

INSTITUT D'OPTIQUE - Graduate School  
Première Année

---

Notes de cours de Mathématiques et Signal  
Probabilités et Variables Aléatoires

Matthieu Boffety  
`matthieu.boffety@institutoptique.fr`

Version du 24 octobre 2023

# Table des matières

<b>1</b>	<b>Probabilités</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	Définitions historiques . . . . .	1
1.2.1	Probabilité « naturelle » . . . . .	2
1.2.2	Fréquences relatives . . . . .	4
1.2.3	Approche axiomatique . . . . .	4
1.3	Théorie des probabilités . . . . .	5
1.3.1	Définitions élémentaires . . . . .	5
1.3.2	Probabilités d'éventualités finies . . . . .	7
1.3.3	Probabilités conditionnelles et événements indépendants . . . . .	7
<b>2</b>	<b>Variables Aléatoires</b>	<b>10</b>
2.1	Définitions élémentaires . . . . .	10
2.2	Espérance, variance et autres moments . . . . .	13
2.2.1	Espérance . . . . .	13
2.2.2	Variance et écart-type . . . . .	16
2.2.3	Moments d'ordres supérieurs . . . . .	17
2.2.4	Médiane . . . . .	18
2.2.5	Mode . . . . .	18
2.3	Quelques lois classiques . . . . .	18
2.4	Fonction d'une variable aléatoire . . . . .	21
2.4.1	Changement de variable aléatoire . . . . .	21
2.4.2	Fonction caractéristique et fonction génératrice . . . . .	23
<b>3</b>	<b>Ensembles de variables aléatoires</b>	<b>25</b>
3.1	Couples de variables aléatoires . . . . .	25
3.1.1	Définitions . . . . .	25
3.1.2	Couples discrets . . . . .	26
3.1.3	Couples (absolument) continus . . . . .	27
3.1.4	Covariance . . . . .	27
3.1.5	Indépendance . . . . .	28
3.2	Fonction d'un couple de variables aléatoires . . . . .	29
3.2.1	Image d'un couple de variables aléatoires . . . . .	29
3.2.2	Somme de variables aléatoires . . . . .	30
3.3	Vecteurs aléatoires . . . . .	32
3.3.1	Définitions . . . . .	32
3.3.2	Espérance et covariance . . . . .	33
3.3.3	Fonction d'un vecteur aléatoire . . . . .	34
3.3.4	Indépendance . . . . .	35
<b>4</b>	<b>VA et vecteurs aléatoires gaussiens</b>	<b>37</b>
4.1	Vecteur aléatoire gaussien . . . . .	37

4.1.1	Définition . . . . .	37
4.1.2	Densité de probabilité . . . . .	38
4.1.3	Indépendance et corrélation . . . . .	39
4.2	Espérance conditionnelle . . . . .	40
4.2.1	Définitions . . . . .	40
4.2.2	Conditionnement de vecteurs aléatoires gaussiens . . . . .	41
<b>5</b>	<b>Théorèmes limites et comportements asymptotiques</b>	<b>42</b>
5.1	Modes de convergence . . . . .	42
5.2	Lois des grands nombres . . . . .	43
5.3	Théorème central limite . . . . .	44
5.4	Comportements asymptotiques . . . . .	44
5.4.1	Loi de Poisson comme limite de loi binomiale . . . . .	44
5.4.2	Approximation normale d'une loi binomiale . . . . .	45
5.4.3	Approximation normale d'une loi de Poisson . . . . .	45
5.4.4	Résumé des approximations . . . . .	45
<b>A</b>	<b>Quelques rappels mathématiques</b>	<b>47</b>
A.1	Rappels sur la théorie des ensembles . . . . .	47
A.2	Dénombrément . . . . .	50
<b>B</b>	<b>Lois usuelles</b>	<b>53</b>
<b>C</b>	<b>La transformée de Fourier dans le cours de variables aléatoires</b>	<b>54</b>
C.1	Résultats classiques . . . . .	54
C.2	TF de fonctions classiques . . . . .	55
<b>D</b>	<b>Exercices d'entraînement</b>	<b>56</b>
D.1	Exercices sur les Probabilités . . . . .	56
D.2	Exercices sur les Variables aléatoires . . . . .	57
<b>E</b>	<b>Correction des exercices</b>	<b>59</b>
E.1	Exercices sur les Probabilités . . . . .	59
E.2	Exercices sur les variables aléatoires . . . . .	63
	<b>Bibliographie</b>	<b>64</b>

# Chapitre 1

## Probabilités

Le contenu de ce chapitre repose principalement sur les notes rédigées par Gaëtan Messin entre 2009 et 2015 et sur les références [1, 2, 4].

### 1.1 Introduction

La théorie des probabilités repose sur l'*observation* que la *moyenne* de certains résultats d'expériences approche une valeur constante lorsque le nombre d'observations augmente et que cette valeur reste la même si la moyenne est évaluée sur une partie arbitraire de la séquence d'observations.

Par exemple, le jet répété d'un dé conduit à l'observation d'un résultat pair dans environ 50% des cas pour toute séquence suffisamment longue ou toute partie suffisamment longue de cette séquence.

La théorie des probabilités trouve évidemment des applications aux jeux de hasard, mais également dans de nombreux domaines dans lesquels de très nombreux phénomènes entrent en jeu séquentiellement ou simultanément : l'émission d'électrons ou de photons, la radioactivité, le contrôle qualité, la mécanique statistique, les pannes, les turbulences, le bruit, les naissances et les décès, les phénomènes de files d'attente, etc.

Le but de la théorie des probabilités est de décrire les moyennes observées en termes de probabilités. La probabilité  $\mathbb{P}$  d'un événement  $\mathcal{A}$  est un nombre associé à cet événement qui peut s'interpréter de la façon suivante :

*Si l'on répète  $n$  fois l'expérience considérée et que l'événement  $\mathcal{A}$  se produit  $n_{\mathcal{A}}$  fois, alors, avec un très grand degré de certitude, la fréquence relative  $n_{\mathcal{A}}/n$  d'occurrence de  $\mathcal{A}$  est proche de  $\mathbb{P}(\mathcal{A})$ , soit :*

$$\mathbb{P}(\mathcal{A}) \simeq \frac{n_{\mathcal{A}}}{n}, \quad (1.1)$$

*pour peu que  $n$  soit suffisamment grand.*

Cette interprétation est évidemment floue, mais cela ne peut être évité et résulte du fait que la probabilité provient de l'*expérience physique*. Il ne s'agit donc que d'un modèle imparfait destiné à décrire la réalité. Cela étant, la *théorie* des probabilités est une discipline développée logiquement à partir d'une série d'axiomes, dans le cadre des mathématiques, et peut être appliquée avec succès à des problèmes réels.

### 1.2 Définitions historiques

Plusieurs approches peuvent être utilisées pour développer la théorie des probabilités. Les trois principales voies reposent sur l'interprétation des probabilités en termes de fréquences d'occurrence relatives, sur l'interprétation en termes de rapport entre nombre de cas favorables et nombre de cas possibles

ou encore sur une description basée sur la théorie des ensembles et la théorie de la mesure. C'est cette dernière voie que nous allons suivre dans la suite, néanmoins nous allons donner un aperçu de chacune de ces approches.

### 1.2.1 Probabilité « naturelle »

L'idée est d'attribuer à un événement  $\mathcal{A}$  une probabilité *a priori* définie par :

$$\mathbb{P}(\mathcal{A}) = \frac{N_{\mathcal{A}}}{N},$$

où  $N$  est le nombre de résultats possibles que peut donner un tirage et  $N_{\mathcal{A}}$  est le nombre de résultats favorables associés à  $\mathcal{A}$ , c'est-à-dire le nombre de résultats possibles de tirage pour lesquels l'événement  $\mathcal{A}$  est réalisé.

Cette approche historique fut la première développée et reste largement utilisée. Elle s'appuie sur des éléments d'analyse combinatoire (cf. Dénombrement p. 50).

Suivant cette interprétation des probabilités, pour un lancer de dé, la probabilité d'avoir un résultat pair vaut donc *a priori* un demi. En effet, c'est le rapport du nombre de cas favorables, c'est-à-dire le cardinal de l'ensemble  $\{\square, \boxtimes, \boxplus\}$ , par le nombre de cas possibles, c'est-à-dire le cardinal de l'ensemble  $\{\square, \square, \square, \boxtimes, \boxtimes, \boxplus\}$ .

Cette approche, malgré sa simplicité, pose un certain nombre de problèmes, comme nous allons le voir au travers de quelques exemples.

#### Exemple 1 : un exemple caricatural

On jette deux dés et on cherche à déterminer la probabilité que la somme des valeurs obtenues soit égale à 7. Plusieurs raisonnements sont envisageables et ne conduisent pas forcément au même résultat. En voici trois :

- On considère que les résultats possibles pour la somme des valeurs de deux dés sont les nombres entiers de 2 à 12, soit 11 cas possibles. Un seul cas est favorable, c'est une somme égale à 7. La probabilité est donc  $p = 1/11$ .
- On peut raisonner sur les paires de résultats possibles (tels que  $\boxtimes$  et  $\boxplus$ , par exemple). Il y a alors 21 cas possibles dont 3 donnent comme somme la valeur 7 ( $\square\boxtimes, \square\boxplus, \boxtimes\square$ ). La probabilité est donc  $p = 3/21 = 1/7$ .
- Enfin, et c'est le raisonnement correct, on considère que les résultats possibles sont tous les couples ordonnés (comme  $(\boxtimes, \boxplus)$  par exemple). Il y a donc 36 possibilités, dont 6 donnent comme somme 7. La probabilité vaut donc  $p = 6/36 = 1/6$ .

#### Exemple 2 : les Grands se trompent aussi

Dans l'article *Croix ou pile (analyse des hasards)* de l'*Encyclopédie* de Diderot et D'Alembert, ce dernier se demandait quelle était la probabilité d'obtenir au moins une fois « face » (ou « croix ») en lançant au maximum deux fois une pièce non biaisée. Il tenait le raisonnement suivant :

*On demande combien il y a à parier qu'on amènera croix en jouant deux coups consécutifs. La réponse qu'on trouvera dans les auteurs, & suivant les principes ordinaires, est celle-ci. Il y a quatre combinaisons : {Croix, Croix}, {Pile, Croix}, {Croix, Pile}, {Pile, Pile} De ces quatre combinaisons, une seule fait perdre & trois font gagner ; il y a donc 3 contre 1 à parier en faveur du joueur qui jette la pièce. [...]*  
*Cependant cela est-il bien exact ? Car, [...] ne faut-il pas réduire à une les deux combinaisons qui donnent croix au premier coup ? Car, dès qu'une fois croix est venu, le jeu est fini, & le second coup est compté pour rien.*

Ainsi, il n'y a proprement que trois combinaisons de possibles :  $\{Croix\}$ , premier coup ;  $\{Pile, Croix\}$ , premier & second coup ;  $\{Pile, Pile\}$ , premier & second coup ; Donc il n'y a que 2 contre 1 à parier. [...]

Ceci est digne, ce me semble, de l'attention des calculateurs, & irait à réformer bien des règles unanimement reçues sur les jeux de hasard.

Néanmoins, Joseph Bertrand affirme en 1889 que

*L'esprit de d'Alembert, habituellement juste et fin, déraisonnait complètement sur le Calcul des probabilités.*

**L'explication**

L'erreur commise dans les deux exemples ci-dessus est de considérer à chaque fois que les événements ont le même nombre de chances de se produire. Cette observation conduit à une définition plus précise de la probabilité :

La probabilité d'un événement  $\mathcal{A}$  vaut :

$$\mathbb{P}(\mathcal{A}) = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}} ,$$

à condition que tous les cas possibles soient équiprobables.

Le principe sous-jacent à cette définition est que le choix qui consiste à considérer comme équiprobables tous les événements possibles est le plus raisonnable en l'absence de plus amples informations. En fait, le principe en question est équivalent au principe de maximum d'entropie dans le cadre de la théorie de l'information dont vous verrez une illustration dans le cours de *Traitement du Signal*.

**Le paradoxe de Bertrand**

On sent donc bien dans les exemples précédents qu'il n'est pas toujours évident de savoir ce que  $N_{\mathcal{A}}$  et  $N$  valent. Un autre exemple emblématique de cette difficulté est le fameux *Paradoxe de Bertrand*. Ce problème a été proposé par Joseph Bertrand en 1889 dans son ouvrage *Calcul des Probabilités*.

Soit un cercle  $(C)$  de rayon  $r$  et  $[A, B]$  une corde de longueur  $\ell$  de ce cercle choisie au hasard. Quelle est la probabilité que  $\ell \geq \sqrt{3}r$ ? Autrement dit, quelle est la probabilité que la corde  $[A, B]$  coupe le cercle  $(C')$  de rayon  $r/2$  concentrique à  $(C)$  ?

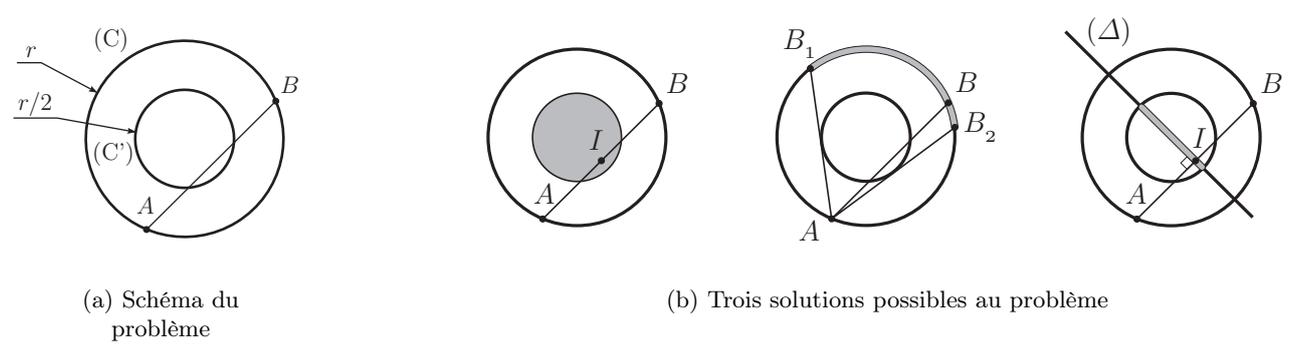


FIGURE 1.1 – Paradoxe de Bertrand : Quelle est la probabilité d'une corde  $[A, B]$  de  $(C)$  choisie au hasard coupe  $(C')$  ?

Trois raisonnements distincts conduisent à trois conclusions différentes.

- a) On considère le point  $I$  milieu du segment  $[A, B]$ . La probabilité que  $\ell \geq \sqrt{3}r$  est la probabilité que le point  $I$  soit dans  $(C')$ , l'ensemble des possibles étant l'intérieur de  $(C)$ . Ainsi :

$$\mathbb{P}(\{\ell \geq \sqrt{3}r\}) = \frac{\text{Aire}(C')}{\text{Aire}(C)} = \frac{1}{4} .$$

b) Etant donné un point  $A$  choisi arbitrairement sur le cercle  $(C)$ , la longueur  $\ell$  ne dépend que du choix de  $B$ . Soient  $B_1$  et  $B_2$  les deux intersections entre le cercle  $(C)$  et les deux droites issues de  $A$  et tangentes à  $(C')$ . La probabilité que  $\ell \geq \sqrt{3}r$  est la probabilité que  $B$  soit sur l'arc le plus court délimité par les points  $B_1$  et  $B_2$ , l'ensemble des possibles étant toute la circonférence de  $(C)$ . Ainsi :

$$\mathbb{P}(\{\ell \geq \sqrt{3}r\}) = \frac{\text{Longueur}(\widehat{B_1 B_2})}{2\pi r} = \frac{1}{3}.$$

c) Considérant la droite  $(\Delta)$ , médiatrice de la corde  $[AB]$  et le milieu  $I$  de  $[AB]$ , la probabilité que  $\ell \geq \sqrt{3}r$  est la probabilité que le point  $I$  soit dans le segment de  $(\Delta)$  enfermé dans  $(C')$ , l'ensemble des possibles étant le segment de  $(\Delta)$  enfermé dans  $(C)$ . Ainsi :

$$\mathbb{P}(\{\ell \geq \sqrt{3}r\}) = \frac{r}{2r} = \frac{1}{2}.$$

En réalité, ces trois cas correspondent à trois situations physiques différentes pour lesquelles l'équiprobabilité porte sur la position d'un point sur une surface, sur un arc ou sur un segment. En l'absence de précision, il n'est pas possible de décider quelle est la réponse correcte.

### 1.2.2 Fréquences relatives

Dans cette approche, la probabilité d'un événement  $\mathcal{A}$  est la limite :

$$\mathbb{P}(\mathcal{A}) = \lim_{n \rightarrow \infty} \frac{N_{\mathcal{A}}(n)}{n},$$

où la suite  $N_{\mathcal{A}}(n)$  représente le nombre d'occurrences de  $\mathcal{A}$  pour  $n$  tirages. Ainsi, la probabilité est posée comme *a priori* égale à la limite de la fréquence d'occurrence relative de l'événement considéré pour une infinité de tirages.

Evidemment, il existe une infinité de suites qui représentent chacune une séquence infinie de tirages possibles. Cela conduit à la manipulation d'objets mathématiques qui sont des classes d'équivalences de suites infinies représentant toutes les séquences possibles associées à la probabilité  $\mathbb{P}(\mathcal{A})$  comprise comme la limite de la fréquence relative d'apparition de  $\mathcal{A}$  lors d'une infinité de tirages.

Bien qu'appréciée des physiciens pour son caractère intuitif, l'approche par fréquence relative (ou *frequentism* en anglais) reste finalement peu utilisée en raison de sa complexité. En pratique on ne réalise jamais une infinité de tirages. De même, certaines situations ne permettent d'accéder qu'à un seul « tirage », ou ne permettent pas du tout d'identifier la notion de probabilité à celle de fréquence d'occurrence.

### 1.2.3 Approche axiomatique

Cette approche s'appuie sur des concepts de théorie des ensembles (voir les rappels du chapitre suivant) et de théorie de la mesure (voir le cours de François Goudail). Elle repose sur trois (et seulement trois) postulats :

- la probabilité d'un événement  $\mathcal{A}$  est un nombre non négatif assigné à cet événement :  $\mathbb{P}(\mathcal{A}) \geq 0$ ,
- la probabilité de l'événement certain  $\mathcal{S}$  vaut  $\mathbb{P}(\mathcal{S}) = 1$ ,
- si deux événements  $\mathcal{A}$  et  $\mathcal{B}$  sont mutuellement exclusifs alors  $\mathbb{P}(\mathcal{A} \cup \mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B})$ .

On la doit à Andrey Nikolaevich Kolmogorov en 1933. Cette approche est donc relativement récente. Néanmoins, elle est aujourd'hui considérée comme supérieure aux deux autres et se retrouve largement présentée dans la plupart des ouvrages mathématiques. C'est cette approche qui sera développée dans la suite.

## 1.3 Théorie des probabilités

Si on peut faire remonter les prémices de l'axiomatisation de la théorie des probabilités à la fin du 19<sup>ème</sup> siècle, c'est à A. N. Kolmogorov que l'on doit la forme moderne de sa formulation mathématique. Néanmoins, il faut garder à l'esprit que la théorie des probabilités donne avant tout une interprétation d'observations expérimentales et qu'elle repose sur des hypothèses empiriques que la formulation mathématique choisit comme axiomes. Ces axiomes ne sont pertinents que dans la mesure où ils décrivent la réalité de manière satisfaisante.

### 1.3.1 Définitions élémentaires

Dans la section précédente, nous utilisons les concepts d'*événement* et de *probabilité* de manière intuitive. Néanmoins, le développement d'une théorie mathématique des probabilités nécessite de définir clairement ces concepts et les notions qui y sont liées.

#### Epreuves et univers des possibles

La théorie des probabilité nécessite d'utiliser un ensemble, noté ici  $\Omega$ , permettant de représenter le « hasard ». Il est en effet possible de s'imaginer qu'il existe une multitude de « mondes » différents, dans lesquels les événements ne sont pas les mêmes. Dans un monde, Julie gagne au Loto ; dans un autre, elle perd. Dans un monde, il pleut à Nantes et fait beau à Marseille ; dans un autre, il grêle dans la cité phocéenne et les nantais profitent d'un soleil radieux.

#### Définition 1

L'ensemble  $\Omega$  de tous ces mondes est appelé *univers (des possibles), espace des épreuves, ou ensemble fondamental*.

On appellera une expérience aléatoire, comme celle qui consiste à jeter un dé ou une pièce, une *épreuve*.

#### Evénements

#### Définition 2

On appelle *tribu sur l'ensemble*  $\Omega$  un ensemble  $\Sigma$  tel que :

- $\Sigma \subset \mathfrak{P}(\Omega)$  (où  $\mathfrak{P}(\Omega)$  désigne l'ensemble des parties de  $\Omega$ )
- $\Omega \in \Sigma$  et  $\emptyset \in \Sigma$
- $\Sigma$  est stable par passage au complémentaire : si  $\mathcal{A} \in \Sigma$ , alors  $\bar{\mathcal{A}} \in \Sigma$
- $\Sigma$  est stable par union dénombrable : si  $\forall n \in \mathbb{N} \quad \mathcal{A}_n \in \Sigma$ , alors  $\bigcup_{n \in \mathbb{N}} \mathcal{A}_n \in \Sigma$

Le couple  $(\Omega, \Sigma)$  est appelé *espace probabilisable*. Un élément de  $\Sigma$  est appelé *événement*. Un élément  $\omega$  de  $\Omega$  est appelé *résultat de l'expérience*  $\Omega$ , *issue*, *éventualité* ou *événement atomique (ou élémentaire)*.

- **Remarque :** Si  $\Omega$  est fini, il est courant de choisir  $\Sigma = \mathfrak{P}(\Omega)$ , l'ensemble des parties de  $\Omega$ .
- **Exemple :**
  - Dans le cas d'un tirage à pile ou face, on peut définir  $\Omega = \{\text{Pile}, \text{Face}\}$  et  $\Sigma = \{\emptyset, \{\text{Pile}\}, \{\text{Face}\}, \{\text{Pile}, \text{Face}\}\}$
  - Dans le cas du tirage d'un dé à 6 faces, on a  $\Omega = \{\square, \square, \square, \square, \square, \square\}$  et  $\Sigma = \mathfrak{P}(\Omega)$ . Alors l'évènement « le tirage est impair » est l'élément  $\{\square, \square, \square\} \in \Sigma$

**Définition 3**

Lorsque ledit résultat  $\omega$  d'une épreuve appartient à un événement  $\mathcal{A} \in \Sigma$ , on dit que  $\mathcal{A}$  est *réalisé*. Deux événements  $\mathcal{A}, \mathcal{B} \in \Sigma$  sont alors dits *réalisés simultanément* si  $\omega \in \mathcal{A} \cap \mathcal{B}$ . Ils sont *incompatibles* si  $\mathcal{A} \cap \mathcal{B} = \emptyset$ .

- **Exemple** : Dans le cas du lancer de dé, si le résultat du tirage est  $\{\ominus\}$ , alors l'événement « le tirage est impair » est réalisé.

**Probabilités**

La théorie des probabilités fait appel à celle de la mesure. Kolmogorov a proposé de définir une probabilité comme la *mesure* d'un événement.

**Définition 4**

Soit  $(\Omega, \Sigma)$  un espace probabilisable. Une application  $\mathbb{P} : \Sigma \rightarrow [0; 1]$  est une *probabilité* sur  $(\Omega, \Sigma)$  si :

- $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1,$
- si  $(\mathcal{A}_n)_{n \in \mathbb{N}}$  est une famille dénombrable d'événements de  $\Sigma$  deux à deux incompatibles, alors
 
$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} \mathcal{A}_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(\mathcal{A}_n).$$

Le triplet  $(\Omega, \Sigma, \mathbb{P})$  est appelé *espace probabilisé*.

- **Remarque** : On peut remarquer qu'un espace probabilisable est un espace mesurable, et qu'une probabilité est une mesure positive normalisée à 1. (On se reportera au cours de François Goudail pour plus d'informations sur la notion de mesure).
- **Exemple** :
  - Soit  $a \in \Omega$ . L'application,  $\delta_a : \mathfrak{P}(\Omega) \rightarrow \{0, 1\}$  qui, à un événement  $\mathcal{A}$ , associe 1 si  $a \in \mathcal{A}$  et 0 sinon, est une probabilité appelée masse de Dirac en  $a$ .
  - Soit  $(p_n)_{n \in \mathbb{N}}$  une suite de réels positifs, vérifiant  $\sum_{n \in \mathbb{N}} p_n = 1$ . Soit  $\Omega$  un ensemble et  $(x_n)_{n \in \mathbb{N}}$  une suite d'éléments de  $\Omega$ . Alors  $\sum_{n \in \mathbb{N}} p_n \delta_{x_n}$  est une probabilité sur  $(\Omega, \mathfrak{P}(\Omega))$ .

**Proposition 1**

Si  $\mathbb{P}$  est une probabilité sur  $(\Omega, \Sigma)$  alors :

- $\mathbb{P}(\bar{\mathcal{A}}) = 1 - \mathbb{P}(\mathcal{A}),$
- $\mathbb{P}(\mathcal{A} \cup \mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B}) - \mathbb{P}(\mathcal{B} \cap \mathcal{A}) \quad (\star),$
- $\mathcal{A} \subset \mathcal{B} \Rightarrow \mathbb{P}(\mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B} \setminus \mathcal{A}),$
- $\mathcal{A} \subset \mathcal{B} \Rightarrow \mathbb{P}(\mathcal{A}) \leq \mathbb{P}(\mathcal{B}).$

☞ **Démonstration** : Soient  $\mathcal{A}$  et  $\mathcal{B}$  deux événements,

- On a par définition  $\mathcal{A} \cup \bar{\mathcal{A}} = \Omega$  et  $\mathcal{A} \cap \bar{\mathcal{A}} = \emptyset$ , et donc  $\mathbb{P}(\bar{\mathcal{A}}) + \mathbb{P}(\mathcal{A}) = 1.$
- De même, on peut voir que  $\mathcal{A} \cup \mathcal{B} = \mathcal{A} \cup (\mathcal{B} \cap \bar{\mathcal{A}})$  or  $\mathcal{A} \cap (\mathcal{B} \cap \bar{\mathcal{A}}) = \emptyset$ , et que  $\mathcal{B} = (\mathcal{B} \cap \mathcal{A}) \cup (\mathcal{B} \cap \bar{\mathcal{A}})$  et  $(\mathcal{B} \cap \mathcal{A}) \cap (\mathcal{B} \cap \bar{\mathcal{A}}) = \emptyset$ . Donc  $\mathbb{P}(\mathcal{A} \cup \mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\bar{\mathcal{A}} \cap \mathcal{B})$  et  $\mathbb{P}(\mathcal{B}) = \mathbb{P}(\mathcal{A} \cap \mathcal{B}) + \mathbb{P}(\bar{\mathcal{A}} \cap \mathcal{B})$ , d'où le résultat.
- On trouve la troisième propriété en écrivant que  $\mathcal{B} = \mathcal{A} \cup (\mathcal{B} \setminus \mathcal{A})$  et que  $\mathcal{A} \cap (\mathcal{B} \setminus \mathcal{A}) = \emptyset$ .
- La dernière propriété se déduit de la précédente et de la non-négativité des probabilités.

