

Télécom SudParis

# Probabilités

SIC 3101

Emmanuel Monfrini  
Édition 2023 - Sandro Franceschi

PUBLISHED BY TELECOM SUDPARIS



*First printing, Septembre 2015*

Copyright © 2015 Emmanuel Monfrini  
Nouvelle édition 2023 Sandro Franceschi



# Table des matières

|   |           |
|---|-----------|
| <b>Introduction</b> .....                                   | <b>7</b>  |
| <b>1 Fondement de la théorie des probabilités</b> .....     | <b>11</b> |
| 1.1 Expériences et évènements aléatoires                    | 11        |
| 1.2 Équiprobabilité   | 15        |
| 1.3 Définition d'une probabilité dans le cas fini           | 17        |
| 1.4 L'idée géniale de Kolmogorov                            | 18        |
| 1.5 Théorie de la mesure et probabilités                    | 20        |
| 1.6 Définition générale d'une probabilité                   | 23        |
| 1.7 Mesures et densité de Radon-Nikodym                     | 25        |
| <b>2 Conditionnement et indépendance d'évènements</b> ..... | <b>27</b> |
| 2.1 Probabilité conditionnelle                              | 27        |
| 2.2 Formule des probabilités totales et formule de Bayes    | 28        |
| 2.3 Indépendance d'évènements                               | 30        |
| 2.4 Lemme de Borel-Cantelli                                 | 31        |
| <b>3 Variables aléatoires réelles</b> .....                 | <b>33</b> |
| 3.1 Loi d'une variable aléatoire                            | 33        |
| 3.2 Fonction de répartition et densité                      | 35        |
| 3.3 Changement de variable aléatoire $Y = \phi(X)$          | 37        |
| 3.4 Lois usuelles discrètes                                 | 38        |

|           |   |           |
|-----------|---|-----------|
| 3.5       | Lois usuelles continues                                       | 42        |
| <b>4</b>  | <b>Variables aléatoires vectorielles</b> .....                | <b>45</b> |
| 4.1       | Fonction de répartition conjointe et marginale                | 45        |
| 4.2       | Mesure produit et Fubini                                      | 47        |
| 4.3       | Changement de variable aléatoire multidimensionnel            | 49        |
| <b>5</b>  | <b>Indépendance de variables aléatoires</b> .....             | <b>51</b> |
| 5.1       | Indépendance de variables aléatoires                          | 51        |
| 5.2       | Somme de variables aléatoires indépendantes et convolution    | 53        |
| <b>6</b>  | <b>Espérance et moments</b> .....                             | <b>59</b> |
| 6.1       | La notion d'espérance mathématique                            | 59        |
| 6.2       | Moments d'une variable aléatoire                              | 61        |
| 6.3       | Cas vectoriel   | 64        |
| 6.4       | Fonction caractéristique                                      | 64        |
| 6.5       | Inégalités de Jensen, Bienaymé-Tchebichev et Markov           | 68        |
| 6.6       | Caractéristiques des variables aléatoires                     | 69        |
| <b>7</b>  | <b>Loi normale et vecteurs gaussiens</b> .....                | <b>73</b> |
| 7.1       | La loi normale  | 73        |
| 7.2       | Vecteurs gaussiens  | 77        |
| <b>8</b>  | <b>Espérance conditionnelle</b> .....                         | <b>83</b> |
| 8.1       | Conditionnement discret                                       | 83        |
| 8.2       | Espérance conditionnelle                                      | 85        |
| 8.3       | Loi conditionnelle  | 86        |
| 8.4       | Le cas gaussien   | 88        |
| <b>9</b>  | <b>Convergence de suites de variables aléatoires</b> .....    | <b>91</b> |
| 9.1       | Les différentes notions de convergence                        | 91        |
| 9.2       | Loi des grands nombres  | 95        |
| 9.3       | Théorème central-limite                                       | 96        |
| <b>10</b> | <b>Simuler des réalisations de variables aléatoires</b> ..... | <b>99</b> |
| 10.1      | Générateurs de nombres pseudo-aléatoires                      | 99        |
| 10.2      | Simulation de variables aléatoires unidimensionnelles         | 101       |
| 10.3      | Le cas des vecteurs gaussiens                                 | 102       |
| 10.4      | La méthode d'acceptation-rejet                                | 103       |
| 10.5      | Méthode de Monte-Carlo  | 107       |

|           |   |            |
|-----------|---|------------|
| <b>11</b> | <b>Exercices</b> .....  | <b>111</b> |
|           | Séance 1 - Évènements et probabilités                         | 112        |
|           | Séance 2 - Probabilités conditionnelles                       | 113        |
|           | Séance 3 - Lois de variables aléatoires                       | 115        |
|           | Séance 4 - Lois et calculs de probabilités                    | 118        |
|           | Séance 5 - Fonctions de répartition et changement de variable | 120        |
|           | Séance 6 - Espérance  | 122        |
|           | Séance 7 - Loi normale et vecteurs gaussiens                  | 124        |
|           | Séance 8 - Espérance conditionnelle                           | 126        |
|           | Séance 9 - Fonctions caractéristiques et convergence          | 128        |
|           | Séance 10 - Convergence de variables aléatoires               | 130        |
|           | Approfondissement - Simulation de variables aléatoires        | 132        |
|           | Soutien   | 135        |
|           | <b>Bibliographie</b> .....                                    | <b>137</b> |
|           | Livres  | 137        |
|           | Articles  | 137        |





## Introduction

*« Le hasard n'est que la mesure de notre ignorance. »*

Alfred Capus, *Les pensées*.

C'est en cherchant à en tirer avantage, en particulier dans les jeux de cartes ou de dés, que l'Homme a débuté sa quête de quantification de ce qu'on appelle le « hasard ». Cette volonté qui, suivant les époques, a tantôt été considérée comme hérétique tantôt comme prétentieuse, semble surtout manquer de réalisme.

*« Le hasard, c'est Dieu qui se promène incognito. »*

Albert Einstein.

C'est pourtant ce qui a progressivement conduit à faire passer la notion de probabilité d'une opinion, apanage d'une élite chez les Grecs anciens, à une idée philosophique à partir du Moyen-âge pour devenir un objet mathématique à la fin du *XVII<sup>e</sup>* siècle avant de converger, durant la première moitié du *XX<sup>e</sup>* siècle, vers la théorie mathématique que nous allons étudier ici.

La puissance de ce concept mathématique permet de modéliser toutes les situations où le hasard intervient, dépassant largement le cadre restreint des jeux, motivation initiale. Comme le précise S. Méléard dans son cours à l'Ecole Polytechnique, *Aléatoire*  
MÉLÉARD 2010

« La modélisation probabiliste est fondamentale dans tous les domaines d'application, qu'ils soient issus des **sciences dures** ou des **sciences humaines**, de la **physique** (physique quantique, physique des particules), de la **climatologie**, de la **biologie** (mutations du génome), de l'**écologie** (variabilité des comportements individuels ou variations environnementales), de l'**informatique** et des **réseaux** de télécommunications, du **traitement du signal et de la parole**, de la **médecine** (imagerie médicale), de l'**économie**, l'**assurance**, la **finance** (marchés boursiers), ou de la **sociologie**. »

Dans notre vie quotidienne, nous avons tendance à considérer que seuls les faits les moins agréables sont le fruit du hasard :

*« Le hasard fait parfois singulièrement les choses - exprès, dirait-on, pour nous empoisonner l'existence. »*

Jean-Yves Soucy, *Un été sans aube*.

Précisons donc que le hasard se doit d'être invoqué dès lors que nous nous trouvons confrontés à un fait que nous sommes incapables de prévoir **avec certitude**. Soumis continuellement à un certain nombre d'incertitudes et à une volonté naturelle de nous rassurer, nous nous sommes fait notre propre idée sur le concept de probabilité et sur son utilisation. Si le « bon sens » et l'intuition nous retiennent, le plus souvent, de commettre de grossières erreurs d'interprétations dans les situations courantes, il est important de garder à l'esprit que le passage d'un questionnement concret à la quantification des choix possibles, requiert bel et bien de passer par une, toujours délicate, **modélisation mathématique**.

Cette difficulté essentielle qui consiste à passer d'une situation de la vie quotidienne à un modèle mathématique suffisamment simple pour autoriser les calculs mais suffisamment complexe pour crédibiliser le choix de la modélisation, est au coeur du **calcul de probabilités**. En effet, dans le cadre du modèle, il est possible de faire de la prédiction, de la prise de décision ou même de l'expérimentation numérique grâce à des simulations.

L'incroyable diversité des phénomènes dans lesquels le hasard intervient a longtemps contribué à restreindre, l'étude des probabilités à un catalogue de méthodes à appliquer au cas par cas. Ce n'est qu'en 1933, qu'Andreï Kolmogorov a offert un cadre théorique suffisamment large pour que soit développée la **théorie des probabilités**. Il s'est, pour cela, inscrit dans la suite des travaux de **théorie de la mesure** d'Emile Borel (1897) et les travaux de **théorie de l'intégration** de Henri-Léon Lebesgue (1902), faisant entrer les probabilités dans le cercle très fermé des domaines fondamentaux des mathématiques.

L'étude des probabilités est donc, avant tout, une étude théorique dans le cadre d'un **modèle probabiliste**. Dans ce cadre, **il n'y a plus de place au hasard**; les objets

et notions mathématiques manipulés obéissent aux règles de l'analyse « classique ». La possibilité de réinvestir certains des théorèmes dans le cadre d'une situation réelle pourrait presque être considérée comme la récompense inattendue d'un travail bien fait.

La simplicité d'une situation réelle ou d'un énoncé concret donne souvent une idée trompeuse de l'effort de modélisation mathématique qui sera nécessaire à sa résolution. Accepter de procéder avec rigueur et méthode, comme on se doit de le faire en mathématiques, vous amènera à découvrir une manière moderne et captivante de faire des mathématiques, idéalement adaptée aux besoins de l'ingénieur. L'objectif de ce cours est de vous aider à ouvrir les yeux!

*« Le bon sens, quoi qu'il fasse, ne peut manquer de se laisser surprendre à l'occasion. Le but de la science est de lui épargner cette surprise [...]. »*

Bertrand Russell.





# 1. Fondement de la théorie des probabilités

*« The theory of probability as mathematical discipline can and should be developed from axioms in exactly the same way as Geometry and Algebra. »*

Andreï Kolmogorov, *Foundations of the Theory of Probability*.

## 1.1 Expériences et événements aléatoires

Les mathématiques « déterministes » nous amènent à distinguer le vrai du faux, à nous prononcer pour le « 1 » ou pour le « 0 » suivant une logique binaire. Pour pouvoir traiter des problèmes dont l'issue ne peut être prévue de façon certaine, les probabilités ouvrent l'intervalle de liberté existant entre le vrai (« 1 ») et le faux (« 0 »). En ce sens, les probabilités apparaissent comme une audacieuse généralisation des modes de raisonnement classiques. Mais ne nous y trompons pas... s'il n'y a rien d'incertain dans l'issue du problème étudié, il est parfaitement inutile de faire appel aux probabilités puisque des siècles de travaux mathématiques nous permettent d'avoir une grande confiance en ce que nous sommes en mesure de faire dans ce cas.

**R** Avant de vous lancer dans une modélisation probabiliste, assurez-vous d'être en présence d'une **expérience aléatoire**.

**Définition 1.1.1 — Expérience aléatoire.** Une expérience aléatoire est une expérience,  $E$ , renouvelable (en théorie si ce n'est en pratique), et qui, renouvelée dans des conditions identiques ne donne pas forcément le même résultat à chaque renouvellement.

Une fois validé l'aspect aléatoire du problème considéré, il faut délimiter les possibilités d'expression du hasard. Pour cela, il faut situer précisément la source de l'aléa et **déterminer toutes les issues possibles de l'expérience aléatoire.**

**Définition 1.1.2 — Espace d'états.** L'ensemble  $\Omega$  de tous les résultats possibles de l'expérience  $E$  est appelé espace d'états (associé à l'expérience  $E$ ). Un résultat possible de l'expérience est noté classiquement  $\omega$ . Ainsi,  $\omega \in \Omega$ .

Les exemples suivants donnent un aperçu de la palette de formes que peut prendre l'espace d'états d'une expérience aléatoire.

- **Exemple 1.1** 1. On lance deux pièces : **expérience aléatoire ✓**

$$\Omega = \{PP, PF, FP, FF\};$$

2. On lance un dé à six faces : **expérience aléatoire ✓**

$$\Omega = \{1, 2, 3, 4, 5, 6\};$$

3. On envoie une fléchette sur une cible circulaire de 20 cm de diamètre et on souhaite caractériser son point d'impact : **Et oui! C'est une expérience aléatoire ✓**

$$\Omega = \left\{ (x, y), \sqrt{x^2 + y^2} \leq 10 \right\};$$

4. On étudie la durée de vie d'une ampoule électrique : **expérience aléatoire ✓**

$$\Omega = \mathbb{R}^+ = [0; \infty[;$$

5. On s'intéresse au temps d'attente devant le bureau du prof de proba sachant que celui-ci pourra se libérer entre 13h et 15h : **expérience aléatoire ✓**

$$\Omega = [0; 2];$$

6. Sans plus de précision concernant la durée d'observation, on s'intéresse aux temps de passage des véhicules à un péage autoroutier : **expérience aléatoire ✓**

$$\Omega = (\mathbb{R}^+)^{\mathbb{N}};$$

7. On observe l'évolution du prix d'un actif financier sur un intervalle de temps  $[t_1, t_2]$ . Afin de pouvoir décrire toutes les situations possibles on considère l'ensemble des fonctions continues sur  $[t_1, t_2]$ , à valeurs réelles positives : **expérience aléatoire ✓**

$$\Omega = \mathcal{C}([t_1, t_2], \mathbb{R}^+),$$

■

Sans y prêter attention, nous avons déjà commencé à modéliser mathématiquement le problème qui nous intéresse. Nous avons éliminé tous les événements impossibles en rapport avec la situation d'intérêt.

**Définition 1.1.3 — Evènement aléatoire.** Un évènement aléatoire est une propriété de l'expérience aléatoire  $E$  dont on peut dire si elle est réalisée ou non une fois connu le résultat de ladite expérience. Un évènement aléatoire (associé à l'expérience  $E$ ) est donc une partie (un sous-ensemble) de  $\Omega$ .

**R** Les plus simples des évènements aléatoires sont les singletons  $\{\omega\} \subset \Omega$ . Ils sont appelés **évènements élémentaires** et ne sont pas les seuls évènements aléatoires.

Voici quelques évènements aléatoires prolongeant les espaces d'états de l'exemple précédent.

■ **Exemple 1.2** 1. On lance deux pièces :

$$A_1 = \{PF, FP\};$$

2. On lance un dé à six faces :

$$A_2 = \{\text{Obtenir au moins } 4\};$$

3. On envoie une fléchette sur une cible [...] :

$$A_3 = \text{« Toucher la cible »} = \Omega;$$

4. On étudie la durée de vie d'une ampoule électrique :

$$A_4 = \{\text{La durée de vie est comprise entre 2 et 3 ans}\},$$

que l'on réécrit mathématiquement :

$$A_4 = ]2;3[;$$

5. On s'intéresse au temps d'attente devant le bureau [...] :

$$A_5 = \text{« Il faut attendre moins d'une heure »},$$

ou encore

$$A_5 = [0;1[;$$

6. On observe l'évolution du prix d'un actif financier sur un intervalle de temps  $[t_1, t_2]$  [...] :

$$A_6 = \{\text{Le prix de l'actif est inférieur à } 100\text{€}\},$$

ou encore

$$A_6 = \left\{ \omega \in \mathcal{C}([t_1, t_2], \mathbb{R}^+), \sup_{t \in [t_1, t_2]} \|\omega(t)\| < 100 \right\},$$

■

## Ensembles ou évènements, une question de vocabulaire...

Considérons un ensemble  $\Omega$  et  $A$  et  $B$  deux parties ou sous-ensembles de  $\Omega$ . Du point de vu probabiliste,  $\Omega$  est l'espace d'états et  $A$  et  $B$  sont des évènements aléatoires.

On note  $A \subseteq \Omega$  et  $B \subseteq \Omega$ .

Le principe de la modélisation mathématique consiste à faire correspondre les opérations ensemblistes avec les opérations sur les évènements aléatoires.

1. L'**évènement contraire** de  $A$  est identifié à l'ensemble **complémentaire** de  $A$  (dans  $\Omega$ ). C'est l'ensemble formé par tous les éléments de  $\Omega$  n'appartenant pas à  $A$ .

l'évènement « NON  $A$  » est noté  $A^c$ .

2. L'**évènement certain** est identifié à l'ensemble  $\Omega$  tout entier puisque tous les résultats possibles de l'expérience composent  $\Omega$ .

$\Omega$  est l'évènement certain.

3. L'**évènement impossible** est identifié à l'ensemble vide  $\emptyset$  puisque tout résultat de l'expérience est nécessairement dans  $\Omega$ .

$\emptyset$  est l'évènement impossible.

4. Les évènements  $A$  et  $B$  sont simultanément réalisés si le résultat de l'expérience appartient aux deux ensembles. C'est le cas si et seulement si ce résultat appartient à l'**intersection** des ensembles  $A$  et  $B$ .

l'évènement «  $A$  ET  $B$  » est noté  $A \cap B$ .

5. Les évènements  $A$  et  $B$  sont **incompatibles** s'il ne peuvent pas être simultanément réalisés. C'est le cas si et seulement l'évènement «  $A$  et  $B$  » est l'évènement impossible.

l'évènement «  $A$  ET  $B$  sont incompatibles » est noté  $A \cap B = \emptyset$ .

6. Les évènements  $A$  **ou**  $B$  sont réalisés si le résultat de l'expérience appartient au moins à l'un des deux ensembles. C'est le cas si et seulement si ce résultat appartient à la **réunion** des ensembles  $A$  et  $B$ .

l'évènement «  $A$  OU  $B$  » est noté  $A \cup B$ .

7. La réalisation de l'évènement  $A$  implique celle de l'évènement  $B$  si le résultat de l'expérience (qui appartient à  $A$  par hypothèse) appartient également à  $B$ . C'est le cas si et seulement si l'ensemble  $A$  est inclus dans  $B$ .

l'évènement «  $A$  IMPLIQUE  $B$  » est noté  $A \subseteq B$ .

Notons  $\mathcal{A}$  l'ensemble de tous les évènements aléatoires. D'après ce que nous venons de dire, il est important que  $\mathcal{A}$  présente des propriétés de stabilité vis à vis des opérations sur les ensembles.

Donc pour  $A, B \in \mathcal{A}$  il serait souhaitable que

$$A^c \in \mathcal{A}, A \cap B \in \mathcal{A}, A \cup B \in \mathcal{A}, \Omega \in \mathcal{A} \text{ et } \emptyset \in \mathcal{A}. \quad (1.1)$$

**R** J'attire votre attention sur le fait que  $A$  et  $B$  sont inclus dans  $\Omega$  ( $A, B \subseteq \Omega$ ) et appartiennent à  $\mathcal{A}$  ( $A, B \in \mathcal{A}$ ). Ce sont des **sous-ensembles** de  $\Omega$  et des **éléments** de  $\mathcal{A}$ .

## 1.2 Équiprobabilité

La naissance du calcul des probabilités remonte aux travaux de Pascal et Fermat en 1654. Ils étaient à cette époque les deux grands mathématiciens européens et c'est tout naturellement vers eux que se tourna le Chevalier de Méré, joueur et « bel esprit », pour modéliser un jeu de hasard dont il était particulièrement friand (cf DELLACHERIE 1994).

Comme le disait, avec une pointe d'ironie, J. Bertoin dans son cours de probabilités à l'Université Paris VI (cf BERTOIN 2008) :

« On va voir comment on résolvait des problèmes simples à la préhistoire de la théorie. En fait, la plupart des notions importantes par la suite y figurent déjà. La modernité fera son apparition dans les chapitres ultérieurs. »

Considérons donc un espace d'états de dimension finie,  $n$ , ( $n$  entier fixé) :

$$\Omega = \{\omega_1, \dots, \omega_n\},$$

et, par analogie avec l'approche intuitive basée sur l'analyse fréquentielle, définissons la probabilité  $Pr$  d'un évènement  $A \subset \Omega$  par :

$$Pr(A) = \frac{Card(A)}{Card(\Omega)}.$$

Nous venons de définir une application de l'ensemble  $\mathcal{P}(\Omega)$  de tous les sous-ensembles de  $\Omega$  dans  $[0; 1]$  :

$$Pr : \mathcal{P}(\Omega) \longrightarrow [0; 1]$$

**R** Notons que pour un ensemble  $\Omega$  fini, on peut montrer que  $\mathcal{P}(\Omega)$  vérifie les propriétés de stabilité espérées en (1.1).

**Exercice 1.1** Démontrer le résultat de la remarque précédente. ■

**Définition 1.2.1 — Équiprobabilité.** Le cas que nous venons de décrire est le cas d'**équiprobabilité**. On montre que tous les évènements élémentaires  $\{\omega_i\} \in \mathcal{P}(\Omega)$  ont alors la même probabilité de se produire :

$$Pr(\{\omega_1\}) = Pr(\{\omega_2\}) = \dots = Pr(\{\omega_n\}) = \frac{1}{n}. \quad (1.2)$$

**Exercice 1.2** Démontrer (1.2). ■

**R** Dans le cas d'équiprobabilité, la probabilité d'un évènement  $A$  est « le nombre de cas favorables divisé par le nombre de cas possibles ». C'est dans ce cadre que les arguments d'analyse combinatoire sont très utiles.

**Propriété 1.2.1** On montre que :

$$Pr(\Omega) = 1. \quad (1.3)$$

Pour deux évènements  $A$  et  $B$  **incompatibles** :

$$A \cap B = \emptyset \Rightarrow Pr(A \cup B) = Pr(A) + Pr(B). \quad (1.4)$$

**Exercice 1.3** Démontrer les résultats de la propriété 1.2.1. ■

Des deux résultats précédents on peut également déduire les propriétés suivantes

**Propriété 1.2.2** On a également

$$Pr(\emptyset) = 0 \quad (1.5)$$

$$Pr(A^c) = 1 - Pr(A) \quad (1.6)$$

Pour deux évènements  $A$  et  $B$  **quelconques**,

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B) \quad (1.7)$$

Pour deux évènements  $A$  et  $B$ ,

$$A \subset B \Rightarrow Pr(A) < Pr(B) \quad (1.8)$$

**Exercice 1.4** Démontrer les résultats de la propriété 1.2.2 grâce aux résultats de la propriété 1.2.1. Il est conseillé de se référer à la FIGURE 1.1. ■

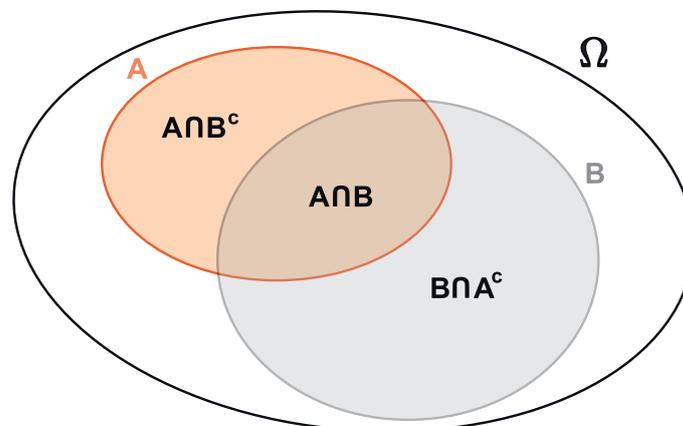


FIGURE 1.1 – Une idée de décomposition

### 1.3 Définition d'une probabilité dans le cas fini

Afin d'éviter tout malentendu par la suite, il est primordial de noter dès maintenant que

**Diviser le nombre de cas possibles par le nombre de cas favorables n'est pas la définition d'une probabilité.**

Il s'agit d'un cas très particulier dans lequel tous les événements élémentaires ont la même probabilité de se produire.

**Définition 1.3.1** Une **probabilité** sur un espace d'états fini  $\Omega = \{\omega_1, \dots, \omega_n\}$  est une application  $Pr : \mathcal{P}(\Omega) \rightarrow [0; 1]$  qui vérifie :

- (i). Pour tout événement  $A \in \mathcal{P}(\Omega)$ ,  $0 \leq Pr(A) \leq 1$  (par définition) ;
- (ii).  $Pr(\Omega) = 1$  ;
- (iii). Pour tous événements  $A$  et  $B$  incompatibles de  $\mathcal{P}(\Omega)$  :

$$A \cap B = \emptyset \Rightarrow Pr(A \cup B) = Pr(A) + Pr(B).$$

Les différentes étapes que nous venons d'enchaîner nous conduisent à la construction d'un **modèle probabiliste**.

**Définition 1.3.2** Un **modèle probabiliste** que l'on appelle également **triplet probabiliste**, est un triplet  $(\Omega, \mathcal{P}(\Omega), Pr)$  constitué de l'espace d'états  $\Omega$ , de l'ensemble de tous les événements possibles  $\mathcal{P}(\Omega)$  et de la probabilité  $Pr$  définie de  $\mathcal{C}$  définie ci-dessus.

**R** Dans le cas d'un espace d'états  $\Omega$  fini, on a  $\mathcal{C} = \mathcal{P}(\Omega)$ .

Les propriétés 1.2.1 et 1.2.2 restent valables dans le cadre de la définition 1.3.1 et on peut préciser (1.4) par la propriété suivante.

**Propriété 1.3.1** Pour toute famille  $(A_i)_{1 \leq i \leq n}$  d'événements **2 à 2 disjoints**, on a :

$$Pr\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n Pr(A_i). \quad (1.9)$$

**Exercice 1.5** Démontrer la propriété 1.3.1. ■

On peut envisager de définir les probabilités sur les ensembles finis avec une approche plus analytique. Les deux propositions suivantes justifient cette démarche.

**Proposition 1.3.2** Dans le modèle probabiliste  $(\Omega = \{\omega_1, \dots, \omega_n\}, \mathcal{P}(\Omega), Pr)$ , la probabilité  $Pr$  est **entièrement définie** par ses valeurs sur les singletons  $\{\omega_i\}$  c'est à dire par les réels

$$p_i = Pr(\{\omega_i\}), 1 \leq i \leq n$$

et pour tout  $A$  dans  $\mathcal{P}(\Omega)$  on a

$$Pr(A) = \sum_{\omega_i \in A} P(\omega_i) = \sum_{\omega_i \in A} p_i.$$

**Proposition 1.3.3** Considérons un espace d'états fini  $\Omega = \{\omega_1, \dots, \omega_n\}$ . La famille de réels  $\{p_i\}_{1 \leq i \leq n}$  définit une **unique probabilité**  $Pr$  sur  $(\Omega, \mathcal{P}(\Omega))$  telle que

$$Pr(\{\omega_i\}) = p_i$$

si et seulement si

$$\forall i \in \{1, \dots, n\}, 0 \leq p_i \leq 1 \text{ et } \sum_{i=1}^n p_i = 1.$$

**Exercice 1.6** Démontrer les résultats des propositions 1.3.2 et 1.3.3. ■

**R** Pour définir une probabilité  $Pr$  sur un espace d'états fini  $\Omega = \{\omega_1, \dots, \omega_n\}$ , « il suffit » de se donner un  $n$ -uplet de réels positifs  $(r_1, \dots, r_n)$ . On peut alors définir la **constante de normalisation**  $C = \sum_{i=1}^n r_i$  qui permet de définir le  $n$ -uplet  $(p_1, \dots, p_n)$  des  $p_i = r_i / C$  vérifiant les hypothèses de la proposition 1.3.3.

## 1.4 L'idée géniale de Kolmogorov

Bien qu'offrant un cadre propice à un calcul des probabilités « conforme à l'intuition », la définition 1.3.1 ne permet pas de modéliser tous les cas d'expériences aléatoires présentés dans l'exemple 1.1. En effet, mis à part les deux premiers cas, les espaces d'états considérés ne sont pas des ensembles finis. Ils ne sont même pas dénombrables!

Il est donc nécessaire d'étendre la définition 1.3.1 à des espaces d'états infinis non nécessairement dénombrables. Il faut également garder à l'esprit les « contraintes » de stabilité (1.1) qui s'étaient imposées d'elles-mêmes compte tenu des opérations sur les évènements aléatoires.

Pour mieux comprendre nos « nouveaux » besoins, intéressons-nous à une expérience aléatoire consistant à étudier la première apparition d'un "Pile" dans des lancers successifs d'une pièce de monnaie non truquée. Deux cas sont envisageables :

1. Le nombre  $n$  de lancers est fixé à l'avance;
2. On lance la pièce jusqu'à l'obtention du premier "Pile".

**Cas 1.** Si le nombre  $n$  de lancers est fixé à l'avance, une expérience se résume à une succession de  $n$  "Pile" ( $P$ ) ou "Face" ( $F$ ) c'est à dire, un mot de  $n$  lettres pris dans un alphabet à deux lettres  $P$  et  $F$ . Mathématiquement, un évènement aléatoire sera un  $n$ -uplet appartenant au produit cartésien :

$$\Omega = \{P, F\} \times \{P, F\} \dots \times \{P, F\} \text{ également noté } \Omega = \{P, F\}^n.$$

$\Omega$  est alors un ensemble fini à  $2^n$  évènements élémentaires dont on peut considérer qu'il sont équiprobables (de probabilité  $\frac{1}{2^n}$ ) puisque la pièce n'est pas truquée. On peut donc définir le modèle probabiliste  $(\Omega, \mathcal{P}(\Omega), Pr)$  pour lequel la probabilité  $Pr : \mathcal{P}(\Omega) \rightarrow [0; 1]$  est définie par :

$$\forall A \in \mathcal{P}(\Omega), Pr(A) = \frac{Card(A)}{2^n}. \quad (1.10)$$

**Cas 2.** Si le nombre  $n$  de lancers n'est pas fixé à l'avance, on ne sait pas combien de lancers vont être nécessaires pour décrire l'expérience aléatoire. Pour être certain de recenser tous les cas possibles il faut envisager tous les mots formés avec les lettres  $P$  et  $F$  et **de longueur infinie**.

L'espace d'états est donc :

$$\Omega = \{P, F\}^{\mathbb{N}^*}.$$

C'est un espace d'états infini **non dénombrable** et les développements précédents ne s'appliquent plus. Intéressons-nous alors simplement à l'évènement :

$A$  : « On ne tire jamais Pile ».

Cet évènement correspond à un unique élément de  $\Omega$ ,  $(F, F, F, \dots, F, \dots)$ , mais comme l'équiprobabilité n'a plus de sens ici, nous ne savons pas en calculer la probabilité. En revanche, nous savons calculer la probabilité de l'évènement :

$A_n$  : « On ne tire jamais Pile lors des  $n$  premiers lancers ».

En effet, puisqu'on se restreint à  $n$  lancers, on a, d'après (1.10) :

$$Pr(A_n) = 2^{-n}.$$

Du point de vue ensembliste, la suite  $(A_n)_{n \geq 1}$  est une suite décroissante puisque  $\forall n \geq 1, A_{n+1} \subset A_n$  et  $A$  est la limite de cette suite puisque :

$$A = \bigcap_{n \geq 1} A_n.$$

Pour des raisons de cohérence avec la modélisation que nous cherchons à étendre, on s'attend à ce que :

$$P(A) = P\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} Pr(A_n) = 0, \quad (1.11)$$

mais pour cela, les axiomes de la définition 1.3.1 et en particulier le troisième permettant une **additivité** (simple), ne sont pas suffisants...

**Exercice 1.7** Démontrer que  $\forall n \in \mathbb{N}, \{P, F\}^n$  est dénombrable.  
Démontrer que  $\{P, F\}^{\mathbb{N}^*}$  est non dénombrable. ■

C'est là qu'intervient l'idée géniale de Kolmogorov : définir le modèle probabiliste  $(\Omega, \mathcal{A}, P)$  comme étant la donnée de :

- un espace d'états  $\Omega$ ;
- une tribu  $\mathcal{A}$  sur cet espace;
- une mesure  $P : \mathcal{A} \rightarrow [0; 1]$ .

## 1.5 Théorie de la mesure et probabilités

Cette partie du cours, essentielle à la compréhension en profondeur de la notion de probabilité. Je vous recommande donc la lecture de trois ouvrages de référence sur le sujet que sont le cours LE GALL pas de date de Jean-François Le Gall à l'ENS, *Intégration, probabilités et processus aléatoire*, le livre BARBE et LEDOUX 2007 de Philippe Barbe et Michel Ledoux *Probabilité* (au singulier!) et l'incontournable ouvrage de Patrick Billingsley *Probability and Measure* BILLINGSLEY 1995.

Soit un ensemble  $\Omega$  et  $(P)(\Omega)$  l'ensemble des parties de  $\Omega$  qui peut être un ensemble **fini, dénombrable** ou **infini non dénombrable**.

La théorie de la mesure distingue alors deux type de sous-ensembles stables de  $\mathcal{P}(\Omega)$ , les algèbres (de Boole) et les tribus.

**Définition 1.5.1 — Algèbre de Boole.** Un sous ensemble  $\mathcal{C}$  de  $\mathcal{P}(\Omega)$  est une **algèbre** (de Boole) si :

- (i).  $\Omega \in \mathcal{C}$ ;
- (ii).  $\mathcal{C}$  est stable par passage au complémentaire :

$$\forall A \in \mathcal{P}(\Omega), A \in \mathcal{C} \Rightarrow A^c \in \mathcal{C};$$

- (iii).  $\mathcal{C}$  est stable par réunion finie :

$$\forall A_1, \dots, A_n \in \mathcal{C}, A_1 \cup \dots \cup A_n \in \mathcal{C}. \quad (1.12)$$

**R** Pour l'axiome (iii) de la définition 1.5.1 on peut se contenter du cas  $n = 2$  puisque le cas général s'en déduit par récurrence.

**Propriété 1.5.1** Par passage au complémentaire dans (1.12), une algèbre de Boole est également stable par intersection finie.

**R** L'algèbre de Boole remplit toutes les conditions de stabilité (1.1) permettant la définition d'un modèle probabiliste sur un espace d'états fini.

**Définition 1.5.2 — Tribu.** Un sous ensemble  $\mathcal{A}$  de  $\mathcal{P}(\Omega)$  est une **tribu** ou une  **$\sigma$ -algèbre** (de Boole) si :

- (i).  $\Omega \in \mathcal{A}$  ;
- (ii).  $\mathcal{A}$  est stable par passage au complémentaire :

$$\forall A \in \mathcal{P}(\Omega), A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A};$$

- (iii).  $\mathcal{C}$  est stable par réunion dénombrable :

$$\text{pour toute suite } (A_n)_{n \in \mathbb{N}} \text{ d'éléments de } \mathcal{A}, \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}. \quad (1.13)$$

**Propriété 1.5.2** Par passage au complémentaire, une tribu est également stable par intersection dénombrable.

**Exercice 1.8** Démontrer les propriétés 1.5.1 et 1.5.2. ■

**R**

- Toute tribu est une algèbre, la réciproque est fausse.
- $\mathcal{A} = \{\emptyset, \Omega\}$  est la plus petite tribu sur  $\Omega$ . C'est la **tribu grossière**.
- $\mathcal{A} = \mathcal{P}(\Omega)$  est la plus grande tribu sur  $\Omega$ . **C'est celle que nous choisirons systématiquement lorsque  $\Omega$  sera fini ou dénombrable.**

**Définition 1.5.3 — Tribu engendrée.** Soit  $C$  une partie de  $\mathcal{P}(\Omega)$ . On appelle **tribu engendrée par  $C$**  la plus petite tribu contenant  $C$ .

**Proposition 1.5.3**

- La tribu engendrée par  $C \subset \mathcal{P}(\Omega)$  existe puisque  $\mathcal{P}(\Omega)$  est une tribu contenant  $C$ .
- Comme une intersection dénombrable de tribus est encore une tribu, la tribu engendrée par  $C \subset \mathcal{P}(\Omega)$  est l'intersection de toutes les tribus contenant  $C$ .

**Exercice 1.9**

1. Montrer qu'une intersection dénombrable de tribus est encore un tribu.
  2. Montrer qu'une réunion dénombrable de tribus est encore une tribu.
  3. Montrer que la tribu engendrée par  $C \subset \mathcal{P}(\Omega)$  est l'intersection de toutes les tribus contenant  $C$ .
-

**Définition 1.5.4 — Tribu borélienne.** Si  $\Omega = \mathbb{R}$ , on appelle **tribu borélienne** la tribu engendrée par les intervalles ouverts de  $\mathbb{R}$ . Elle est notée  $\mathcal{B}(\mathbb{R})$ .

**Proposition 1.5.4**

- La tribu borélienne  $\mathcal{B}(\mathbb{R})$  contient tous les intervalles ouverts, fermés ou semi-ouverts de  $\mathbb{R}$ . Il en est de même de toute réunion, finie ou dénombrable, d'intervalles (ouverts, fermés ou semi-ouverts) de  $\mathbb{R}$
- La tribu borélienne  $\mathcal{B}(\mathbb{R})$  peut être engendrée par l'une quelconques des 4 classes suivantes d'intervalles de  $\mathbb{R}$  :
  - (a)  $\mathcal{K}_1 = ]-\infty; x]$  pour  $x \in \mathbb{R}$ ;
  - (b)  $\mathcal{K}_1 = ]-\infty; a]$  pour  $a \in \mathbb{Q}$ ;
  - (c)  $\mathcal{K}_1 = ]-\infty; x[$  pour  $x \in \mathbb{R}$ ;
  - (d)  $\mathcal{K}_1 = ]-\infty; a[$  pour  $a \in \mathbb{Q}$ ;

**Exercice 1.10** Démontrer la proposition 1.5.4. ■

Le fait de pouvoir associer à notre espace d'états un ensemble de parties « suffisamment » stable, nous permet de définir un espace mesurable.

**Définition 1.5.5 — Espace mesurable - ensemble mesurable.**

- Le couple  $(\Omega, \mathcal{A})$  formé par un ensemble  $\Omega$  et une tribu  $\mathcal{A}$  sur  $\Omega$  est appelé **espace mesurable**.
- Un élément  $A$  de  $\mathcal{A}$  est appelé **ensemble mesurable**.

On peut ensuite, introduire la notion de **mesure** de laquelle découlera celle de **(mesure de) probabilité**.

**Définition 1.5.6 — Mesure.** Une mesure sur l'espace mesurable  $(\Omega, \mathcal{A})$  est une application  $\mu: \mathcal{A} \rightarrow [0; +\infty[$  qui vérifie les propriétés suivantes

- (i).  $\mu(\emptyset) = 0$ ;
- (ii). Pour toute famille  $(A_n)_{n \in \mathbb{N}}$  d'ensembles mesurables deux à deux disjoints,

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n) \quad (1.14)$$

Le triplet  $(\Omega, \mathcal{A}, \mu)$  est appelé **espace mesuré**.

- R** Parmi les propriétés des mesures, il y en a une qui est particulièrement intéressante lorsqu'on se souvient du questionnement sur lequel nous étions restés en (1.11). En effet, pour toute famille d'ensembles mesurables  $(B_n)_{n \in \mathbb{N}}$ , décroissante (au sens de l'inclusion), on a bien :

$$\mu\left(\bigcap_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow \infty} \downarrow \mu(B_n).$$

Merci Kolmogorov!

