchy de paramètres 0 et 1) si elle admet pour densité :

$$f(t) = \frac{1}{\pi(1+t^2)}.$$

Notation : $X \sim \text{Cau}(0, 1)$.

Cette loi est *symétrique*, ce qui signifie que X et $-X$ ont même loi, ceci résultant ici de la parité de f . La fonction de répartition F est donnée par :

$$F(x) = \int_{-\infty}^x \frac{dt}{\pi(1+t^2)} = \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan x \right),$$

où $\arctan x$ est l'unique réel $y \in]-\pi/2, \pi/2[$ tel que $\tan y = x$.

Si $Y = a + bX$, avec X de loi $\text{Cau}(0, 1)$, $a \in \mathbb{R}$ et $b \in \mathbb{R}_+^*$, on dit encore que Y suit une loi de Cauchy, de paramètres (a, b) , notation $Y \sim \text{Cau}(a, b)$. La densité est alors

$$f_{a,b}(t) = \frac{1}{\pi b} \frac{1}{1 + \left(\frac{t-a}{b}\right)^2}.$$

Chapitre 3

Espérance

L'espérance d'une variable aléatoire est, lorsqu'elle existe, la *moyenne des valeurs de cette variable, pondérées par leurs probabilités de réalisation*. On voit bien comment traduire cette définition informelle dans le cas d'une variable aléatoire discrète X en posant :

$$\mathbf{E}X := \sum_{x \in X(\Omega)} xP(X = x). \quad (3.1)$$

Cette formule n'a de sens que si la famille de réels $\{xP(X = x) ; x \in X(\Omega)\}$ est sommable, ce qui se traduit par la condition suivante pour l'existence de l'espérance de la v.a. discrète X :

$$\sum_{x \in X(\Omega)} |x|P(X = x) < +\infty. \quad (3.2)$$

La théorie de l'intégration au sens de Lebesgue permet de généraliser facilement cette définition en posant pour X variable aléatoire réelle définie sur l'espace probabilisable (Ω, \mathcal{F}) , muni de la mesure de probabilité P :

$$\mathbf{E}X = \int_{\Omega} X \, dP,$$

lorsque cette intégrale existe.

Sauf mention explicite du contraire, toutes les variables aléatoires considérées dans ce chapitre seront supposées définies sur le même espace probabilisable (Ω, \mathcal{F}) , muni de la mesure de probabilité P .

3.1 Définition et formules de calcul

Définition 3.1 (intégrabilité d'une v.a.). *La variable aléatoire X est dite intégrable (plus précisément P -intégrable) si l'intégrale $\int_{\Omega} |X| \, dP$ est finie.*

Cette intégrale de la v.a. *positive* X a toujours un sens comme élément de $\overline{\mathbb{R}}_+$. Nous la notons $\mathbf{E}|X|$. Avec cette notation, on peut dire que X est intégrable si et seulement si $\mathbf{E}|X| < +\infty$.

Remarque 3.2. Toute variable aléatoire X *bornée* sur Ω est intégrable (quelle que soit la mesure de probabilité P). En effet, s'il existe une constante réelle positive M telle que pour tout $\omega \in \Omega$, $|X(\omega)| \leq M$, alors

$$\int_{\Omega} |X| \, dP \leq \int_{\Omega} M \, dP = MP(\Omega) = M < +\infty.$$

On a la même conclusion pour une v.a. P -presque sûrement bornée (c.-à-d. telle que $P(|X| \leq M) = 1$ pour une certaine constante M), mais là, l'intégrabilité dépend de P car une v.a. peut être p.s. bornée relativement à une mesure P sur Ω et ne plus l'être relativement à une autre mesure de probabilité¹.

Définition 3.3. Soit X une v.a. réelle intégrable. On appelle espérance de X l'intégrale :

$$\mathbf{E}X := \int_{\Omega} X \, dP = \int_{\Omega} X(\omega) \, dP(\omega). \quad (3.3)$$

Insistons sur le fait que si X n'est pas une v.a. positive, la notation $\mathbf{E}X$ n'a de sens que si X est intégrable.

Exemple 3.4 (espérance d'une constante). Si X est une v.a. constante sur Ω (il existe un réel c tel que $\forall \omega \in \Omega, X(\omega) = c$), alors X est intégrable et $\mathbf{E}X = c$. En effet puisque $P(\Omega) < +\infty$, les constantes sont P -intégrables et $\int_{\Omega} X \, dP = \int_{\Omega} c \, dP = cP(\Omega) = c$. Ainsi,

$$\text{pour toute constante réelle } c, \quad \mathbf{E}c = c. \quad (3.4)$$

Exemple 3.5 (espérance d'une indicatrice). Soit A un évènement ($A \in \mathcal{F}$), son indicatrice 1_A est intégrable car bornée par 1 et on peut écrire

$$\mathbf{E}1_A = \int_{\Omega} 1_A \, dP = \int_A 1 \, dP + \int_{A^c} 0 \, dP = P(A).$$

Ainsi,

$$\forall A \in \mathcal{F}, \quad \mathbf{E}(1_A) = P(A). \quad (3.5)$$

Outre sa simplicité, cette formule a un intérêt intrinsèque : elle permet d'écrire toute probabilité d'évènement comme une espérance. C'est très utile en statistique lorsque l'on veut estimer une probabilité inconnue par une suite de fréquences observées via la loi des grands nombres.

Si on ajoute aux exemples 3.4 et 3.5 le cas où X est une combinaison linéaire finie d'indicatrices, on aura quasiment fait le tour des espérances que l'on peut calculer par une application directe de la formule (3.3). Cette dernière est très utile pour démontrer des propriétés générales de l'espérance (notamment la linéarité), elle l'est beaucoup moins pour le calcul explicite du réel $\mathbf{E}X$, car généralement on ne sait intégrer sur Ω que théoriquement. Nous allons donc ramener le calcul de $\mathbf{E}X$ à celui d'une intégrale sur \mathbb{R} par rapport à une mesure qui n'est autre que la loi de X . La clé de cette transformation est le théorème de transfert que nous « rappelons » ici.

Théorème 3.6 (de transfert). Soient $(\Omega_1, \mathcal{F}_1)$ et $(\Omega_2, \mathcal{F}_2)$ deux espaces mesurables et $\varphi : \Omega_1 \rightarrow \Omega_2$ une application mesurable \mathcal{F}_1 - \mathcal{F}_2 . Soit μ une mesure sur $(\Omega_1, \mathcal{F}_1)$ et $\nu := \mu \circ \varphi^{-1}$ sa mesure image sur $(\Omega_2, \mathcal{F}_2)$. Soit h une application $h : \Omega_2 \rightarrow \mathbb{K}$, $\mathbb{K} = \mathbb{R}, \mathbb{R} \text{ ou } \mathbb{C}$, mesurable \mathcal{F} -Bor(\mathbb{K}). Alors h est ν -intégrable si et seulement si $h \circ \varphi$ est μ -intégrable et dans ce cas,

$$\int_{\Omega_1} (h \circ \varphi) \, d\mu = \int_{\Omega_2} h \, d\nu. \quad (3.6)$$

1. Exemple : $\Omega = \mathbb{R}$, X étant l'identité sur \mathbb{R} est une v.a. P -p.s. bornée par 1 si P est la loi uniforme sur $[0, 1]$, mais elle n'est pas Q -p.s. bornée si Q est une loi exponentielle.

♣ IFP th. 4.13 p. 112 et th. 3.19 p. 98.

En appliquant ce théorème avec $\Omega_1 = \Omega$, $\mathcal{F}_1 = \mathcal{F}$, $\mu = P$, $\Omega_2 = \mathbb{R}$, $\mathcal{F}_2 = \text{Bor}(\mathbb{R})$, $\varphi = X$ et $h = \text{I} : t \mapsto t$, identité sur \mathbb{R} , nous transférons la condition d'existence et le calcul de $\mathbf{E}X$ de Ω à \mathbb{R} .

Proposition 3.7. *La variable aléatoire réelle X est intégrable si et seulement si*

$$\int_{\mathbb{R}} |t| dP_X(t) < +\infty, \quad (3.7)$$

où $P_X = P \circ X^{-1}$ désigne la loi de X sous P . Si cette condition est réalisée, elle a une espérance donnée par

$$\mathbf{E}X = \int_{\mathbb{R}} \text{I} dP_X = \int_{\mathbb{R}} t dP_X(t). \quad (3.8)$$

Remarque 3.8.

- a) Les formules (3.7) et (3.8) montrent que l'existence et la valeur de $\mathbf{E}X$ ne dépendent en fait que de P_X , c'est à dire de la loi de X . Au lieu d'espérance de X , on pourrait donc tout aussi bien parler de l'espérance de la loi de X (sous P).
- b) Il en résulte que si deux variables aléatoires X et Y définies sur le même (Ω, \mathcal{F}) ont même loi sous P et que l'une des deux est P -intégrable, l'autre l'est aussi et elles ont même espérance.
- c) Plus audacieux encore : si X est une v.a. sur (Ω, \mathcal{F}, P) et Y une v.a. sur $(\Omega', \mathcal{G}, Q)$ et si elles ont même loi (c.-à-d. $P_X = Q_Y$), alors $\mathbf{E}X = \mathbf{E}Y$, sous réserve de d'existence de l'une des deux espérances².
- d) Au fond, pour calculer $\mathbf{E}X$, on peut complètement oublier Ω et même X , seule compte la loi de X . Ceci amène à interpréter directement $\mathbf{E}X$ comme le *barycentre* ou centre de gravité de la mesure P_X vue comme une répartition de masse sur la droite réelle. Cette interprétation peut rendre de grands services pour prévoir ou contrôler le résultat de bien des calculs d'espérance.

Voyons maintenant comment se traduit la proposition 3.7 dans les deux cas usuels de lois discrètes ou à densité.

Corollaire 3.9 (espérance d'une loi discrète). *Supposons que la loi de X soit discrète et que $X_P(\Omega) = \{x \in \mathbb{R} ; P(X = x) > 0\} = \{x_i, i \in I\}$, avec I au plus dénombrable. Alors X est P -intégrable si et seulement si*

$$\sum_{i \in I} |x_i| P(X = x_i) < +\infty \quad (3.9)$$

et dans ce cas l'espérance de X sous P peut se calculer par

$$\mathbf{E}X = \sum_{i \in I} x_i P(X = x_i). \quad (3.10)$$

En utilisant la théorie des familles sommables, on pourrait réécrire (3.10) sous la forme $\mathbf{E}X = \sum_{x \in \mathbb{R}} x P(X = x)$, mais il faut bien comprendre que la sommabilité de cette famille (qui équivaut ici dans le cas discret à l'existence de $\mathbf{E}X$) implique que l'ensemble des termes non nuls dans cette « somme » est au plus dénombrable.

2. Pour être correct, il faudrait en fait noter $\mathbf{E}_P X = \mathbf{E}_Q Y$.

Exemple 3.10 (espérance d'un loi uniforme discrète). Supposons que la v.a. X soit de loi uniforme sur l'ensemble fini $\{x_1, \dots, x_n\}$ de cardinal n (pas d'ex-aequo dans la liste des x_i). Alors $I = \llbracket 1, n \rrbracket$ et pour tout $i \in I$, $P(X = x_i) = 1/n$, d'où

$$\mathbf{E}X = \sum_{i=1}^n x_i \frac{1}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Ceci peut s'énoncer : *L'espérance de la loi uniforme sur un ensemble de n valeurs réelles distinctes est égale à la moyenne arithmétique de ces valeurs.* C'est conforme à l'intuition barycentrique évoquée à la remarque 3.8 d).

Corollaire 3.11 (espérance d'une loi à densité). *Soit X une variable aléatoire réelle de densité f . Alors X est intégrable si et seulement si*

$$\int_{-\infty}^{+\infty} |t|f(t) dt < +\infty \quad (3.11)$$

et dans ce cas,

$$\mathbf{E}X = \int_{-\infty}^{+\infty} tf(t) dt. \quad (3.12)$$

♣ IFP prop. 4.12 p. 112 et th. 3.17 p. 95.

Exemple 3.12 (espérance de la loi uniforme sur $[a, b]$). Avant de se lancer dans le calcul, on remarque que la loi uniforme sur le segment $[a, b]$ peut être vue comme une répartition de masse homogène sur $[a, b]$. On peut ainsi voir $[a, b]$ comme une tige homogène de masse 1. À l'évidence, son centre de gravité est le milieu de $[a, b]$. On s'attend donc à trouver $(a + b)/2$ comme valeur de l'espérance. Vérifions le en appliquant la formule (3.12). Si X suit la loi uniforme sur $[a, b]$, sa loi admet pour densité $f = (b - a)^{-1} \mathbf{1}_{[a, b]}$, d'où

$$\mathbf{E}X = \int_{-\infty}^{+\infty} \frac{1}{b - a} \mathbf{1}_{[a, b]}(t) dt = \frac{1}{b - a} \int_a^b t dt = \frac{1}{b - a} \left[\frac{t^2}{2} \right]_a^b = \frac{b^2 - a^2}{2(b - a)} = \frac{a + b}{2}.$$

Nous allons voir maintenant des formules exprimant l'espérance à l'aide de la fonction de répartition ou de la fonction de survie. Elles sont très utiles quand la loi de X n'est ni discrète ni à densité.

Proposition 3.13. *Si X est une variable aléatoire positive sur (Ω, \mathcal{F}) muni de la probabilité P , l'espérance de X sous P vérifie*

$$\mathbf{E}X = \int_0^{+\infty} P(X > t) dt = \int_0^{+\infty} (1 - F(t)) dt, \quad (3.13)$$

où F désigne la fonction de répartition de X .

Les égalités dans (3.13) doivent se comprendre comme égalités dans $\overline{\mathbb{R}}_+$, on ne suppose pas ici que X est intégrable.

Preuve. On exprime d'abord $\int_0^{+\infty} P(X > t) dt$ comme une intégrale double :

$$\int_0^{+\infty} P(X > t) dt = \int_{\mathbb{R}_+} P_X([t, +\infty[) d\lambda(t) = \int_{\mathbb{R}_+} \left\{ \int_{\mathbb{R}} \mathbf{1}_{[t, +\infty[}(x) dP_X(x) \right\} d\lambda(t).$$

En remarquant que pour tous réels x, t , $\mathbf{1}_{]t, +\infty[}(x) = \mathbf{1}_{]-\infty, x[}(t)$ et en appliquant le théorème de Fubini-Tonelli (IFP, th. 5.11 p. 154), on obtient :

$$\begin{aligned} \int_0^{+\infty} P(X > t) dt &= \int_{\mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{]t, +\infty[}(x) d\lambda(t) dP_X(x) = \int_{\mathbb{R}} \left\{ \int_{\mathbb{R}_+} \mathbf{1}_{]-\infty, x[}(t) d\lambda(t) \right\} dP_X(x) \\ &= \int_{\mathbb{R}} \lambda([0, x[) dP_X(x) \\ &= \int_{\mathbb{R}} x dP_X(x) = \mathbf{E}X, \end{aligned}$$

d'après (3.8). □

En séparant partie positive et partie négative d'une variables aléatoire réelle X :

$$X = X^+ - X^-, \quad \text{où} \quad X^+ := \max(X, 0), \quad X^- := \max(-X, 0),$$

et en se rappelant que par définition, $\int_{\Omega} X dP = \int_{\Omega} X^+ dP - \int_{\Omega} X^- dP$, sous réserve que ces deux dernières intégrales soient *finies*³, on peut généraliser la formule (3.13) au cas des variables aléatoires réelles, à condition qu'elles soient *intégrables*. On aboutit ainsi au résultat suivant.

Proposition 3.14. *Si X est une variable aléatoire intégrable sur (Ω, \mathcal{F}) muni de la probabilité P , l'espérance de X sous P vérifie*

$$\mathbf{E}X = \int_0^{+\infty} P(X > t) dt - \int_{-\infty}^0 P(X < t) dt = \int_0^{+\infty} (1 - F(t)) dt - \int_{-\infty}^0 F(t) dt, \quad (3.14)$$

où F désigne la fonction de répartition de X .

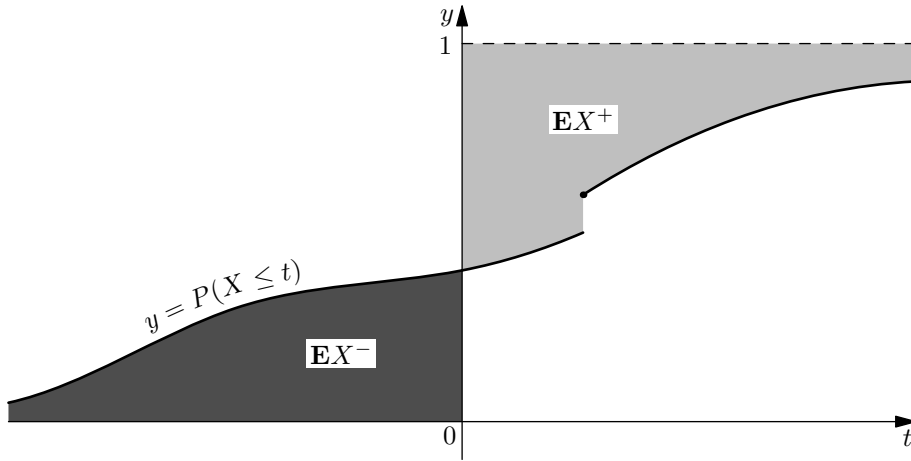


FIGURE 3.1 – Espérance d'une v.a. réelle $\mathbf{E}X := \mathbf{E}(X^+) - \mathbf{E}(X^-)$

Preuve. On remarque d'abord que pour tout $t > 0$, les événements $\{X > t\}$ et $\{X^+ > t\}$ sont identiques, d'où

$$\mathbf{E}X^+ = \int_0^{+\infty} P(X^+ > t) dt = \int_0^{+\infty} P(X > t) dt.$$

3. Cette formule est utilisée dans la construction de l'intégrale abstraite sur Ω , elle ne nécessite pas que l'on ait déjà démontré la linéarité de l'intégrale, cf. IFP chapitre 4.

De même, pour tout $s > 0$, $\{X^- > s\} = \{X < -s\}$, d'où

$$\mathbf{E}X^- = \int_0^{+\infty} P(X^- > s) ds = \int_0^{+\infty} P(X < -s) ds = \int_{-\infty}^0 P(X < t) dt,$$

en effectuant le changement de variable $t = -s$ (au sens Riemann ou Lebesgue, les deux sont valides ici). À ce stade, nous avons justifié la première égalité dans (3.14). Pour la deuxième, il suffit de remarquer que la fonction $t \mapsto P(X < t)$ diffère de F seulement aux éventuels points de discontinuité de F . Comme leur ensemble est au plus dénombrable, les deux fonctions sont égales λ -presque partout, donc

$$\int_{-\infty}^0 P(X < t) dt = \int_{-\infty}^0 P(X \leq t) dt = \int_{-\infty}^0 F(t) dt,$$

ce qui complète la vérification de (3.14). \square

Nous avons vu ci-dessus qu'une condition suffisante pour qu'une variable aléatoire soit intégrable est qu'elle soit bornée. La formule (3.13) appliquée à la v.a. positive $|X|$ nous fournit une C.N.S. d'intégrabilité. On en déduit immédiatement les conditions suffisantes suivantes : si la variable aléatoire réelle X , vérifie $P(|X| > t) \leq Ct^{-\alpha}$ pour un certain $\alpha > 1$ et tout $t \geq t_0 > 0$, ou si $P(|X| > t) \leq t^{-1}(\ln t)^{-\beta}$ pour un $\beta > 1$ et tout $t \geq t_0 > 0$, alors X est intégrable. Réciproquement, l'intégrabilité de X nous donne un renseignement sur la vitesse de convergence vers 0 de $P(|X| > t)$ quand t tend vers $+\infty$, grâce à l'inégalité suivante due à Markov.

Proposition 3.15 (inégalité de Markov). *Si X est une variable aléatoire positive,*

$$\forall x > 0, \quad P(X \geq x) \leq \frac{\mathbf{E}X}{x}. \quad (3.15)$$

Même si l'inégalité est vraie dans $\overline{\mathbb{R}}_+$ sans hypothèse d'intégrabilité de X , elle n'a d'intérêt que lorsque le second membre est inférieur à 1, c'est-à-dire lorsque $\mathbf{E}X < +\infty$ et $x > \mathbf{E}X$.