**Définition 5**

Si  $\mathbb{P}(\mathcal{A}) = 0$ , on dit que l'événement  $\mathcal{A}$  est *presque impossible*. Si  $\mathbb{P}(\mathcal{A}) = 1$ , on dit que  $\mathcal{A}$  est *presque certain*.

La généralisation de la propriété  $(\star)$  de la Proposition 1 (et une démonstration par récurrence) permet d'obtenir la formule bien connue suivante :

**Proposition 2**

**Formule de Poincaré, ou du crible :** Soient  $n \in \mathbb{N}^*$  et  $\mathcal{A}_1, \dots, \mathcal{A}_n$  des événements quelconques, alors :

$$\mathbb{P}\left(\bigcup_{i=1}^n \mathcal{A}_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(\mathcal{A}_{i_1} \cap \dots \cap \mathcal{A}_{i_k})$$

■ **Exemple :** Soient  $\mathcal{A}, \mathcal{B}$  et  $\mathcal{C}$  trois événements, alors  $\mathbb{P}(\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B}) + \mathbb{P}(\mathcal{C}) - \mathbb{P}(\mathcal{A} \cap \mathcal{B}) - \mathbb{P}(\mathcal{B} \cap \mathcal{C}) - \mathbb{P}(\mathcal{A} \cap \mathcal{C}) + \mathbb{P}(\mathcal{A} \cap \mathcal{B} \cap \mathcal{C})$

**1.3.2 Probabilités d'éventualités finies**

Dans la Section 1.2.1, nous avons défini intuitivement la notion de probabilité d'un événement  $\mathcal{A}$ . Il est possible de retrouver ce résultat dans le cadre de l'approche axiomatique à partir d'une définition précise de l'espace  $(\Omega, \Sigma, \mathbb{P})$ . Pour cela il suffit de considérer un ensemble  $\Omega = \{\omega_1, \dots, \omega_n\}$  fini avec  $n \in \mathbb{N}$ , de prendre  $\Sigma = \mathfrak{P}(\Omega)$  et de définir  $\mathbb{P}$  telle que  $\mathbb{P}(\{\omega_i\}) = p_i = 1/n$ . On retrouve alors

$$\forall \mathcal{A} \in \Sigma, \mathbb{P}(\mathcal{A}) = \frac{\text{Card}\mathcal{A}}{\text{Card}\Omega}$$

Les problèmes de ce type se ramènent alors à l'évaluation de  $\text{Card}\mathcal{A}$  et  $\text{Card}\Omega$ , c'est à dire à des problèmes de dénombrement.

**1.3.3 Probabilités conditionnelles et événements indépendants**

Lors d'un tirage de deux dés à 6 faces, la probabilité que la somme des deux nombres soit un 9 est égale à  $1/9$ . Sachant que l'on a obtenu un 5 sur l'un des deux dés, la probabilité d'obtenir 9 passe alors à  $1/6$ .

Dans cet exemple, la probabilité de l'événement  $\mathcal{A} = \text{« obtenir un 9 avec deux dés »}$  est  $\mathbb{P}(\mathcal{A}) = 1/9$ , mais la probabilité de  $\mathcal{A}$  sachant que l'événement  $\mathcal{B} = \text{« le premier dé donne un 5 »}$  est donnée par  $\mathbb{P}(\mathcal{A}|\mathcal{B}) = 1/6$ . Ainsi la probabilité d'un événement peut être modifiée par la connaissance de la réalisation d'un second événement.

Ces considérations amènent à définir les concepts de conditionnement et d'indépendance.

**Probabilités conditionnelles**

**Définition 6**

**Conditionnement :** Soient  $\mathcal{A}$  et  $\mathcal{B}$  deux événements. On suppose que  $\mathbb{P}(\mathcal{B}) \neq 0$ . On définit la *probabilité conditionnelle* de l'événement  $\mathcal{A}$  sachant l'événement  $\mathcal{B}$ , que l'on note  $\mathbb{P}(\mathcal{A}|\mathcal{B})$  ou  $\mathbb{P}_{\mathcal{B}}(\mathcal{A})$ , comme la quantité :

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

Si  $\mathbb{P}(\mathcal{B}) = 0$ , on peut poser  $\mathbb{P}_{\mathcal{B}}(\mathcal{A}) = \mathbb{P}(\mathcal{A})$ . L'application  $\mathbb{P}_{\mathcal{B}} : \Sigma \rightarrow [0; 1]$  ainsi définie est une probabilité sur  $(\Omega, \Sigma)$ , appelée *probabilité conditionnée par  $\mathcal{B}$* .

Si on a à la fois  $\mathbb{P}(\mathcal{A}) \neq 0$  et  $\mathbb{P}(\mathcal{B}) \neq 0$ , on peut aussi définir la probabilité de  $\mathcal{B}$  sachant l'événement  $\mathcal{A}$  et on a alors

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

De cette formule, on déduit le théorème publié par Thomas Bayes en 1764 :

**Théorème 1**

**Formule de Bayes :** Soient  $\mathcal{A}$  et  $\mathcal{B}$  deux événements de probabilité non nulle. On a alors

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

Si on suppose maintenant que l'ensemble  $\Omega$  se décompose en une partition d'événements dénombrables  $(\mathcal{A}_n)_{n \in \mathbb{N}}$  tels que

$$\bigcup_{n \in \mathbb{N}} \mathcal{A}_n = \Omega \quad \text{et} \quad i \neq j \Rightarrow \mathcal{A}_i \cap \mathcal{A}_j = \emptyset \quad (*),$$

on a alors naturellement :

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} \mathcal{A}_n\right) = 1 \quad \text{et} \quad \mathbb{P}(\mathcal{A}_i \cap \mathcal{A}_j) = 0 \quad \text{si} \quad i \neq j \quad (*).$$

Le système d'événements  $(\mathcal{A}_n)_{n \in \mathbb{N}}$  est alors dit complet. Si on considère un événement  $\mathcal{B}$  on obtient alors :

**La formule des probabilités totales :**

$$\mathbb{P}(\mathcal{B}) = \sum_{n \in \mathbb{N}} \mathbb{P}(\mathcal{A}_n \cap \mathcal{B}) = \sum_{n \in \mathbb{N}} \mathbb{P}(\mathcal{A}_n)\mathbb{P}(\mathcal{B} | \mathcal{A}_n)$$

qui permet de généraliser la formule de Bayes :

**Théorème 2**

**Formule de Bayes :** Soient  $(\mathcal{A}_n)_{n \in \mathbb{N}}$  un système complet d'événements et  $\mathcal{B}$  un événement de probabilité non nulle. Si on suppose que  $\mathbb{P}(\mathcal{A}_n) \neq 0$  pour tout  $n \in \mathbb{N}$ , on a alors

$$\forall p \in \mathbb{N}, \quad \mathbb{P}(\mathcal{A}_p | \mathcal{B}) = \frac{\mathbb{P}(\mathcal{A}_p)\mathbb{P}(\mathcal{B} | \mathcal{A}_p)}{\sum_{n \in \mathbb{N}} \mathbb{P}(\mathcal{A}_n)\mathbb{P}(\mathcal{B} | \mathcal{A}_n)}$$

- **Remarque :** Un système d'événements qui vérifient les propriétés (\*) sans vérifier les propriétés (\*) est dit quasi-complet. Les formules des probabilités totales et de Bayes s'appliquent aussi dans le cadre de systèmes quasi-complets.

**Indépendance**

Intuitivement, on dira que deux événements  $\mathcal{A}$  et  $\mathcal{B}$  sont indépendants si  $\mathbb{P}(\mathcal{B} | \mathcal{A}) = \mathbb{P}(\mathcal{B})$  et  $\mathbb{P}(\mathcal{A} | \mathcal{B}) = \mathbb{P}(\mathcal{A})$ . Cela conduit à la définition suivante :

**Définition 7**

**Indépendance :** Soient  $(\Omega, \Sigma, \mathbb{P})$  un espace probabilisé et  $\mathcal{A}$  et  $\mathcal{B}$  deux événements de  $\Sigma$ . On dit que  $\mathcal{A}$  et  $\mathcal{B}$  sont indépendants si

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

**Définition 8**

**Indépendance mutuelle :** Soient  $(\Omega, \Sigma, \mathbb{P})$  un espace probabilisé et  $(\mathcal{A}_n)_{n \in \llbracket 1, N \rrbracket}$   $N$  événements de  $\Sigma$ . On dit que les événements  $(\mathcal{A}_n)$  sont *mutuellement indépendants* si et seulement si, pour tout  $0 < p \leq N$  et pour tout  $p$ -uplet  $(i_1, \dots, i_p)$  distincts, on a

$$\mathbb{P}\left(\bigcap_{k=1}^p \mathcal{A}_{i_k}\right) = \prod_{k=1}^p \mathbb{P}(\mathcal{A}_{i_k})$$

■ **Remarque :** Attention ! L'indépendance deux à deux n'implique pas l'indépendance mutuelle.

■ **Exemple :** On considère le lancer de deux dés et les événements suivants :

- $\mathcal{A}_1 = \llcorner$  la somme des deux dés est paire  $\llcorner$ ,
- $\mathcal{A}_2 = \llcorner$  le résultat du premier dé est pair  $\llcorner$ ,
- $\mathcal{A}_3 = \llcorner$  le résultat du second dé est pair  $\llcorner$ .

Ces événements sont bien deux à deux indépendants, on vérifie en effet que

$$\mathbb{P}(\mathcal{A}_i) = 1/2 \text{ et } \mathbb{P}(\mathcal{A}_i \cap \mathcal{A}_j) = 1/4. \text{ Or } \mathbb{P}(\mathcal{A}_1)\mathbb{P}(\mathcal{A}_2)\mathbb{P}(\mathcal{A}_3) = 1/8 \text{ mais}$$

$$\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3) = 1/4. \text{ Les événements ne sont donc pas mutuellement indépendants.}$$

**Tirages répétés**

Les concepts de probabilité conditionnelle et d'indépendance révèlent leur intérêt dans les situations où une épreuve globale se compose en  $n$  sous-épreuves successives ou simultanées, on parle généralement de tirages répétés ou simultanés.

Cela amène à considérer  $n$  espaces probabilisés  $(\Omega_1, \Sigma_1, \mathbb{P}_1), \dots, (\Omega_n, \Sigma_n, \mathbb{P}_n)$ . Les résultats possibles forment alors des  $n$ -uplets  $(\omega_1, \omega_2, \dots, \omega_n)$ . Il est alors possible de définir une probabilité  $\mathbb{P}$  sur  $(\Omega_1 \times \dots \times \Omega_n, \Sigma_1 \times \dots \times \Sigma_n)$  de la manière suivante :

$$\forall i \in \llbracket 1, n \rrbracket, \mathcal{A}_i \subset \Omega_i, \text{ et } \mathbb{P}(\Omega_1 \times \dots \times \Omega_{i-1} \times \mathcal{A}_i \times \Omega_{i+1} \times \dots \times \Omega_n) = \mathbb{P}_i(\mathcal{A}_i)$$

Si on considère des espaces indépendants, on a alors

$$\mathbb{P}(\mathcal{A}_1 \times \dots \times \mathcal{A}_n) = \mathbb{P}_1(\mathcal{A}_1) \cdot \mathbb{P}_2(\mathcal{A}_2) \cdot \dots \cdot \mathbb{P}_n(\mathcal{A}_n)$$

Un cas très courant est celui où l'on répète  $n$  fois la même expérience de manière indépendante. Les espaces  $(\Omega_1, \Sigma_1, \mathbb{P}_1), \dots, (\Omega_n, \Sigma_n, \mathbb{P}_n)$  sont alors identiques et on définit l'espace probabilisé  $(\Omega^n, \Sigma_n, \mathbb{P}_n)$ , où  $\mathbb{P}_n$  est définie par

$$\begin{aligned} \mathbb{P}_n : \quad \Sigma_n &\longrightarrow [0, 1] \\ \mathcal{A}_1 \times \dots \times \mathcal{A}_n &\longmapsto \mathbb{P}(\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_2) \cdot \dots \cdot \mathbb{P}(\mathcal{A}_n). \end{aligned}$$

# Chapitre 2

## Variables Aléatoires

Au cours du chapitre précédent, nous nous sommes intéressés aux probabilités de parties de  $\Omega$  appelées événements. Cependant, il est courant en physique de s'intéresser plutôt à des données numériques issues / déduites des éléments  $\omega$  de  $\Omega$ . Par exemple, si  $\Omega$  désigne l'ensemble des couples (vitesse, position) possibles des particules d'un gaz, il est plus courant de s'intéresser à la pression associée plutôt qu'à la probabilité d'obtenir une certaine combinaison de ces couples (vitesse, position). Ainsi, si on note  $X$  cette donnée numérique,  $X$  sera une application de  $\Omega$  dans  $\mathbb{R}$  et sera appelée *variable aléatoire*. Afin d'obtenir un cadre théorique rigoureux complet, il est aussi nécessaire de s'appuyer sur des notions issues de la théorie de la mesure et/ou du cours d'intégration de François Goudail.

### 2.1 Définitions élémentaires

#### Définition 9

**Variable aléatoire :** Soient l'espace probabilisé  $(\Omega, \Sigma, \mathbb{P})$  et l'espace  $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$  où  $\mathfrak{B}(\mathbb{R})$  désigne la tribu des Boréliens. On appelle *variable aléatoire* (ou VA) toute application mesurable  $X : \Omega \rightarrow \mathbb{R}$ , c'est à dire si  $\forall A \in \mathfrak{B}(\mathbb{R}), X^{-1}(A) \in \Sigma$ .

Quand on attribue une valeur à une variable aléatoire, on parle de *tirage* ou de *réalisation*.

- **Remarque :** • Bien qu'il soit possible de définir des variables aléatoires sur  $\mathbb{C}$ , nous ne considérerons que des VA réelles dans la suite de ce cours.
  - En pratique, on vérifie rarement l'hypothèse de mesurabilité (même si cette notion peut être importante, en physique statistique notamment). Ainsi, à notre niveau, n'importe quelle fonction de  $\Omega$  dans  $\mathbb{R}$  pourra être une VA.
  - Si  $\omega$  est un élément de  $\Omega$ , on trouve régulièrement la notation  $X_\omega$  pour désigner  $X(\omega)$ . De même, il est aussi courant de noter  $\lambda$  les éléments de  $\Omega$ , de sorte que l'on peut aussi utiliser la notation  $X_\lambda$ .

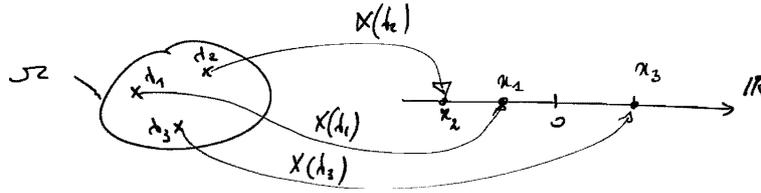
#### Définition 10

**Loi de probabilité :** Soient  $(\Omega, \Sigma, \mathbb{P})$  un espace probabilisé et  $X$  une variable aléatoire. La *loi de probabilité* (ou parfois *distribution*) de la variable aléatoire  $X$  est la probabilité

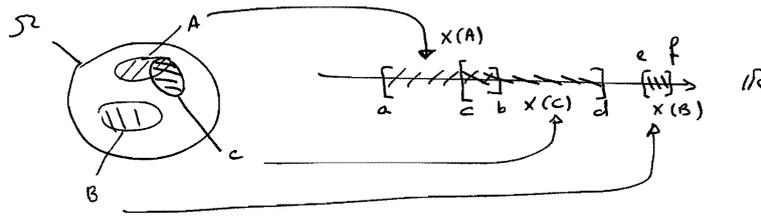
$$\begin{aligned} \mathbb{P}_X : \mathfrak{B}(\mathbb{R}) &\longrightarrow [0, 1] \\ \mathcal{B} &\longmapsto \mathbb{P}(\{X \in \mathcal{B}\}) \end{aligned}$$

où on définit  $\{X \in \mathcal{B}\} = \{\omega \in \Omega \mid X(\omega) \in \mathcal{B}\} = X^{-1}(\mathcal{B})$ .

Autrement dit, la loi de  $X$  est la probabilité image de  $\mathbb{P}$  par  $X$ .



(a) Variable aléatoire sur un ensemble dénombrable



(b) Variable aléatoire sur un ensemble infini

FIGURE 2.1 – Variable aléatoire - schématisation. Sur la Figure (a),  $X$  est telle que  $\mathbb{P}_X(x_i) = \mathbb{P}(\{\omega_i\})$ . Si la Figure (b),  $X$  est définie de sorte que  $\mathbb{P}_X([a, b]) = \mathbb{P}(\mathcal{A})$ ,  $\mathbb{P}_X([c, d]) = \mathbb{P}(\mathcal{C})$  et  $\mathbb{P}_X([e, f]) = \mathbb{P}(\mathcal{B})$ .

On a en particulier :

- $\mathbb{P}_X([a, b]) = \mathbb{P}(\{a < X \leq b\}) = \mathbb{P}(X^{-1}([a, b])) = \mathbb{P}(\{\omega \in \Omega \mid a < X(\omega) \leq b\})$ ,
- $\mathbb{P}_X(]-\infty, x]) = \mathbb{P}(\{X \leq x\}) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \leq x\})$

- **Remarque :**
  - Deux variables aléatoires égales presque partout ont la même loi de probabilité mais la réciproque est fautive.
  - En pratique, l'ensemble  $\Omega$  n'est généralement pas connu. De même, il est très rare dans les cas abordés de spécifier l'espace  $(\Omega, \Sigma, \mathbb{P})$ . En effet, on manipule presque toujours les valeurs prises par  $X$ , de sorte que les modélisations mathématiques d'un problème physique se contentent généralement de donner la loi de  $X$ . Cela est justifié par un théorème (appelé théorème de représentation) qui stipule que pour toute probabilité  $\mathbb{P}^*$  sur  $\mathbb{R}$ , on peut trouver un espace probabilisé  $(\Omega, \Sigma, \mathbb{P})$  et une variable aléatoire  $X$  tels que  $\mathbb{P}^* = \mathbb{P}_X$ . On pourra ainsi poser un problème en définissant les variables aléatoires d'intérêt « sans prendre de précaution ».

**Définition 11**

**Fonction de répartition :** On appelle *fonction de répartition* de la variable aléatoire  $X$  la fonction

$$F_X : \mathbb{R} \rightarrow \mathbb{R}$$

$$x \mapsto F_X(x) = \mathbb{P}(\{X \leq x\}) = \mathbb{P}_X(]-\infty, x]).$$

On a alors pour tous  $a, b \in \mathbb{R}$ ,  $\mathbb{P}(\{a < X \leq b\}) = F_X(b) - F_X(a)$ .

- **Remarque :**  $F_X$  caractérise entièrement la loi de probabilité de  $X$  : toute l'information sur la loi  $\mathbb{P}_X$  est contenue dans  $F_X$ .

**Proposition 3**

Soit  $X$  une variable aléatoire de fonction de répartition  $F_X$ . On retiendra les propriétés suivantes :

- $F_X$  est croissante.
- $F_X$  est continue à droite :  $\forall x \in \mathbb{R}, F_X(x^+) = F_X(x)$ .
- $F_X$  n'est pas forcément continue à gauche et  $\forall x \in \mathbb{R}, F_X(x) - F_X(x^-) = \mathbb{P}(\{X = x\})$ .
- $\forall x \in \mathbb{R}, 0 \leq F_X(x) \leq 1, \lim_{x \rightarrow -\infty} F_X(x) = 0$  et  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .
- $\mathbb{P}_X([a, b]) = F_X(b) - F_X(a^-)$  avec  $a < b$ .
- Si  $F_X(x_0) = 0$  alors  $\forall x < x_0, F_X(x) = 0$ .
- $\mathbb{P}(\{X > x\}) = 1 - F_X(x)$ .

On rappelle que la notation  $F_X(a^-)$  est équivalente à  $\lim_{\substack{x \rightarrow a \\ x < a}} F_X(x)$ . De même, on utilisera parfois la notation  $F_X(+\infty)$  pour désigner  $\lim_{x \rightarrow +\infty} F_X(x)$

☞ **Démonstration :**

- Pour tous  $a$  et  $b$  tels que  $a \leq b$ , on a  $\{x \leq a\} \subset \{x \leq b\}$  et donc  $\mathbb{P}(\{x \leq a\}) \leq \mathbb{P}(\{x \leq b\})$ .
- Pour tout  $\epsilon$ , on a  $F_X(x + \epsilon) - F_X(x) = \mathbb{P}(\{x < X \leq x + \epsilon\})$ . Un passage à la limite quand  $\epsilon$  tend vers 0 permet de voir que  $\{x < X \leq x + \epsilon\}$  tend vers l'ensemble vide et de montrer le résultat.
- La troisième propriété se montre de manière similaire en posant  $F_X(x) - F_X(x - \epsilon) = \mathbb{P}(\{x - \epsilon < X \leq x\})$ .
- Les autres propriétés se déduisent de la définition de  $F_X$  et des propriétés précédentes.

En pratique, la forme de la fonction  $F_X$  permet de définir la nature de  $X$ .

**Définition 12**

Soient  $X$  une variable aléatoire et  $F_X$  sa fonction de répartition.

- $X$  est dite discrète si  $F_X$  est en escalier (constante par morceaux). On note alors  $\mathbb{P}_X(x_i) = p_i$  où les  $x_i$  sont les points de discontinuité de  $F_X$ . On a alors  $\sum_i p_i = 1$ .
- $X$  est dite continue si  $F_X$  est continue.
- $X$  est dite absolument continue si  $F_X$  est dérivable presque partout, c'est à dire s'il existe une fonction  $p_X$  telle que

$$F_X(x) = \int_{-\infty}^x p_X(t) dt$$

sauf sur un ensemble de mesure nul. La fonction  $p_X$  est alors appelée *la densité de probabilité* de  $X$  et vérifie donc

$$\int_{\mathbb{R}} p_X(t) dt = 1$$

■ **Remarque :** En pratique, on considèrera principalement des VA absolument continues, de sorte qu'on omettra l'adverbe *absolument*. De plus, on utilisera surtout le résultat  $p_X(x) = dF_X(x)/dx$ .

On peut aussi distinguer le cas « mixte » qui peut être vu comme la généralisation des deux situations précédentes. Il s'agit de variables aléatoires  $X$  où la fonction  $F_X$  est dérivable par morceaux et possède un ensemble de points de discontinuité  $\{x_0, x_1, x_2, \dots\}$ . On alors  $\mathbb{P}_X(x_i) = F_X(x_i) - F_X(x_i^-) = p_i$  et partout ailleurs  $p_X(x) = dF_X(x)/dx$ .

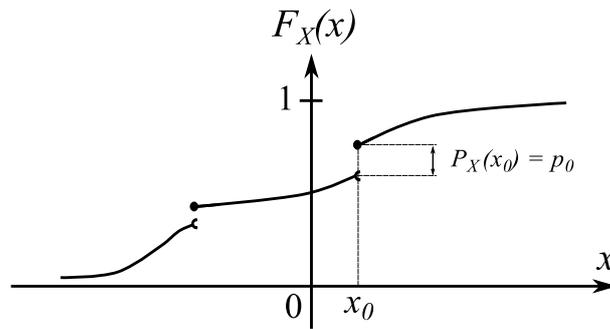


FIGURE 2.2 – Allure générale d’une fonction de répartition.

■ **Remarque :** • On trouve plus classiquement la définition suivante d’une VA discrète :

“Soient  $(\Omega, \Sigma, \mathbb{P})$  un espace probabilisé et  $X$  une variable aléatoire.  $X$  est dite discrète si  $X(\Omega)$  est dénombrable et si,  $\forall x \in \mathbb{R}$ ,  $\{\omega \in \Omega \mid X(\omega) = x\}$  est un événement.”

- Les lois de probabilité de variables discrètes sont généralement représentées graphiquement par des « pics » d’une certaine hauteur, appelés *pics de Dirac*.
- La définition d’une VA continue  $X$  revient à supposer que  $F_X$  est dérivable. De plus, on peut interpréter la densité de probabilité en disant que la probabilité que la variable aléatoire  $X$  prenne ses valeurs dans l’intervalle  $]x, x + dx]$  est égale (au premier ordre) à  $p_X(x)dx$  :  $\mathbb{P}(\{x < X \leq x + dx\}) \approx p_X(x)dx$ .

## 2.2 Espérance, variance et autres moments

Bien que la connaissance de la loi complète d’une variable aléatoire soit cruciale dans de nombreux problèmes physiques, il est parfois plus pratique ou utile de ne considérer qu’une quantité numérique dérivée de cette loi, appelée moment. Les plus utilisées de ces quantités sont l’espérance et la variance.

### 2.2.1 Espérance

Il convient dans cette section de distinguer le cas discret du cas continu.

#### Définition 13

**Espérance d’une VA discrète :** Soit  $X$  une variable aléatoire discrète de loi de probabilité  $\{\mathbb{P}_X(x_i) = p_i \mid i \in \mathbb{N}\}$ . Si la série  $\sum x_i p_i$  converge absolument, alors  $X$  admet une espérance notée  $\mathbb{E}[X]$  et définie par

$$\mathbb{E}[X] = \sum_{i=0}^{+\infty} x_i p_i$$

On rappelle qu’une série  $\sum a_n$  converge absolument si la suite de terme  $S_n = \sum_{k=0}^n |a_k|$  converge. De plus, une série absolument convergente est commutativement convergente (maintient de la convergence quand on change l’ordre des termes).