## 1.6 Définition générale d'une probabilité

L'idée de Kolmogorov est donc de présenter la probabilité comme une mesure dont les propriétés correspondent à celles que nous attendons mais la probabilité n'est pas une mesure quelconque. En particulier, le fait que la probabilité d'un évènement quelconque ne puisse dépasser 1 nous oblige à préciser la mesure à laquelle nous nous intéressons.

**Définition 1.6.1 — Mesures particulières.** Considérons un espace mesurable  $(\Omega, \mathcal{A})$  et une mesure  $\mu$  sur cet espace.

- (i).  $\mu$  est dite  $\sigma$ -**finie** s'il existe un recouvrement  $\bigcup_{n \in \mathbb{N}} A_n$  de  $\Omega$  par une suite  $(A_n)_{n \in \mathbb{N}}$  d'éléments de  $\mathcal{A}$  de mesure finie ( $\forall n \in \mathbb{N}, \mu(A_n) < +\infty$ );
- (ii).  $\mu$  est dite **finie** si  $\mu(\Omega) < +\infty$ ;
- (iii).  $\mu$  est une **probabilité** si  $\mu(\Omega) = 1$ .

Nous sommes désormais en « mesure »<sup>1</sup> de modéliser les cas plus exotiques d'expériences aléatoires, grâce à la notion de probabilité que l'on définit simplement par

**Définition 1.6.2 — Probabilité.** Une probabilité sur l'espace mesurable  $(\Omega, \mathcal{A})$  est une application  $P: \mathcal{A} \rightarrow [0; 1]$  qui vérifie les deux axiomes suivants :

- (i).  $P(\Omega) = 1$ ;
- (ii).  $\sigma$ -additivité : Pour toute suite dénombrable  $(A_n)_{n \in \mathbb{N}}$  d'évènements de  $\mathcal{A}$  **deux à deux disjoints**,

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n) \quad (1.15)$$

**R** Une probabilité est une mesure finie majorée par 1.

La définition suivante est fondamentale en Probabilités. Elle introduit une notion de “vrai” et de “faux” qui abandonne le caractère absolu qu'on lui connaît habituellement pour dépendre de la probabilité choisie sur l'espace d'états.

**Définition 1.6.3** Soit  $(\Omega, \mathcal{A}, P)$  un modèle probabiliste.

- Un évènement de probabilité nulle est dit **négligeable** ( pas impossible) ou **faux P-presque sûrement**;
- Un évènement de probabilité 1 est dit **P-presque certain** ( pas certain) ou **vrai P-presque sûrement**.

**R** En abrégé,  $P$ -presque sûrement sera noté  $P$ -p.s. ou même p.s. si le contexte ne laisse pas de place au doute sur la probabilité  $P$  de référence.

1. Ndlr : humour de probabiliste...

Le résultat suivant, que nous avons déjà évoqué en référence à (1.11) est un élément important pour le calcul des probabilités

**Proposition 1.6.1** Soit  $P$ , une probabilité définie sur  $(\Omega, \mathcal{A})$ . On a alors les deux propriétés **équivalentes** suivantes :

(i). Pour toute suite dénombrable d'évènements  $(A_n)_{n \in \mathbb{N}}$ , croissante (au sens de l'inclusion), on a :

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \uparrow P(A_n). \quad (1.16)$$

(ii). Pour toute suite dénombrable d'évènements  $(B_n)_{n \in \mathbb{N}}$ , décroissante (au sens de l'inclusion), on a :

$$P\left(\bigcap_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow \infty} \downarrow P(B_n). \quad (1.17)$$

**Exercice 1.11** Démontrer la proposition 1.6.1. Il est conseillé de montrer (i) puis (i)  $\Leftrightarrow$  (ii). ■

Un exemple d'application de cette proposition est la démonstration de la propriété suivante.

**Propriété 1.6.2** Soit  $P$ , une probabilité définie sur  $(\Omega, \mathcal{A})$ . Soit  $I$  un sous ensemble fini ou dénombrable de  $\mathbb{N}$ . Considérons alors la famille  $(A_n)_{n \in I}$ , finie ou dénombrable, d'évènements de  $\mathcal{A}$ . On a alors :

$$P\left(\bigcup_{n \in I} A_n\right) \leq \sum_{n \in I} P(A_n). \quad (1.18)$$

**Exercice 1.12** Démontrer la propriété 1.6.2. Il est conseillé de commencer par le cas où  $I$  est un ensemble fini et d'en déduire le cas où  $I$  est dénombrable. ■

## Probabilité sur un espace d'états dénombrable

**R** Grâce aux propriétés des séries à termes positifs, ce cas a été étudié en classes prépa. En effet, grâce à la proposition 1.6.3 qui étend les propositions 1.3.2 et 1.3.3, définir une probabilité sur un espace d'états dénombrable est une chose simple.

Considérons donc un espace d'états dénombrable  $\Omega = \{\omega_0, \omega_1, \dots, \omega_n, \dots\}$  pour tout  $n \in \mathbb{N}$ . Comme nous l'avons précisé plus haut, la tribu naturellement associée à un espace d'états fini ou dénombrable est  $\mathcal{P}(\Omega)$ .

**Proposition 1.6.3**

- Une probabilité  $P$  sur un espace dénombrable  $(\Omega, \mathcal{P}(\Omega))$  est **entièrement caractérisée par sa valeur sur les singletons**  $\{\omega_n\}$ .

- Etant donnée une suite  $(p_n)_{n \in \mathbb{N}}$  de réels tels que :

$$0 \leq p_n \leq 1, \text{ et } \sum_{n \in \mathbb{N}} p_n = 1,$$

il **existe** une **unique** probabilité  $P$  telle que pour tout  $A \in \mathcal{P}(\Omega)$ ,

$$P(A) = \sum_{\omega_n \in A} P(\{\omega_n\}) = \sum_{\omega_i \in A} p_n. \quad (1.19)$$

**Exercice 1.13** Démontrer la proposition 1.6.3. Cette démonstration requiert une certaine maîtrise du calcul des séries. Je vous invite donc à effectuer les révisions nécessaires à cette occasion! ■

## 1.7 Mesures et densité de Radon-Nikodym

### La mesure de comptage

**Définition 1.7.1 — Mesure de comptage.** Soit  $(E, \mathcal{E})$  un espace mesurable pour lequel  $E$  est un sous ensemble fini ou dénombrable de  $\mathbb{R}^d$ . La mesure  $\mu$  définie par :

$$\forall A \in \mathcal{E}, \mu(A) = \sum_{x \in E} \delta_x(A) \quad (1.20)$$

est appelée **mesure de comptage** sur  $E$ .

**Propriété 1.7.1** La mesure de comptage sur un ensemble fini ou dénombrable est  $\sigma$ -finie.

**R** En particulier, la mesure de comptage sur  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  est  $\sigma$ -finie.

**Exercice 1.14** Démontrer la propriété 1.7.1. ■

### La mesure de Lebesgue

**Définition 1.7.2 — Mesure de Lebesgue..** Considérons l'espace mesurable  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . La **mesure de Lebesgue**,  $\lambda$ , est définie par :

$$\forall A \in \mathcal{B}(\mathbb{R}), \lambda(A) = \inf \left\{ \sum_{i \in \mathbb{N}} (b_i - a_i) \mid A \subset \bigcup_{i \in \mathbb{N}} ]a_i; b_i[ \right\}. \quad (1.21)$$

L'infimum porte sur tous les recouvrements dénombrables de  $A$  par des intervalles ouverts  $]a_i; b_i[$  avec  $a_i < b_i$ .

**R** En particulier, pour tout intervalle  $I$  de la forme  $]a; b[, ]a; b], [a; b[$  ou  $[a; b]$  avec

$a \leq b$  la mesure de Lebesgue de  $I$  est la longueur de l'intervalle :

$$\lambda(I) = b - a.$$

**Propriété 1.7.2** La mesure de Lebesgue est  $\sigma$ -finie.

**Exercice 1.15** Démontrer la propriété 1.7.2. ■

### Le Théorème de Radon-Nikodym

**Définition 1.7.3** Soient  $\mu$  et  $\nu$  deux mesures définies sur un espace mesurable  $(E, \mathcal{E})$ . On dit que  $\mu$  est **absolument continue** par rapport à  $\nu$  si :

$$\forall A \in \mathcal{E}, \nu(A) = 0 \Rightarrow \mu(A) = 0. \quad (1.22)$$

On note  $\mu \ll \nu$ .  $\nu$  est la *mesure dominante* et  $\mu$  est la *mesure dominée*.

**Théorème 1.7.3 — Corollaire du théorème de Radon-Nikodym.** Considérons  $\mu$  et  $\nu$ , deux mesures positives  $\sigma$ -finie sur un espace mesurable  $(E, \mathcal{E})$ . Alors  $\mu$  est absolument continue par rapport à  $\nu$  **si et seulement si** il existe une fonction mesurable positive,  $f$ , telle que :

$$\forall A \in \mathcal{E}, \mu(A) = \int_A f d\nu \left( = \int_E f \mathbb{1}_A d\nu \right). \quad (1.23)$$

On note :  $f = \frac{d\mu}{d\nu}$ .

**R** Le corollaire du théorème de Radon-Nikodym ci-dessus est une **condition nécessaire et suffisante pour qu'une mesure  $\mu$  admette une densité par rapport à la mesure  $\nu$ .**

En probabilités, ce théorème est très important puisqu'il affirme qu'une mesure de probabilité  $P$  (qui est une mesure positive finie... donc  $\sigma$ -finie) admet une densité si et seulement si on peut exhiber une mesure positive  $\sigma$ -finie qui domine  $P$ .

Or toute mesure de probabilité *discrète* sur  $(E, \mathcal{E})$  est dominée par la mesure de comptage sur  $E$  et s'écrit sous la forme  $P(A) = \sum_{x \in E} p_x \delta_x(A)$ .

D'une manière générale, une mesure de probabilité  $P$  sur l'espace probablisable  $(E, \mathcal{E})$  peut s'écrire sous forme intégrale grâce à sa **densité**  $f$  :

$$\forall A \in \mathcal{E}, P(A) = \int_A f d\nu,$$

où  $\nu$  est une mesure  $\sigma$ -finie dominant  $P$ .

C'est le prolongement naturel de l'idée géniale de Kolmogorov : donner un cadre calculatoire à la théorie de probabilités grâce à la théorie de l'intégration de Lebesgue!



## 2. Conditionnement et indépendance d'évènements

« *Toute connaissance dégénère en probabilité.* »

David Hume, *Traité de la nature humaine*.

Le concept d'indépendance est sans doute la première notion importante où la théorie des probabilités se différencie nettement de l'intégration. Afin de l'introduire nous devons tout d'abord parler de conditionnement. Ce chapitre se concentre sur le conditionnement et l'indépendance *d'évènements*. La notion d'indépendance sera approfondie dans le Chapitre 5 qui définit l'indépendance entre des variables aléatoires ou entre des tribus. La notion de conditionnement sera elle approfondie dans le Chapitre 8 qui introduit l'espérance conditionnelle.

### 2.1 Probabilité conditionnelle

La notion de probabilité conditionnelle permet de prendre en compte, en cours d'expérience, une information supplémentaire dont on ne disposait pas au moment de la modélisation. En particulier lorsqu'un évènement  $B$  se réalise, on peut utiliser cette information pour affiner la probabilité d'un évènement  $A$ .

**Définition 2.1.1** Considérons le modèle probabiliste  $(\Omega, \mathcal{A}, P)$ . Soient alors deux évènements  $A$  et  $B$  de  $\mathcal{A}$ .

La **probabilité conditionnelle** de l'évènement  $A$  sachant que l'évènement  $B$  est réalisé est notée :

$$P_B(A) = P(A|B),$$

et définie par

$$P_B(A) = \begin{cases} \frac{P(A \cap B)}{P(B)} & \text{si } P(B) \neq 0 \\ P(A) & \text{sinon} \end{cases}. \quad (2.1)$$

**R** Lorsqu'on sait que l'évènement  $B$  est réalisé, on peut dire avec certitude que le résultat de l'expérience aléatoire est dans  $B$ . Tout se passe donc comme si l'espace n'était plus  $\Omega$  mais  $B$ . Quel que soit l'évènement  $A$  considéré, il est alors naturel de ne s'intéresser qu'à la partie de cet évènement compatible avec  $B$ , c'est à dire  $A \cap B$ . L'actualisation de la probabilité de  $A$  sera donc proportionnelle à  $P(A \cap B)$ . Dans (2.1),  $P(B)$  joue alors le rôle de la constante de normalisation déjà évoquée dans la remarque suivant la proposition 1.3.3.

**Proposition 2.1.1** Soit  $B$  un évènement de  $(\Omega, \mathcal{A}, P)$  de probabilité strictement positive. Alors  $P_B$  est une probabilité sur  $(\Omega, \mathcal{A})$  et  $(\Omega, \mathcal{A}, P_B)$  est donc un modèle probabiliste.

**Exercice 2.1** Démontrer la proposition 2.1.1. ■

En utilisant le conditionnement de façon séquentielle, il est possible de caractériser complètement la probabilité d'une intersection finie d'évènements quelconques.

**Propriété 2.1.2 — Formule des probabilités composées.** Considérons une famille finie  $(A_k)_{k=1}^n$ , d'évènements de  $\mathcal{A}$  tels que  $P\left(\bigcap_{k=1}^n A_k\right) > 0$ . Alors :

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}). \quad (2.2)$$

**R** La Formule des probabilités composées s'avère particulièrement utile lorsque l'on s'intéresse à un enchaînement d'expériences aléatoires dont les issues tiennent compte de ce qui s'est déjà passé lors des étapes précédentes.

**Exercice 2.2** Démontrer par récurrence la propriété 2.1.2. ■

## 2.2 Formule des probabilités totales et formule de Bayes

Soit  $(\Omega, \mathcal{A}, P)$  un modèle probabiliste. Considérons alors un évènement  $A$  de  $\mathcal{A}$  dont on cherche à connaître la probabilité  $P(A)$ .

Le calcul « direct » de cette probabilité peut être très complexe et une décomposition de  $A$  sur une famille finie ou dénombrable d'évènements plus « simples » peut s'avérer nécessaire.

**Propriété 2.2.1 — Formule des probabilités totales.** Considérons une famille, finie ou dénombrable  $(B_n)_{n \in I}$ ,  $I \subseteq \mathbb{N}$ , d'évènements de  $\mathcal{A}$  formant une partition de  $\Omega$ .

Alors, pour tout évènement  $A$  de  $\mathcal{A}$  on a :

$$P(A) = \sum_{n \in I} P(A \cap B_n). \quad (2.3)$$

**R** Rappelons que la famille  $(B_n)_{n \in I}$ ,  $I \subseteq \mathbb{N}$ , est une partition de  $\Omega$  si :

- Les  $B_n$  sont deux à deux disjoints;
- $\bigcup_{n \in I} B_n = \Omega$ , (la notation  $\bigcup$  rappelle que la réunion est disjointe).

**Exercice 2.3** Démontrer la propriété 2.2.1 qui est une conséquence immédiate de la définition 1.6.2. ■

**Propriété 2.2.2 — Formule de Poincaré.** Considérons une famille finie  $(A_m)_{m=1}^n$ , d'évènements quelconques de  $\mathcal{A}$ . Alors :

$$P\left(\bigcup_{m=1}^n A_m\right) = \sum_{k=1}^n (-1)^{k-1} p_k = p_1 - p_2 + \dots + (-1)^{n-1} p_n. \quad (2.4)$$

où, pour tout  $k \in \{1, \dots, n\}$  les

$$p_k = \sum_{\substack{i_1, \dots, i_k \in \{1, \dots, n\} \\ 1 \leq i_1 < \dots < i_k \leq n}} P(A_{i_1} \cap \dots \cap A_{i_k})$$

sont les sommes, sur toutes les combinaisons possibles, des probabilités des intersections de  $k$  éléments de  $(A_m)_{m=1}^n$ .

**Exercice 2.4** Démontrer la propriété 2.2.2. D'un point de vue mathématique, cette démonstration ne requiert que la maîtrise technique de la démonstration par récurrence. Il est cependant fortement recommandé de se munir de feuilles A3 et d'une quantité suffisante de tisane apaisante. ■

Grâce aux probabilités conditionnelles, il est également possible de décomposer un évènement quelconque sur une suite finie ou dénombrable d'évènements disjoints de probabilités connues.

**Propriété 2.2.3 — Enoncé conditionnel de la Formule des probabilités totales.** Considérons une famille, finie ou dénombrable  $(B_n)_{n \in I}$ ,  $I \subseteq \mathbb{N}$ , d'évènements de  $(\Omega, \mathcal{A}, P)$  formant une partition de  $\Omega$ . Alors, pour tout évènement  $A$  de  $\mathcal{A}$  on a :

$$P(A) = \sum_{n \in I} P(A|B_n) P(B_n). \quad (2.5)$$

**Exercice 2.5** Démontrer la propriété 2.2.3 qui est une conséquence immédiate de la propriété 2.2.1. ■

Le résultat suivant, issu des travaux du révérend Bayes au XVIII<sup>e</sup> siècle, est un résultat très utile en probabilités et davantage encore en statistique puisqu'il a donné naissance

à une méthodologie connue sous le nom de *statistique bayésienne* que vous étudierez au second semestre.

**Propriété 2.2.4 — Formule de Bayes simple.** Considérons deux évènements  $A$  et  $B$  de  $(\Omega, \mathcal{A}, P)$ ,  $B$  étant de probabilité strictement positive. Alors,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (2.6)$$

**R** Roughly speaking, comme disent les anglophones, la formule de Bayes permet de passer du conditionnement « dans un sens » au conditionnement « dans l'autre sens ». C'est « calculatoirement » exact.

Une interprétation plus « statistique » de cette formule est de retenir qu'elle permet d'améliorer la compréhension,  $P(A)$ , que nous avons de l'évènement  $A$ , pris isolément, en intégrant l'information issue de la réalisation préalable de l'évènement  $B$  sous la forme  $P(A|B)$ . Ainsi, la *probabilité a posteriori*,  $P(A|B)$ , sera liée à la *probabilité a priori*,  $P(A)$ , grâce à  $P(B|A)$ .

**Exercice 2.6** Démontrer la propriété 2.2.4. ■

Le théorème suivant généralise la formule de Bayes simple.

**Théorème 2.2.5 — Formule de Bayes.** Considérons une famille, finie ou dénombrable  $(B_n)_{n \in I}$ ,  $I \subseteq \mathbb{N}$ , d'évènements de  $(\Omega, \mathcal{A}, P)$  formant une partition de  $\Omega$ . Alors, pour tout évènement  $A$  de  $\mathcal{A}$  on a :

$$\forall i \in I, P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{n \in I} P(A|B_n)P(B_n)}. \quad (2.7)$$

**Exercice 2.7** Démontrer le théorème 2.2.5. ■

## 2.3 Indépendance d'évènements

La notion d'indépendance est absolument fondamentale en probabilités et on verra par la suite toutes ses implications dans la modélisation de l'aléatoire. Intuitivement, deux évènements  $A$  et  $B$  sont indépendants si le fait de savoir que  $A$  est réalisé ne donne aucune information sur la réalisation de  $B$  et réciproquement. En d'autres termes, si  $A$  et  $B$  sont indépendants, on a :

$$P(A|B) = P(A) \text{ et } P(B|A) = P(B). \quad (2.8)$$

En fait, ces deux égalités sont équivalentes et nous amènent à la même définition de l'indépendance de deux évènements :

**Définition 2.3.1** Deux évènements aléatoires  $A$  et  $B$  de probabilités strictement positives sont indépendants si :

$$P(A \cap B) = P(A)P(B). \quad (2.9)$$

**Exercice 2.8** Montrer l'équivalence entre les deux égalités de (2.8) puis montrer l'équivalence de ces égalités et de (2.9). ■

**Proposition 2.3.1** Considérons deux événements aléatoires  $A$  et  $B$  indépendants. Alors

- $A$  et  $B^c$  sont indépendants;
- $A^c$  et  $B$  sont indépendants;
- $A^c$  et  $B^c$  sont indépendants.

**Exercice 2.9** Démontrer la proposition 2.3.1. ■

La définition 2.3.1 permet de donner une existence mathématique à l'intuition que l'on peut avoir de la notion d'indépendance. Dès qu'il y a plus de deux événements aléatoires en jeu, il est moins évident de faire appel à l'intuition. La définition suivante éclaire la question.

**Définition 2.3.2** Considérons une famille, finie ou dénombrable  $(A_n)_{n \in I}$ ,  $I \subseteq \mathbb{N}$ , d'événements de  $(\Omega, \mathcal{A}, P)$ . Ces événements sont indépendants si, pour tout entier positif  $k \leq \text{Card}(I)$ , pour toute suite  $i_1, \dots, i_k$  d'éléments deux à deux distincts de  $I$  on a :

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}). \quad (2.10)$$

**R** D'après la définition 3.1.1, il ne suffit pas que les événements soient indépendants deux à deux pour avoir l'indépendance globale. L'exercice suivant fait intervenir trois événements globalement « liés » qui sont pourtant indépendants 2 à 2.

**Exercice 2.10** On joue 2 fois à Pile ou Face et on considère les événements

- $A = \{ \text{Face au premier lancé} \};$
- $B = \{ \text{Face au deuxième lancé} \};$
- $C = \{ \text{les deux tirages donnent le même résultat} \}.$

Montrer que ces événements sont deux à deux indépendants mais que :

$$P(A \cap B \cap C) \neq P(A)P(B)P(C).$$

## 2.4 Lemme de Borel-Cantelli

Soit  $(A_n)_{n \in \mathbb{N}}$  une suite d'événements, on pose :

$$\limsup_{n \in \mathbb{N}} A_n = \bigcap_{n \geq 0} \left( \bigcup_{k \geq n} A_k \right)$$

et

$$\liminf_{n \in \mathbb{N}} A_n = \bigcup_{n \geq 0} \left( \bigcap_{k \geq n} A_k \right).$$

**Théorème 2.4.1 — Borel-Cantelli.** Soit  $(A_n)_{n \in \mathbb{N}}$  une suite d'évènements :

1. Si  $\sum_{n \in \mathbb{N}} P(A_n) < +\infty$  alors

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

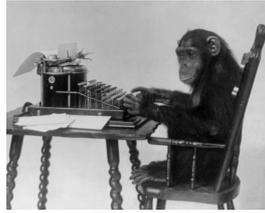
C'est à dire que presque sûrement  $\{n \in \mathbb{N} : \omega \in A_n\}$  est fini.

2. Si  $\sum_{n \in \mathbb{N}} P(A_n) = \infty$  ET si les évènements sont indépendants alors

$$P(\limsup_{n \in \mathbb{N}} A_n) = 1.$$

C'est à dire que presque sûrement  $\{n \in \mathbb{N} : \omega \in A_n\}$  est infini.

■ **Exemple 2.1 Le singe dactylographe.** Supposons qu'un singe, placé devant une machine à écrire, tape au hasard sur toutes touches. Supposons que ce singe appuie sur chaque touche avec une fréquence égale. Alors, tôt ou tard, il tapera n'importe quel texte choisi à l'avance et il le tapera même une infinité de fois! Si le texte fait  $N$  caractères il suffit de prendre  $A_n = \{\text{Le singe écrit le texte entre sa } nN+1 \text{ et sa } (n+1)N \text{ touche du clavier}\}$ , ces évènements sont indépendants et sont tous de probabilité identique non nulle. Même si cette probabilité est très faible on a donc bien  $\sum_{n \in \mathbb{N}} P(A_n) = \infty$  et le lemme de Borel-Cantelli permet de conclure. ■



*Démonstration de Borel-Cantelli.* 1. Si  $\sum_{n \in \mathbb{N}} P(A_n) < +\infty$  alors  $\mathbb{E} \left[ \sum_{n \in \mathbb{N}} \mathbf{1}_{A_n} \right] = \sum_{n \in \mathbb{N}} P(A_n) < +\infty$  et donc  $\sum_{n \in \mathbb{N}} \mathbf{1}_{A_n}$  est fini p.s.

2. Soit  $n_0 < n$ , alors

$$P \left( \bigcap_{k=n_0}^n A_k^c \right) = \prod_{k=n_0}^n P(A_k^c) = \prod_{k=n_0}^n (1 - P(A_k)).$$

La divergence de la série  $\sum_{n \in \mathbb{N}} P(A_n) = \infty$  implique que  $\prod_{k=n_0}^{\infty} (1 - P(A_k)) = 0$  (cf résultat de classe prépa) et donc que

$$P \left( \bigcap_{k=n_0}^{\infty} A_k^c \right) = 0.$$

On en déduit donc que

$$P \left( \bigcup_{n_0=0}^{\infty} \bigcap_{k=n_0}^{\infty} A_k^c \right) = 0,$$

et on conclut en passant au complémentaire que

$$P \left( \limsup_{n \in \mathbb{N}} A_n \right) = P \left( \bigcap_{n \geq 0} \left( \bigcup_{k \geq n} A_k \right) \right) = 1.$$

■



## 3. Variables aléatoires réelles

*« Le nom seul de calcul des probabilités est un paradoxe : la probabilité, opposée à la certitude, c'est ce qu'on ne sait pas, et comment peut-on calculer ce que l'on ne connaît pas. »*

Henri Poincaré, *La science et l'hypothèse*.

Bien que nous progressions dans la construction d'une théorie des probabilités, ce que nous avons vu jusqu'ici pourrait donner raison à Poincaré. En effet, le travail sur les événements, et donc sur les ensembles, semble se faire « au cas par cas » et n'apparaît pas comme la mise en œuvre d'une théorie mathématique globale. La notion de variable aléatoire permet un pas supplémentaire dans la bonne direction!

### 3.1 Loi d'une variable aléatoire

Vous avez déjà rencontré la notion de variable aléatoire durant votre scolarité. En effet, dans un premier temps, on a appelé variable aléatoire ou *aléa numérique* une variable  $X$  pouvant prendre différentes valeurs, chacune avec une certaine probabilité. Dans ce cas nous restons très proche de ce que nous savons faire dans  $(\mathbb{N}, \mathcal{P}(\mathbb{N}), P)$  ou  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ .

Satisfaisante lorsqu'on manie une seule variable aléatoire, cette définition est totalement insuffisante si l'on traite simultanément plusieurs variables aléatoires. Là encore, c'est la théorie de la mesure qui nous permet de généraliser et théoriser la notion de variable aléatoire.

**Définition 3.1.1 — Variables aléatoires.** On appelle **variable aléatoire**, en abrégé **v.a.**, toute application mesurable définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et à valeurs dans l'espace mesurable  $(E, \mathcal{E})$ .

**R** Une variable aléatoire n'est donc pas une variable au sens habituel du terme. Comme c'est une fonction (mesurable)... elle n'est pas aléatoire non plus!!! Une variable aléatoire est un **objet mathématique déterministe**.

Nous nous intéresserons en particulier aux **variables aléatoires réelles**, c'est à dire les variables aléatoires définies de  $(\Omega, \mathcal{A}, P)$  dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Il est d'usage de noter les variables aléatoires en lettres majuscules  $X, Y, Z$  etc... Une réalisation d'une variable aléatoire est habituellement notée en lettre minuscule  $X(\omega) = x$  pour un  $\omega \in \Omega$ .

Nous nous intéresserons également aux **vecteurs aléatoires réels**, c'est à dire les variables aléatoires définies de  $(\Omega, \mathcal{A}, P)$  dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . Il est d'usage de noter les vecteurs aléatoires en lettres majuscules grasses et droites avec une indication de « début » et de « fin »  $\mathbf{X}_1^d = (X_1, \dots, X_d)$ . Une réalisation est notée  $\mathbf{X}_1^d(\omega) = \mathbf{x}_1^d$ .

Enfin, nous serons amenés à nous intéresser à des **suites de variables aléatoires réelles**, c'est à dire aux variables aléatoires définies de  $(\Omega, \mathcal{A}, P)$  dans  $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$  que l'on notera en lettres majuscules grasses et droites  $\mathbf{X} = (X_1, \dots, X_k, \dots)$ . Une réalisation est notée  $\mathbf{X}(\omega) = \mathbf{x}$ .

La théorie de la mesure, nous apprend qu'il est possible d'associer une mesure « image » à toute application mesurable d'un espace mesuré dans un espace mesurable.

**Définition 3.1.2 — et proposition.** Soit une fonction mesurable  $\phi$  définie sur  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $(E, \mathcal{E})$ . L'application  $P_\phi$  de  $\mathcal{E}$  dans  $\mathbb{R}^+$  définie par :

$$\forall B \in \mathcal{E}, P_\phi(B) = P(\phi^{-1}(B)), \quad (3.1)$$

est une mesure. Cette mesure est appelée **mesure image** de  $P$  par  $\phi$ .

Dans le cas où  $\phi$  est une variable aléatoire, nous pouvons donc « transporter », au moyen de cette variable aléatoire, la mesure de probabilité de l'espace probabilisé initial vers le nouvel espace.

**Définition 3.1.3 — Loi d'une v.a..** Soit une variable aléatoire  $X$  définie sur  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $(E, \mathcal{E})$ . La loi de la variable aléatoire  $X$  est la mesure-image de  $P$  par  $X$ . C'est donc la mesure de probabilité sur  $(E, \mathcal{E})$ , notée  $P_X$ , définie par :

$$\forall A \in \mathcal{E}, P_X(A) = P(X^{-1}(A)), \quad (3.2)$$

En pratique, le lien avec les règles de calcul énoncées dans le chapitre 2 apparaît plus clairement si l'on écrit :

$$P_X(A) = P(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

La loi de  $X$  est souvent notée  $\mathcal{L}(X)$ .

**Exercice 3.1** Montrer que (3.2) définit bien une loi de probabilité. ■

**R** On peut considérer de manière complémentaire les variables aléatoires et leur lois. Dans le langage des probabilités, on aura souvent tendance à confondre les deux notions.



il ne suffit pas que deux variables aléatoires aient la même loi pour être égales!

Une variable aléatoire scalaire  $X$ , resp. multidimensionnelle  $\mathbf{Y}_1^d$ , définit un nouvel espace probabilisé  $(\mathcal{B}(\mathbb{R}), \mathbb{R}, P_X)$ , resp.  $(\mathcal{B}(\mathbb{R}^d), \mathbb{R}^d, P_{\mathbf{Y}_1^d})$ , propre à la variable aléatoire considérée. Pour contraster avec ce nouvel espace, l'espace probabilisé de référence,  $(\Omega, \mathcal{A}, P)$ , est alors appelé *espace probabilisé fondamental*.

## 3.2 Fonction de répartition et densité

**Définition 3.2.1 — fonction de répartition.** Soit  $X$  une v.a. réelle. On appelle **fonction de répartition** (f.d.r.) de  $X$  la fonction  $F_X$  définie par :

$$\begin{aligned} F_X &: \mathbb{R} \rightarrow [0; 1] \\ x &\mapsto F_X(x) = P_X([-\infty, x]) = P(X \leq x) \end{aligned}$$

Le théorème suivant est une conséquence de la définition 3.2.1.

**Théorème 3.2.1** Soit  $X$  une variable aléatoire réelle. Alors

- $F_X$  est croissante;
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$ ,  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ ;
- $F_X$  est càdlàg (continue à droite et admet une limite à gauche) et cette limite est  $\lim_{x \rightarrow x_0, x < x_0} F_X(x) = P(X < x_0)$ .

**Propriété 3.2.2** La fonction de répartition  $F_X$  caractérise la loi  $P_X$ . En particulier :

$$\text{pour tous réels } a \leq b, \quad P_X([a; b]) = F_X(b) - F_X(a).$$

**R**  $F_X$  est continue en  $x_0$  si, et seulement si,  $P(X = x_0) = 0$ .

### Variables aléatoires continues réelles

**Définition 3.2.2** Une variable aléatoire  $X$  est dite **continue** ou **à densité** (par rapport à la mesure de Lebesgue) si il existe une fonction  $f_X$  définie de  $\mathbb{R}$  dans  $\mathbb{R}$  vérifiant :

- $f_X$  est positive et intégrable sur  $\mathbb{R}$ ,
- $f_X$  est continue sauf peut-être en un nombre fini de points où elle admet une limite à gauche et une limite à droite finies ou non.
- $\int_{-\infty}^{+\infty} f_X(x) dx = 1$ ,

et telle que pour tous réels  $a \leq b$  (avec éventuellement  $a$  et  $b$  infinis) :

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx.$$

On dit que  $f$  est une densité de  $X$ .

**Réciproquement** pour toute fonction  $f$  vérifiant les propriétés ci dessus, il existe une variable aléatoire  $X$  définie sur un espace de probabilité convenable et admettant  $f$  comme densité.

**R** Soit  $X$  une variable aléatoire continue réelle à densité  $f_X$ . Pour tout réel  $x \in \mathbb{R}$  :

$$P(X = x) = 0.$$

On a de plus

$$\forall x \in \mathbb{R}, F_X(x) = \int_{]-\infty; x]} f_X(x) dx.$$

Plus généralement, si  $\mu$  domine la mesure de probabilité  $P_X$ , la **densité**  $f_X$  par rapport à  $\mu$  est définie comme la dérivée de Radon-Nikodym :

$$\forall x \in \mathbb{R}, F_X(x) = \int_{]-\infty; x]} f_X(t) d\mu(t).$$

**Propriété 3.2.3** La fonction de répartition est la primitive de la densité par rapport à la mesure dominante et réciproquement, la densité est la dérivée de la fonction de répartition par rapport à cette mesure dominante.

**R** Compte tenu de cette dernière propriété, la connaissance de la fonction de répartition est équivalente à la connaissance de la loi de  $P_X$ .

### Variables aléatoires discrètes réelles

Soit une variable aléatoire  $X$  définie sur  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $(E, \mathcal{E})$ . Supposons alors que l'ensemble  $E$  est un sous-ensemble fini ou dénombrable de  $\mathbb{R}$ . Comme nous l'avons signalé précédemment, dans ce cas, la tribu  $\mathcal{E}$  la plus pertinente est alors  $\mathcal{P}(E)$ , l'ensemble des sous-ensembles de  $E$ .

On a alors :

$$\forall A \in \mathcal{E}, P_X(A) = P(X \in A) = P\left(\bigcup_{x \in A} \{X = x\}\right) = \sum_{x \in A} P(\{X = x\}) = \sum_{x \in A} P_X(\{x\}).$$

**Définition 3.2.3** Une variable aléatoire  $X$  définie sur  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $(E, \mathcal{E})$  où  $E$  est un ensemble fini ou dénombrable est une **variable aléatoire discrète** et sa loi est de la forme :

$$P_X = \sum_{x \in E} p_x \delta_x,$$

où  $\delta_x$  est la mesure de Dirac en  $x$  et  $p_x = P_X(\{x\})$ .

**R** En pratique, définir la loi d'une variable aléatoire discrète c'est donner toutes les probabilités  $P_X(\{x\})$  pour  $x \in E$ .

Nous l'avons signalé plus haut, dans le cas d'une variable aléatoire discrète, la mesure de comptage  $\mu$  est une mesure dominante. Comment s'écrit la fonction de répartition dans ce cas ?

**Propriété 3.2.4** Considérons donc une variable aléatoire discrète,  $X$ , définie sur l'espace probabilisé  $(E, \mathcal{P}(E), P_X)$ , où  $E$  est un sous-ensemble fini ou dénombrable de  $\mathbb{R}$ .

On a :

$$P_X = \sum_{x \in E} p_x \delta_x \text{ et } \mu = \sum_{x \in E} \delta_x$$

et si  $f_X = \frac{dP_X}{d\mu}$ , alors :

$$\forall t \in \mathbb{R}, F_X(t) = \int_{]-\infty; t]} f_X(x) d\mu(x) = \sum_{x \in E} \mathbb{1}_{[x; +\infty[}(t) p_x. \quad (3.3)$$

**Exercice 3.2** Démontrer la propriété 3.2.4. ■

### 3.3 Changement de variable aléatoire $Y = \phi(X)$

Soit  $X$  une variable aléatoire continue qui admet une densité  $f$  et une fonction de répartition  $F$ . Soit  $\phi$  une fonction dérivable. On cherche à déterminer  $g$  et  $G$ , la densité et la fonction de répartition de  $Y = \phi(X)$ .

**Si  $\phi$  est bijective.**

**Propriété 3.3.1** Si  $\phi$  est bijective on a

$$g(y) = \frac{f(\phi^{-1}(y))}{|\phi'(\phi^{-1}(y))|}. \quad (3.4)$$

*Démonstration.*  $\phi$  est dérivable et bijective donc monotone. Si  $\phi$  est croissante, on a :

$$Y \leq y \Leftrightarrow \phi(X) \leq y \Leftrightarrow X \leq \phi^{-1}(y)$$

D'où

$$G(y) = P(Y \leq y) = P(\phi(X) \leq y) = P(X \leq \phi^{-1}(y)) = F(\phi^{-1}(y)),$$

c'est à dire  $G(\phi(x)) = F(x)$ . En dérivant, on obtient :

$$g(\phi(x)) \cdot \phi'(x) = f(x)$$

soit

$$g(y) = \frac{f(x)}{\phi'(x)} = \frac{f(\phi^{-1}(y))}{\phi'(\phi^{-1}(y))}.$$

Si  $\phi$  est décroissante, on a de même  $Y \leq y \Leftrightarrow \phi(X) \leq y \Leftrightarrow X \geq \phi^{-1}(y)$  et donc  $G(y) = 1 - F(\phi^{-1}(y))$ . Ainsi

$$g(y) = -\frac{f(x)}{\phi'(x)} = -\frac{f(\phi^{-1}(y))}{\phi'(\phi^{-1}(y))}.$$

Puisque  $\phi$  est décroissante,  $\phi' < 0$ , et on obtient la formule générale pour une application bijective  $\phi$  quelconque. ■

### Cas général.

Le principe consiste toujours à identifier la fonction de répartition  $G$ , en recherchant l'antécédent pour  $X$  de l'événement «  $Y \leq y$  ».

■ **Exemple 3.1** Considérons le cas  $Y = X^2$

$$Y \leq y \Leftrightarrow |X| \leq y$$

On obtient

$$G(y) = \begin{cases} 0 & \text{si } y \leq 0, \\ F(\sqrt{y}) - F(-\sqrt{y}) & \text{si } y > 0. \end{cases}$$

et enfin

$$g(y) = \begin{cases} 0 & \text{si } y \leq 0, \\ \frac{1}{2\sqrt{y}}(f(\sqrt{y}) + f(-\sqrt{y})) & \text{si } y > 0. \end{cases}$$

Si de plus, on suppose que  $X$  suit une loi normale centrée réduite (i.e.  $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ ) alors pour  $y > 0$ ,  $g(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$ . Cette loi est très importante en statistique, c'est la loi du  $\chi^2$ . ■

## 3.4 Lois usuelles discrètes

### 3.4.1 Variable de Bernoulli ou variable indicatrice.

**Définition 3.4.1 — Variable de Bernoulli.** Une variable aléatoire discrète,  $X$ , qui ne prend que les valeurs 1 et 0 avec les probabilités respectives  $p$  et  $1 - p$  est appelée *variable de Bernoulli*. Sa loi est notée  $\mathcal{B}(p)$ .