Preuve. Une preuve « muette » est donnée par la figure 3.2. \square

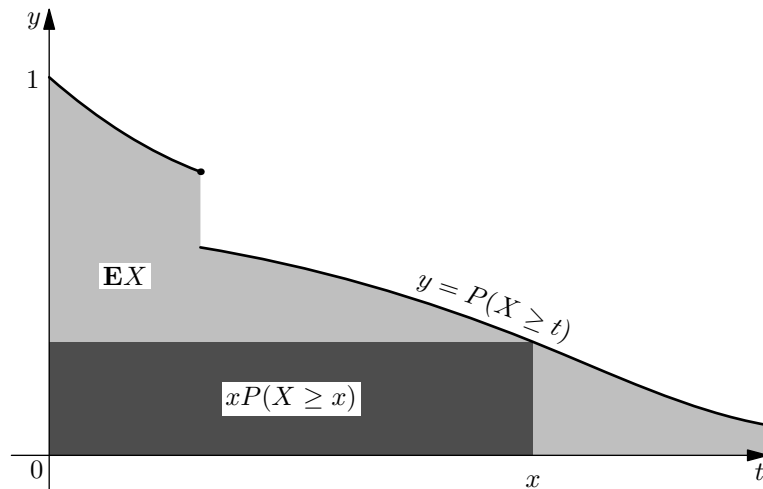


FIGURE 3.2 – Inégalité de Markov : $xP(X \geq x) \leq \int_0^{+\infty} P(X \geq t) dt = \mathbf{E}X$.



Deux autres preuves de l'inégalité de Markov sont proposées dans PVIR p. 261–262.

Corollaire 3.16. *Si X est une variable aléatoire positive, on a l'équivalence*

$$\mathbf{E}X = 0 \Leftrightarrow P(X = 0) = 1,$$

autrement dit, $\mathbf{E}X$ est nulle si et seulement si X est presque sûrement nulle.

Preuve. Supposons d'abord que $\mathbf{E}X = 0$. La v.a. X étant positive, l'égalité $P(X = 0) = 1$ équivaut à $P(X > 0) = 0$. Introduisons la suite des événements $A_n := \{X \geq 1/n\}$, $n \in \mathbb{N}^*$. Cette suite est croissante de réunion $A := \{X > 0\}$. Par continuité séquentielle croissante de P , $P(A) = \lim_{n \rightarrow +\infty} P(A_n)$. Or l'inégalité de Markov appliquée avec $x = 1/n$ nous montre que pour tout $n \in \mathbb{N}^*$, $0 \leq P(A_n) \leq n\mathbf{E}X = 0$. Ainsi $P(A_n) = 0$ pour tout n et $P(A) = 0$ comme limite de la suite nulle.

Réciproquement, si $P(X = 0) = 1$, alors pour tout $t \geq 0$, $P(X > t) = 0$, d'où $\mathbf{E}X = \int_0^{+\infty} 0 \, dt = 0$. \square

3.2 Propriétés de l'espérance

Les deux propositions suivantes nous donnent des propriétés de l'espérance qui découlent directement des propriétés de l'intégrale sur Ω relativement à P .

Proposition 3.17 (Linéarité de l'espérance).

a) *L'espérance des v.a. réelles intégrables est additive : si X et Y v.a. réelles définies sur le même (Ω, \mathcal{F}, P) sont intégrables, alors $X + Y$ l'est aussi et*

$$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y. \quad (3.16)$$

b) *Si X est intégrable, cX l'est aussi pour toute constante réelle c et*

$$\mathbf{E}(cX) = c\mathbf{E}X. \quad (3.17)$$

Voici deux exemples où l'utilisation de la linéarité apporte une simplification significative au calcul d'espérance.

Exemple 3.18 (espérance d'une loi binomiale). Soit X une v.a. de loi binomiale $\text{Bin}(n, p)$. Elle est évidemment intégrable puisque positive et bornée par n . Elle a même loi et donc même espérance que la v.a.

$$S_n = \sum_{i=1}^n \mathbf{1}_{A_i},$$

définie sur un espace probabilisé $(\Omega', \mathcal{F}', P')$, où les $A_i \in \mathcal{F}'$ sont n événements mutuellement indépendants et de même probabilité p . L'additivité vue ci-dessus pour l'espérance de la somme de deux v.a. s'étendant par une récurrence immédiate aux sommes d'un nombre fini de v.a. intégrables, nous avons donc

$$\mathbf{E}X = \mathbf{E}S_n = \sum_{i=1}^n \mathbf{E}\mathbf{1}_{A_i} = \sum_{i=1}^n P(A_i) = np.$$

En résumé :

$$\text{Si } X \sim \text{Bin}(n, p), \quad \mathbf{E}X = np. \quad (3.18)$$

Exemple 3.19 (espérance d'une loi hypergéométrique). Soit X une v.a. de loi hypergéométrique $\text{Hypg}(N, M, n)$. Elle est intégrable puisque positive et bornée par n . Rappelons que X a même loi que la variable Y définie comme suit. On considère une population totale d'effectif N et une sous-population d'intérêt d'effectif M . On prélève sans remise un échantillon de n individus dans la population totale. Y est le nombre d'individus de la sous-population d'intérêt dans cet échantillon. Puisque X et Y ont même loi, $\mathbf{E}Y = \mathbf{E}X$. Pour calculer $\mathbf{E}Y$, nous allons l'exprimer comme une somme d'indicatrices. Pour cela, numérotions de 1 à M les individus de la sous-population d'intérêt et notons

$$A_i = \{\text{l'individu n}^\circ i \text{ est pris dans l'échantillon}\}.$$

On remarque ensuite que

$$Y = \sum_{i=1}^M \mathbf{1}_{A_i}, \quad \text{d'où} \quad \mathbf{E}Y = \sum_{i=1}^M P(A_i).$$

Pour calculer $P(A_i)$, on utilise l'hypothèse d'équiprobabilité de choix de tous les échantillons possibles (il y en a C_N^n) et on dénombre les échantillons de taille n contenant l'individu n° i : il y en a C_{N-1}^{n-1} . Ainsi

$$\forall i \in \llbracket 1, M \rrbracket, \quad P(A_i) = \frac{C_{N-1}^{n-1}}{C_N^n} = \frac{(N-1)!}{(n-1)!(N-n)!} \times \frac{n!(N-n)!}{N!} = \frac{n}{N}.$$

En résumé :

$$\text{Si } X \sim \text{Hypg}(N, M, n), \quad \mathbf{E}X = n \frac{M}{N}. \quad (3.19)$$

On remarque que si les individus de l'échantillon avaient été prélevés avec remise, la loi du nombre d'individus de la sous-population d'intérêt dans l'échantillon serait la binomiale $\text{Bin}(n, p)$ avec $p = M/N$ et d'espérance $np = nM/N = \mathbf{E}Y$. Ainsi, que les tirages aient lieu avec ou sans remise, l'espérance est la même.

Proposition 3.20 (espérance et ordre).

- a) L'espérance des v.a. réelles intégrables est croissante : si X et Y v.a. réelles définies sur le même (Ω, \mathcal{F}, P) sont intégrables et vérifient $X \leq Y$ p.s., alors $\mathbf{E}X \leq \mathbf{E}Y$.
- b) Si X est intégrable, $|X|$ l'est aussi et

$$|\mathbf{E}X| \leq \mathbf{E}|X|. \quad (3.20)$$

Nous allons voir maintenant les théorèmes d'interversion limite-espérance qui sont hérités de la théorie de l'intégrale de Lebesgue. Cela nécessite une mise au point préalable sur la notion de variable aléatoire positive. Dans ce cours, une variable aléatoire positive X est une application de Ω dans \mathbb{R}_+ , mesurable \mathcal{F} - $\text{Bor}(\mathbb{R}_+)$. Donc en particulier, $X(\omega)$ ne peut valoir $+\infty$ pour aucun $\omega \in \Omega$. L'ennui c'est que les théorèmes d'intégration pour les fonctions mesurables positives concernent généralement des applications de Ω dans $\overline{\mathbb{R}}_+$, mesurables \mathcal{F} - $\text{Bor}(\overline{\mathbb{R}}_+)$. Pour éviter la confusion, nous introduisons la définition suivante.

Définition 3.21. Soit (Ω, \mathcal{F}) un espace probabilisable. On dit que X est une variable aléatoire sur cet espace, à valeurs dans $\overline{\mathbb{R}}_+$ (en abrégé X est une v.a. dans $\overline{\mathbb{R}}_+$) si c'est une application $\Omega \rightarrow \overline{\mathbb{R}}_+$, mesurable \mathcal{F} - $\text{Bor}(\overline{\mathbb{R}}_+)$.

Une variable aléatoire dans $\overline{\mathbb{R}}_+$ a toujours une espérance définie dans $\overline{\mathbb{R}}_+$ par $\mathbf{E}X = \int_{\Omega} X \, dP$.

♣ Pour la mesurabilité relativement à $\text{Bor}(\overline{\mathbb{R}}_+)$, voir IFP chapitre 2.

Remarque 3.22. Nous avons donc maintenant 3 sortes de variables aléatoires :

- les variables aléatoires réelles ;
- les variables aléatoires positives ;
- les variables aléatoires dans $\overline{\mathbb{R}}_+$.

Toute variable aléatoire positive est aussi une variable aléatoire réelle et une variable aléatoire dans $\overline{\mathbb{R}}_+$. Par contre, une variable aléatoire X dans $\overline{\mathbb{R}}_+$ n'est une v.a. positive que si $X^{-1}(\{+\infty\}) = \emptyset$, autrement dit si pour tout $\omega \in \Omega$, $X(\omega) < +\infty$.

Remarque 3.23. Soit X une v.a. dans $\overline{\mathbb{R}}_+$ telle que $P(X < +\infty) = 1$. Alors X est égale p.s. à une variable aléatoire positive X' . En effet, notons $\Omega' := \{\omega \in \Omega ; X(\omega) < +\infty\}$, alors $\Omega' \in \mathcal{F}$ et on peut prendre X' égale à X sur Ω' et à 0 sur son complémentaire.

Remarque 3.24. Si X est une v.a. dans $\overline{\mathbb{R}}_+$ telle que $\mathbf{E}X < +\infty$, alors $P(X < +\infty) = 1$, cf. IFP prop. 4.19 p. 116.

Théorème 3.25 (de Beppo Levi). *Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires dans $\overline{\mathbb{R}}_+$, définie sur l'espace probabilisé (Ω, \mathcal{F}, P) et croissante, c'est-à-dire : pour tout n , $X_n \leq X_{n+1}$. Notons X la v.a. dans $\overline{\mathbb{R}}_+$ définie par $X(\omega) = \lim_{n \rightarrow +\infty} X_n(\omega)$ pour tout $\omega \in \Omega$. Alors la suite $(\mathbf{E}X_n)_{n \geq 1}$ converge en croissant vers $\mathbf{E}X$ dans $\overline{\mathbb{R}}_+$.*

Notons que dans ce théorème, même si on prend pour X_n des v.a. positives au lieu de v.a. dans $\overline{\mathbb{R}}_+$, X est toujours une v.a. dans $\overline{\mathbb{R}}_+$ et on ne peut pas affirmer en général que X est une v.a. positive. Bien sûr, si $\lim_{n \rightarrow +\infty} \mathbf{E}X_n$ est finie, $\mathbf{E}X$ est finie et par les remarques 3.24 et 3.23, X est égale p.s. à une v.a. positive.

Corollaire 3.26 (intersion série-espérance pour les v.a. dans $\overline{\mathbb{R}}_+$). *Pour toute suite $(X_k)_{k \in \mathbb{N}}$ de v.a. dans $\overline{\mathbb{R}}_+$ définies sur le même espace probabilisé,*

$$\mathbf{E} \left(\sum_{k=0}^{+\infty} X_k \right) = \sum_{k=0}^{+\infty} \mathbf{E}X_k, \quad \text{égalité dans } \overline{\mathbb{R}}_+.$$

Si les X_k sont des v.a. positives et si $\sum_{k=0}^{+\infty} \mathbf{E}X_k$ converge dans \mathbb{R}_+ , alors $\sum_{k=0}^{+\infty} X_k$ est finie p.s., autrement dit, la série converge p.s. dans \mathbb{R}_+ .

Corollaire 3.27 (lemme de Fatou pour les espérances). *Pour toute suite $(X_k)_{k \in \mathbb{N}}$ de v.a. positives définies sur le même espace probabilisé,*

$$\mathbf{E} \liminf_{n \rightarrow +\infty} X_n \leq \liminf_{n \rightarrow +\infty} \mathbf{E}X_n$$

♣ Sur le th. de B. Levi et ses corollaires, IFP pp. 90–94.

Théorème 3.28 (convergence dominée). *On suppose que les variables aléatoires réelles Y_n ($n \geq 0$), Y et Z définies sur le même espace probabilisé (Ω, \mathcal{F}, P) vérifient*

- a) Y_n converge presque-sûrement vers Y ;
- b) pour tout $n \geq 1$, $|Y_n| \leq Z$ p.s. ;
- c) Z est intégrable.

Dans ces conditions,

1. les Y_n et Y sont intégrables ;
2. Y_n converge vers Y au sens L^1 : $\mathbf{E}|Y_n - Y| \rightarrow 0$;
3. on a l'interversion limite espérance : $\lim_{n \rightarrow +\infty} \mathbf{E}Y_n = \mathbf{E}Y$.



IFP th. 4.39 p. 124.

3.3 Moments

On étudie dans cette section les $\mathbf{E}h(X)$, où h est une fonction réelle. Pour que l'expression $\mathbf{E}h(X)$ ait un sens, il est nécessaire que $Y := h(X)$ soit une variable aléatoire réelle. Cette condition sera réalisée si $h : \mathbb{R} \rightarrow \mathbb{R}$ est *borélienne*, c'est-à-dire si $B' = h^{-1}(B) \in \text{Bor}(\mathbb{R})$ pour tout $B \in \text{Bor}(\mathbb{R})$. Alors en effet,

$$Y^{-1}(B) = \{\omega \in \Omega ; h(X(\omega)) \in B\} = \{\omega \in \Omega ; X(\omega) \in h^{-1}(B)\} = X^{-1}(B') \in \mathcal{F},$$

en raison de la mesurabilité $\mathcal{F} - \text{Bor}(\mathbb{R})$ de la variable aléatoire X . Comme le borélien B ci-dessus est quelconque, ceci montre que Y est elle aussi mesurable $\mathcal{F} - \text{Bor}(\mathbb{R})$ et est donc bien une variable aléatoire sur (Ω, \mathcal{F}) . Ainsi si h est borélienne, $\mathbf{E}|h(X)|$ existe toujours comme élément de $\overline{\mathbb{R}}_+$ et si $\mathbf{E}|h(X)| < +\infty$, $\mathbf{E}h(X)$ existe (et $\mathbf{E}h(X) \in \mathbb{R}$). On pourra désigner $\mathbf{E}|h(X)|$ et $\mathbf{E}h(X)$ respectivement par l'appellation *h -moment absolu* de X et *h -moment* de X . Bien entendu si h est borélienne positive, *h -moment absolu* et *h -moment* sont confondus et ce dernier existe toujours dans $\overline{\mathbb{R}}_+$. Nous utiliserons aussi l'appellation générique de *moments fonctionnels* pour désigner les *h -moments*⁴.

Le cas le plus utile est celui où h est une fonction puissance, $h(x) = x^r$, on parle alors de *moment d'ordre r de X* .

Définition 3.29. Soit r un réel positif. On appelle *moment absolu d'ordre r de la variable aléatoire réelle X* la quantité $\mathbf{E}(|X|^r)$, élément de $\overline{\mathbb{R}}_+$. Si r est entier et X^r intégrable, donc si le *moment absolu d'ordre r de X* est fini, on appelle *moment d'ordre r de X* le réel $\mathbf{E}(X^r)$. On notera $\mathbf{E}|X|^r$ pour $\mathbf{E}(|X|^r)$ et $\mathbf{E}X^r$ pour $\mathbf{E}(X^r)$ en prenant garde de ne pas confondre ces quantités avec $(\mathbf{E}|X|)^r$ et $(\mathbf{E}X)^r$ respectivement.


Remarquons qu'on ne définit pas le *moment d'ordre r non entier* pour X v.a. réelle, même si $\mathbf{E}|X|^r < +\infty$. En effet dans ce cas, X^r n'est pas définie sur l'évènement $\{X < 0\}$ et si $P(X < 0) \neq 0$, X^r ne peut être égale presque sûrement à une v.a. définie sur tout Ω . Bien entendu, si X est une v.a. positive, $\mathbf{E}X^r$ existe toujours dans $\overline{\mathbb{R}}_+$.

Théorème 3.30 (inégalité de Hölder). Soient $p \geq 1$ et $q \geq 1$ deux réels exposants conjugués : $\frac{1}{p} + \frac{1}{q} = 1$. Pour toutes variables aléatoires réelles X et Y définies sur le même espace probabilisé,

$$\mathbf{E}|XY| \leq (\mathbf{E}|X|^p)^{1/p} (\mathbf{E}|Y|^q)^{1/q}. \quad (3.21)$$

Corollaire 3.31 (comparaison des normes L^p des v.a.). Pour toute v.a. réelle X , l'application $[1, +\infty[\rightarrow \overline{\mathbb{R}}_+$, $p \mapsto (\mathbf{E}|X|^p)^{1/p}$ est croissante.

4. Ces appellations ne sont pas standard, nous les adoptons par confort de rédaction.

 IFP, preuve du th. 6.18 p. 200, avec $\mu(\Omega) = 1$.

L'existence d'un moment absolu d'ordre r fini donne un renseignement sur la vitesse de convergence vers 0 de $P(|X| \geq t)$ quand t tend vers $+\infty$. On a alors $P(|X| \geq t) = O(t^{-r})$ par le corollaire suivant de l'inégalité de Markov.

Proposition 3.32 (inégalité de Markov avec moment). *Pour toute variable aléatoire réelle X , pour tout réel $r > 0$,*

$$\forall t > 0, \quad P(|X| \geq t) \leq \frac{\mathbf{E}|X|^r}{t^r}. \quad (3.22)$$

Bien entendu, cette inégalité n'a d'intérêt que si $\mathbf{E}|X|^r < +\infty$ et $t^{-r}\mathbf{E}|X|^r < 1$.

Preuve. Il suffit de noter que $P(|X| \geq t) \leq P(|X|^r \geq t^r) \leq t^{-r}\mathbf{E}|X|^r$, en utilisant la croissance de l'application $\mathbb{R}_+ \rightarrow \mathbb{R}_+$, $x \mapsto x^r$ et l'inégalité de Markov pour la v.a. positive $|X|^r$. \square

Proposition 3.33 (moments d'une v.a. discrète). *Si X est une variable aléatoire discrète, pour tout réel $r \geq 0$,*

$$\mathbf{E}|X|^r = \sum_{x_i \in X_P(\Omega)} |x_i|^r P(X = x_i). \quad (3.23)$$

De plus si r est entier et $\mathbf{E}|X|^r < +\infty$,

$$\mathbf{E}X^r = \sum_{x_i \in X_P(\Omega)} x_i^r P(X = x_i). \quad (3.24)$$

Cette proposition n'est qu'une application de la formule de calcul de $\mathbf{E}h(X)$ pour X discrète qui est établie dans toute sa généralité à la proposition 3.35 ci-dessous.

Proposition 3.34 (moments d'une v.a. à densité). *Si X est une variable aléatoire réelle à densité f , pour tout réel $r \geq 0$,*

$$\mathbf{E}|X|^r = \int_{-\infty}^{+\infty} |x|^r f(x) dx. \quad (3.25)$$

Si de plus r est entier et $\mathbf{E}|X|^r < +\infty$,

$$\mathbf{E}X^r = \int_{-\infty}^{+\infty} x^r f(x) dx. \quad (3.26)$$

Là aussi, il s'agit d'un cas particulier d'une formule générale pour $\mathbf{E}h(X)$ lorsque X est à densité, donnée ci-dessous.