De manière intuitive, l’espérance correspond à la valeur « moyenne » attendue de la variable aléatoire, où la moyenne est définie par

$$S_n = \frac{1}{n} \sum_{i=0}^{n-1} x_i,$$

on parle alors de moyenne arithmétique. On montrera au Chapitre 5 que la suite des  $(S_n)_{n \in \mathbb{N}}$  « converge » vers  $\mathbb{E}[X]$  quand  $n$  augmente. C’est la loi des grands nombres.

**Théorème 3**

**Formule de transfert :** Soit une fonction mesurable  $g$  de  $\mathbb{R}$  dans  $\mathbb{R}$ . Si  $X$  est une VA discrète alors l'application  $g \circ X$ , notée  $g(X)$ , est une VA discrète. Son espérance, si la série suivante converge absolument, est

$$\mathbb{E}[g(X)] = \sum_{k=0}^{+\infty} g(x_k) \mathbb{P}(\{X = x_k\})$$

☞ **Démonstration :** On note  $Y = g(X)$ . L'ensemble des valeurs de  $X$  étant dénombrable, son image par  $g$  l'est aussi et donc l'ensemble des valeurs de  $Y$  est dénombrable. On note  $y_j$  les valeurs que peut prendre  $Y$  et

$$E_j = \{k \in \mathbb{N} \mid g(x_k) = y_j\}$$

La famille des  $(E_j)_j$  est une partition de  $\mathbb{N}$ . De plus, on obtient la loi de  $Y$  en remarquant que

$$\mathbb{P}(\{Y = y_j\}) = \sum_{k \in E_j} \mathbb{P}(\{X = x_k\})$$

Si on suppose que  $\sum_{k=0}^{+\infty} g(x_k) \mathbb{P}(\{X = x_k\})$  converge absolument, on a alors

$$\begin{aligned} \sum_{k=0}^{+\infty} |g(x_k)| \mathbb{P}(\{X = x_k\}) &= \sum_j \left( \sum_{k \in E_j} |g(x_k)| \mathbb{P}(\{X = x_k\}) \right) \\ &= \sum_j |y_j| \left( \sum_{k \in E_j} \mathbb{P}(\{X = x_k\}) \right) = \sum_j |y_j| \mathbb{P}(\{Y = y_j\}) \end{aligned}$$

La série  $\sum_j y_j \mathbb{P}(\{Y = y_j\})$  converge donc absolument et  $Y$  admet donc une espérance  $\mathbb{E}[Y]$ . Un calcul analogue au précédent permet alors de trouver que

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum_{k=0}^{+\infty} g(x_k) \mathbb{P}(\{X = x_k\})$$

Ces derniers résultats se généralisent au cas des variables aléatoires continues :

**Définition 14**

**Espérance d'une VA continue :** Soit  $X$  une variable aléatoire (absolument) continue de densité de probabilité  $p_X$ . Si la fonction  $t \mapsto t p_X(t)$  est intégrable ( $\int |t| p_X(t) < +\infty$ ), alors  $X$  admet une espérance notée  $\mathbb{E}[X]$  et définie par

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} t p_X(t) dt$$

**Théorème 4**

**Formule de transfert (admise) :** Soit une fonction mesurable  $g$  de  $\mathbb{R}$  dans  $\mathbb{R}$ . Si  $X$  est une VA (absolument) continue alors l'application  $g \circ X$ , notée  $g(X)$ , est une VA (absolument) continue. Son espérance, si l'intégrale suivante converge absolument, est

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) p_X(x) dx$$

Il est aussi possible de généraliser la notion d'espérance aux cas « mixtes » si les conditions de convergence et d'intégrabilité sont vérifiées :

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} t p_X(t) dt + \sum_i x_i p_i$$

La justification de cette expression se trouve dans la théorie des distributions ou dans la définition d'un type d'intégrales dites de Riemann-Stieltjes, concepts que nous n'aborderons pas dans ce cours. Dans la suite, les cas discret et continu ne seront plus séparés, et il sera sous-entendu que la densité de probabilité de  $X$  peut admettre des pics de Dirac.

- **Remarque :**
  - S'assurer de l'absolue convergence (pas seulement de la simple convergence) permet de s'assurer de l'existence de l'espérance indépendamment de l'ordre des termes de la série.
  - L'espérance est aussi parfois appelée *moyenne statistique*, *moyenne d'ensemble*, ou tout simplement *moyenne*, mais dans ce dernier cas une ambiguïté pourra exister avec la notion de moyenne arithmétique.
  - On note aussi l'espérance de  $X$  sous la forme  $\langle X \rangle$ . On trouve aussi parfois la notation  $\overline{X}$ , mais cette dernière sera réservée plutôt à la moyenne temporelle (cf. cours de Traitement du signal du semestre 2).

**Proposition 4**

**Propriétés de l'espérance :** Soient  $X$  et  $Y$  deux variables aléatoires admettant une espérance et soient  $\alpha, \beta \in \mathbb{R}$ .

- Linéarité :  $\alpha X + \beta Y$  est une variable aléatoire qui admet une espérance et  $\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$ .
- Si  $a$  est une constante, alors  $\mathbb{E}[a] = a$ .
- Positivité : si  $X \geq 0$  alors  $\mathbb{E}[X] \geq 0$ .
- Monotonie : si  $X \leq Y$  alors  $\mathbb{E}[X] \leq \mathbb{E}[Y]$ .

☞ **Démonstration :**

- La démonstration rigoureuse de la première propriété nécessite des résultats qui ne seront vus qu'au chapitre suivant. Néanmoins, à titre d'illustration, je vous présente ici une démonstration pour les VA discrètes. Ce résultat s'étend aux VA continues en utilisant par exemple les théorèmes de la section 3.2.2.

On note  $Z = \alpha X + \beta Y$ , de sorte que les valeurs prises par  $Z$  sont  $z_k = \alpha x_i + \beta y_j$  et que

$$\mathbb{P}(\{Z = z_k\}) = \sum_{(i,j) \in E_k} \mathbb{P}(\{X = x_i, Y = y_j\})$$

avec  $E_k = \{(i, j) \in \mathbb{N}^2 \mid z_k = \alpha x_i + \beta y_j\}$ . On peut alors écrire que

$$\sum_k |z_k| \mathbb{P}(\{Z = z_k\}) = \sum_k \sum_{(i,j) \in E_k} |\alpha x_i + \beta y_j| \mathbb{P}(\{X = x_i, Y = y_j\})$$

et donc par inégalité triangulaire que

$$\begin{aligned} \sum_k |z_k| \mathbb{P}(\{Z = z_k\}) &\leq \sum_i \sum_j (|\alpha| |x_i| + |\beta| |y_j|) \mathbb{P}(\{X = x_i, Y = y_j\}) \\ &\leq |\alpha| \sum_i |x_i| \sum_j \mathbb{P}(\{X = x_i, Y = y_j\}) + |\beta| \sum_j |y_j| \sum_i \mathbb{P}(\{X = x_i, Y = y_j\}) \end{aligned}$$

On peut en effet commuter les sommes grâce aux propriétés de commutativité des séries absolument convergentes, l'absolue convergence étant garantie par les résultats de la section 3.1.2. De plus, on peut remarquer que

$$\sum_j \mathbb{P}(\{X = x_i, Y = y_j\}) = \mathbb{P}(\{X = x_i\}) \text{ et } \sum_i \mathbb{P}(\{X = x_i, Y = y_j\}) = \mathbb{P}(\{Y = y_j\})$$

on a donc

$$\sum_k |z_k| \mathbb{P}(\{Z = z_k\}) \leq |\alpha| \sum_i |x_i| \mathbb{P}(\{X = x_i\}) + |\beta| \sum_j |y_j| \mathbb{P}(\{Y = y_j\})$$

la série  $\sum_k z_k \mathbb{P}(\{Z = z_k\})$  converge donc absolument et par le même raisonnement on trouve que

$$\sum_k z_k \mathbb{P}(\{Z = z_k\}) = \alpha \sum_i x_i \mathbb{P}(\{X = x_i\}) + \beta \sum_j y_j \mathbb{P}(\{Y = y_j\})$$

et donc que  $\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$ .

- Les deuxième et troisième propriétés se déduisent des définitions de l'espérance et de la loi de probabilité / densité de probabilité.
- La dernière propriété se déduit de la positivité et de la linéarité en travaillant sur la VA  $(X - Y)$ .

## 2.2.2 Variance et écart-type

### Définition 15

**Variance et écart-type d'une VA :** Soit  $X$  une variable aléatoire. On appelle *variance* de  $X$ , lorsqu'elle existe, la quantité

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

On appelle *écart-type* de  $X$  la quantité  $\sigma_x = \sqrt{\mathbb{V}[X]}$ .

- **Remarque :**
  - Si  $X$  est une quantité possédant une dimension physique (longueur, tension électrique, etc.), alors  $\mathbb{E}[X]$  et  $\sigma_x$  possèdent la même dimension. Pensez donc à bien vérifier l'homogénéité de vos calculs !
  - Si  $X$  admet une variance, elle admet une espérance et on déduit de la Proposition 4 une autre expression de la variance :

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

- La formule du transfert permet d'explicitier l'expression de la variance dans le cas discret :

$$\mathbb{V}[X] = \sum_{k=0}^{+\infty} (x_k - \mathbb{E}[X])^2 p_k \quad \text{en utilisant les notation de la Définition 13}$$

et dans le cas continu :

$$\mathbb{V}[X] = \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2 p_X(x) dx \quad \text{en utilisant les notation de la Définition 14.}$$

### Proposition 5

**Propriétés de la variance :** Soit  $X$  une variable aléatoire qui admet une variance, et soit  $a$  et  $b$  deux scalaires. Alors :

$$\mathbb{V}[aX + b] = a^2 \mathbb{V}[X]$$

Ce résultat vient directement de la définition de la variance.

**Définition 16**

**Variable centrée, réduite :** Soit  $X$  une variable aléatoire.

On dit que  $X$  est une *variable aléatoire centrée* si et seulement si  $\mathbb{E}[X] = 0$ .

On dit que  $X$  est une *variable aléatoire réduite* si et seulement si  $\sigma_X = 1$ .

- **Remarque :** Si  $X$  est une variable aléatoire d'espérance égale à  $m$  et de variance égale à  $\sigma^2$ , la variable  $Y = (X - m)/\sigma$  est une variable centrée réduite.  
A l'inverse, si  $X$  est une variable centrée réduite,  $Y = \sigma X + m$  est une variable aléatoire d'espérance égale à  $m$  et de variance égale à  $\sigma^2$ .  
Ces considérations sont particulièrement utiles en programmation quand on utilise des générateurs de nombres aléatoires basés sur des VA ayant une certaine espérance et une certaine variance.

**Interprétation de l'écart-type**

L'écart-type donne la tendance qu'a une variable aléatoire à s'écarter de son espérance, il donne ainsi une idée de la distribution des valeurs de la variable aléatoire autour de sa valeur moyenne. L'inégalité de Bienaymé-Tchebychev permet de quantifier cette tendance en donnant une borne supérieure à la probabilité que  $X$  diffère de sa moyenne d'une certaine quantité.

**Théorème 5**

**Inégalité de Bienaymé-Tchebychev :** Soit  $X$  une variable aléatoire admettant une espérance notée  $m$  et une variance notée  $\sigma^2$ . Pour tout  $\epsilon > 0$ , la probabilité que lors d'un tirage de  $X$ , l'écart à la moyenne  $|X - m|$  dépasse  $\epsilon$  est toujours plus petit que  $\sigma^2/\epsilon^2$  :

$$\mathbb{P}(\{|X - m| \geq \epsilon\}) \leq \frac{\sigma^2}{\epsilon^2}$$

ou en posant  $\epsilon = k\sigma$  :

$$\mathbb{P}(\{|X - m| \geq k\sigma\}) \leq \frac{1}{k^2}$$

Ainsi, pour toute VA qui possède une variance  $\sigma^2$ , la probabilité de s'écarter de  $2\sigma$  de la moyenne ne peut dépasser 25%. Celle de s'écarter de  $3, 3\sigma$  ne dépasse pas 10% et celle de s'écarter de  $10\sigma$ , 1%.

☞ **Démonstration :** Soit  $Y$  une VA positive admettant une espérance et soit  $a > 0$ , on peut alors écrire

$$\mathbb{E}[Y] = \int y p_Y(y) dy \geq \int_a^{+\infty} y p_Y(y) dy \geq a \int_a^{+\infty} p_Y(y) dy = a \mathbb{P}(\{Y \geq a\})$$

Il s'agit de l'inégalité de Markov. Si on applique cette inégalité à la VA  $Y = (X - m)^2$  avec  $a = \epsilon^2$  on obtient l'inégalité de Bienaymé-Tchebychev.

**2.2.3 Moments d'ordres supérieurs****Définition 17**

**Moments d'une VA :** Soit  $X$  une variable aléatoire. On dit que  $X$  admet *un moment d'ordre  $k$*  ( $k \in \mathbb{N}$ ) si et seulement si la variable  $X^k$  admet une espérance. Son moment d'ordre  $k$  est alors

$$\mu_k = \mathbb{E}[X^k]$$

On appelle *moment centré d'ordre  $k$*  de la variable aléatoire  $X$  d'espérance  $m$  la quantité :

$$\mu_k^c = \mathbb{E}[(X - m)^k]$$

- **Remarque :** Le moment d'ordre 0 vaut 1. L'espérance est le moment d'ordre 1. La variance est le moment centré d'ordre 2.

### 2.2.4 Médiane

#### Définition 18

**Médiane d'une VA :** Soit  $X$  une variable aléatoire absolument continue de fonction caractéristique  $F_X$ . On appelle *médiane* de  $X$  tout réel noté  $x_{1/2}$  vérifiant  $F_X(x) = 1/2$ .

Dans le cas général (c'est-à-dire où  $X$  n'est pas forcément absolument continue), la médiane  $x_{1/2}$  est définie par

$$F_X(x_{1/2}) \geq \frac{1}{2} \quad \text{et} \quad F_X(x_{1/2-}) \leq \frac{1}{2}$$

- **Remarque :** En physique nucléaire, le temps de vie des particules radioactives est modélisé par une loi exponentielle (voir plus loin). Le temps de vie médian est alors appelé *demi-vie* et le temps de vie moyen est appelé *durée de vie*.

### 2.2.5 Mode

#### Définition 19

**Mode d'une loi de probabilité :** Soit  $X$  une variable aléatoire absolument continue de densité de probabilité  $p_X$ . On appelle *mode* de  $X$  le(s) réel(s) noté(s)  $x_p$  vérifiant, quand il(s) existe(nt),

$$x_p = \operatorname{argmax}_x [p_X(x)].$$

Dans le cas discret, le mode correspond à la valeur la plus probable de  $X$ .

## 2.3 Quelques lois classiques

Cette partie présente la définition et les propriétés de quelques lois de probabilité usuelles. D'autres sont présentées dans l'Annexe B.

↪ **Exercice :** Retrouvez, pour chacune des lois, l'espérance et la variance associées.

### Loi de Bernouilli, loi binomiale

#### Définition 20

Soient  $n \in \mathbb{N}$  et  $p \in [0, 1]$ . On dit qu'une variable aléatoire  $X$  suit la loi de Bernouilli de paramètre  $p$ , et on écrit  $X \sim \mathcal{B}(p)$ , si  $X(\Omega) = \{0, 1\}$  et

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

On dit qu'une variable aléatoire  $X$  suit la loi binomiale de paramètres  $n$  et  $p$ , et on écrit  $X \sim \mathcal{B}(n, p)$ , si  $X(\Omega) = \llbracket 0, n \rrbracket$  et si

$$\mathbb{P}(\{X = k\}) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & \forall k \in \llbracket 0, n \rrbracket \\ 0 & \text{sinon} \end{cases}$$

- **Remarque :** On peut montrer que la loi binomiale correspond au résultat de la somme de  $n$  tirages indépendants d'une variable aléatoire suivant la loi de Bernouilli.

**Proposition 6**

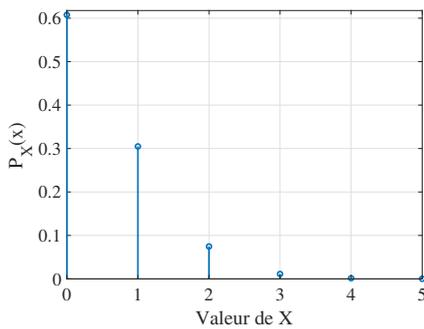
Si  $X \sim \mathcal{B}(p)$  alors  $\mathbb{E}[X] = p$  et  $\sigma_x^2 = p(1 - p)$ .  
 Si  $X \sim \mathcal{B}(n, p)$  alors  $\mathbb{E}[X] = np$  et  $\sigma_x^2 = np(1 - p)$ .

**Loi de Poisson**

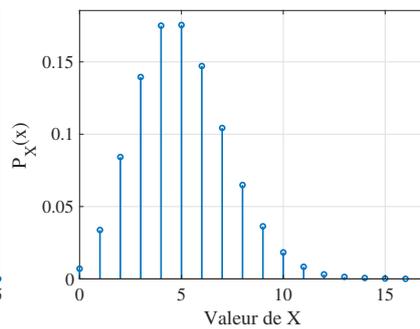
**Définition 21**

Soit  $\lambda \in \mathbb{R}^*$ . On dit qu'une variable aléatoire  $X$  suit la loi de Poisson de paramètre  $\lambda$ , et on écrit  $X \sim \mathcal{P}(\lambda)$ , si  $X(\Omega) = \mathbb{N}$  et

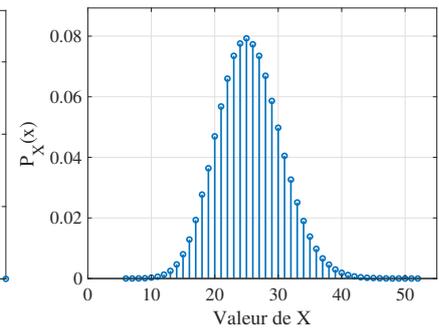
$$\mathbb{P}(\{X = n\}) = e^{-\lambda} \frac{\lambda^n}{n!}$$



(a)  $\lambda = 0.5$



(b)  $\lambda = 5$



(c)  $\lambda = 25.5$

FIGURE 2.3 – Exemples de loi de Poisson

■ **Remarque :** On verra (au chapitre 5) que la loi de Poisson peut être vue comme la limite de la loi binomiale quand  $n \rightarrow +\infty$ .

**Proposition 7**

Si  $X \sim \mathcal{P}(\lambda)$  alors  $\mathbb{E}[X] = \lambda$  et  $\sigma_x^2 = \lambda$ .

**Loi uniforme**

**Définition 22**

Soit  $[a, b]$  un segment de  $\mathbb{R}$ . On dit qu'une variable aléatoire  $X$  suit la loi uniforme sur  $[a, b]$  si sa densité de probabilité est de la forme

$$p_x(x) = \begin{cases} \frac{1}{b-a} & \text{sur } [a, b] \\ 0 & \text{ailleurs} \end{cases}$$

**Proposition 8**

Si  $X$  suit la loi uniforme alors  $\mathbb{E}[X] = (a + b)/2$  et  $\sigma_x^2 = (b - a)^2/12$ .

**Loi normale (ou gaussienne)**

**Définition 23**

Soient  $m \in \mathbb{R}$  et  $\sigma > 0$ . On dit qu'une variable aléatoire  $X$  suit la loi normale de paramètres  $m$  et  $\sigma$ , et on écrit  $X \sim \mathcal{N}(m, \sigma)$ , si sa densité de probabilité est de la forme

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

$X$  suit la loi normale réduite centrée si  $X \sim \mathcal{N}(0, 1)$ .

**Proposition 9**

Si  $X \sim \mathcal{N}(m, \sigma)$  alors  $\mathbb{E}[X] = m$  et  $\sigma_X^2 = \sigma^2$ .

**Loi exponentielle**

**Définition 24**

Soit  $\alpha > 0$ . On dit qu'une variable aléatoire  $X$  suit la loi exponentielle de paramètre  $\alpha$ , et on écrit  $X \sim \mathcal{E}(\alpha)$ , si sa densité de probabilité est de la forme

$$p_X(x) = \begin{cases} \alpha e^{-\alpha x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

■ **Remarque :** La définition de la loi exponentielle donnée ici est celle rencontrée couramment en mathématiques. En physique, elle est plus classiquement définie comme

$$p_X(x) = \begin{cases} \frac{1}{a} e^{-x/a} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

**Proposition 10**

Si  $X \sim \mathcal{E}(\alpha)$  alors  $\mathbb{E}[X] = 1/\alpha$  et  $\sigma_X^2 = 1/\alpha^2$ .

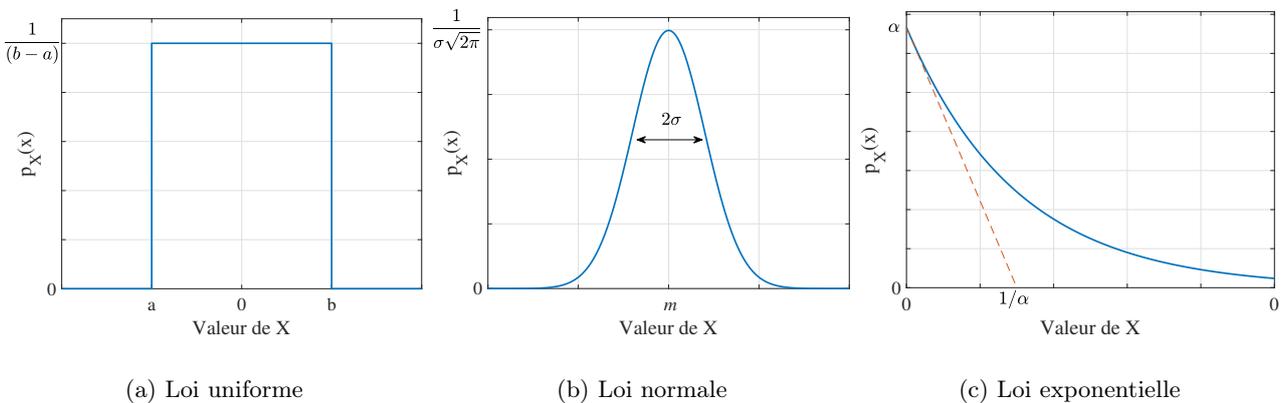


FIGURE 2.4 – Formes générales de lois classiques

**Histogramme**

En pratique, il n'est pas toujours évident de connaître de manière analytique l'expression de la loi ou de la densité de probabilité d'une variable aléatoire. Une solution consiste à répéter un grand nombre de

fois les mesures pour obtenir un grand nombre de réalisations de la variable aléatoire en question, puis à partir de ces réalisations de tracer leur histogramme. Un histogramme est un graphique, généralement représenté sous la forme de barres, qui donne le nombre d'occurrences (indiqué en ordonnées) des valeurs prises par la variable aléatoire (indiquées en abscisses).

Vous verrez en 3A, si vous suivez le parcours *Signal et Images*, que, sous certaines conditions, un histogramme peut constituer « une bonne approximation » de la densité de probabilité de la variable aléatoire

## 2.4 Fonction d'une variable aléatoire

Soit une fonction mesurable  $g$  de  $\mathbb{R}$  dans  $\mathbb{R}$ . Si  $X$  est une variable aléatoire alors  $g(X)$  est une variable aléatoire. Si on pose  $Y = g(X)$ , on a affaire à un changement de variable aléatoire et dans de nombreux problèmes on sera amené à déterminer les propriétés statistiques de  $Y$

### 2.4.1 Changement de variable aléatoire

#### Fonction de répartition d'une fonction d'une VA

Une manière d'avoir accès aux propriétés statistiques de la variable  $Y$  est de déterminer sa fonction de répartition. L'idée est pour cela de partir de sa définition.

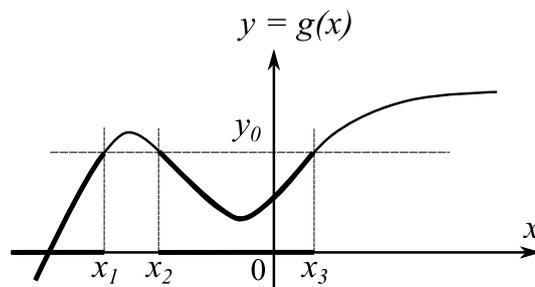


FIGURE 2.5 – Exemple de changement de variable aléatoire

Considérons l'exemple de la figure 2.5 qui illustre une situation classique où la fonction  $g$  continue et dérivable. On cherche dans un premier temps à déterminer la fonction de répartition pour tout réel  $y_0$  comme indiqué sur la figure. On voit alors que l'évènement  $\{Y \leq y_0\}$  est équivalent à l'évènement  $\{X \leq x_1\} \cup \{x_2 \leq X \leq x_3\}$ , on a donc

$$\mathbb{P}(\{Y \leq y_0\}) = \mathbb{P}(\{X \leq x_1\} \cup \{x_2 \leq X \leq x_3\}) = \mathbb{P}(\{X \leq x_1\}) + \mathbb{P}(\{x_2 \leq X \leq x_3\})$$

Et par définition de la fonction de répartition

$$\forall y_0, F_Y(y_0) = F_X(x_1) + [F_X(x_3) - F_X(x_2)]$$

Plusieurs autres cas se présentent en pratique et seront vus directement en classe ou en TD.

#### Densité de probabilité d'une fonction d'une VA continue

Si l'on a déterminé la fonction caractéristique  $F_Y$ , partout où elle est dérivable, il suffit de la dériver pour obtenir la densité de probabilité  $p_Y$  de la variable aléatoire  $Y$ . Aux discontinuités éventuelles, on dérive au sens des distributions (voir cours de signaux certains) en faisant apparaître la distribution de Dirac. La figure 2.6 donne une illustration de cette situation. La densité de probabilité correspondante s'écrit alors

$$p_Y(y) = \frac{dF_Y(y)}{dy} + p_i \delta(y - y_i).$$

Néanmoins, dans de nombreuses situations, il sera possible d'utiliser le théorème suivant.