On a :

$$\forall \omega \in \Omega, X(\omega) = 0 \times \mathbb{1}_{\{0\}}(\omega) + 1 \times \mathbb{1}_{\{1\}}(\omega) = \mathbb{1}_{\{1\}}(\omega),$$

ainsi que :

$$P(X = 0) = 1 - p \text{ et } P(X = 1) = p. \quad (3.5)$$

■ **Exemple 3.2** Une urne contient deux boules rouges et trois boules vertes. On tire une boule de l'urne. La variable aléatoire  $X =$  nombre de boules rouges tirées est une variable de Bernoulli. On a :  $p(X = 1) = 2/5 = p$ ,  $p(X = 0) = 3/5 = 1 - p$ . ■

**R** On utilisera une variable de Bernoulli dès qu'on effectuera une épreuve qui n'a que deux issues : le succès ou l'échec. On affectera alors la valeur 1 à la variable en cas de succès et 0 en cas d'échec :

$$\forall \omega \in \Omega, X(\omega) = \mathbb{1}_{\{\text{Succès}\}}(\omega).$$

- R** La variable de Bernoulli est la plus simple des variables aléatoires puisque s'il y a moins de deux issues possibles, nous n'avons plus à faire à un phénomène aléatoire.

### 3.4.2 Variable discrète uniforme.

**Définition 3.4.2 — Variable discrète uniforme.** Une variable aléatoire discrète,  $X$ , qui prend ses valeurs dans le sous-ensemble  $\{a_1, \dots, a_n\}$  de  $\mathbb{R}$  avec des probabilités égales est appelée *variable discrète uniforme* sur  $\{a_1, \dots, a_n\}$ .

On a :

$$\forall x \in \{a_1, \dots, a_n\}, P(X = x) = \frac{1}{n}. \quad (3.6)$$

### 3.4.3 Distribution binomiale.

- On effectue une épreuve à deux issues : le succès est obtenu avec une probabilité  $p$  et l'échec avec une probabilité  $1 - p$ .
- On répète  $n$  fois cette épreuve.
- Les  $n$  épreuves sont *indépendantes* entre elles, ce qui signifie que la probabilité de réalisation de l'événement « succès » est toujours égale à  $p$ .
- On s'intéresse à la variable :

$S =$  nombre de succès au cours des  $n$  épreuves .

Notons  $X_1, \dots, X_n$ ,  $n$  variables de Bernoulli indépendantes de paramètre  $p$  représentant les  $n$  expériences précédentes.

Calculatoirement on a donc  $S = X_1 + \dots + X_n$ . L'ensemble des valeurs prises par  $S$  est  $0, 1, \dots, n$ .

**Définition 3.4.3 — Variable binomiale.** Une variable aléatoire discrète,  $S$ , est une *variable binomiale* de paramètres  $n$  et  $p$  si :

$$\forall k \in \{0, \dots, n\}, P(S = k) = \binom{n}{k} p^k (1 - p)^{n-k}. \quad (3.7)$$

Sa loi est notée  $\mathcal{B}(n, p)$ .

**Exercice 3.3** Montrer que la somme de  $n$  variables aléatoires de Bernoulli indépendantes et de même loi suit une loi binomiale  $\mathcal{B}(n, p)$ . ■

**Exercice 3.4** Soient  $X_1$  et  $X_2$ , deux variables aléatoires indépendantes de lois respectives  $\mathcal{B}(n_1, p_1)$  et  $\mathcal{B}(n_2, p_2)$ .

Quelle est la loi de la variable aléatoire  $X_1 + X_2$ ? ■

### 3.4.4 Distribution géométrique.

- On effectue une épreuve à deux issues : le succès est obtenu avec une probabilité  $p$  et l'échec avec une probabilité  $1 - p$ .

- On répète cette épreuve jusqu'à obtenir le premier succès.
- Toutes les épreuves sont indépendantes entre elles.
- On s'intéresse à la variable :

$X$  : nombre de fois qu'il faut répéter l'épreuve pour obtenir le premier succès .

- R** On se place donc sous les mêmes hypothèses que pour la loi binomiale, mais le nombre d'épreuves n'est pas fixé à l'avance. On s'arrête au premier succès.  $X$  peut donc prendre toutes les valeurs entières positives soit  $\mathbb{N}^*$  tout entier.

**Définition 3.4.4 — Variable géométrique.** Une variable aléatoire discrète,  $X$ , définie sur  $\mathbb{N}^*$ , est une *variable géométrique* de paramètre  $p$  si :

$$\forall n \in \mathbb{N}^*, P(X = n) = p(1 - p)^{n-1}. \quad (3.8)$$

Sa loi est souvent notée  $G(p)$ .

- R** L'appellation *géométrique* est donnée par analogie avec la série géométrique qui permet de justifier que  $\sum_{k \in \mathbb{N}^*} P(X = k) = 1$ .

**Exercice 3.5** Soit  $X$  une variable géométrique de paramètre  $p$ . Montrer que  $\sum_{k \in \mathbb{N}^*} P_X(k) = 1$  et que  $P_X$  est une loi de probabilité. ■

- R** La loi géométrique est une loi **sans mémoire** au sens où la loi de probabilité du nombre d'épreuves à répéter jusqu'à l'obtention d'un premier succès, dans une suite d'épreuves de Bernoulli identiques et indépendantes, est la même quel que soit le nombre d'échecs accumulés auparavant. Mathématiquement, cela se traduit par :

$$p(X > n + m | X > n) = p(X > m).$$

**Exercice 3.6** Montrer que la loi géométrique est une loi sans mémoire. ■

**Exercice 3.7** Montrer que la loi géométrique est la seule loi discrète sans mémoire, au sens décrit dans la remarque précédente. ■

### 3.4.5 Variable aléatoire de Poisson.

La loi de Poisson est attribuée à Siméon Denis Poisson, mathématicien français (1781-1840). Cette loi fut proposée par Poisson dans un ouvrage qu'il publia en 1837 sous le titre « Recherche sur la probabilité de jugements en matière criminelle et en matière civile ».

Beaucoup de situations sont liées à l'étude de la réalisation d'un événement dans un intervalle de temps donné (arrivée de clients qui se présentent à un guichet d'une

banque en une heure, apparitions de pannes d'un réseau informatique en une année, arrivée de malades aux urgences d'un hôpital en une nuit,...). Les phénomènes ainsi étudiés sont des phénomènes d'attente. La loi de Poisson permet de modéliser ces *temps d'attente* et permet, plus généralement, de modéliser les événements « rares ».

**Définition 3.4.5** Une variable aléatoire discrète,  $X$ , définie sur  $\mathbb{N}$  est une variable de Poisson de paramètre  $\lambda \in \mathbb{R}^{*+}$  si :

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

Sa loi est notée  $\mathcal{P}(\lambda)$ .

**R** La loi de Poisson est « construite » sur la série exponentielle.

**Exercice 3.8** Soit  $X$  une variable de Poisson de paramètre  $\lambda \in \mathbb{R}^{*+}$ . Montrer que  $P_X$  est bien une loi de probabilité. ■

**Proposition 3.4.1** La loi de Poisson  $\mathcal{P}(\lambda)$  de paramètre  $\lambda$ , est la loi limite d'une loi binomiale  $B(n, \lambda/n)$  lorsque  $n$  tend vers l'infini.

**Exercice 3.9** Démontrer la proposition 3.4.1. ■

**R** Lorsque  $n$  est grand et  $p$  petit, le produit  $np = \lambda$  reste petit par rapport à  $n$ . Alors, la loi binomiale  $\mathcal{B}(n, p)$  peut être approchée par la loi de Poisson  $\mathcal{P}(\lambda)$ . Cette approximation s'appliquant lorsque  $p$  est petit, la loi de Poisson est appelée la loi des événements rares.

En pratique :

On peut approcher la loi  $\mathcal{B}(n, p)$  par la loi  $\mathcal{P}(np)$  dès que  $n > 20$ ,  $p \leq 0.1$  et  $np \leq 5$ .

**Exercice 3.10** Soient  $X_1$  et  $X_2$  deux variables aléatoires indépendantes de lois respectives  $\mathcal{P}(\lambda_1)$  et  $\mathcal{P}(\lambda_2)$ . Quelle est la loi de la variable aléatoire  $X_1 + X_2$ ? ■

En pratique, la loi de Poisson est très utilisée. Le physicien et le technologue la rencontrent dans de nombreuses circonstances.

### Le comptage en radioactivité, microanalyse X, analyse SIMS...

Lorsqu'un détecteur compte le nombre de particules qu'il reçoit pendant un temps de comptage  $t_0$ , on constate que ce nombre  $X$  est bien modélisé par une loi de Poisson dont le paramètre  $\lambda$  est le produit du flux moyen de particules  $\alpha$  pendant le temps de comptage  $t_0$ .

- R** Comme vous le verrez durant le cours sur les « Files d'attente », la loi de Poisson permet de modéliser les temps d'arrivée (voitures à un péage, appels téléphoniques à un standard etc...).

### Contrôle de qualité

Lorsqu'on effectue un contrôle de qualité sur des composants fabriqués par une unité de production, un composant est réputé bon (ou mauvais) s'il satisfait (ou non) à un certain nombre de spécifications. Le critère de choix étant **qualitatif**, on s'intéresse au nombre  $N$  de pièces défectueuses dans un échantillon de taille  $n$ . Si on note  $p$  le taux moyen de rebuts, la loi de la v.a.  $N$  peut alors être modélisée par une loi de Poisson  $\mathcal{P}(np)$ . C'est un exemple dans lequel on approche la loi binomiale par la loi de Poisson (*loi des événements rares*).

- R** Lorsque  $np$  est petit, le rapport devient grand et la variabilité sur le nombre de pièces défectueuses rend le contrôle imprécis. Ceci explique que, dans un contrôle de qualité, la taille des échantillons tirés de la population des pièces fabriquées doit être au moins de l'ordre de 100.

### Dénombrement d'événements survenant dans un espace délimité

La loi de Poisson permet de modéliser aussi bien le nombre d'événements survenant pendant un temps donné que le nombre d'événements survenant dans un espace délimité.

Par exemple, si on appelle  $X$  le nombre de particules bombardant une cible de surface  $S$  soumise à une irradiation de fluence  $\mathcal{F}$  (mesurée en  $m^{-2}$ ), alors  $X$  suit une loi  $\mathcal{P}(\mathcal{F}S)$ .  $\mathcal{F}S$  est donc analogue au  $\alpha t_0$  du comptage en temps.

- R** La loi de Poisson sert donc aussi à modéliser des phénomènes de localisation spatiale et non plus seulement temporelle, c'est-à-dire qu'elle modélisera aussi bien le nombre d'accidents qui peuvent survenir en une matinée que le nombre d'accidents qui peuvent survenir sur une section donnée d'autoroute.

## 3.5 Lois usuelles continues

### 3.5.1 Loi uniforme continue.

**Définition 3.5.1** Une variable aléatoire  $U$  suit une loi uniforme sur  $[a, b]$  notée  $\mathcal{U}([a, b])$  si sa densité sur  $\mathbb{R}$  s'écrit :

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

- R** La loi uniforme continue prolonge, à un espace d'états non dénombrable, la loi uniforme discrète en attribuant la même valeur de la densité à tous les éléments de  $[a; b]$ .

### 3.5.2 Loi exponentielle.

**Définition 3.5.2** Une variable aléatoire  $X$  suit une loi exponentielle de paramètre  $\lambda \in \mathbb{R}^{+*}$  notée  $\mathcal{E}(\lambda)$ , si sa densité sur  $\mathbb{R}$  s'écrit :

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x).$$

**R** Comme la loi géométrique, cette loi est **sans mémoire** :

$$P(X > t + h | X > t) = P(X > h).$$

**Exercice 3.11** Montrer que la loi exponentielle est une loi sans mémoire. ■

**Exercice 3.12** Montrer que la loi exponentielle est la seule loi continue sans mémoire, au sens décrit dans la remarque précédente. ■

### 3.5.3 Loi normale.

**Définition 3.5.3** Une variable aléatoire  $X$  suit une loi normale de paramètres  $\mu \in \mathbb{R}$  et  $\sigma^2 \in \mathbb{R}^+$  notée  $\mathcal{N}(\mu, \sigma^2)$  si sa densité sur  $\mathbb{R}$  s'écrit :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

**R** Si  $X$  suit la loi normale  $\mathcal{N}(\mu, \sigma^2)$ , alors  $Z = \frac{X-\mu}{\sigma}$  suit la **loi normale centrée réduite**  $\mathcal{N}(0, 1)$ .

Nous consacrerons le chapitre 7 à la loi normale et aux vecteurs gaussiens.

### 3.5.4 Loi gamma.

**Définition 3.5.4** Une variable aléatoire  $X$  suit une loi gamma de paramètres  $k \in \mathbb{N}^{*+}$  et  $\theta \in \mathbb{R}^{*+}$  notée  $\Gamma(k, \theta)$  si sa densité sur  $\mathbb{R}$  s'écrit

$$f(t) = \frac{1}{\Gamma(k)\theta^k} t^{k-1} e^{-\frac{t}{\theta}} \mathbb{1}_{\mathbb{R}^+}(t),$$

avec  $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$  pour tout  $x > 0$ .

**R** La loi  $\Gamma(1, \frac{1}{\lambda})$  a la même densité que la loi exponentielle de paramètre  $\lambda$ .





## 4. Variables aléatoires vectorielles

« [...] Dans le petit nombre de choses que nous pouvons savoir avec certitude [...], les principaux moyens de parvenir à la vérité [...] se fondent sur les probabilités. »

Pierre Simon de Laplace.

### 4.1 Fonction de répartition conjointe et marginale

Considérons une famille d'espaces  $(E_i, \mathcal{E}_i)_{1 \leq i \leq n}$  pour un entier  $n \geq 2$ . Le produit cartésien  $F = E_1 \times \dots \times E_n$  est l'ensemble des  $n$ -uplets  $(x_1, \dots, x_n)$  où pour chaque  $i \in \{1, \dots, n\}$ ,  $x_i$  parcourt l'ensemble  $E_i$ .

On appelle  $j^{\text{ième}}$  application coordonnée l'application  $f_j$  définie par :

$$f_j : \begin{array}{l} F \rightarrow E_j \\ (x_1, \dots, x_n) \mapsto x_j \end{array} .$$

**Définition 4.1.1 — Tribu produit.** La **tribu produit** des  $\mathcal{E}_i$  est la plus petite tribu  $\mathcal{F}$  de  $F$  pour laquelle chaque application coordonnée  $f_i$  est mesurable de  $(F, \mathcal{F})$  dans  $(E_i, \mathcal{E}_i)$ .

C'est la tribu engendrée par  $\bigcup_{i=1}^n f_i^{-1}(\mathcal{E}_i)$ . On la note :

$$\mathcal{F} = \otimes_{i=1}^n \mathcal{E}_i = \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$$

Lorsque tous les  $(E_i, \mathcal{E}_i)$  sont égaux à un même  $(E, \mathcal{E})$ , on note aussi :

$$F = E^d \text{ et } \mathcal{F} = \mathcal{E}^{\otimes n}.$$

- R** Sur  $\mathbb{R}$ , on considère la famille  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))_{1 \leq i \leq n}$ . L'espace produit est donc  $(\mathbb{R}^n, (\mathcal{B}(\mathbb{R}))^{\otimes n})$  pour lequel on peut montrer que  $(\mathcal{B}(\mathbb{R}))^{\otimes n} = \mathcal{B}(\mathbb{R}^n)$ .  
On peut également montrer que pour tous entiers  $m$  et  $n$ ,  $\mathcal{B}(\mathbb{R}^{m+n}) = \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n)$ .

La théorie de la mesure nous permet donc d'étendre la notion de mesure à des produits d'espaces mesurables. En termes de probabilités, cela signifie que nous pouvons envisager d'étendre la notion de *variables aléatoires scalaires*, sur laquelle nous avons travaillé jusqu'ici, à celle de *variables aléatoires vectorielles*. Il nous faut pour cela comprendre ce qu'est une mesure définie sur un espace produit.

Le résultat suivant établit, en particulier, une Condition Nécessaire et Suffisante (CNS) d'existence d'une variable aléatoire vectorielle.

**Proposition 4.1.1** Soit  $(\Omega, \mathcal{A})$  un espace mesurable.

Une application  $f$  de  $\Omega$  dans  $F = E_1 \times \dots \times E_n$  est mesurable relativement à  $\mathcal{A}$  et  $\mathcal{F} = \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$  **si et seulement si** toutes les applications coordonnées  $f_i$  sont mesurables de  $(\Omega, \mathcal{A})$  dans  $(E_i, \mathcal{E}_i)$ .

- R** D'un point de vu probabiliste, on a une variable aléatoire vectorielle si et seulement si toutes les composantes sont des variables aléatoires.

**Définition 4.1.2** On appelle **vecteur aléatoire**  $\mathbf{X}_1^n$  sur un espace produit  $(\prod_{i=1}^n E_i, \otimes_{i=1}^n \mathcal{E}_i)$  tout  $n$ -uplet  $(X_1, \dots, X_n)$  formé de v.a.  $X_1, \dots, X_n$  définies sur  $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ .

On peut alors étendre la notion de fonction de répartition à des variables aléatoires vectorielles réelles.

**Définition 4.1.3** Soit  $\mathbf{X}_1^n$  un vecteur aléatoire réel. La **fonction de répartition conjointe** de  $\mathbf{X}_1^n$  est la fonction  $F_{\mathbf{X}_1^n}$  définie par :

$$\begin{aligned} F_{\mathbf{X}_1^n} &: \mathbb{R}^n \rightarrow [0; 1] \\ \mathbf{x}_1^n &\mapsto F_{\mathbf{X}_1^n}(x_1, \dots, x_n) = P_{\mathbf{X}_1^n}(\prod_{i=1}^n ]-\infty, x_i]) = P(\{X_1 \leq x_1\} \cap \dots \cap \{X_n \leq x_n\}). \end{aligned} \quad (4.1)$$

- R** Les propriétés du théorème 3.2.1 s'étendent naturellement au cas vectoriel.

Lorsqu'on fait tendre  $x_i$  vers  $+\infty$  dans (4.1), l'évènement concerné  $\{X_i \leq x_i\} = X_i^{-1} ]-\infty, x_i])$  devient l'évènement certain  $\Omega$ . L'intersection  $\{X_1 \leq x_1\} \cap \dots \cap \{X_n \leq x_n\}$  se restreint alors à l'intersection des  $n-1$  autres évènements  $\bigcap_{\substack{j=1 \\ j \neq i}}^n \{X_j \leq x_j\}$ . On récupère ainsi la fonction de répartition du vecteur aléatoire marginal  $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ .

**Proposition 4.1.2** La connaissance de la loi d'un vecteur aléatoire  $\mathbf{X}_1^n = (X_1, \dots, X_n)$  donne accès aux lois de tous les vecteurs marginaux formés de  $k$  composantes ( $k \in \{1, \dots, n-1\}$ ) choisies parmi les  $X_1, \dots, X_n$ . En particulier, pour tout  $i$  dans  $\{1, \dots, n\}$ , on retrouve la **fonction de répartition marginale** de  $X_i$  :

$$\forall x \in \mathbb{R}, F_{X_i}(x) = \lim_{\substack{x_j \rightarrow +\infty \\ \forall j \neq i}} F_{\mathbf{X}_1^n}(x_1, \dots, x_n) = P(\{X_i \leq x\}).$$

## 4.2 Mesure produit et Fubini

D'après le théorème de Radon-Nikodym, une variable aléatoire, même vectorielle, admet une densité par rapport à une mesure dominante appropriée  $\mu$ . Pour un *vecteur aléatoire discret* il s'agit de la généralisation de la mesure de comptage et pour un *vecteur aléatoire continu*, il s'agit de la mesure de Lebesgue sur  $\mathcal{B}(\mathbb{R}^n)$ . On a donc :

$$\forall \mathbf{x}_1^n \in \mathbb{R}^n, F_{\mathbf{X}_1^n}(x_1, \dots, x_n) = \int_{\prod_{i=1}^n ]-\infty; x_i]} f_{\mathbf{X}_1^n}(\mathbf{t}_1^n) d\mu(\mathbf{t}_1^n).$$

La proposition suivante nous éclaire d'une part sur la mesurabilité des densités marginales des variables aléatoires vectorielles (fonctions mesurables positives) et permet, d'autre part, de donner un cadre aux opérations sur les variables aléatoires vectorielles pour additionner ou multiplier les composantes ou effectuer toute transformation mesurable plus complexe.

**Proposition 4.2.1** Soit  $f$  une transformation mesurable de  $(\prod_{i=1}^n E_i, \otimes_{i=1}^n \mathcal{E}_i)$  dans  $(G, \mathcal{G})$ , un espace mesurable. Si  $k \in \{1, \dots, n-1\}$  et si les  $x_i \in E_i$  sont fixés pour  $i \in \{k+1, \dots, n\}$  alors l'application marginale :

$$h_{x_{k+1}, \dots, x_n} : (x_1, \dots, x_k) \mapsto f(x_1, \dots, x_k, x_{k+1}, \dots, x_n)$$

est mesurable de  $(\prod_{i=1}^k E_i, \otimes_{i=1}^k \mathcal{E}_i)$  dans  $(G, \mathcal{G})$ .



La réciproque est fautive. Il ne suffit pas que les applications marginales soient toutes mesurables pour que  $f$  le soit.

- R** D'un point de vu probabiliste, l'absence de réciproque à la proposition 4.2.1 nous renseigne sur le fait que la densité d'un vecteur aléatoire « concentre » plus d'information que ne peuvent en apporter toutes les densités marginales réunies.

Le théorème suivant permet de décomposer des mesures définies sur des espaces produits ce qui se montre particulièrement utile pour simplifier le calcul des intégrales.

**Théorème 4.2.2** Soient  $\mu_1, \dots, \mu_n$  une famille de mesures  $\sigma$ -finies définies respectivement sur  $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ . Alors **il existe une unique** mesure  $\mu$  sur l'espace produit  $(E_1 \times \dots \times E_n, \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n)$  telle que :

$$\forall A_1 \in \mathcal{E}_1, \dots, \forall A_n \in \mathcal{E}_n, \mu(A_1 \times \dots \times A_n) = \mu_1(A_1) \dots \mu_n(A_n).$$

Cette mesure notée  $\mu = \mu_1 \otimes \dots \otimes \mu_n$  est appelée **mesure produit**.

**R** La mesure de Lebesgue  $\lambda_n$  sur  $\mathcal{B}(\mathbb{R}^n)$  est en fait la mesure produit  $\lambda \otimes \dots \otimes \lambda = \otimes_{k=1}^n \lambda$  de  $n$  mesures de Lebesgue  $\lambda$  sur  $\mathcal{B}(\mathbb{R})$ .

Le célèbre théorème suivant que vous connaissez dans le cas de l'intégrale de Riemann, s'étend à l'intégrale de Lebesgue et est particulièrement utile en probabilités.

**Théorème 4.2.3 — Fubini.** Soient  $\mu_1, \dots, \mu_n$ ,  $n$  mesures  $\sigma$ -finies définies respectivement sur les espaces  $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ . Soit  $f$  une fonction positive mesurable sur l'espace produit  $(F, \mathcal{F}) = (E_1 \times \dots \times E_n, \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n)$  et soit également  $\mu = \mu_1 \otimes \dots \otimes \mu_n$  la mesure produit définie sur le même espace. Alors

•

$$\int_F f(\mathbf{x}_1^n) d\mu(\mathbf{x}_1^n) = \int_{E_n} \dots \left( \int_{E_2} \left( \int_{E_1} f(x_1, \dots, x_n) d\mu_1(x_1) \right) d\mu_2(x_2) \right) \dots d\mu_n(x_n);$$

- Les intégrales successives du membre de droite sont calculables dans n'importe quel ordre;
- Toutes les intégrales partielles sont mesurables (sur l'espace idoine).

Grâce au Théorème de Fubini, si on peut écrire la mesure sur l'espace produit comme une mesure produit (ce qui est le cas, sur  $\mathbb{R}^n$ , pour la mesure de Lebesgue et la mesure de comptage), on peut réécrire la fonction de répartition d'un vecteur aléatoire. Si  $\mathbf{X}_1^n$  admet une densité par rapport à une mesure produit  $\mu$ , alors

$$\forall \mathbf{x}_1^n \in \mathbb{R}^n, F_{\mathbf{X}_1^n}(\mathbf{x}_1^n) = \int_{]-\infty; x_n]} \dots \left( \int_{]-\infty; x_2]} \left( \int_{]-\infty; x_1]} f(x_1, \dots, x_n) d\mu_1(x_1) \right) d\mu_2(x_2) \right) \dots d\mu_n(x_n), \quad (4.2)$$

et toutes les intégrales partielles sont mesurables,

Intéressons-nous à ce qui se passe lorsque l'on fait tendre vers  $+\infty$  tous les  $x_j$  pour  $i \neq j$ . Comme nous pouvons intégrer « dans l'ordre que nous voulons » le membre de droite de (4.2), gardons l'intégrale par rapport à la  $i^{\text{ème}}$  composante  $x_i$  de  $\mathbf{x}_1^n$  pour la fin. On retrouve évidemment le résultat sur les fonctions de répartition :

$$\forall x_i \in \mathbb{R}, F_{X_i}(x_i) = \int_{]-\infty; x_i]} \left[ \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(x_1, \dots, x_i, \dots, x_n) d\mu_1(x_1) \dots d\mu_n(x_n) \right] d\mu_i(x_i), \quad (4.3)$$

et l'intégrale « entre crochets » est donc tout simplement la densité  $f_{X_i}$  de la variable  $X_i$  :

$$\int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(x_1, \dots, x_i, \dots, x_n) d\mu_1(x_1) \dots d\mu_n(x_n) = f(x_i) \quad (4.4)$$

De manière générale, lorsqu'on intègre, sur  $\mathbb{R}$ , la densité du vecteur  $\mathbf{X}_1^n$  suivant l'une des composantes,  $X_d$ , on obtient la **densité marginale** du vecteur  $(X_1, \dots, X_{d-1}, X_{d+1}, \dots, X_n)$ .

### 4.3 Changement de variable aléatoire multidimensionnel

Après le théorème de Fubini, la formule de changement de variable est le deuxième outil fondamental du calcul des probabilités.

**Définition 4.3.1** —  $C^1$ -difféomorphisme. On rappelle qu'une application  $\phi : U \rightarrow V$  est un  $C^1$ -difféomorphisme si  $\phi$  est bijective et de classe  $C^1$  sur  $U$  et que  $\phi^{-1}$  est bijective et de classe  $C^1$  sur  $V$ .

En probabilités, nous nous intéressons en particulier au cas des  $C^1$ -difféomorphismes pour lesquels  $U$  et  $V$  sont des ouverts de  $\mathbb{R}^d$ . Rappelons, dans ce contexte, ce que sont la **matrice jacobienne** et le **jacobien** de  $\phi$ . Pour cela, notons  $\phi_i$ ,  $1 \leq i \leq d$ , la  $i^{\text{ème}}$  composante de  $\phi$ .

**Définition 4.3.2** La matrice  $\phi'(x) = \left( \frac{\partial \phi_i}{\partial x_j}(x) \right)_{1 \leq i, j \leq d}$  définie au point  $x$  de  $U$  est appelée **matrice jacobienne** de  $\phi$ .

Le déterminant au point  $x \in U$  de la matrice jacobienne  $D\phi(x) = \det(\phi'(x))$  est appelé **jacobien** de  $\phi$  en  $x$ . Le jacobien de  $\phi$  en  $x$  est aussi noté  $J_\phi(x)$ .

**R** Souvent, par analogie avec la notation de la dérivée dans  $\mathbb{R}$  dont c'est une extension, le jacobien  $D\phi(x)$  est présenté sous la forme :

$$D\phi(x) = \frac{D(\phi_1(x), \dots, \phi_d(x))}{D(x_1, \dots, x_d)}. \quad (4.5)$$

Cette notation est cohérente avec le fait qu'en « dérivant »  $\phi \circ \phi^{-1}(x) = x$ , on s'aperçoit que  $D\phi \times D\phi^{-1} = 1$ .

**R** Le jacobien de  $\phi$  en  $x$  peut aussi être simplement noté  $J_\phi(x)$ .

Le théorème de changement de variable s'énonce alors comme suit :

**Théorème 4.3.1** Soit  $\phi : U \rightarrow V$  un  $C^1$ -difféomorphisme. Alors, pour toute fonction borélienne  $f : V \rightarrow \mathbb{R}^+$  :

$$\int_V f(y) dy = \int_U f(\phi(x)) |D\phi(x)| dx. \quad (4.6)$$

**R** Attention à bien prendre la valeur absolue du déterminant 

Si l'on reprend la notation de l'égalité (4.5) en posant pour tout  $i$ ,  $\phi_i(x) = y_i$ , on a :

$$D\phi(x) = \frac{D(y_1, \dots, y_d)}{D(x_1, \dots, x_d)},$$

et la formule (4.6) de changement de variable devient

$$\int_V f(y_1, \dots, y_d) dy_1 \dots dy_d = \int_U f(\phi(x_1, \dots, x_d)) \left| \frac{D(y_1, \dots, y_d)}{D(x_1, \dots, x_d)} \right| dx_1 \dots dx_d.$$

Compte tenu de la définition des  $y_i$ , formellement, l'élément différentiel  $dy_1 \dots dy_d$  est remplacé par  $\left| \frac{D(y_1, \dots, y_d)}{D(x_1, \dots, x_d)} \right| dx_1 \dots dx_d$ . On peut donc également remplacer  $dx_1 \dots dx_d$  par :

$$\left| \frac{D(x_1, \dots, x_d)}{D(y_1, \dots, y_d)} \right| dy_1 \dots dy_d = \frac{1}{|D\phi(\mathbf{x})|} d\mathbf{y} = |D\phi^{-1}(\mathbf{y})| d\mathbf{y}.$$

Pour toute fonction borélienne  $g : U \rightarrow \mathbb{R}^+$  on a donc :

$$\int_U g(x) dx = \int_V g(\phi^{-1}(y)) \left| \frac{D(x_1, \dots, x_d)}{D(y_1, \dots, y_d)} \right| dy_1 \dots dy_d = \int_V g(\phi^{-1}(y)) \frac{1}{|D\phi(x)|} dy.$$

En probabilités, ce qui nous intéresse, c'est de pouvoir calculer la densité de la variable aléatoire obtenue par un changement de variable. Pour cela, souvenons-nous de la définition 3.1.2 de la mesure image, qui intervient dans l'énoncé du corollaire du théorème 4.3.1 qui règle la question dans le cas d'un changement de variable **bijectif**.

**Corollaire 4.3.2** Soit une mesure  $\mu$  admettant une densité  $g$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$ . Soit  $\phi$  un  $C^1$ -difféomorphisme de  $\mathbb{R}^d$  dans lui-même. Notons enfin  $\nu$  la mesure image de  $\mu$  par  $\phi$ , (i.e. pour tout ensemble  $A$  mesurable,  $\nu(\phi(A)) = \mu(A)$ ). Alors :

- $\nu$  admet une densité  $f$  par rapport à la mesure de Lebesgue du  $\mathbb{R}^d$  qui vérifie  $\int_{\phi(A)} f = \nu(\phi(A)) = \mu(A) = \int_A g$ ;
- $\forall x \in \mathbb{R}^d$  :

$$f(x) = g(\phi^{-1}(x)) |(D\phi^{-1}(x))|. \quad (4.7)$$

**R** Nous verrons en TD que cette méthode peut également être utilisée pour traiter des cas de changement de variables qui ne sont pas bijectifs au départ.

## 5. Indépendance de variables aléatoires

*« Quels que soient les progrès des connaissances humaines, il y aura toujours place pour l'ignorance et par suite pour le hasard et la probabilité. »*

Émile Borel.

### 5.1 Indépendance de variables aléatoires

Avant de pouvoir définir la notion d'indépendance de variables aléatoires, il est nécessaire de lui donner un cadre qui étende la notion d'indépendance d'évènements que nous connaissons. C'est la notion d'indépendance de tribus qui joue ce rôle.

**Définition 5.1.1** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé,  $\mathcal{A}_1$  et  $\mathcal{A}_2$  deux sous-tribus de  $\mathcal{A}$ . On dit que les tribus  $\mathcal{A}_1$  et  $\mathcal{A}_2$  sont indépendantes sous  $P$  si  $\forall A_1 \in \mathcal{A}_1$  et  $\forall A_2 \in \mathcal{A}_2$ ,  $P(A_1 \cap A_2) = P(A_1)P(A_2)$ .

Plus généralement, lorsque  $I$  est un sous-ensemble fini ou dénombrable de  $\mathbb{N}$ , si  $(\mathcal{A}_i)_{i \in I}$  est une famille de sous-tribus de  $\mathcal{A}$ , on dit que cette famille est indépendante sous  $P$  si pour tout ensemble fini  $J \subset I$ , on a :

$$\forall (A_j)_{j \in J} \in \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_j, \quad P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

**R** Si les tribus  $\mathcal{A}_1$  et  $\mathcal{A}_2$  sont indépendantes sous  $P$ , et que  $\mathcal{A}'_1$  est une sous tribu de  $\mathcal{A}_1$  et  $\mathcal{A}'_2$  une sous-tribu de  $\mathcal{A}_2$ , alors  $\mathcal{A}'_1$  et  $\mathcal{A}'_2$  sont indépendantes sous  $P$ .

**Exercice 5.1** Soient  $A_1, A_2 \in \mathcal{A}$ . Montrer que  $A_1$  et  $A_2$  sont indépendants si et seulement si les tribus  $\sigma(A_1)$  et  $\sigma(A_2)$  sont indépendantes. ■

Nous avons défini plus haut la notion de tribu engendrée par un ensemble. Étendons cette notion aux variables aléatoires.

**Proposition 5.1.1** Considérons une variable aléatoire,  $X$ , définie de  $(\Omega, \mathcal{A}, P)$  dans  $(\Omega', \mathcal{A}', P_X)$ . La famille :

$$\sigma(X) = \{X^{-1}(A), A \in \mathcal{A}'\}$$

est une tribu.

**Exercice 5.2** Démontrer la proposition 5.1.1. ■

**Définition 5.1.2** La **tribu engendrée** par la variable aléatoire  $X$  définie de  $(\Omega, \mathcal{A}, P)$  dans  $(\Omega', \mathcal{A}', P_X)$  est la famille  $\sigma(X) = \{X^{-1}(A), A \in \mathcal{A}'\}$ .

On peut alors étendre la notion d'indépendance aux variables aléatoires.

**Définition 5.1.3** Soit  $I$ , un sous-ensemble fini ou dénombrable de  $\mathbb{N}$ . Les variables aléatoires  $(X_i)_{i \in I}$  sont **indépendantes** si les tribus de la famille  $(\sigma(X_i))_{i \in I}$  sont indépendantes.

**R** Dans cette définition la notion de variable aléatoire peut être entendue au sens large et inclut donc les vecteurs aléatoires.

**R** L'indépendance des variables aléatoires  $X$  et  $Y$  est souvent notée  $X \perp Y$ .

Le théorème suivant, appelé **lemme des coalitions**, est très important lorsque l'on fait des transformations mesurables des variables aléatoires.

**Théorème 5.1.2** Soit  $I$ , un sous-ensemble fini ou dénombrable de  $\mathbb{N}$  et  $(X_i)_{i \in I}$  une famille de variables aléatoires indépendantes pour lesquelles  $X_i$  est à valeurs dans  $\mathbb{R}^{n_i}$ . Considérons également une famille,  $(f_i)_{i \in I}$ , d'applications mesurables ( $f_i$  est mesurable de  $\mathbb{R}^{n_i}$  dans  $\mathbb{R}^{m_i}$ ).

Alors les variables aléatoires :

$$Y_i = f_i(X_i)$$

sont indépendantes.

**R** Ce qu'il faut retenir du théorème précédent, c'est que les transformations mesurables de variables aléatoires indépendantes permettent de préserver l'indépendance.

Par exemple si  $X, Y$  et  $Z$  sont indépendantes alors  $X^2$  et  $Y + Z$  sont indépendantes.

Un autre résultat éclaire la notion d'indépendance sous l'angle de la mesure produit, ce qui se révèle également très utile pour le calcul des probabilités.

**Théorème 5.1.3** Soit  $(X_1, \dots, X_n)$  un vecteur aléatoire.

Alors :

$$\begin{aligned} X_1, \dots, X_n \text{ sont indépendantes} &\Leftrightarrow P_{X_1, \dots, X_n} = P_{X_1} \otimes \dots \otimes P_{X_n} \\ &\Leftrightarrow F_{X_1, \dots, X_n} = F_{X_1} \cdots F_{X_n}. \end{aligned}$$

**R** Dans ce théorème, rien n'empêche les  $X_i$  d'être des vecteurs aléatoires.

Le théorème 5.1.3 qui nous donne une CNS d'indépendance nous permet, par exemple, d'envisager de montrer l'indépendance de variables aléatoires sans passer par la fastidieuse (et même impossible) vérification des conditions d'indépendance des tribus engendrées. En effet, nous savons désormais que les  $(X_i)_{i \in \{1, \dots, n\}}$  sont indépendantes si et seulement si la fonction de répartition  $F_{X_1, \dots, X_n}$  de  $(X_1, \dots, X_n)$  est le produit des fonctions de répartition  $F_{X_i}$  des  $X_i$ . Pensez-y!

Le corolaire suivant nous permet de conclure ce paragraphe sur l'indépendance par une condition suffisante d'indépendance.

**Corollaire 5.1.4** Soit  $(X_1, \dots, X_n)$  un vecteur aléatoire de  $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ . Soient également  $\mu_1, \dots, \mu_n$  des mesures  $\sigma$ -finies définies respectivement sur les  $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$  et telles que  $P_{X_1, \dots, X_n} = \mu_1 \otimes \dots \otimes \mu_n$  alors les  $X_1, \dots, X_n$  sont indépendantes.

**R** Le corollaire 5.1.4 affirme qu'il suffit de pouvoir décomposer la loi du vecteur  $(X_1, \dots, X_n)$  en un produit de mesures « séparant les variables » pour avoir l'indépendance des  $X_i$ .



les  $\mu_i$  ne sont pas nécessairement des mesures de probabilité.

## 5.2 Somme de variables aléatoires indépendantes et convolution

D'un manière générale, la loi de probabilité de toute variable aléatoire issue d'un changement de variable (même non bijectif) peut être obtenue en passant par sa fonction de répartition. Le cas particulier de la somme de deux variables aléatoires en est un exemple très important. Considérons, donc, deux variables aléatoires  $X$  et  $Y$  et intéressons-nous au changement de variable mesurable :

$$\begin{aligned} \Sigma : (\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d)) &\rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \\ (X, Y) &\mapsto X + Y \end{aligned}$$

Un tel changement de variable ne peut pas être bijectif puisque les espaces de départ et d'arrivée ne sont pas de même dimension.

**Définition 5.2.1** Si  $\mu$  et  $\nu$  sont deux mesures  $\sigma$ -finies définies sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ , on appelle **produit de convolution** de  $\mu$  et  $\nu$  la mesure image de la mesure produit  $\mu \otimes \nu$  par l'application  $\Sigma$ .