Proposition 3.35 (calcul de $\mathbf{E}h(X)$, X discrète). *Si X est une variable aléatoire discrète et $h : \mathbb{R} \rightarrow \mathbb{R}$ une application quelconque,*

$$\mathbf{E}|h(X)| = \sum_{x_i \in X_P(\Omega)} |h(x_i)| P(X = x_i). \quad (3.27)$$

De plus, si $\mathbf{E}|h(X)| < +\infty$,

$$\mathbf{E}h(X) = \sum_{x_i \in X_P(\Omega)} h(x_i) P(X = x_i). \quad (3.28)$$



IFP Cor. 4.16, pp. 113–114.

Proposition 3.36 (calcul de $\mathbf{E}h(X)$, X à densité). *Si X est une variable aléatoire de densité f et $h : \mathbb{R} \rightarrow \mathbb{R}$ une application borélienne,*

$$\mathbf{E}|h(X)| = \int_{-\infty}^{+\infty} |h(x)|f(x) \, dx. \quad (3.29)$$

De plus, si $\mathbf{E}|h(X)| < +\infty$,

$$\mathbf{E}h(X) = \int_{-\infty}^{+\infty} h(x)f(x) \, dx. \quad (3.30)$$

Preuve. En appliquant le théorème de transfert (th. 3.6) avec $\Omega_1 = \Omega$, $\Omega_2 = \mathbb{R}$, $\varphi = X$, $\mu = P$, $\nu = P_X$, on obtient

$$\mathbf{E}|h(X)| = \int_{\mathbb{R}} |h(x)| \, dP_X(x).$$

Par la formule d'intégration par rapport à une mesure à densité (IFP prop. 4.12 p. 112), cette dernière intégrale vaut $\int_{-\infty}^{+\infty} |h(x)|f(x) \, dx$. On procède de même pour (3.30), une fois garantie l'intégrabilité de la v.a. $h(X)$. \square

Nous présentons maintenant une formule permettant de calculer des moments fonctionnels à partir de la fonction de survie. Son intérêt est de permettre un tel calcul pour des variables aléatoires qui ne sont ni discrètes ni à densité.

Proposition 3.37. *Soient X une variable aléatoire positive et g une application continue croissante $\mathbb{R}_+ \rightarrow \mathbb{R}_+$, de classe C^1 sur \mathbb{R}_+^* . Alors*

$$\mathbf{E}g(X) = g(0) + \int_0^{+\infty} g'(s)P(X > s) \, ds, \quad (3.31)$$

l'égalité ayant lieu dans $\overline{\mathbb{R}}_+$.

La preuve, analogue à celle de la proposition 3.13, est laissée en exercice.

Corollaire 3.38. *Pour toute variable aléatoire réelle X et tout réel $p > 0$,*

$$\mathbf{E}|X|^p = \int_0^{+\infty} ps^{p-1}P(|X| > s) \, ds, \quad (3.32)$$

l'égalité ayant lieu dans $\overline{\mathbb{R}}_+$.

Le h -moment $\mathbf{E}h(X)$ pour $h : x \mapsto (x - \mathbf{E}X)^2$ occupe une place particulière dans la théorie des probabilités.

Définition 3.39 (variance et écart type). *Si X est de carré intégrable ($\mathbf{E}X^2 < +\infty$), on appelle variance de X le réel positif noté $\text{Var } X$ défini par*

$$\text{Var } X := \mathbf{E}(X - \mathbf{E}X)^2. \quad (3.33)$$

On appelle alors écart type de X le réel $\sigma(X) := (\text{Var } X)^{1/2}$.

Remarquons que si $\mathbf{E}X^2$ est fini, $\mathbf{E}|X|$ l'est aussi (puisque $\mathbf{E}|X| \leq (\mathbf{E}X^2)^{1/2}$ par le corollaire 3.31), donc $\mathbf{E}X$ est bien défini⁵. De plus $(X - \mathbf{E}X)^2 = X^2 - 2(\mathbf{E}X)X + (\mathbf{E}X)^2$ apparaît alors comme une combinaison linéaire de trois variables⁶ intégrables, donc est aussi intégrable. Ainsi la v.a. positive $(X - \mathbf{E}X)^2$ est intégrable et $\mathbf{E}(X - \mathbf{E}X)^2$ est bien un réel positif, ce qui justifie la définition 3.39. Notons aussi que si X représente une grandeur physique, X , $\mathbf{E}X$ et $\sigma(X)$ ont la même unité, mais pas $\text{Var } X$.

Lorsqu'elle existe, la variance de X est une façon de mesurer la *dispersion* de la loi de X autour de l'espérance. Les raisons de l'importance de la variance apparaîtront ultérieurement dans ce cours (inégalité de Tchebycheff, théorème limite central). L'application des propositions 3.35 et 3.36 nous donne (sous réserve d'intégrabilité de X^2) les formules respectives :

$$\text{Var } X = \sum_{x_i \in X(\Omega)} (x_i - \mathbf{E}X)^2 P(X = x_i) \quad (\text{cas } X \text{ discrète}), \quad (3.34)$$

$$\text{Var } X = \int_{-\infty}^{+\infty} (x - \mathbf{E}X)^2 f(x) dx \quad (\text{cas } X \text{ à densité } f). \quad (3.35)$$

Dans la pratique, ces formules sont rarement utilisées, on leur préfère la formule suivante qui simplifie les calculs.

Proposition 3.40 (formule de Koenig pour la variance). *Si la variable aléatoire X est de carré intégrable,*

$$\text{Var } X = \mathbf{E}X^2 - (\mathbf{E}X)^2. \quad (3.36)$$

Preuve. Rappelons que nous notons $\mathbf{E}X^2$ pour $\mathbf{E}(X^2)$ et que le second membre de la formule ci-dessus n'est donc généralement pas nul. On pose $c = \mathbf{E}X$.

$$\begin{aligned} \text{Var } X &= \mathbf{E}(X - c)^2 = \mathbf{E}(X^2 - 2cX + c^2) \\ &= \mathbf{E}X^2 - 2c\mathbf{E}X + \mathbf{E}c^2 \\ &= \mathbf{E}X^2 - 2c^2 + c^2 = \mathbf{E}X^2 - c^2, \end{aligned}$$

en utilisant la linéarité de l'espérance et l'espérance d'une constante. □

Proposition 3.41 (translation et changement d'échelle). *Si $\mathbf{E}X^2 < +\infty$,*

$$\forall a \in \mathbb{R}, \forall b \in \mathbb{R}, \quad \text{Var}(aX + b) = a^2 \text{Var } X, \quad \sigma(aX + b) = |a|\sigma(X). \quad (3.37)$$

Preuve. En utilisant la définition 3.39, la linéarité de l'espérance et le fait que l'espérance de la constante b est b :

$$\begin{aligned} \text{Var}(aX + b) &= \mathbf{E}(aX + b - \mathbf{E}(aX + b))^2 = \mathbf{E}(aX + b - a\mathbf{E}X - b)^2 \\ &= \mathbf{E}(a(X - \mathbf{E}X))^2 = \mathbf{E}(a^2(X - \mathbf{E}X)^2) \\ &= a^2 \mathbf{E}((X - \mathbf{E}X)^2) = a^2 \text{Var } X, \end{aligned}$$

ce qui nous donne les formules (3.37). □

Il est clair, d'après la définition de la variance, que la variance d'une constante est nulle. La réciproque est *presque* vraie :

5. Une façon plus simple de le vérifier est d'utiliser l'inégalité $|X| \leq 1 + |X|\mathbf{1}_{\{|X|>1\}} \leq 1 + X^2$.

6. À savoir X^2 , X et la v.a. constante $(\mathbf{E}X)^2$.

Proposition 3.42 (nullité de la variance et constance p.s.).

$$\text{Var } X = 0 \Leftrightarrow X = \mathbf{E}X \text{ p.s.} \Leftrightarrow X \text{ est presque sûrement constante.} \quad (3.38)$$

Preuve. Les implications de droite à gauche dans (3.38) sont déjà acquises. On sait en effet que l'espérance d'une constante est cette constante et que si la v.a. $Y := (X - \mathbf{E}X)^2$ vaut 0 avec probabilité 1, son espérance est nulle.

Pour la première implication de gauche à droite, il suffit d'appliquer le corollaire 3.16 à la v.a. positive Y . La deuxième implication est triviale. \square

Espérance et variance des lois usuelles

Lois discrètes

Loi	$X_P(\Omega)$	$P(X = k), k \in X_P(\Omega)$	$\mathbf{E}X$	$\text{Var } X$
Dirac δ_c	$\{c\}$	$P(X = c) = 1$	c	0
Bern(p)	$\{0, 1\}$	$P(X = 1) = p$	p	$p(1 - p)$
Unif($\{x_1, \dots, x_n\}$)	$\{x_1, \dots, x_n\}$	$P(X = x_k) = \frac{1}{n}$	$\mu = \frac{1}{n} \sum_{k=1}^n x_k$	$\frac{1}{n} \sum_{k=1}^n x_k^2 - \mu^2$
Bin(n, p)	$\llbracket 0, n \rrbracket$	$C_n^k p^k (1 - p)^{n-k}$	np	$np(1 - p)$
Hypg(N, M, n)		$\frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n}$	$n \frac{M}{N}$	$n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N - n}{N - 1}$
Geom(p)	\mathbb{N}^*	$(1 - p)^{k-1} p$	$\frac{1}{p}$	$\frac{1 - p}{p^2}$
Pois(α)	\mathbb{N}	$\frac{e^{-\alpha} \alpha^k}{k!}$	α	α

Lois à densité

Loi	Densité $f(t)$	$\mathbf{E}X$	$\text{Var } X$
Unif($[a, b]$)	$\frac{1}{b - a} \mathbf{1}_{[a, b]}(t)$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$
Exp(a)	$a \exp(-at) \mathbf{1}_{[0, +\infty[}(t)$	$\frac{1}{a}$	$\frac{1}{a^2}$
$\mathfrak{N}(m, \sigma)$	$\frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{-(t - m)^2}{2\sigma^2}\right)$	m	σ^2
Cau(0, 1)	$\frac{1}{\pi(1 + t^2)}$	sans	sans

Chapitre 4

Vecteurs aléatoires et indépendance

L'information pertinente résultant d'une expérience aléatoire ne se résume pas toujours à la valeur prise par une seule variable aléatoire réelle. On a souvent besoin de connaître les valeurs d'une suite finie de variables aléatoires. Par exemple au jeu de 421, on lance trois dés et on a besoin de connaître les points affichés par chacun des dés, le résultat sera donc décrit par un vecteur $(X_1(\omega), X_2(\omega), X_3(\omega))$. Si on tire sur une cible, le résultat sera décrit par les coordonnées $(X(\omega), Y(\omega))$ du point d'impact. Si on étudie le fonctionnement d'un guichet en observant les n premiers clients, le résultat sera décrit par la suite des $X_1, Y_1, Z_1, X_2, Y_2, Z_2, \dots, X_n, Y_n$ où X_i est le temps d'attente au guichet du i^{e} client, Y_i son temps de service et Z_i le temps s'écoulant entre le départ du i^{e} client et l'arrivée du $(i + 1)^{\text{e}}$. Ces suites finies de variables aléatoires sont appelées des *vecteurs aléatoires*. De même qu'une variable aléatoire peut être vue comme un procédé de choix d'un nombre réel au hasard, un vecteur aléatoire de dimension d , $X = (X_1, \dots, X_d)$ est un procédé de choix au hasard d'un point de \mathbb{R}^d . Ses composantes X_1, \dots, X_d sont alors autant de variables aléatoires réelles. Arrivé là, le lecteur peut se demander légitimement s'il y a un intérêt à consacrer tout un chapitre aux vecteurs aléatoires puisque ces objets ne sont que des suites finies de variables aléatoires et que ces dernières sont maintenant bien connues. L'intérêt de cette étude repose sur la remarque informelle suivante à laquelle nous donnerons bientôt un sens mathématique précis : la *connaissance probabiliste globale du vecteur* $X = (X_1, \dots, X_d)$ *apporte davantage d'information que la connaissance probabiliste individuelle de chacune de ses composantes* X_i . Au premier abord, cette idée peut paraître choquante car une lecture rapide de la phrase précédente laisse croire que la connaissance de $X(\omega)$ apporte quelque chose de plus que celle de tous les $X_i(\omega)$, $i = 1, \dots, d$, ce qui n'est évidemment pas vrai. La clé de l'énigme est dans l'expression « connaissance *probabiliste* » que nous remplacerons bientôt par connaissance de la *loi*, dès que nous aurons défini la loi d'un vecteur aléatoire. En attendant voici une image qui peut nous aider à comprendre ce dont il s'agit. Considérons un ensemble de 10 coureurs de fond, chacun muni d'un dossard numéroté de 1 à 10. Si on les rassemble sur une même piste pour une épreuve de 5 000 mètres, on peut représenter le résultat de la course par le vecteur (X_1, \dots, X_{10}) , où X_i désigne le temps mis par le coureur numéroté i pour parcourir les 5 000 mètres. Tout amateur d'athlétisme sait bien que cette expérience n'est pas équivalente à faire courir isolément un 5 000 mètres à chacun des 10 coureurs sur des stades séparés. La différence ici vient de la compétition, de la tactique de course, etc. En revanche, dans d'autres situations, le comportement global du vecteur des d composantes se réduit au comportement individuel de chacune d'elles. On admet généralement que c'est le cas lorsqu'on lance trois dés en considérant qu'il revient au même de les lancer ensemble sur la même table ou séparément sur trois tables. On parle alors d'indépendance

des composantes. Cette notion d'indépendance des composantes X_i du vecteur aléatoire X est reliée à celle d'une suite d'évènements A_i , où la réalisation ou non de A_i ne dépend que des valeurs de X_i . L'étude des suites finies de variables aléatoires indépendantes prend donc place naturellement dans ce chapitre comme le cas particulier des vecteurs aléatoires à composantes indépendantes.

Sauf mention explicite du contraire, toutes les variables aléatoires et tous les vecteurs aléatoires considérés dans ce chapitre seront définis sur le même espace probabilisé (Ω, \mathcal{F}, P) .

4.1 Vecteurs aléatoires

4.1.1 Généralités

Définition 4.1 (vecteur aléatoire de \mathbb{R}^d). *On dit que l'application $X : \Omega \rightarrow \mathbb{R}^d$ est un vecteur aléatoire sur (Ω, \mathcal{F}) si c'est une application mesurable $\mathcal{F} - \text{Bor}(\mathbb{R}^d)$.*

Proposition 4.2. *Soit X un vecteur aléatoire sur (Ω, \mathcal{F}) et P une probabilité sur (Ω, \mathcal{F}) . La fonction d'ensembles $P_X = P \circ X^{-1}$ définie sur $\text{Bor}(\mathbb{R}^d)$ par*

$$\forall B \in \text{Bor}(\mathbb{R}^d), \quad P_X(B) := P(X^{-1}(B)) = P(X \in B) \quad (4.1)$$

est une probabilité sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$.

 IFP prop. 2.25, p. 74.

Définition 4.3 (loi d'un vecteur aléatoire). *Soient (Ω, \mathcal{F}, P) un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire sur (Ω, \mathcal{F}) . On appelle loi de X sous P , ou plus simplement loi de X , la probabilité P_X sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$ définie par (4.1).*

Proposition 4.4 (lois marginales). *Si $X = (X_1, \dots, X_d)$ est un vecteur aléatoire sur (Ω, \mathcal{F}) , chacune de ses composantes X_i ($1 \leq i \leq d$) est une variable aléatoire réelle sur (Ω, \mathcal{F}) . La loi de X_i est appelée i^e loi marginale de X et est donnée par :*

$$\forall B_i \in \text{Bor}(\mathbb{R}), \quad P_{X_i}(B_i) = P(X_i \in B_i) = P(X \in \mathbb{R}^{i-1} \times B_i \times \mathbb{R}^{d-i}). \quad (4.2)$$

Preuve. La mesurabilité $\mathcal{F} - \text{Bor}(\mathbb{R})$ de X_i s'obtient par composition à partir de la mesurabilité $\mathcal{F} - \text{Bor}(\mathbb{R}^d)$ de X . En effet $X_i = \pi_i \circ X$, où $\pi_i : (x_1, \dots, x_d) \mapsto x_i$ est la i^e projection canonique de \mathbb{R}^d sur \mathbb{R} ; comme π_i est continue, elle est borélienne, c'est-à-dire ici mesurable $\text{Bor}(\mathbb{R}^d) - \text{Bor}(\mathbb{R})$. Donc X_i est bien une variable aléatoire. Pour vérifier (4.2), il suffit de remarquer que l'équivalence

$$X_i \in B_i \Leftrightarrow X \in \mathbb{R}^{i-1} \times B_i \times \mathbb{R}^{d-i}$$

entraîne l'égalité des évènements correspondants et de leur probabilité. □

Exemple 4.5. Soit $X = (X_1, X_2)$ un vecteur aléatoire de loi uniforme sur $[0, 1]^2$. Alors $P(X \in B) = \lambda_2(B \cap [0, 1]^2)$ pour tout borélien B de \mathbb{R}^2 et on vérifie (exercice) que les lois marginales P_{X_1} et P_{X_2} sont égales à la loi uniforme sur $[0, 1]$.

Remarque 4.6 (d'importance capitale). Une conséquence de la proposition 4.4 est que la connaissance de la loi du vecteur aléatoire X détermine complètement celle de ses lois marginales. La *reciproque est fausse*. On peut même affirmer sans hésiter qu'il est impossible de comprendre la notion de vecteur aléatoire tant que l'on n'a pas assimilé ce fait. La comparaison de l'exemple 4.7 ci-dessous avec l'exemple 4.5 permet de voir que la connaissance des lois marginales d'un vecteur ne détermine pas la loi du vecteur.

Exemple 4.7. Prenons une variable aléatoire réelle Y_1 de loi uniforme sur $[0, 1]$ et posons $Y_2 := Y_1$ et $Y := (Y_1, Y_2)$. Le vecteur aléatoire Y a par construction les mêmes lois marginales que le vecteur X de l'exemple 4.5. Notons $\Delta := \{(s, t) \in [0, 1]^2; s = t\}$ la première diagonale du carré unité. Il est clair par construction que $P(Y \in \Delta) = 1$. D'un autre côté, $P(X \in \Delta) = 0$, car le segment Δ est de λ_2 mesure nulle. Ceci empêche que X et Y aient même loi.



Pour une simulation illustrative, PVIR figure 8.2 p. 298.

Définition 4.8 (vecteur aléatoire discret). *Le vecteur aléatoire X de \mathbb{R}^d est dit discret si $X(\Omega)$ est une partie au plus dénombrable de \mathbb{R}^d .*

Il est clair que la loi de X s'écrit

$$P_X = \sum_{x \in X(\Omega)} P(X = x) \delta_x.$$

Les variables aléatoires marginales X_i de $X = (X_1, \dots, X_d)$ sont alors des variables aléatoires discrètes. En effet, soit π_i , la restriction à $X(\Omega)$ de la projection canonique sur la i -ième composante de \mathbb{R}^d . Cette application réalise une surjection de $X(\Omega)$ sur $X_i(\Omega)$. On en déduit que $X_i(\Omega)$ est au plus dénombrable.