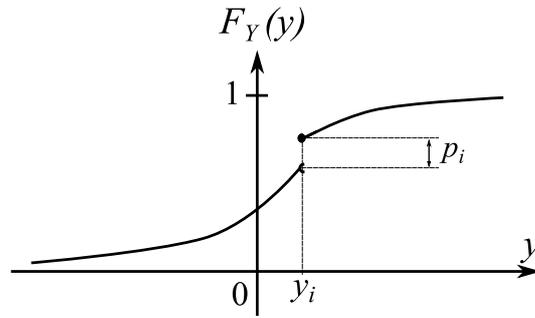


FIGURE 2.6 – Fonction de répartition présentant une discontinuité

**Théorème 6**

Soit  $X$  une VA continue de densité  $p_X$ . Soient  $g : X \mapsto Y$  un changement de variable aléatoire et  $y \in \mathbb{R}$ , tels que l'équation

$$y = g(x)$$

admette un ensemble au plus dénombrable de solutions, noté  $S_y$ , et tel que  $\forall x \in S_y, g'(x) \neq 0$ . Alors la densité de probabilité  $p_Y$  de  $Y$  est liée à celle de  $X$  par

$$p_Y(y) = \sum_{x \in S_y} \frac{p_X(x)}{|g'(x)|}$$

Dans le cas où la fonction  $g$  est bijective, le résultat précédent se simplifie puisque l'équation  $y = g(x)$  admet une unique solution :

$$p_Y(y) = \left. \frac{p_X(x)}{|g'(x)|} \right|_{x=g^{-1}(y)}$$

☞ **Démonstration** : Dans le cas bijectif : si  $g$  est croissante, on a

$$F_Y(y) = \mathbb{P}(\{Y \leq y\}) = \mathbb{P}(\{g(X) \leq y\}) = \mathbb{P}(\{X \leq g^{-1}(y)\}) = F_X(g^{-1}(y))$$

et donc

$$p_Y(y) = F'_Y(y) = F'_X(g^{-1}(y))(g^{-1})'(y) = p_X(x)/g'(x).$$

Si  $g$  est décroissante, cela devient

$$F_Y(y) = \mathbb{P}(\{Y \leq y\}) = \mathbb{P}(\{g(X) \leq y\}) = \mathbb{P}(\{X \geq g^{-1}(y)\}) = 1 - F_X(g^{-1}(y))$$

et donc

$$p_Y(y) = F'_Y(y) = -F'_X(g^{-1}(y))(g^{-1})'(y) = -p_X(x)/g'(x)$$

d'où le résultat.

Dans le cas général, on considère la situation de la figure 2.7. On voit que l'on peut écrire

$$\mathbb{P}(\{y_0 \leq Y \leq y_0 + dy\}) = \sum_{i \in I+} \mathbb{P}(\{x_i \leq X \leq x_i + dx\}) + \sum_{i \in I-} \mathbb{P}(\{x_i + dx \leq X \leq x_i\})$$

où  $I+$  regroupe les indices des antécédents de  $y_0$  autour desquels  $g$  est croissante et  $I-$  ceux des antécédents autour desquels  $g$  est décroissante.

Or si  $dy \rightarrow 0$ , on a  $\forall i, dx_i \rightarrow 0$  et  $dy/dx_i \rightarrow g'(x_i)$ . De plus,

$$\mathbb{P}(\{y_0 \leq Y \leq y_0 + dy\}) \sim p_Y(y_0)dy$$

$$\mathbb{P}(\{x_i \leq X \leq x_i + dx\}) \sim p_X(x_i)dx_i \quad \text{et} \quad \mathbb{P}(\{x_i + dx \leq X \leq x_i\}) \sim -p_X(x_i)dx_i$$

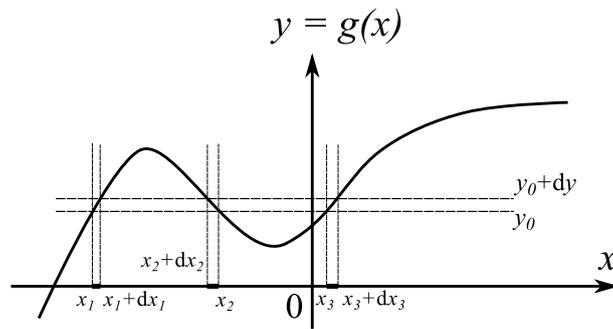


FIGURE 2.7 – Démonstration, cas général

En combinant ces différents éléments, il vient  $\forall y_0$

$$p_Y(y_0) \sim \sum_{i \in I+U \cup I-} p_X(x_i) |dx_i/dy|$$

et un passage à la limite avec  $dy \rightarrow 0$  permet de retrouver le résultat du théorème.

### 2.4.2 Fonction caractéristique et fonction génératrice

Deux types de fonction de variables aléatoires sont particulièrement utiles, notamment lors de l'étude de sommes de variables aléatoires : la fonction caractéristique et la fonction génératrice.

#### Définition 25

**Fonction caractéristique :** Soit  $X$  une variable aléatoire réelle, sa *fonction caractéristique*  $\varphi_X$  est définie par

$$\forall \nu \in \mathbb{R}, \varphi_X(\nu) = \mathbb{E} \left[ e^{i\nu X} \right]$$

Pour une variable aléatoire continue, on en déduit donc que la fonction caractéristique  $\varphi_X$  est la transformée de Fourier de sa densité de probabilité :

$$\varphi_X(\nu) = \int_{-\infty}^{+\infty} e^{i\nu x} p_X(x) dx$$

- **Remarque :**
  - La définition de la TF est ici légèrement différente de celle utilisée dans le cours de *Signaux certains* et reprend des conventions plus usuelles en probabilité. Néanmoins, tous les résultats et notions du cours de *Signaux certains* peuvent être appliqués à l'étude des fonctions caractéristiques.
  - On admettra que la fonction caractéristique caractérise entièrement la loi d'une VA, c'est à dire que deux variables aléatoires ont même fonction caractéristique si et seulement si leurs fonctions de répartition sont égales.
  - Dans le cas le plus général, la fonction caractéristique d'une variable aléatoire est la transformée de Fourier au sens des distributions de sa densité de probabilité, notion que vous verrez dans le cours d'intégration de François Goudail.

**Théorème 7**

Soit  $X$  une variable aléatoire réelle de fonction caractéristique  $\varphi_X$ . Si  $n$  est un entier naturel tel que  $\mathbb{E}[X^n]$  existe alors

$$\varphi_X(\nu) = \sum_{k=0}^n \frac{i^k \nu^k}{k!} \mathbb{E}[X^k] + o(\nu^n).$$

☞ **Démonstration** : Ce résultat est simplement la conséquence de la formule de Taylor appliquée à  $\varphi$  en 0, en remarquant que

$$\varphi_X^{(k)}(\nu) = i^k \mathbb{E}[X^k e^{i\nu X}].$$

Dans le cas d'une variable aléatoire discrète, il n'est pas toujours naturel de travailler avec la transformée de Fourier, c'est pourquoi dans ce cas on préférera parfois utiliser les fonctions génératrices.

**Définition 26**

**Fonction génératrice** : Soit  $X$  une variable aléatoire discrète à valeurs dans  $\mathbb{N}$  telle  $p_n = \mathbb{P}(\{X = n\})$ , sa *fonction génératrice*  $G_X$  est définie par

$$\forall t \in [-1, 1], G_X(t) = \mathbb{E}[t^X] = \sum_{n=0}^{\infty} p_n t^n$$

La définition de fonction génératrice  $G_X(t) = \mathbb{E}[t^X]$  peut s'appliquer à toute VA et pour tout  $t$  où l'espérance est définie.

**Définition 27**

**Fonction génératrice des moments** : Soit  $X$  une variable aléatoire, la *fonction génératrice des moments*  $M_X$  de  $X$  est définie par

$$M_X(t) = \mathbb{E}[e^{tX}]$$

# Chapitre 3

## Ensembles de variables aléatoires

Il est courant dans certains problèmes physiques de devoir considérer un ensemble de variables aléatoires. C'est par exemple le cas pour des problèmes en traitement d'images où l'on doit modéliser une image de  $N \times N$  pixels comme un  $N^2$ -uplet de variables aléatoires. Dans ces problèmes, il est alors nécessaire de prendre en compte d'éventuels liens entre les variables aléatoires du  $N^2$ -uplet. Ces considérations sont au centre du présent chapitre.

### 3.1 Couples de variables aléatoires

#### 3.1.1 Définitions

**Définition 28**

**Couples de variables aléatoires :** Soit l'espace probabilisé  $(\Omega, \Sigma, \mathbb{P})$  et soient deux variables aléatoires  $X$  et  $Y$  définies sur cet espace. On appelle *couple de variables aléatoires* le couple  $(X, Y)$  si l'application

$$(X, Y) : \Omega \longrightarrow \mathbb{R}^2 \\ \omega \longmapsto (X(\omega), Y(\omega))$$

est mesurable.

Dans le chapitre précédent, nous avons admis que toute l'information sur  $X$  et  $Y$  était contenue dans leurs fonctions de répartition respectives  $F_X$  et  $F_Y$ . Néanmoins, ces fonctions ne permettent pas d'avoir d'information simultanée sur les propriétés statistiques de  $X$  et  $Y$ . Il est donc nécessaire de faire la distinction entre les caractéristiques statistiques des deux variables aléatoires individuellement (on parlera alors de caractéristiques marginales) et leurs caractéristiques dites conjointes.

**Définition 29**

**Fonction de répartition conjointe :** Soit  $(X, Y)$  un couple de variables aléatoires. On appelle *fonction de répartition (conjointe)* du couple  $(X, Y)$  la fonction définie par

$$F_{XY}(x, y) = \mathbb{P}(\{(X \leq x) \cap (Y \leq y)\})$$

On a alors, pour tout  $x, y \in \mathbb{R}$ ,  $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0$  et  $F_{XY}(+\infty, +\infty) = 1$ .

On retrouve les mêmes propriétés que dans le cas 1D :

$$\mathbb{P}(\{(x_1 < X \leq x_2) \cap (y_1 < Y \leq y_2)\}) = F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1).$$

En effet, on a

$$\{(x_1 < X \leq x_2) \cap (Y \leq y_2)\} = \{(x_1 < X \leq x_2) \cap (Y \leq y_1)\} \cup \{(x_1 < X \leq x_2) \cap (y_1 < Y \leq y_2)\}$$

Or les deux derniers événements sont exclusifs de sorte que :

$$\mathbb{P}(\{(x_1 < X \leq x_2) \cap (Y \leq y_2)\}) = \mathbb{P}(\{(x_1 < X \leq x_2) \cap (Y \leq y_1)\}) + \mathbb{P}(\{(x_1 < X \leq x_2) \cap (y_1 < Y \leq y_2)\})$$

et donc

$$\mathbb{P}(\{(x_1 < X \leq x_2) \cap (y_1 < Y \leq y_2)\}) = \mathbb{P}(\{(x_1 < X \leq x_2) \cap (Y \leq y_2)\}) - \mathbb{P}(\{(x_1 < X \leq x_2) \cap (Y \leq y_1)\}) \quad (*)$$

De même, on a

$$\forall y_i, \quad \{(X \leq x_2) \cap (Y \leq y_i)\} = \{(X \leq x_1) \cap (Y \leq y_i)\} \cup \{(x_1 < X \leq x_2) \cap (Y \leq y_i)\}$$

comme les deux événements du terme de droite sont exclusifs, on a

$$\mathbb{P}(\{(X \leq x_2) \cap (Y \leq y_i)\}) = \mathbb{P}(\{(X \leq x_1) \cap (Y \leq y_i)\}) + \mathbb{P}(\{(x_1 < X \leq x_2) \cap (Y \leq y_i)\})$$

et donc

$$\mathbb{P}(\{(x_1 < X \leq x_2) \cap (Y \leq y_i)\}) = \mathbb{P}(\{(X \leq x_2) \cap (Y \leq y_i)\}) - \mathbb{P}(\{(X \leq x_1) \cap (Y \leq y_i)\}) \quad (*)$$

En utilisant l'équation (\*) dans (\*), on trouve :

$$\begin{aligned} \mathbb{P}(\{(x_1 < X \leq x_2) \cap (y_1 < Y \leq y_2)\}) &= \mathbb{P}(\{(X \leq x_2) \cap (Y \leq y_2)\}) - \mathbb{P}(\{(X \leq x_1) \cap (Y \leq y_2)\}) \\ &\quad - \left[ \mathbb{P}(\{(X \leq x_2) \cap (Y \leq y_1)\}) - \mathbb{P}(\{(X \leq x_1) \cap (Y \leq y_1)\}) \right] \\ &= \mathbb{P}(\{(X \leq x_2) \cap (Y \leq y_2)\}) - \mathbb{P}(\{(X \leq x_1) \cap (Y \leq y_2)\}) \\ &\quad - \mathbb{P}(\{(X \leq x_2) \cap (Y \leq y_1)\}) + \mathbb{P}(\{(X \leq x_1) \cap (Y \leq y_1)\}) \end{aligned}$$

qui donne le résultat voulu en appliquant la définition de la fonction de répartition.

### 3.1.2 Couples discrets

Dans la suite  $X$  et  $Y$  sont deux variables aléatoires discrètes de sorte que  $X(\Omega) = \{x_i \mid i \in \mathbb{N}\}$  et  $Y(\Omega) = \{y_i \mid i \in \mathbb{N}\}$ . La loi du couple  $(X, Y)$  est alors déterminée par l'ensemble des  $p_{ij}$  tels que

$$p_{ij} = \mathbb{P}(\{(X = x_i) \cap (Y = y_j)\}) \quad \text{et} \quad \sum_{i,j} p_{ij} = 1$$

#### Définition 30

**Loi marginale :** Soit  $(X, Y)$  un couple de variables aléatoires discrètes.

On appelle *loi marginale de X* la loi de la seule variable aléatoires  $X$ . Si  $p_i = \mathbb{P}(\{X = x_i\})$ , on a

$$p_i = \sum_{j=0}^{+\infty} \mathbb{P}(\{(X = x_i) \cap (Y = y_j)\}) = \sum_{j=0}^{+\infty} p_{ij}$$

De même, la *loi marginale de Y* est définie, par

$$p_j = \mathbb{P}(\{Y = y_j\}) = \sum_{i=0}^{+\infty} p_{ij}$$

### 3.1.3 Couples (absolument) continus

**Définition 31**

**Densité de probabilité conjointe :** Soit  $(X, Y)$  un couple de variables aléatoires absolument continues de densités de probabilité  $p_X$  et  $p_Y$ . On appelle *densité de probabilité conjointe* de  $(X, Y)$  la fonction positive sur  $\mathbb{R}^2$

$$p_{XY}(x, y) = \frac{\partial^2 F_{XY}}{\partial x \partial y}(x, y) \quad \text{telle que} \quad \iint p_{XY}(x, y) dx dy = 1.$$

On a donc

$$F_{XY}(x, y) = \mathbb{P}(\{(X \leq x) \cap (Y \leq y)\}) = \int_{-\infty}^x \int_{-\infty}^y p_{XY}(u, v) du dv$$

et pour tout borélien  $\mathcal{A} \in \mathfrak{B}(\mathbb{R}^2)$ ,

$$\mathbb{P}(\{(X, Y) \in \mathcal{A}\}) = \iint_{\mathcal{A}} p_{XY}(u, v) du dv$$

Or comme  $\{(X \leq x) \cap (Y \leq +\infty)\} = \{(X \leq x)\}$  on a ainsi  $F_{XY}(x, +\infty) = F_X(x)$  et donc, par définition des fonctions de répartition,

$$F_{XY}(x, +\infty) = \int_{-\infty}^x \left( \int_{-\infty}^{+\infty} p_{XY}(u, v) dv \right) du = F_X(x) = \int_{-\infty}^x p_X(u) du.$$

Un raisonnement similaire pour  $Y$  permet d'écrire que  $F_{XY}(+\infty, y) = F_Y(y)$ .

**Théorème 8**

**Densités de probabilité marginales :** Soit  $(X, Y)$  un couple de variables aléatoires absolument continues. Leurs *densités de probabilité marginales*  $p_X$  et  $p_Y$  sont données par

$$p_X(x) = \int_{-\infty}^{+\infty} p_{XY}(x, y) dy \quad \text{et} \quad p_Y(y) = \int_{-\infty}^{+\infty} p_{XY}(x, y) dx$$

### 3.1.4 Covariance

Tout comme pour les variables aléatoires simples, il est parfois plus pratique d'utiliser des quantités numériques issues de la loi ou densité conjointe de couple de V.A. que la loi / densité en elle-même. On peut alors définir aussi la notion de moments pour un couple de variables aléatoires.

**Définition 32**

Soit  $(X, Y)$  un couple de variables aléatoires et soient  $n$  et  $k$  deux entiers naturels. Les *moments d'ordre  $(n, k)$*  du couples, s'ils existent, sont les quantités

$$\mu_{n,k} = \iint_{\mathbb{R}^2} x^n y^k p_{XY}(x, y) dx dy = \mathbb{E} [X^n Y^k]$$

Si on note  $m_X$  et  $m_Y$  les espérance de  $X$  et  $Y$ , on définit les moments centrés par

$$\mu_{n,k}^c = \iint_{\mathbb{R}^2} (x - m_X)^n (y - m_Y)^k p_{XY}(x, y) dx dy = \mathbb{E} [(X - m_X)^n (Y - m_Y)^k]$$

**Définition 33**

**Covariance :** On appelle *covariance* du couple  $(X, Y)$  le moment centré d'ordre  $(1,1)$ . On le note  $\text{Cov}[X, Y]$  :

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

**Proposition 11**

**Propriétés de la covariance :** Soient  $X$  et  $Y$  deux variables aléatoires admettant des moments d'ordre 1 et 2, alors  $(X, Y)$  admet une covariance. De plus, on a les propriétés suivantes :

- $\text{Cov}[X, Y] = \text{Cov}[Y, X]$ ,
- $\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ ,
- $\text{Cov}[X, X] = \mathbb{V}[X]$ ,
- $\forall (a, b, c, d) \in \mathbb{R}^4, \text{Cov}[aX + b, cY + d] = ac\text{Cov}[X, Y]$ ,
- $\text{Cov}[X, Y] \leq \sqrt{\mathbb{V}[X]}\sqrt{\mathbb{V}[Y]}$ .

☞ **Démonstration :** Les quatre premières propriétés se déduisent facilement des définitions de la covariance, de la variance et des propriétés de l'espérance. La dernière propriété se déduit du Théorème 11 (voir plus bas) en remplaçant  $X$  et  $Y$  respectivement par  $(X - m_X)$  et  $(Y - m_Y)$ .

**Définition 34**

**Coefficient de corrélation :** On appelle *coefficient de corrélation* du couple  $(X, Y)$  la quantité :

$$\rho_{XY} = \frac{\text{Cov}[X, Y]}{\sqrt{\mathbb{V}[X]}\sqrt{\mathbb{V}[Y]}}$$

On a alors  $|\rho_{XY}| \leq 1$ .

On dit que les variables aléatoires  $X$  et  $Y$  sont *décorréliées* si leur coefficient de corrélation (ou leur covariance) est nul(le).

- **Remarque :**
  - Le coefficient de corrélation donne une mesure de la relation linéaire qui peut exister entre deux variables. Deux variables aléatoires dont le coefficient de corrélation vaut 1 sont presque sûrement reliées par une relation affine. Il est possible d'avoir une idée du niveau de corrélation entre deux variables aléatoires en regardant l'allure graphique de leur densité conjointe : plus celle-ci possèdera une forme symétrique par rapports aux axes  $OX$  et  $OY$  plus les variables seront décorréliées. Quelques exemples seront vus en TD.
  - Il convient d'être prudent dans l'utilisation du coefficient de corrélation. Suivant le contexte, un coefficient de  $\rho = 0.75$  pourra signifier ou non une relation intéressante entre les variables. La manière de calculer le coefficient et le choix des variables considérées restent cruciaux pour interpréter ce chiffre.

**3.1.5 Indépendance**

Soient  $X$  et  $Y$  deux variables aléatoires et soient  $\mathcal{A}$  et  $\mathcal{B}$  deux boréliens. On dira que  $X$  et  $Y$  sont indépendantes si et seulement si, pour tout  $\mathcal{A}$  et tout  $\mathcal{B}$ , les événements  $\{X \in \mathcal{A}\}$  et  $\{Y \in \mathcal{B}\}$  sont indépendants. Dit autrement, elles sont indépendantes si et seulement si on peut écrire

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

Néanmoins, en pratique on utilisera assez peu cette définition pour déterminer l'indépendance de variables aléatoires pour lui préférer les théorèmes qui en découlent.

**Théorème 9**

**Indépendance :** Les variables aléatoires  $X$  et  $Y$  sont indépendantes si et seulement si pour tous  $x, y \in \mathbb{R}$

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

Pour des variables aléatoires continues, cela est équivalent au fait que la densité conjointe est égale au produit des densités marginales :

$$p_{XY}(x, y) = p_X(x)p_Y(y), \text{ pour tous } x, y \in \mathbb{R}$$

☞ **Démonstration :** Ce théorème découle simplement de la définition précédente de l'indépendance et de la définition de la fonction de répartition conjointe.

■ **Remarque :** On verra en TD que la représentation graphique donne des indications sur les propriétés d'un couple aléatoire. On verra par exemple que si la représentation graphique de la densité conjointe de deux V.A. uniformes est un carré alors ces V.A. seront indépendantes.

**Théorème 10**

Si les variables aléatoires  $X$  et  $Y$  sont indépendantes alors elles sont décorrélées.

La réciproque est fausse.

☞ **Démonstration :** Ce résultat se démontre simplement en s'appuyant sur les définitions d'indépendance et de coefficient de corrélation. Un contre-exemple de la réciproque sera vu en TD.

■ **Remarque :** Il existe deux cas particuliers notables pour lesquels indépendance et décorrélation sont équivalentes : pour des variables aléatoires de Bernoulli et des variables aléatoires gaussiennes. Dans ces deux cas, les variables sont indépendantes si et seulement si leur covariance est nulle.

**Théorème 11**

**Inégalité de Cauchy-Schwartz :** Soient  $X$  et  $Y$  deux variables aléatoires admettant des moments d'ordre 1 et 2. Alors

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

et l'égalité a lieu si  $X$  et  $Y$  sont linéairement dépendants.

☞ **Démonstration :** Il faut voir que  $\mathbb{E}[(\lambda X - Y)^2]$  est un polynôme du second degré en  $\lambda$  qui n'est jamais négatif. Son discriminant  $\Delta$  est donc négatif :  $\Delta = 4\mathbb{E}[XY]^2 - 4\mathbb{E}[X^2]\mathbb{E}[Y^2] \leq 0$ , ce qui donne directement le résultat.

## 3.2 Fonction d'un couple de variables aléatoires

### 3.2.1 Image d'un couple de variables aléatoires

**Théorème 12**

Soient  $f$  et  $h$  deux fonctions mesurables de  $\mathbb{R}$  dans  $\mathbb{R}$ . Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes, alors  $f(X)$  et  $h(Y)$  sont aussi des variables aléatoires indépendantes.

☞ **Démonstration :** Ce théorème découle des Théorèmes 9 et 4, et du théorème du Fubini [6, 8].

**Théorème 13**

**Formule de transfert (admise) :** Soient  $X$  et  $Y$  deux variables aléatoires de densité conjointe  $p_{XY}$ . Soit  $g$  une fonction mesurable définie sur  $X(\Omega) \times Y(\Omega)$ . Alors  $g(X, Y)$  est une variable aléatoire et, si elle admet une espérance, on a

$$\mathbb{E}[g(X, Y)] = \iint_{X(\Omega) \times Y(\Omega)} g(x, y) p_{XY}(x, y) dx dy$$

De l'ensemble des résultats on déduit une propriété fondamentale de l'espérance d'un couple de VA indépendantes :

Soient  $X$  et  $Y$  deux variables aléatoires indépendantes admettant une espérance, alors  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

**3.2.2 Somme de variables aléatoires**

**Théorème 14**

**Cas continu :** Soient  $X$  et  $Y$  deux variables aléatoires de densités marginales respectives  $p_X$  et  $p_Y$  et de densité conjointe  $p_{XY}$ . Soit  $Z = X + Y$ . Alors une densité de  $Z$  est

$$p_Z(z) = \int_{-\infty}^{\infty} p_{XY}(u, z - u) du$$

Si  $X$  et  $Y$  sont indépendantes, alors  $p_Z = p_X * p_Y$  où  $*$  est le produit de convolution.

☞ **Démonstration :** On a par définition

$$\begin{aligned} F_Z(z) &= \mathbb{P}(\{Z \leq z\}) = \int_{\mathcal{D}} p_{XY}(x, y) dx dy \quad \text{avec } \mathcal{D} = \{(x, y) \in \mathbb{R}^2 | x + y \leq z\} \\ &= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{z-x} p_{XY}(x, y) dy \right) dx \\ &= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^z p_{XY}(x, u - x) du \right) dx \quad \text{avec le changement de variable } u = x + y \\ &= \int_{-\infty}^z \left( \int_{-\infty}^{+\infty} p_{XY}(x, u - x) dx \right) du \quad \text{en appliquant le théorème de Fubini [6, 8]} \end{aligned}$$

La dérivation par  $z$  permet d'obtenir le résultat.

Dans le cas de variables aléatoires discrètes, on se ramène à des sommes discrètes, de sorte que si les variables sont indépendantes on trouve que

$$\mathbb{P}(\{X + Y = n\}) = \sum_{k \in \mathbb{Z}} \mathbb{P}(\{X = k\}) \cdot \mathbb{P}(\{Y = n - k\})$$

**Théorème 15**

**Variance d'une somme :** Soient  $X$  et  $Y$  deux variables aléatoires admettant une variance. Alors  $(X, Y)$  admet une covariance et

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}[X, Y]$$

☞ **Démonstration :** Il suffit d'utiliser les définitions de la variance et de la covariance, et les propriétés de l'espérance.

De ce théorème, on retrouve le résultat fondamental que si  $X$  et  $Y$  sont deux variables aléatoires indépendantes admettant une variance, on a

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$$

### Stabilité

On dit qu'une loi est *stable par la somme* quand la somme de deux variables suivant cette loi est elle-même une variable aléatoire suivant la dite loi. On montre en particulier que, dans le cas de variables aléatoires indépendantes, la loi de Poisson et la loi normale sont stables par la somme.

#### Théorème 16

**Somme de lois de Poisson :** Soient  $X$  et  $Y$  deux variables aléatoires indépendantes suivant des lois de Poisson de paramètres respectifs  $\lambda$  et  $\mu$ . Alors  $X + Y$  suit une loi de Poisson de paramètre  $\lambda + \mu$

☞ **Démonstration :** Il faut montrer que la loi de  $Z = X + Y$  est de la forme

$$\mathbb{P}(\{Z = k\}) = \frac{e^{-(\lambda+\mu)}}{k!} (\lambda + \mu)^k$$

en utilisant l'extension du Théorème 14 aux VA discrètes. Une autre méthode plus rapide consiste à utiliser les Théorèmes 18 ou 19.