Pour tout  $A \in \mathcal{B}(\mathbb{R}^d)$  on a donc :

$$\mu \star \nu(A) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_A(x+y) d(\mu \otimes \nu)(x, y).$$

Le produit de convolution de  $\mu$  et  $\nu$  est noté  $\mu \star \nu$ .

**R** A titre personnel, je trouve que cette définition dans laquelle nous réinvestissons toutes les notions importantes que nous avons vues jusqu'ici est d'une élégance remarquable.

Un problème courant consiste à trouver la loi de probabilité d'une somme de deux variables indépendantes  $Z = X + Y$ . Si le couple  $(X, Y)$  admet une densité par rapport à la mesure de Lebesgue, la fonction de répartition de la variable  $Z = X + Y$  est, pour tout  $z = (z_1, \dots, z_d)$  :

$$F_Z(z) = \mu \star \nu \left( \prod_{i=1}^d ]-\infty; z_i] \right) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_{\prod_{i=1}^d ]-\infty; z_i]}(x+y) f_{X,Y}(x,y) dx dy,$$

et on montre que :

$$f_Z(z) = f_X \star f_Y = \int_{\mathbb{R}^d} f_X(x) f_Y(z-x) dx = \int_{\mathbb{R}^d} f_X(z-y) f_Y(y) dy. \quad (5.1)$$

**Exercice 5.3** Démontrer (5.1). ■

On détaille ci dessous le cas discret et le cas continu.

### 5.2.1 Cas discret

Le théorème des probabilités totales donne la solution du problème

$$P(Z = z) = \sum_x P(X = x \text{ et } Y = z - x) = \sum_y P(X = z - y \text{ et } Y = y)$$

Lorsque  $X$  et  $Y$  sont indépendantes, on a

$$P(Z = z) = \sum_x P(X = x) \cdot P(Y = z - x) = \sum_y P(X = z - y) \cdot P(Y = y) \quad (5.2)$$

**R** Certains termes peuvent être nuls.  $x$  ne prend pas nécessairement toutes les valeurs possibles de  $X$  mais uniquement celles compatibles avec l'événement  $Z = z$  (de même pour  $y$  valeurs de  $Y$ ).

**Proposition 5.2.1** La somme de deux variables aléatoires indépendantes de lois de POISSON  $\mathcal{P}(\lambda)$  et  $\mathcal{P}(\mu)$  suit la loi de POISSON  $\mathcal{P}(\lambda + \mu)$ .

*Démonstration.* Soient  $X$  une variable aléatoire de loi  $\mathcal{P}(\lambda)$ ,  $Y$  une variable aléatoire de loi  $\mathcal{P}(\mu)$ , avec  $X$  et  $Y$  indépendantes. Soit  $Z = X + Y$ , la variable aléatoire somme.  $X$  et  $Y$  prennent toutes les valeurs entières positives,  $Z$  prend donc aussi toutes les valeurs

entières positives. Pour tout entier positif  $k$

$$\begin{aligned}
 P(Z = k) &= \sum_{i \in \mathbb{N}} P(X = i) \cdot P(Y = k - i) \\
 &= \sum_{i=0}^k \left( e^{-\lambda} \frac{\lambda^i}{i!} \right) \times \left( e^{-\mu} \frac{\mu^{k-i}}{(k-i)!} \right) \\
 &= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{i=0}^k \frac{k!}{i!(k-i)!} \lambda^i \mu^{k-i} \\
 &= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{i=0}^k \binom{k}{i} \lambda^i \mu^{k-i} \\
 P(Z = k) &= \frac{e^{-(\lambda+\mu)}}{k!} (\lambda + \mu)^k \quad \text{par le binôme de NEWTON}
 \end{aligned}$$

Donc  $Z$  suit la loi de POISSON  $P(\lambda + \mu)$ . ■

On a un résultat analogue pour la somme de loi binomiale. On démontrerait

**Proposition 5.2.2** La somme de deux variables aléatoires indépendantes de lois binomiale  $\mathcal{B}(n, p)$  et  $\mathcal{B}(m, p)$  suit une loi binomiale  $\mathcal{B}(n + m, p)$ .

La loi  $\mathcal{B}(n, p)$  peut être considérée comme la somme de  $n$  variables aléatoires mutuellement indépendantes de BERNOULLI  $\mathcal{B}(1, p)$ . Dès lors, si  $X_i$  désigne  $n$  variables aléatoires indépendantes, alors  $S_n = \sum_{i=1}^n X_i$  suit une loi binomiale et on retrouve

$$\mathbb{E}(S_n) = \sum_{i=1}^n \mathbb{E}(X_i) = \sum_{i=1}^n p = np \quad \text{et} \quad \text{Var}(S_n) = \sum_{i=1}^n \text{Var}(X_i) = npq$$

### 5.2.2 Cas continu

Rappelons la définition d'un produit de convolution  $*$ . Sous réserve de convergence

$$(f * g)(x) = \int_{\mathbb{R}} f(y - x) g(y) dy \quad (5.3)$$

**Proposition 5.2.3** La densité de probabilité de la somme de deux variables aléatoires à densité indépendantes est le produit de convolution des densités des deux variables aléatoires. C'est à dire

$$f_{X+Y} = f_X * f_Y \quad (5.4)$$

**R** Cette formule est la traduction « continue » de la formule du cas discret. Pour la loi (fonction de densité ou probabilité ponctuelle) de  $Z$  en  $z$ , on ajoute (intégrer ou sommer) tous les produits (des fonctions de densité ou des probabilités ponctuelles) pour les valeurs  $x$  et  $y$  de  $X$  et  $Y$  telles que  $x + y = z$ .

■ **Exemple 5.1** La somme de deux lois Gamma indépendantes  $\Gamma(r, \lambda)$  et  $\Gamma(s, \lambda)$  suit la loi Gamma  $\Gamma(r + s, \lambda)$ . Soient  $X$  et  $Y$  deux variables aléatoires de loi respective  $\Gamma(r, \lambda)$

et  $\Gamma(s, \lambda)$  avec  $X$  et  $Y$  indépendantes. Soit  $Z = X + Y$ , la variable aléatoire somme.  $X$  et  $Y$  prennent toutes les valeurs réelles positives,  $Z$  prend donc aussi toutes les valeurs réelles positives. Pour tout réel  $z$

$$f_Z(z) = (f_X * f_Y)(z) = \begin{cases} 0 & \text{si } z < 0 \\ \int_0^z f_X(t) \cdot f_Y(z-t) dt & \text{sinon} \end{cases}$$

La fonction  $f_Z$  n'est pas forcément définie en 0, on peut choisir  $f_Z(0) = 0$  et si  $z > 0$

$$\begin{aligned} f_Z(z) &= \frac{\lambda^{r+s}}{\Gamma(r) \cdot \Gamma(s)} \int_0^z t^{r-1} e^{-\lambda t} \cdot (z-t)^{s-1} e^{-\lambda(z-t)} dt \\ &= \frac{\lambda^{r+s} e^{-\lambda z}}{\Gamma(r) \cdot \Gamma(s)} \int_0^z t^{r-1} (z-t)^{s-1} dt \\ &= \frac{\lambda^{r+s} e^{-\lambda z}}{\Gamma(r) \cdot \Gamma(s)} \int_0^1 (uz)^{r-1} (z-uz)^{s-1} z du && \text{par } t = uz \\ f_Z(z) &= \frac{\lambda^{r+s} z^{r+s-1} e^{-\lambda z}}{\Gamma(r) \cdot \Gamma(s)} \beta(r, s) && \text{où } \beta(r, s) = \int_0^1 u^{r-1} (1-u)^{s-1} du \end{aligned}$$

$f_Z$  est une densité, la valeur du réel  $\beta(r, s)$  est telle que  $\int_{\mathbb{R}} f_Z = 1$ , donc **la fonction bêta** est

$$\beta(r, s) = \frac{\Gamma(r) \cdot \Gamma(s)}{\Gamma(r+s)} \quad (5.5)$$

$$f_Z(z) = (f_X * f_Y)(z) = \begin{cases} 0 & \text{si } z < 0 \\ \lambda^{r+s} \frac{z^{r+s-1} e^{-\lambda z}}{\Gamma(r+s)} & \text{sinon} \end{cases} \quad \text{La loi de } Z \text{ est donc la loi } \Gamma(r+s, \lambda). \quad \blacksquare$$

■ **Exemple 5.2** Soient  $Y_1, \dots, Y_n$  des variables aléatoires indépendantes suivant des lois normales centrées réduites et  $Z$  définie par

$$Z = Y_1^2 + \dots + Y_k^2$$

Nous avons vu à l'exemple 3.1 que  $Y_i$  suit une loi du  $\chi^2$  à 1 degré de liberté. Or cette dernière est aussi une loi Gamma de paramètre  $(\frac{1}{2}; \frac{1}{2})$ . D'après ce qui précède,  $Z$  suit donc une loi  $\Gamma(\frac{k}{2}; \frac{1}{2})$ . C'est l'expression de la loi du  $\chi^2$  à  $k$  degré de liberté. En particulier, on en déduit l'expression de l'espérance et de la variance pour la loi du  $\chi^2$

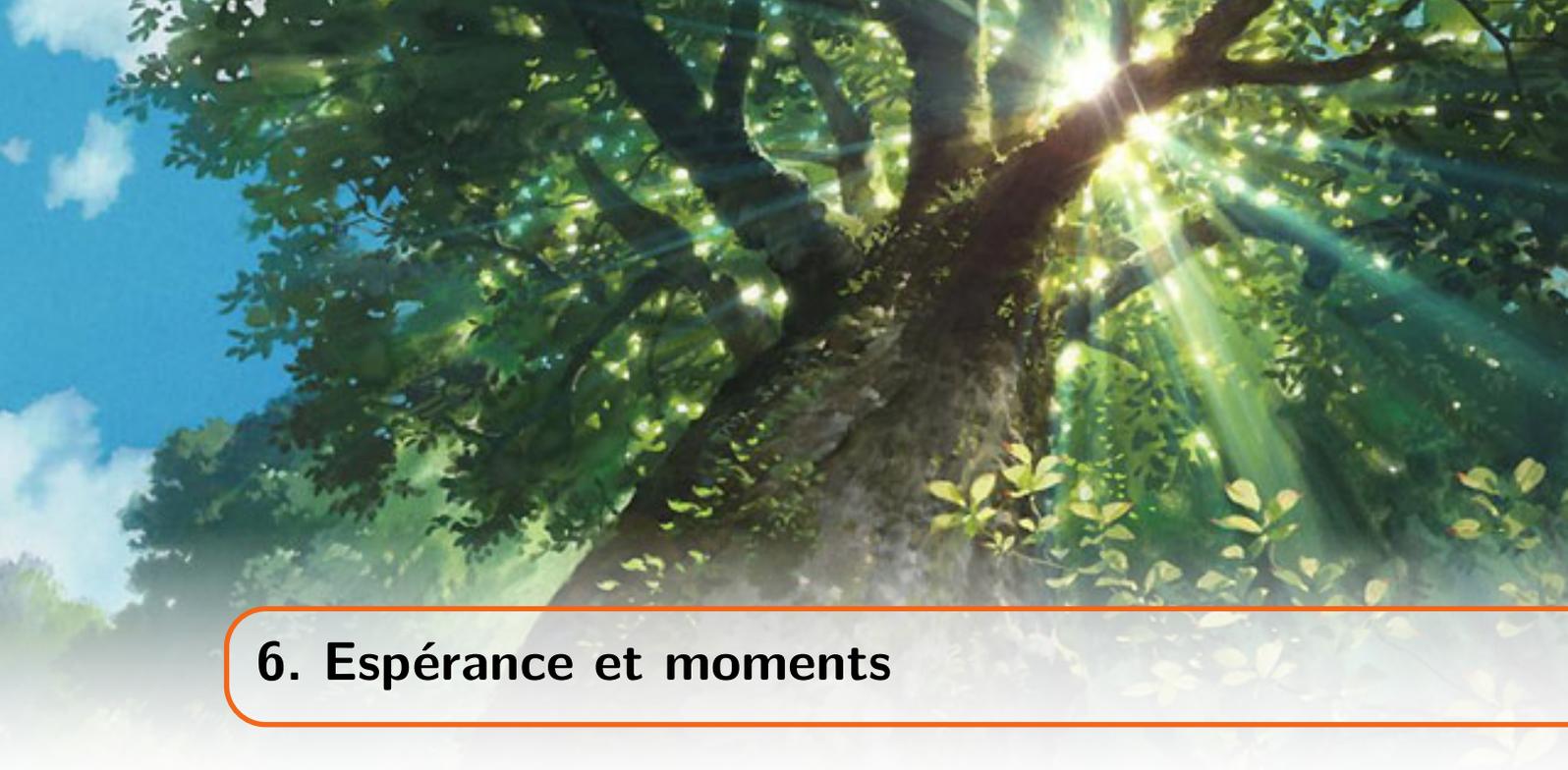
$$\mathbb{E}(Z) = k\mathbb{E}(Y_1^2) = k \quad \text{et} \quad \text{Var}(Z) = k\text{Var}(Y_1^2) = 2k \quad \blacksquare$$

De même que pour la somme de loi  $\Gamma$ , on démontrerait le résultat fondamental suivant

**Proposition 5.2.4** La somme de deux lois normales indépendantes  $\mathcal{N}(\mu, \sigma)$  et  $\mathcal{N}(\mu', \sigma')$  suit la loi normale  $\mathcal{N}(\mu + \mu', \sqrt{\sigma^2 + \sigma'^2})$ .

Ce résultat se démontre en faisant le produit de convolution de deux gaussiennes, ou plus simplement en utilisant les résultats connus sur les fonctions caractéristiques, voir la section 6.4.





## 6. Espérance et moments

« *L'accumulation met fin à l'impression de hasard.* »

Sigmund Freud.

### 6.1 La notion d'espérance mathématique

**Définition 6.1.1** Soit  $X$  v.a. de  $(\Omega, \mathcal{A}, P)$  dans  $(E, \mathcal{E})$ . L'**espérance mathématique** de la v.a.  $X$ , si elle existe, est définie par :

$$\mathbb{E}(X) = \int_{\Omega} X(\omega)P(d\omega).$$

**R** Lorsque l'espérance de  $X$  est définie ( $\mathbb{E}(|X|) < \infty$ ), on dit que  $X$  est intégrable ce qui est le cas lorsque :

- $X \geq 0$  (et dans ce cas  $\mathbb{E}(X) \in [0, +\infty]$ );
- $X$  de signe quelconque et  $\int_{\Omega} |X(\omega)|P(d\omega) < \infty$ .

**Propriété 6.1.1** L'espérance est une intégrale au sens de Lebesgue. A ce titre, elle hérite de toutes les propriétés de celle-ci que nous rappelons dans le cas des variables aléatoires réelles (v.a.r.) :

- **Linéarité** : si  $X$  et  $Y$  sont deux v.a.r. et  $a, b \in \mathbb{R}$ ,  $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ .
- **Croissance** : si  $X$  et  $Y$  sont deux v.a.r. telles que  $X \leq Y$  (c'est à dire  $\forall \omega \in \Omega, X(\omega) \leq Y(\omega)$ ) alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .
- **Variable aléatoire constante** : si  $X$  v.a.r. et  $a \in \mathbb{R}$  tels que  $X(\omega) = a, \forall \omega$ ,

alors  $\mathbb{E}(X) = a$ .

- Si  $X$  et  $Y$  sont deux v.a.r. telles que  $X = Y$  p.s. alors  $\mathbb{E}(X) = \mathbb{E}(Y)$ .

**Exercice 6.1** Démontrer les quatre assertions de la propriété 6.1.1. ■

**Théorème 6.1.2 — Théorème de transfert ou de transport.** Soit  $X$  une variable aléatoire à valeurs dans  $(E, \mathcal{E})$ . Soit  $\phi$  une application mesurable de  $(E, \mathcal{E})$  dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  telle que  $\phi(X)$  soit une variable aléatoire intégrable. Nous avons alors :

$$\mathbb{E}(\phi(X)) = \int_{\Omega} \phi(X(\omega)) P(d\omega) = \int_E \phi(x) P_X(dx).$$

En particulier, pour une variable aléatoire définie sur  $E = \mathbb{R}^d$  admettant une densité  $f_X$  on a :

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

- **Exemple 6.1** Soit  $X$  v.a.r. de densité exponentielle définie  $\forall x \in \mathbb{R}$  par  $f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x)$ ,  $\lambda > 0$ . On a alors

$$\begin{aligned} \mathbb{E}(X) &= \int_{\mathbb{R}} \lambda x e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x) dx \\ &= \int_0^{+\infty} \lambda x e^{-\lambda x} dx \\ (\text{intégration par parties}) &= [-x e^{-\lambda x}]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx \\ &= 0 + \frac{1}{\lambda} \int_0^{+\infty} \lambda e^{-\lambda x} dx \\ &= \frac{1}{\lambda}. \end{aligned}$$

- **Exemple 6.2** Soit  $X$  v.a. binomiale à valeurs dans  $\{0, \dots, n\}$  ( $n$  un entier fixé) avec  $\forall 0 \leq k \leq n$ ,  $P(X = k) = C_n^k p^k (1-p)^{n-k}$  ( $p \in [0, 1]$  fixé). Alors

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^n k P(X = k) \\ &= \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n \frac{n(n-1)!}{(k-1)!(n-1-(k-1))!} p^k (1-p)^{n-k} \\ (\text{changement d'indice en } q = k-1) &= n \sum_{q=0}^{n-1} C_{n-1}^q p^{q+1} (1-p)^{n-1-q} \\ &= np(p+1-p)^{n-1-q} = np. \end{aligned}$$

La remarque suivante souligne à nouveau le fait que le calcul des probabilités est un calcul d'intégrales.

**R** Soit  $X$  une variable aléatoire réelle de densité  $f_X$  et  $B \in \mathcal{B}(\mathbb{R})$ .

$$\begin{aligned} P(X \in B) &= \int_B f_X(x) dx \\ &= \int_{\mathbb{R}} \mathbb{1}_B(x) f_X(x) dx \\ &= \mathbb{E}(\mathbb{1}_B(X)). \end{aligned}$$

## 6.2 Moments d'une variable aléatoire

**Définition 6.2.1 — Moments d'une v.a.** Si  $X$  est une v.a. telle que pour  $k \in \mathbb{N}$ ,  $X^k$  est intégrable alors le **moment d'ordre**  $k$  de  $X$  est défini par :

$$m_k(X) = \mathbb{E}(X^k).$$

Le moment centré d'ordre  $k$  de  $X$  est défini par :

$$\mu_k(X) = \mathbb{E}\left((X - \mathbb{E}(X))^k\right).$$

Parmi ces moments, le moment centré d'ordre 2 de  $X$  joue un rôle particulier. On le connaît également sous le nom de **variance** de  $X$ .

**Définition 6.2.2 — Variance.** Si  $X$  est une v.a. telle que  $X^2$  est intégrable alors la variance de  $X$  est le moment centré d'ordre 2 de  $X$  c'est à dire :

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

$\text{Var}(X)$  est souvent notée  $\sigma_X^2$  voire  $\sigma^2$  s'il n'y a pas d'ambiguïté.

**Propriété 6.2.1** La variance de  $X$  est également :

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

**Exercice 6.2** Démontrer la propriété 6.2.1. ■

**R** Cette forme de la variance est particulièrement utile pour les calculs de variance.

■ **Exemple 6.3** Soit  $X$  v.a.r. de densité exponentielle définie  $\forall x \in \mathbb{R}$  par  $f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x)$ ,  $\lambda >$

0. On calcule :

$$\begin{aligned}
 \mathbb{E}(X^2) &= \int_{\mathbb{R}} \lambda x^2 e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x) dx \\
 &= \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx \\
 \text{(intégration par parties)} &= [-x^2 e^{-\lambda x}]_0^{+\infty} + \int_0^{+\infty} 2x e^{-\lambda x} dx \\
 &= 0 + \frac{2}{\lambda} \mathbb{E}(X) \\
 &= \frac{2}{\lambda^2}.
 \end{aligned}$$

La variance de  $X$  est donc :

$$\begin{aligned}
 \text{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
 &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \\
 &= \frac{1}{\lambda^2}.
 \end{aligned}$$

■  
**Exemple 6.4** Soit  $X$  v.a. binomiale à valeurs dans  $\{0, \dots, n\}$  ( $n$  un entier fixé) avec  $\forall 0 \leq k \leq n, P(X = k) = C_n^k p^k (1-p)^{n-k}$  ( $p \in [0, 1]$  fixé). On calcule :

$$\begin{aligned}
 \mathbb{E}(X) &= \sum_{k=0}^n k^2 P(X = k) \\
 &= \sum_{k=0}^n k^2 C_n^k p^k (1-p)^{n-k} \\
 &= \sum_{k=0}^n k(k-1) C_n^k p^k (1-p)^{n-k} + \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} \\
 &= \sum_{k=2}^n \frac{n(n-1)(n-2)!}{(k-2)!(n-2-(k-2))!} p^k (1-p)^{n-k} + \mathbb{E}(X) \\
 \text{(changement d'indice)} &= n(n-1)p^2 + np.
 \end{aligned}$$

La variance de  $X$  est donc :

$$\begin{aligned}
 \text{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
 &= n(n-1)p^2 + np - (np)^2 = -np^2 + np \\
 &= np(1-p).
 \end{aligned}$$

■  
 Avant d'aller plus loin, introduisons une généralisation de la variance appelée **covariance**.

**Définition 6.2.3** Considérons deux variables aléatoires  $X$  et  $Y$  de carré intégrable. La **covariance** de  $X$  et  $Y$  est définie par :

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

**R** La covariance de  $X$  et  $X$  n'est rien d'autre que la variance de  $X$  :

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

**Propriété 6.2.2** Comme pour la variance, on peut développer la covariance sous la forme :

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

**Exercice 6.3** Démontrer la propriété 6.2.2. ■

Grâce à la notion de covariance, nous pouvons préciser le comportement du moment centré d'ordre 2 vis à vis des transformations linéaires.

**Propriété 6.2.3** Contrairement à l'espérance, la variance et la covariance ne sont pas des opérateurs linéaires. Plus précisément :

- si  $X$  est une v.a. et  $a, b \in \mathbb{R}$ ,  $\text{Var}(aX + b) = a^2\text{Var}(X)$  ;
- si  $X$  et  $Y$  sont deux v.a. et  $a, a', b, b' \in \mathbb{R}$ ,  $\text{Cov}(aX + b, a'Y + b') = aa'\text{Cov}(X, Y)$  . ;
- si  $X$  et  $Y$  sont deux v.a.,  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$ .

**Exercice 6.4** Démontrer les trois résultats de la propriété 6.2.3. ■

Dans le cas général, la variance d'une somme de variables aléatoires n'est donc pas égale à la somme des variances de ces variables aléatoires. C'est le cas, si et seulement si,  $\text{Cov}(X, Y) = 0$  c'est à dire si et seulement si  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ .

**Définition 6.2.4** On dit que les variables  $X$  et  $Y$  sont **non corrélées** si leur covariance est nulle c'est à dire si  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ .

**Théoreme 6.2.4** Soient  $X_1, \dots, X_n$  des variables aléatoires indépendantes à valeurs (respectivement) dans des espaces mesurables  $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ . Alors  $\forall \phi_1 : E_1 \rightarrow \mathbb{R}$  mesurable,  $\dots, \forall \phi_n : E_n \rightarrow \mathbb{R}$  mesurable :

$$\mathbb{E}(\phi_1(X_1) \dots \phi_n(X_n)) = \mathbb{E}(\phi_1(X_1)) \times \dots \times \mathbb{E}(\phi_n(X_n)) .$$

**Corollaire 6.2.5** Deux variables aléatoires indépendantes sont non-corrélées.

**R** La réciproque est fautive en général. 

**Corollaire 6.2.6** Si  $X$  et  $Y$  sont deux variables aléatoires **indépendantes** alors :

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

### 6.3 Cas vectoriel

Pour une variable aléatoire multidimensionnelle  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^n$ , on fait appel à un vecteur espérance mathématique défini comme le vecteur constitué par les différentes espérances mathématiques des variables marginales, on le note  $\mathbb{E}(\mathbf{X})$  ou  $\mathbf{m}_X$  :

$$\mathbb{E}(\mathbf{X}) = \mathbf{m}_X = [\mathbb{E}(X_1), \dots, \mathbb{E}(X_k), \dots, \mathbb{E}(X_n)]^T,$$

car il est commode, comme on le verra plus loin, de le noter sous la forme d'un **vecteur colonne**.

Pour apprécier la dispersion d'une variable aléatoire multidimensionnelle, on pourrait considérer le vecteur  $(\sigma_{X_1}^2, \dots, \sigma_{X_k}^2, \dots, \sigma_{X_n}^2)^T$  mais l'information fournie serait insuffisante car elle ne concernerait que les dispersions des projections de  $X$  sur les axes de coordonnées, axes dont le choix est souvent arbitraire. En réalité, la notion de variance que vous connaissez est la restriction au cas unidimensionnel d'une matrice appelée **matrice de covariance**.

**Définition 6.3.1** La matrice de covariance du vecteur  $\mathbf{X}$  est la matrice carrée notée  $\mathbf{C}_X$  et définie par :

$$\mathbf{C}_X = \mathbb{E}[(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T] = \mathbb{E}(\mathbf{X}\mathbf{X}^T) - \mathbf{m}_X\mathbf{m}_X^T,$$

**Propriété 6.3.1** Le terme  $\mathbf{C}_{i,j}$  de la matrice  $\mathbf{C}_X$  est  $\mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))] = \text{Cov}(X_i, X_j)$ , c'est à dire la covariance des variables marginales  $X_i$  et  $X_j$ .

**R** Comme  $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ , la matrice de covariance est une matrice symétrique réelle.

Sur sa diagonale, on retrouve les  $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$ .

**Propriété 6.3.2** Quel que soit  $\mathbf{a} = (a_1, \dots, a_n)^T$  un vecteur de  $\mathbb{R}^n$ , on a :

$$\mathbf{a}^T \mathbf{C}_X \mathbf{a} = \mathbb{E}[\mathbf{a}^T (\mathbf{X} - \mathbf{m}_X) (\mathbf{X} - \mathbf{m}_X)^T \mathbf{a}] = \mathbb{E} \left[ (\mathbf{a}^T (\mathbf{X} - \mathbf{m}_X))^2 \right] \geq 0.$$

La matrice de covariance est une **matrice semi-définie positive**.

### 6.4 Fonction caractéristique

#### Définition et principales propriétés

Tout d'abord la notion de variable aléatoire réelle s'étend au cas complexe en considérant  $X$  à valeurs dans  $\mathbb{C}$  et non plus seulement dans  $\mathbb{R}$ . Ainsi si  $X$  est une variable aléatoire réelle et  $t$  un paramètre réel,  $e^{itX}$  définit une variable aléatoire *complexe* définie sur  $(\Omega, \mathcal{A})$ .

**Définition 6.4.1 — Fonction caractéristique.** On appelle **fonction caractéristique** de la variable aléatoire  $X$  la fonction  $\rho_X$  de la variable  $t$  définie par l'espérance mathématique de la variable aléatoire  $e^{itX}$ .

$$\rho_X(t) = \mathbb{E}\left(e^{itX}\right) \quad (6.1)$$

— si  $X$  est une variable aléatoire discrète, alors

$$\rho_X(t) = \sum_k e^{itx_k} P(X = x_k) \quad (6.2)$$

— si  $X$  est une variable aléatoire continue ayant pour densité  $f$ , alors

$$\rho_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx \quad (6.3)$$

**R** Dans le cas continu,  $\rho_X$  coïncide avec la transformée de FOURIER (inverse) de la densité  $f$  pour un choix particulier des paramètres. Pour rappel, la transformée de  $f$  est

$$\mathcal{F}(f)(u) = \int_{-\infty}^{+\infty} e^{-iut} f(t) dt \quad (6.4)$$

et elle transforme le produit de convolution en produit  $\mathcal{F}(f * g) = \mathcal{F}(f) \times \mathcal{F}(g)$ .

**Proposition 6.4.1** —  $\rho_X$  est continue sur  $\mathbb{R}$  et bornée par  $\rho_X(0) = 1$ .

— Pour tout couple  $(a, b) \in \mathbb{R}^2$ , on a  $\rho_{aX+b}(t) = e^{ibt} \rho_X(at)$ .

— Si  $X$  et  $Y$  sont indépendantes alors

$$\rho_{X+Y} = \rho_X \rho_Y \quad (6.5)$$

*Démonstration.* La première propriété provient du fait que  $X$  étant une variable aléatoire réelle alors

$$|\mathbb{E}(e^{itX})| \leq \mathbb{E}(|e^{itX}|) = \mathbb{E}(1) = P(\Omega) = 1 = \rho_X(0)$$

Les autres propriétés sont immédiates, en utilisant les résultats connus sur la transformée de FOURIER et celui sur la somme de variables aléatoires indépendantes. ■

### Fonctions caractéristiques et moments

**Proposition 6.4.2** — Si  $\mathbb{E}(X^s) < +\infty$  pour  $s \in \mathbb{N}^*$  alors

$$\rho_X^{(s)}(0) = i^s \mathbb{E}(X^s) \quad (6.6)$$

— En particulier

$$\mathbb{E}(X) = -i\rho_X'(0) \text{ et } \text{Var}(X) = \rho_X'^2(0) - \rho_X''(0) \quad (6.7)$$

*Démonstration.* La condition  $\mathbb{E}(X^s) < +\infty$  permet de dériver  $s$  fois la fonction caractéristique. Par exemple, dans le cas continu, on a

$$\rho_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx$$

D'où

$$\rho_X^{(s)}(t) = \int_{\mathbb{R}} \frac{\partial^s}{\partial t^s} \left( e^{itx} f(x) \right) dx = i^s \int_{\mathbb{R}} x^s \cdot e^{itx} f(x) dx$$

En particulier

$$\rho_X^{(s)}(0) = i^s \int_{\mathbb{R}} x^s \cdot f(x) dx = i^s \mathbb{E}(X^s)$$

La seconde relation découle des définitions. ■

**R** Si la fonction caractéristique est développable en série entière, alors par la formule de TAYLOR

$$\rho_X(t) = \sum_{s=0}^{+\infty} \frac{t^s}{s!} \cdot i^s \mathbb{E}(X^s) \quad (6.8)$$

En particulier

$$\rho_X(t) = 1 + it\mathbb{E}(X) - \frac{t^2}{2} \mathbb{E}(X^2) + t^2 \varepsilon(t) \text{ avec } \varepsilon(t) \xrightarrow[t \rightarrow 0]{} 0 \quad (6.9)$$

Réciproquement, d'après les propriétés de la transformée de FOURIER, on peut montrer que deux variables aléatoires ayant la même fonction caractéristique, ont même loi de probabilité car la transformée de Fourier définie sur l'espace des mesures de probabilité sur  $\mathbb{R}^d$  est injective : on retiendra donc que

La fonction caractéristique détermine la loi de la variable aléatoire  $X$ .

En pratique, pour montrer que deux variables aléatoires ont même loi, on peut tenter de prouver qu'elles ont même fonction caractéristique. C'est ainsi que l'on démontrera le théorème central-limite.

**R** En utilisant les fonctions caractéristiques, on retrouve les résultats suivants.

- La somme de deux lois binomiales  $\mathcal{B}(n, p)$  et  $\mathcal{B}(n', p)$  indépendantes est une loi binomiale  $\mathcal{B}(n + n', p)$ .
- La somme de deux lois de POISSON  $\mathcal{P}(\lambda)$  et  $\mathcal{P}(\lambda')$  indépendantes est une loi de POISSON  $\mathcal{P}(\lambda + \lambda')$ .
- La somme de deux lois normales indépendantes est encore une loi normale.

## Fonctions caractéristiques des lois usuelles

|                       |                            |  |
|-----------------------|----------------------------|--|
| Loi Binomiale         | $\mathcal{B}(n; p)$        | $\rho_X(t) = (q + pe^{it})^n$                        |
| Loi de POISSON        | $\mathcal{P}(\lambda)$     | $\exp(\lambda(e^{it} - 1))$                          |
| Loi géométrique       | $\mathcal{G}(p)$           | $\frac{pe^{it}}{1 - qe^{it}}$                        |
| Loi uniforme discrete | $\mathcal{U}_{[[a, b]]}$   | $\frac{e^{iat}}{b - a + 1} \sum_{k=0}^{b-a} e^{ikt}$ |
| Loi uniforme continue | $\mathcal{U}_{[a, b]}$     | $\frac{e^{itb} - e^{ita}}{it(b - a)}$                |
| Loi exponentielle     | $\mathcal{E}(\lambda)$     | $\left(1 - \frac{it}{\lambda}\right)^{-1}$           |
| Loi normale           | $\mathcal{N}(\mu, \sigma)$ | $\exp\left(\mu it - \frac{\sigma^2 t^2}{2}\right)$   |

Justifions rapidement ces valeurs. Pour la loi binomiale

$$\rho_X(t) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} e^{itk} = \sum_{k=0}^n \binom{n}{k} (e^{it} p)^k q^{n-k} = (q + pe^{it})^n$$

Pour la loi de POISSON,

$$\rho_X(t) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{itk} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(e^{it} \lambda)^k}{k!} = \exp(\lambda(e^{it} - 1))$$

Pour la loi géométrique,

$$\rho_X(t) = \sum_{k=1}^{+\infty} pq^{k-1} e^{itk} = pe^{it} \sum_{k=0}^{+\infty} (e^{it} q)^k = \frac{pe^{it}}{1 - qe^{it}}$$

Pour la loi uniforme continue,

$$\rho_X(t) = \frac{1}{b-a} \int_a^b e^{itx} dx = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

Pour la loi exponentielle,

$$\rho_X(t) = \lambda \int_0^{+\infty} e^{-(\lambda-it)x} dx = \frac{\lambda}{\lambda-it}$$

Terminons par la loi normale centrée réduite. Nous savons que la transformée de FOURIER d'une fonction gaussienne est encore de type gaussien (voir (6.4) pour la définition). Plus précisément,

$$\mathcal{F}(t \mapsto \exp(-\pi t^2))(u) = \exp(-\pi u^2)$$

donc en effectuant un changement de variable, la fonction caractéristique de la loi normale centrée réduite est

$$\rho_X(t) = \exp\left(-\frac{t^2}{2}\right)$$

En utilisant la proposition 6.4.1, on obtient la fonction caractéristique d'une loi normale  $\mathcal{N}(\mu, \sigma)$ .

## 6.5 Inégalités de Jensen, Bienaymé-Tchebichev et Markov

Les trois théorèmes suivants permettent d'encadrer le comportement incertain des variables aléatoires. Elles limitent en effet les domaines de variabilité des variables en fonction des moments qui les caractérisent. Elles sont, en particulier, très utiles en statistique.

**Théorème 6.5.1 — Inégalité de Jensen.** Soit  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  mesurable **convexe**. Soit  $X$  v.a.r. intégrable telle que  $\phi(X)$  est intégrable. Alors :

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)) .$$

La démonstration de ce théorème, est un peu technique mais tout à fait accessible à des étudiants tels que vous.

**Théorème 6.5.2 — Inégalité de Markov.** Soit  $X$  v.a.r. positive, intégrable. Soit  $\lambda > 0$ . Alors :

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

**Corollaire 6.5.3 — Inégalité de Bienaymé-Tchebichev.** Soit  $X$  v.a.r. telle que  $X^2$  est intégrable. Alors :

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

*Démonstration du théorème 6.5.2.* Pour tout  $\omega$ ,  $X(\omega) \geq \lambda \mathbb{1}_{X(\omega) \geq \lambda}$  donc, par la propriété de croissance (cf. propriété 6.1.1),

$$\begin{aligned} \mathbb{E}(X) &\geq \mathbb{E}(\lambda \mathbb{1}_{X \geq \lambda}) \\ &= \lambda P(X \geq \lambda) . \end{aligned}$$

■

*Démonstration du corollaire 6.5.3.*

$$\begin{aligned} P(|X - \mathbb{E}(X)| \geq \lambda) &= P((X - \mathbb{E}(X))^2 \geq \lambda^2) \\ (\text{par inégalité de Bienaymé-Tchebichev}) &\leq \frac{1}{\lambda^2} \mathbb{E}((X - \mathbb{E}(X))^2). \end{aligned}$$

■

**Théorème 6.5.4** — **Inégalité de Markov généralisée.** Si  $X$  v.a.r. avec  $X^2$  intégrable et si  $\lambda > 0$  alors :

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

*Démonstration du théorème 6.5.4.*

$$\begin{aligned} P(|X| \geq \lambda) &= P(X^2 \geq \lambda^2) \\ (\text{par inégalité de Bienaymé-Tchebichev}) &\leq \frac{\mathbb{E}(X^2)}{\lambda^2}. \end{aligned}$$

■

## 6.6 Caractéristiques des variables aléatoires

Comme nous avons eu l'occasion de le dire à plusieurs reprises, la connaissance de la loi  $f_X$  d'une variable aléatoire  $X$  nous donne accès à tous les secrets de celle-ci. Cependant, il est parfois très utile de pouvoir résumer cette information, de la synthétiser en la réduisant à une ou un petit nombre de valeurs numériques caractéristiques. On distinguera ainsi :

- les caractéristiques de **position**, encore appelées caractéristiques de **localisation** qui visent à indiquer un « centre » ou un « repère » autour duquel se positionnent les valeurs prises par la variable aléatoire  $X$  ;
- les caractéristiques de **dispersion** qui donnent, comme leur nom l'indique, un ordre de grandeur de la dispersion des valeurs prises par la variable aléatoire  $X$ , permettant ainsi de caractériser leur concentration ;
- les caractéristiques de **forme** qui permettent de caractériser l'asymétrie et l'aplatissement de la distribution de la loi  $f_X$  de  $X$ .

### 6.6.1 Caractéristiques de position.

Il est important de retenir que les caractéristiques de position sont homogènes à  $X$ .

La plus utilisée des caractéristiques de localisation est l'espérance mathématique, souvent appelée **moyenne**. Pourtant, ce n'est pas nécessairement la plus pertinente et d'autres que nous définissons maintenant sont également très utilisées. La définition de ces caractéristiques permettant de comprendre ce qu'elles caractérisent, nous ne ferons pas de commentaire particulier quand à leur utilisation qui sera illustrée, en TD et, plus tard, en statistique descriptive.

**Définition 6.6.1** Le **mode**,  $mode(X)$ , de la variable aléatoire  $X$  est la valeur pour laquelle la densité  $f_X$  (au sens de Radon-Nikodym) de  $X$  atteint son maximum.

**R**  $mode(X)$  est la valeur la plus souvent « observée » de la variable aléatoire  $X$ .

**Définition 6.6.2** On appelle **quantile** ou **fractile d'ordre**  $\alpha$ , ( $0 \leq \alpha \leq 1$ ) de la variable aléatoire  $X$ , de fonction de répartition  $F_X$ , la valeur  $x_\alpha$  telle que  $F(x_\alpha) = \alpha$ .