Exemple 4.9 (lois multinomiales). Le vecteur aléatoire N suit la loi multinomiale de paramètres n et (p_1, \dots, p_d) où $n \in \mathbb{N}^*$ et les p_i sont strictement positifs et de somme 1 si pour tout d -uplet (j_1, j_2, \dots, j_d) d'entiers tels que $j_1 + j_2 + \dots + j_d = n$,

$$P\{N = (j_1, j_2, \dots, j_d)\} = \frac{n!}{j_1! j_2! \dots j_d!} p_1^{j_1} p_2^{j_2} \dots p_d^{j_d}.$$

Ici l'ensemble $N(\Omega) = \{(j_1, j_2, \dots, j_d) \in \mathbb{N}^d; j_1 + j_2 + \dots + j_d = n\}$ est fini et on vérifie grâce à la formule du multinôme que

$$\sum_{x \in N(\Omega)} P(N = x) = \sum_{j_1 + \dots + j_d = n} \frac{n!}{j_1! j_2! \dots j_d!} p_1^{j_1} p_2^{j_2} \dots p_d^{j_d} = (p_1 + \dots + p_d)^n = 1^n = 1.$$

La loi multinomiale est celle du vecteur des résultats d'une suite d'épreuves répétées indépendantes ayant chacune d issues possibles de probabilités respectives p_1, \dots, p_d . On pourra justifier cette affirmation en exercice. Par exemple considérons 20 tirages d'une boule avec remise dans une urne contenant 1 boule bleue, 3 jaunes, 4 rouges et 2 vertes. Notons $N = (N_1, N_2, N_3, N_4)$ où N_i est le nombre de boules de la couleur i en numérotant les couleurs par ordre alphabétique (b,j,r,v). On a $(p_1, p_2, p_3, p_4) = (\frac{1}{10}, \frac{3}{10}, \frac{4}{10}, \frac{2}{10})$. La probabilité d'obtenir en 20 tirages 3 bleues, 5 jaunes, 10 rouges et 2 vertes est

$$P(N = (3, 5, 10, 2)) = \frac{20!}{3! 5! 10! 2!} \left(\frac{1}{10}\right)^3 \left(\frac{3}{10}\right)^5 \left(\frac{4}{10}\right)^{10} \left(\frac{2}{10}\right)^2 \simeq 0,004745.$$

Il est facile de voir que les lois des vecteurs aléatoires discrets sur \mathbb{R}^d sont exactement les mesures de probabilité ponctuelles sur \mathbb{R}^d (au sens de l'exemple 1.8). Autrement dit, une mesure μ sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$ est la loi d'un vecteur aléatoire discret si et seulement si elle peut s'écrire $\mu := \sum_{i \in I} p_i \delta_{x_i}$ où $\{x_i, i \in I\}$ est une famille de vecteurs de \mathbb{R}^d et $\{p_i, i \in I\}$ est une famille *sommable* de réels *strictement* positifs de somme 1.

Certaines lois de vecteurs aléatoires ont une « densité » par rapport à la mesure de Lebesgue de \mathbb{R}^d .

Définition 4.10 (densité de probabilité sur \mathbb{R}^d). On appelle *densité de probabilité* sur \mathbb{R}^d toute fonction borélienne positive définie sur \mathbb{R}^d et vérifiant :

$$\int_{\mathbb{R}^d} f(t) dt = \int_{\mathbb{R}^d} f(t_1, \dots, t_d) dt_1 \dots dt_d = 1.$$

Définition 4.11 (vecteur aléatoire à densité). Soit f une densité de probabilité sur \mathbb{R}^d . On dit que le vecteur aléatoire X de \mathbb{R}^d a pour densité f si sa loi P_X a pour densité f par rapport à la mesure de Lebesgue λ_d de \mathbb{R}^d , autrement dit si

$$\forall B \in \text{Bor}(\mathbb{R}^d), \quad P(X \in B) = \int_B f d\lambda_d = \int_B f(t_1, \dots, t_d) dt_1 \dots dt_d. \quad (4.3)$$

Pour que X admette pour densité f , il suffit de vérifier (4.3) pour tout B dans la classe des pavés fermés bornés : $B = [a_1, b_1] \times \dots \times [a_d, b_d]$.

Voici un premier exemple de vecteur aléatoire à densité. D'autres seront vus ultérieurement.

Exemple 4.12 (densité de la loi uniforme sur un borélien de \mathbb{R}^d). Soit B un borélien de \mathbb{R}^d tel que $0 < \lambda_d(B) < +\infty$. Si le vecteur aléatoire X de \mathbb{R}^d suit la loi uniforme sur B , cf. exemple 1.23, il admet pour densité la fonction

$$f = \frac{1}{\lambda_d(B)} \mathbf{1}_B.$$

 IFP, exemple 3.3., p. 97.


Comme dans le cas $d = 1$, on dispose d'une condition suffisante pratique pour vérifier que deux vecteurs aléatoires à densité n'ont pas même loi.

Lemme 4.13. Soient X et Y deux vecteurs aléatoires de \mathbb{R}^d admettant respectivement pour densité les fonctions f et g . On suppose qu'il existe $t_0 \in \mathbb{R}^d$ tel que $f(t_0) \neq g(t_0)$ et que de plus, f et g sont toutes deux continues au point t_0 . Alors X et Y n'ont pas même loi.

La preuve est essentiellement la même que celle du lemme 2.25, en remplaçant les intervalles ouverts par des pavés ouverts.

Proposition 4.14 (densités marginales). Si le vecteur aléatoire X de \mathbb{R}^d est à densité f , ses lois marginales sont aussi à densité et pour $i \in \llbracket 1, d \rrbracket$, une densité de X_i est donnée par

$$x_i \longmapsto f_{X_i}(x_i) = \int_{\mathbb{R}^{i-1} \times \mathbb{R}^{d-i}} f(t_1, \dots, t_{i-1}, x_i, t_{i+1}, \dots, t_d) dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_d.$$

 IFP, prop. 5.21, p. 162.

Exemple 4.15 (loi uniforme sur un disque). Soit D le disque unité de \mathbb{R}^2 et $X = (X_1, X_2)$ un vecteur aléatoire suivant la loi uniforme sur D . D'après l'exemple 4.12, nous savons qu'il admet pour densité $f = \pi^{-1} \mathbf{1}_D$. Calculons la densité marginale f_{X_1} fournie par la proposition 4.14.

$$f_{X_1}(x_1) = \int_{\mathbb{R}} \frac{1}{\pi} \mathbf{1}_D(x_1, t_2) dt_2 = \dots = \frac{2}{\pi} (1 - x_1^2)^{1/2} \mathbf{1}_{[-1, 1]}(x_1). \quad (4.4)$$

Par raison de symétrie il est clair que X_1 et X_2 ont même loi et donc même densité, d'où

$$f_{X_2}(x_2) = \frac{2}{\pi}(1 - x_2^2)^{1/2} \mathbf{1}_{[-1,1]}(x_2).$$



Détails dans PVIR, exemple 8.15, pp. 301–302.

On peut définir la fonction de répartition F d'un vecteur aléatoire par

$$\forall x = (x_1, \dots, x_d), \quad F(x) = P(X \in]-\infty, x_1] \times \dots \times]-\infty, x_d]).$$

Comme en dimension 1, la fonction de répartition caractérise la loi. Ceci est lié au fait que la tribu borélienne de \mathbb{R}^d est engendrée par la classe des ensembles de la forme $] -\infty, x_1] \times \dots \times] -\infty, x_d]$. Néanmoins le rôle des f.d.r. en dimension $d > 1$ est bien moindre qu'en dimension 1. On préfère caractériser la loi d'un vecteur aléatoire par une collection de h -moments $\mathbf{E}h(X)$, $h : \mathbb{R}^d \rightarrow \mathbb{R}$, au sens suivant.

Proposition 4.16 (caractérisation par les moments fonctionnels). *La loi d'un vecteur aléatoire X de \mathbb{R}^d est caractérisée par la famille de moments fonctionnels $\{\mathbf{E}h(X) ; h \in \mathcal{H}\}$, où \mathcal{H} est une classe « suffisamment riche » de fonctions boréliennes $\mathbb{R}^d \rightarrow \mathbb{R}$. Autrement dit, les deux vecteurs aléatoires X et Y ont même loi si et seulement si $\mathbf{E}h(X) = \mathbf{E}h(Y)$ pour toute $h \in \mathcal{H}$. Comme famille \mathcal{H} « suffisamment riche », on peut prendre :*

- l'ensemble des fonctions boréliennes positives $\mathbb{R}^d \rightarrow \mathbb{R}_+$,
- l'espace $C^b(\mathbb{R}^d)$ des fonctions continues bornées $\mathbb{R}^d \rightarrow \mathbb{R}$,
- l'espace $C^c(\mathbb{R}^d)$ des fonctions continues à support compact $\mathbb{R}^d \rightarrow \mathbb{R}$,
- le sous-ensemble $C_+^c(\mathbb{R}^d)$ des fonctions positives de $C^c(\mathbb{R}^d)$.



PVIR, prop. 8.16, p. 303.

Il importe de savoir calculer les $\mathbf{E}h(X)$ quand on connaît la loi de X . Les formules sont analogues à celles déjà données en dimension 1 et sont des applications immédiates du théorème de transfert.

Proposition 4.17 (calcul de $\mathbf{E}h(X)$).

- a) *Si la loi du vecteur aléatoire X est discrète et si h est une fonction borélienne $\mathbb{R}^d \rightarrow \mathbb{R}$,*

$$\mathbf{E}|h(X)| = \sum_{x \in X_P(\Omega)} |h(x)| P(X = x),$$

où $X_P(\Omega) = \{x \in \mathbb{R}^d ; P(X = x) \neq 0\}$. Si de plus $\mathbf{E}|h(X)| < +\infty$,

$$\mathbf{E}h(X) = \sum_{x \in X_P(\Omega)} h(x) P(X = x).$$

- b) *Si X est à densité f et $h : \mathbb{R}^d \rightarrow \mathbb{R}$ est borélienne et telle que $\int_{\mathbb{R}^d} |h| f d\lambda_d < +\infty$, alors la variable aléatoire réelle $h(X)$ est intégrable et*

$$\mathbf{E}h(X) = \int_{\mathbb{R}^d} h(x) f(x) dx = \int_{\mathbb{R}^d} h(x_1, \dots, x_d) f(x_1, \dots, x_d) dx_1 \dots dx_d.$$

Nous donnons maintenant une formule de calcul de densité du vecteur aléatoire $g(X)$ où X est un vecteur aléatoire à densité.

Proposition 4.18 (densité d'un vecteur aléatoire image). *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de \mathbb{R}^d ayant une densité f_X . On suppose de plus que D est un ouvert de \mathbb{R}^d tel que $P(X \in D) = 1$ et que $g : D \rightarrow \mathbb{R}^d$ est C^1 , injective avec un déterminant jacobien $\text{Jac}(g)$ qui ne s'annule en aucun point de D . Alors g réalise une bijection C^1 d'inverse C^1 (autrement dit un C^1 -difféomorphisme) entre D et son image $D' := g(D)$. Le vecteur aléatoire $Y := g(X)$ admet pour densité*

$$f_Y(y) = f_X(g^{-1}(y)) |\text{Jac}(g^{-1})(y)| \mathbf{1}_{D'}(y).$$

Le fait que g soit un C^1 -difféomorphisme découle d'un théorème d'analyse (théorème d'inversion globale). Rappelons que le jacobien de g est donné par

$$\text{Jac}(g) = \det \left(\left[\frac{\partial g_i}{\partial x_j} \right]_{1 \leq i, j \leq d} \right),$$

où les g_i sont les applications coordonnées de $g = (g_1, \dots, g_d)$, c.-à-d. $g_i(x_1, \dots, x_d) = \pi_i(g(x_1, \dots, x_d))$, π_i étant la projection canonique sur la i^{e} coordonnée dans \mathbb{R}^d . Pour calculer $\text{Jac}(g^{-1})(y)$, on peut soit utiliser la formule ci-dessus en remplaçant g par g^{-1} et les x_j par les y_j , soit utiliser la relation

$$\text{Jac}(g^{-1})(y) = \frac{1}{\text{Jac}(g)(g^{-1}(y))}.$$

Un cas particulier facile et important est celui où g est une application linéaire bijective de \mathbb{R}^d sur \mathbb{R}^d . Dans ce cas, soit $A = [a_{i,j}]$ la matrice $d \times d$ de g relativement à la base canonique de \mathbb{R}^d . Cette matrice est inversible, son déterminant est donc non nul. La i^{e} composante de $g(x)$ est ici $g_i(x_1, \dots, x_d) = \sum_{j=1}^d a_{i,j} x_j$, d'où

$$\forall (i, j) \in \llbracket 1, d \rrbracket^2, \quad \frac{\partial g_i}{\partial x_j} = a_{i,j}.$$

On en déduit que le jacobien de g comme celui de g^{-1} sont constants et valent

$$\text{Jac}(g) = \det A = \det(g), \quad \text{Jac}(g^{-1}) = \frac{1}{\det A} = \frac{1}{\det g}.$$

Nous obtenons ainsi le corollaire suivant de la proposition 4.18, avec ici $D = \mathbb{R}^d$ et $D' = g(D) = \mathbb{R}^d$.

Corollaire 4.19 (changement de variable linéaire). *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de \mathbb{R}^d ayant une densité f_X et $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ une application linéaire bijective. Alors le vecteur aléatoire $Y := g(X)$ admet pour densité*

$$f_Y(y) = \frac{1}{|\det g|} f_X(g^{-1}(y)).$$

 IFP, toute la section 5.4 et notamment la sous-section 5.4.7, pp. 173–176.

4.1.2 Covariance de deux variables aléatoires

Le calcul de la variance de la somme de deux variables aléatoires de carré intégrable X et Y ne peut se faire en général qu'à partir de la loi du couple (X, Y) . Ceci nous amène à introduire la covariance de (X, Y) .

Proposition 4.20 (inégalité de Cauchy-Schwarz). *Si les variables aléatoires réelles X et Y ont des moments d'ordre 2, alors la variable aléatoire XY est intégrable et*

$$|\mathbf{E}(XY)| \leq (\mathbf{E}X^2)^{1/2}(\mathbf{E}Y^2)^{1/2}. \quad (4.5)$$

Définition 4.21 (covariance). *Si les variables aléatoires réelles X et Y ont des moments d'ordre 2, on appelle covariance du couple aléatoire (X, Y) la quantité :*

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

Remarquons que $\text{Cov}(X, X) = \text{Var } X$.

Proposition 4.22 (propriétés de la covariance). *Les propriétés suivantes sont vérifiées pour tout couple (X, Y) de v.a. réelles ayant des moments d'ordre 2.*

- (i) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- (ii) Pour tous réels a, b, c, d : $\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$.
- (iii) $|\text{Cov}(X, Y)| \leq \sigma(X)\sigma(Y)$.

La vérification est laissée au lecteur.

Définition 4.23 (coefficient de corrélation). *Si X et Y sont des variables aléatoires réelles non constantes ayant des moments d'ordre 2, on appelle coefficient de corrélation entre X et Y la quantité :*

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

D'après (iii) on a toujours $-1 \leq \rho(X, Y) \leq 1$. D'autre part il résulte facilement du cas d'égalité dans l'inégalité de Cauchy-Schwarz que $|\rho|$ est maximal lorsque Y est une fonction affine de X : $Y = aX + b$. Quand $\rho = 0$, ce qui arrive en particulier lorsque X et Y sont indépendantes, on dit que X et Y sont *non corrélées*.

Proposition 4.24 (formule de Koenig pour la covariance). *Si la covariance de X et Y existe, elle peut se calculer par :*

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

Preuve. La vérification est analogue à celle de la formule de Koenig pour la variance (qui n'est que le cas particulier $Y = X$) et est laissée en exercice. \square

Remarque 4.25 (calcul explicite de la covariance). Les formules de calcul des h -moments appliquées au vecteur aléatoire (X, Y) (prop. 4.17) et aux variables aléatoires réelles X et Y (prop. 3.35 et 3.36) nous donnent pour la covariance (lorsqu'elle existe) les formules explicites suivantes.

Si (X, Y) est discret,

$$\text{Cov}(X, Y) = \sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega)}} xyP(X = x, Y = y) - \sum_{x \in X(\Omega)} xP(X = x) \sum_{y \in Y(\Omega)} yP(Y = y). \quad (4.6)$$

Si (X, Y) est à densité f , en notant f_X et f_Y les densités marginales,

$$\text{Cov}(X, Y) = \int_{\mathbb{R}^2} xyf(x, y) \, dx \, dy - \int_{-\infty}^{+\infty} xf_X(x) \, dx \int_{-\infty}^{+\infty} yf_Y(y) \, dy. \quad (4.7)$$

Proposition 4.26 (variance d'une somme). *Si X_1, \dots, X_n sont des variables aléatoires réelles ayant des moments d'ordre 2,*

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j) \quad (4.8)$$

$$= \sum_{i=1}^n \text{Var } X_i + \sum_{\substack{i,j=1 \\ i \neq j}}^n \text{Cov}(X_i, X_j). \quad (4.9)$$

Dans le cas $n = 2$ (4.9) s'écrit :

$$\text{Var}(X + Y) = \text{Var } X + \text{Var } Y + 2 \text{Cov}(X, Y). \quad (4.10)$$

4.1.3 Espérance et covariance d'un vecteur aléatoire

En dimension 1, l'espérance et la variance d'une variable aléatoire, lorsqu'elles existent, permettent de se faire une idée de la localisation de la loi et de sa dispersion. Elles jouent un rôle important notamment dans le théorème limite central. Nous allons étendre ces notions au cas des vecteurs aléatoires.

Définition 4.27 (espérance d'un vecteur aléatoire). *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de \mathbb{R}^d . On dit qu'il est intégrable si la variable aléatoire positive $\|X\|$ est intégrable ($\mathbf{E}\|X\| < +\infty$), ce qui équivaut à l'intégrabilité de chacune des composantes ($\mathbf{E}|X_i| < +\infty$ pour tout $i \in \llbracket 1, d \rrbracket$). Dans ce cas on appelle espérance de X ou vecteur d'espérances de X le vecteur*

$$\mathbf{E}X := (\mathbf{E}X_1, \dots, \mathbf{E}X_d). \quad (4.11)$$

C'est délibérément que nous n'avons pas précisé le choix de la norme dans cette définition. En effet toutes les normes sur \mathbb{R}^d sont équivalentes. Si donc $\|\cdot\|$ est une norme sur \mathbb{R}^d elle est équivalente en particulier à la norme $\|\cdot\|_1$ définie par $\|x\|_1 := |x_1| + \dots + |x_d|$. On en déduit l'existence de deux constantes a et b strictement positives telles que

$$a(|X_1| + \dots + |X_d|) \leq \|X\| \leq b(|X_1| + \dots + |X_d|). \quad (4.12)$$

De la première inégalité on tire $|X_i| \leq a^{-1}\|X\|$, ce qui montre que l'intégrabilité de $\|X\|$ implique celle de chaque X_i . De la seconde inégalité on déduit que si les v.a. X_i sont toutes intégrables, les $|X_i|$ le sont aussi, ainsi que leur somme finie indexée par i , d'où l'intégrabilité de $\|X\|$. Nous venons de vérifier que l'intégrabilité de $\|X\|$ équivaut à celle de toutes les X_i , ce qui montre aussi que l'intégrabilité de X ne dépend pas du choix de la norme.

Une propriété importante de l'espérance des variables aléatoires est la *linéarité*. Sa généralisation aux *vecteurs* aléatoires est immédiate : si X et Y sont des vecteurs aléatoires intégrables de \mathbb{R}^d , a et b des scalaires quelconques, le vecteur aléatoire $aX + bY$ est intégrable et $\mathbf{E}(aX + bY) = a\mathbf{E}X + b\mathbf{E}Y$. Pour le vérifier, il suffit d'appliquer composante par composante, la linéarité de l'espérance des variables aléatoires réelles.

Le résultat suivant nous dit *grosso modo* que l'espérance commute avec les applications linéaires.

Proposition 4.28 (espérance et applications linéaires). *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire intégrable de \mathbb{R}^d .*

a) Pour toute forme linéaire $u : \mathbb{R}^d \rightarrow \mathbb{R}$, la v.a. réelle $u(X)$ est intégrable et

$$\mathbf{E}u(X) = u(\mathbf{E}X). \quad (4.13)$$

b) $\mathbf{E}X$ est le seul vecteur z de \mathbb{R}^d vérifiant $\mathbf{E}u(X) = u(z)$ pour toute forme linéaire u sur \mathbb{R}^d .

c) Si $A : \mathbb{R}^d \rightarrow \mathbb{R}^j$, $x \mapsto A(x)$ est une application linéaire, le vecteur aléatoire AX de \mathbb{R}^j est intégrable et

$$\mathbf{E}A(X) = A(\mathbf{E}X). \quad (4.14)$$



Pour la preuve, voir PVIR p. 309.

Définition 4.29 (matrice de covariance). Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de carré intégrable, c.-à-d. $\mathbf{E}\|X\|^2 < +\infty$. On appelle matrice de covariance de X la matrice carrée K de terme général

$$K_{i,j} := \text{Cov}(X_i, X_j), \quad (i, j) \in \llbracket 1, d \rrbracket^2.$$

En utilisant (4.12) comme ci-dessus on a $|X_i| \leq a^{-1}\|X\|$ ce qui montre que les composantes X_i d'un vecteur de carré intégrable sont des v.a. de carré intégrable. Ceci justifie l'existence des $\text{Cov}(X_i, X_j)$. La connaissance de la matrice de covariance K de X permet le calcul de $\text{Var } u(X)$ pour toute forme linéaire u sur \mathbb{R}^d .