#### Théorème 17

**Somme de lois normales :** Soient  $X$  et  $Y$  deux variables aléatoires indépendantes suivant des lois normales  $\mathcal{N}(m_X, \sigma_X)$  et  $\mathcal{N}(m_Y, \sigma_Y)$ . Alors  $X + Y$  suit une loi normale  $\mathcal{N}(m_X + m_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})$ .

☞ **Démonstration :** Il faut montrer que la densité de  $Z = X + Y$  est de la forme

$$p_Z(z) = \frac{1}{\sqrt{2\pi(\sigma_X^2 + \sigma_Y^2)}} \exp\left(-\frac{(z - m_X - m_Y)^2}{2(\sigma_X^2 + \sigma_Y^2)}\right)$$

en utilisant l'extension du Théorème 14. Là encore, une méthode plus rapide consiste à utiliser le Théorème 18.

### Fonction caractéristique et fonction génératrice

#### Théorème 18

La fonction caractéristique de la somme de deux variables aléatoires indépendantes est le produit de leurs fonctions caractéristiques :

$$\varphi_{X+Y}(\nu) = \varphi_X(\nu) \cdot \varphi_Y(\nu)$$

#### Théorème 19

De même, la fonction génératrice de la somme de deux variables aléatoires indépendantes est le produit de leurs fonctions génératrices :

$$G_{X+Y}(t) = G_X(t) \cdot G_Y(t)$$

☞ **Démonstration :** Ces deux résultats découlent du Théorème 12 et des propriétés de l'espérance pour des produits de VA indépendantes.

### 3.3 Vecteurs aléatoires

Il est possible de généraliser les résultats précédents aux ensembles de  $n$  variables aléatoires.

#### 3.3.1 Définitions

##### Définition 35

**Vecteur aléatoire :** Soit l'espace probabilisé  $(\Omega, \Sigma, \mathbb{P})$ . On appelle *vecteur aléatoire à  $n$  dimensions* un  $n$ -uplet de variables aléatoires, c'est-à-dire la fonction  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  définie par

$$\begin{aligned} \mathbf{X} : \Omega &\longrightarrow \mathbb{R}^n \\ \omega &\longmapsto (X_1(\omega), X_2(\omega), \dots, X_n(\omega))^T \end{aligned}$$

où  $T$  représente l'opérateur de transposition<sup>1</sup>.

##### Définition 36

**Fonction de répartition conjointe :** On appelle fonction de répartition conjointe du vecteur  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  la fonction  $F_{\mathbf{X}}$  de  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  définie par

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = \mathbb{P}(\{(X_1 \leq x_1) \cap \dots \cap (X_n \leq x_n)\})$$

On a alors :

- $F_{\mathbf{X}}(-\infty, x_2, \dots, x_n) = \dots = F_{\mathbf{X}}(x_1, \dots, -\infty, \dots, x_n) = \dots = F_{\mathbf{X}}(x_1, x_2, \dots, -\infty) = 0$ ,
- $F_{\mathbf{X}}(+\infty, +\infty, \dots, +\infty) = 1$ .

##### Définition 37

**Densité de probabilité conjointe :** On appelle *densité de probabilité conjointe* du vecteur aléatoire  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ , la fonction définie par

$$p_{\mathbf{X}}(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_{\mathbf{X}}(x_1, x_2, \dots, x_n)$$

On a alors

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} p_{\mathbf{X}}(u_1, u_2, \dots, u_n) du_1 du_2 \dots du_n$$

et l'on peut généraliser la notion de densité marginale vue à la section précédente :

##### Proposition 12

**Densité marginale :** Soient  $\mathbf{X}$  et  $\mathbf{Y}$  deux vecteurs aléatoires à valeurs dans  $\mathbb{R}^l$  et  $\mathbb{R}^m$  respectivement, tels que le vecteur aléatoire  $(\mathbf{X}, \mathbf{Y})$  de  $\mathbb{R}^n = \mathbb{R}^{l+m}$  admette la densité de probabilité  $p_{(\mathbf{X}, \mathbf{Y})}$ . Alors les vecteurs  $\mathbf{X}$  et  $\mathbf{Y}$  admettent les densités de probabilités  $p_{\mathbf{X}}$  et  $p_{\mathbf{Y}}$  données respectivement

1. On choisit ici la convention classique en mathématiques qui consiste à considérer des vecteurs colonnes, d'où la nécessité d'utiliser la transposition dans la définition ci-dessus. Il s'agit en fait d'une astuce courante pour gagner de la place dans les textes puisque

$$(x_1, x_2, \dots, x_n)^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

par

$$\begin{cases} p_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} p_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, y_1, \dots, y_m) dy_1 \cdots dy_m = \int_{\mathbb{R}^m} p_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ p_{\mathbf{Y}}(\mathbf{y}) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} p_{(\mathbf{X}, \mathbf{Y})}(x_1, \dots, x_l, \mathbf{y}) dx_1 \cdots dx_l = \int_{\mathbb{R}^l} p_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y}) d\mathbf{x} \end{cases}$$

On appelle  $p_{\mathbf{X}}(\mathbf{x})$  et  $p_{\mathbf{Y}}(\mathbf{y})$  les *densités marginales* de  $(\mathbf{X}, \mathbf{Y})$  sur les  $l$  premières composantes et les  $m = n - l$  dernières composantes respectivement.

☞ **Démonstration** : On admettra ce résultat dont la démonstration s'appuie sur le théorème de Fubini. On notera que l'énoncé précédent se généralise à la définition des densités marginales sur des composantes quelconques.

■ **Remarque** : On retrouve aussi la généralisation de la propriété sur la densité de probabilité d'un vecteur aléatoire à valeur dans  $\mathbb{R}^n$  :

$$\int_{\mathbb{R}^n} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = 1.$$

De plus, on notera que la notation  $(\mathbf{X}, \mathbf{Y})$  désigne ici le vecteur  $\mathbb{R}^{l+m}$  créé à partir de la concaténation des composantes de  $\mathbf{X}$  et  $\mathbf{Y}$ . Il en sera ainsi dans toute la suite du cours.

### 3.3.2 Espérance et covariance

#### Définition 38

**Espérance d'un vecteur aléatoire** : Soit  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  un vecteur aléatoire dont les composantes admettent une espérance. On appelle *espérance du vecteur aléatoire  $\mathbf{X}$*  le vecteur (colonne) défini par

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix}$$

La formule du transfert se généralise aussi à un vecteur aléatoire de taille quelconque :

#### Théorème 20

**Formule de transfert (admise)** : Soit  $\mathbf{X}$  un vecteur aléatoire de densité conjointe  $p_{\mathbf{X}}$ . Soit  $g$  une fonction mesurable définie de  $\mathbb{R}^n$  dans  $\mathbb{R}$ . Alors  $g(\mathbf{X})$  est une variable aléatoire et, si elle admet une espérance, on a

$$\mathbb{E}[g(\mathbf{X})] = \int_{\mathbb{R}^n} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

La formule reste vraie dans le cas où  $g$  est une fonction de  $\mathbb{R}^n$  dans  $\mathbb{C}$ . Néanmoins on se montrera très prudent dans les cas où l'on doit manipuler des variables aléatoires définies comme complexes (cf. TD).

#### Définition 39

**Matrice de covariance** : Soit  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  un vecteur aléatoire dont les composantes admettent des moments d'ordre 1 et 2. On appelle *matrice de covariance* la matrice définie par

$$\text{Cov}[\mathbf{X}] = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T].$$

La notion de matrice de covariance est la généralisation aux vecteurs aléatoires de la notion de covariance. En particulier en dimension 2, on peut voir que

$$\text{Cov}[\mathbf{X}] = \begin{pmatrix} \mathbb{V}[X_1] & \text{Cov}[X_1, X_2] \\ \text{Cov}[X_1, X_2] & \mathbb{V}[X_2] \end{pmatrix}$$

- **Remarque :** On retrouve la notion de matrice de covariance en traitement du signal (notamment en théorie de l'estimation), en polarisation et en physique statistique. La matrice de covariance d'un vecteur aléatoire est symétrique et positive :

$$\forall \mathbf{u} \in \mathbb{R}^n, \mathbf{u}^T \text{Cov}[\mathbf{X}] \mathbf{u} \geq 0.$$

Cela vient du fait que  $\mathbf{u}^T \text{Cov}[\mathbf{X}] \mathbf{u} = \mathbb{V}[u_1 X_1 + u_2 X_2 + \dots + u_n X_n]$ .

**Définition 40**

**Matrice de corrélation :** Soient  $\mathbf{X}$  et  $\mathbf{Y}$  deux vecteurs aléatoires d'espérances  $\mathbf{m}_\mathbf{X}$  et  $\mathbf{m}_\mathbf{Y}$ . Les inter-corrélations de  $\mathbf{X}$  et  $\mathbf{Y}$  sont décrites par la *matrice de corrélation (ou d'intercorrélations)*

$$\Gamma_{\mathbf{X}\mathbf{Y}} = \mathbb{E}[(\mathbf{X} - \mathbf{m}_\mathbf{X})(\mathbf{Y} - \mathbf{m}_\mathbf{Y})^T]$$

On dit que  $\mathbf{X}$  et  $\mathbf{Y}$  sont non corrélés si leur matrice de corrélation est nulle.

La matrice de covariance (ou matrice d'autocorrélations) est un cas particulier de matrice d'intercorrélations pour lequel  $\mathbf{X} = \mathbf{Y}$ .

### 3.3.3 Fonction d'un vecteur aléatoire

#### Fonction caractéristique

**Définition 41**

Soit  $\mathbf{X}$  un vecteur variable aléatoire réel, sa *fonction caractéristique*  $\varphi_\mathbf{X}$  est définie par

$$\forall \boldsymbol{\nu} \in \mathbb{R}^n, \varphi_\mathbf{X}(\boldsymbol{\nu}) = \varphi_\mathbf{X}(\nu_1, \dots, \nu_n) = \mathbb{E} \left[ e^{i \sum_{k=1}^n \nu_k X_k} \right] = \mathbb{E} \left[ e^{i \boldsymbol{\nu}^T \mathbf{X}} \right]$$

Comme dans le cas d'une variable aléatoire, la fonction  $\varphi_\mathbf{X}$  caractérise la loi de  $\mathbf{X}$  de sorte que deux vecteurs aléatoires ayant la même fonction caractéristique ont la même fonction de répartition.

#### Image d'un vecteur aléatoire

**Théorème 21**

Soit  $g : \mathbf{X} \mapsto \mathbf{Y}$  un changement de vecteur aléatoire de  $\mathbb{R}^n$  dans  $\mathbb{R}^n$ . Soit  $\mathbf{y} \in \mathbb{R}^n$  tel que l'équation  $\mathbf{y} = g(\mathbf{x})$  admette un nombre dénombrable de solutions  $\mathbf{x}_k, k \in I$ . Alors la densité de probabilité  $p_\mathbf{Y}$  de  $\mathbf{Y}$  est liée à celle de  $\mathbf{X}$  par

$$p_\mathbf{Y}(\mathbf{y}) = \sum_{k \in I} \frac{p_\mathbf{X}(\mathbf{x}_k)}{|\det(\mathbf{J}_k)|}$$

où  $\mathbf{J}$  est la matrice jacobienne du changement de variable :

$$\mathbf{J}_k = \left( \begin{array}{ccc} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{array} \right) \Bigg|_{\mathbf{x}_k} .$$

■ **Remarque :** Ce théorème, que nous admettrons, ne peut s'appliquer que si le jacobien (déterminant de la matrice jacobienne) du changement est non nul.

De plus, comme nous le verrons en TD, ce théorème peut être pratique pour retrouver les densités de probabilité de fonction de couples de VA (somme, différence, quotient, produit, etc.)

### 3.3.4 Indépendance

La notion d'indépendance vue pour un couple de VA ainsi que les résultats qui en découlent peuvent aussi être généralisées aux vecteurs aléatoires

#### **Théorème 22**

Soit  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  un vecteur aléatoire. Les propositions suivantes sont équivalentes :

- $X_1, X_2, \dots, X_n$  sont mutuellement indépendantes,
- pour tout ensemble de boréliens  $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ ,

$$\mathbb{P}(\{X_1 \in \mathcal{A}_1\} \cap \dots \cap \{X_n \in \mathcal{A}_n\}) = \mathbb{P}(\{X_1 \in \mathcal{A}_1\}) \mathbb{P}(\{X_2 \in \mathcal{A}_2\}) \cdots \mathbb{P}(\{X_n \in \mathcal{A}_n\})$$

- la fonction de répartition conjointe est le produit des fonctions de répartition marginales

$$\forall \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n, F_{\mathbf{X}}(x_1, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n),$$

- pour des V.A. continues, la densité conjointe est le produit des densités marginales

$$\forall \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n, p_{\mathbf{X}}(x_1, \dots, x_n) = p_{X_1}(x_1) p_{X_2}(x_2) \cdots p_{X_n}(x_n).$$

- la fonction caractéristique du vecteur  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  est égale au produit des fonctions caractéristiques de ses composantes :

$$\forall \boldsymbol{\nu} \in \mathbb{R}^n, \varphi_{\mathbf{X}}(\boldsymbol{\nu}) = \varphi_{\mathbf{X}}(\nu_1, \dots, \nu_n) = \varphi_{X_1}(\nu_1) \varphi_{X_2}(\nu_2) \cdots \varphi_{X_n}(\nu_n)$$

Si  $X_1, X_2, \dots, X_n$  sont mutuellement indépendantes, on a alors que l'espérance du produit est égale au produit des espérances

$$\mathbb{E}[X_1 X_2 \cdots X_n] = \mathbb{E}[X_1] \mathbb{E}[X_2] \cdots \mathbb{E}[X_n]$$

■ **Remarque :** Si  $X_1, X_2, \dots, X_n$  sont mutuellement indépendantes, alors elles sont indépendantes deux à deux. Mais encore une fois, la réciproque est fautive.

Les propositions précédentes s'appliquent aussi entre deux vecteurs aléatoires  $\mathbf{X}$  et  $\mathbf{Y}$ . Par exemple, si  $\mathbf{X}$  et  $\mathbf{Y}$  sont indépendants alors  $\mathbb{E}[\mathbf{X}\mathbf{Y}] = \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{Y}]$ .

**Théorème 23**

**Additivité des variances :** Soient  $X_1, X_2, \dots, X_n$  des variables aléatoires, et  $Y$  définie par

$$Y = \sum_{k=1}^n a_k X_k$$

où les  $\{a_k\}$  sont des réels. Alors

$$\mathbb{E}[Y] = \sum_{k=1}^n a_k \mathbb{E}[X_k] \quad \text{et} \quad \mathbb{V}[Y] = \sum_{k=1}^n a_k^2 \mathbb{V}[X_k] + 2 \sum_{1 \leq i < j \leq n} a_i a_j \text{Cov}[X_i, X_j]$$

Si les variables sont indépendantes deux à deux, on a en particulier

$$\mathbb{V}[Y] = \sum_{k=1}^n a_k^2 \mathbb{V}[X_k]$$

Là encore il s'agit de la généralisation des mêmes résultats pour un couple de VA. On retrouve aussi le résultat suivant :

**Théorème 24**

Soient  $X_1, X_2, \dots, X_n$  des variables aléatoires, et  $Y$  définie par

$$Y = \sum_{k=1}^n X_k.$$

Alors si les variables  $X_k$  sont mutuellement indépendantes, on a

$$\forall \nu \in \mathbb{R}, \varphi_Y(\nu) = \varphi_{X_1}(\nu) \varphi_{X_2}(\nu) \cdots \varphi_{X_n}(\nu)$$

# Chapitre 4

## VA et vecteurs aléatoires gaussiens

Les variables et vecteurs aléatoires gaussiens tiennent une place particulièrement intéressante en physique. Ils ont un rôle fondamental dans un certain nombre de résultats importants, comme nous le verrons au chapitre suivant, et sont très utilisés notamment en traitement du signal et en automatique, comme vous le verrez plus tard dans l'année ou en 2ème année. De plus, les vecteurs gaussiens jouent un rôle particulier dans des problèmes d'estimation faisant intervenir la notion d'espérance conditionnelle (le filtrage de Kalman que vous verrez en 2ème année). Cette notion est développée dans une deuxième partie du chapitre. La plupart des notions et explications données dans ce chapitre proviennent de la Référence [7].

### 4.1 Vecteur aléatoire gaussien

#### 4.1.1 Définition

Les variables aléatoires gaussiennes ont été définies au chapitre 2. Nous rappelons ici quelques définitions et résultats intéressants avant de nous concentrer sur les vecteurs aléatoires gaussiens.

■ **Rappel :** Une variable aléatoire réelle  $X$  est dite gaussienne (ou normale) réduite si elle admet la densité de probabilité

$$p_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right].$$

Pour tout couple  $(\sigma, m) \in \mathbb{R}^{+*} \times \mathbb{R}$ , la variable aléatoire  $Y = \sigma X + m$  est gaussienne, et admet alors pour espérance  $m$ , pour variance  $\sigma^2$  et sa densité de probabilité est donnée par

$$p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-m)^2}{2\sigma^2}\right].$$

La fonction caractéristique de la variable aléatoire  $Y$  est

$$\varphi_Y(\nu) = \exp\left[i\nu m - \frac{\sigma^2 \nu^2}{2}\right].$$

#### Définition 42

Soit  $\mathbf{X} = (X_1, \dots, X_N)^T$  un vecteur aléatoire réel de dimension  $N$ . On dit que  $\mathbf{X}$  est un vecteur aléatoire gaussien si toute combinaison affine de ses coordonnées  $Y = a_0 + \sum_{k=1}^N a_k X_k$ ,  $\forall (a_0, \dots, a_N) \in \mathbb{R}^{N+1}$  est une variable aléatoire gaussienne.

■ **Remarque :** Soit  $\mathbf{X}$  un vecteur gaussien. Selon la définition 41, sa fonction caractéristique est donnée par

$$\forall \boldsymbol{\nu} \in \mathbb{R}^n, \varphi_{\mathbf{X}}(\boldsymbol{\nu}) = \varphi_{\mathbf{X}}(\nu_1, \dots, \nu_n) = \mathbb{E} \left[ \exp \left( i \sum_{k=1}^n \nu_k X_k \right) \right].$$

La variable aléatoire  $Y = \sum_{k=1}^n \nu_k X_k$  est donc gaussienne et selon le rappel précédent, on a

$$\varphi_{\mathbf{X}}(\nu_1, \dots, \nu_n) = \varphi_Y(1) = \exp \left[ i m_Y - \frac{\sigma_Y^2}{2} \right]$$

avec

$$m_Y = \mathbb{E}[Y] = \mathbb{E} \left[ \sum_{k=1}^n \nu_k X_k \right] = \sum_{k=1}^n \nu_k \mathbb{E}[X_k] = \boldsymbol{\nu}^T \mathbf{m}_{\mathbf{X}}$$

et

$$\sigma_Y = \mathbb{E}[(Y - m_Y)^2] = \mathbb{E} \left[ \left( \sum_{k=1}^n \nu_k [X_k - m_{X_k}] \right)^2 \right] = \sum_{k=1}^n \sum_{l=1}^n \nu_k \nu_l \mathbb{E}[(X_k - m_{X_k})(X_l - m_{X_l})] = \boldsymbol{\nu}^T \boldsymbol{\Gamma}_{\mathbf{X}} \boldsymbol{\nu}$$

On a donc

$$\forall \boldsymbol{\nu} \in \mathbb{R}^n, \varphi_{\mathbf{X}}(\boldsymbol{\nu}) = \exp \left[ i \boldsymbol{\nu}^T \mathbf{m}_{\mathbf{X}} - \frac{1}{2} \boldsymbol{\nu}^T \boldsymbol{\Gamma}_{\mathbf{X}} \boldsymbol{\nu} \right]$$

où  $\boldsymbol{\Gamma}_{\mathbf{X}}$  est la matrice de covariance du vecteur  $\mathbf{X}$ . Si cette matrice est inversible, on peut montrer qu'il existe une matrice  $\mathbf{A}$  inversible telle que  $\boldsymbol{\Gamma}_{\mathbf{X}} = \mathbf{A} \mathbf{A}^T$ . On supposera que les matrices de covariance  $\boldsymbol{\Gamma}_{\mathbf{X}}$  sont inversibles dans la suite de ce cours.

### 4.1.2 Densité de probabilité

Soit  $\mathbf{X}$  un vecteur aléatoire gaussien de dimension  $N$ , d'espérance  $\mathbf{m}_{\mathbf{X}} = \mathbb{E}[\mathbf{X}]$  et de matrice de covariance  $\boldsymbol{\Gamma}_{\mathbf{X}}$  inversible.

Soit  $\mathbf{Y}$  un vecteur aléatoire tel que  $\mathbf{Y} = \mathbf{A}^{-1}(\mathbf{X} - \mathbf{m}_{\mathbf{X}})$ . On a alors  $\mathbf{m}_{\mathbf{Y}} = \mathbf{0}$  et

$$\boldsymbol{\Gamma}_{\mathbf{Y}} = \mathbf{A}^{-1} \boldsymbol{\Gamma}_{\mathbf{X}} (\mathbf{A}^{-1})^T = \mathbf{A}^{-1} \mathbf{A}^T \mathbf{A} (\mathbf{A}^{-1})^T = \mathbf{I}.$$

De plus, les composantes de  $\mathbf{Y}$  – étant des combinaisons affines des composantes de  $\mathbf{X}$  – sont des VA gaussiennes centrées réduites par définition de  $\mathbf{Y}$ . Ce qui amène à

$$\forall \boldsymbol{\nu} \in \mathbb{R}^n, \varphi_{\mathbf{Y}}(\boldsymbol{\nu}) = \prod_{k=1}^N e^{-\frac{1}{2} \nu_k^2} = \prod_{k=1}^N \varphi_{Y_k}(\nu_k)$$

Ce qui revient à dire que les VA  $Y_k$  sont mutuellement indépendantes et donc que la densité de probabilité de  $\mathbf{Y}$  est égale au produit des densités de probabilité de ses composantes. Les VA  $Y_k$  étant gaussiennes centrées réduites on a

$$p_{\mathbf{Y}}(\mathbf{y}) = \prod_{k=1}^N \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{y_k^2}{2} \right] = \frac{1}{\sqrt{2\pi}^N} e^{-\frac{1}{2} \sum_{k=1}^N y_k^2} = \frac{1}{\sqrt{2\pi}^N} \exp \left[ -\frac{1}{2} \mathbf{y}^T \mathbf{y} \right].$$

En utilisant la relation entre  $\mathbf{Y}$  et  $\mathbf{X}$  et le théorème 21, on trouve que

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{|\det(\mathbf{A}^{-1})|}{(\sqrt{2\pi})^N} \exp \left[ -\frac{1}{2} [\mathbf{A}^{-1}(\mathbf{x} - \mathbf{m}_{\mathbf{X}})]^T [\mathbf{A}^{-1}(\mathbf{x} - \mathbf{m}_{\mathbf{X}})] \right]$$

---

1. Notamment comme  $\boldsymbol{\Gamma}_{\mathbf{X}}$  est positive, on peut la diagonaliser sous la forme  $\boldsymbol{\Gamma}_{\mathbf{X}} = \mathbf{P} \mathbf{D} \mathbf{P}^T$  avec  $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$ . Il suffit alors de prendre  $\mathbf{A} = \mathbf{P} \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ .

et donc que

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^N \sqrt{\det \mathbf{\Gamma}_{\mathbf{x}}}} \exp \left[ -\frac{1}{2}(\mathbf{x} - \mathbf{m}_{\mathbf{x}})^T \mathbf{\Gamma}_{\mathbf{x}}^{-1}(\mathbf{x} - \mathbf{m}_{\mathbf{x}}) \right].$$

Ce qui permet de déduire les résultats suivants :

**Théorème 25**

Un vecteur aléatoire gaussien  $\mathbf{X}$  de dimension  $N$ , d'espérance  $\mathbf{m}_{\mathbf{x}} = \mathbb{E}[\mathbf{X}]$  et de matrice de covariance  $\mathbf{\Gamma}_{\mathbf{x}}$  a une densité de probabilité définie par

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^N \sqrt{\det \mathbf{\Gamma}_{\mathbf{x}}}} \exp \left[ -\frac{1}{2}(\mathbf{x} - \mathbf{m}_{\mathbf{x}})^T \mathbf{\Gamma}_{\mathbf{x}}^{-1}(\mathbf{x} - \mathbf{m}_{\mathbf{x}}) \right]$$

et sa fonction caractéristique est donnée par

$$\forall \boldsymbol{\nu} \in \mathbb{R}^N, \varphi_{\mathbf{x}}(\boldsymbol{\nu}) = \exp [i\boldsymbol{\nu}^T \mathbf{m}_{\mathbf{x}}] \exp \left[ -\frac{1}{2}\boldsymbol{\nu}^T \mathbf{\Gamma}_{\mathbf{x}} \boldsymbol{\nu} \right]$$

- **Remarque :** En pratique on prendra ce théorème comme "définition" d'un vecteur aléatoire gaussien.

### 4.1.3 Indépendance et corrélation

**Théorème 26**

Soit un vecteur aléatoire gaussien  $\mathbf{X}$  de dimension  $N$  de matrice de covariance  $\mathbf{\Gamma}_{\mathbf{x}}$  alors  $\mathbf{\Gamma}_{\mathbf{x}} = \text{diag}(\sigma_1^2, \dots, \sigma_N^2)$  si et seulement si les coordonnées  $X_k$  du vecteur sont des VA gaussiennes mutuellement indépendantes de variances respectivement  $\sigma_k^2$ .

☞ **Démonstration :** Ce résultat provient des définitions de la matrice de covariance et de la fonction caractéristique d'une VA gaussienne, combinées au résultat du théorème 22 sur les fonctions caractéristiques.

**Définition 43**

On dit que deux vecteurs aléatoires  $\mathbf{X}$  et  $\mathbf{Y}$  sont gaussiens dans leur ensemble si toute combinaison affine de leurs coordonnées est une variable aléatoire gaussienne.