Certains quantiles sont particulièrement utilisés :

- $x_0$  caractérise la valeur **minimale** de  $X$  ;
- $x_1$  caractérise la valeur **maximale** de  $X$  ;
- $med(X) = x_{0,5}$  est la **médiane** de  $X$ . Elle sépare en deux parties « égales » les valeurs prise par  $X$ . C'est une caractéristique de **tendance centrale** au même titre que l'espérance. C'est la valeur pour laquelle la fonction de répartition de  $X$  vaut 1/2 :

$$F_X(med(X)) = \frac{1}{2}.$$

**R** Les quartiles  $(x_0, x_{\frac{1}{4}}, x_{0,5}, x_{\frac{3}{4}}, x_1)$ , les déciles  $(x_{\frac{i}{10}})_{0 \leq i \leq 10}$  et les percentiles (ou centiles)  $(x_{\frac{i}{100}})_{0 \leq i \leq 100}$  sont les plus utilisés des quantiles.

### 6.6.2 Caractéristiques de dispersion.

Les moment centrés et non centrés d'ordre  $p > 1$  sont les caractéristiques de dispersion les plus utilisées. La variance  $\text{Var}(X)$  est, en particulier, considérée comme étant la mesure de dispersion « de référence » par la plupart des utilisateurs des statistiques. On interprète en effet  $\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$  comme étant une mesure de la dispersion de  $X$  autour de sa « moyenne »,  $\mathbb{E}(X)$ .

Toutefois, comme la variance est homogène à  $X^2$ , on ne peut pas la comparer directement à  $\mathbb{E}(X)$  (qui rappelons-le, est homogène à  $X$ ). Pour cela, on introduit l'**écart-type**

**Définition 6.6.3** L'**écart-type de**  $X$  est défini par :

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

Un moyen usuel de mesurer la dispersion, relativement à la moyenne est le **coefficient de variation**.

**Définition 6.6.4** Le coefficient de variation  $CV(X)$  de  $X$  est sans unité comme tout bon « coefficient ». C'est le rapport de l'écart-type à la valeur absolue de la moyenne :

$$CV(X) = \frac{\sigma_X}{|\mathbb{E}(X)|}.$$

**R** Le coefficient de variation est parfois exprimé en pourcentages.

- R** Le coefficient de variation permet de comparer la dispersion relative de distributions qui ne sont pas du même ordre de grandeur. Plus  $CV(X)$  est proche de 0, moins la distribution est dispersée, plus il est proche de 1 (pire... si il dépasse 1) plus la distribution est dispersée.

Une autre mesure de la dispersion très utilisée en statistique est basée sur la notion d'**étendue entre fractiles**.

**Définition 6.6.5** On appelle **étendue** la différence  $x_1 - x_0$  entre la valeur maximale et la valeur minimale de  $X$ .

- R** l'étendue caractérise donc la dispersion de l'ensemble des valeurs prises par  $X$ .

**Définition 6.6.6** On appelle **étendue inter-quartiles** la différence  $x_{\frac{3}{4}} - x_{\frac{1}{4}}$  entre la valeur maximale et la valeur minimale de  $X$ .

- R** L'étendue inter-quartiles mesure l'« étalement » de la moitié centrale de la distribution des valeurs de  $X$ .

Pour une variable aléatoire bivariée,  $(X, Y)$  la covariance est une mesure de la dispersion des « points » dans l'espace.

**Définition 6.6.7** Lorsque les deux variables aléatoires sont non constantes (c'est à dire lorsque  $\sigma_X > 0$  et  $\sigma_Y > 0$ ), on normalise souvent la covariance par le produit des écarts-types. On obtient alors ce qu'on appelle le **coefficient de corrélation linéaire entre  $X$  et  $Y$** . On le note :

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Le coefficient de corrélation linéaire présente quelques propriétés importantes :

- $|\rho_{X,Y}| \leq 1$ ;
- Si  $\rho_{X,Y} > 0$  [resp.  $\rho_{X,Y} < 0$ ], on dit que les variables aléatoires  $X$  et  $Y$  sont positivement [resp. négativement] corrélées;
- $\rho_{X,Y} = 0 \Leftrightarrow \text{Cov}(X, Y) = 0$ , les variables sont non corrélées.

Ce coefficient donne une mesure du degré de « dépendance » linéaire entre les variables aléatoires  $X$  et  $Y$ .

### 6.6.3 Caractéristiques de forme.

Deux coefficients de forme dus à Fischer,  $\gamma_1$ , l'asymétrie, et  $\gamma_2$ , l'aplatissement, permettent de caractériser la forme de la densité de la variable  $X$ . Ils s'expriment en fonction des 4 premiers moments de  $X$  (s'ils existent!).

**Définition 6.6.8** Le **coefficient d'asymétrie**,  $\gamma_1$  est une valeur sans dimension définie par :

$$\gamma_1 = \frac{\mu_3(X)}{\sigma_X^3} = \frac{\mathbb{E}((X - \mathbb{E}(X))^3)}{\sigma_X^3}.$$

Il n'est pas affecté par un changement d'origine et d'échelle.

Les informations apportées par le coefficient d'asymétrie également appelé **skewness** sont résumées sur la figure ci-dessous.

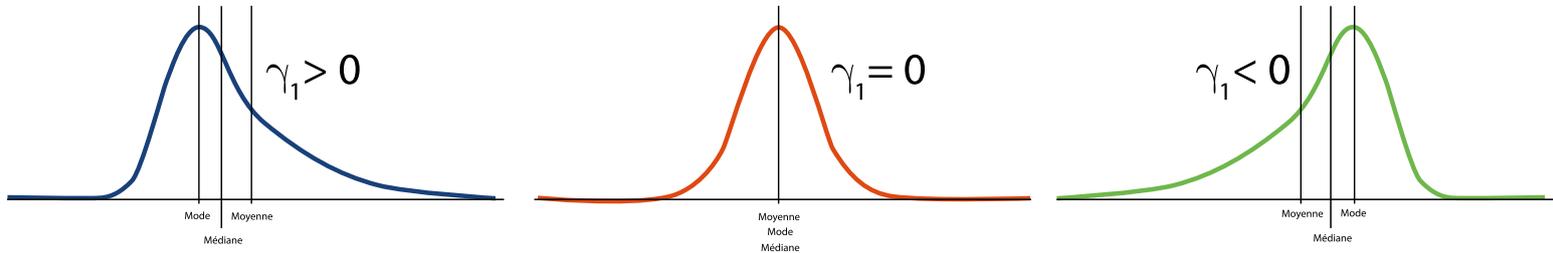


FIGURE 6.1 – Discussion de l'asymétrie suivant le signe de  $\gamma_1$ .

**Définition 6.6.9** Le **coefficient d'aplatissement**,  $\gamma_2$  est une valeur sans dimension qui permet de situer la hauteur de la courbe de la densité  $f_X$  de  $X$  par rapport à une loi de référence, la loi normale. Il est défini par :

$$\gamma_2 = \frac{\mu_4(X)}{\sigma_X^4} - 3.$$

La constante « 3 » permet d'avoir  $\gamma_2 = 0$  pour la loi normale centrée (espérance nulle) et réduite (variance égale à 1).

Les informations apportées par le coefficient d'aplatissement également appelé **kurtosis** sont résumées sur la figure ci-dessous.

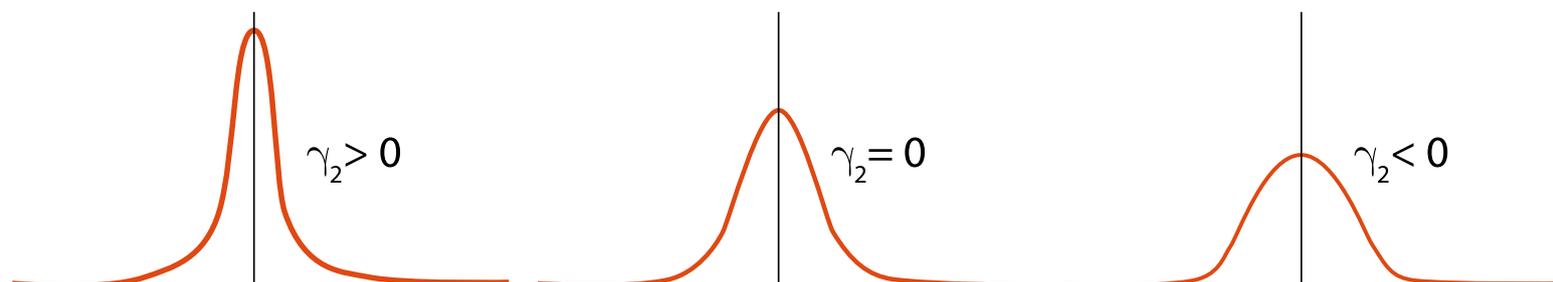


FIGURE 6.2 – Discussion de l'aplatissement suivant le signe de  $\gamma_2$ .



## 7. Loi normale et vecteurs gaussiens

*« Les charmes enchanteurs de cette sublime science ne se décèlent dans toute leur beauté qu'à ceux qui ont le courage de l'approfondir. »*

Karl Friedrich Gauss, *Lettre de Gauss à Sophie Germain du 30 avril 1807.*

### 7.1 La loi normale

Étudiée depuis les années 1770 par De Moivre et Laplace, c'est seulement en 1812 pour Karl Friedrich Gauss qu'elle prit sa forme définitive. On l'appelle tantôt loi de Laplace, tantôt loi de Gauss, tantôt loi de Laplace-Gauss. On trouve aussi la dénomination de loi normale au sens où elle caractérise la distribution d'équilibre de l'ordre, ou plutôt du désordre « naturel » (en l'absence de toute intervention extérieure). Du point de vue mathématique ceci se traduit par le théorème de la limite centrale que nous verrons plus loin. D'un point de vue physique, c'est l'entropie maximale au sens de Shannon qui est modélisée par la loi normale.

Pour faire ressortir toute son importance et sa forme, W.J. Youden, du National Bureau of Standards, a eu l'ingénieuse idée de la présenter telle qu'elle apparaît ci-dessous.

#### 7.1.1 Définition et propriété

**Définition 7.1.1** Une variable aléatoire suit une **loi normale** si sa densité de probabilité est de la forme :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \forall x \in \mathbb{R}.$$

Cette loi que l'on note  $\mathcal{N}(\mu, \sigma)$  ne dépend que des deux paramètres  $\mu$  et  $\sigma^2$ .

THE  
 NORMAL  
 LAW OF ERROR  
 STANDS OUT IN THE  
 EXPERIENCE OF MANKIND  
 AS ONE OF THE BROADEST  
 GENERALIZATIONS OF NATURAL  
 PHILOSOPHY ● IT SERVES AS THE  
 GUIDING INSTRUMENT IN RESEARCHES  
 IN THE PHYSICAL AND SOCIAL SCIENCES AND  
 IN MEDICINE AGRICULTURE AND ENGINEERING ●  
 IT IS AN INDISPENSABLE TOOL FOR THE ANALYSIS AND THE  
 INTERPRETATION OF THE BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT

FIGURE 7.1 – Présentation de l'importance de la loi normale par WJ. Youden.

- R** Pour que  $f$  soit une densité de probabilité il est nécessaire que  $\sigma > 0$ .  
 Lorsque  $\sigma \rightarrow 0$ , on parle de **loi normale dégénérée**. Cette loi se résume à une mesure de Dirac  $\delta_{\mu}(\cdot)$  en  $\mu$ .

**Propriété 7.1.1** Si  $X$  est une v.a. normale  $\mathcal{N}(\mu, \sigma^2)$ , alors :

$$E(X) = \mu, \quad \text{Var}(X) = \sigma^2, \quad \sigma(X) = \sigma.$$

**Exercice 7.1** Démontrer la propriété 7.1.1. ■

### 7.1.2 Caractéristiques de la distribution normale.

La fonction de densité de la loi normale a la forme d'une « courbe en cloche ». En fait il ne s'agit pas d'une courbe unique mais plutôt d'une famille de courbes dépendant de  $\mu$  et  $\sigma^2$ .

- R** Modifier le paramètre  $\mu$  revient à effectuer une translation de la courbe.  
 Plus  $\sigma$  est grand, plus la densité est aplatie.

**Propriété 7.1.2** La densité est symétrique par rapport à la droite d'équation  $x = \mu$ . L'aire sous la courbe de part et d'autre de cette droite est donc égale à  $\frac{1}{2}$ .  $\mu$  est la **médiane** de la densité normale.

**Exercice 7.2** Démontrer la propriété 7.1.2. ■

**Propriété 7.1.3** La densité atteint son maximum pour  $x = \mu$ .  $\mu$  est le **mode** de la densité normale.

**Exercice 7.3** Démontrer la propriété 7.1.3. ■

- R** Pour la loi normale, **l'espérance, le mode et la médiane sont confondus**. C'est l'une des caractéristiques de cette loi et une interprétation que l'on peut en donner est que « l'ordre (ou le désordre) naturel » s'équilibre autour d'une moyenne, d'un mode et d'une médiane confondus.

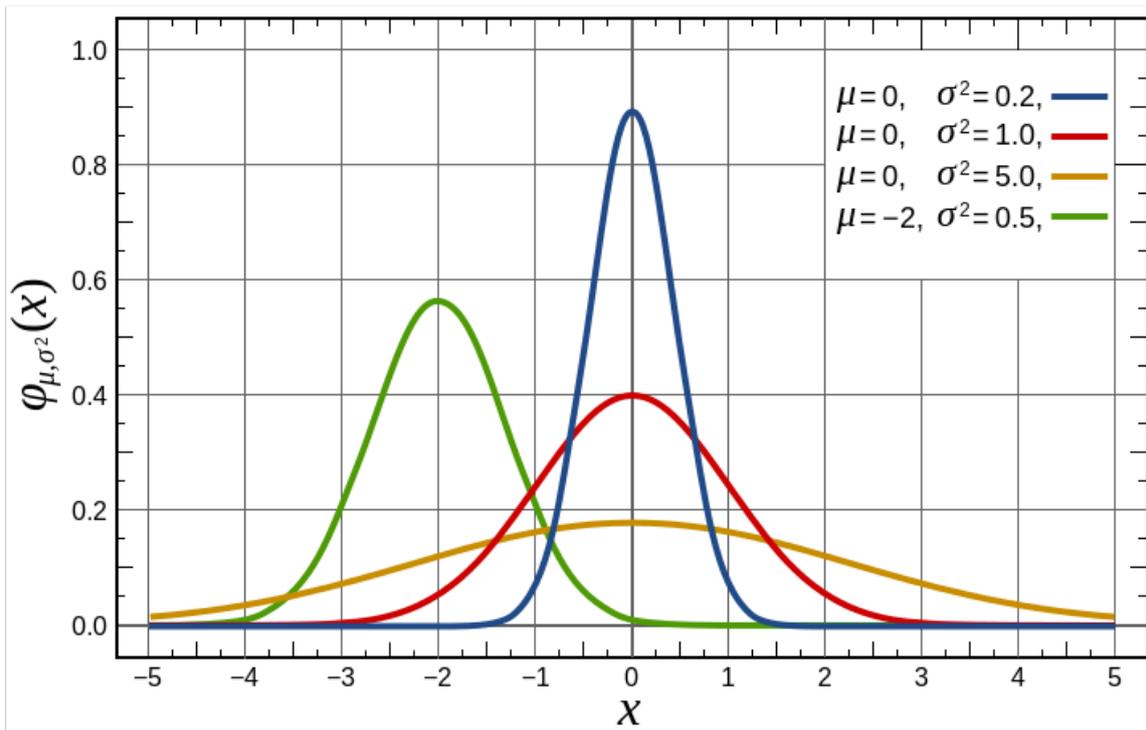


FIGURE 7.2 – Exemple de courbes de la densité normale.

La fonction de répartition de la loi normale est donnée par :

$$\forall x \in \mathbb{R}, \Phi_{\mu, \sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

- R** Il n'existe pas d'expression explicite connue de la f.d.r. de la loi normale. Pour un  $x_0$  donné,  $\Phi_{\mu, \sigma^2}(x_0)$  peut être calculée, avec la précision désirée, par des méthodes numériques.

La représentation graphique de la figure 7.3 met en évidence le fait que la fonction de répartition  $\Phi$  de la loi normale est symétrique par rapport au point  $(\mu, \frac{1}{2})$ , point pour lequel la pente de la tangente à la courbe est la plus élevée.

- R** Modifier le paramètre  $\mu$  revient à effectuer une translation de la courbe. Plus  $\sigma$  est petit, plus la pente de la fonction de répartition autour de la moyenne est marquée.

Comme le montre la figure 7.4, on peut construire des intervalles, centrés autour de  $\mu$ , faisant intervenir l'écart-type et permettant de résumer la distribution des valeurs d'une variable aléatoire suivant une  $\mathcal{N}(\mu, \sigma)$ . Ainsi on a :

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &= 0.6826; \\ P(\mu - 2\sigma < X < \mu + 2\sigma) &= 0.9544; \\ P(\mu - 3\sigma < X < \mu + 3\sigma) &= 0.9974. \end{aligned}$$

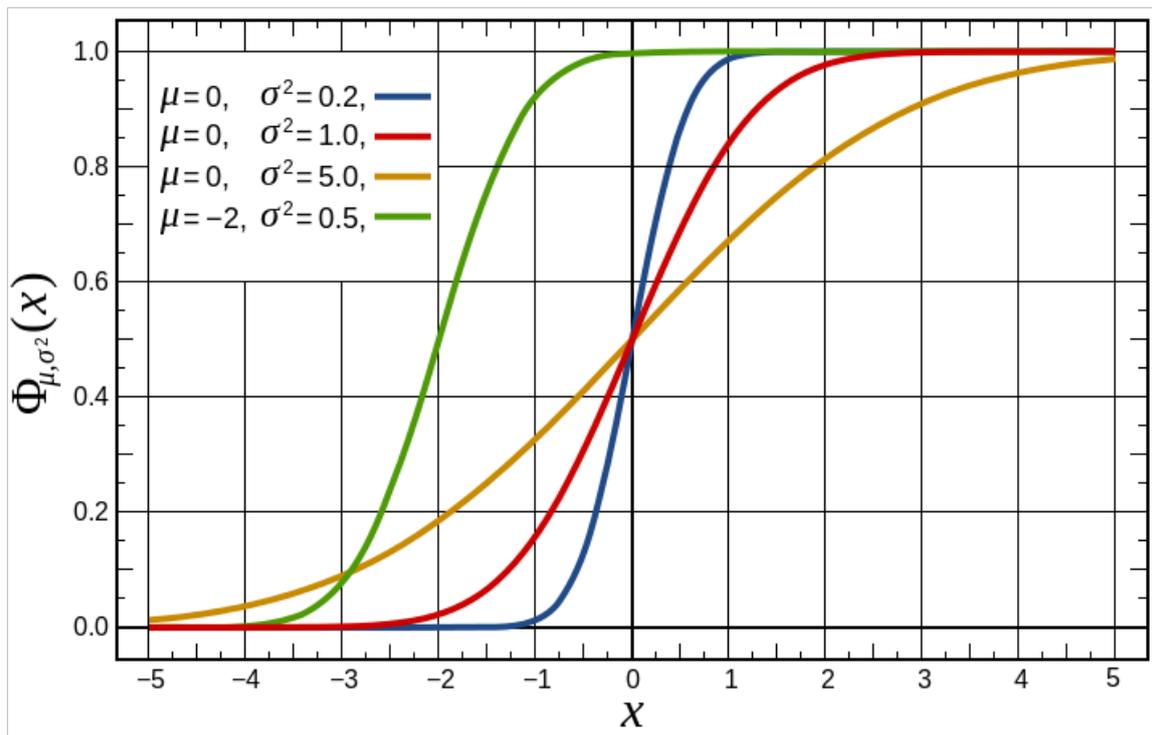


FIGURE 7.3 – Exemple de courbes de fonctions de répartition de la loi normale.

### 7.1.3 Opérations sur les v.a. normales

**Proposition 7.1.4** Considérons une variable aléatoire  $X$  de loi normale  $\mathcal{N}(\mu, \sigma^2)$ . Soient également  $a$  et  $b$  deux réels quelconques. Alors :

$$\mathcal{L}(aX + b) = \mathcal{N}(a\mu + b, a^2\sigma^2).$$

- R** Lorsqu'on effectue le changement de variable affine  $Y = aX + b$  d'une variable aléatoire  $X$ , les propriétés de l'espérance et de la variance permettent d'affirmer que la variable aléatoire  $Y$  aura pour espérance  $\mathbb{E}(Y) = a\mathbb{E}(X) + b$  et pour variance  $\text{Var}(Y) = a^2\text{Var}(X)$ .

**Exercice 7.4** Démontrer la proposition 7.1.4. ■

**Proposition 7.1.5** Considérons deux variables aléatoires  $X_1$  et  $X_2$  de lois normales respectives  $\mathcal{N}(\mu_1, \sigma_1^2)$  et  $\mathcal{N}(\mu_2, \sigma_2^2)$ . Si  $X_1$  et  $X_2$  **sont indépendantes** alors :

$$\mathcal{L}(X_1 + X_2) = \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

- R** Rappelons que lorsqu'on somme deux variables aléatoires indépendantes  $X_1$  et  $X_2$ , les propriétés de l'espérance et de la variance permettent d'affirmer que la variable aléatoire  $X_1 + X_2$  aura pour espérance  $\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)$  et pour variance  $\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)$ .

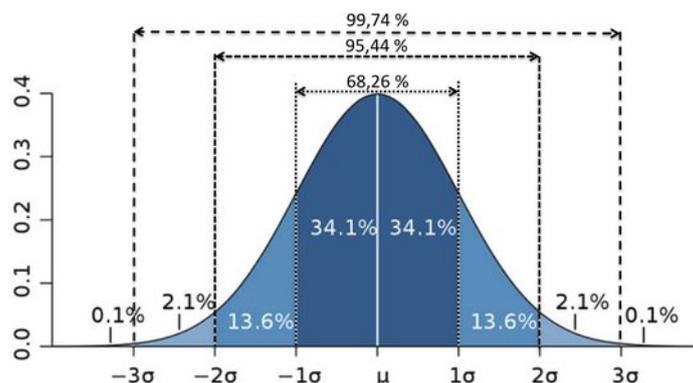


FIGURE 7.4 – Intervalles remarquables de la loi normale.

**Exercice 7.5** Démontrer la proposition 7.1.5. ■

**Exercice 7.6** Trouver un exemple montrant que la somme  $X + Y$  de deux v.a. normales  $X$  et  $Y$  non indépendantes n'est pas nécessairement une v.a. normale. ■

Le fait que l'on préserve la normalité par transformation affine permet d'associer à toute variable aléatoire normale  $X$  sa version standardisée  $Z = \frac{X - E(X)}{\sigma(X)}$  qui est une loi normale d'espérance nulle et de variance unité.

**Pour toute variable aléatoire  $X$  suivant une loi normale  $\mathcal{N}(\mu, \sigma^2)$**  on peut donc se ramener, par centrage et réduction, à une v.a.  $Z = \frac{X - \mu}{\sigma}$  suivant la loi normale standardisée  $\mathcal{N}(0, 1)$ .

Réciproquement si  $Z$  suit une loi normale centrée et réduite  $\mathcal{N}(0, 1)$ , la variable aléatoire  $X = \sigma Z + \mu$  suit une loi normale  $\mathcal{N}(\mu, \sigma^2)$ .

**Propriété 7.1.6** Il suffit de connaître parfaitement la fonction de répartition d'une seule v.a. normale pour connaître celle de toutes les autres. C'est la loi normale standardisée qui joue ce rôle de référence. On retrouve la description de cette loi dans des tables de valeurs du type de celle de la figure 7.5.

## 7.2 Vecteurs gaussiens

La notion de loi normale telle que nous l'avons décrite plus haut nous restreint à des v.a. à valeurs dans  $\mathbb{R}$ . L'extension vectorielle de cette loi amène un certain nombre de (bonnes) surprises que nous nous proposons de développer maintenant.

**Définition 7.2.1** Un vecteur aléatoire  $\mathbf{X} = (X_1, \dots, X_n)^T$  à valeurs dans  $\mathbb{R}^n$  est un vecteur gaussien si toute combinaison linéaire de ses composantes suit une loi normale (avec la convention que la mesure de Dirac,  $\delta_\mu$ , au point  $\mu$  est la « loi normale »  $\mathcal{N}(\mu, 0)$ ).

Autrement dit,  $\mathbf{X}$  est un vecteur gaussien si :

$$\forall a \in \mathbb{R}^n, \text{ le produit scalaire } \langle a, \mathbf{X} \rangle \text{ suit une loi normale.}$$

**R** En particulier, chaque composante  $X_i$  de  $\mathbf{X}$  suit donc une loi normale.



La réciproque est fautive en général!

Il ne suffit pas que toutes les composantes  $X_i$  de  $\mathbf{X}$  suivent des lois normales pour que  $\mathbf{X}$  soit un vecteur gaussien.

### 7.2.1 Transformation affine d'un vecteur gaussien

**Proposition 7.2.1** Considérons un vecteur gaussien  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^n$ . Son espérance est notée  $m_{\mathbf{X}}$  et sa matrice de covariance est notée  $\Sigma_{\mathbf{X}}$ .

L'image :

$$\mathbf{Y} = A\mathbf{X} + b$$

de  $\mathbf{X}$  par une transformation affine où  $A$  est une matrice  $k \times n$  et  $b$  un vecteur à  $k$  lignes est encore un vecteur gaussien d'espérance :

$$m_{\mathbf{Y}} = Am_{\mathbf{X}} + b,$$

et de matrice de covariance :

$$\Sigma_{\mathbf{Y}} = A\Sigma_{\mathbf{X}}A^T.$$

**R** En particulier, pour tout sous-ensemble  $I \subset \{1, \dots, n\}$ , le vecteur extrait  $(X_i)_{i \in I}$  de  $\mathbf{X}$  est donc également un vecteur gaussien.

**Exercice 7.7** Démontrer la proposition 7.2.1. ■

### 7.2.2 Vecteurs à composantes gaussiennes indépendantes.

**Proposition 7.2.2** Soient  $X_1, \dots, X_n$ ,  $n$  variables aléatoires **gaussiennes et indépendantes**. Alors  $\mathbf{X} = (X_1, \dots, X_n)^T$  est un vecteur gaussien.

**Exercice 7.8** Démontrer, par récurrence, la proposition 7.2.2. ■

**R** Il suffit donc que les composantes d'un vecteur soient gaussiennes et indépendantes pour que le vecteur soit un vecteur gaussien.



l'ajout de l'hypothèse d'indépendance est fondamental dans ce résultat!

### 7.2.3 Caractérisation d'un vecteur gaussien.

Nous nous intéressons ici aux conditions d'existence et aux paramètres permettant une caractérisation univoque des vecteurs gaussiens. Pour cela, nous établissons deux théorèmes essentiels.

**Théorème 7.2.3** Considérons un vecteur  $m \in \mathbb{R}^n$  et une matrice  $\Sigma \in \mathbb{R}^{n \times n}$ , **symétrique et semi-définie positive**.

On peut alors construire un vecteur gaussien  $\mathbf{X} = (X_1, \dots, X_n)^T$  d'espérance  $m$  et de

matrice de covariance  $\Sigma$ .

- R** Il est donc possible de trouver au moins un vecteur gaussien pour lequel on se donne un vecteur moyenne et une matrice de covariance. Rien n'assure cependant que ce vecteur soit unique.

**Exercice 7.9** Démontrer le théorème 7.2.2.

Cette démonstration se fait en trois temps. Après avoir justifié la possibilité de décomposer  $\Sigma$  sous la forme  $\Sigma = AA^T$ , montrer qu'il existe un vecteur gaussien  $Z$  à valeurs dans  $\mathbb{R}^n$  dont le vecteur moyenne est  $O_{\mathbb{R}^n}$  et la matrice de covariance est  $Id_{\mathbb{R}^n}$ . Conclure alors la démonstration en passant par une transformation affine bien choisie. ■

le théorème suivant répond à la question que nous avons soulevée dans la remarque précédente puisqu'il permet d'affirmer que la loi d'un vecteur gaussien est entièrement définie par son espérance et sa matrice de covariance.

**Théorème 7.2.4** Soient  $X$  et  $Y$  deux vecteurs gaussiens de même espérance et de même matrice de covariance. Alors  $X$  et  $Y$  ont même loi.

On peut donc définir de manière univoque la loi d'un vecteur gaussien à partir de la seule donnée de son espérance et de sa matrice de covariance. C'est l'extension aux vecteurs aléatoires du résultat que nous connaissons sur les lois normales sur  $\mathbb{R}$ .

**Définition 7.2.2** Soient un vecteur  $m \in \mathbb{R}^n$  et une matrice  $\Sigma \in \mathbb{R}^{n \times n}$ , symétrique et semi-définie positive. On note  $\mathcal{N}(m, \Sigma)$  la loi commune à tous les vecteurs gaussiens admettant  $m$  comme espérance et  $\Sigma$  comme matrice de covariance.

Avant de pouvoir décrire explicitement la forme de la distribution des vecteurs gaussiens, il est important de compléter nos connaissances sur l'effet de l'indépendance sur les vecteurs gaussien **et réciproquement!**

#### 7.2.4 Vecteurs gaussiens et indépendance.

Intéressons-nous pour commencer à la distribution produit de plusieurs vecteurs gaussiens dont on montre qu'elle est encore une distribution gaussienne.

**Proposition 7.2.5** Soient  $d_1, \dots, d_n$ ,  $n$  entiers positifs et notons  $d = \sum_{i=1}^n d_i$ . Soient également  $m_1 \in \mathbb{R}^{d_1}, \dots, m_n \in \mathbb{R}^{d_n}$ ,  $n$  vecteurs et  $\Sigma_1 \in \mathbb{R}^{d_1 \times d_1}, \dots, \Sigma_n \in \mathbb{R}^{d_n \times d_n}$ ,  $n$  matrices symétriques semi-définies positives. Alors :

$$\mathcal{N}(m_1, \Sigma_1) \otimes \dots \otimes \mathcal{N}(m_n, \Sigma_n) = \mathcal{N}(m, \Sigma),$$

avec :

$$m = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} \text{ et } \Sigma = \begin{pmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_n \end{pmatrix}.$$

**Exercice 7.10** Démontrer la proposition 7.2.5 en exploitant le lien entre indépendance et mesure produit. ■

**R** En d'autres termes, on forme des vecteurs gaussiens entièrement définis en rassemblant, dans un même vecteur, des vecteurs gaussiens indépendants.

La proposition 7.2.6 est la réciproque de la proposition 7.2.5.

**Proposition 7.2.6** Considérons un vecteur gaussien  $\mathbf{X} = (X_1, \dots, X_d)^T$  dont la matrice de covariance est diagonale par blocs :

$$\Sigma_{\mathbf{X}} = \begin{pmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_n \end{pmatrix}$$

où  $\Sigma_i \in \mathbb{R}^{d_i \times d_i}$  avec  $\sum_{i=1}^n d_i = d$ .

Alors, si  $\mathbf{Y}_1 = (X_1, \dots, X_{d_1})$ ,  $\mathbf{Y}_2 = (X_{d_1+1}, \dots, X_{d_1+d_2})$ , ...,  $\mathbf{Y}_n = (X_{d_1+\dots+d_{n-1}+1}, \dots, X_d)$ , les vecteurs  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  sont des vecteurs gaussiens et indépendants.

**Exercice 7.11** Démontrer la proposition 7.2.6 en exploitant la proposition 7.2.5. ■

Avec le corollaire suivant, on perçoit peut-être mieux l'intérêt de la proposition 7.2.6 et en particulier le lien entre non-corrélation et indépendance dans le cas des vecteurs gaussiens.

**Corollaire 7.2.7** Si le vecteur gaussien  $\mathbf{X} = (X_1, \dots, X_n)^T$  a une matrice de covariance dont tous les termes non-diagonaux sont nuls, alors  $X_1, \dots, X_n$  sont des variables normales et indépendantes.

**Exercice 7.12** Démontrer le corollaire 7.2.7 de la proposition 7.2.6. ■

Il y a deux résultats importants à retenir dans ce paragraphe :

- La première, c'est que dans le cadre des vecteurs gaussiens, les notions d'**indépendance** et de **non corrélation** sont **équivalentes**;
- La seconde, qui est une extension de la première dans le cadre des vecteurs gaussiens, c'est que l'indépendance mutuelle de  $n$  composantes est équivalente à l'indépendance 2 à 2 de ces composantes.

En conclusion de ce chapitre, nous avons tous les éléments pour donner la forme de la distribution d'un vecteur gaussien qui, comme nous l'avons établi au paragraphe 7.2.3, est entièrement défini par la donnée de son vecteur moyenne et de sa matrice de covariance.

### 7.2.5 Densité d'un vecteur gaussien.

Le résultat suivant, dont la restriction aux v.a. scalaires est la loi normale sur  $\mathbb{R}$  est la conséquence de tous les résultats établis jusqu'ici.

**Théorème 7.2.8** Soit un vecteur  $m \in \mathbb{R}^n$  et une matrice symétrique définie positive  $\Sigma \in \mathbb{R}^{n \times n}$ . La loi  $\mathcal{N}(m, \Sigma)$  sur  $\mathbb{R}^n$  admet comme densité par rapport à la mesure de Lebesgue :

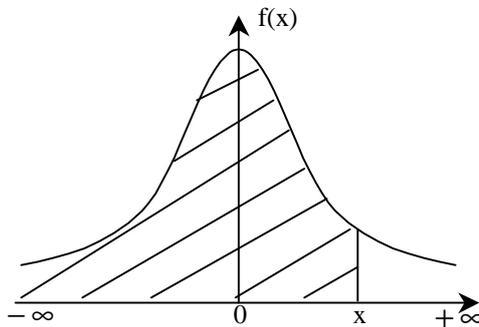
$$f_{m, \Sigma}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2} (\mathbf{x} - m)^T \Sigma^{-1} (\mathbf{x} - m)\right).$$

**Exercice 7.13** Démontrer le théorème 7.2.8.

Cette démonstration fait méthodiquement intervenir les connaissances vues dans ce chapitre. A ce titre, elle est très formatrice. ■

## Loi Normale centrée réduite

Probabilité de trouver une valeur inférieure à  $x$ .



$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2p}} e^{-\frac{u^2}{2}} du$$

| X          | 0,00   | 0,01   | 0,02   | 0,03   | 0,04   | 0,05   | 0,06   | 0,07   | 0,08   | 0,09   |
|------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| <b>0,0</b> | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 | 0,5359 |
| <b>0,1</b> | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 | 0,5753 |
| <b>0,2</b> | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 | 0,6141 |
| <b>0,3</b> | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 | 0,6517 |
| <b>0,4</b> | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 | 0,6879 |
| <b>0,5</b> | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 | 0,7224 |
| <b>0,6</b> | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 | 0,7549 |
| <b>0,7</b> | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 | 0,7852 |
| <b>0,8</b> | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 | 0,8133 |
| <b>0,9</b> | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8264 | 0,8289 | 0,8315 | 0,8340 | 0,8365 | 0,8389 |
| <b>1,0</b> | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 | 0,8621 |
| <b>1,1</b> | 0,8643 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 | 0,8830 |
| <b>1,2</b> | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 | 0,9015 |
| <b>1,3</b> | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 | 0,9177 |
| <b>1,4</b> | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 | 0,9319 |
| <b>1,5</b> | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 | 0,9441 |
| <b>1,6</b> | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 | 0,9545 |
| <b>1,7</b> | 0,9554 | 0,9564 | 0,9573 | 0,9582 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 | 0,9633 |
| <b>1,8</b> | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 | 0,9706 |
| <b>1,9</b> | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 | 0,9767 |
| <b>2,0</b> | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 | 0,9817 |
| <b>2,1</b> | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 | 0,9857 |
| <b>2,2</b> | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 | 0,9890 |
| <b>2,3</b> | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 | 0,9916 |
| <b>2,4</b> | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 | 0,9936 |
| <b>2,5</b> | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 | 0,9952 |
| <b>2,6</b> | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 | 0,9964 |
| <b>2,7</b> | 0,9965 | 0,9966 | 0,9967 | 0,9968 | 0,9969 | 0,9970 | 0,9971 | 0,9972 | 0,9973 | 0,9974 |
| <b>2,8</b> | 0,9974 | 0,9975 | 0,9976 | 0,9977 | 0,9977 | 0,9978 | 0,9979 | 0,9979 | 0,9980 | 0,9981 |
| <b>2,9</b> | 0,9981 | 0,9982 | 0,9982 | 0,9983 | 0,9984 | 0,9984 | 0,9985 | 0,9985 | 0,9986 | 0,9986 |
| <b>3,0</b> | 0,9987 | 0,9987 | 0,9987 | 0,9988 | 0,9988 | 0,9989 | 0,9989 | 0,9989 | 0,9990 | 0,9990 |
| <b>3,1</b> | 0,9990 | 0,9991 | 0,9991 | 0,9991 | 0,9992 | 0,9992 | 0,9992 | 0,9992 | 0,9993 | 0,9993 |
| <b>3,2</b> | 0,9993 | 0,9993 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9995 | 0,9995 | 0,9995 |
| <b>3,3</b> | 0,9995 | 0,9995 | 0,9995 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9997 |
| <b>3,4</b> | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9998 |
| <b>3,5</b> | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 |

Table pour les grandes valeurs de  $x$  :

| x    | 3          | 3,2        | 3,4        | 3,6        | 3,8        | 4          | 4,2        | 4,4        | 4,6        | 4,8        |
|------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| F(x) | 0,99865003 | 0,99931280 | 0,99966302 | 0,99984085 | 0,99992763 | 0,99996831 | 0,99998665 | 0,99999458 | 0,99999789 | 0,99999921 |

FIGURE 7.5 – table des valeurs de la fonction de répartition de la loi normale standardisée.



## 8. Espérance conditionnelle

« - *À quoi tu penses ?*

- *Je pense que si en ouvrant un dictionnaire au hasard, on tombait sur le mot hasard, ce serait un miracle, alors que si on tombait sur le mot miracle, ce serait un hasard. »*

H. Le Tellier, *Les amnésiques n'ont rien vécu d'inoubliable.*

Comme nous l'avons vu au chapitre 2, l'idée du conditionnement est de réévaluer les probabilités d'évènements en tenant compte d'une information auxiliaire permettant de compléter la connaissance de l'expérience aléatoire. Dans de nombreux problèmes de statistiques (prédiction, observation incomplète, etc...) il est nécessaire de pouvoir estimer une variable aléatoire sur laquelle on n'a qu'une information partielle que l'on doit compléter grâce à la connaissance d'une autre variable aléatoire à laquelle elle est liée. La notion d'espérance conditionnelle joue un rôle crucial dans ce cas.

### 8.1 Conditionnement discret

On se donne toujours un espace probabilisé  $(\Omega, \mathcal{A}, P)$ .

**Définition 8.1.1** Soient  $A, B \in \mathcal{A}$  avec  $P(B) > 0$ , la probabilité de  $B$  sachant  $A$  est :

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

$P(\cdot|B)$  est une mesure de probabilité sur l'espace probabilisé  $(\Omega, \mathcal{A})$ .