Proposition 4.30. Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de carré intégrable et $u : x = (x_1, \dots, x_d) \mapsto a_1x_1 + \dots + a_dx_d$ une forme linéaire sur \mathbb{R}^d . Alors la variable aléatoire réelle $u(X)$ est de carré intégrable et

$$\text{Var } u(X) = \sum_{i,j=1}^d a_i a_j \text{Cov}(X_i, X_j). \quad (4.15)$$

Outre son utilité pour le calcul de la variance d'une combinaison linéaire des composantes de X , la matrice de covariance K est porteuse d'informations sur la loi de X . Elle permet notamment de savoir quel sous-espace affine de \mathbb{R}^d est « l'habitat naturel » de X . En particulier, si le déterminant de K est nul, la loi de X est portée par un sous-espace affine de dimension strictement inférieure à d . Et dans ce cas, on peut montrer que la loi de X ne peut avoir de densité au sens de la définition 4.11.

Proposition 4.31 (support d'un vecteur aléatoire de carré intégrable). Soit X un vecteur aléatoire de carré intégrable, $\mathbf{E}X$ son vecteur d'espérances et K sa matrice de covariance. Alors la loi de X est portée par le sous-espace affine $\mathbf{E}X + \text{Im}(K)$, où $\text{Im}(K)$ désigne le s.e.v. image de \mathbb{R}^d par l'application linéaire de matrice K dans la base canonique. Autrement dit,

$$P(X - \mathbf{E}X \in \text{Im}(K)) = 1. \quad (4.16)$$

Bien entendu, ce résultat n'a d'intérêt que si $\text{Im}(K)$ est un sous-espace vectoriel strict de \mathbb{R}^d , autrement dit si l'application linéaire associée à K n'est pas bijective.



PVIR, pp. 310–311.

4.2 Indépendance de variables et vecteurs aléatoires

4.2.1 Indépendance de familles d'évènements

Définition 4.32. Soit I un ensemble quelconque d'indices.

a) Les évènements d'une famille $(A_i)_{i \in I}$ sont mutuellement indépendants si

$$\forall J \text{ fini } \subset I \text{ tel que } \text{card } J \geq 2, \quad P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

b) Les classes d'évènements $(\mathcal{C}_i)_{i \in I}$ (i.e. $\forall i \in I, \mathcal{C}_i \subset \mathcal{F}$) sont mutuellement indépendantes si pour tout choix d'un A_i dans chaque \mathcal{C}_i , les $(A_i)_{i \in I}$ sont mutuellement indépendants au sens du a).

Dans toute la suite, sauf mention explicite du contraire, nous allègerons « mutuellement indépendant(e)s » en « indépendant(e)s ». D'autre part il est clair d'après cette définition que l'indépendance ne dépend pas de la façon dont sont indexés les A_i . En particulier, elle est conservée par toute permutation sur I . La partie b) de la définition permet notamment de définir l'indépendance de tribus et par la suite l'indépendance de variables aléatoires. Cependant tester l'indépendance de tribus en appliquant directement cette définition peut s'avérer infaisable car on ne sait généralement pas décrire explicitement tous les éléments d'une tribu. Il est donc crucial de pouvoir tester cette indépendance sur des classes plus petites mais « suffisamment riches ». C'est l'objet de la proposition suivante.

Proposition 4.33. Si $\mathcal{C}_1, \dots, \mathcal{C}_n$ sont des π -classes indépendantes d'évènements, alors les tribus engendrées $\sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_n)$ sont indépendantes.

Rappelons qu'une π -classe est une famille de parties de Ω stable par intersection finie. Par exemple la classe des intervalles de \mathbb{R} est une π -classe¹, de même que la classe des pavés boréliens de \mathbb{R}^d . Le théorème $\pi - \lambda$ ou théorème de Dynkin selon lequel si une λ -classe de parties de Ω contient une π -classe \mathcal{C} , elle contient aussi la tribu engendrée par \mathcal{C} joue un rôle central dans la preuve de la proposition 4.33.

♣ IFP th. 1.15 p. 21 et prop. 5.35 p. 177.

Pour une application élémentaire de la proposition 4.33, examinons le cas où les \mathcal{C}_i sont réduites à un seul évènement A_i (ce sont alors trivialement des π classes). Les tribus engendrées sont alors les $\sigma(\mathcal{C}_i) = \{A_i, A_i^c, \Omega, \emptyset\}$. On en déduit le corollaire suivant (que l'on pourra aussi démontrer directement à titre d'exercice).

Corollaire 4.34. Si A_1, \dots, A_n est une suite finie d'évènements indépendants, alors toute suite B_1, \dots, B_n telle que pour chaque i , $B_i = A_i$ ou $B_i = A_i^c$ est encore une suite d'évènements indépendants. La même propriété reste valable pour les suites infinies d'évènements indépendants.

4.2.2 Indépendance de variables aléatoires

Nous définissons maintenant l'indépendance d'une famille quelconque $(X_i)_{i \in I}$ de variables ou vecteurs aléatoires. Cette indépendance est celle de la famille de sous-tribus de \mathcal{F} engendrées par les X_i . Rappelons que si X est une application $(\Omega, \mathcal{F}) \rightarrow (E, \mathcal{B})$, la tribu engendrée par X est

$$\sigma(X) := X^{-1}(\mathcal{B}) = \{X^{-1}(B); B \in \mathcal{B}\}.$$

1. En considérant \emptyset comme un intervalle ouvert dont les bornes sont égales.

Si de plus, X est \mathcal{F} - \mathcal{B} mesurable, alors $\sigma(X)$ est une sous-tribu de \mathcal{F} . On peut dire alors que X est une variable aléatoire à valeurs dans l'espace mesurable (E, \mathcal{B}) . En pratique, E peut être \mathbb{R}_+ , \mathbb{R} , \mathbb{C} , \mathbb{N} , \mathbb{R}^d (on parle alors de vecteur aléatoire), un espace vectoriel normé (y compris de dimension infinie). . . et la tribu \mathcal{B} est la tribu borélienne correspondante ou la tribu $\mathcal{P}(E)$ lorsque E est dénombrable.

Jusqu'à la fin du chapitre, toutes les variables aléatoires et vecteurs aléatoires apparaissant dans un même énoncé sont *réputées définies sur un même espace probabilisable* (Ω, \mathcal{F}) , *muni d'une mesure de probabilité P fixée*. Toutes les notions d'indépendance introduites dans cette section sont relatives à cette mesure et en toute rigueur il faudrait parler à chaque fois de P -indépendance. Nous nous en abstiendrons conformément à l'usage.

Définition 4.35. Soit I une famille quelconque d'indices et pour tout $i \in I$, X_i une variable aléatoire à valeurs dans l'espace mesurable (E_i, \mathcal{B}_i) . On dit que $(X_i)_{i \in I}$ est une famille de variables aléatoires indépendantes si la famille de tribus $(\sigma(X_i))_{i \in I}$ est indépendante, autrement dit si

$$\forall J \text{ fini } \subset I \text{ tel que } \text{card } J \geq 2, \forall j \in J, \forall B_j \in \mathcal{B}_j, \quad P(\forall j \in J, X_j \in B_j) = \prod_{j \in J} P(X_j \in B_j).$$

Notons que cette définition est assez souple pour englober l'indépendance d'une collection complètement hétéroclite de « variables aléatoires », certaines pouvant être des variables aléatoires réelles, d'autres des vecteurs aléatoires (de dimensions diverses), d'autres des variables aléatoires discrètes, . . .

La proposition suivante qui établit le lien entre indépendance et probabilité produit est un outil fondamental pour vérifier l'indépendance d'une famille de vecteurs aléatoires et pour établir des propriétés essentielles de l'indépendance.

Proposition 4.36. Un vecteur aléatoire (X_1, \dots, X_d) à valeurs dans \mathbb{R}^d est à composantes indépendantes si et seulement si sa loi est le produit de ses lois marginales, i.e.

$$P_{(X_1, \dots, X_d)} = P_{X_1} \otimes \dots \otimes P_{X_d}.$$



IFP, prop. 5.39 p. 180.

Nous examinons maintenant quelques conséquences de la proposition 4.36. Toutes les propositions figurant dans le reste de cette sous-section sont en fait des corollaires de la proposition 4.36.

Une propriété bien commode de l'indépendance des v.a. est *l'hérédité*. Avant de l'énoncer formellement, voyons sa signification sur un exemple. Supposons que les v.a. réelles X_1, \dots, X_5 soient indépendantes. Alors les trois variables aléatoires Y_1, Y_2, Y_3 suivantes sont indépendantes

$$Y_1 := X_1 + X_2, \quad Y_2 := X_3 \sin X_4, \quad Y_3 := \exp(X_5^2 - X_5).$$

Il en va de même pour les vecteurs aléatoires Z_1, Z_2, Z_3

$$Z_1 := (X_1, X_2), \quad Z_2 := (X_3 + \cos X_4, X_4^2), \quad Z_3 := (X_5^3, 2X_5, X_5^2).$$

Nous énonçons le résultat à partir d'une suite de variables aléatoires indépendantes pour ne pas trop alourdir les notations, mais il se généralise à une famille finie hétéroclite de vecteurs aléatoires indépendants au sens de la définition 4.35.

Proposition 4.37 (hérédité de l'indépendance). *Soient X_1, \dots, X_n des variables aléatoires réelles indépendantes. Découpons $\llbracket 1, n \rrbracket$ en k blocs disjoints, en notant m_j le cardinal du j^e bloc et $n_j = \sum_{1 \leq l \leq j} m_l$, avec $n_0 := 0$. Pour $j \in \llbracket 1, k \rrbracket$, associons au j^e bloc une fonction borélienne $h_j : \mathbb{R}^{m_j} \rightarrow \mathbb{R}^{d_j}$. Posons enfin $Z_j := h_j(X_{n_{j-1}+1}, \dots, X_{n_j})$. Alors la suite finie $(Z_j)_{1 \leq j \leq k}$ de vecteurs aléatoires est indépendante (au sens de la définition 4.35).*

En prenant pour chaque j , $d_j = m_j$ et h_j égale à l'identité $\mathbb{R}^{m_j} \rightarrow \mathbb{R}^{m_j}$, on voit que la proposition 4.37 contient en particulier le résultat suivant.

Proposition 4.38 (indépendance de blocs disjoints). *Si X_1, \dots, X_n sont des variables aléatoires réelles indépendantes, les « blocs disjoints » $Y_j := (X_{n_{j-1}+1}, \dots, X_{n_j})$, où $0 = n_0 < n_1 < n_2 < \dots < n_k = n$, sont des vecteurs aléatoires indépendants.*

Le point clé pour démontrer ces deux corollaires est la remarque suivante. Si $X : \Omega \rightarrow E$ est \mathcal{F} - \mathcal{B} mesurable et $h : E \rightarrow E'$ est \mathcal{B} - \mathcal{B}' mesurable, alors la tribu sur Ω engendrée par $h \circ X$ est une sous-tribu de celle engendrée par X , ce qui implique l'hérédité.



IFP p. 181.

Remarque 4.39. Compte-tenu de la remarque page 62 sur l'ordre d'indexation, on généralise facilement les propositions 4.37 et 4.38 au cas où les « blocs » ne sont plus forcément indexés par des entiers consécutifs, pourvu que les blocs d'indices correspondants restent deux à deux disjoints. Ainsi par exemple si X_1, \dots, X_7 sont indépendantes, (X_3, X_1) , (X_2, X_5, X_7) et (X_6, X_4) sont des vecteurs aléatoires indépendants.

Voici maintenant des règles concrètes pour établir l'indépendance des composantes d'un vecteur aléatoire, découlant de la proposition 4.36.

Proposition 4.40.

a) *Les variables aléatoires réelles X_1, \dots, X_d sont indépendantes si et seulement si*

$$\forall (x_1, \dots, x_d) \in \mathbb{R}^d, \quad P(X_1 \leq x_1, \dots, X_d \leq x_d) = P(X_1 \leq x_1) \dots P(X_d \leq x_d). \quad (4.17)$$

b) *Soit (X_1, \dots, X_d) un vecteur aléatoire dans \mathbb{R}^d , de densité f . Ses composantes X_i sont indépendantes si et seulement s'il existe $f_1, \dots, f_d, \mathbb{R} \rightarrow \overline{\mathbb{R}}_+$, mesurables telles que*

$$f = f_1 \otimes \dots \otimes f_d.$$

Dans ce cas les densités marginales f_{X_i} sont de la forme $c_i f_i$ où les constantes $c_i > 0$ sont liées par $c_1 \dots c_d = 1$.

c) *Soient (X_1, \dots, X_d) un vecteur aléatoire discret à valeurs dans \mathbb{N}^d . Ses composantes X_i sont indépendantes si et seulement si*

$$\forall (k_1, \dots, k_d) \in \mathbb{N}^d, \quad P(X_1 = k_1, \dots, X_d = k_d) = P(X_1 = k_1) \dots P(X_d = k_d). \quad (4.18)$$



IFP prop. 5.42, pp. 181–184. Pour des exemples du b), PVIR pp. 319–320.

La dernière propriété importante de l'indépendance qu'il nous reste à examiner est « l'interversion espérance-produit ». Les énoncés précis sont les suivants.

Proposition 4.41. *Soit $(X_i)_{i \in I}$ une famille quelconque de variables aléatoires réelles indépendantes. Alors pour toute partie finie J de I et toute famille $\{h_j, j \in J\}$ de fonctions boréliennes $\mathbb{R} \rightarrow \mathbb{C}$ telles que les $h_j(X_j)$ soient P -intégrables, la variable aléatoire produit $\prod_{j \in J} h_j(X_j)$ est P -intégrable et*

$$\mathbf{E} \left(\prod_{j \in J} h_j(X_j) \right) = \prod_{j \in J} \mathbf{E} h_j(X_j).$$

Corollaire 4.42. *Si les variables aléatoires réelles (ou complexes) X_1, \dots, X_n sont indépendantes et intégrables, leur produit est aussi intégrable et*

$$\mathbf{E}(X_1 \dots X_n) = (\mathbf{E}X_1) \dots (\mathbf{E}X_n).$$

♣ IFP prop. 5.43 et cor. 5.44, pp. 184–185.

Le cas le plus simple pour la formule ci-dessus est $n = 2$ avec $X_1 = \mathbf{1}_A$ et $X_2 = \mathbf{1}_B$. Alors l'indépendance de ces deux variables aléatoires indicatrices équivaut à l'indépendance des événements A et B (regardez la tribu engendrée par chacune pour comprendre). Et comme $\mathbf{1}_A \mathbf{1}_B = \mathbf{1}_{A \cap B}$, l'interversion espérance produit nous donne simplement ici $\mathbf{E}(\mathbf{1}_{A \cap B}) = (\mathbf{E}\mathbf{1}_A)(\mathbf{E}\mathbf{1}_B)$, ce qui s'écrit encore $P(A \cap B) = P(A)P(B)$, soit la définition de l'indépendance de A et B . On peut d'ailleurs faire le chemin à l'envers et partir de cette égalité pour prouver le corollaire 4.42.

♣ PVIR pp. 325–327.

Il est facile de voir que la réciproque du corollaire 4.42 est fautive en construisant un couple (X_1, X_2) de variables aléatoires réelles non indépendantes telles que $\mathbf{E}(X_1 X_2) = \mathbf{E}X_1 \mathbf{E}X_2$. Prenons par exemple X_1 de loi uniforme sur $[-1, +1]$ et $X_2 := X_1^2$. On a alors

$$\mathbf{E}X_1 X_2 = \mathbf{E}X_1^3 = \int_{[-1, +1]} x^3 d\lambda(x) = 0.$$

D'autre part $\mathbf{E}X_1 = 0$ donc $\mathbf{E}X_1 \mathbf{E}X_2 = 0$. Il est clair intuitivement que X_1 et X_2 ne sont pas indépendantes puisque X_2 est une fonction déterministe de X_1 . Pour vérifier cette non-indépendance par le calcul, on peut remarquer que d'une part

$$\begin{aligned} P(X_1 \in [0, 1/2] \text{ et } X_2 \in [0, 1/4]) &= P(X_1 \in [0, 1/2] \text{ et } X_1 \in [-1/2, 1/2]) \\ &= P(X_1 \in [0, 1/2]) = \frac{1}{4} \end{aligned}$$

et d'autre part

$$P(X_1 \in [0, 1/2])P(X_2 \in [0, 1/4]) = \frac{1}{4}P(X_1 \in [-1/2, 1/2]) = \frac{1}{8}.$$

Corollaire 4.43 (indépendance et covariance). *Si X et Y sont deux variables aléatoires de carré intégrable indépendantes, leur covariance est nulle.*

4.2.3 Sommes de variables aléatoires

Lorsque l'on connaît la loi de la v.a. X et celle de la v.a. Y , on ne peut pas en déduire en général la loi de $X + Y$. Mais si l'on sait que X et Y sont indépendantes, alors on peut déterminer la loi du vecteur (X, Y) et à partir de cette dernière calculer la loi de $X + Y$.

♣ PVIR pp. 320–324.

Nous nous contenterons d'énoncer le cas où X et Y sont à densité.

Proposition 4.44 (somme de deux v.a. à densité indépendantes). *Si les variables aléatoires réelles indépendantes X et Y admettent pour densités respectives f et g , leur somme $S = X + Y$ admet pour densité le produit de convolution $f * g$ défini sur \mathbb{R} par*

$$(f * g)(s) = \int_{-\infty}^{+\infty} f(s - t)g(t) dt = \int_{-\infty}^{+\infty} f(t)g(s - t) dt. \quad (4.19)$$



Pour une vue plus générale de la convolution de deux probabilités, IFP pp. 187–189.

Nous donnons maintenant deux inégalités importantes qui permettent de contrôler le comportement en probabilité d'une somme de n variables aléatoires sans avoir à calculer la loi de cette somme. La première est valide dans un cadre bien plus large que l'indépendance.

Les X_k sont dites deux à deux *non-corrélées* si $\text{Cov}(X_i, X_j) = 0$ pour tous i, j distincts. Ceci se produit en particulier lorsque les X_k sont deux à deux indépendantes, cf. cor. 4.43. Pour des X_k deux à deux non-corrélées, on déduit de (4.9) l'égalité

$$\text{Var } S_n = \sum_{k=1}^n \text{Var } X_k. \quad (4.20)$$

Proposition 4.45 (inégalité de Bienaymé-Tchebycheff). *Si les X_k sont de carré intégrable et deux à deux non-corrélées,*

$$\forall t > 0, \quad P\left(\left|\sum_{k=1}^n (X_k - \mathbf{E}X_k)\right| \geq t\right) \leq \frac{1}{t^2} \sum_{k=1}^n \text{Var } X_k. \quad (4.21)$$



IFP p. 239.

Théorème 4.46 (Inégalité de Kolmogorov). *Soient X_1, \dots, X_n des variables aléatoires réelles indépendantes, de carré intégrable $S_k := \sum_{i=1}^k X_i$. Alors*

$$\forall t > 0, \quad P\left(\max_{1 \leq k \leq n} |S_k - \mathbf{E}S_k| \geq t\right) \leq \frac{1}{t^2} \text{Var } S_n. \quad (4.22)$$

Remarque 4.47. L'inégalité de Kolmogorov ressemble formellement à celle de Bienaymé-Tchebycheff. Elle est cependant beaucoup plus puissante, puisqu'elle permet de contrôler en probabilité les déviations de *toute la suite* finie $(S_k - \mathbf{E}S_k)_{1 \leq k \leq n}$ au lieu de seulement son dernier terme $S_n - \mathbf{E}S_n$ pour Bienaymé-Tchebycheff. L'hypothèse est aussi plus restrictive (indépendance mutuelle au lieu de non-corrélation deux à deux).



IFP th. 8.8 pp. 240–241.

Chapitre 5

Théorèmes limite

Ce chapitre débute l'étude du comportement asymptotique de suites de variables aléatoires. Après avoir vu les différents modes de convergence de ces suites, nous abordons la loi des grands nombres. Ce résultat essentiel nous dit que les moyennes arithmétiques d'une suite de v.a. X_i indépendantes et de même loi ayant une espérance, convergent en un certain sens, vers cette espérance :

$$M_n := \frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{} \mathbf{E}X_1. \quad (5.1)$$

Cette convergence est très utile en statistique pour *estimer* des paramètres d'une loi inconnue, sur la base de l'observation d'un *échantillon* X_1, \dots, X_n de grande taille.