- **Remarque :** Soient  $\mathbf{X}$  et  $\mathbf{Y}$  deux vecteurs aléatoires gaussiens dans leur ensemble. Le vecteur aléatoire  $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$  est un vecteur aléatoire gaussien de matrice de covariance

$$\mathbf{\Gamma}_{\mathbf{z}} = \begin{pmatrix} \mathbf{\Gamma}_{\mathbf{x}} & \mathbf{\Gamma}_{\mathbf{xY}} \\ \mathbf{\Gamma}_{\mathbf{YX}} & \mathbf{\Gamma}_{\mathbf{Y}} \end{pmatrix}$$

**Proposition 13**

Soient  $\mathbf{X}$  et  $\mathbf{Y}$  deux vecteurs aléatoires gaussiens dans leur ensemble qui sont non corrélés. Alors  $\mathbf{X}$  et  $\mathbf{Y}$  sont indépendants.

☞ **Démonstration :** Si on définit le vecteur  $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ , par définition de  $\mathbf{X}$  et  $\mathbf{Y}$  et de leur non corrélation,  $\mathbf{Z}$  est un vecteur aléatoire gaussien de matrice de covariance

$$\mathbf{\Gamma}_{\mathbf{z}} = \begin{pmatrix} \mathbf{\Gamma}_{\mathbf{x}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{\mathbf{Y}} \end{pmatrix}$$

On a alors pour tout  $\mathbf{w} = (\mathbf{u}, \mathbf{v})$ ,

$$\mathbf{w}^T \mathbf{\Gamma}_{\mathbf{z}} \mathbf{w} = \mathbf{u}^T \mathbf{\Gamma}_{\mathbf{x}} \mathbf{u} + \mathbf{v}^T \mathbf{\Gamma}_{\mathbf{Y}} \mathbf{v} \quad \text{et} \quad \mathbf{w}^T \mathbf{m}_{\mathbf{z}} = \mathbf{u}^T \mathbf{m}_{\mathbf{x}} + \mathbf{v}^T \mathbf{m}_{\mathbf{Y}}$$

et donc

$$\varphi_{\mathbf{Z}}(\mathbf{w}) = \varphi_{\mathbf{X}}(\mathbf{u})\varphi_{\mathbf{Y}}(\mathbf{v})$$

ce qui permet de conclure à l'indépendance de  $\mathbf{X}$  et  $\mathbf{Y}$ .

- **Remarque :** Il est indispensable d'avoir l'hypothèse de vecteurs gaussiens dans leur ensemble pour obtenir le résultat précédent. Si  $\mathbf{X}$  et  $\mathbf{Y}$  sont simplement gaussiens, il est possible de trouver des contre-exemple qui mettent à mal le théorème (cf. Ref. [7]).

## 4.2 Espérance conditionnelle

La notion de conditionnement vue pour les probabilités au premier chapitre se retrouve aussi pour les variables et vecteurs aléatoires. Elle sera notamment utilisée dans le cours d'Automatique 2A via le concept de filtrage de Kalmann. Dans la suite, nous nous placerons dans le cas de variables aléatoires à densité car seule l'application aux VA gaussiennes sera vue ici, mais les résultats présentés dans cette partie peuvent se généraliser aux VA discrètes.

### 4.2.1 Définitions

#### Définition 44

**Densité conditionnelle :** Soient  $\mathbf{X}$  et  $\mathbf{Y}$  deux vecteurs aléatoires tels que  $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$  soit un vecteur aléatoire de  $\mathbb{R}^l \times \mathbb{R}^m$  admettant une densité de probabilité conjointe  $p_{(\mathbf{X}, \mathbf{Y})}$ . La fonction de  $\mathbf{x}$  et  $\mathbf{y}$  définie par

$$p_{(\mathbf{X}|\mathbf{Y})}(\mathbf{x}|\mathbf{y}) = \begin{cases} p_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y})/p_{\mathbf{Y}}(\mathbf{y}) & \text{si } p_{\mathbf{Y}}(\mathbf{y}) > 0 \\ 0 & \text{si } p_{\mathbf{Y}}(\mathbf{y}) = 0 \end{cases}$$

est appelée *densité conditionnelle de  $\mathbf{X}$  sachant que  $\mathbf{Y} = \mathbf{y}$* , avec  $p_{\mathbf{Y}}$  qui est la densité marginale de  $\mathbf{Y}$ . On peut noter que

$$\int_{\mathbb{R}^l} p_{(\mathbf{X}|\mathbf{Y})}(\mathbf{x}|\mathbf{y})d\mathbf{x} = \int_{\mathbb{R}^l} p_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y})/p_{\mathbf{Y}}(\mathbf{y})d\mathbf{x} = p_{\mathbf{Y}}(\mathbf{y})/p_{\mathbf{Y}}(\mathbf{y}) = 1 \quad (4.1)$$

- **Remarque :** La définition précédente suppose que le vecteur  $(\mathbf{X}, \mathbf{Y})$  admette une densité de probabilité conjointe. Cependant, il est aussi possible de définir une densité conditionnelle si l'un des vecteurs  $\mathbf{X}$  ou  $\mathbf{Y}$  est un vecteur aléatoire discret. Néanmoins, nous n'aborderons pas ce cas dans ce cours.

#### Définition 45

**Espérance conditionnelle :** Soit  $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$  un vecteur aléatoire de  $\mathbb{R}^l \times \mathbb{R}^m$  admettant une densité de probabilité conjointe  $p_{(\mathbf{X}, \mathbf{Y})}$ , et soit  $g$  une fonction mesurable de  $\mathbb{R}^l \times \mathbb{R}^m$  sur  $\mathbb{R}$  telle que

$$\iint_{\mathbb{R}^l \times \mathbb{R}^m} |g(\mathbf{x}, \mathbf{y})|p_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y})d\mathbf{x}d\mathbf{y} < +\infty.$$

Alors la grandeur définie par

$$\mathbb{E}[g(\mathbf{X}, \mathbf{Y})|\mathbf{Y}] = \int_{\mathbb{R}^l} g(\mathbf{x}, \mathbf{y})p_{(\mathbf{X}|\mathbf{Y})}(\mathbf{x}|\mathbf{y})d\mathbf{x}$$

est une variable aléatoire appelée *espérance conditionnelle de  $g(\mathbf{X}, \mathbf{Y})$  sachant  $\mathbf{Y}$* .

On retiendra notamment l'expression de l'espérance conditionnelle d'un vecteur aléatoire

$\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  par rapport à un autre vecteur  $\mathbf{Y}$  :

$$\mathbb{E}[\mathbf{X}|\mathbf{Y}] = \begin{pmatrix} \mathbb{E}[X_1|\mathbf{Y}] \\ \mathbb{E}[X_2|\mathbf{Y}] \\ \dots \\ \mathbb{E}[X_n|\mathbf{Y}] \end{pmatrix}$$

■ **Remarque** : L'espérance conditionnelle possède les mêmes propriétés de linéarité et de monotonie que l'espérance d'une VA réelle. De plus, on admettra que

$$\mathbb{E}[\mathbf{Y}|\mathbf{Y}] = \mathbf{Y} \quad \text{et} \quad \mathbb{E}[\mathbb{E}[g(\mathbf{X}, \mathbf{Y})|\mathbf{Y}]] = \mathbb{E}[g(\mathbf{X}, \mathbf{Y})]. \quad (4.2)$$

## 4.2.2 Conditionnement de vecteurs aléatoires gaussiens

### Théorème 27

Soient  $\mathbf{X}$  et  $\mathbf{Y}$  deux vecteurs aléatoires gaussiens dans leur ensemble d'espérances  $\mathbf{m}_\mathbf{X}$  et  $\mathbf{m}_\mathbf{Y}$ , et de matrices de covariance  $\Gamma_\mathbf{X}$  et  $\Gamma_\mathbf{Y}$ . L'espérance conditionnelle  $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$  est alors donnée par

$$\mathbb{E}[\mathbf{X}|\mathbf{Y}] = \mathbf{m}_\mathbf{X} + \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}(\mathbf{Y} - \mathbf{m}_\mathbf{Y})$$

et la matrice de covariance de  $\tilde{\mathbf{X}} = \mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]$  est

$$\Gamma_{\tilde{\mathbf{X}}} = \Gamma_\mathbf{X} - \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}\Gamma_{\mathbf{Y}\mathbf{X}}$$

D'autre part les vecteurs aléatoires  $\tilde{\mathbf{X}}$  et  $\mathbf{Y}$  sont indépendants.

☞ **Démonstration** : Définissons le vecteur aléatoire  $\mathbf{U}$  par  $\mathbf{U} = \mathbf{X} - \mathbf{m}_\mathbf{X} - \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}(\mathbf{Y} - \mathbf{m}_\mathbf{Y})$ . Un simple calcul permet de trouver que  $\mathbb{E}[\mathbf{U}] = 0$ . De plus,  $\mathbf{X}$  et  $\mathbf{Y}$  étant deux vecteurs aléatoires gaussiens dans leur ensemble,  $\mathbf{U}$  qui en est une combinaison linéaire est aussi gaussien. Or

$$\begin{aligned} \Gamma_{\mathbf{U}\mathbf{Y}} &= \mathbb{E}[\mathbf{U}(\mathbf{Y} - \mathbf{m}_\mathbf{Y})^T] = \mathbb{E}[(\mathbf{X} - \mathbf{m}_\mathbf{X} - \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}(\mathbf{Y} - \mathbf{m}_\mathbf{Y}))(\mathbf{Y} - \mathbf{m}_\mathbf{Y})^T] \\ &= \mathbb{E}[(\mathbf{X} - \mathbf{m}_\mathbf{X})(\mathbf{Y} - \mathbf{m}_\mathbf{Y})^T] - \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}\mathbb{E}[(\mathbf{Y} - \mathbf{m}_\mathbf{Y})(\mathbf{Y} - \mathbf{m}_\mathbf{Y})^T] \\ &= \Gamma_{\mathbf{X}\mathbf{Y}} - \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}\Gamma_\mathbf{Y} \\ &= 0 \end{aligned}$$

$\mathbf{U}$  et  $\mathbf{Y}$  sont donc non corrélés donc indépendants. Or en utilisant les équations (4.2) et (4.1), on montre que l'indépendance de  $\mathbf{U}$  et  $\mathbf{Y}$  amène  $\mathbb{E}[\mathbf{U}|\mathbf{Y}] = \mathbb{E}[\mathbf{U}] = 0$ , et donc que

$$\begin{aligned} \mathbb{E}[\mathbf{U}|\mathbf{Y}] &= \mathbb{E}[\mathbf{X} - \mathbf{m}_\mathbf{X} - \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}(\mathbf{Y} - \mathbf{m}_\mathbf{Y})|\mathbf{Y}] \\ &= \mathbb{E}[\mathbf{X}|\mathbf{Y}] - \mathbf{m}_\mathbf{X} - \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}(\mathbb{E}[\mathbf{Y}|\mathbf{Y}] - \mathbf{m}_\mathbf{Y}) \\ &= \mathbb{E}[\mathbf{X}|\mathbf{Y}] - \mathbf{m}_\mathbf{X} - \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}(\mathbf{Y} - \mathbf{m}_\mathbf{Y}) = 0 \end{aligned}$$

Soit  $\mathbb{E}[\mathbf{X}|\mathbf{Y}] = \mathbf{m}_\mathbf{X} + \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}(\mathbf{Y} - \mathbf{m}_\mathbf{Y})$ , et donc que  $\mathbf{U} = \tilde{\mathbf{X}}$ . On a alors

$$\begin{aligned} \Gamma_{\tilde{\mathbf{X}}} &= \mathbb{E}[\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T] = \mathbb{E}[(\mathbf{X} - \mathbf{m}_\mathbf{X} - \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}(\mathbf{Y} - \mathbf{m}_\mathbf{Y}))(\mathbf{X} - \mathbf{m}_\mathbf{X} - \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}(\mathbf{Y} - \mathbf{m}_\mathbf{Y}))^T] \\ &= \mathbb{E}[(\mathbf{X} - \mathbf{m}_\mathbf{X})(\mathbf{X} - \mathbf{m}_\mathbf{X})^T] - \mathbb{E}[(\mathbf{X} - \mathbf{m}_\mathbf{X})(\mathbf{Y} - \mathbf{m}_\mathbf{Y})^T\Gamma_\mathbf{Y}^{-1}\Gamma_{\mathbf{Y}\mathbf{X}}] \\ &\quad - \mathbb{E}[\Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}(\mathbf{Y} - \mathbf{m}_\mathbf{Y})(\mathbf{X} - \mathbf{m}_\mathbf{X})^T] + \mathbb{E}[\Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}(\mathbf{Y} - \mathbf{m}_\mathbf{Y})(\mathbf{Y} - \mathbf{m}_\mathbf{Y})^T\Gamma_\mathbf{Y}^{-1}\Gamma_{\mathbf{Y}\mathbf{X}}] \\ &= \Gamma_\mathbf{X} - \Gamma_{\mathbf{X}\mathbf{Y}}\Gamma_\mathbf{Y}^{-1}\Gamma_{\mathbf{Y}\mathbf{X}} \end{aligned}$$

# Chapitre 5

## Théorèmes limites et comportements asymptotiques

Certains problèmes physiques amènent à considérer le comportement asymptotique de suites de variables aléatoires et donc à analyser leur convergence. Or, comme on va l'illustrer dans ce chapitre, dire que la suite de variables aléatoires  $(X_n)_{n \in \mathbb{N}}$  tend vers la variable aléatoire  $X$  quand  $n$  tend vers  $+\infty$  peut avoir plusieurs significations. Nous nous appuyerons ensuite sur ces définitions pour établir plusieurs résultats fondamentaux, qui ont des applications très intéressantes en physique.

### 5.1 Modes de convergence

Nous présentons dans cette partie trois modes de convergence. Si tous les modes interviennent dans la suite de ce chapitre, seules les convergences en loi et en probabilité sont à bien connaître.

#### Définition 46

**Convergence presque sûrement :** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires définies sur l'espace  $(\Omega, \Sigma, \mathbb{P})$  et soit  $X$  une variable aléatoire définie sur l'espace  $(\Omega, \Sigma, \mathbb{P})$ . On dit que  $(X_n)_{n \in \mathbb{N}}$  converge presque sûrement vers  $X$  si l'événement

$$\mathcal{A} = \{\omega \in \Omega \mid \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\}$$

est presque certain,  $\mathbb{P}(\mathcal{A}) = 1$ .

**Convergence en probabilité :** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires définies sur l'espace  $(\Omega, \Sigma, \mathbb{P})$  et soit  $X$  une variable aléatoire définies sur l'espace  $(\Omega, \Sigma, \mathbb{P})$ . On dit que  $(X_n)_{n \in \mathbb{N}}$  converge en probabilité vers  $X$  si, pour tout  $\epsilon > 0$ ,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| > \epsilon\}) = 0$$

**Convergence en loi :** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires de fonction de répartition  $(F_{X_n})_{n \in \mathbb{N}}$  et soit  $X$  une variable aléatoire de fonction de répartition  $F_X$ . On dit que  $(X_n)_{n \in \mathbb{N}}$  converge en loi (ou en distribution) vers  $X$  si, pour tout  $x \in \mathbb{R}$  où  $F_X$  est continue,

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$$

- **Remarque :** Pour des VA continues, il convient de noter que la convergence en loi n'implique pas la convergence des densités de probabilité  $(p_{X_n})_{n \in \mathbb{N}}$  vers  $p_X$ .

On admettra les résultats suivants :

**Théorème 28**

La convergence presque sûrement implique la convergence en probabilité. La convergence en probabilité implique la convergence en loi. Les réciproques sont fausses.

Lorsque l'on manipule des suites de variables aléatoires, il peut s'avérer plus pratique de considérer les fonctions caractéristiques. La continuité de la transformation de Fourier permet de déduire le théorème suivant que l'on admettra.

**Théorème 29**

**Continuité des fonctions caractéristiques (ou théorème de convergence de Lévy) :** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires, cette suite converge en loi vers la variable aléatoire  $X$  si et seulement si, pour tout réel  $\nu$ , la suite  $(\varphi_{X_n}(\nu))_{n \in \mathbb{N}}$  converge vers  $\varphi_X(\nu)$ .

## 5.2 Lois des grands nombres

**Théorème 30**

**Loi faible des grands nombres :** Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires, deux à deux indépendantes et de même loi, admettant une espérance  $m$ . Si on définit la suite  $(Y_n)_{n \in \mathbb{N}}$  par

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k$$

alors cette suite converge en probabilité vers la constante  $m$  :

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P}(\{|Y_n - m| > \epsilon\}) = 0$$

Si on suppose que les variables aléatoires  $X_n$  admettent aussi une variance, la démonstration se fait assez simplement en utilisant l'inégalité de Bienaymé-Tchebychev. Dans le cas contraire, une méthode consiste à définir à partir des  $X_n$  des *variables aléatoires tronquées*, c'est à dire pour lesquelles il existe une variance finie si on ne considère qu'une partie de la suite  $(X_n)_{n \in \mathbb{N}}$ . Dans le cadre de ce cours, on ne se placera pas dans le deuxième cas.

- **Remarque :** Il est possible de trouver une autre version de la loi faibles des grands nombres qui n'impose pas d'avoir la même loi, mais qui nécessite que les  $X_n$  aient la même espérance et la même variance.

Il existe une autre version du théorème précédent, plus difficile à démontrer, qui stipule que la suite converge presque sûrement :

**Théorème 31**

**Loi forte des grands nombres :** Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires indépendantes et de même loi, admettant une espérance  $m$ . Si on définit la suite  $(Y_n)_{n \in \mathbb{N}}$  par

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k$$

alors cette suite converge presque sûrement vers la constante  $m$  :  $\mathbb{P}\left(\left\{\lim_{n \rightarrow +\infty} Y_n = m\right\}\right) = 1$ .

### 5.3 Théorème central limite

Les théorèmes précédents décrivent vers quelle valeur la suite des moyennes d'une variable aléatoire converge. Néanmoins, il est intéressant pour certains problèmes physiques de savoir comment cette suite converge. Dit autrement, il est intéressant de connaître la loi des variables aléatoires que définissent ces moyennes. Cela est donné par un résultat fondamental en théorie des probabilités, appelé *théorème central limite*.

#### Théorème 32

**Théorème central limite :** Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires indépendantes et de même loi, admettant une espérance  $m$  et une variance  $\sigma^2$ . Si on définit les variables

$$S_n = \sum_{k=1}^n X_k \quad \text{et} \quad Y_n = \frac{S_n - n m}{\sigma \sqrt{n}}$$

alors la suite des variables centrées réduites  $(Y_n)_{n \in \mathbb{N}^*}$  converge en loi vers la loi normale centrée réduite  $\mathcal{N}(0, 1)$  :

$$\forall y \in \mathbb{R}, \quad \mathbb{P}(\{Y_n \leq y\}) \xrightarrow{n \rightarrow +\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-t^2/2} dt$$

☞ **Démonstration :** On pose  $U_k = (X_k - m)/\sigma$  de sorte que, par indépendance des  $X_k$ , les  $U_k$  sont aussi indépendantes et que

$$Y_n = \frac{1}{\sqrt{n}} \sum_{k=0}^n U_k$$

De plus, si on note  $\varphi_U$  la fonction caractéristique commune aux  $U_k$  (les  $X_k$  ayant la même loi) et si on note  $\varphi_n$  celle associée à  $Y_n$ , on a par indépendance des  $U_k$  et en utilisant la formule de Taylor appliquée à  $\varphi$  en 0 :

$$\varphi_n(\nu) = \left[ \varphi_U \left( \frac{\nu}{\sqrt{n}} \right) \right]^n = \left[ 1 - \frac{\nu^2}{2n} + o \left( \frac{\nu^2}{n} \right) \right]^n \xrightarrow{n \rightarrow +\infty} e^{-\nu^2/2}$$

or la fonction  $e^{-\nu^2/2}$  est la fonction caractéristique d'une loi normale centrée réduite, ce qui montre le résultat.

■ **Remarque :** On trouve aussi les appellations : théorème limite central, ou théorème de la limite centrée ou centrale. Cette variété de dénominations viendrait du nom anglo-saxon *central limit theorem* et de l'ambiguïté qui en découle : l'adjectif « central » pouvant s'appliquer à la limite ou au théorème. Il s'avère ici qu'il s'applique au théorème, le terme « central » étant synonyme de « fondamental ». L'appellation correcte devrait être alors « théorème central de la limite ». Néanmoins, nous utiliserons indifféremment, dans ce cours, les appellations que l'on rencontre régulièrement dans la littérature : *théorème central limite* ou *théorème limite central*. L'expression *théorème de la limite centrale* pourra être aussi parfois employée, même si elle ne se justifie pas vraiment.

## 5.4 Comportements asymptotiques

### 5.4.1 Loi de Poisson comme limite de loi binomiale

#### Théorème 33

Soit  $X_n \sim \mathcal{B}(n, \lambda/n)$  une variable aléatoire binomiale de paramètres  $n$  et  $\lambda/n$ , où  $\lambda \in \mathbb{R}$ . Alors, pour tout  $k \in \mathbb{N}$ ,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\{X_n = k\}) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Dit autrement, une suite de variables aléatoires binomiales  $\mathcal{B}(n, \lambda/n)$  converge en loi vers une loi de Poisson  $\mathcal{P}(\lambda)$  de paramètre  $\lambda$ .

☞ **Démonstration** : Ce résultat se démontre en passant par les fonctions caractéristiques (utilisez l'Annexe B) et en utilisant le Théorème 29.

■ **Remarque** : Ce théorème est utile en physique nucléaire (pour calculer la probabilité d'émission de particules) ou en physique statistique (pour calculer la probabilité d'avoir un certain nombre de molécules de gaz contenues dans un volume donné).

### 5.4.2 Approximation normale d'une loi binomiale

#### Théorème 34

**Théorème de Moivre-Laplace** : Soient  $p \in ]0, 1[$  et  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires avec  $X_n \sim \mathcal{B}(n, p)$ , pour tout  $n \in \mathbb{N}^*$ . On définit les variables centrées réduites  $Y_n$  par

$$Y_n = \frac{X_n - np}{\sqrt{npq}}.$$

Alors la suite  $(Y_n)_{n \in \mathbb{N}^*}$  converge en loi vers la variable aléatoire  $Y \sim \mathcal{N}(0, 1)$ .

☞ **Démonstration** : Il suffit de montrer qu'une variable binomiale peut s'écrire comme une somme de variables de Bernoulli indépendantes (en passant par les fonctions caractéristiques des lois usuelles et en utilisant le Théorème 29) puis d'appliquer le théorème central limite.

### 5.4.3 Approximation normale d'une loi de Poisson

#### Théorème 35

Soient  $\lambda > 0$  et  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires avec  $X_n \sim \mathcal{P}(n\lambda)$ , pour tout  $n \in \mathbb{N}^*$ . On définit les variables centrées réduites  $Y_n$  par

$$Y_n = \frac{X_n - n\lambda}{\sqrt{n\lambda}}.$$

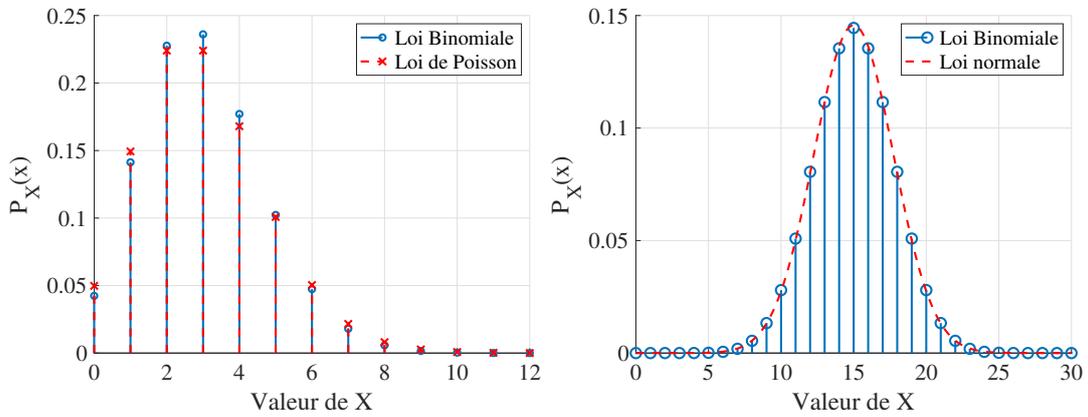
Alors la suite  $(Y_n)_{n \in \mathbb{N}^*}$  converge en loi vers la variable aléatoire  $Y \sim \mathcal{N}(0, 1)$ .

☞ **Démonstration** : La démonstration de ce théorème est similaire à celle du précédent. Elle s'appuie sur le théorème central limite et sur la stabilité de la loi de Poisson à la somme.

### 5.4.4 Résumé des approximations

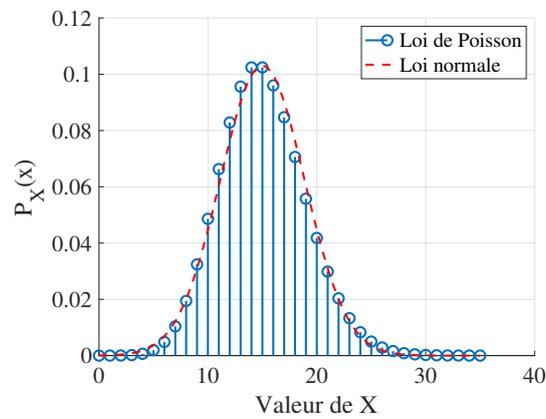
De manière empirique, il est possible de résumer les résultats des approximations par le tableau suivant [1]. Ils sont illustrés sur la Figure 5.1.

Loi	Loi approchée	Conditions pratiques
$\mathcal{B}(n, p)$	$\mathcal{P}(np)$	$p \leq 0,1 \quad n \geq 30$
$\mathcal{P}(n\mu)$	$\mathcal{N}(n\mu, \sqrt{n\mu})$	$n\mu \geq 15$
$\mathcal{B}(n, p)$	$\mathcal{N}(np, \sqrt{npq})$	$np \geq 15 \quad nq \geq 15$



(a) Loi de Poisson  $\mathcal{P}(np)$  comme limite d'une loi Binomiale  $\mathcal{B}(n, p)$  pour  $p = 0, 1$  et  $n = 30$ .

(b) Approximation d'une loi Binomiale  $\mathcal{B}(n, p)$  par une loi normale  $\mathcal{N}(np, \sqrt{np(1-p)})$  pour  $np = n(1-p) = 15$ .