**Définition 8.1.2** Si  $X$  est une v.a. et  $B \in \mathcal{A}$  avec  $P(B) > 0$ , l'espérance de  $X$  sachant  $B$  est l'espérance par rapport à la probabilité conditionnelle  $P(\cdot|B)$ .

$$\mathbb{E}(X|B) = \int_{\Omega} X dP(\cdot|B) = \frac{1}{P(B)} \int_B X dP = \frac{\mathbb{E}(X\mathbb{1}_B)}{P(B)}.$$

On peut interpréter cette espérance comme la « moyenne » de  $X$  lorsque  $B$  est réalisé. Au stade où nous en sommes, nous disposons de tous les éléments permettant d'étendre cette notion d'espérance conditionnelle à un conditionnement par rapport à une variable aléatoire. Pour cela, considérons une v.a.r.  $X$  et une v.a.  $Y$  définie sur un espace  $E$  dénombrable. On note  $E'$  le sous-ensemble de  $E$  tel que  $E' = \{y \in E, P(Y = y) > 0\}$ . Pour tout  $y \in E'$  on définit l'évènement :

$$\{Y = y\} = \{\omega \in \Omega, Y(\omega) = y\},$$

et on peut définir l'espérance conditionnellement à cet évènement :

$$\mathbb{E}(X|Y = y) = \frac{\mathbb{E}(X\mathbb{1}_{Y=y})}{P(Y = y)}.$$

Ceci nous amène alors à définir l'espérance conditionnellement à la variable  $Y$ .

**Définition 8.1.3** On définit l'espérance conditionnelle de  $X$  sachant  $Y$  de la manière suivante :  $\mathbb{E}(X|Y)$  est une **variable aléatoire** qui peut s'écrire  $\mathbb{E}(X|Y) = \phi(Y)$  avec :

$$\begin{aligned} \phi : E &\rightarrow \mathbb{R} \\ y &\mapsto \begin{cases} \mathbb{E}(X|Y = y) = \frac{\mathbb{E}(X\mathbb{1}_{Y=y})}{P(Y=y)} & \text{si } y \in E' \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

**R** Le choix de la valeur de  $\phi$  sur  $E \setminus E'$  est arbitraire. En fait, ce choix n'influence la définition de  $\mathbb{E}(X|Y = y)$  que sur un ensemble de probabilité nulle, puisque :

$$P(Y \in E \setminus E') = \sum_{y \in E \setminus E'} P(Y = y) = 0$$

On pourrait changer la définition de  $\phi$  sur  $E \setminus E'$  et cela donnerait la même v.a.  $\mathbb{E}(X|Y)$  à un ensemble de mesure nulle près. Dans les situations plus générales que nous rencontrerons plus tard, les espérances conditionnelles (sachant une v.a. ou une tribu) seront toujours définies à un ensemble de probabilité nulle près.

Il est important de noter que l'espérance conditionnelle est une variable aléatoire. Plus précisément,  $\mathbb{E}(X|Y)$  est la moyenne de  $X$  lorsqu'on connaît  $Y$  p.s.

On a :

$$\forall \omega \in \Omega \quad \mathbb{E}(X|Y)(\omega) = \mathbb{E}(X|Y(\omega)) = \mathbb{E}(X|Y = y) \text{ si } Y(\omega) = y.$$

Cette nouvelle variable aléatoire est intégrable. En effet,

$$\mathbb{E}[|\mathbb{E}(X|Y)|] = \sum_{y \in E} |\mathbb{E}(X|Y = y)| P(Y = y) = \sum_{y \in E} \left| \frac{\mathbb{E}(X\mathbb{1}_{Y=y})}{P(Y = y)} \right| P(Y = y).$$

On a donc :

$$\mathbb{E}[|\mathbb{E}(X|Y)|] = \sum_{y \in E} |\mathbb{E}(X \mathbb{1}_{Y=y})| \leq \sum_{y \in E} \mathbb{E}(|X| \mathbb{1}_{Y=y}) = \mathbb{E}(|X|).$$

Considérons alors une variable aléatoire  $Z$ ,  $\sigma(Y)$ -mesurable et bornée. La tribu  $\sigma(Y)$  représentant les événements relatifs à  $Y$  (tous ceux qui peuvent se décrire en terme de « il est arrivé telle chose à  $Y$  »), dire que  $Z$  est  $Y$ -mesurable revient à dire que tous les événements relatifs à  $Z$  peuvent se décrire comme des événements relatifs à  $Y$ . Cette variable  $Z$  peut donc s'écrire sous la forme  $Z = \psi(Y)$  et :

$$\mathbb{E}[Z \mathbb{E}(X|Y)] = \sum_{y \in E} \psi(y) \mathbb{E}(X \mathbb{1}_{Y=y}) = \sum_{y \in E} \mathbb{E}(X \psi(Y) \mathbb{1}_{Y=y}),$$

ce qui conduit à :

$$\mathbb{E}[Z \mathbb{E}(X|Y)] = \mathbb{E}(X \psi(Y)) = \mathbb{E}(ZX).$$

On peut alors montrer que s'il existe une variable aléatoire discrète  $Y'$  telle que  $\sigma(Y) = \sigma(Y')$ , on a :

$$\mathbb{E}(X|Y) = \mathbb{E}(X|Y') \text{ p.s.,}$$

ce qui met en évidence que le conditionnement « naturel » est celui par rapport à une tribu.

## 8.2 Espérance conditionnelle

Le résultat/définition suivant offre deux propriétés caractéristiques de l'espérance conditionnelle.

**Théoreme 8.2.1 — et définition.** Soit  $\mathcal{B}$  une sous-tribu de  $\mathcal{A}$ . Soit  $X$  une v.a.r. intégrable sur  $(\Omega, \mathcal{A}, P)$ . Il existe une et une seule v.a.r. intégrable sur  $(\Omega, \mathcal{B}, P)$ , appelée **espérance conditionnelle de  $X$  sachant  $\mathcal{B}$**  et notée  $\mathbb{E}(X|\mathcal{B})$ , qui vérifie :

$$\forall B \in \mathcal{B}, \mathbb{E}(X \mathbb{1}_B) = \mathbb{E}(\mathbb{E}(X|\mathcal{B}) \mathbb{1}_B).$$

La variable  $\mathbb{E}(X|\mathcal{B})$  vérifie en outre que  $\forall Z$  v.a. à valeurs dans  $\mathbb{R}^d$ ,  $\mathcal{B}$ -mesurable et bornée :

$$\mathbb{E}(XZ) = \mathbb{E}(\mathbb{E}(X|\mathcal{B})Z).$$

**R** Dans le cas particulier où  $\mathcal{B} = \sigma(Y)$ ,  $Y$  étant une v.a. quelconque, l'espérance conditionnelle sera notée indifféremment :

$$\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X|\sigma(Y)) = \mathbb{E}(X|Y),$$

ce qui est cohérent avec le cas discret que nous avons étudié précédemment.

L'espérance conditionnelle admet un ensemble de propriétés, très utiles pour les calculs, que nous regroupons dans la proposition suivante.

**Proposition 8.2.2** Soit  $X, Y$  des v.a.r. et  $\mathcal{B}$  tribu  $\subset \mathcal{A}$ ,

- si  $X$  est  $\mathcal{B}$ -mesurable alors  $\mathbb{E}(X|\mathcal{B}) = X$ ;
- **linéarité** :  $\forall a, b \in \mathbb{R}, \mathbb{E}(aX + bY|\mathcal{B}) = a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B})$ ;
- $\mathbb{E}(\mathbb{E}(X|\mathcal{B})) = \mathbb{E}(X)$ ;
- $|\mathbb{E}(X|\mathcal{B})| \leq \mathbb{E}(|X||\mathcal{B})$ ;
- **croissance** :  $X \geq Y \Rightarrow \mathbb{E}(X|\mathcal{B}) \geq \mathbb{E}(Y|\mathcal{B})$ , p.s.;
- si  $X \perp Y$ ,  $\mathbb{E}(XY|\mathcal{B}) = \mathbb{E}(X|\mathcal{B})\mathbb{E}(Y|\mathcal{B})$ ;
- si  $X \perp Y$ ,  $\mathbb{E}(X|\sigma(Y)) = \mathbb{E}(X)$ .

**Exercice 8.1** Démontrer la proposition 8.2.2. C'est un exercice essentiel pour la compréhension de l'espérance conditionnelle. ■

**Proposition 8.2.3** • Si  $X, Y$  v.a.r. avec  $Y$   $\mathcal{B}$ -mesurable ( $\mathcal{B}$  tribu  $\subset \mathcal{A}$ ), alors :

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

- Si  $\mathcal{B}_1, \mathcal{B}_2$  tribus  $\subset \mathcal{A}$  avec  $\mathcal{B}_1 \subset \mathcal{B}_2$ , alors pour toute v.a.r.  $X$  :

$$\mathbb{E}(\mathbb{E}(X|\mathcal{B}_2)|\mathcal{B}_1) = \mathbb{E}(X|\mathcal{B}_1).$$

**Exercice 8.2** Démontrer la proposition 8.2.3 grâce aux résultats de la proposition 8.2.2. ■

**R** On peut, dans le cas des variables aléatoires de carré intégrable, interpréter l'espérance conditionnelle d'une variable aléatoire  $X$  comme la **projection orthogonale** de  $X$  sur l'espace vectoriel des variables aléatoires  $\mathcal{B}$ -mesurables, et, partant de là, comme la meilleure approximation qu'on puisse donner de la variable  $X$  à l'aide d'une variable aléatoire  $\mathcal{B}$ -mesurable. En effet, l'espérance conditionnelle possède la propriété suivante : pour toute variable aléatoire  $Y$  intégrable  $\mathcal{G}$ -mesurable,

$$\mathbb{E}[(X - Y)^2] \geq \mathbb{E}[(X - \mathbb{E}[X|\mathcal{B}])^2].$$

C'est-à-dire que, parmi les variables aléatoires  $Y$  intégrables  $\mathcal{B}$ -mesurables, la plus proche de  $X$  (pour la distance induite par le produit scalaire  $(X, Y) \mapsto \mathbb{E}[XY]$ ) est  $Y_0 = \mathbb{E}[X|\mathcal{B}]$ .

### 8.3 Loi conditionnelle

Commençons par le cas discret que nous connaissons bien. Considérons donc une v.a.  $Y$  à valeurs dans un espace dénombrable  $E$  et soit  $X$  une variable aléatoire intégrable.

On a déjà vu qu'alors :

$$\mathbb{E}(X|Y) = \phi(Y),$$

avec :

$$\phi(Y) = \frac{\mathbb{E}(X\mathbb{1}_{Y=y})}{P(Y=y)},$$

pour tout  $y \in E$  tel que  $P(Y = y) > 0$  ( $\phi(y)$  pouvant être choisie arbitrairement lorsque  $P(Y = y) = 0$ ).

Comme nous en avons désormais l'habitude, l'étude du cas « continue » est plus riche que celle du cas discret.

Considérons donc deux v.a.  $X$  et  $Y$  à valeurs respectivement dans  $\mathbb{R}^m$  et  $\mathbb{R}^n$  et supposons que le couple a pour densité  $f_{X,Y}(x, y)$ . Notons enfin  $f_Y(y) = \int_{\mathbb{R}^m} f_{X,Y}(x, y) dx$ , la densité marginale de  $Y$ .

Alors, pour toute fonction borélienne  $\psi : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^+$  on a, par le théorème de transfert :

$$\mathbb{E}(\psi(X, Y)) = \int_{\mathbb{R}^m \times \mathbb{R}^n} \psi(x, y) f_{X,Y}(x, y) dx dy.$$

Intéressons-nous alors à l'espérance conditionnelle  $\mathbb{E}(h(X) | Y)$  où  $h$  est une fonction mesurable  $h : \mathbb{R}^m \rightarrow \mathbb{R}^+$ . Pour le calcul de cette espérance conditionnelle, passons, pour toute fonction borélienne  $g : \mathbb{R}^n \rightarrow \mathbb{R}^+$ , par le calcul de l'espérance  $\mathbb{E}(h(X)g(Y))$ . D'après le théorème de transfert on a donc :

$$\mathbb{E}(h(X)g(Y)) = \int_{\mathbb{R}^m \times \mathbb{R}^n} h(x)g(y) f_{X,Y}(x, y) dx dy,$$

que l'on réécrit :

$$\mathbb{E}(h(X)g(Y)) = \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^m} h(x) f_{X,Y}(x, y) dx \right) g(y) dy.$$

Posons alors :

$$\forall y \in \mathbb{R}^n, \phi(y) = \frac{1}{f_Y(y)} \int_{\mathbb{R}^m} h(x) f_{X,Y}(x, y) dx \mathbb{1}_{\{f_Y(y) > 0\}} + h(0) \mathbb{1}_{\{f_Y(y) = 0\}},$$

pour obtenir, pour toute fonction borélienne  $g : \mathbb{R}^n \rightarrow \mathbb{R}^+$  :

$$\mathbb{E}(h(X)g(Y)) = \int_{\mathbb{R}^n} \phi(y) g(y) f_Y(y) dy,$$

ou encore, par le théorème de transfert sur la loi de  $Y$  :

$$\mathbb{E}(h(X)g(Y)) = \mathbb{E}(\phi(Y)g(Y)). \quad (8.1)$$

D'après le théorème 8.2.1,  $\mathbb{E}(X | \mathcal{B})$  est une espérance conditionnelle si et seulement si pour toute v.a.  $Z$ ,  $\mathcal{B}$ -mesurable et bornée :

$$\mathbb{E}(XZ) = \mathbb{E}(\mathbb{E}(X | \mathcal{B})Z).$$

D'après (8.1), on a donc :

$$\mathbb{E}(h(X) | Y) = \phi(Y).$$

Ce résultat, anodin au premier abord, constitue le fondement de la définition de la **densité conditionnelle** :

**Proposition 8.3.1** Pour tout  $y \in \mathbb{R}^n$ , soit  $f_{X|Y=y}$  la densité de probabilité sur  $\mathbb{R}^m$  définie par

$$f_{X|Y=y}(x) = \frac{1}{f_Y(y)} f_{X,Y}(x, y) \mathbb{1}_{\{f_Y(y) > 0\}} + \delta_0(x) \mathbb{1}_{\{f_Y(y) = 0\}}.$$

Alors pour toute fonction mesurable  $h : \mathbb{R}^m \rightarrow \mathbb{R}^+$  :

$$\mathbb{E}(h(X) | Y = y) = \int h(x) f_{X|Y=y}(x) dx = \frac{1}{f_Y(y)} \int h(x) f_{X,Y}(x, y) dx.$$

La fonction  $f_{X|Y=y}$  est appelée **densité conditionnelle** de  $X$  sachant  $Y = y$ .

## 8.4 Le cas gaussien

Comme le montre la proposition suivante, dans le cas des vecteurs gaussiens, le conditionnement préserve la gaussianité.

**Proposition 8.4.1** Soit  $(Y_1, \dots, Y_n, X)$  un vecteur gaussien centré (c'est à dire que  $\mathbb{E}(Y_1) = \dots = \mathbb{E}(Y_n) = \mathbb{E}(X) = 0$ ). Alors,  $\exists \lambda_1, \dots, \lambda_n \in \mathbb{R}$  tels que

$$\mathbb{E}(X | \sigma(Y_1, \dots, Y_n)) = \sum_{j=1}^n \lambda_j Y_j.$$

De plus, pour toute fonction mesurable  $h : \mathbb{R} \rightarrow \mathbb{R}^+$ ,

$$\mathbb{E}(h(X) | \sigma(Y_1, \dots, Y_n)) = \int_{\mathbb{R}} h(x) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx$$

avec

$$\sigma^2 = \mathbb{E}\left(\left(X - \sum_{j=1}^n \lambda_j Y_j\right)^2\right), \quad m = \sum_{j=1}^n \lambda_j y_j.$$

On peut aisément calculer les  $\lambda_i$  apparaissant dans la proposition 8.4.1. Pour cela, notons  $Z = \mathbb{E}(X | \sigma(Y_1, \dots, Y_n))$ . Nous avons alors,  $\forall i \in \{1, \dots, n\}$  :

$$\mathbb{E}(Z Y_i) = \mathbb{E}(X Y_i) = \text{Cov}(X, Y_i).$$

Par ailleurs, on a :

$$\begin{aligned} \mathbb{E}(Z Y_i) &= \mathbb{E}\left(\sum_{k=1}^n \lambda_k Y_i Y_k\right) \\ &= \sum_{k=1}^n \lambda_k \text{Cov}(Y_i, Y_k). \end{aligned}$$

Donc :

$$\begin{aligned} \Sigma_Y \times \begin{bmatrix} \lambda_1 \\ \dots \\ \lambda_n \end{bmatrix} &= \begin{bmatrix} \text{Cov}(X, Y_1) \\ \dots \\ \text{Cov}(X, Y_n) \end{bmatrix} \\ \begin{bmatrix} \lambda_1 \\ \dots \\ \lambda_n \end{bmatrix} &= \Sigma_Y^{-1} \times \begin{bmatrix} \text{Cov}(X, Y_1) \\ \dots \\ \text{Cov}(X, Y_n) \end{bmatrix}. \end{aligned}$$

En particulier, dans le cas d'un vecteur gaussien centré  $(X, Y)$ , on a donc  $\mathbb{E}(X | Y) = \lambda Y$  avec  $\lambda = \frac{1}{\sigma_Y^2} \text{Cov}(X, Y)$  ou encore  $\lambda = \rho_{X,Y} \frac{\sigma_X}{\sigma_Y}$ .

La densité conditionnelle  $f_{X|Y=y}$  suit une loi normale de moyenne :

$$m = \rho_{X,Y} \frac{\sigma_X}{\sigma_Y} y,$$

et de variance :

$$\begin{aligned} \sigma^2 &= \mathbb{E} \left( \left( X - \rho_{X,Y} \frac{\sigma_X}{\sigma_Y} Y \right)^2 \right) = \mathbb{E}(X^2) + \rho_{X,Y}^2 \frac{\sigma_X^2}{\sigma_Y^2} \mathbb{E}(Y^2) - 2\rho_{X,Y} \frac{\sigma_X}{\sigma_Y} \mathbb{E}(XY) \\ &= \sigma_X^2 + \rho_{X,Y}^2 \frac{\sigma_X^2}{\sigma_Y^2} \sigma_Y^2 - 2\rho_{X,Y} \frac{\sigma_X}{\sigma_Y} \text{Cov}(X, Y) \\ &= (1 - \rho_{X,Y}^2) \sigma_X^2. \end{aligned}$$

La densité conditionnelle  $f_{X|Y=y}$  suit donc une loi normale :

$$\mathcal{N} \left( \rho_{X,Y} \frac{\sigma_X}{\sigma_Y} y; (1 - \rho_{X,Y}^2) \sigma_X^2 \right).$$

Si  $(X, Y)$  n'est pas centré, la densité conditionnelle  $f_{X|Y=y}$  suit une loi normale :

$$\mathcal{N} \left( m_X + \rho_{X,Y} \frac{\sigma_X}{\sigma_Y} (y - m_Y); (1 - \rho_{X,Y}^2) \sigma_X^2 \right).$$





## 9. Convergence de suites de variables aléatoires

« *Le hasard aime à servir celui qui suit toujours une même pensée.* »

Jules Michelet, *Le Peuple*.

### 9.1 Les différentes notions de convergence

Dans la mesure où les variables aléatoires sont des fonctions mesurables, s'intéresser à la convergence d'une suite de variables aléatoires c'est s'intéresser à la convergence d'une suite de fonctions. Cependant, la notion d'égalité en probabilité amène à introduire quelques subtilités supplémentaires qui, je vous l'assure, ne sont pas dépourvues de charme!

Pour toute la suite du chapitre, on se donne une v.a.  $X$  ainsi qu'une suite de v.a.  $(X_n)_{n \geq 0}$  à valeurs dans  $\mathbb{R}^d$ .

**Définition 9.1.1** Soit  $p > 0$ , on dit que  $X_n$  **converge dans  $L^p$**  vers  $X$  et on note

$X_n \xrightarrow[n \rightarrow +\infty]{L^p} X$  si  $\mathbb{E}(\|X - X_n\|^p) \xrightarrow[n \rightarrow +\infty]{} 0$  (ici,  $\|\cdot\|$  est la norme usuelle sur  $\mathbb{R}^d$ ).

On dit également que  $X_n$  converge vers  $X$  **en moyenne d'ordre  $p$** .

**R** Cette convergence est une convergence classique par rapport à la norme d'ordre  $p$ . La particularité des probabilités est liée au choix de la mesure par rapport à laquelle l'intégrale (l'espérance) est calculée.

**Proposition 9.1.1** Soient  $p$  et  $q$ , tels que  $p \geq q$  alors :

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

**Exercice 9.1** Grâce à une utilisation bien choisie de l'inégalité de Hölder, démontrer la proposition 9.1.1. ■

**Définition 9.1.2** On dit que  $X_n$  **converge presque sûrement** vers  $X$  et on note  $X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}}$   $X$  si :

$$P(\{\omega \in \Omega : X(\omega) = \lim_{n \rightarrow +\infty} X_n(\omega)\}) = 1 .$$

**R** La convergence presque sûre est une convergence simple (au sens que vous connaissez) sauf éventuellement sur un ensemble de mesure nulle. A ce titre, elle est « plus faible » que la convergence simple.

La définition suivante est propre à la convergence des suites de variables aléatoires. Contrairement à la convergence presque sûre, c'est sur la mesure de probabilité que porte la limite et donc l'idée de convergence.

**Définition 9.1.3** On dit que  $X_n$  **converge en probabilité** vers  $X$  et on note  $X_n \xrightarrow[n \rightarrow +\infty]{P}$   $X$  si :

$$\forall \epsilon > 0, P(\|X - X_n\| \geq \epsilon) \xrightarrow[n \rightarrow +\infty]{} 0.$$

**R** La convergence en probabilité assure que la probabilité que  $X_n$  reste éloignée de  $X$  tend vers 0 quand  $n \rightarrow +\infty$ .

**Proposition 9.1.2** La convergence de  $X_n$  vers  $X$  en moyenne d'ordre  $p$  assure la convergence en probabilité de  $X_n$  vers  $X$ .

**Exercice 9.2** Grâce à l'inégalité de Markov, démontrer la proposition 9.1.2. ■

**R** La proposition 9.1.2 n'a pas de réciproque.

**Proposition 9.1.3** La convergence presque sûre de  $X_n$  vers  $X$  assure la convergence en probabilité de  $X_n$  vers  $X$ .

**Exercice 9.3** Démontrer la proposition 9.1.3.

Pour cela on introduit, pour  $\epsilon > 0$  fixé, la suite  $f_n = \mathbb{1}_{\{|X - X_n| > \epsilon\}}$  et on montre que  $\mathbb{E}(f_n) \rightarrow 0$  ■

La proposition 9.1.3 admet une réciproque partielle :

**Proposition 9.1.4** Si la suite  $(X_n)_n$  converge en probabilité vers  $X$  alors elle admet une sous suite  $(X_{n_k})_k$  qui converge presque sûrement vers  $X$  (quand  $k \rightarrow +\infty$ ).

**R** La démonstration de la proposition 9.1.4 est un peu technique. Elle est basée sur le lemme de Borel-Cantelli.

Si la notion de convergence simple est claire pour nous, ce qu'assure la convergence en probabilité l'est moins. En particulier, nous ne savons pas si nous pouvons compter sur l'unicité de la limite ou encore si cette limite est continue... La CNS suivante nous permet de répondre à ces questions.

**Proposition 9.1.5** La suite  $(X_n)_n$  converge en probabilité vers  $X$  si et seulement si, de toute sous suite  $(X_{n_k})_k$  de  $(X_n)_n$  on peut extraire une sous suite  $(X_{n_{k(i)}})_i$  qui converge presque sûrement vers  $X$  (quand  $i \rightarrow +\infty$ ).

Grâce à ce résultat, on peut établir que :

- La limite en probabilité est unique :

$$X_n \xrightarrow[n \rightarrow +\infty]{P} X \text{ et } Y_n \xrightarrow[n \rightarrow +\infty]{P} Y \Rightarrow Y = X \text{ p.s.}$$

- La limite en probabilité est continue :

$$X_n \xrightarrow[n \rightarrow +\infty]{P} X \Rightarrow f(X_n) \xrightarrow[n \rightarrow +\infty]{P} f(X).$$

**Définition 9.1.4** Soient  $F_n$  la fonction de répartition de  $X_n$  et  $F$  la fonction de répartition de  $X$ . On dit que  $X_n$  **converge en loi** vers  $X$  si :

$$F_n(t) \xrightarrow[n \rightarrow +\infty]{} F(t)$$

**en tout point  $t$  où  $F$  est continue** (c'est à dire  $P(X = t) = 0$ ).

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

**R** Pour une variable aléatoire « continue », il n'y a plus d'hypothèse à faire sur les points de discontinuité de  $F$  puisque pour tout  $t$ ,  $P(X = t) = 0$ .

La convergence en loi n'est pas simple à interpréter si on se réfère aux propriétés arithmétiques usuelles. En effet :

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X \not\Rightarrow (X_n - X) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} 0.$$

Pour s'en convaincre, il suffit de considérer une v.a.  $Z$  de loi  $\mathcal{N}(0, 1)$  et une suite  $(Z_n)$  telle que  $Z_n = Z$ ,  $\forall n \in \mathbb{N}$  et qui converge évidemment en loi vers  $Z$ . Or  $Z$  et  $-Z$  ont même loi donc  $Z_n$  converge également en loi vers  $-Z$ . Pourtant,  $Z - (-Z)$  suit une loi  $\mathcal{N}(0, 2)$  qui n'est évidemment pas celle de la v.a. constante  $Z - Z = 0$ .

Toutefois il est possible d'établir une proposition plus faible qui va dans le sens souhaité.

**Proposition 9.1.6** Si la suite  $(X_n)_n$  converge en loi vers  $X$  et si la suite  $(Y_n)_n$  converge en loi vers 0, alors  $(X_n + Y_n)$  converge en loi vers  $X$ .

Il est possible de reformuler la définition de la convergence en loi sous forme d'une proposition caractéristique basée sur l'inversion de la limite et de l'opérateur  $\mathbb{E}$ .

**Proposition 9.1.7** Une suite de v.a.  $(X_n)_n$  converge en loi vers  $X$  si et seulement si, pour toute fonction  $g$  de  $\mathbb{R}$  dans  $\mathbb{R}$  continue et bornée on a

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

**R** Cette caractérisation de la convergence en loi n'est pas pratique à utiliser. Elle éclaire cependant sur la préservation des « moments » au sens large par convergence uniforme.

L'un des intérêts de la convergence en loi est de permettre, en statistiques par exemple, d'approcher les lois inconnues des  $X_n$  (ou connues mais difficilement calculables) par celle (souvent plus simple) de la limite  $X$ . L'approximation de la loi binomiale par la loi de Poisson ou la loi normale en sont deux exemples.

Nous l'avons sous-entendu depuis le début, la convergence en loi est plus faible que les autres modes de convergence. Les deux propositions suivantes le montrent.

**Proposition 9.1.8** La convergence presque sûre de  $X_n$  vers  $X$  assure la convergence en loi de  $X_n$  vers  $X$ .

**Exercice 9.4** Démontrer la proposition 9.1.8 en utilisant la proposition 9.1.7. ■

**Proposition 9.1.9** La convergence en probabilité de  $X_n$  vers  $X$  assure la convergence en loi de  $X_n$  vers  $X$ .

**Exercice 9.5** Démontrer la proposition 9.1.9.

Pour cela, montrer que pour tout  $t$ , point où  $F_X$  est continue, on a, pour  $\epsilon > 0$  quelconque :

$$P(X_n \leq t) \leq P(X \leq t + \epsilon) + P(|X - X_n| > \epsilon),$$

et, de même que :

$$P(X \leq t - \epsilon) \leq P(X_n \leq t) + P(|X - X_n| > \epsilon).$$

On peut alors conclure (avec précautions quand même). ■

Cette proposition n'admet de réciproque que dans le cas des limites constantes.

**Proposition 9.1.10** La convergence en Loi de  $X_n$  vers la constante  $c$  assure la convergence en probabilités de  $X_n$  vers  $c$ .

**Exercice 9.6** Démontrer la proposition 9.1.10.

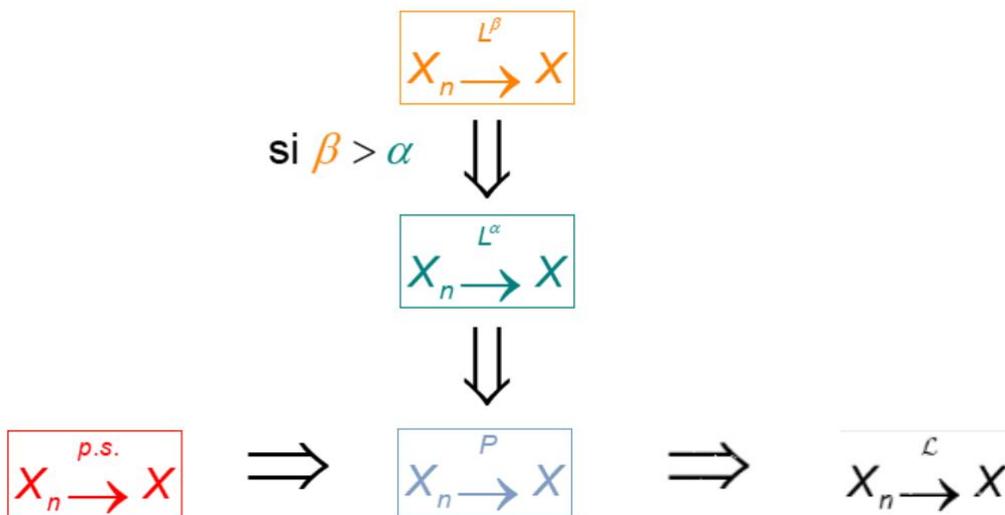
Pour cela, introduisons pour  $\epsilon > 0$ , les fonctions :

$$h = \mathbb{1}_{]-\infty; c-\epsilon[} \cup ]]c+\epsilon; +\infty[ \text{ et } g = h + \frac{c-x}{\epsilon} \mathbb{1}_{]c-\epsilon; c[} + \frac{x-c}{\epsilon} \mathbb{1}_{]c; c+\epsilon[}.$$

$g$  est une fonction continue, bornée et majorant  $h$ . On peut alors montrer (en procédant toujours avec précautions) que :

$$P(|X_n - c| \leq \epsilon) \xrightarrow{n \rightarrow +\infty} 0.$$

Le schéma ci dessous récapitule les diverses implications entre les différentes notions de convergence.



## 9.2 Loi des grands nombres

La loi des grands nombres est la formulation rigoureuse de l'intuition que nous avons de l'effet d'un « grand nombre » de répétitions d'une expérience aléatoire. En effet, si on lance un « grand nombre » de fois une pièce équilibrée on s'attend à observer en moyenne 50% de piles et 50% de faces. De même, si on lance un « grand nombre » de fois un dé à 6 faces on s'attend à observer l'apparition de chaque face avec une fréquence de  $\frac{1}{6}$ .

Il existe une première version de la LGN (dite faible) pour la convergence en probabilité et la convergence  $L^2$ .

**Proposition 9.2.1** — **Loi faible des grands nombres.** Soient  $X_1, X_2, \dots$  des **variables aléatoires réelles indépendantes et identiquement distribuées (v.a.r.i.i.d.)**. Si  $\mathbb{E}(X_1^2) <$

$\infty$ , on a :

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n \rightarrow +\infty]{P} \mathbb{E}(X_1).$$

**Exercice 9.7** Démontrer la proposition 9.2.1 en montrant, grâce à l'inégalité de Bienaymé-Tchebychev que :

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X_1)\right| \geq \epsilon\right) \xrightarrow[n \rightarrow +\infty]{} 0.$$

**Proposition 9.2.2** — **Loi faible des grands nombres.** Soient  $X_1, X_2, \dots$  des **variables aléatoires réelles indépendantes et identiquement distribuées (v.a.r.i.i.d.)**. Si  $\mathbb{E}(X_1^2) < \infty$ , on a

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n \rightarrow +\infty]{L^2} \mathbb{E}(X_1).$$

**Exercice 9.8** Démontrer la proposition 9.2.2 en montrant que

$$\mathbb{E}\left(\left(\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X_1)\right)^2\right) \xrightarrow[n \rightarrow +\infty]{} 0$$

**R** En plus de la convergence, les démonstrations mettent en évidence la vitesse en  $O(\frac{1}{n})$  de ces convergences. Ceci est utile pour déterminer des intervalles de confiance autour de la limite.

La version forte de la loi des grands nombres permet d'alléger l'hypothèse exigeant que  $X$  soit 2 fois intégrable.

**Proposition 9.2.3** — **Loi forte des grands nombres.** Soient  $X_1, X_2, \dots$  des v.a.r. i.i.d. Si  $\mathbb{E}(|X_1|) < \infty$  (en d'autres termes. Si  $X_1$  est intégrable) alors :

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}(X_1).$$

**R** Comme la convergence presque sûre entraîne la convergence en proba, il existe une version forte de la loi des grands nombres pour la convergence en proba.

### 9.3 Théorème central-limite

Ce sont les travaux de Paul Levy qui ont permis d'établir ce résultat fondamental en probabilité et en statistique.

**Théorème 9.3.1 — Théorème central-limite (aussi noté TCL).** Soit  $(X_n)$  une suite de v.a.r. i.i.d. avec  $\mathbb{E}(X_1) = m$  et  $\text{Var}(X_1) = \sigma^2$  ( $m, \sigma^2 < \infty$ ). Alors :

$$\frac{X_1 + \cdots + X_n - nm}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Z \text{ de loi } \mathcal{N}(0, 1).$$

Autrement dit, si l'on somme des v.a.r.i.i.d.  $(X_n)_n$  dont on ignore la loi commune mais dont on connaît l'espérance et la variance, la loi limite de cette somme, lorsque le nombre de ses termes tend vers l'infini, est une loi normale de moyenne  $n\mathbb{E}(X_1)$  et de variance  $n\text{Var}(X_1)$ .

La compréhension et l'interprétation de ce théorème peuvent prendre des formes variées, mathématiques ou physiques. C'est sûrement grâce au lien avec la notion de **fonction caractéristique** (que nous n'aborderons pas dans ce cours) que la compréhension mathématique de l'« apparition » de la loi normale est la plus aisée. Pour une compréhension plus fine, il peut être intéressant de creuser le lien entre TCL et entropie de Shannon.

**R** Sous les hypothèses du théorème précédent, prenons  $a < b$ ,  $f(x) = \mathbb{1}_{[a,b]}(x)$ . Alors,

$$\mathbb{E}\left(f\left(\frac{X_1 + \cdots + X_n - nm}{\sigma\sqrt{n}}\right)\right) \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}(f(Z)),$$

c'est à dire

$$P\left(a \leq \frac{X_1 + \cdots + X_n - nm}{\sigma\sqrt{n}} \leq b\right) \xrightarrow[n \rightarrow +\infty]{} \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

Cette propriété est très utile pour les calculs.

De nombreux auteurs ont cherché (avec succès) à affaiblir les hypothèses « i.i.d. » du théorème central limite. Nous citerons en particulier les travaux de Lyapounov et Lindeberg ainsi que ceux de Gnedenko et Kolmogorov pour les variables non nécessairement identiquement distribuées ou pour lesquelles il existe une « dépendance faible ».

Signalons, pour conclure, que le théorème central limite s'étend sous une forme multidimensionnelle très proche de celle que nous avons vue en dimension 1.





## 10. Simuler des réalisations de variables aléatoires

*« -I'm afraid I don't understand the method of characteristics...  
-Young man, in mathematics you don't understand things. You just get  
used to them. »*

*John Von Neumann, en réponse à un ami physicien.*

Le défi que nous abordons dans ce chapitre n'est pas simple à relever. Il s'agit en effet de mettre au point des méthodes permettant de simuler, par ordinateur, des **réalisations** de variables aléatoires. Pour ce faire, il y a deux étapes à franchir. La première consiste à « apprendre » à un ordinateur, dont le fonctionnement obéit à des méthodes arithmétiques, à **générer des nombres aux hasard**. La seconde, consiste à utiliser cette « graine » de hasard pour générer une réalisation des fonctions mesurables que sont les variables aléatoires **en respectant leur loi**.

### 10.1 Générateurs de nombres pseudo-aléatoires

C'est sous l'impulsion de John Von Neumann que s'est effectué le développement des méthodes de simulation de variables aléatoires, dans l'objectif de mettre au point des **méthodes stochastiques** de calcul pour accélérer le développement de la bombe H durant la seconde guerre mondiale.

Afin d'éviter toute ambiguïté, accordons-nous dès maintenant sur le fait que générer le hasard à partir de méthodes purement arithmétiques, sans l'aide auxiliaire d'outil physique, est impossible. Jamais avare d'un bon mot, von Neumann qui est à l'origine, en 1946, de la **middle-square method**, première méthode de génération de **nombres pseudo-aléatoires**, déclarait à ce sujet :

« *Anyone who considers arithmetical methods of producing random numbers is, of course, in a state of sin.* »

Voyons donc comment offenser les mathématiques!

La discussion de la justification de l'aspect aléatoire des nombres générés par les méthodes que nous détaillons ci-dessous n'est pas aisée. Nous les résumerons (un peu rapidement) en disant qu'il existe des méthodes statistiques appelées **tests statistiques** (que nous verrons en 2A) qui permettent de trancher la question de « l'adéquation à une loi » et de « l'indépendance de variables aléatoires ». Dans la suite, nous pourrions donc considérer que nous disposons de méthodes permettant de générer des suites de nombres aléatoires indépendants.

**R** On dira donc que le générateur (déterministe) de la suite  $(u_1, \dots, u_m, \dots)$  est aléatoire si cette suite a tout d'un comportement aléatoire, « similaire à celui d'une suite de variables aléatoires indépendantes ».

En pratique, un générateur de nombres aléatoires est cyclique et après  $L$  appels, il redonne la valeur initiale. Il est primordial que cette période soit la plus grande possible pour éviter les répétitions dans les appels successifs à ce générateur.

La méthode du **générateur linéaire congruentiel** permet de générer des suites de nombres entiers positifs pour lesquels « on ne voit pas de lien déterministe entre les éléments de la suite » et pour lesquelles on peut avoir des périodicités très grandes. Cette méthode dépend de trois paramètres  $a$ ,  $b$  et  $L$  et se base sur la récurrence :

$$x_{m+1} \equiv ax_m + b \pmod{L};$$

Pour atteindre des périodes maximales égales à  $L$ , on peut choisir :

$$a = 75, \quad b = 0 \quad \text{et} \quad L = 2^{31} - 1 \quad (= 2147483647).$$

On peut alors ramener la suite de nombres entiers positifs indépendants à des réalisations d'une loi uniforme sur  $[0; 1[$  par :

$$u_m = \frac{x_m}{L}.$$

Citons également le **générateur Mersenne-Twister** qui est un générateur plus récent, robuste et rapide, avec pour période, le nombre premier de Mersenne  $10^{19937} - 1$ . Il est encore peu utilisé.

Dans la suite de ce cours, nous partirons du principe que tous les langages usuels de calcul sont munis d'un générateur de nombre aléatoire donnant accès à une suite de réels  $(u_1, \dots, u_m, \dots)$ , réalisations d'une loi uniforme sur  $[0; 1[$ . Selon le langage de programmation et les bibliothèques utilisées, l'appel au générateur passe par la commande *drand48* en C, *random* en C++ et Java, *rand* en Matlab, *grand* en Scilab avec différents paramètres d'appel.

**R** La suite de nombres aléatoires générés est toujours la même pour un jeu de paramètres donné. La graine (seed) du générateur (c'est-à-dire l'indice du premier

nombre renvoyé) est par défaut égale à 1. La suite récupérée par l'utilisateur est donc toujours la même, ce qui n'a finalement plus rien d'aléatoire...