5.1 Convergences de suites de variables aléatoires

5.1.1 Convergence presque sûre et en probabilité

Quand on envisage la question de la convergence d'une suite de v.a. (Y_n) vers une v.a. Y , la première notion de convergence qui vient à l'esprit est la convergence simple sur tout Ω , au sens de l'analyse¹ :

$$\forall \omega \in \Omega, \quad Y_n(\omega) \xrightarrow[n \rightarrow +\infty]{} Y(\omega).$$

On voit immédiatement que cette notion n'est pas satisfaisante pour la convergence de la suite (M_n) donnée en (5.1). En effet considérons le modèle probabiliste infini le plus simple possible, à savoir le jeu de pile ou face infini. On peut prendre ici $\Omega = \{\mathbf{f}, \mathbf{p}\}^{\mathbb{N}^*}$ et ω est une suite infinie $\omega = (u_i)_{i \geq 1}$, avec $u_i \in \{\mathbf{f}, \mathbf{p}\}$ pour tout i . En prenant pour X_i l'indicatrice de l'évènement obtention de pile au i^{e} lancer, M_n est la fréquence d'apparition de pile au cours des n premiers lancers. Si la pièce est équilibrée, on s'attend à ce que M_n converge vers $1/2$. Or il est clair qu'il y a une infinité d'évènements élémentaires ω pour lesquels $M_n(\omega)$ ne converge pas vers $1/2$. On peut même construire facilement une infinité de ω pour lesquels $M_n(\omega)$ n'a *aucune* limite². Ce simple exemple montre que la notion de convergence simple n'est pas pertinente en théorie des probabilités. Pour dépasser ce problème, on introduit la notion de convergence presque sûre, qui est la convergence simple sur un sous-ensemble Ω' de probabilité 1 de Ω .

1. Ceci suppose que les Y_n et Y sont définies sur le même Ω .

2. Pour approfondir cette question, voir la section « 6.5 Discussion » dans ICP.

Définition 5.1 (convergence presque sûre). Soient $(Y_n)_{n \geq 1}$ et Y des variables aléatoires définies sur le même espace probabilisé (Ω, \mathcal{F}, P) . Notons

$$\Omega' := \{\omega \in \Omega ; \lim_{n \rightarrow +\infty} Y_n(\omega) = Y(\omega)\}. \quad (5.2)$$

Si $P(\Omega') = 1$, on dit qu' Y_n converge presque sûrement vers Y , notation $Y_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} Y$.

Cette définition soulève immédiatement une question. Pour que l'égalité $P(\Omega') = 1$ ait un sens, encore faut-il que Ω' appartienne à la tribu \mathcal{F} sur laquelle est définie la fonction d'ensembles P . Pour établir l'appartenance à \mathcal{F} de Ω' , il est naturel de s'appuyer sur la mesurabilité des Y_n et de Y dont héritent les v.a. positives $|Y_n - Y|$. Pour cela, on commence par écrire avec des quantificateurs l'appartenance à Ω' :

$$\omega \in \Omega' \Leftrightarrow \forall \varepsilon > 0, \exists j = j(\omega, \varepsilon) \in \mathbb{N}, \forall k \geq j, |Y_k(\omega) - Y(\omega)| < \varepsilon. \quad (5.3)$$

En utilisant la traduction automatique des quantificateurs en opérations ensemblistes, on en déduit que

$$\Omega' = \bigcap_{\varepsilon > 0} \bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon\}. \quad (5.4)$$

En lisant (5.4) de droite à gauche, on obtient les appartenances successives à \mathcal{F} , d'abord des ensembles $\{|Y_k - Y| < \varepsilon\}$ en raison de la mesurabilité des v.a. $|Y_n - Y|$, puis de l'intersection dénombrable sur k , puis de l'union dénombrable sur j . Arrivés là, on est coincé car la dernière intersection sur ε a pour ensemble d'indexation $]0, +\infty[$ qui est infini *non*-dénombrable. On ne peut donc pas en déduire l'appartenance à \mathcal{F} de Ω' . Pour franchir cet obstacle, il suffit de revenir à (5.3) et d'écrire une version « discrétisée » de la convergence de $Y_n(\omega)$ vers $Y(\omega)$. Pour cela on choisit une suite de réels strictement positifs $(\varepsilon_i)_{i \geq 1}$, tendant vers 0 et on écrit que

$$\omega \in \Omega' \Leftrightarrow \forall i \geq 1, \exists j = j(\omega, i) \in \mathbb{N}, \forall k \geq j, |Y_k(\omega) - Y(\omega)| < \varepsilon_i. \quad (5.5)$$

La traduction automatique des quantificateurs nous donne maintenant

$$\Omega' = \bigcap_{i \in \mathbb{N}^*} \bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon_i\}. \quad (5.6)$$

Sous cette forme, il est maintenant clair que Ω' appartient à \mathcal{F} et ceci légitime la définition 5.1. Nous avons établi au passage que

$$Y_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} Y \Leftrightarrow P\left(\bigcap_{i \in \mathbb{N}^*} \bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon_i\}\right) = 1, \quad (5.7)$$

pour une suite de réels $\varepsilon_i > 0$, tendant vers 0.

Pour l'instant, (5.7) n'est pas directement exploitable comme méthode pratique pour montrer une convergence presque sûre, mais on peut progresser dans cette direction en « faisant sortir le $\forall i$ » de la probabilité. Cette opération est légitimée par le lemme suivant.

Lemme 5.2. Si $(B_i)_{i \geq 1}$ est une suite d'événements, on a l'équivalence

$$P\left(\bigcap_{i \geq 1} B_i\right) = 1 \Leftrightarrow \forall i \geq 1, P(B_i) = 1. \quad (5.8)$$

 PVIR p. 335.

En appliquant le lemme 5.2 aux évènements $B_i := \bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon_i\}$, nous obtenons à partir de (5.7) l'équivalence entre la convergence presque sûre de Y_n vers Y et la condition

$$\forall i \in \mathbb{N}^*, P\left(\bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon_i\}\right) = 1, \quad (5.9)$$

Maintenant que nous avons réussi à sortir le « $\forall \varepsilon$ » de la probabilité, on peut laisser tomber la discrétisation pour aboutir à la caractérisation suivante de la convergence presque sûre.

Proposition 5.3 (une c.n.s. de convergence p.s.). *La suite de variables aléatoires $(Y_n)_{n \geq 1}$ converge presque sûrement vers la variable aléatoire Y si et seulement si*

$$\forall \varepsilon > 0, \quad P\left(\bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon\}\right) = 1, \quad (5.10)$$

ou de manière équivalente,

$$\forall \varepsilon > 0, \quad P\left(\bigcap_{j \in \mathbb{N}} \bigcup_{k \geq j} \{|Y_k - Y| \geq \varepsilon\}\right) = 0. \quad (5.11)$$

 PVIR p. 336.

Intéressons nous maintenant à (5.11). En notant $A_k := \{|Y_k - Y| \geq \varepsilon\}$, on aimerait trouver une condition qui nous assure que $P(\bigcap_{j \in \mathbb{N}} \bigcup_{k \geq j} A_k) = 0$. En pratique, on a souvent une majoration des probabilités $P(A_k)$ via une inégalité du type inégalité de Markov et il est donc naturel de chercher une condition portant sur les $P(A_k)$. Ce problème a un intérêt propre, indépendant de la définition des A_k , provenant de l'interprétation suivante de l'évènement $\bigcap_{j \in \mathbb{N}} \bigcup_{k \geq j} A_k$:

$$\begin{aligned} \bigcap_{j \in \mathbb{N}} \bigcup_{k \geq j} A_k &= \{\omega \in \Omega ; \omega \text{ appartient à une infinité de } A_k\} \\ &= \{\omega \in \Omega ; \omega \text{ réalise une infinité de } A_k\} \\ &= \{\text{réalisation d'une infinité de } A_k\}. \end{aligned}$$

L'étude de la probabilité de cet évènement conduit aux deux lemmes de Borel Cantelli.

Lemme 5.4 (Borel Cantelli I). *Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'évènements telle que*

$$\sum_{n=0}^{+\infty} P(A_n) < +\infty. \quad (5.12)$$

Alors

$$P(\text{réalisation d'une infinité de } A_n) = 0.$$

 PVIR lemme 9.4 p. 336.

Le lemme de Borel Cantelli I est d'une portée très générale, puisqu'on obtient la conclusion $P(A_\infty) = 0$ sous la seule hypothèse de convergence de la série $\sum_{n \in \mathbb{N}} P(A_n)$, sans rien supposer sur la structure de dépendance de la suite (A_n) . Pour les suites d'évènements indépendants, on a le résultat complémentaire suivant.

Lemme 5.5 (Borel Cantelli II). *Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'évènements indépendants telle que*

$$\sum_{n=0}^{+\infty} P(A_n) = +\infty. \quad (5.13)$$

Alors

$$P(\text{réalisation d'une infinité de } A_n) = 1.$$

♣ PVIR lemme 9.5 p. 337 ou IFP pp. 178–179.

Revenons à notre quête d'une condition pratique de convergence presque sûre. Le premier lemme de Borel Cantelli nous permet d'aboutir au résultat suivant.

Proposition 5.6 (condition suffisante de convergence p.s.). *Soient Y et $(Y_n)_{n \geq 1}$ des variables aléatoires réelles définies sur le même espace probabilisé (Ω, \mathcal{F}, P) et vérifiant*

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{+\infty} P(|Y_n - Y| \geq \varepsilon) < +\infty. \quad (5.14)$$

Alors Y_n converge presque sûrement vers Y .

Si $(Y_n)_{n \geq 1}$ vérifie (5.14), on dit qu'elle converge *presque complètement* vers Y .

Nous introduisons maintenant un nouveau mode de convergence de Y_n vers Y , la convergence en probabilité. Cette notion nous sera utile pour la loi faible des grands nombres.

Définition 5.7 (convergence en probabilité). *La suite Y_n converge en probabilité vers Y (notation $Y_n \xrightarrow[n \rightarrow +\infty]{\text{Pr}} Y$) si*

$$\forall \varepsilon > 0, \quad P(|Y_n - Y| \geq \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0.$$

La convergence ci-dessus devant avoir lieu pour *tout* $\varepsilon > 0$, il est clair qu'on obtient une définition équivalente en remplaçant $P(|Y_n - Y| \geq \varepsilon)$ par $P(|Y_n - Y| > \varepsilon)$.

Notons la différence de point de vue par rapport à la convergence p.s. Dans la convergence p.s., on reste proche de la notion de convergence simple de l'analyse. Il s'agit de la convergence de la suite de réels $(Y_n(\omega))_{n \geq 1}$, pour tous les ω d'un même évènement de probabilité 1. La convergence en probabilité ne concerne pas le comportement individuel de chaque suite $(Y_n(\omega))_{n \geq 1}$, mais plutôt celui de la suite d'évènements $D_{n,\varepsilon} := \{|Y_n - Y| < \varepsilon\}$ dont la probabilité $P(D_{n,\varepsilon})$ doit tendre vers 1, pour tout ε . En pratique, établir la convergence en probabilité de Y_n vers Y est souvent un travail préliminaire pour prouver la convergence p.s. de Y_n vers Y . En effet si on utilise la condition suffisante de convergence p.s. (5.14), pour que la série converge, il faut que son terme général tende vers 0 pour tout ε , ce qui est précisément la convergence en probabilité de Y_n vers Y .

Proposition 5.8. *La convergence presque-sûre implique la convergence en probabilité.*

♣ PVIR pp. 339–340 ou IFP p. 236.

Remarque 5.9. La convergence en probabilité n'implique pas la convergence presque-sûre. Pour un contre exemple, voir PVIR p. 340.

Remarque 5.10. Il est possible d'obtenir la convergence presque sûre à partir de la convergence en probabilité, à condition d'avoir une *bonne vitesse* de convergence en probabilité. Le sens précis de cette affirmation est donné par la proposition 5.6.

La proposition 5.6 permet aussi, même sans bonne vitesse de convergence en probabilité, d'obtenir de la convergence p.s. pour une *sous-suite*.

Proposition 5.11 (convergence p.s. d'une sous-suite). *Si la suite $(Y_n)_{n \geq 1}$ converge en probabilité vers Y , on peut en extraire une sous-suite $(Y_{n_i})_{i \geq 1}$ qui converge presque sûrement vers Y .*



IFP pp. 237–238 ou PVIR p. 341.

5.1.2 Convergence en moyenne d'ordre p

Nous introduisons maintenant un nouveau mode de convergence, utile notamment dans les problèmes d'interversion limite espérance.

Définition 5.12. Soit $p \geq 1$ un réel et $(Y_n)_{n \geq 1}$ une suite de v.a. ayant un moment absolu d'ordre p fini. On dit que cette suite converge en moyenne d'ordre p (ou au sens L^p) vers la v.a. Y si

$$\lim_{n \rightarrow +\infty} \mathbf{E}(|Y_n - Y|^p) = 0. \quad (5.15)$$

Notation : $Y_n \xrightarrow[n \rightarrow +\infty]{L^p} Y$.

Remarque 5.13. Si Y_n converge vers Y en moyenne d'ordre p , Y a nécessairement un moment absolu d'ordre p fini. Pour le voir, on utilise la *convexité* de la fonction $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, $x \mapsto x^p$, pour $p \geq 1$. Ceci signifie que la « corde » entre deux points de son graphe est « au dessus » de l'arc de courbe correspondant, ou encore que « l'image du barycentre » est majorée par « le barycentre des images », voir la figure 5.1. Cette convexité implique notamment que pour tous $a, b \geq 0$, $\varphi\left(\frac{a+b}{2}\right) \leq \frac{1}{2}\varphi(a) + \frac{1}{2}\varphi(b)$, ce qui s'écrit encore

$$\forall a, b \geq 0, \quad (a + b)^p \leq 2^{p-1}(a^p + b^p). \quad (5.16)$$

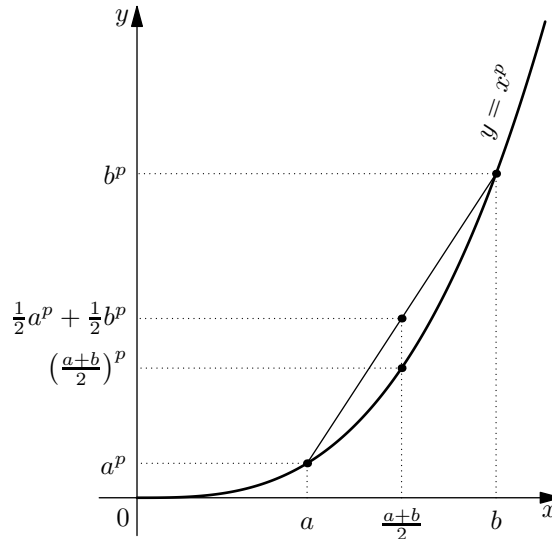


FIGURE 5.1 – Convexité de $x \mapsto x^p$ et inégalité $\left(\frac{a+b}{2}\right)^p \leq \frac{1}{2}(a^p + b^p)$

Par croissance de φ , inégalité triangulaire et (5.16) appliquée avec $a = |Y_n(\omega)|$ et $b = |Y(\omega) - Y_n(\omega)|$, on voit que pour tout $\omega \in \Omega$,

$$|Y(\omega)|^p \leq (|Y_n(\omega)| + |Y(\omega) - Y_n(\omega)|)^p \leq 2^{p-1}|Y_n(\omega)|^p + 2^{p-1}|Y(\omega) - Y_n(\omega)|^p.$$

Par croissance de l'espérance, on en déduit :

$$\mathbf{E}|Y|^p \leq 2^{p-1}\mathbf{E}|Y_n|^p + 2^{p-1}\mathbf{E}|Y - Y_n|^p.$$

Ce majorant est fini car $\mathbf{E}|Y_n|^p$ est fini par hypothèse et $\mathbf{E}|Y - Y_n|^p$ tend vers 0 donc est fini au moins pour n assez grand.

Proposition 5.14. *La convergence en moyenne d'ordre p implique la convergence en probabilité.*

Preuve. C'est une conséquence immédiate de l'inégalité de Markov avec moment, cf. la proposition 3.32, puisque

$$\forall \varepsilon > 0, \quad P(|Y_n - Y| \geq \varepsilon) \leq \frac{\mathbf{E}(|Y_n - Y|^p)}{\varepsilon^p}.$$

□

Proposition 5.15 (hiérarchie des convergences L^p). *Si $1 \leq p < r < +\infty$ et si $Y_n \xrightarrow[n \rightarrow +\infty]{L^r} Y$, alors $Y_n \xrightarrow[n \rightarrow +\infty]{L^p} Y$.*

La démonstration de la proposition 5.15 est une conséquence immédiate de l'inégalité

$$(\mathbf{E}|X|^p)^{1/p} \leq (\mathbf{E}|X|^r)^{1/r}, \quad 1 \leq p < r < +\infty.$$

5.1.3 Bilan sur les modes de convergence

Arrivé à ce stade, il est bon de faire le point sur les différents modes de convergence étudiés. Le diagramme de la figure 5.2 résume les relations entre ces modes de convergence. Les flèches en trait plein représentent des implications (si la suite converge selon le mode de la case de départ, alors elle converge aussi selon celui de la case d'arrivée). Les flèches en tirets signalent l'existence d'une sous-suite convergente selon le mode de la case d'arrivée.

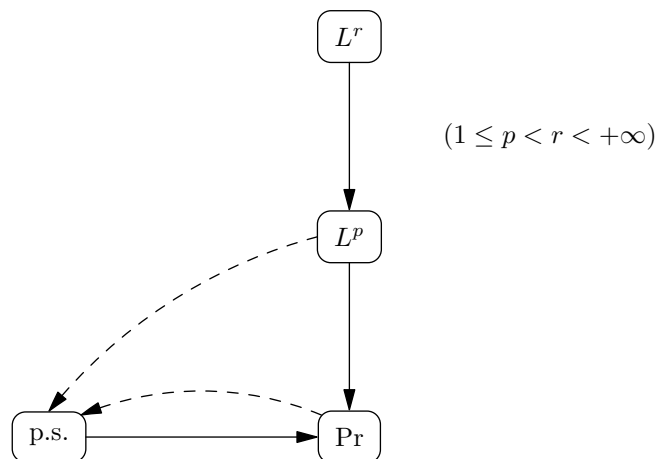


FIGURE 5.2 – Diagramme des convergences des suites de v.a.

Signalons deux propriétés communes aux trois modes de convergence étudiés jusqu'ici (vérification laissée au lecteur).

D'abord la limite pour ces modes de convergence n'est pas à strictement parler unique. Elle l'est *modulo* l'égalité presque sûre. Notons (m) , (m') l'un des trois modes de convergence (p.s., en probabilité ou L^p). On peut vérifier que si Y_n converge vers Y au sens (m) et vers Y' au sens (m) alors $Y = Y'$ presque-sûrement (la réciproque est évidente). D'autre part si Y_n converge au sens (m) vers Y et au sens (m') vers Y' , ces deux convergences impliquent la convergence en probabilité de Y_n vers Y et vers Y' , donc l'égalité p.s. de Y et Y' .

Chacun des trois modes de convergence est compatible avec la structure d'espace vectoriel. Si $X_n \xrightarrow{(m)} X$ et $Y_n \xrightarrow{(m)} Y$, $X_n + Y_n \xrightarrow{(m)} X + Y$ et $aX_n \xrightarrow{(m)} aX$ pour tout $a \in \mathbb{R}$.

5.2 Loi des grands nombres

Nous abordons maintenant les lois des grands nombres. Il s'agit d'étudier la convergence des moyennes arithmétiques S_n/n construites à partir d'une suite de variables aléatoires $(X_i)_{i \geq 1}$ en posant $S_n := X_1 + \dots + X_n$. Si S_n/n converge en probabilité, on parle d'une loi faible des grands nombres, tandis que si elle converge presque-sûrement on parle d'une loi forte.