(c) Approximation d'une loi de Poisson  $\mathcal{P}(\lambda)$  par une loi normale  $\mathcal{N}(\lambda, \sqrt{\lambda})$  pour  $\lambda = 15$ .

FIGURE 5.1 – Approximations de lois. Calculs réalisés sous Matlab. Les courbes sont tracées à partir de l'expression analytique des lois.

# Annexe A

## Quelques rappels mathématiques

La plupart de ces rappels sont issus de la référence [3].

### A.1 Rappels sur la théorie des ensembles

#### A.1.1 Ensembles

On appelle *ensemble* une collection d'objets appelés *éléments*. La notation  $x \in E$  se lit «  $x$  appartient à  $E$  » ou «  $x$  est un élément de l'ensemble  $E$  ».

Un ensemble  $F$  est *inclus* dans un ensemble  $E$  si tout élément de  $F$  est aussi élément de  $E$ , on écrit alors  $F \subset E$ . L'inclusion est transitive : si  $E$ ,  $F$  et  $G$  sont trois ensembles, alors  $G \subset F$  et  $F \subset E$  impliquent que  $G \subset E$ .

Il existe un ensemble, appelé *ensemble vide* et noté  $\emptyset$ , qui ne contient aucun élément. Il est inclus dans tout ensemble.

Soit  $E$  un ensemble, il existe un ensemble des *parties* de  $E$  et noté  $\mathfrak{P}(E)$  ou  $\mathcal{P}(E)$ , dont les éléments sont les ensembles inclus dans  $E$ . Il vérifie donc pour tout ensemble  $F : F \in \mathfrak{P}(E) \Leftrightarrow F \subset E$

■ **Exemple** : Si  $E = \{x, y, z\}$  est l'ensemble dont les éléments sont  $x$ ,  $y$  et  $z$ , alors

$$\mathfrak{P}(E) = \{\emptyset, \{x\}, \{y\}, \{z\}, \{x, y\}, \{x, z\}, \{y, z\}, E\}$$

Deux ensembles  $E$  et  $F$  sont dit *égaux*  $E = F$  s'ils ont les mêmes éléments, c'est à dire si les deux propriétés  $E \subset F$  et  $F \subset E$  sont vérifiées simultanément.

#### A.1.2 Quantifieurs

On rappelle la définition des deux symboles suivants :

- $\forall$  : « pour tout » ou « quel que soit l'élément », ainsi  $\forall x \in E$  se lit « pour tout élément  $x$  appartenant à  $E$  ... »,
- $\exists$  : « il existe », ainsi  $\exists x \in E$  se lit « il existe  $x$  appartenant à  $E$  ... »,

#### A.1.3 Assertions et prédicats

Une *assertion* est un énoncé auquel on peut donner une valeur de vérité (vraie ou fausse).

Un *prédicat* est un énoncé  $A(x, y, \dots)$  dépendant des variables  $x, y, \dots$  tel que l'on obtienne une assertion si on remplace ces variables par des éléments de certains ensembles.

■ **Exemple** : « 2 est un entier pair » est une assertion (vraie).

«  $x$  est un réel impair » ou «  $x \leq 2$  » sont des prédicats, on ne peut savoir si ces énoncés sont vrais ou faux avant de donner une valeur à  $x$ .

Si  $A$  est un prédicat à une variable et si  $E$  est un ensemble, on peut définir la partie de  $E$  constituée des éléments de  $E$  vérifiant  $A$ . On la note alors :

$$F = \{x \in E \mid A(x)\}$$

la barre verticale  $\mid$  signifiant « tels que ».

■ **Exemple** : L'ensemble des entiers naturels pairs peut être ainsi défini par :

$$\{n \in \mathbb{N} \mid \exists k \in \mathbb{N} : n = 2k\}$$

#### A.1.4 Opérations sur les ensembles

Soient  $F$  et  $G$  deux parties d'un ensemble  $E$ . On appelle

- *intersection* de  $F$  et  $G$ , la partie de  $E$  définie par

$$F \cap G = \{x \in E \mid x \in F \text{ ET } x \in G\}$$

- *union* de  $F$  et  $G$ , la partie de  $E$  définie par

$$F \cup G = \{x \in E \mid x \in F \text{ OU } x \in G\}$$

- *différence* de  $F$  et  $G$ , la partie de  $E$  définie par

$$F \setminus G = \{x \in F \mid x \notin G\}$$

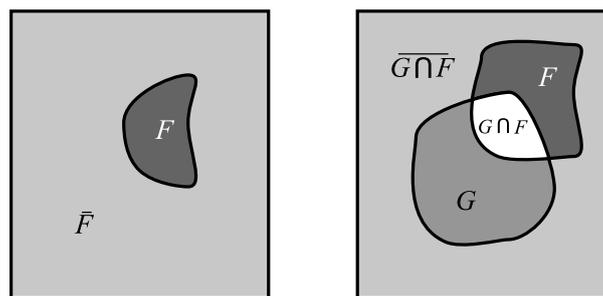
$F$  et  $G$  sont dits *mutuellement exclusifs* si  $F \cap G = \emptyset$ .

$F$  et  $G$  forment une *partition* de  $E$  si  $F \cap G = \emptyset$  et  $F \cup G = E$ .

On appelle *complémentaire* de  $F$  dans  $E$  la partie de  $E$  définie par

$$\bar{F} = F \setminus E = \{x \in E \mid x \notin F\}$$

On a alors les propriétés suivantes :  $F \cup \bar{F} = E$ ,  $F \cap \bar{F} = \emptyset$ ,  $\bar{\bar{F}} = F$ ,  $\bar{\emptyset} = E$ ,  $\bar{E} = \emptyset$ .



(a)  
Complémentaire

(b) Intersection

FIGURE A.1 – Illustrations de quelques opérations sur les ensembles

### A.1.5 Produit cartésien

A partir de deux éléments  $x$  et  $y$ , on peut construire le couple  $(x, y)$  avec la propriété fondamentale suivante :

$$(x, y) = (x', y') \Leftrightarrow (x = x' \text{ ET } y = y')$$

Soient deux ensembles  $E$  et  $F$ , l'ensemble des couples  $(x, y)$ , où  $x \in E$  et  $y \in F$ , est appelé produit cartésien de  $E$  par  $F$  et se note  $E \times F$  :

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

Lorsque  $E = F$ , le produit cartésien  $E \times F$  se note aussi  $E^2$ .

Ces considérations se généralisent aussi pour des  $n$ -uplets, ou  $n$ -listes,  $(x_1, x_2, \dots, x_n)$  avec  $n \in \mathbb{N}$ . On a alors le produit cartésien

$$E_1 \times E_2 \times \dots \times E_n = \{(x_1, x_2, \dots, x_n) \mid x_1 \in E_1 \text{ ET } x_2 \in E_2 \text{ ET } \dots \text{ ET } x_n \in E_n\}$$

En cas d'égalité des différents ensembles  $E_i = E$ ,  $\forall i \in \llbracket 1, n \rrbracket$ , le produit se note  $E^n$ .

Par convention, il n'existe qu'une unique 0-liste, appelée liste vide.

■ **Remarque** : Si  $a$  et  $b$  sont deux entiers relatifs, la notation  $\llbracket a, b \rrbracket$  désigne l'intervalle sur  $\mathbb{Z}$  incluant les entiers relatifs compris entre  $a$  et  $b$ .

### A.1.6 Applications

Soient  $E$  et  $F$  deux ensembles quelconques.

On appelle *graphe* de  $E$  vers  $F$  toute partie du produit cartésien  $E \times F$ .

Une application (ou fonction) est un triplet  $u = (E, F, \Gamma)$  où  $\Gamma$  est un graphe de  $E$  vers  $F$  tel que :

$$\forall x \in E, \exists y \in F, \text{ unique} : (x, y) \in \Gamma$$

On dit aussi que  $u$  est une application de  $E$  dans  $F$  ou de  $E$  vers  $F$ .

■ **Remarque** : Avec les notations précédentes,

- $E$  est appelé ensemble de départ ou ensemble de définition de  $u$ ,
- $F$  est appelé ensemble d'arrivée de  $u$ ,
- pour  $x \in E$ , l'unique  $y \in F$  tel que  $(x, y) \in \Gamma$  s'appelle image de  $x$  par  $u$  et se note  $u(x)$ ,
- pour  $y \in F$ , tout  $x \in E$  tel que  $y = u(x)$  est appelé antécédent de  $y$ ,
- $\Gamma = \{(x, u(x)) \mid x \in E\}$
- l'ensemble  $\{y \in F \mid \exists x \in E : y = u(x)\}$  noté aussi  $\{u(x) \mid x \in E\}$  est appelé ensemble image de  $u$ ; c'est une partie de  $F$ ,
- l'application  $u$  se note aussi

$$E \xrightarrow{u} F \quad \text{ou} \quad u : E \longrightarrow F \quad \text{ou} \quad u : \begin{array}{l} E \longrightarrow F \\ x \longmapsto u(x). \end{array}$$

On dit qu'une application  $u$  de  $E$  vers  $F$  est une *injection* ou est *injective* si tout élément de  $F$  a au plus un antécédent par  $u$ .

On dit qu'une application  $u$  de  $E$  vers  $F$  est une *surjection* ou est *surjective* si tout élément de  $F$  a au moins un antécédent par  $u$ .

On dit qu'une application  $u$  de  $E$  vers  $F$  est une *bijection* ou est *bijective* s'il est injective et surjective, c'est-à-dire si tout élément de  $F$  a *un et un seul* antécédent par  $u$ . Une application bijective de  $E$  sur  $E$  est aussi appelée une *permutation* de  $E$ .

### A.1.7 Cardinal

Un ensemble  $E$  est appelé ensemble fini s'il existe un entier naturel  $n$  tel qu'il existe une bijection entre  $E$  et  $\llbracket 1, n \rrbracket$ . Il y a alors unicité d'un tel entier  $n$ , que l'on appelle cardinal de  $E$  et que l'on notera  $\text{card}(E)$  ou  $\text{card}E$ . Un ensemble qui n'est pas fini est appelé ensemble infini.

Si  $F$  est une partie d'un ensemble fini  $E$ , alors  $F$  est un ensemble fini et  $\text{card}(F) \leq \text{card}(E)$ .

■ **Remarque** : De manière plus naïve, le cardinal est souvent vu comme « le nombre d'éléments que contient l'ensemble  $E$  ». Néanmoins, cela amène parfois à des raisonnements non rigoureux voire faux. Par exemple, l'écriture  $E = \{x_1, x_2, \dots, x_n\}$  n'implique pas que  $\text{card}(E) = n$  selon la définition précédente. Il faut en effet que les  $x_k$  soient distincts deux à deux, pour que l'application  $i \mapsto x_i$  soit une bijection de  $\llbracket 1, n \rrbracket$  dans  $E$ .

■ **Exemple** : L'ensemble vide est un ensemble fini de cardinal 0.

Si  $n$  et  $p$  sont des entiers naturels, l'application  $k \mapsto k + p$  est une bijection de  $\llbracket 1, n \rrbracket$  dans  $\llbracket p + 1, p + n \rrbracket$ . L'intervalle  $\llbracket p + 1, p + n \rrbracket$  est donc un ensemble fini de cardinal égal à  $n$ .

Si  $p$  et  $q$  sont des entiers relatifs tels que  $p \leq q$ , l'ensemble  $\llbracket p, q \rrbracket$  est fini de cardinal  $q - p + 1$ .

■ **Remarque** : Il convient ici de différencier les notions d'*ensembles finis* et d'*ensembles dénombrables*. Un ensemble est dit *dénombrable* quand il existe une bijection entre cet ensemble et  $\mathbb{N}$ . Une autre définition stipule qu'un ensemble est dénombrable s'il est en bijection avec une partie de  $\mathbb{N}$ . Cette définition inclut la précédente et la notion d'ensembles finis défini précédemment. Dans ce recueil de notes de cours, on utilisera la seconde définition, de sorte que l'on distinguera des ensembles dénombrables finis et des ensembles dénombrables infinis. On notera enfin, contrairement à un ensemble infini, un ensemble fini sera toujours dénombrable.

Soient  $E$  et  $F$  deux ensembles finis. On peut montrer que :

- $\text{card}(E \cup F) = \text{card}(E) + \text{card}(F) - \text{card}(E \cap F)$
- $\text{card}(E \times F) = \text{card}(E) \cdot \text{card}(F)$

## A.2 Dénombrement

Le dénombrement, aussi appelé analyse combinatoire, s'emploie à étudier les regroupements de parties ou de  $n$ -listes d'ensembles finis. On doit son développement à l'étude des jeux de hasard et des probabilités.

On considère dans la suite un ensemble fini  $E$  de cardinal égal à  $n \in \mathbb{N}$ . On rappelle aussi qu'une  $p$ -liste (ou  $p$ -uplet) d'éléments de  $E$  est une partie dénombrable de  $E^p$ . Une  $p$ -liste n'est pas nécessairement un ensemble fini puisque les  $p$  éléments ne sont pas nécessairement distincts.

### A.2.1 Permutations

Tout classement ordonné des  $n$  éléments distincts de  $E$  est une permutation de  $E$ . Autrement dit, une permutation de  $E$  est une bijection de  $E$  sur  $E$ .

Il existe  $n!$  ( $n$  factorielle) permutations de  $E$ .

■ **Remarque** : Il est important ici de se souvenir que la notion d'ensemble fini suppose que les éléments de  $E$  sont tous distincts.

■ **Exemple** : Les  $4! = 24$  permutations de  $\{\clubsuit, \diamond, \heartsuit, \spadesuit\}$  sont :

$$\begin{array}{cccccccc} \{\clubsuit, \diamond, \heartsuit, \spadesuit\} & \{\clubsuit, \diamond, \spadesuit, \heartsuit\} & \{\clubsuit, \heartsuit, \spadesuit, \diamond\} & \{\clubsuit, \heartsuit, \diamond, \spadesuit\} & \{\clubsuit, \spadesuit, \heartsuit, \diamond\} & \{\clubsuit, \spadesuit, \diamond, \heartsuit\} \\ \{\diamond, \clubsuit, \heartsuit, \spadesuit\} & \{\diamond, \clubsuit, \spadesuit, \heartsuit\} & \{\diamond, \heartsuit, \spadesuit, \clubsuit\} & \{\diamond, \heartsuit, \clubsuit, \spadesuit\} & \{\diamond, \spadesuit, \heartsuit, \clubsuit\} & \{\diamond, \spadesuit, \clubsuit, \heartsuit\} \\ \{\heartsuit, \clubsuit, \diamond, \spadesuit\} & \{\heartsuit, \clubsuit, \spadesuit, \diamond\} & \{\heartsuit, \diamond, \spadesuit, \clubsuit\} & \{\heartsuit, \diamond, \clubsuit, \spadesuit\} & \{\heartsuit, \spadesuit, \clubsuit, \diamond\} & \{\heartsuit, \spadesuit, \diamond, \clubsuit\} \\ \{\spadesuit, \clubsuit, \diamond, \heartsuit\} & \{\spadesuit, \clubsuit, \heartsuit, \diamond\} & \{\spadesuit, \diamond, \clubsuit, \heartsuit\} & \{\spadesuit, \diamond, \heartsuit, \clubsuit\} & \{\spadesuit, \heartsuit, \clubsuit, \diamond\} & \{\spadesuit, \heartsuit, \diamond, \clubsuit\} \end{array}$$

Soit  $L$  une  $n$ -liste d'éléments de  $E$  n'ayant que  $k \leq n$  éléments distincts, de sorte que ces  $k$  éléments apparaissent chacun un nombre  $n_1, n_2, \dots, n_k$  de fois, avec  $n_1 + n_2 + \dots + n_k = n$ .

On appelle par extension permutation avec répétitions tout classement des éléments de  $L$ . Le nombre de ces permutations est donné par

$$\frac{n!}{n_1!n_2!\dots n_k!}$$

■ **Exemple** : Considérons la 4-liste suivante  $\{\clubsuit, \diamond, \diamond, \spadesuit\}$ .

Les 12 permutations que l'on peut constituer sont :

$$\begin{array}{cccccccc} \{\clubsuit, \diamond, \diamond, \spadesuit\} & \{\clubsuit, \diamond, \spadesuit, \diamond\} & \{\clubsuit, \spadesuit, \diamond, \diamond\} & \{\spadesuit, \clubsuit, \diamond, \diamond\} & \{\spadesuit, \diamond, \clubsuit, \diamond\} & \{\spadesuit, \diamond, \diamond, \clubsuit\} \\ \{\diamond, \clubsuit, \diamond, \spadesuit\} & \{\diamond, \clubsuit, \spadesuit, \diamond\} & \{\diamond, \spadesuit, \clubsuit, \diamond\} & \{\diamond, \spadesuit, \diamond, \clubsuit\} & \{\diamond, \spadesuit, \diamond, \clubsuit\} & \{\diamond, \spadesuit, \clubsuit, \diamond\} \\ \{\diamond, \spadesuit, \diamond, \clubsuit\} & \{\diamond, \spadesuit, \clubsuit, \diamond\} & \{\diamond, \clubsuit, \spadesuit, \diamond\} & \{\diamond, \clubsuit, \diamond, \spadesuit\} & \{\diamond, \clubsuit, \diamond, \spadesuit\} & \{\diamond, \clubsuit, \spadesuit, \diamond\} \end{array}$$

## A.2.2 Arrangements

Soit  $L$  une  $p$ -liste d'éléments de  $E$ .

Si les éléments de  $L$  sont deux à deux distincts,  $L$  est appelé un *arrangement d'ordre  $p$*  (ou arrangement simple) de  $E$ . Il existe  $A_n^p = n!/(n-p)!$  arrangements de  $E$ .

Si les éléments de  $L$  ne sont pas distincts, on parle d'arrangement avec répétitions. Il existe alors  $n^p$  arrangements.

■ **Exemple** : • Il y a 12 arrangements simples de 2 éléments choisis parmi  $\{\clubsuit, \diamond, \heartsuit, \spadesuit\}$  :

$$\{\clubsuit, \diamond\} \{\clubsuit, \heartsuit\} \{\clubsuit, \spadesuit\} \{\diamond, \clubsuit\} \{\diamond, \heartsuit\} \{\diamond, \spadesuit\} \{\heartsuit, \clubsuit\} \{\heartsuit, \diamond\} \{\heartsuit, \spadesuit\} \{\spadesuit, \clubsuit\} \{\spadesuit, \diamond\} \{\spadesuit, \heartsuit\}$$

• Il y a 16 arrangements avec répétitions de 2 éléments choisis parmi  $\{\clubsuit, \diamond, \heartsuit, \spadesuit\}$  :

$$\begin{array}{cccccccc} \{\clubsuit, \clubsuit\} & \{\clubsuit, \diamond\} & \{\clubsuit, \heartsuit\} & \{\clubsuit, \spadesuit\} & \{\diamond, \clubsuit\} & \{\diamond, \diamond\} & \{\diamond, \heartsuit\} & \{\diamond, \spadesuit\} \\ \{\heartsuit, \clubsuit\} & \{\heartsuit, \diamond\} & \{\heartsuit, \heartsuit\} & \{\heartsuit, \spadesuit\} & \{\spadesuit, \clubsuit\} & \{\spadesuit, \diamond\} & \{\spadesuit, \heartsuit\} & \{\spadesuit, \spadesuit\} \end{array}$$

## A.2.3 Combinaisons

Soit  $F$  une partie de  $p$  d'éléments de  $E$ .  $F$  est appelée *combinaison d'ordre  $p$*  de  $E$ .

Le nombre de combinaisons d'ordre  $p$  de  $E$  est donné par :

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

■ **Remarque** : Une combinaison consiste à choisir  $p$  éléments différents parmi  $n$  objets distincts sans prendre en compte l'ordre d'apparition.

■ **Exemple** : Il y a 6 combinaisons de 2 éléments choisis parmi  $\{\clubsuit, \diamond, \heartsuit, \spadesuit\}$  :

$$\{\clubsuit, \diamond\} \{\clubsuit, \heartsuit\} \{\clubsuit, \spadesuit\} \{\diamond, \heartsuit\} \{\diamond, \spadesuit\} \{\heartsuit, \spadesuit\}$$

Soient  $n$  et  $p$  deux entiers naturels, on a alors les relations suivantes :

- $C_n^p = C_n^{n-p}$  si  $p \leq n$
- $C_n^p = C_{n-1}^p + C_{n-1}^{p-1}$  si  $p \geq 1$  et  $n \geq 1$  (Relation de Pascal)

- $C_n^p = \frac{n}{p} C_{n-1}^{p-1}$  si  $n \geq p \geq 1$
- $C_n^p = \frac{n-p+1}{p} C_n^{p-1}$  si  $n \geq p \geq 1$

Tout comme pour les arrangements ou les permutations, on définit parfois les « *combinaisons avec répétitions* ». Le nombre possible de ces combinaisons de  $p$  éléments parmi  $n$  objets est donné par

$$\bar{C}_n^p = C_{n+p-1}^p = \binom{n+p-1}{p} = \frac{(n+p-1)!}{p!(n-1)!}$$

■ **Exemple** : Il y a 10 combinaisons avec répétitions de 2 éléments choisis parmi  $\{\clubsuit, \diamond, \heartsuit, \spadesuit\}$  :

$\{\clubsuit, \clubsuit\}$   $\{\clubsuit, \diamond\}$   $\{\clubsuit, \heartsuit\}$   $\{\clubsuit, \spadesuit\}$   $\{\diamond, \diamond\}$   $\{\diamond, \heartsuit\}$   $\{\diamond, \spadesuit\}$   $\{\heartsuit, \heartsuit\}$   $\{\heartsuit, \spadesuit\}$   $\{\spadesuit, \spadesuit\}$

# Annexe B

## Lois usuelles

### Lois discrètes

Nom	Loi	Espérance	Variance
Loi uniforme sur $\llbracket 1, n \rrbracket$ , $n \in \mathbb{N}$	$\mathbb{P}(\{X = k\}) = \frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Loi de Bernouilli $\mathcal{B}(p)$ , $p \in [0, 1]$	$\begin{cases} \mathbb{P}(\{X = 1\}) = p \\ \mathbb{P}(\{X = 0\}) = 1 - p \end{cases}$	$p$	$p(1 - p)$
Loi binomiale $\mathcal{B}(n, p)$ , $n \in \mathbb{N}$ , $p \in [0, 1]$	$\mathbb{P}(\{X = k\}) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & \forall k \in \llbracket 0, n \rrbracket \\ 0 & \text{sinon} \end{cases}$	$np$	$np(1 - p)$
Loi géométrique $\mathcal{G}(p)$ , $p \in [0, 1]$	$\mathbb{P}(\{X = k\}) = p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Loi de Poisson $\mathcal{P}(\lambda)$ , $\lambda \in \mathbb{R}^*$	$\begin{cases} \mathbb{P}(\{X = k\}) = e^{-\lambda} \frac{\lambda^k}{k!}, & \forall k \in \llbracket 0, n \rrbracket \\ 0 & \text{sinon} \end{cases}$	$\lambda$	$\lambda$

### Lois continues

Nom	Densité	Espérance	Variance	Fonc. carac.
Loi uniforme sur $\llbracket a, b \rrbracket$ , $a, b \in \mathbb{R}$	$p_X(x) = \frac{1}{b-a}$ , sur $[a, b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{1}{b-a} \frac{e^{i\nu b} - e^{i\nu a}}{i\nu}$
Loi gaussienne $\mathcal{N}(m, \sigma)$ , $m \in \mathbb{R}, \sigma > 0$	$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	$m$	$\sigma^2$	$e^{im\nu} e^{-\sigma^2\nu^2/2}$
Loi exponentielle $\mathcal{E}(\alpha)$ , $\alpha > 0$	$p_X(x) = H(x)\alpha e^{-\alpha x}$	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$	$(1 - i\frac{\nu}{\alpha})^{-1}$
Loi de Cauchy de paramètre $a > 0$	$p_X(x) = \frac{a}{\pi} \frac{1}{a^2 + (x-m)^2}$	$\infty$	$\infty$	$e^{im\nu -  \alpha\nu }$
Loi du $\chi^2$ de paramètre $n \in \mathbb{N}^*$	$p_X(x) = \frac{H(x)}{2\Gamma(n/2)} \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-x/2}$	$n$	$2n$	$(1 - 2i\nu)^{-n/2}$
Loi Gamma $\Gamma(a, \lambda)$ , $a > 0, \lambda > 0$	$p_X(x) = H(x) \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$	$(1 - i\frac{\nu}{\lambda})^{-a}$

où  $\forall x \in \mathbb{R}, \Gamma(x)$  est définie par  $\Gamma(x) = \int_0^{+\infty} u^{x-1} e^{-u} du$  et où  $H(x)$  est donnée par la fonction d'Heaviside.

## Annexe C

# La transformée de Fourier dans le cours de variables aléatoires

Les cours de signaux certains et de variables aléatoires utilisent deux conventions différentes (mais équivalentes) pour la transformée de Fourier. En effet, la définition 25 définit la transformée de Fourier par

$$\tilde{f}(\nu) = \int_{-\infty}^{\infty} e^{i\nu x} f(x) dx$$

Passer de cette convention à celle de signaux certains peut se faire par un changement de variable, mais cela nécessite de faire attention à la définition de la transformée de Fourier inverse qui dans le cas des conventions du cours de variables aléatoires s'écrit alors :

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\nu x} \tilde{f}(\nu) d\nu$$

où un terme de normalisation en  $1/2\pi$  apparaît.

### C.1 Résultats classiques

Si on respecte ces conventions, il est possible de retrouver les mêmes résultats qu'en signaux certains.

#### C.1.1 Cas de la dilatation

Soit  $a$  un réel, on a alors

$$\mathbf{TF}[f(ax)](\nu) = \int_{-\infty}^{+\infty} e^{i\nu x} f(ax) dx = \int_{-\infty}^{+\infty} e^{i\nu u/a} f(u) \frac{du}{|a|} = \frac{1}{|a|} \tilde{f}\left(\frac{\nu}{a}\right)$$

De même,

$$\mathbf{TF}^{-1}\left[\tilde{f}\left(\frac{\nu}{a}\right)\right](x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\nu x} \tilde{f}\left(\frac{\nu}{a}\right) d\nu = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iakx} \tilde{f}(k) |a| dk = |a| f(ax)$$

#### C.1.2 Cas de la translation

Soit  $x_0$  un réel, on a alors

$$\mathbf{TF}[f(x - x_0)](\nu) = \int_{-\infty}^{+\infty} e^{i\nu x} f(x - x_0) dx = \int_{-\infty}^{+\infty} e^{i\nu(u+x_0)} f(u) du = e^{i\nu x_0} \tilde{f}(\nu)$$

De même, pour  $\nu_0$  réel,

$$\mathbf{TF}^{-1}\left[\tilde{f}(\nu - \nu_0)\right](x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\nu x} \tilde{f}(\nu - \nu_0) d\nu = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i(k+\nu_0)x} \tilde{f}(k) dk = e^{-i\nu_0 x} f(x)$$

On notera notamment ici que le terme de modulation qui découle de la translation s'exprime de manière cohérente avec la définition de la TF.