**Il est primordial de changer la graine du générateur lorsqu'on relance un programme utilisant un générateur de nombres aléatoires!**

Ceci peut, par exemple, se faire en indexant cette graine sur l'horloge interne de la machine utilisée, ce qui ajoute une part d'aléa dans l'initialisation de la graine qui dépendra donc de l'instant où le programme est lancé.

## 10.2 Simulation de variables aléatoires unidimensionnelles

Malgré son titre, cette partie et celles qui suivent sont intégralement basées sur des résultats de la théorie des probabilités que nous avons établis jusqu'ici.

La possibilité d'exploiter ces résultats informatiquement pour simuler des réalisations de variables aléatoires à partir d'un générateur de loi uniforme sur  $]0; 1[$  est une conséquence heureuse.

**Proposition 10.2.1** Soit  $F_X$  la fonction de répartition d'une variable aléatoire  $X$  continue.

Alors :

$$F_X(X) = U \text{ avec } U \text{ de loi } \mathcal{U}_{[0;1]}.$$

**Exercice 10.1** Démontrer la proposition 10.2.1.

Pour cela, il est recommandé de s'intéresser à la variable aléatoire  $V = F_X^{-1}(U)$ , après en avoir justifié l'existence. ■



Ce résultat est particulièrement intéressant pour la problématique qui nous occupe. En effet, il « suffit » de pouvoir inverser la fonction de répartition de  $X$  pour obtenir une transformation fonctionnelle de  $U$  en  $X$  et permettre ainsi la simulation de réalisations  $x$  de  $X$  à partir de réalisations  $u$  de  $U$ .

Cette méthode n'est pas générale puisque les propriétés élémentaires de la fonction de répartition ne suffisent pas à assurer de son inversibilité. Ce n'est, en particulier, pas le cas pour les fonctions de répartition des variables aléatoires discrètes.

**Définition 10.2.1** Si  $F_X$  est la fonction de répartition d'une variable aléatoire  $X$ , on appelle fonction quantile la fonction définie pour tout  $u \in ]0; 1[$  par :

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

**Proposition 10.2.2** Si  $F_X$  est la fonction de répartition d'une variable aléatoire  $X$ , et que  $F^{-1}$  est la fonction quantile associée, alors, si  $U$  est une variable aléatoire uniforme sur  $[0; 1]$ ,

$$V = F^{-1}(U) \text{ suit la même loi que } X,$$

et  $F^{-1}$  est appelée **fonction inverse généralisée** de  $F_X$ .

**Exercice 10.2** Démontrer la proposition 10.2.2.

Pour cela, il est recommandé de commencer par montrer que :

$$\forall x \in \mathbb{R}, \forall u \in ]0; 1[, F^{-1}(u) \leq x \Leftrightarrow u \leq F_X(x),$$

avant de s'intéresser à la fonction de répartition de  $V$ . ■

Pour une variable aléatoire discrète  $X$  à valeurs dans  $E$  fini ou dénombrable, rappelons que :

$$P_X = \sum_{x \in E} p_x \delta_x \text{ et que } F_X(t) = \sum_{x \in E} \mathbb{1}_{]-\infty; x]}(t) p_x, \quad \forall t \in \mathbb{R}.$$

Ordonnons alors les éléments de  $E$  pour les noter  $x_0, \dots, x_n$  si  $E$  est fini et  $(x_n)_{n \in \mathbb{N}}$  si  $E$  est dénombrable. Alors :

$$V = x_0 \mathbb{1}_{[0; p_{x_0}[}(U) + \sum_{k \geq 1} x_k \mathbb{1}_{[\sum_{i=0}^{k-1} p_{x_i}; \sum_{i=0}^k p_{x_i}[}(U). \quad (10.1)$$

suit la même loi que  $X$ . Alors pour simuler une réalisation  $x$  de  $X$  :

- simuler une réalisation  $u$  de  $U$ ;
- en déduire  $v$  grâce à (10.1);
- alors  $x = v$ .

### 10.3 Le cas des vecteurs gaussiens

La loi normale et, plus généralement, les vecteurs gaussiens n'entrent pas dans le cadre des variables que l'on peut simuler en utilisant la méthode de la fonction de répartition inverse et ce, pour une raison très simple... il n'existe pas de formulation explicite de la fonction de répartition de la loi normale, ce qui annihile évidemment tout espoir de l'inverser... Cependant, il est possible de trouver un **changement de variables** qui transforme deux variables indépendantes  $U_1, U_2$ , chacune étant de la loi uniforme sur  $[0, 1]$ , en un couple de variables gaussiennes indépendantes. C'est la méthode de **Box-Muller** basée sur le fait que les variables :

$$Z_1 = \sqrt{-2 \log(U_1)} \cos(2\pi U_2) \text{ et } Z_2 = \sqrt{-2 \log(U_1)} \sin(2\pi U_2)$$

sont des variables aléatoires normales indépendantes de moyenne 0 et de variance 1. Par conséquent :

$$X = \sigma_X \sqrt{-2 \log(U_1)} \cos(2\pi U_2) + \mu_X \text{ et } Y = \sigma_Y \sqrt{-2 \log(U_1)} \sin(2\pi U_2) + \mu_Y \quad (10.2)$$

suivent respectivement des **lois normales**  $\mathcal{N}(\mu_X, \sigma_X^2)$  et  $\mathcal{N}(\mu_Y, \sigma_Y^2)$  **indépendantes**.

La méthode de Box-Muller permet donc de simuler des réalisations de gaussiennes indépendantes :

- Simuler deux réalisations  $u_1$  et  $u_2$  de  $U_1$  et  $U_2$  variables indépendantes et uniformes sur  $[0; 1]$ ;

- En déduire  $x$  et  $y$  réalisations de gaussiennes indépendantes grâce à (10.2).

Le fait qu'un changement de variables nous permette de passer d'un vecteur formé de deux v.a. indépendantes et uniformes sur  $[0; 1]$  à un vecteur formé de deux gaussiennes indépendantes est extrêmement intéressant puisqu'un tel vecteur est un vecteur gaussien!

Considérons alors  $U_1, \dots, U_n$ ,  $n$  v.a. uniformes sur  $[0; 1]$  mutuellement indépendantes. En réitérant autant de fois que nécessaire le changement de variable décrit ci-dessus on obtient un vecteur  $\mathbf{X} = (X_1, \dots, X_n)^T$  composé de v.a. gaussiennes indépendantes deux à deux. C'est donc un vecteur gaussien caractérisé par

$$m = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \text{ et } \Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{pmatrix}$$

et en particulier, on sait obtenir le vecteur gaussien centré et réduit  $\mathbf{Z} = (Z_1, \dots, Z_n)^T$  tel que

$$m = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \text{ et } \Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Un tel vecteur, bien que gaussien, ne correspond pas au cas le plus général de vecteur gaussien puisque sa matrice de covariance n'est pas sous la forme la plus générale. Rappelons que si  $\mathbf{X}$  est un vecteur gaussien, le vecteur  $\mathbf{Y} = \mathbf{A}\mathbf{X} + b$  est aussi gaussien d'espérance  $\mathbb{E}(\mathbf{Y}) = \mathbf{A}\mathbb{E}(\mathbf{X}) + b$  et de matrice de covariance  $\Sigma_{\mathbf{Y}} = \mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}^T$ .

Dans le cas du vecteur  $\mathbf{Z}$ , par la transformation  $\mathbf{Y} = \mathbf{A}\mathbf{Z} + b$  on obtient un vecteur gaussien caractérisé par  $\mathbb{E}(\mathbf{Z}) = b$  et  $\Sigma_{\mathbf{Y}} = \mathbf{A}\mathbf{A}^T$ .

Pour générer un vecteur gaussien de moyenne  $M$  et de matrice de covariance  $\Sigma$  « il suffit » de déterminer la matrice  $A$  telle que  $\Sigma = \mathbf{A}\mathbf{A}^T$  et faire le changement de variable  $\mathbf{Y} = \mathbf{A}\mathbf{Z} + M$ . Numériquement, il sera intéressant de choisir  $A$  triangulaire en utilisant, par exemple, une décomposition de Cholesky.

Il existe une autre méthode pour générer des réalisations d'un vecteur gaussien. En effet, si  $\mathbf{X} = (X_1, \dots, X_n)$  est un vecteur gaussien de densité  $f_{\mathbf{X}}$  sur  $\mathbb{R}^n$ . Nous pouvons alors utiliser la décomposition séquentiel de Bayes :

$$f(x) = f(x_1, \dots, x_n) = f(x_1)f(x_2|x_1)f(x_3|x_1, x_2)\dots f(x_n|x_1, x_2, \dots, x_{n-1}). \quad (10.3)$$

Comme nous savons que chacune des densités conditionnelles figurant dans le second membre est une densité normale sur  $\mathbb{R}$ , on peut simuler ainsi les composantes du vecteur de proche en proche : on commence par  $X_1 = x_1$  en utilisant (10.2), ensuite on calcule la moyenne et la variance de  $f(x_2|x_1)$  et on simule  $X_2 = x_2$  (toujours par (10.2)), et ainsi de suite ...

## 10.4 La méthode d'acceptation-rejet

La méthode d'acceptation-rejet est utile lorsqu'on ne peut pas recourir à la fonction de répartition inverse ou si l'on ne connaît pas de changement de variables permettant

de simuler la v.a. d'intérêt. Pour cela, nous allons exploiter ce que nous savons de la densité  $f$  de cette v.a. afin de l'« imiter » avec une densité  $g$  que l'on sait simuler.

Pour cela procédons par étapes en commençant par le résultat suivant.

**Proposition 10.4.1** Soit  $f$  une densité de probabilité sur  $\mathbb{R}$ .

Si le vecteur :

$$(X; Y) \in (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$$

suit une loi uniforme sur :

$$B = \{(x; y) \in \mathbb{R}^2; 0 < y < f(x)\},$$

alors  $X$  a pour densité  $f$ .

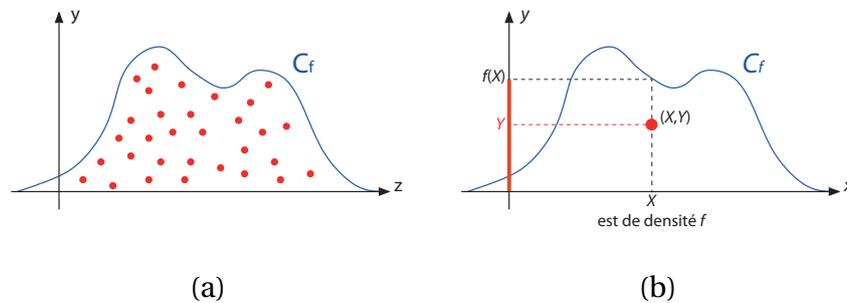


FIGURE 10.1 – Caractérisation des points uniformément répartis sous la courbe  $C_f$ .

**Exercice 10.3** Démontrer la proposition 10.4.1.

Pour cela, il est recommandé de calculer la mesure de Lebesgue de  $B$  avant de s'intéresser à la fonction de répartition de  $X$ . ■

Il semble donc qu'il « suffise » de pouvoir simuler des réalisations uniformément réparties sous la courbe de la densité  $f$  pour que les abscisses de ces points soient des réalisations suivant  $f$ . La proposition suivante montre que cela est possible pour les densités à support borné.

**Proposition 10.4.2** Soient  $A$  et  $B$  deux boréliens de  $\mathbb{R}^2$  de mesure de Lebesgue finie. Supposons de plus que  $B \subset A$ . On a donc  $0 \leq \lambda(B) \leq \lambda(A) < \infty$ .

Considérons également une suite  $(X_n)_{n \geq 0}$  de variables aléatoires **indépendantes et de loi uniforme sur  $A$**  et définissons, pour tout  $n \geq 0$ , la suite :

$$Y_n = \mathbb{1}_B(X_n).$$

Alors :

- La suite  $(Y_n)_{n \geq 0}$  est une suite de v.a.i.i.d. de loi de Bernoulli de paramètre

$$p = \frac{\lambda(B)}{\lambda(A)};$$

- La variable aléatoire  $N = \min \{n \in \mathbb{N}^*, X_n \in B\}$ , appelée **temps d'attente**, suit une loi géométrique de paramètre  $p$ ;
- Les variables aléatoires  $N$  et  $X_N$  sont indépendantes;
- La variable aléatoire  $X_N$  **suit une loi uniforme sur  $B$** .

Autrement dit, à partir d'une loi uniforme sur un borélien  $A$  on peut simuler des réalisations d'une loi uniforme sur un borélien  $B \subset A$ . On bénéficie, de plus, d'informations complémentaires extrêmement précieuses. En effet, on sait que pour obtenir une réalisation suivant la loi uniforme sur  $B$ , cela prendra « d'autant plus de temps » que  $A$  est « plus gros » que  $B$ , puisque le paramètre  $p$  de la variable  $N$  sera d'autant plus petit.

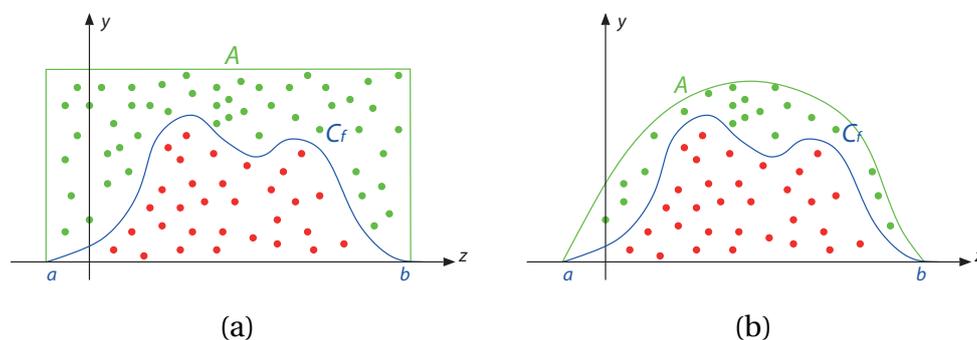


FIGURE 10.2 – Propagation de l'uniforme répartition d'un borélien à un autre.

La figure 10.2 illustre graphiquement la proposition 10.4.2. On simule des points dans le borélien  $A$  et on « rejette » tous les points (en vert) qui ne sont pas sous la courbe  $C_f$ . Les points acceptés (en rouge) sont donc des points uniformément répartis sous la courbe. On comprend alors pourquoi l'on rejette beaucoup plus de points dans le cas (a) que dans le cas (b).

**Exercice 10.4** Démontrer la proposition 10.4.2. ■

Le cas (b) de la figure 10.2 amène à s'interroger sur la possibilité d'étendre le cadre strict de la proposition 10.4.2 pour laquelle on se base sur des lois uniformes sur les pavés de  $\mathbb{R}^2$  (que nous savons simuler facilement). Par extension, se pose également la question de la simulation de réalisations de variables aléatoires dont la densité n'est pas de support fini.

La proposition suivante qui est une forme de réciproque de la proposition 10.4.1 répond à ces deux questions. En effet, elle permet d'affirmer que si on sait simuler des réalisations d'une variable aléatoire de densité  $g$ , alors il est possible de simuler des points uniformément répartis sous la courbe de  $g$  et même celle de  $cg$  où  $c \in \mathbb{R}^{+*}$ .

**Proposition 10.4.3** Soit  $X$  une variable aléatoire de densité  $g$  sur  $\mathbb{R}$  et soit  $U$  une variable aléatoire qui suit une loi uniforme sur  $[0; 1]$ . Supposons  $X$  et  $U$  **indépendantes**. Alors pour tout  $c > 0$ , le vecteur :

$$(X; cg(X)U)$$

suit une loi uniforme sur :

$$A = \{(x; y) \in \mathbb{R}^2; 0 < y < cg(x)\}.$$

Le cas (b) de la figure 10.2 est alors possible si l'on se dote d'une densité  $g$  suivant laquelle on est capable de simuler et que l'on trouve un réel positif  $c$  pour lequel  $cg$  majore  $f$  sur  $\mathbb{R}$ . Comme le montre la figure 10.3 on peut même dépasser le cadre des densité à support fini.

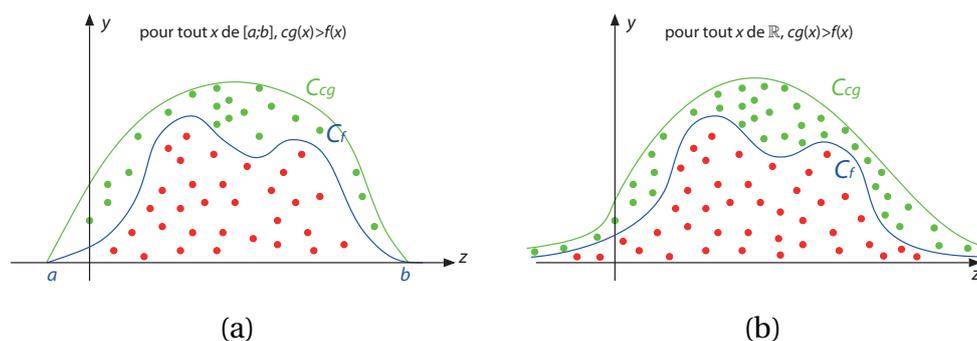


FIGURE 10.3 – Propagation de l'uniforme répartition d'une densité à une autre.

**Exercice 10.5** Démontrer la proposition 10.4.3. ■

Nous avons alors tous les éléments en main pour établir la forme générale du théorème d'acceptation-rejet.

**Théorème 10.4.4** • Soit  $f$  une densité de probabilité sur  $\mathbb{R}$ . Supposons alors qu'il existe une densité  $g$  sur  $\mathbb{R}$  et une constante  $M \in \mathbb{R}^{+*}$  telles que :

$$\forall x \in \mathbb{R}, f(x) \leq Mg(x);$$

- Soit  $(U_n)_{n>0}$  une suite de v.a.i.i.d. de loi **uniforme** sur  $[0; 1]$  ;
- Soit  $(X_n)_{n>0}$  une suite de v.a.i.i.d. de **densité  $g$ , indépendantes de  $(U_n)$**  ;

Alors :

- La variable aléatoire  $N = \min \{n \in \mathbb{N}^*, Mg(X_n)U_n < f(X_n)\}$  suit une loi géométrique de paramètre  $p = \frac{1}{M}$  ;
- La variable aléatoire  $X_N$  **a pour densité  $f$** .

**Exercice 10.6** Démontrer le théorème 10.4.4. ■

Notons la grande généralité de cette méthode. En particulier, la densité  $f$  peut être connue à un facteur multiplicatif près puisque l'algorithme sera utilisable dès lors que le rapport  $\frac{f(z)}{Mg(z)}$  est calculable.

Enfin, la variable  $N$  étant une loi géométrique,  $\mathbb{E}(N) = \frac{1}{p} = M$ , pour attendre « le moins longtemps possible » on recherchera une fonction  $g$  simulable pour laquelle  $M$  est aussi petit que possible. C'est un point crucial de la méthode.

## 10.5 Méthode de Monte-Carlo

**Définition 10.5.1** On appelle **méthode de Monte-Carlo** toute méthode visant à calculer une valeur numérique en utilisant des procédés aléatoires.

Le nom de ces méthodes fait allusion aux jeux de hasard pratiqués à Monte-Carlo et leur développement s'est effectué, durant la Seconde Guerre mondiale, dans le cadre des recherches sur la fabrication de la bombe atomique pour résoudre numériquement des équations aux dérivées partielles particulièrement complexes. Ces méthodes sont très employées pour le calcul des intégrales en dimensions plus grandes que 1 (surfaces, volumes, etc...) et une application importante est le filtrage adaptatif qui permet, par exemple, de faire de la trajectographie ou de la prédiction en finance.

L'intégration par une méthode de Monte-Carlo est fondée sur deux éléments essentiels de la théorie des probabilités :

- la **loi des grands nombres**;
- le **théorème de transfert**.

Lorsqu'on cumule ces deux notions, on établit que, pour une densité de probabilité  $f$  sur  $\mathbb{R}^d$  et une application  $h$  sur  $\mathbb{R}^d$  telle que  $fh$  soit intégrable :

$$\frac{h(X_1) + \dots + h(X_n)}{n} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}[h(X_1)] = \int_{\mathbb{R}^d} h(x)f(x)dx, \quad (10.4)$$

où les  $X_1, \dots, X_n, \dots$  sont des v.a.i.i.d. dont la loi commune admet  $f$  pour densité.

L'idée que l'on retient de cette approche est qu'elle offre la possibilité d'approcher l'intégrale  $I = \int_{\mathbb{R}^d} h(x)f(x)dx$  par la moyenne arithmétique  $\frac{1}{n} \sum_{i=1}^n h(X_i)$  lorsque  $n$  est « très grand ».

D'un point de vue pratique, en simulant un « grand nombre » de réalisations suivant la loi de densité  $f$  on peut, par un simple calcul de moyenne, approcher l'intégrale  $I$ . De plus, lorsque les variables  $h(X_1), \dots, h(X_n)$  sont de carré intégrable on peut appliquer le théorème central limite et il devient alors possible de construire des intervalles de confiance pour caractériser la qualité de l'approximation en fonction de  $n$  ainsi que la vitesse de convergence de la méthode.

La formulation de (10.4) laisse penser qu'on ne peut pas calculer n'importe quelle intégrale par une méthode de Monte-Carlo puisqu'une espérance est une intégrale sur  $\mathbb{R}^d$  d'une part, par rapport à la densité  $f$  d'autre part. Pourtant, si, pour  $A \subset \mathbb{R}^d$  de mesure de Lebesgue finie ( $\lambda(A) < \infty$ ) et pour  $h$  intégrable on souhaite calculer l'intégrale  $J = \int_A h(x)dx$ , il suffit de calculer :

$$\frac{J}{\lambda(A)} = \frac{1}{\lambda(A)} \int_{\mathbb{R}^d} h(x)\mathbb{1}_A dx,$$

ce qui permet d'affirmer que, si les  $V_i$  sont des variables uniformes sur  $A$  :

$$\frac{h(V_1) + \dots + h(V_n)}{n} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \frac{J}{\lambda(A)}$$

Une illustration simple de l'emploi de cette méthode est l'estimation de l'aire d'une surface. Prenons l'exemple de la figure 10.4.

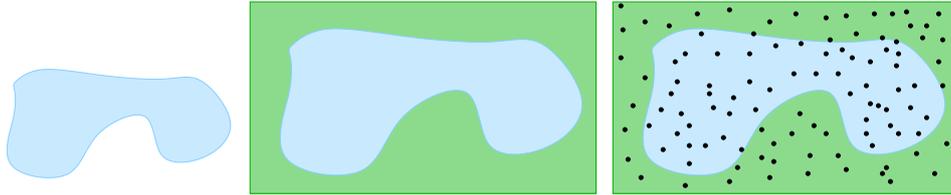


FIGURE 10.4 – Illustration de la méthode de Monte-Carlo pour estimer une surface.

La technique consiste à englober la surface dans un rectangle dans lequel on tire des points uniformément répartis. Ensuite la méthode de Monte Carlo approxime l'aire de la façon suivante :

$$\frac{\text{Nb de points dans } S}{\text{Nb total de points}} \approx \frac{\text{Aire de } S}{\text{Aire du rectangle}}.$$

Dans ce cas,  $f$  est la loi uniforme sur le rectangle et  $h$  est la fonction indicatrice de l'appartenance à  $S$ .

Cette méthode peut être utilisée dans un grand nombre de situations ; cependant, elle ne donne pas des résultats plus intéressants que certaines méthodes d'intégration numérique pour des intégrales simples voire doubles. C'est véritablement pour les calculs d'intégrales en grande dimension que les méthodes de Monte-Carlo sont particulièrement intéressantes.

**R** Pour  $X_1 = x_1, \dots, X_n = x_n$ , la quantité  $\frac{1}{n} \sum_{i=1}^n h(X_i)$  peut être vue comme une intégrale par rapport à la mesure « empirique »,  $\mu(\cdot) = \frac{1}{n} \sum_{i=1}^n \delta(\cdot)$ , cette dernière consistant à associer à chaque ensemble  $B \subset \mathbb{R}^d$  la proportion des  $x_1, \dots, x_k$  qu'il contient.

Si nous avons évoqué la question de la vitesse de convergence des méthodes de Monte-Carlo, nous n'avons, en revanche, pas cherché à l'optimiser. Nous allons donc réparer ce manquement grave à la vérité mathématique en commençant par constater que si  $g$  est une densité de probabilité dont le support contient celui de  $f$ , l'intégrale d'intérêt  $\int_{\mathbb{R}^d} h(x) f(x) dx$  peut aussi s'écrire :

$$\int_{\mathbb{R}^d} h(x) f(x) dx = \int_{\mathbb{R}^d} \frac{h(x) f(x)}{g(x)} g(x) dx = \mathbb{E} \left[ \frac{h(X_1) f(X_1)}{g(X_1)} \right],$$

et donc pour une suite  $X_1, \dots, X_n$  de variables aléatoires indépendantes dont la loi commune admet  $g$  pour densité, nous avons, par application du théorème de transfert

et de la loi des grands nombres :

$$\frac{1}{n} \left( \frac{h(X_1)f(X_1)}{g(X_1)} + \dots + \frac{h(X_n)f(X_n)}{g(X_n)} \right) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \int_{\mathbb{R}^d} h(x)f(x)dx. \quad (10.5)$$

On peut alors chercher à déterminer la densité  $g$  pour laquelle la vitesse de la convergence de la méthode de Monte-Carlo vers  $\int_{\mathbb{R}^d} h(x)f(x)dx$  est la plus rapide. Cette vitesse étant proportionnelle à la variance des variables aléatoires  $\frac{1}{n} \left( \frac{h(X_1)f(X_1)}{g(X_1)} + \dots + \frac{h(X_n)f(X_n)}{g(X_n)} \right)$ , on se base sur la proposition 10.5.1 pour choisir, si possible, la méthode la plus rapide.

**Proposition 10.5.1** La densité :

$$g^*(x) = \frac{|h(x)|f(x)}{\int_{\mathbb{R}^d} |h(t)|f(t)dt}$$

est celle qui minimise la variance de  $\frac{1}{n} \left( \frac{h(X_1)f(X_1)}{g(X_1)} + \dots + \frac{h(X_n)f(X_n)}{g(X_n)} \right)$ .

 Le plus souvent, l'expression explicite de  $g$  n'est pas atteignable.

### L'exemple des chaînes de Markov cachées.

Considérons une suite de variables aléatoires réelles  $X_1, \dots, X_n, \dots$  dont la loi admet, pour chaque  $n \in \mathbb{N}$ , une densité de la forme :

$$f(x_1, \dots, x_n) = f(x_1)f(x_2|x_1)f(x_3|x_2) \dots f(x_n|x_{n-1}) \quad (10.6)$$

Une telle suite est appelée **chaîne de Markov**.

On montre que l'on a la factorisation (10.6) (qui est un cas particulier de (10.3)) si et seulement si pour tout  $n \in \mathbb{N}$  la loi de  $X_n$  conditionnelle à  $X_1, \dots, X_{n-1}$  est égale à sa loi conditionnellement à  $X_{n-1}$ .

Une telle suite est également dite à **mémoire d'ordre 1**. En effet, si l'on observe  $X_{n-1} = x_{n-1}$ , la loi de  $X_n$  ne dépend que de  $x_{n-1}$  et l'information contenue dans  $x_1, \dots, x_{n-2}$ , qui est la mémoire de ce qui s'est passé avant l'instant  $n-1$ , n'a aucune influence sur le comportement de  $X_n$ . Lorsqu'on y réfléchit, on réalise que cette hypothèse, est la plus faible qu'il soit possible de faire pour que les  $X_i$  ne soient pas indépendantes.

Une autre manière de voir les choses est de mettre un ordre temporel sur les variables  $X_1, \dots, X_n$ . Si l'on suppose que  $n-1$  est l'instant présent, les instants  $1, \dots, n-2$  sont le passé, et l'instant  $n$  est le futur, on dit aussi que **le futur et le passé sont indépendants conditionnellement au présent**.

 si  $x_{n-1}$  n'est pas observé (on supprime le conditionnement),  $X_n$  n'est plus indépendant de  $X_1, \dots, X_{n-1}$ .

Supposons désormais que les variables  $X_1, \dots, X_n, \dots$  de la chaîne de Markov ne peuvent pas être observées directement mais qu'il est possible d'observer les  $Y_1, \dots, Y_n, \dots$ , reliées aux  $X_i$ . Plus précisément, on suppose que pour tout  $n \in \mathbb{N}$ , la loi de  $(Y_1, \dots, Y_n)$  conditionnellement à  $(X_1, \dots, X_n) = (x_1, \dots, x_n)$  est donnée par :

$$f(y_1, \dots, y_n | x_1, \dots, x_n) = f(y_1 | x_1) f(y_2 | x_2) \dots f(y_n | x_n). \quad (10.7)$$

La question qui se pose alors et qui admet une multitude d'applications, est alors de retrouver  $X_1, \dots, X_n$  à partir de  $Y_1, \dots, Y_n$ .

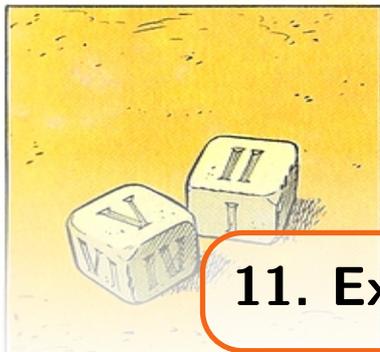
Une approche particulière consiste à calculer, de proche en proche,  $f(x_n | y_1, \dots, y_n)$  à partir de  $f(x_{n-1} | y_1, \dots, y_{n-1})$  et  $y_n$ . Dans une telle approche **appelée filtrage adaptatif** on considère généralement deux étapes :

$$\begin{aligned} f(x_n | y_1, \dots, y_{n-1}) &= \int_{\mathbb{R}} f(x_{n-1}, x_n | y_1, \dots, y_{n-1}) dx_{n-1} \\ &= \int_{\mathbb{R}} f(x_n | x_{n-1}) f(x_{n-1} | y_1, \dots, y_{n-1}) dx_{n-1}, \end{aligned} \quad (10.8)$$

puis :

$$f(x_n | y_1, \dots, y_n) = \frac{f(x_n, y_n | y_1, \dots, y_{n-1})}{f(y_n | y_1, \dots, y_{n-1})} = \frac{f(x_n | y_1, \dots, y_{n-1}) f(y_n | x_n)}{f(y_n | y_1, \dots, y_{n-1})}. \quad (10.9)$$

- Dans les cas simples, typiquement **linéaires et Gaussiens**, (10.9) admet une solution analytique et on applique une méthode déterministe, le célèbre **filtre de Kalman**.
- Dans le cas général on doit approcher l'intégration dans (10.8) en utilisant des tirages stochastiques. De telles méthodes sont dites **filtrage particulaire** en référence aux points, ou particules, obtenues par les tirages. Pour ces méthodes, très étudiée par la recherche actuelle, que vous pourrez étudier en 2A, l'optimisation de la vitesse de convergence des méthodes de Monte Carlo est très importante.



## 11. Exercices

*« J'ai peur en avion... J'ai toujours eu peur en avion... »*

*Pour qu'un avion se crashe il faut une succession d'emmerdements qui dépasse l'entendement, une loi des séries de l'emmerdement qui dépasse toutes les lois... y compris la loi de Murphy, la loi de l'emmerdement maximal... Non... Pour qu'un avion se crashe il faut penser à la théorie des tranches de gruyères...*

*La première fois que j'ai entendu quelqu'un me parler de cette théorie je devais avoir une dizaine d'années; Un collègue de travail de mon père a tenté de m'expliquer pourquoi je ne devais pas avoir peur en avion :*

*La probabilité d'un crash aérien est équivalente à la probabilité qu'on puisse tirer un trait bien droit qui traverserait tous les trous d'un gruyère coupé en tranches...*

*J'avoue que ça ne m'avait pas convaincu plus que ça... surtout, quand bien des années plus tard, le 25 mai 1991 à 18h20, le vol ITZ5632 Paris-Zurich s'est écrasé... et que le collègue de mon père était dedans...*

*Des fois on peut tirer des traits bien droits à travers les trous de plusieurs tranches de gruyère... c'est comme ça... c'est suisse... »*

Yann Kerbec (Vincent Elbaz),  
« Ma vie en l'air » de Rémi Bezançon

## Séance 1 - Évènements et probabilités

### Exercice 1 — Formule de Bayes Séquentielle.

Considérons  $N$  évènements  $A_k$  d'un espace probabilisé  $(\Omega, \mathcal{A}, P)$  pour lesquels on a  $P\left(\bigcap_{i=1}^{N-1} A_i\right) \neq 0$ .

Démontrer la relation suivante :

$$P\left(\bigcap_{i=1}^N A_i\right) = P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2) \dots P\left(A_k \left| \bigcap_{i=1}^{k-1} A_i \right.\right) \dots P\left(A_N \left| \bigcap_{i=1}^{N-1} A_i \right.\right)$$

### Exercice 2 — Formule de Poincaré.

Considérons  $N$  évènements  $A_i$  d'un espace probabilisé  $(\Omega, \mathcal{A}, P)$ .

1. Démontrer que :

$$P\left(\bigcup_{i=1}^N A_i\right) = \sum_{i=1}^N P(A_i) - \sum_{1 \leq i_1 < i_2 \leq N} P(A_{i_1} \cap A_{i_2}) + \dots \\ + (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq N} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) + \dots + (-1)^{N+1} P\left(\bigcap_{i=1}^N A_i\right).$$

2. Dans une réception, chacun des  $n$  invités a donné son chapeau au vestiaire et on suppose qu'à la fin de la réception, les  $n$  chapeaux sont distribués au hasard entre les invités.

- Calculer la probabilité  $P_n(A)$  qu'au moins l'un des invités reçoive le chapeau qui lui appartient.
- Quelle est la limite de cette probabilité lorsque  $n$  tend vers l'infini.

### Exercice 3 — Axiome de $\sigma$ -additivité.

Deux joueurs  $A$  et  $B$  lancent deux dés à tour de rôle, le joueur  $A$  commence. Ce joueur sera dit gagnant s'il obtient lors d'un lancer un total de six points avant que  $B$  n'obtienne un total de sept points.  $B$  est dit gagnant dans le cas contraire.

A quel joueur le jeu est-il le plus favorable ?

## Séance 2 - Probabilités conditionnelles

### Exercice 1 — Formule de Bayes - Probabilités des causes.

On considère un test médical permettant de dépister une maladie donnée dans une population donnée. Il s'agit d'une expérience aléatoire au sens où l'on ne connaît pas, a priori, le résultat du test pratiqué sur une personne de la population étudiée, ni si celle-ci est atteinte de la maladie que l'on veut dépister. (Il faut avoir à l'esprit qu'un test n'est pas infaillible et qu'il peut donner un "faux-positif" ou un "faux-négatif").

Une personne étant choisie, on appellera

- $M$  l'événement : la personne est atteinte de la maladie dépistée
- $T^+$  l'événement : le test pratiqué est positif.

On notera  $\bar{M}$  et  $T^-$  les événements complémentaires de  $M$  et  $T^+$  respectivement.

On désire s'assurer de la qualité de ce test qui pour nous, utilisateurs, est mesurée par  $P(M|T^+)$  et  $P(\bar{M}|T^-)$ . Le test est d'autant plus fiable que ces probabilités sont proches de 1.

Lors de la phase de développement du test, c'est par rapport aux probabilités conditionnelles  $P(T^+|M)$  et  $P(T^-|\bar{M})$  que les médecins peuvent faire l'étalonnage.

Dans la suite, on supposera par souci de simplification que

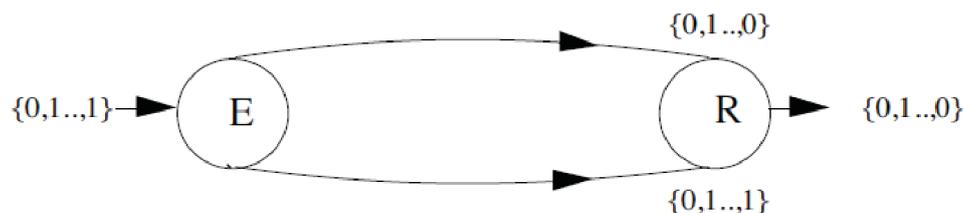
$$P(T^+|M) = P(T^-|\bar{M}) = r \text{ où } r \in [0; 1].$$

1. Quelle doit être la valeur de  $r$  si l'on veut s'assurer que  $P(M|T^+) = 0,95$  pour  $P(M) = 5 \cdot 10^{-3}$  ?
2. Calculer  $P(\bar{M}|T^-)$ . Comment interpréter ce résultat ?

### Exercice 2 — Introduction à la détection, formule de Bayes.

Des symboles 0 ou 1 sont émis par une source  $E$  d'information binaire et reçus par un récepteur  $R$  au moyen de deux canaux de transmissions imparfaits. Les probabilités d'émission sont :

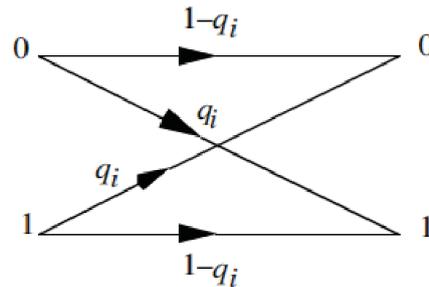
$$P(0) = 0,3 \text{ et } P(1) = 0,7.$$



On suppose que les canaux de transmission peuvent être entachés d'erreurs en réception de façon indépendante. On supposera de plus que chaque canal est un **canal binaire symétrique** (i.e. les probabilités d'erreur conditionnellement au symbole émis sont égales).

Les probabilités d'erreur (conditionnellement au symbole émis) de chacun de ces canaux sont respectivement :

$$q_1 = 10^{-7} \text{ et } q_2 = 2 \cdot 10^{-7}.$$



1. Le récepteur reçoit à chaque instant un symbole de chaque canal de transmission. Il reçoit donc un couple parmi  $(0, 0)$ ,  $(0, 1)$ ,  $(1, 0)$ ,  $(1, 1)$  et doit prendre une « décision », c'est-à-dire fournir un symbole 0 ou 1 au destinataire.

On décide donc de choisir le symbole qui, connaissant le signal reçu, a la probabilité la plus grande d'avoir été émis (règle appelée du **maximum a posteriori**).

Préciser la règle de décision.

2. Calculer la probabilité d'erreur globale d'un tel système de transmission.
3. Que gagne-t-on à mettre en place un système à deux canaux de transmission?

### Exercice 3 — Sécurité aérienne.

On considère deux avions, un biréacteur  $B$  et un quadriréacteur  $Q$ .

On suppose que chaque réacteur a la même probabilité  $p$  de tomber en panne (au cours d'un vol). De plus, on admet que les pannes des différents réacteurs sont indépendantes.

Soit  $X$  (resp  $Y$ ) la variable aléatoire « nombre de réacteurs de  $B$  (resp  $Q$ ) en panne lors d'un vol ».

1. Déterminer les lois de  $X$  et de  $Y$ .
2. On estime en outre qu'un avion ne peut achever son vol que si la moitié au moins des réacteurs fonctionnent normalement.