5.2.1 Loi faible des grands nombres

Rappelons que si X est de carré intégrable ($\mathbf{E}X^2 < +\infty$), sa variance est définie par

$$\text{Var } X := \mathbf{E}(X - \mathbf{E}X)^2.$$

Les X_k sont dites deux à deux *non-corrélées* si $\text{Cov}(X_i, X_j) = 0$ pour tous i, j distincts. Ceci se produit en particulier lorsque les X_k sont deux à deux indépendantes, cf. cor. 4.43. Pour des X_k deux à deux non-corrélées, on déduit de (4.9) l'égalité

$$\text{Var } S_n = \sum_{k=1}^n \text{Var } X_k. \quad (5.17)$$

Théorème 5.16 (loi faible des grands nombres). *Si les X_k sont de même loi, de carré intégrable et deux à deux non-corrélées, on a la convergence en probabilité :*

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow +\infty]{\text{Pr}} \mathbf{E}X_1. \quad (5.18)$$

 IFP p. 239.

5.2.2 Lois fortes des grands nombres

Théorème 5.17 (loi forte des grands nombres de Kolmogorov). *On suppose que $(X_k)_{k \geq 1}$ est une suite de variables aléatoires réelles indépendantes, de carré intégrable, d'espérance nulle et que*

$$\sum_{j=1}^{+\infty} \frac{\text{Var } X_j}{j^2} < +\infty. \quad (5.19)$$

Alors

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0. \quad (5.20)$$

Remarque 5.18. Si les X_k ne sont pas centrées, en appliquant (5.20) aux $X'_k := X_k - \mathbf{E}X_k$, on obtient

$$\frac{1}{n}S_n - \frac{1}{n} \sum_{k=1}^n \mathbf{E}X_k \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0. \quad (5.21)$$

Ce résultat est intéressant notamment dans le cas où la suite déterministe $(\mathbf{E}X_n)_{n \geq 1}$ a une limite finie ℓ quand n tend vers l'infini. En effet le théorème de Césaro nous donne alors la convergence vers ℓ de $n^{-1} \sum_{k=1}^n \mathbf{E}X_k$, d'où la convergence presque sûre de $n^{-1}S_n$ vers ℓ . En particulier si les X_k ont même espérance, $n^{-1}S_n$ converge p.s. vers $\mathbf{E}X_1$.

♣ IFP pp. 242–243.

Théorème 5.19 (loi forte des grands nombres pour des v.a. i.i.d.). *On suppose les X_k indépendantes, de même loi et $\mathbf{E}|X_1| < +\infty$. Alors*

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbf{E}X_1. \quad (5.22)$$

♣ IFP pp. 244–246.

Le théorème précédent est le résultat optimal pour les variables aléatoires indépendantes et de même loi (i.i.d.) en raison de la « réciproque » suivante.

Théorème 5.20. *Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires indépendantes et de même loi telle que S_n/n converge presque sûrement. Alors $\mathbf{E}|X_1| < +\infty$.*

La preuve ci-dessous est mieux adaptée à ce cours que celle qui figure dans IFP p. 247.

Preuve. Par hypothèse, S_n/n converge p.s. vers une certaine variable aléatoire Y . Notons

$$\Omega' = \left\{ \omega \in \Omega ; \frac{S_n(\omega)}{n} \xrightarrow[n \rightarrow +\infty]{} Y(\omega) \right\}.$$

Alors $P(\Omega') = 1$ et pour tout $\omega \in \Omega'$,

$$\frac{X_n(\omega)}{n} = \frac{S_n(\omega) - S_{n-1}(\omega)}{n} = \frac{S_n(\omega)}{n} - \frac{n-1}{n} \times \frac{S_{n-1}(\omega)}{n-1} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0,$$

puisque $S_n(\omega)/n$ et $S_{n-1}(\omega)/(n-1)$ convergent vers le même réel $Y(\omega)$. Considérons l'évènement

$$B = \left\{ \omega \in \Omega ; \exists n_0 = n_0(\omega), \forall n \geq n_0, \frac{|X_n(\omega)|}{n} < 1 \right\}.$$

Puisque $P(\Omega') = 1$ et $\Omega' \subset B$, l'évènement B a pour probabilité 1, d'où en passant au complémentaire

$$P\left(\frac{|X_n|}{n} \geq 1 \text{ une infinité de fois}\right) = 0.$$

Il résulte alors du second lemme de Borel-Cantelli (lemme 5.5) que

$$\sum_{n \geq 1} P\left(\frac{|X_n|}{n} \geq 1\right) < +\infty.$$

Les X_i ayant même loi, ceci s'écrit aussi

$$\sum_{n=1}^{+\infty} P(|X_1| \geq n) < +\infty. \quad (5.23)$$

Pour finir la preuve, on observe (faites un dessin !) que

$$\begin{aligned}\mathbf{E}|X_1| &= \int_0^{+\infty} P(|X_1| > t) \, dt = \sum_{n=0}^{+\infty} \int_n^{(n+1)} P(|X_1| > t) \, dt \\ &\leq \sum_{n=0}^{+\infty} P(|X_1| \geq n) = 1 + \sum_{n=1}^{+\infty} P(|X_1| \geq n) < +\infty.\end{aligned}$$

□



Pour des applications des lois des grands nombres, voir IFP section 8.4.


Chapitre 6

Théorème limite central

Le théorème limite central nous dit qu'une somme d'un grand nombre de variables aléatoires indépendantes, de carré intégrable, convenablement normalisée, se comporte asymptotiquement « en loi » comme une variable aléatoire gaussienne. Il explique l'importance centrale des lois gaussiennes dans la théorie des probabilités et la statistique. Il complète la loi des grands nombres en donnant une sorte de vitesse de convergence, permettant notamment de construire des « intervalles de confiance » pour l'estimation d'un paramètre.

Pour donner un sens mathématique précis à cette notion de « comportement asymptotique en loi », il nous faut d'abord introduire la convergence en loi.

6.1 Convergence en loi

 Discussion introductive dans IFP pp. 271–272.

Nous admettrons l'équivalence des deux définitions suivantes.

Définition 6.1 (convergence en loi). *Notons F_n et F les fonctions de répartition respectives des variables aléatoires réelles Y_n ($n \geq 1$) et Y . On dit que la suite $(Y_n)_{n \geq 1}$ converge en loi vers Y si*

$$\text{pour tout } x \text{ point de continuité de } F, \quad F_n(x) \xrightarrow[n \rightarrow +\infty]{} F(x). \quad (6.1)$$

Rappelons que x est point de continuité de la fonction de répartition F si et seulement si $F(x-) = F(x)$ ou encore $P(Y = x) = 0$.

Définition 6.2 (convergence en loi). *On dit que la suite $(Y_n)_{n \geq 1}$ de variables aléatoires réelles converge en loi vers la variable aléatoire réelle Y si*

$$\text{pour toute } h \text{ continue bornée } \mathbb{R} \rightarrow \mathbb{R}, \quad \mathbf{E}h(Y_n) \xrightarrow[n \rightarrow +\infty]{} \mathbf{E}h(Y). \quad (6.2)$$

Remarquons que si h est continue bornée, les $h(Y_n)$ et $h(Y)$ sont des v.a. bornées, donc intégrables. Nous noterons la convergence en loi de Y_n vers Y par

$$Y_n \xrightarrow[n \rightarrow +\infty]{\text{loi}} Y.$$

 Pour une preuve de l'équivalence de ces deux définitions, cf. IFP th. 9.32 p. 275.

La définition 6.1 est la plus concrète, surtout lorsque F est continue sur tout \mathbb{R} , cas souvent rencontré en pratique. En effet dans ce cas, la convergence en loi équivaut à la convergence simple sur \mathbb{R} des fonctions de répartition et nous donne, pour tous réels $a < b$, la convergence des $P(Y_n \in I(a, b))$ vers les $P(Y \in I(a, b))$, où $I(a, b)$ désigne n'importe lequel des 4 intervalles d'extrémités a et b .

La définition 6.2 est souvent plus commode pour établir les propriétés de la convergence en loi et a l'intérêt d'une généralisation immédiate aux vecteurs aléatoires de \mathbb{R}^d .

Définition 6.3 (convergence en loi de vecteurs aléatoires). *On dit que la suite $(Y_n)_{n \geq 1}$ de vecteurs aléatoires de \mathbb{R}^d converge en loi vers le vecteur aléatoire Y de \mathbb{R}^d si*

$$\text{pour toute } h \text{ continue bornée } \mathbb{R}^d \rightarrow \mathbb{R}, \quad \mathbf{E}h(Y_n) \xrightarrow{n \rightarrow +\infty} \mathbf{E}h(Y). \quad (6.3)$$

Remarques 6.4 (les pièges de la convergence en loi). Pointons d'emblée des différences importantes entre la convergence en loi et les autres modes de convergence vus jusqu'ici.

1. Il n'est pas nécessaire, pour la convergence en loi de Y_n vers Y , que ces variables aléatoires soient définies *sur le même* (Ω, \mathcal{F}, P) .
2. Il n'y a pas unicité de la v.a. limite en loi. Si $(Y_n)_{n \geq 1}$ converge en loi vers Y , elle converge aussi en loi vers *n'importe quelle variable aléatoire Z ayant même loi* que Y (éventuellement définie sur un autre espace probabilisé). Ceci se voit facilement sur chacune des deux définitions de la convergence en loi¹. Réciproquement si Y_n converge en loi vers Y et aussi vers Z , alors Y et Z ont même loi. En effet en utilisant la définition 6.2 et l'unicité de la limite d'une suite convergente de réels, on voit que $\mathbf{E}h(Y) = \mathbf{E}h(Z)$ pour toute $h : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée. Par la caractérisation des lois par leurs h -moments, cf. la proposition 4.16 avec $d = 1$, on en déduit que Y et Z ont même loi. En résumé, s'il n'y a pas unicité de la v.a. limite en loi, il y a unicité de sa loi, que l'on appellera *loi limite*².
3. La convergence en loi n'est pas compatible avec l'addition. Si X_n converge en loi vers X et si Y_n converge en loi vers Y , il est faux en général que $X_n + Y_n$ converge en loi vers $X + Y$. En effet si c'était le cas, comme X_n converge en loi vers n'importe quel X' ayant même loi que X , $X_n + Y_n$ devrait converger aussi en loi vers $X' + Y$. Le *hic* c'est que $X + Y$ n'a pas forcément même loi que $X' + Y$.

 Pour un exemple intéressant de convergence en loi établie à l'aide des f.d.r., voir « Une loi limite d'extrêmes », IFP p. 282.

Une propriété bien commode de la convergence en loi est sa conservation par image continue.

Proposition 6.5 (convergence en loi par image continue). *Si Y_n converge en loi vers Y , alors pour toute f continue $\mathbb{R} \rightarrow \mathbb{R}$, $f(Y_n)$ converge en loi vers $f(Y)$.*

Noter que l'on ne suppose pas f bornée sur \mathbb{R} .

1. Cette non-unicité de la limite est bien plus générale que pour les autres modes de convergence vus jusqu'ici où l'on avait convergence vers n'importe quelle Z égale p.s. à Y . Bien sûr, si Y et Z sont définies sur le même espace et sont égales p.s., elles ont même loi, mais la réciproque est grossièrement fausse. Quand on lance deux dés, on n'est pas sûr d'obtenir un double!

2. Ceci incite à voir la convergence en loi de Y_n vers Y comme la *convergence de la loi* P_{Y_n} vers la loi P_Y . On pourrait d'ailleurs donner un sens mathématique précis à cette convergence, appelée *convergence étroite des mesures de probabilité* en notant que $\mathbf{E}h(Y_n)$ ne dépend que de h et de P_{Y_n} , cf. IFP p. 272.

Preuve. D'après la définition 6.2, il nous faut vérifier que pour toute fonction *continue bornée* $g : \mathbb{R} \rightarrow \mathbb{R}$, $\mathbf{E}g(f(Y_n))$ tend vers $\mathbf{E}g(f(Y))$ quand n tend vers l'infini. Or la fonction $g \circ f$ est continue sur \mathbb{R} par composition et bornée sur \mathbb{R} par $\sup_{t \in \mathbb{R}} |g(t)|$. On sait par hypothèse que $\mathbf{E}h(Y_n)$ converge vers $\mathbf{E}h(Y)$ pour toute h continue bornée sur \mathbb{R} . En appliquant ceci avec $h = g \circ f$, on obtient la conclusion souhaitée. \square

La preuve ci-dessus se généralise immédiatement aux vecteurs aléatoires.

Proposition 6.6 (convergence en loi de vecteurs par image continue). *Si les Y_n et Y sont des vecteurs aléatoires de \mathbb{R}^d tels que Y_n converge en loi vers Y , alors pour toute f continue $\mathbb{R}^d \rightarrow \mathbb{R}^j$, $f(Y_n)$ converge en loi vers $f(Y)$ dans \mathbb{R}^j .*

En appliquant la proposition 6.6 avec pour f les projections canoniques $\pi_k : \mathbb{R}^d \rightarrow \mathbb{R}$, $x = (x_1, \dots, x_d) \mapsto x_k$, on en déduit le corollaire suivant.

Corollaire 6.7 (convergence en loi des composantes). *Si le vecteur aléatoire Y_n converge en loi dans \mathbb{R}^d vers un vecteur aléatoire Y , alors chacune de ses composantes converge en loi dans \mathbb{R} vers la composante correspondante de Y .*

Il importe de comprendre que la réciproque *est fausse*. Pour s'en convaincre, examinons le cas $d = 2$ en supposant que les variables aléatoires réelles X_n et Y_n convergent en loi respectivement vers X et Y . Si ces convergences en loi impliquaient celle du couple (X_n, Y_n) vers le couple (X, Y) , alors en prenant l'image par l'application continue $(s, t) \mapsto s + t$, on en déduirait la convergence en loi de $X_n + Y_n$ vers $X + Y$. Ceci serait en contradiction avec la remarque 6.4-3.

Corollaire 6.8. *Si le vecteur aléatoire (X_n, Y_n) de \mathbb{R}^2 converge en loi vers le vecteur aléatoire (X, Y) , alors*

$$a) \quad \forall (a, b) \in \mathbb{R}^2, \quad aX_n + bY_n \xrightarrow[n \rightarrow +\infty]{\text{loi}} aX + bY,$$

$$bn) \quad X_n Y_n \xrightarrow[n \rightarrow +\infty]{\text{loi}} XY.$$

Le diagramme des convergences de la figure 6.1, qui reprend et complète celui de la figure 5.2 avec les mêmes conventions, indique que la convergence en loi est la plus faible des convergences de suites de variables aléatoires. Cette affirmation se justifie par le résultat suivant.

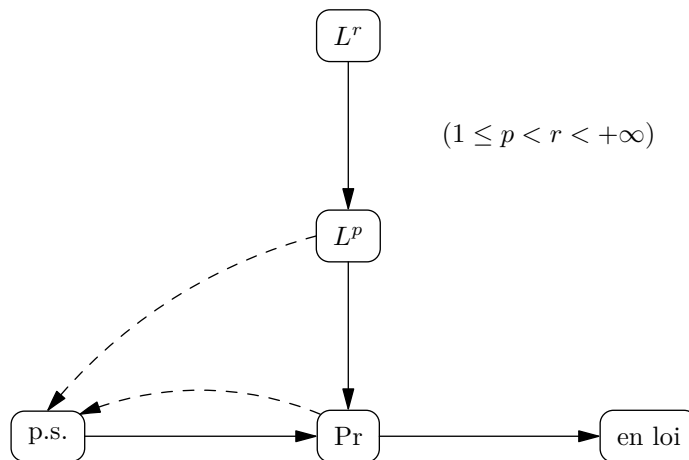


FIGURE 6.1 – Diagramme des convergences des suites de v.a.

Proposition 6.9. *La convergence en probabilité implique la convergence en loi : si les Y_n ($n \geq 1$) et Y sont des variables aléatoires réelles définies sur le même espace probabilisé (Ω, \mathcal{F}, P) telles que Y_n converge en probabilité vers Y , alors Y_n converge aussi en loi vers Y .*



Pour la preuve, voir PVIR p. 367 de préférence à IFP p. 283.

Remarques 6.10 (convergences en probabilité et en loi).

1. La convergence en loi n'implique pas la convergence en probabilité. Voici un contre exemple qui est l'un des plus simples possibles. On prend $\Omega = \{-1, 1\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et P l'équiprobabilité sur Ω . On définit sur Ω la variable aléatoire Y_1 comme l'identité sur Ω . On construit la suite (Y_n) en posant pour $n \geq 2$, $Y_n = (-1)^{n-1}Y_1$. Alors la loi sous P de chaque Y_n est la loi de Rademacher : $P(Y_n = -1) = P(Y_n = 1) = 1/2$. Donc Y_n converge en loi trivialement vers Y_1 (et aussi vers $-Y_1$ qui a même loi). Mais il n'y a pas convergence en probabilité car pour tout $k \in \mathbb{N}^*$ et tout $\varepsilon \in]0, 2[$, $P(|Y_{2k} - Y_1| > \varepsilon) = P(2|Y_1| > \varepsilon) = 1$.
2. La réciproque de la proposition 6.9 est vraie dans le cas très particulier où la v.a. limite est *constante*. Supposons en effet que Y_n converge en loi vers la constante c . La fonction de répartition de la variable aléatoire c est $\mathbf{1}_{[c, +\infty[}$ qui est continue partout sauf au point c . Désignant par F_n la fonction de répartition de Y_n , on note que pour tout $\varepsilon > 0$,

$$\begin{aligned} P(|Y_n - c| \geq \varepsilon) &= P(Y_n \leq c - \varepsilon) + P(Y_n \geq c + \varepsilon) \\ &\leq P(Y_n \leq c - \varepsilon) + P\left(Y_n > c + \frac{\varepsilon}{2}\right) \\ &= F_n(c - \varepsilon) + 1 - F_n\left(c + \frac{\varepsilon}{2}\right). \end{aligned}$$

La convergence de $P(|Y_n - c| \geq \varepsilon)$ vers 0 résulte de cette majoration puisque $F_n(x)$ converge vers $\mathbf{1}_{[c, +\infty[}(x)$ pour tout $x \neq c$.

Le lemme suivant est d'une grande utilité dans les problèmes de convergence en loi rencontrés en statistique mathématique. Pour en voir l'intérêt, rappelons que la convergence en loi de chacune des composantes d'un vecteur aléatoire n'implique pas la convergence en loi du vecteur.

Lemme 6.11 (Slutsky). *Soient $(Y_n)_{n \geq 1}$ et $(Z_n)_{n \geq 1}$ deux suites de variables aléatoires définies sur le même espace probabilisé et telles que*

$$Y_n \xrightarrow[n \rightarrow +\infty]{\text{loi}} Y \quad \text{et} \quad Z_n \xrightarrow[n \rightarrow +\infty]{\text{Pr}} c,$$

où c est une constante. Alors,

$$(Y_n, Z_n) \xrightarrow[n \rightarrow +\infty]{\text{loi}} (Y, c). \quad (6.4)$$



Pour une preuve, PVIR p. 392.

En pratique, le lemme de Slutsky est utilisé le plus souvent *via* le corollaire suivant.

Corollaire 6.12 (Slutsky). *Soient $(Y_n)_{n \geq 1}$ et $(Z_n)_{n \geq 1}$ deux suites de v.a. définies sur le même espace probabilisé et telles que Y_n converge en loi vers Y et Z_n converge en loi vers une constante c . Alors*

- a) $Y_n + Z_n$ converge en loi vers $Y + c$;
- b) $Y_n Z_n$ converge en loi vers cY .

6.2 Convergence en loi et fonctions caractéristiques

La transformation de Fourier est un outil important, aussi bien en analyse qu'en théorie des probabilités. Après avoir introduit les propriétés fondamentales de la transformée de Fourier des mesures finies et donc des lois de probabilité, nous verrons comment la transformée de Fourier permet l'étude de la *convergence en loi*.

Dans toute cette section, \mathbb{R}^d est muni du produit scalaire et de la norme euclidienne standards notés :

$$\langle t, x \rangle = \sum_{j=1}^d t_j x_j, \quad \|x\| = (x_1^2 + \cdots + x_d^2)^{1/2},$$

où $t = (t_1, \dots, t_d)$ et $x = (x_1, \dots, x_d)$ désignent deux vecteurs quelconques de \mathbb{R}^d .



Pour toute cette section, IFP pp. 255–266 et pp. 284–286.