## C.2 TF de fonctions classiques

L'impact du changement de convention se fait principalement sentir dans les expressions des TF et TF inverse des fonctions classiques. Par exemple, le résultat classique

$$\mathbf{TF} \left[ e^{-\pi x^2} \right] = e^{-\pi \nu^2}$$

n'est plus valable.

### C.2.1 Cas de la gaussienne $f(x) = e^{-\pi x^2}$

Cherchons comment s'exprime la TF de  $e^{-\pi x^2}$  avec les nouvelles conventions, sachant que l'on a toujours le résultat classique  $\int e^{-x^2} dx = \sqrt{\pi}$  qui s'écrit aussi  $\int e^{-\pi x^2} dx = 1$

On pose

$$\mathbf{TF} [f(x)] = \int_{-\infty}^{+\infty} e^{i\nu x - \pi x^2} dx = F(\nu)$$

Par définition de  $f$ , on a  $\forall x \in \mathbb{R}$ ,  $f'(x) = -2\pi x f(x)$  et donc que  $\mathbf{TF} [f'(x)] = -2\pi \mathbf{TF} [x f(x)]$ .

Or, selon le cours de *Signaux Certains*, on trouve facilement que

$$\mathbf{TF} [x f(x)] = \frac{1}{i} F'(\nu) \quad \text{et} \quad \mathbf{TF} [f'(x)] = -i\nu F(\nu)$$

On a donc que  $\frac{-2\pi}{i} F'(\nu) = -i\nu F(\nu)$  ce qui amène à la résolution de l'équation différentielle

$$F'(\nu) + \frac{\nu}{2\pi} F(\nu) = 0$$

On trouve ainsi que  $F(\nu) = F(0) \exp[-\nu^2/(4\pi)]$  avec par définition

$$F(0) = \int_{-\infty}^{+\infty} e^{-\pi x^2} dx = 1$$

et donc que

$$\mathbf{TF} \left[ e^{-\pi x^2} \right] = e^{-\nu^2/(4\pi)}$$

### C.2.2 Cas de la gaussienne normalisée

Si on cherche maintenant à calculer la TF de la gaussienne

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

il suffit d'appliquer les règles des TF des opérations classiques, ici la dilatation avec  $a = 1/\sqrt{2\pi}\sigma$  puis une translation de  $m$ , pour retrouver un résultat classique :

$$\tilde{f}(\nu) = e^{-\nu^2\sigma^2/2} e^{i\nu m}$$

# Annexe D

## Exercices d'entraînement

### D.1 Exercices sur les Probabilités

#### D.1.1 Probabilités uniformes

On souhaite répondre aux questions suivantes :

**Loto** Quel est la probabilité de gagner au loto lors d'un tirage ? Et à Euromillions ?

**Dés** Aux lancer de deux dés, quelle est la probabilité de faire un double six ? Que la somme des valeurs indiquées par les dés soit égale à six ?

**Dates d'anniversaire** Quelle est la probabilité que deux personnes (prises au hasard évidemment) n'aient pas la même date d'anniversaire ? Que parmi trois personnes il n'y ait pas de coïncidence de dates ? À partir de combien de personnes dans une salle, la probabilité que au moins deux personnes aient la même date d'anniversaire devient supérieure à 0,5 ?, à 0,9 ?

**Dates d'anniversaire bis** À partir de quel nombre d'élèves dans une classe, la probabilité que l'un d'entre eux ait la même date d'anniversaire que l'enseignant(e) devient supérieure à 0,5 ?

#### D.1.2 Subtil !

- Mes voisins ont deux enfants dont une fille, quelle est la probabilité que l'autre soit un garçon ?
- Mes voisins ont deux enfants dont le plus jeune est une fille, quelle est la probabilité que l'autre soit un garçon ?

#### D.1.3 Le danger des faux positifs <sup>1</sup>

Supposons, quand on est une vache, que la probabilité d'être « folle » est de  $1/10000$  et qu'on dispose d'un test bon marché qui produit  $1/1000$  faux positifs et autant de faux négatifs. Ce qui veut dire que si vous êtes une vache folle (ou saine), la probabilité que vous soyez mal cataloguée est de  $1/1000$ .

Quelle est alors la probabilité qu'avec ce test une vache soit déclarée folle ?

#### D.1.4 Chasse aux trésors

Trois coffrets contiennent respectivement deux pièces d'or, deux pièces d'argent et enfin une pièce d'or et une pièce d'argent. On choisit un coffret au hasard et on tire une pièce au hasard dans celui-ci. La pièce est en or, quelle est la probabilité que la deuxième le soit aussi ?

---

1. Exercice tiré du *Petit traité de hasardologie* de Hubert Krivine.

### D.1.5 Une question de stratégie

Un jeu télévisé proposait de gagner une voiture. Le candidat était placé face à trois portes. La voiture était cachée derrière l'une d'elles.

Le candidat choisissait une des trois portes. Le présentateur ouvrait alors l'une des deux portes restantes, derrière laquelle ne se trouvait pas la voiture. Le candidat pouvait alors :

- soit maintenir son choix,
- soit choisir la porte laissée fermée par le présentateur.

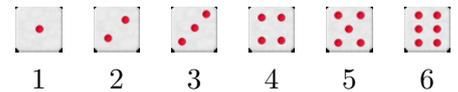
Qu'auriez-vous fait ?

## D.2 Exercices sur les Variables aléatoires

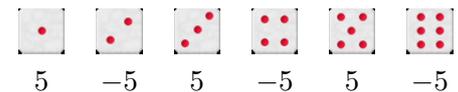
### VA discrète

On considère le lancer d'un dé à six faces, les six faces étant équiprobables. On associe au tirage deux variables aléatoires :

**La variable aléatoire  $A$**  associe au lancer la valeur affichée par le dé



**La variable aléatoire  $B$**  associe au lancer la valeur  $\pm 5$  selon la parité



1. Quelle est la loi de probabilité du lancer de dé ?
2. Quelles sont les lois de probabilités des variables aléatoires  $A$  et  $B$  ?
3. Quelles sont leurs fonctions de répartition ?
4. Quelles sont leurs espérances mathématiques ?
5. Quelles sont leurs variances ?

### VA continue

On considère le résultat de la mesure de l'éclairement solaire au sol. Sans information supplémentaire, on suppose que la valeur de cet éclairement est compris entre  $500\text{W}/\text{m}^2$  (météo médiocre) et  $1,5\text{kW}/\text{m}^2$  (grand beau temps) de façon équiprobable (!).

On associe à cette mesure trois variables aléatoires :

**La variable aléatoire  $A$**  associe à la mesure lue sa valeur numérique  $\in \mathbb{R}$ .

**La variable aléatoire  $B$**  associe à cette mesure la valeur du courant fourni par une cellule photovoltaïque carrée de côté  $10\text{cm}$  dont la réponse est de  $0.3\text{A}/\text{W}$ .

**La variable aléatoire  $C$**  associe à cette mesure la valeur  $1$  si il fait beau, c'est à dire si l'éclairement est supérieur à  $1\text{kW}/\text{m}^2$  et la valeur  $0$  sinon.

Déterminer la fonction de répartition, la loi ou densité de probabilité, l'espérance mathématique et la variance de chacune de ces variables.

**Propriété d'une VA**

On considère la fonction  $f$  définie par :

$$\begin{aligned} f(x) &= \frac{1}{4} - \lambda x^2 \text{ pour } x \in [-1, 1] \\ f(x) &= 0 \text{ sinon} \end{aligned}$$

1. Déterminer  $\lambda$  pour que  $f$  soit une densité de probabilité.
2. Soit  $X$  une variable aléatoire de densité  $f$ . Calculer  $\mathbb{E}[X]$  et  $\text{Var}[X]$ .

# Annexe E

## Correction des exercices

### E.1 Exercices sur les Probabilités

#### E.1.1 Probabilités uniformes

**Loto** La probabilité de gagner au loto lors d'un tirage où comment choisir 5 nombres parmi 49 (sans ordre et sans remise) et 1 parmi 10 :

$$P = \frac{1}{C_{49}^5} \frac{1}{C_{10}^1} = \frac{1}{\sim 19 \cdot 10^6} = 5 \cdot 10^{-8}$$

**Euromillions** La probabilité de gagner à Euromillions lors d'un tirage où comment choisir 5 nombres parmi 50 (sans ordre et sans remise) et 2 parmi 12 :

$$P = \frac{1}{C_{50}^5} \frac{1}{C_{12}^2} = \frac{1}{\sim 1.4 \cdot 10^8} = 7.2 \cdot 10^{-9}$$

**Dés** Aux lancer de deux dés, la probabilité de faire un double six est :

$$P = \frac{1}{36}$$

que la somme des valeurs indiquées par les dés soit égale à six est :

$$P = \frac{5}{36}$$

car 5 cas sont favorables sur 36 :  $\{(1, 5) (2, 4) (3, 3) (4, 2) (5, 1)\}$

**Dates d'anniversaire** La probabilité que deux personnes (prises au hasard évidemment) n'aient pas la même date d'anniversaire :

$$P = \frac{364}{365}$$

(Les cas possibles peuvent être décrits par un pavage du plan par une grille 365x365. Les cas favorables sont décrits par tous les éléments du pavage moins la diagonale.)

Que parmi trois personnes il n'y ait pas de coïncidence de dates :

$$P = \frac{364}{365} \frac{363}{365}$$

(Tous les cas possibles peuvent être décrits par les éléments d'un 'rubik's cube' , chaque face contenant 365x365 éléments... )

**Attention aux raisonnements erronés** La probabilité que 3 personnes aient la même date d'anniversaire est plus difficile à évaluer :

$$\begin{aligned}
 P' &= P(1=2)P(3 \neq 1,2) \\
 &\quad \underbrace{\frac{1}{365} \cdot \frac{364}{365}} \\
 &+ P(1=3)P(2 \neq 1,3) \\
 &+ P(2=3)P(1 \neq 2,3) \\
 &+ \underbrace{P(1=2=3)} \\
 &\quad \underbrace{\frac{1}{365^2}} \\
 &= \frac{3}{365} \cdot \frac{364}{365} + \frac{1}{365^2} \\
 &= \frac{3}{365} \left(1 - \frac{1}{365}\right) + \frac{1}{365^2} \\
 &= \frac{3}{365} - \frac{2}{365^2} = \frac{3 \times 365 - 2}{365^2}
 \end{aligned}$$

or :

$$\begin{aligned}
 1 - P &= 1 - \frac{364}{365} \frac{363}{365} \\
 &= \frac{365^2 - 364 \times 363}{365^2} \\
 &= \frac{365^2 - (365 - 1) \times (365 - 2)}{365^2} \\
 &= \frac{3 \times 365 - 2}{365^2}
 \end{aligned}$$

Pour  $N$  personnes :

$$\begin{aligned}
 P &= \frac{(365 - 1)(365 - 2) \cdots (365 - N + 1)}{365^{N-1}} \\
 &= \frac{365!}{\underbrace{365^N (365 - N)!}}
 \end{aligned}$$

Difficile à évaluer sous cette forme!

On peut le réécrire :

$$\begin{aligned}
 P &= \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \\
 &\quad \cdots \left(1 - \frac{N-1}{365}\right) \\
 \ln(P) &= \sum_{k=1}^{N-1} \ln\left(1 - \frac{k}{365}\right)
 \end{aligned}$$

En approchant  $\ln(1 - x) \approx -x$  on obtient :

$$\ln(p) = -\frac{N(N-1)}{2 \times 365} \approx -\frac{N^2}{730}$$

Le nombre de personnes dans une salle à partir duquel la probabilité que au moins deux personnes aient la même date d'anniversaire devient supérieure à 0,5 :

$$N = \sqrt{730 \ln(2)} = 22,3$$

soit 23 personnes ; à 0,9 :

$$N = \sqrt{730 \ln(10)} = 40,9$$

soit 41 personnes.

**Dates d'anniversaire bis** Dans une classe de  $N$  élèves, la probabilité que l'un d'entre eux ait la même date d'anniversaire que l'enseignant(e) est donnée par :

$$P = 1 - \left(\frac{364}{365}\right)^N$$

donc  $P > 0.5 \rightarrow N > 253$

### E.1.2 Subtil!

Evènements équiprobables  $\{(FG)(GF)(FF)(GG)\}$   $P(FG) = P(GF) = P(FF) = P(GG)$ . C'est le point de départ le plus important!

Première façon de faire : on définit l'espace des possibles uniquement avec les évènements pertinents.

- Possibles =  $\left\{ \begin{matrix} (FG) & (GF) & (FF) \\ \text{Fav.} & \text{Fav.} & \text{X} \end{matrix} \right\}$  donc  $p = \frac{2}{3}$
- Possibles =  $\left\{ \begin{matrix} (GF) & (FF) \\ \text{Fav.} & \text{X} \end{matrix} \right\}$  donc  $p = \frac{1}{2}$

Seconde façon de faire : on définit l'espace des possibles en gardant tous les évènements mais en définissant aussi les évènements suivants

- $A =$  « avoir au moins 1 fille »
- $B =$  « avoir au moins 1 garçon »
- $C =$  « avoir 1 fille cadette »

L'évènement  $A \cap B$  équivaut alors à « avoir un garçon et une fille » et l'évènement  $C \cap B$  équivaut alors à « avoir un garçon aîné et une fille cadette »

La première question revient à calculer  $\mathbb{P}_A(B) = \mathbb{P}(A \cap B)/\mathbb{P}(A) = (1/2)/(3/4) = 2/3$ .

La seconde question revient à calculer  $\mathbb{P}_C(B) = \mathbb{P}(C \cap B)/\mathbb{P}(C) = (1/4)/(1/2) = 1/2$

### E.1.3 Le danger des faux positifs

On définit

- $F =$  « être classée folle par le test »
- $C =$  « être bien classée »

La probabilité d'être classée comme une vache folle est la somme de la probabilité d'être vraiment folle et d'être bien classée par le test, plus la probabilité d'être vraiment saine et d'être mal classée par le test :

$$\mathbb{P}(F) = \mathbb{P}(F \cap C) + \mathbb{P}(F \cap \bar{C}) = \mathbb{P}(F|C)\mathbb{P}(C) + \mathbb{P}(F|\bar{C})\mathbb{P}(\bar{C})$$

La probabilité d'être bien classé en tant que vache folle ou non est de  $\mathbb{P}(C) = 1 - \mathbb{P}(\bar{C}) = 1 - \frac{1}{1000}$ . De même la probabilité d'être vraiment folle est de  $\mathbb{P}(F|C) = \frac{1}{10000}$  et celle d'être bien vraiment saine est de  $\mathbb{P}(F|\bar{C}) = 1 - \frac{1}{10000}$

La probabilité d'être classée comme folle est donc

$$\mathbb{P}(F) = \frac{1}{10000} \left(1 - \frac{1}{1000}\right) + \left(1 - \frac{1}{10000}\right) \frac{1}{1000} = \frac{1}{1000} + \frac{1}{10000} - \frac{2}{10^7} \approx \frac{1}{1000}$$

On trouverait 10 fois plus de vaches folles que prévu.

**E.1.4 Chasse aux trésors**

On sait que :

- $\mathbb{P}(2 \text{ pièces en or dans le coffret}) = \mathbb{P}(\text{on a choisi le coffret n}^\circ 1) = \frac{1}{3}$
- $\mathbb{P}(2 \text{ pièces en argent dans le coffret}) = \mathbb{P}(\text{on a choisi le coffret n}^\circ 2) = \frac{1}{3}$
- $\mathbb{P}(1 \text{ pièce de chaque matériau dans le coffret}) = \mathbb{P}(\text{on a choisi le coffret n}^\circ 3) = \frac{1}{3}$

On définit les évènements :

- $1\text{Or} = \text{"on a tiré une pièce en or lors du premier tirage"}$
- $2\text{Or} = \text{"il y a 2 pièces en or dans le coffret"}$

et on cherche à déterminer :  $P_{2\text{Or}/1\text{Or}}$

La formule de Bayes donne

$$P_{2\text{Or}/1\text{Or}} = \frac{P_{2\text{Or} \cap 1\text{Or}}}{P_{1\text{Or}}}$$

or

$$\begin{aligned} P_{2\text{Or}} &= \mathbb{P}(\text{on a choisi le coffret n}^\circ 1) = \frac{1}{3} \\ P_{1\text{Or}} &= \mathbb{P}(\text{Tirer une pièce en or dans le coffret}) \\ &= \mathbb{P}(\text{Tirer une pièce en or} \mid \text{on a choisi le coffret 1}) \cdot \mathbb{P}(\text{Choisir le coffret 1}) \\ &+ \mathbb{P}(\text{Tirer une pièce en or} \mid \text{on a choisi le coffret 3}) \cdot \mathbb{P}(\text{Choisir le coffret 3}) \\ &= 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2} \end{aligned}$$

Finalement :

$$P_{2\text{Or}/1\text{Or}} = 2/3$$

**E.1.5 Une question de stratégie**

Il faut changer de porte car la probabilité de gagner passe de  $1/3$  à  $2/3$ .

**Explication rapide (extraite d'un article de Pour la Science)**

Supposons (sans perte de généralité) que l'on choisisse la porte 1, le présentateur va ouvrir l'une des deux autres portes qui sera forcément vide. Trois cas sont possibles :

- Cas 1 : porte 1 voiture ; porte 2 rien ; porte 3 rien.
- Cas 2 : porte 1 rien ; porte 2 voiture ; porte 3 rien.
- Cas 3 : porte 1 rien ; porte 2 rien ; porte 3 voiture.

Chacun de ces cas a une probabilité  $1/3$  de se produire. Dans le cas 1, quelle que soit la porte que le présentateur ouvre (2 ou 3) si le candidat change, il perd. Dans les cas 2 et 3, en revanche, le candidat, en changeant, gagne à chaque fois. Donc, au total, en changeant de porte, le candidat gagne 2 fois sur 3.

**Autre démonstration en passant par les probabilité conditionnelles**

Supposons une nouvelle fois que l'on choisisse la porte 1. Notons  $\mathcal{V}_i$  l'événement « la voiture est derrière la porte  $i$  » et  $\mathcal{O}_i$  l'événement « le présentateur ouvre la porte  $i$  ». On a de manière assez évidente  $\mathbb{P}(\mathcal{V}_i) = 1/3$  et donc 1 chance sur 3 de gagner. L'idée est maintenant de calculer la probabilité de gagner si on change de porte c'est à dire de regarder, par exemple, la probabilité de l'événement « la voiture se trouve derrière la porte 2 sachant que le présentateur a ouvert la porte 3, noté  $\mathbb{P}(\mathcal{V}_2|\mathcal{O}_3)$ . Pour cela on utilise la formule de Bayes :

$$\mathbb{P}(\mathcal{V}_2 | \mathcal{O}_3) = \frac{\mathbb{P}(\mathcal{V}_2)\mathbb{P}(\mathcal{O}_3 | \mathcal{V}_2)}{\sum_{i=1}^3 \mathbb{P}(\mathcal{V}_i)\mathbb{P}(\mathcal{O}_3 | \mathcal{V}_i)}$$

- si la voiture se trouve derrière la porte 1, le présentateur va ouvrir une autre porte au hasard, on a alors  $\mathbb{P}(\mathcal{O}_3|\mathcal{V}_1) = 1/2$ ,
- si la voiture se trouve derrière la porte 2, le présentateur va forcément ouvrir la porte 3, on a alors  $\mathbb{P}(\mathcal{O}_3|\mathcal{V}_2) = 1$ ,
- si la voiture se trouve derrière la porte 3, le présentateur va forcément ouvrir la porte 2, on a alors  $\mathbb{P}(\mathcal{O}_3|\mathcal{V}_3) = 0$ ,

On a donc

$$\mathbb{P}(\mathcal{V}_2 | \mathcal{O}_3) = \frac{1/3 \times 1}{1/3 \times 1/2 + 1/3 \times 1 + 1/3 \times 0} = \frac{2}{3}$$

**Illustration en images**

Voir la vidéo extraite du film *Las Vegas 21* (en VF) : <https://www.youtube.com/watch?v=huLoJcTppXk>

**E.2 Exercices sur les variables aléatoires****VA discrète**

1. La loi de probabilité du lancer de dé est la fonction

$$\mathbb{P}\left(\text{lancer donne } \begin{array}{|c|} \hline \bullet \\ \hline \end{array}\right) = \mathbb{P}\left(\text{lancer donne } \begin{array}{|c|} \hline \bullet \bullet \\ \hline \end{array}\right) = \dots = 1/6$$

2.  $\mathbb{P}_A(i) = 1/6$  pour  $i \in \{1, 2, 3, 4, 5, 6\}$  et  $\mathbb{P}_A(i) = 0$  ailleurs.  
 $\mathbb{P}_B(i) = 1/2$  pour  $i \in \{-5, 5\}$  et  $\mathbb{P}_B(i) = 0$  ailleurs.

3. 
$$F_A(x) = \begin{cases} i/6 \text{ avec } i \in \{1, 2, 3, 4, 5, 6\} & \text{pour } x \in [i, i + 1[ \\ 0 & \text{pour } x < 1 \\ 1 & \text{pour } x \geq 6. \end{cases}$$

$$F_B(x) = \begin{cases} 1/2 \text{ avec } & \text{pour } x \in [-5, 5[ \\ 0 & \text{pour } x < -5 \\ 1 & \text{pour } x \geq 5. \end{cases}$$

4.  $\mathbb{E}[A] = \sum_{i \in \{1, 2, 3, 4, 5, 6\}} i\mathbb{P}_A(i) = 3,5$   
 $\mathbb{E}[B] = \sum_{i \in \{-5, 5\}} i\mathbb{P}_B(i) = 0$

5.  $\mathbb{V}[A] = \sum_{i \in \{1, 2, 3, 4, 5, 6\}} i^2\mathbb{P}_A(i) - \mathbb{E}[A]^2 \approx 2,92$   
 $\mathbb{V}[B] = \sum_{i \in \{-5, 5\}} i^2\mathbb{P}_B(i) = 25$

**VA continue**1. Variable  $A$ 

(a)  $p_A(x) = \frac{1}{1000} \text{rect}\left(\frac{x-1000}{1000}\right)$ .

(b)  $F_A(x) = \frac{(x-500)}{1000}$  pour  $x \in [500 - 1500]$

(c)  $\mathbf{E}[A] = \int x \cdot p_A(x) dx = \int_{500}^{1500} x \cdot \frac{1}{1000} dx = \frac{1}{2000} (1500^2 - 500^2) = \frac{1}{2000} (2000)(1000) = 1000$   
(!)

2. Variable  $B$ 

(a)  $p_B(x) = \frac{1}{I_0} \text{rect}\left(\frac{x-I_0}{I_0}\right)$  où  $I_0 = 0.3 \text{ (A/W)} \times 10^3 \text{ (W/m}^2) \times 10^{-2} \text{ (m}^2) = 3 \text{ (A)}$ .

(b)  $F_B(x) = \frac{(x-1.5)}{3}$  pour  $x \in [1.5, 4.5]$ .

(c)  $\mathbf{E}[B] = 3$

3. Variable  $C$ 

(a)  $p_C(i) = \frac{1}{2}$  si  $i \in [0; 1]$ , 0 sinon.

(b)  $\mathbf{E}[C] = \frac{1}{2}$ .

**Propriété d'une VA**La fonction  $f$  définie par :

$$f(x) = \frac{1}{4} - \lambda x^2 \text{ pour } x \in [-1, 1]$$

$$f(x) = 0 \text{ sinon}$$

1. est une densité de probabilité si :

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

$$\iff$$

$$\int_{-1}^{+1} \left(\frac{1}{4} - \lambda x^2\right) dx = 1$$

$$\iff$$

$$\frac{1}{2} - \lambda \left[\frac{x^3}{3}\right]_{-1}^{+1} = \frac{1}{2} - \lambda \frac{2}{3} = 1$$

$$\iff$$

$$\lambda = -\frac{3}{4}$$

On vérifie que  $f(x) \geq 0, \forall x$ 2.  $\mathbf{E}[X] = 0$  car la fonction  $f$  est paire.

$$\begin{aligned} \text{var}[X] &= \mathbf{E}[X^2] \\ &= \int_{-1}^{+1} x^2 \left(\frac{1}{4} + \frac{3}{4}x^2\right) dx \\ &= \frac{1}{4} \left[\frac{x^3}{3}\right]_{-1}^{+1} + \frac{3}{4} \left[\frac{x^5}{5}\right]_{-1}^{+1} \\ &= \frac{1}{6} + \frac{3}{10} = \frac{7}{15} \end{aligned}$$

# Bibliographie

- [1] W. Appel (2008), *Mathématiques pour la physique et les physiciens*, 4ème éd., H&K éditions, Paris.
- [2] N. Boccara (1995), *Probabilités*, Ellipses, Paris.
- [3] C. Deschamps et A. Warusfel (1999), *Mathématiques 1ère année, cours et exercices corrigés - MPSI, PCSI, PTSI*, Dunod, Paris.
- [4] A. Papoulis et S. U. Pillai (2002), *Probability, Random Variables, and Stochastic Processes*, 4ème éd., International edition, McGraw-Hill, New York.
- [5] Le cours de MATHS ET SIGNAL donné par Gaëtan Messin de 2009 à 2015.
- [6] N. Boccara (1995), *Intégration*, Ellipses, Paris.
- [7] P. Brémaud (1997), *Introduction aux Probabilités – Modélisation des Phénomènes Aléatoires*, Springer, Berlin.
- [8] Note : Le théorème de Fubini est un théorème fondamental de la théorie de l'intégration qui autorise l'inversion des sommes dans le cas de l'intégrale de fonction à deux variables. Le théorème ne sera pas rappelé ici mais on notera juste que la nature des densités de probabilité et des espaces probabilisés permet de l'appliquer. Ceux qui sont intéressés par sa démonstration complète sont invités à regarder – par exemple – la référence [6].