Indiquer suivant les valeurs de  $p$  l'avion qui offre les meilleures garanties.

## Séance 3 - Lois de variables aléatoires

### Exercice 1 — Approximation de la loi Binomiale par la loi de Poisson.

On considère une suite de variables aléatoires  $X_n$  ayant une loi de probabilité Binomiale de paramètres  $n$  et  $p_n = \frac{\lambda}{n}$ .

1. Montrer que,  $\forall k \in \{0, \dots, n\}$ ,

$$\lim_{n \rightarrow \infty} P(X_n = k) = \exp - \lambda \frac{\lambda^k}{k!}.$$

On dit que la suite de variables aléatoires  $X_n$  **converge en loi** vers une variable aléatoire  $Y$  de loi de Poisson de paramètre  $\lambda$ .

2. En déduire une approximation, pour tout  $k \in \{0, \dots, n\}$ , de

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

3. Une compagnie d'assurance assure une flotte de 500 navires dont la valeur unitaire est de 5 millions d'euros. Le bail stipule que le sinistre que l'on prend en considération est la perte totale d'un navire, événement dont la probabilité est estimée à 0,001 pour une année. Les risques de perte des navires sont indépendants. On appelle  $X$  la variable aléatoire ayant pour valeur le nombre de navires assurés perdus au cours de l'année.

Quelle est la loi de  $X$ ?

4. La compagnie règle ses indemnités à la fin de chaque année.

A combien doivent s'élever ses réserves pour qu'il y ait une probabilité égale à 0,999 que la compagnie honore ses engagements?

5. Une autre compagnie est placée dans les mêmes conditions que la première. Les deux compagnies envisagent de fusionner. La nouvelle compagnie prendra en charge les 1000 navires.

Refaire les calculs de la question 4. pour la nouvelle compagnie et interpréter le résultat.

**Exercice 2 — Loi géométrique, loi sans mémoire.**

On considère une suite d'expériences aléatoires de Bernoulli indépendantes. Celle-ci est modélisée par l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  avec

- $\Omega = \Omega_0^{\mathbb{N}}$  où  $\Omega_0 = \{\bar{a}, a\}$ ;
- $\mathcal{A} = [\mathcal{P}(\Omega_0)]^{\mathbb{N}}$ ;
- $P = \otimes_{i=1}^{\infty} P_0$  avec  $P_0(a) = p$  et  $P_0(\bar{a}) = 1 - p$ .

1. On s'intéresse à la variable aléatoire  $N$  constituée par l'ordre  $k$  dans la suite où l'événement  $\{a\}$  apparaît pour la première fois.

Déterminer la loi  $P_N$  de  $N$ .

(Cette loi est appelée loi géométrique de paramètre  $p$ .)

2. Supposons que nous n'ayons pas observé l'événement  $\{a\}$  à l'issue des  $n$  premières expériences aléatoires de Bernoulli.

Quelle est alors la loi de probabilité du nombre  $N-n$  d'expériences aléatoires représentant l'instant de la première apparition de  $\{a\}$  à partir de ce nouvel instant origine?

3. Justifier l'appellation de loi discrète sans mémoire pour la loi géométrique.

4. Soit une variable aléatoire  $X$  entière et strictement positive telle que :

$$P(X-n = k | X > n) \text{ ne dépend pas de } n.$$

Quelle est sa loi de probabilité  $P_X$ ?

5. Comment interpréter ce résultat?

**Exercice 3 — Loi exponentielle, loi sans mémoire.**

La durée  $T$  en minutes d'une communication téléphonique est une variable aléatoire de densité de probabilité sur  $\mathbb{R}$  :

$$f(t) = \alpha e^{-Kt} \mathbb{1}_{\mathbb{R}^+}(t), \text{ où } K \text{ est un réel positif fixé.}$$

1. Déterminer  $\alpha$ .
2. Quelle est la fonction de répartition  $F$  de  $T$ .
3. Calculer la probabilité pour qu'une communication dure plus de 3 minutes.
4. Quelle est la probabilité pour qu'une communication dure entre 3 et 6 minutes?
5. Quelle valeur de  $K$  faudrait-il choisir pour que la probabilité qu'une communication dure plus de 3 minutes soit égale à 0,1?
6. Quelle est la probabilité qu'une communication dépasse  $t_0 + t$  étant donné

- 
- que sa durée a déjà dépassé  $t_0$ ?
7. Comment interpréter ce résultat?
  8. Établir la réciproque du résultat précédent.

## Séance 4 - Lois et calculs de probabilités

### Exercice 1 — Logarithme et loi uniforme.

Soit  $\lambda \in \mathbb{R}^{+*}$ , soit  $U$  une v.a de loi uniforme sur  $[0, 1]$ . On considère  $X = -\frac{\ln(U)}{\lambda}$ .

1. Pourquoi peut-on dire que  $X$  existe presque sûrement?
2. Quelle est la loi de  $X$ ?

### Exercice 2 — Coiffeur.

John entre chez le coiffeur. Celui-ci est occupé avec un client. La coupe dure (exactement) 30 min, et celle-ci a débuté selon une durée aléatoire uniformément répartie entre 0 et 30 min.

Calculer la probabilité que  $t$  minutes après l'entrée de John, le coiffeur n'ait pas fini la coupe.

### Exercice 3 — Autobus et taxi.

Jane s'impatiente à un arrêt d'autobus : elle décide de prendre le taxi, si un taxi libre venait à passer devant l'arrêt avant le prochain autobus.

Soient  $X$  la variable donnant le temps d'arrivée du prochain autobus,  $Y$  la variable donnant le temps de passage du prochain taxi libre devant l'arrêt. Ces temps sont exprimés en minutes.

On suppose que

- $X$  est une variable discrète prenant trois valeurs :

$$P(X = 5) = \frac{1}{4}, \quad P(X = 15) = \frac{1}{2}, \quad P(X = 25) = \frac{1}{4};$$

- $Y$  suit une loi  $Exp\left(\frac{1}{5}\right)$ ;
- $X$  et  $Y$  sont indépendantes.

1. Quelle est la probabilité pour que Jane attende plus de dix minutes?
2. Quelle est la probabilité pour que Jane prenne le taxi plutôt que l'autobus?
3. Quelle est la probabilité pour que Jane prenne le taxi plutôt que l'autobus, si l'on sait en outre qu'elle a attendu plus de dix minutes?

**Exercice 4 — Bus.**

John, qui arbore désormais une toute nouvelle coupe de cheveux, doit prendre le prochain bus.

1. John est à l'arrêt de bus. On modélise le temps d'attente avant que le bus n'arrive par une v.a.  $X$  de loi  $Exp(\lambda_1)$ .

Quelle est la probabilité que John attende plus de 10 minutes?

2. John voudrait bien prendre le bus en compagnie de son amie, Jane, mais celle-ci n'est pas encore arrivée. On modélise le temps qu'elle met pour arriver par une v.a.  $Y$  suivant une loi  $Exp(\lambda_2)$ .

Quelle est la probabilité que John puisse effectuer le voyage avec Jane?

## Séance 5 - Fonctions de répartition et changement de variable

### Exercice 1 — Changement de variables non bijectif.

On considère une famille de  $n$  variables aléatoires continues  $X_1, X_2, \dots, X_n$  indépendantes et de même densité de probabilité  $f_X$  sur  $\mathbb{R}$ .

1. Montrer que

$$I_n = \inf_{k \in \{1, \dots, n\}} (X_k) \text{ et } S_n = \sup_{k \in \{1, \dots, n\}} (X_k)$$

sont des variables aléatoires continues, dont on déterminera les densités de probabilité.

2. Calculer la fonction de répartition  $F_{I_n, S_n}$  du couple aléatoire  $(I_n, S_n)$  (on pourra l'exprimer avec profit à l'aide de la probabilité  $P(I_n > x, S_n \leq y)$ ).
3. En déduire la densité de probabilité  $f_{I_n, S_n}$ .
4. Les variables aléatoires  $I_n$  et  $S_n$  sont-elles indépendantes?
5. On appelle « étendue » des  $n$  variables aléatoires  $X_1, X_2, \dots, X_n$ , la variable aléatoire

$$R_n = S_n - I_n.$$

Quelle est la densité de probabilité de la variable aléatoire  $R_n$ ?

6. Calculer de deux façons différentes la probabilité pour que la valeur prise par la variable aléatoire  $X_n$  soit plus grande que la valeur prise par les  $n-1$  autres variables aléatoires  $X_k$  pour  $k \in \{1, 2, \dots, n-1\}$ .

### Application :

- a) Quelle sont les lois de probabilité des variables aléatoires  $I_n, S_n$  et  $R_n$  et du vecteur aléatoire  $(I_n, S_n)$  dans le cas particulier où  $f_X$  est la loi uniforme sur  $[0, 1]$ ?
- b) Examiner

$$\lim_{n \rightarrow \infty} F_{I_n}, \lim_{n \rightarrow \infty} F_{S_n} \text{ et } \lim_{n \rightarrow \infty} F_{R_n}.$$

- c) Comment interpréter ces résultats?

**Exercice 2 — Changement de variables bijectif.**

On considère une chaîne moléculaire dont la longueur  $L$  est une variable aléatoire continue admettant une densité de probabilité  $f_L$ .

On note  $L_1$  et  $L_2$  les variables aléatoires représentant les longueurs des deux chaînes résultant d'une rupture de la chaîne initiale.

1. Comment peut-on modéliser ce phénomène de rupture aléatoire de la chaîne moléculaire, sachant qu'elle est indépendante de la longueur  $L$  de la chaîne initiale et qu'elle est due au « pur hasard »?
2. Déterminer la loi de probabilité du vecteur aléatoire  $(L_1, L_2)$
3. En déduire les lois de probabilités des variables aléatoires  $L_1$  et  $L_2$ .
4. On se place dans le cas où

$$f_L(x) = \lambda^2 x e^{-\lambda x} \mathbb{1}_{]0, \infty[}(x).$$

- a) Que peut-on dire du vecteur aléatoire  $(L_1, L_2)$ ?
- b) Déterminer la loi de probabilité de la variable  $Z = \text{Inf}(L_1, L_2)$  représentant la longueur du tronçon le plus court.
- c) Comparer cette loi de probabilité à celle de  $L_1$  et de  $L_2$ .

## Séance 6 - Espérance

**Exercice 1 — Calculs d'espérances et de variances.**

1. Soit  $X$  une variable aléatoire de loi de Bernoulli  $\mathcal{B}(p)$ ,  $p \in ]0; 1[$ .  
Calculer  $\mathbb{E}[X]$  et  $\text{Var}(X)$ .
2. Soit  $S$  une variable aléatoire de loi binomiale  $\mathcal{B}(n, p)$ ,  $p \in ]0; 1[$ ,  $n \in \mathbb{N}^*$ .  
Calculer  $\mathbb{E}[S]$  et  $\text{Var}(S)$ .
3. Soit  $Y$  une variable aléatoire de loi de Poisson  $\mathcal{P}(\lambda)$ ,  $\lambda \in ]0; +\infty[$ .  
Calculer  $\mathbb{E}[Y]$  et  $\text{Var}(Y)$ .
4. Soit  $U$  une variable aléatoire de loi uniforme sur  $[0; 1]$   $\mathcal{U}([0; 1])$ .  
Calculer  $\mathbb{E}[U]$  et  $\text{Var}(U)$ .
5. Soit  $V$  une variable aléatoire de loi uniforme sur  $[a; b]$   $\mathcal{U}([a; b])$ ,  $a, b \in \mathbb{R}$ .  
Calculer  $\mathbb{E}[V]$  et  $\text{Var}(V)$ .
6. Soit  $T$  une variable aléatoire de loi exponentielle  $\mathcal{E}(\lambda)$ ,  $\lambda \in \mathbb{R}^{+*}$ .  
Calculer  $\mathbb{E}[T]$  et  $\text{Var}(T)$ .
7. Soit  $N$  une variable aléatoire de loi normale  $\mathcal{N}(\mu, \sigma^2)$  avec  $\mu \in \mathbb{R}$  et  $\sigma^2 \in \mathbb{R}^+$ .  
Calculer  $\mathbb{E}[N]$  et  $\text{Var}(N)$ .

**Exercice 2 — Une évaluation de la fiabilité.**

On considère un système  $S$  réalisé à partir de deux éléments  $E_1$  et  $E_2$ .  
Plusieurs configurations sont envisageables :

- Connexion série :  $S$  fonctionne si et seulement si  $E_1$  et  $E_2$  fonctionnent.
- Connexion parallèle :  $S$  fonctionne si et seulement si  $E_1$  ou  $E_2$  fonctionne(nt).
- Connexion stand-by :  $S$  fonctionne tant que  $E_1$  fonctionne. Lorsque  $E_1$  tombe en panne,  $E_2$  prend le relais et  $S$  fonctionne tant que  $E_2$  fonctionne.

Sachant que les instants de panne  $T_1$  et  $T_2$  de  $E_1$  et  $E_2$  sont des variables aléatoires exponentielles indépendantes de moyennes respectives  $\nu_1$  et  $\nu_2$ , comparer la fiabilité des trois systèmes en calculant dans chaque cas, l'espérance de la variable aléatoire  $T$  caractérisant l'instant de panne de  $S$ .

**Exercice 3 — Loi de probabilité Log-normale.**

On considère une variable aléatoire  $X$  dont la loi de probabilité est une loi normale  $\mathcal{N}(\mu, \sigma^2)$  de moyenne  $\mu$  et de variance  $\sigma^2$  :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

On effectue le changement de variable :  $Y = \exp(X)$ .

1. Déterminer la loi de probabilité (appelée loi de probabilité Log-Normale de paramètres  $\mu$  et  $\sigma$  de  $Y$ ).
2. Calculer  $\mathbb{E}[Y]$ .
3. Calculer  $Var(Y)$ .
4. Calculer le mode et la médiane de  $Y$ .
5. Quelle est la loi de probabilité de  $Z = z_0 + Y$  (appelée loi de probabilité Log-Normale généralisée de paramètres  $z_0$ ,  $\mu$  et  $\sigma$ ).
6. On fait l'hypothèse que cette loi de probabilité modélise le revenu mensuel d'une personne prise au hasard dans une population observée donnée.

Déterminer les valeurs des paramètres  $z_0$ ,  $\mu$  et  $\sigma$  sachant que sur les 1000 personnes observées :

- la moitié a des revenus inférieurs à 2000€ ;
- 159 gagnent plus de 3000€ ;
- 159 gagnent moins de 1500€.

## Séance 7 - Loi normale et vecteurs gaussiens

**Exercice 1** — Loi de probabilité Log-normale.

On considère une variable aléatoire  $X$  dont la loi de probabilité est une loi normale  $\mathcal{N}(\mu, \sigma^2)$  de moyenne  $\mu$  et de variance  $\sigma^2$  :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

On effectue le changement de variable :  $Y = \exp(X)$ .

1. Déterminer la loi de probabilité (appelée loi de probabilité Log-Normale de paramètres  $\mu$  et  $\sigma$  de  $Y$ ).
2. Calculer  $\mathbb{E}[Y]$ .
3. Calculer  $Var(Y)$ .
4. Calculer le mode et la médiane de  $Y$ .
5. Quelle est la loi de probabilité de  $Z = z_0 + Y$  (appelée loi de probabilité Log-Normale généralisée de paramètres  $z_0$ ,  $\mu$  et  $\sigma$ ).
6. On fait l'hypothèse que cette loi de probabilité modélise le revenu mensuel d'une personne prise au hasard dans une population observée donnée.

Déterminer les valeurs des paramètres  $z_0$ ,  $\mu$  et  $\sigma$  sachant que sur les 1000 personnes observées :

- la moitié a des revenus inférieurs à 2000€;
- 159 gagnent plus de 3000€;
- 159 gagnent moins de 1500€.

**Exercice 2**

Considérons le vecteurs gaussien  $X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$  de moyenne  $m_X = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$  et de matrice

de covariance  $\Sigma_X = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}$ .

1. Quelles sont les lois de  $X_2$  et de  $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$ ?
2.  $X_2$  et  $X_3$  sont-elles indépendantes?
3. Quelle est la loi du vecteur  $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$  avec  $\begin{cases} Y_1 = X_2 + X_3 \\ Y_2 = X_1 + X_2 \end{cases}$  ?
4.  $Y_1$  et  $Y_2$  sont-elles indépendantes?

### Exercice 3

Considérons le vecteurs gaussien centré  $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$  tel que

- $\mathbb{E}[X^2] = 4$ ;
- $\mathbb{E}[Y^2] = 1$ ;
- $2X + Y$  et  $X - 3Y$  sont indépendantes.

1. Quelle est la loi de  $Z$ ?
2. Quelle est la loi de  $\begin{pmatrix} X + Y \\ 2X - Y \end{pmatrix}$ ?

### Exercice 4

Considérons deux variables aléatoires (v.a.) indépendantes  $X$  et  $Y$ .

- $X$  est une v.a. normale centrée et réduite,
- $Y$  est une v.a. discrète telle que

$$P(Y = 1) = P(Y = -1) = 0,5.$$

1. Quelle est la loi de  $Z = XY$ ?
2. Calculer  $\text{Cov}(X, Z)$ .
3.  $\begin{pmatrix} X \\ Z \end{pmatrix}$  est-il un vecteur gaussien?
4.  $X$  et  $Z$  sont-elles indépendantes?

## Séance 8 - Espérance conditionnelle

### Exercice 1 — Variables Discrètes.

Soit  $(X, Y)$  un couple de variables aléatoires discrètes de loi :

$$P\{X = m, Y = n\} = \frac{e^{-2\lambda} \lambda^m}{n!(m-n)!} \mathbb{1}_{\{m \geq n\}}, \quad (m, n) \in \mathbb{N}^2,$$

où  $\lambda$  est un nombre réel positif donné.

Déterminer les lois conditionnelles et les espérances conditionnelles  $\mathbb{E}[Y/X]$  et  $\mathbb{E}[X/Y]$ .

### Exercice 2 — Couple de Poisson.

Soient  $X_1$  et  $X_2$  deux variables aléatoires indépendantes de loi de Poisson de paramètres respectifs  $\lambda_1$  et  $\lambda_2$  réel strictement positifs. On pose  $Y = X_1 + X_2$ .

Calculer  $\mathbb{E}[Y/X_1]$ .

### Exercice 3 — Variables continues.

On considère un couple  $(X, Y)$  de variables aléatoires réelles continues de densité :

$$f_{X,Y}(x, y) = \frac{1}{\pi} e^{-y(1+x^2)} \mathbb{1}_{\mathbb{R}^{+*}}(y), \quad (x, y) \in \mathbb{R}^2.$$

1. Calculer les lois marginales.
2. Calculer les lois conditionnelles et les reconnaître.
3. En déduire les espérances conditionnelles  $\mathbb{E}[Y/X]$  et  $\mathbb{E}[X/Y]$ .

### Exercice 4 — Estimation d'un signal.

On suppose qu'un signal d'intensité  $s$  émis en  $A$  est reçu en  $B$  avec une intensité normalement distribuée, de moyenne  $s$  et d'écart-type  $\sigma_B$ . On modélise donc le signal émis en  $A$  par une variable aléatoire normale  $S$  de moyenne  $m$  et d'écart-type  $\sigma_A$  et le signal observé en  $B$  par une variable aléatoire  $R$ .

1. Déterminer la meilleure, approximation  $\phi(R)$  de  $S$  au sens de l'erreur quadratique moyenne.
2. Calculer  $\mathbb{E}[\phi(R)]$  et en déduire  $\mathbb{E}[R]$ .
3. Calculer les expressions limites de l'estimateur  $\phi(R)$  lorsque  $\sigma_A \gg \sigma_B$  et  $\sigma_A \ll \sigma_B$ .

**Exercice 5 — Couple gaussien.**

Soit  $(X, Y)$  un couple gaussien centré de matrice de covariance  $\Sigma = \begin{pmatrix} 4/3 & -1 \\ -1 & 1 \end{pmatrix}$ .

1. Calculer  $E[X/Y - X]$ .
2. En déduire la loi de  $E[X/Y - X]$ .

## Séance 9 - Fonctions caractéristiques et convergence

**Exercice 1 — Fonctions caractéristiques.**

1. Déterminer les fonctions caractéristiques des lois de probabilité suivantes :
  - (a) binomiale de paramètres  $n$  et  $p$ ,
  - (b) poisson de paramètre  $\lambda$ ,
  - (c) exponentielle de paramètre  $\lambda$ ,
  - (d) normale centré réduite, puis  $\mathcal{N}(\mu, \sigma)$ ,
  - (e) uniforme sur  $] - 1, 1[$ .
2. Soit  $X$  et  $Y$  deux variables aléatoires indépendantes de fonctions caractéristiques  $\phi_X$  et  $\phi_Y$ .
  - (a) Déterminer la fonction caractéristique de  $aX + b$  où  $a, b \in \mathbb{R}$ .
  - (b) Déterminer  $\phi_{X+Y}$  en fonction de  $\phi_X$  et  $\phi_Y$ .
3. En déduire les lois des v.a. suivantes :
  - (a)  $X+Y$  où  $X$  et  $Y$  sont indépendantes de lois respectives  $\mathcal{B}(m, p)$  et  $\mathcal{B}(n, p)$ ,
  - (b)  $X+Y$  où  $X$  et  $Y$  sont indépendantes de lois respectives  $P(\lambda)$  et  $P(\mu)$ ,
  - (c)  $\sigma X + m$  où  $X$  suit une loi  $\mathcal{N}(0, 1)$ ,
  - (d)  $X - Y + 2Z$  où  $X, Y$  et  $Z$  sont indépendantes de loi  $\mathcal{N}(0, 1)$ .

**Exercice 2 — Méthode de Monte-Carlo.** 1. Soit  $X_n$  une suite de variables aléatoires indépendantes suivant des lois uniformes sur  $[0, 1]$  et une fonction  $f$  mesurable bornée sur  $[0, 1]$  (une fonction continue par exemple). Calculer pour la convergence presque-sûre

$$\lim_n \frac{f(X_1) + \dots + f(X_n)}{n}$$

2. En tirant des points uniformément dans un carré  $[0, 1]^2$  trouver une méthode pour approximer  $\frac{\pi}{4}$  et démontrer sa validité.

**Exercice 3** Soient  $X$  et  $Y$  deux variables aléatoires indépendantes de carré intégrable, de même loi et centrée. Soit  $\phi$  leur fonction caractéristique commune. On suppose que la variable aléatoire  $\frac{X+Y}{\sqrt{2}}$  a même loi que  $X$  et  $Y$ . Montrer que ces variables sont nécessairement de loi normale.

**Exercice 4 — Application non probabiliste d'un théorème probabiliste.** On rappelle que  $X$  suit une loi de Poisson de paramètre  $\lambda > 0$  noté  $\mathcal{P}(\lambda)$  si pour tout  $k \in \mathbb{N}$   $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ . On a  $\mathbb{E}[X] = \lambda$  et  $\text{Var}[X] = \lambda$ .

1. Soit  $X$  et  $Y$  deux variables aléatoires indépendantes de loi Poisson de paramètre respectif  $\lambda$  et  $\mu$ . On pose  $Z = X + Y$ . Calculer  $\mathbb{P}(Z = k)$  et en déduire la loi de  $Z$ .
2. Soient  $X_1, \dots, X_n$   $n$  variables aléatoires indépendantes de loi de poisson de paramètre 1. Donner la loi de  $S_n = X_1 + \dots + X_n$  et préciser  $\mathbb{P}(S_n = k)$ . Calculer  $\mathbb{P}(S_n \leq n)$ .

3. D'après un théorème du cours donner la limite en loi de  $Z_n = \frac{S_n - n}{\sqrt{n}}$ . On notera la variable aléatoire limite  $Z$ . Donner la limite de  $\mathbb{P}(Z_n \leq 0)$  quand  $n$  tend vers l'infini.
4. En déduire la limite de  $e^{-n} \sum_{k=0}^n \frac{n^k}{k!}$ .

## Séance 10 - Convergence de variables aléatoires

### Exercice 1 — Sondage présidentiel.

Une semaine avant le second tour d'une élection présidentielle, on effectue un sondage dans une ville réputée représentative de l'opinion nationale.

Sur les 10000 électeurs de cette ville, 5002 se sont prononcés en faveur du candidat  $A$ .

Sachant que la population nationale comporte 42 millions d'électeurs, calculer la probabilité pour que  $A$  soit élu.

### Exercice 2 — Rando proba.

En terrain plat, un pas mesure en moyenne 80 cm avec un écart-type de 4 cm.

A partir d'un même point, Jane et John font chacun 1000 pas dans la même direction.

Quelle est la probabilité pour qu'à l'arrivée, la différence entre les distances parcourues par les deux amis soit inférieure à un mètre?

### Exercice 3 — Proba et nombres dyadiques.

Considérons une suite  $X_n$  de variables aléatoires discrètes indépendantes :

$$X_n \in \left\{0; \frac{1}{2^n}\right\} \text{ avec } P(X_n = 0) = P\left(X_n = \frac{1}{2^n}\right) = \frac{1}{2}.$$

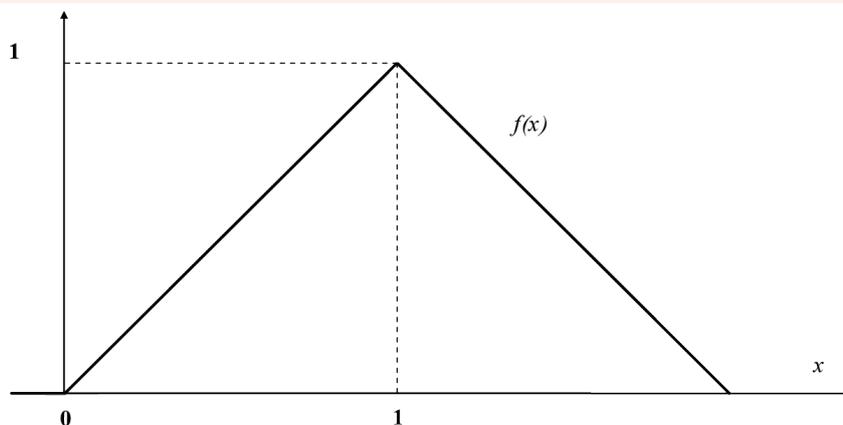
On définit alors la suite de variables aléatoires :

$$U_n = \sum_{k=1}^n X_k.$$

1. Montrer que la suite de variables aléatoires  $U_n$  **converge en loi** vers une variable aléatoire  $U$  de loi de probabilité uniforme sur  $[0,1]$ .
2.  $U_n$  converge-t-elle **presque sûrement** vers une variable aléatoire  $U$  de loi uniforme sur  $[0,1]$ ?
3. Interpréter et commenter ce résultat.

**Exercice 4 — Densité triangle.**

Une variable aléatoire continue  $X$  a une densité  $f(x)$  représentée graphiquement ci-dessous.



1. Calculer et tracer la fonction de répartition de  $X$ .
2. Déterminer sa fonction réciproque et en déduire un procédé de simulation de réalisations de  $X$ .
3. Soient  $U$  et  $V$  deux v.a.i.i.d de loi uniforme sur  $[0;1]$ .  
Quelle est la loi de  $Y = U + V$ ?
4. Proposer une autre méthode de simulation de réalisations de  $X$ .

## Approfondissement - Simulation de variables aléatoires

### Exercice 1

De nombreuses simulations exigent la génération d'un couple de variables aléatoires gaussiennes  $(X, Y)$  indépendantes, d'espérances mathématiques respectives  $\mu_X$  et  $\mu_Y$ , et d'écart types respectifs  $\sigma_X$  et  $\sigma_Y$ .

On propose ici une méthode qui nécessite un générateur de variables aléatoires indépendantes de loi de probabilité uniforme sur  $[0;1]$ .

1. Soient donc  $X$  et  $Y$  des variables aléatoires indépendantes de lois de probabilité respectives  $\mathcal{N}(\mu_X, \sigma_X^2)$  et  $\mathcal{N}(\mu_Y, \sigma_Y^2)$ , et des variables aléatoires  $\Theta$  et  $R$  définies par :

$$X = \mu_X + \sigma_X R \cos(\Theta) \quad \text{et} \quad Y = \mu_Y + \sigma_Y R \sin(\Theta)$$

avec  $R > 0$  et  $\theta \in [0; 2\pi[$ .

Déterminer les densités de probabilités  $f_R(r)$  et  $f_\Theta(\theta)$  après avoir calculé  $f_{R,\Theta}(r, \theta)$ .

2. Quelles remarques peut-on faire sur ces lois de probabilité?
3. Les variables aléatoires  $R$  et  $\Theta$  sont elles indépendantes?
4. Déterminer des changements de variables aléatoires  $R = h(U)$  et  $\Theta = g(V)$ ,  $U$  et  $V$  étant des variables aléatoires indépendantes de loi de probabilité uniforme sur  $[0;1]$ , de telle façon que le couple de variables aléatoires  $(R, \Theta)$  ait la densité de probabilité  $f_{R,\Theta}(r, \theta)$  obtenue à la question 1.
5. En déduire que les variables aléatoires définies par :

$$X = \mu_X + \sigma_X \sqrt{-2\ln(U)} \cos(2\pi V) \quad \text{et} \quad Y = \mu_Y + \sigma_Y \sqrt{-2\ln(U)} \sin(2\pi V)$$

satisfont l'objectif visé.

6. En déduire une méthode de génération d'une suite de variables aléatoire  $X_n$  indépendantes et de loi normale  $\mathcal{N}(\mu, \sigma^2)$  à partir d'une suite de variables  $U_n$  indépendantes et de loi de probabilité uniforme sur  $[0;1]$ .

### Exercice 2

Considérons la loi de Cauchy de densité définie sur  $\mathbb{R}$  par :

$$\forall x \in \mathbb{R}, \quad g(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

1. Vérifier que  $g$  est bien une densité de probabilité et donner l'expression de sa fonction de répartition  $G$ .
2. En déduire une méthode de simulation de réalisations de cette loi.
3. Soit  $f$  la densité de la loi normale centrée réduite. Montrer que l'on peut utiliser  $g$  pour simuler des réalisations de la loi normale centrée réduite par la méthode d'acceptation-rejet.  
(Pour cela on cherchera  $M$  tel que  $\forall x \in \mathbb{R}, f(x) \leq Mg(x)$  et on exhibera, de préférence, la valeur de  $M$  minimisant la probabilité de rejet).
4. Proposer une méthode permettant de simuler des réalisations d'un mélange de deux variables aléatoires de lois normales  $\mathcal{N}(\mu_1, \sigma_1^2)$  et  $\mathcal{N}(\mu_2, \sigma_2^2)$  en proportions égales.

### Exercice 3

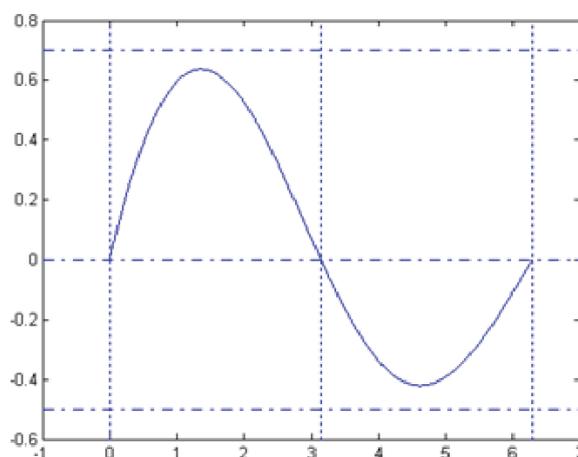
L'objectif de cet exercice est de proposer une méthode stochastique de calcul numérique de l'intégrale :

$$I = \int_0^{2\pi} \frac{\sin(x)}{\sqrt{1+x}} dx.$$

La représentation graphique de la fonction

$$f(x) = \frac{\sin(x)}{\sqrt{1+x}}, \quad x \in [0; 2\pi]$$

est donnée ci-dessous.



1. Exprimer  $I$  sous la forme d'une espérance.
2. Proposer une méthode de Monte-Carlo permettant le calcul de  $I$  et détailler les différentes étapes du calcul.

## Soutien

### Exercice 1

On considère un groupe de personnes. Parmi ces personnes figurent des médecins (8%), des profs de maths (15%), et des gens normaux (77%).

C'est bien connu, les profs de maths écrivent très mal mais dans ce domaine ils sont battus par les médecins, qui écrivent encore plus mal. Ainsi 50% des profs de maths écrivent de façon illisible, mais cette proportion atteint 70% des médecins. Quand aux gens normaux, seuls 30% écrivent de façon incompréhensible.

On considère un mot écrit par une personne du groupe.

1. Quelle est la probabilité que ce mot soit illisible ?
2. Le mot est en effet illisible. Quelle est la probabilité qu'il ait été écrit par une main noble ? (c'est à dire par un prof de math)

### Exercice 2

On considère une maladie dont est atteinte 1% de la population.

Si on est malade, on meurt avec la probabilité 0.5. Il existe un traitement contre la maladie, qui fait qu'un individu malade et traité n'a plus que 10% de chances de mourir. Le test de dépistage permet de détecter 80% des malades, mais désigne aussi à tort 3% de personnes saines. Or si une personne saine est traitée, elle meurt dans 2% des cas.

1. Si on n'effectue aucun test de dépistage, quelle est la probabilité de mourir de cette maladie ?
2. On décide de procéder à un dépistage généralisé et à un traitement des individus désignés comme malades. Quelle est la probabilité de mourir dans ce cas ? (à cause de la maladie ou du traitement)

### Exercice 3

Un quart d'une population a été vaccinée. Si on est vacciné, on tombe malade avec une probabilité de  $1/20$ . Parmi les malades, il y a 4 non-vaccinés pour 1 vacciné. Quelle est la probabilité pour un non-vacciné de tomber malade ?

### Exercice 4

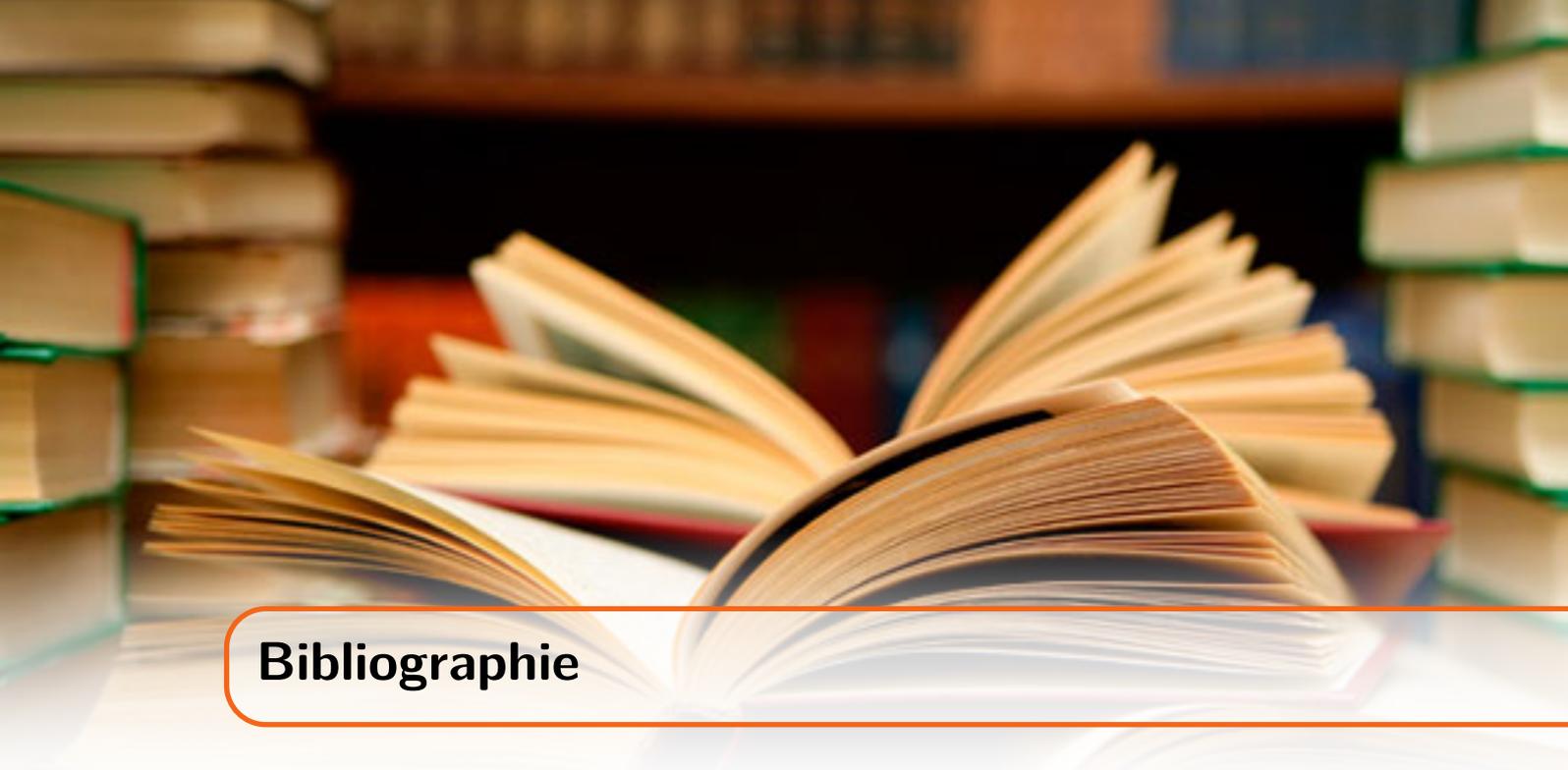
Jane doit se rendre à la poste. Le temps qu'elle passe à attendre dans la queue suit une loi exponentielle de paramètre  $l_1$ . Le temps qu'elle passe au guichet suit une loi exponentielle de paramètre  $l_2$ .

Jane attend un appel téléphonique très important de son voisin John. Le temps qu'elle doit attendre avant que son téléphone ne sonne suit une loi exponentielle de

paramètre  $l_3$ . On considère que ces trois phénomènes aléatoires sont indépendants.  
Quelle est la probabilité que Jane ne puisse pas répondre à l'appel de John parce qu'elle sera encore à la poste?

#### Exercice 5

Soient  $X$  et  $Y$  deux variables aléatoires indépendantes suivant des lois exponentielle de paramètres respectifs  $a$  et  $b$ .  
Quelle est la loi de  $Z = X + Y$ ?



## Bibliographie

### Livres

- BARBE, P. et M. LEDOUX (2007). *Probabilité*. Collection Enseignement sup. Mathématiques. EDP Sciences. ISBN : 9782868839312 (cf. page 20).
- BERTOIN, J. (2008). *Probabilités : Cours de licence de mathématiques appliquées*. Université Paris VI (cf. page 15).
- BILLINGSLEY, P. (1995). *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley. ISBN : 9780471007104 (cf. page 20).
- LE GALL, J.F. (pas de date). *Intégration, probabilités et processus aléatoire*. Ecole Normale Supérieure (cf. page 20).
- MÉLÉARD, S. (2010). *Aléatoire : Introduction à la théorie et au calcul des probabilités*. Mathématiques appliquées. Ecole Polytechnique. ISBN : 9782730215756 (cf. page 7).

### Articles

- DELLACHERIE, C. (1994). "Pascal et Fermat. La naissance du calcul des probabilités." In : Actes (cf. page 15).