Définition 6.13. Soit μ une mesure finie sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$. On appelle transformée de Fourier de μ la fonction $\hat{\mu} : \mathbb{R}^d \rightarrow \mathbb{C}$, définie par :

$$\forall t \in \mathbb{R}^d, \quad \hat{\mu}(t) := \int_{\mathbb{R}^d} \exp(i\langle t, x \rangle) d\mu(x). \quad (6.5)$$

Lorsque μ est la loi d'un vecteur aléatoire, $\hat{\mu}$ est la fonction caractéristique de X (en fait de P_X).

$$\text{Si } \mu = P_X, \quad \hat{\mu}(t) = \varphi_X(t) = \mathbf{E} \exp(i\langle t, X \rangle). \quad (6.6)$$

La fonction caractéristique d'une variable aléatoire réelle X (en fait de sa loi P_X avec le même abus de langage que pour les fonctions de répartition) s'écrit plus simplement :

$$\varphi_X(t) = \mathbf{E} (e^{itX}).$$

On notera que l'hypothèse μ finie (donc $\mu(\mathbb{R}^d) < +\infty$) rend automatiquement μ -intégrable sur \mathbb{R}^d la fonction $x \mapsto \exp(i\langle t, x \rangle)$, de module constant 1. Il est clair que cette fonction n'est λ_d -intégrable sur \mathbb{R}^d pour aucun t et donc (6.5) ne permet pas de définir $\hat{\lambda}_d$.

Les deux tableaux suivants donnent les fonctions caractéristiques de quelques lois usuelles, discrètes ou à densité. Les calculs de vérification sont laissés en exercice.

Exemple 6.14 (Fonctions caractéristiques de lois discrètes usuelles, $d = 1$).

Loi de X	Paramètres	$P(X = k)$	$\varphi_X(t)$
δ_a	$a \in \mathbb{R}$	$P(X = a) = 1$	e^{ita}
Bern(p)	$p \in]0, 1[$	$P(X = 0) = 1 - p, P(X = 1) = p$	$pe^{it} + 1 - p$
Bin(n, p)	$p \in]0, 1[, n \in \mathbb{N}^*$	$C_n^k p^k (1 - p)^{n-k}, \quad 0 \leq k \leq n$	$(pe^{it} + (1 - p))^n$
Geom(p)	$p \in]0, 1[$	$(1 - p)^{k-1} p, \quad k \in \mathbb{N}^*$	$\frac{pe^{it}}{1 - (1 - p)e^{it}}$
Pois(α)	$\alpha \in]0, +\infty[$	$\frac{e^{-\alpha} \alpha^k}{k!}, \quad k \in \mathbb{N}$	$\exp(\alpha(e^{it} - 1))$

Exemple 6.15 (Fonctions caractéristiques de lois usuelles à densité, $d = 1$).

Loi	Paramètres	Densité $f_X(x)$	$\varphi_X(t)$
Unif $[a, b]$	$a, b \in \mathbb{R}, a < b$	$\frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$	$\frac{e^{ibt} - e^{iat}}{(b-a)it}$
Exp(a)	$a \in]0, +\infty[$	$ae^{-ax} \mathbf{1}_{[0,+\infty[}(x)$	$\frac{1}{1 - \frac{it}{a}}$
Cau(a, b)	$a \in \mathbb{R}, b \in]0, +\infty[$	$\frac{1}{\pi b} \frac{1}{1 + \left(\frac{x-a}{b}\right)^2}$	$e^{iat - b t }$
Triangulaire		$(1 - x) \mathbf{1}_{[-1,1]}(x)$	$\frac{2(1 - \cos t)}{t^2}$
$\mathfrak{N}(m, \sigma)$	$m \in \mathbb{R}, \sigma \in]0, +\infty[$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$	$\exp\left(imt - \frac{\sigma^2 t^2}{2}\right)$

Proposition 6.16. *La transformée de Fourier $\hat{\mu}$ d'une mesure finie μ est une fonction continue et bornée sur \mathbb{R}^d . On a*

$$\forall t \in \mathbb{R}^d, \quad |\hat{\mu}(t)| \leq \mu(\mathbb{R}^d) = \hat{\mu}(0). \quad (6.7)$$

Proposition 6.17. *Soit X une variable aléatoire réelle ayant un moment d'ordre $r \in \mathbb{N}^*$ (i.e. $\mathbf{E}|X|^r < +\infty$) et ϕ sa fonction caractéristique. Alors ϕ est r fois dérivable sur \mathbb{R} et*

$$\forall t \in \mathbb{R}, \quad \phi^{(k)}(t) = i^k \mathbf{E} (X^k e^{itX}), \quad 1 \leq k \leq r.$$

En particulier

$$\phi^{(k)}(0) = i^k \mathbf{E} X^k.$$

Proposition 6.18. *Si la mesure μ est une mesure produit $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ de mesures finies sur \mathbb{R} ,*

$$\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d, \quad \hat{\mu}(t) = \prod_{j=1}^d \hat{\mu}_j(t_j), \quad (6.8)$$

autrement dit,

$$\hat{\mu} = \hat{\mu}_1 \otimes \cdots \otimes \hat{\mu}_d.$$

Dans le langage des fonctions caractéristiques, la proposition 6.18 se traduit comme suit.

Corollaire 6.19. *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire à composantes indépendantes, notons φ sa fonction caractéristique et φ_j celle de la variable aléatoire réelle X_j pour $j \in \llbracket 1, d \rrbracket$. Alors*

$$\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d, \quad \varphi(t) = \varphi_1(t_1) \cdots \varphi_d(t_d),$$

autrement dit

$$\varphi = \varphi_1 \otimes \cdots \otimes \varphi_d.$$

Une propriété essentielle de la transformation de Fourier est qu'elle permet de ramener le produit de convolution à un produit ordinaire de deux fonctions.

Théorème 6.20. Si μ et ν sont deux mesures finies sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$,

$$\forall t \in \mathbb{R}^d, \quad \widehat{\mu * \nu}(t) = \widehat{\mu}(t) \widehat{\nu}(t). \quad (6.9)$$

En particulier, si X et Y sont deux vecteurs aléatoires dans \mathbb{R}^d , indépendants, la fonction caractéristique φ_{X+Y} de leur somme est donnée par

$$\forall t \in \mathbb{R}^d, \quad \varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t). \quad (6.10)$$

Une autre propriété essentielle de la transformation de Fourier est la caractérisation des mesures finies.

Théorème 6.21. Si deux mesures finies sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$ ont même transformée de Fourier, elles sont égales. En particulier si deux vecteurs aléatoires X et Y de \mathbb{R}^d ont même fonction caractéristique, ils ont même loi.

Voici une application à la caractérisation de l'indépendance des variables aléatoires par les fonctions caractéristiques.

Proposition 6.22. Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire, notons φ sa fonction caractéristique et φ_j celle de X_j pour $j \in \llbracket 1, d \rrbracket$. Si

$$\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d, \quad \varphi(t) = \varphi_1(t_1) \dots \varphi_d(t_d), \quad (6.11)$$

alors les composantes de X sont mutuellement indépendantes.

Comme la proposition 6.22 est la réciproque de la proposition 6.18, l'égalité (6.11) est une condition *nécessaire et suffisante* pour l'indépendance mutuelle de X_1, \dots, X_d .

Une autre application des théorèmes 6.20 et 6.21 est de simplifier considérablement certains calculs de lois de sommes de v.a. indépendantes en évitant de recourir à la convolution, cf. IFP exemple 9.3 p. 266.

Voici enfin la propriété fondamentale qui explique la présence d'une section sur les fonctions caractéristiques dans ce chapitre.

Théorème 6.23. Soient X_n ($n \in \mathbb{N}^*$) et X des vecteurs aléatoires de \mathbb{R}^d de fonctions caractéristiques respectives φ_{X_n} et φ_X . Alors X_n converge en loi vers X si et seulement si

$$\forall t \in \mathbb{R}^d, \quad \varphi_{X_n}(t) \xrightarrow{n \rightarrow +\infty} \varphi_X(t). \quad (6.12)$$

6.3 Normalité asymptotique

Nous sommes maintenant en mesure d'aborder l'étude de la convergence en loi de sommes de variables aléatoires indépendantes convenablement normalisées vers une limite gaussienne.

6.3.1 Sommes de variables aléatoires i.i.d.

Théorème 6.24 (théorème limite central, cas i.i.d.). Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires définies sur le même espace probabilisé (Ω, \mathcal{F}, P) , indépendantes, de même loi et de carré intégrable et non p.s. constantes. Notons $\mu := \mathbf{E}X_1$, $\sigma^2 := \text{Var } X_1$ avec $\sigma > 0$ et $S_n = \sum_{k=1}^n X_k$. Alors

$$S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\text{Var } S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} Z, \quad (6.13)$$

où Z est une variable de loi gaussienne $\mathfrak{N}(0, 1)$.

Il est possible d'énoncer ce théorème de manière plus élémentaire, sans parler de convergence en loi, ni même de loi gaussienne. En exploitant la *continuité* sur \mathbb{R} de la f.d.r. Φ de la loi $\mathfrak{N}(0, 1)$ et la définition 6.1 de la convergence en loi, on voit en effet qu'une formulation équivalente de la conclusion (6.13) du théorème est que

$$\forall x \in \mathbb{R}, \quad P(S_n^* \leq x) \xrightarrow{n \rightarrow +\infty} \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt. \quad (6.14)$$

Une conséquence pratique de (6.14) est que pour tous réels a, b , tels que $a < b$,

$$P(S_n^* \in I(a, b)) \xrightarrow{n \rightarrow +\infty} \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{t^2}{2}\right) dt, \quad (6.15)$$

où $I(a, b)$ est n'importe lequel des quatre intervalles d'extrémités a et b . Noter que sous cette forme on pourrait énoncer une version du théorème limite central compréhensible par un public ne connaissant que la notion d'intégrale de Riemann ordinaire³.


Corollaire 6.25 (théorème de de Moivre-Laplace). *Si S_n est une variable aléatoire de loi binomiale de paramètres n et $p \in]0, 1[$, on a avec $q := 1 - p$,*

$$S_n^* := \frac{S_n - np}{\sqrt{npq}} = \sqrt{\frac{n}{pq}} \left(\frac{S_n}{n} - p \right) \xrightarrow[n \rightarrow +\infty]{\text{loi}} Z,$$


où Z est une variable de loi gaussienne $\mathfrak{N}(0, 1)$.

Preuve. C'est une conséquence immédiate du théorème 6.24 en remarquant que S_n a même loi⁴ que $X_1 + \dots + X_n$, où les X_k sont des variables aléatoires de Bernoulli indépendantes et de même paramètre p et en rappelant que l'espérance et la variance de la loi $\text{Bin}(n, p)$ sont respectivement np et npq . \square

La démonstration historique du théorème de de Moivre-Laplace repose sur un bon contrôle des coefficients binomiaux *via* la formule de Stirling. L'intérêt de cette approche « élémentaire » est de donner une idée de la vitesse de convergence qui est en $O(n^{-1/2})$.

 ICP chapitre 7.

Il existe plusieurs démonstrations du théorème 6.24. Aucune n'est simple. Celles qui peuvent sembler simples en première lecture ne le sont que parce que l'essentiel du travail est contenu dans une « marche d'approche ». La méthode « classique » repose sur la caractérisation de la convergence en loi par la convergence des fonctions caractéristiques (c'est la marche d'approche assez ardue). Une méthode plus « moderne » utilise un bon contrôle de moments fonctionnels de fonctions C^3 tendant vers 0 à l'infini.

 Pour une preuve par les fonctions caractéristiques, IFP pp. 284–288. Pour une preuve par les moments fonctionnels, PVIR pp. 393–399 démontre plus généralement les TLC de Liapounov et de Lindeberg (pour des v.a. X_i pas forcément de même loi).

Le théorème 6.24 a de multiples applications, notamment en statistique. À ce stade, on peut souligner deux idées.

D'abord, on peut noter que le comportement asymptotique en loi de S_n^* ne dépend pas de la loi de X_1 . La seule condition pour que la loi de S_n^* soit approximativement

3. On peut même laisser tomber la forme intégrale de la limite dans (6.15) en se contentant de dire que Φ est une fonction croissante continue que l'on a tabulée.

4. Il est clair d'après la définition de la convergence en loi que si Y_n converge en loi vers Y et si pour chaque n , Y'_n a même loi que Y_n , Y'_n converge aussi en loi vers Y .

gaussienne pour les grandes valeurs de n est que X_1 soit de carré intégrable. Ceci donne un caractère universel aux lois gaussiennes et explique la fréquence de l'utilisation de ces lois en modélisation⁵. On peut dire que le comportement asymptotique en loi de sommes S_n^* et donc aussi de S_n « oublie » tout de la loi des X_i , sauf le paramètre de localisation $\mu = \mathbf{E}X_1$ et le paramètre de dispersion $\sigma^2 = \text{Var } X_1$. C'est l'une des raisons de l'importance donnée à ces deux paramètres en théorie des probabilités.

La deuxième idée importante est que le théorème limite central donne une idée de la vitesse de convergence dans la loi des grands nombres. *Grosso modo*, on peut dire que dans le bon cas où $\mathbf{E}X_1^2 < +\infty$, cette vitesse est en $O(n^{-1/2})$. Précisons le sens de cette affirmation. Par (6.15) appliqué avec $[a, b] = [-t, t]$, on obtient :

$$\lim_{n \rightarrow +\infty} P(S_n^* \in [-t, t]) = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1, \quad (6.16)$$

en utilisant la relation $\Phi(-t) = 1 - \Phi(t)$ due à la parité de la densité de $\mathfrak{N}(0, 1)$. En remarquant maintenant que

$$S_n^* = \frac{S_n - n\mathbf{E}X_1}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - \mathbf{E}X_1 \right), \quad (6.17)$$

on peut réécrire (6.16) sous la forme

$$P\left(\left|\frac{S_n}{n} - \mathbf{E}X_1\right| \leq \frac{\sigma t}{\sqrt{n}}\right) = 2\Phi(t) - 1 + \varepsilon_n, \quad (6.18)$$

où ε_n est une suite de réels (pas forcément positifs), convergente vers 0. Pour tout $\delta > 0$, on peut choisir un $t = t(\delta)$ assez grand pour que $2\Phi(t) - 1 > 1 - \delta/2$ car $2\Phi(t) - 1$ tend vers 1 quand t tend vers $+\infty$. Ensuite pour $n \geq n_0(\delta)$, on aura $|\varepsilon_n| < \delta/2$ et finalement

$$\forall \delta > 0, \exists t(\delta), n(\delta), \forall n \geq n(\delta), \quad P\left(\left|\frac{S_n}{n} - \mathbf{E}X_1\right| \leq \frac{\sigma t(\delta)}{\sqrt{n}}\right) > 1 - \delta. \quad (6.19)$$

C'est au sens de (6.19) que l'on peut dire que S_n/n converge vers $\mathbf{E}X_1$ avec une vitesse en $O(n^{-1/2})$. On peut résumer (6.19) par l'écriture $|n^{-1}S_n - \mathbf{E}X_1| = O_P(n^{-1/2})$, dont le deuxième membre se lit « grand O en probabilité de $n^{-1/2}$ ».

Dans l'utilisation pratique du théorème 6.24, on travaille souvent avec n « grand » fixé et on approxime la loi de S_n^* par la loi $\mathfrak{N}(0, 1)$, ou ce qui revient au même, *on approxime la loi de S_n par la loi gaussienne $\mathfrak{N}(n\mathbf{E}X_1, \sigma\sqrt{n})$ de même espérance et même variance que S_n* . Plus précisément, en notant que $g_n : x \mapsto \frac{x - n\mathbf{E}X_1}{\sigma\sqrt{n}}$ est une bijection croissante de \mathbb{R} sur \mathbb{R} et en posant pour $a < b$ réels,

$$a_n = g_n(a) = \frac{a - n\mathbf{E}X_1}{\sigma\sqrt{n}}, \quad b_n = g_n(b) = \frac{b - n\mathbf{E}X_1}{\sigma\sqrt{n}},$$

on a

$$P(a \leq S_n \leq b) = P(a_n \leq S_n^* \leq b_n) = \Phi(b_n) - \Phi(a_n) + \varepsilon_n. \quad (6.20)$$

On néglige alors le terme d'erreur ε_n et on termine le calcul en utilisant la table des valeurs de Φ .

La question qui se pose dans le calcul précédent est « que signifie *n grand* ? », ou encore « comment peut-on contrôler l'erreur ε_n ? », autrement dit, quelle est la vitesse de convergence vers 0 de ε_n ? La réponse est que dans le « bon cas » où X_1 a un moment d'ordre 3, la vitesse de convergence dans le théorème limite central est en $O(n^{-1/2})$.

5. D'autant plus qu'il existe de nombreuses généralisations du théorème limite central, avec des v.a. indépendantes mais de lois différentes, avec des vecteurs aléatoires, avec des v.a. « faiblement dépendantes » ...

Théorème 6.26 (Berry-Esséen, 1941–42). *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. telle que $\mathbf{E}|X_i|^3 < +\infty$. On note $\sigma^2 := \text{Var } X_1$, $\rho^3 := \mathbf{E}|X_1 - \mathbf{E}X_1|^3$, avec $\sigma > 0$ et $\rho > 0$. Il existe alors une constante universelle $C > 0$ telle que pour tout $n \geq 1$,*

$$\Delta_n := \sup_{x \in \mathbb{R}} |P(S_n^* \leq x) - \Phi(x)| \leq C \frac{\rho^3}{\sigma^3} \frac{1}{\sqrt{n}}.$$

♣ PVIR pp. 372–373 pour voir que l’on ne peut espérer mieux que $\Delta_n = O(n^{-1/2})$. ICP pp. 172–173 pour les théorèmes d’Uspensky sur la vitesse de convergence dans le théorème de de Moivre-Laplace.

L’obtention de la meilleure constante C a été l’objet d’une longue quête. La valeur initiale de Esséen était $C = 7,59$. Une valeur plus moderne et proche de l’optimale est $C = 0,7975$ (Van Beek (1972)).

La version suivante du théorème limite central est directement motivée par les applications statistiques. Pour construire à l’aide du TLC un intervalle de confiance pour une espérance inconnue, on a besoin de bornes qui ne dépendent pas de la variance lorsque cette dernière est aussi inconnue. Une technique courante est alors de remplacer cette variance inconnue par un estimateur de celle-ci. La légitimation mathématique repose sur le théorème suivant.

Théorème 6.27 (TLC avec autonormalisation). *Soient X_1, \dots, X_n, \dots des variables aléatoires indépendantes et de même loi telle que $\mathbf{E}X_1^2 < +\infty$ et $\sigma^2 := \text{Var } X_1 > 0$. On note $S_n := X_1 + \dots + X_n$. On suppose de plus que $(V_n)_{n \geq 1}$ est une suite de variables aléatoires positives qui converge en probabilité vers σ^2 . Alors*

$$T_n := \sqrt{\frac{n}{V_n}} \left(\frac{S_n}{n} - \mathbf{E}X_1 \right) \xrightarrow[n \rightarrow +\infty]{\text{loi}} Z,$$

où Z suit la loi gaussienne standard $\mathfrak{N}(0, 1)$.

L’énoncé est volontairement légèrement incorrect pour respecter l’usage. La preuve combine le théorème 6.24 avec le lemme de Slutsky.

♣ PVIR pp. 399–400 pour un énoncé rectifié et une preuve détaillée.

En général on applique le TLC avec autonormalisation en prenant pour V_n la « variance empirique ». Cette variance empirique est pour chaque ω , la variance calculée sur la série statistique réellement observée $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$. C’est donc la variable aléatoire :

$$V_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2,$$

où $\bar{X} = S_n/n$. En appliquant deux fois la loi forte des grands nombres, on voit facilement que la variance empirique converge presque sûrement (donc aussi en probabilité) vers la variance théorique $\sigma^2 = \mathbf{E}X_1^2 - (\mathbf{E}X_1)^2$.

Dans le cas particulier important où les X_i sont des variables de Bernoulli, on peut appliquer directement le théorème ci-dessus avec $V_n = \bar{X}(1 - \bar{X})$ en notant que par la loi forte des grands nombres, V_n converge presque-sûrement vers $p(1 - p) = \sigma^2$. On peut aussi remarquer que pour des variables de Bernoulli, la variance empirique n’est autre que $\bar{X}(1 - \bar{X})$. La vérification de cette affirmation est facile une fois que l’on a noté que $X_i = X_i^2$ car X_i ne prend que les valeurs 0 et 1.