

# Probabilités

Matthieu Kowalski





# Table des matières

<b>I. Probabilités discrètes</b>	<b>5</b>
<b>1. Issues, Probabilités, Variables Aléatoires</b>	<b>7</b>
1.1. Introduction . . . . .	7
1.2. Espace des Issues . . . . .	8
1.3. Probabilités . . . . .	8
1.4. Variable aléatoires . . . . .	10
1.4.1. Définition . . . . .	10
1.4.2. Loi d'une variable aléatoire . . . . .	11
1.4.3. Fonction indicatrice . . . . .	14
<b>2. Indépendance, Probabilités conditionnelles</b>	<b>17</b>
2.1. Indépendance . . . . .	17
2.2. De la loi de Bernoulli à la loi Binomiale . . . . .	18
2.2.1. Loi de Bernoulli . . . . .	18
2.2.2. Loi géométrique . . . . .	19
2.2.3. Loi binomiale . . . . .	19
2.3. Probabilités Conditionnelles . . . . .	21
2.3.1. Définition . . . . .	21
2.3.2. Formules des probabilités totales et composées . . . . .	22
<b>3. Espérance, Variance</b>	<b>23</b>
3.1. Espérance . . . . .	23
3.1.1. Définition . . . . .	23
3.1.2. Propriétés . . . . .	25
3.2. Variance . . . . .	26
3.3. Lois classiques et leurs moments . . . . .	28
3.3.1. Loi de Bernoulli . . . . .	28
3.3.2. Loi géométrique . . . . .	28
3.3.3. Loi binomiale . . . . .	29
3.3.4. Loi de Poisson . . . . .	30
<b>II. Probabilités continues</b>	<b>33</b>
<b>4. Probabilités et variables aléatoires continus</b>	<b>35</b>
4.1. Axiomatique des probabilités . . . . .	35
4.2. Variable aléatoires continus . . . . .	38
4.2.1. Variable absolument continue . . . . .	39

Table des matières

4.2.2. Indépendance . . . . .	40
4.2.3. Moments . . . . .	41
4.3. Lois continues classiques . . . . .	42
4.3.1. Loi uniforme sur un intervalle $[a,b]$ . . . . .	42
4.3.1.1. Définition et propriétés . . . . .	42
4.3.1.2. Preuves . . . . .	42
4.3.2. Loi exponentielle de paramètre $\lambda$ . . . . .	43
4.3.2.1. Définition et propriétés . . . . .	43
4.3.2.2. Preuves . . . . .	43
4.3.3. Loi normale de moyenne $\mu$ et variance $\sigma^2$ . . . . .	44
4.3.3.1. Définition et propriétés . . . . .	44
4.4. Fonction caractéristique . . . . .	44
<b>5. Vecteurs de variables aléatoires</b> . . . . .	<b>47</b>
5.1. Couple de variables aléatoires . . . . .	47
5.1.1. Loi jointe, lois marginales . . . . .	47
5.1.2. Variables aléatoires indépendantes . . . . .	48
5.1.3. Moments d'un couple . . . . .	49
5.1.4. Fonction caractéristique d'un couple . . . . .	51
5.2. Vecteur de variables aléatoires . . . . .	52
<b>6. Suite de variables aléatoires</b> . . . . .	<b>57</b>
6.1. Inégalités . . . . .	57
6.2. Modes de convergence . . . . .	58
6.2.1. Limite d'une suite de variable aléatoire . . . . .	58
6.2.2. Convergence en loi . . . . .	59
6.3. Loi des grands nombres . . . . .	60
6.4. Théorème de la limite centrale . . . . .	61

Première partie

Probabilités discrètes



# 1

## ■ Issues, Probabilités, Variables Aléatoires

### Objectifs:

- *Maitrise du vocabulaire propre aux probabilités*
- *Notion d'évènement, d'univers*
- *Notion de Variables Aléatoires*

### Contents

---

<b>1.1. Introduction</b> . . . . .	<b>7</b>
<b>1.2. Espace des Issues</b> . . . . .	<b>8</b>
<b>1.3. Probabilités</b> . . . . .	<b>8</b>
<b>1.4. Variable aléatoires</b> . . . . .	<b>10</b>
1.4.1. Définition . . . . .	10
1.4.2. Loi d'une variable aléatoire . . . . .	11
1.4.3. Fonction indicatrice . . . . .	14

---

### 1.1. Introduction

But : modéliser des expériences où plusieurs issues sont possibles, mais où leurs réalisations ne sont pas déterminées à l'avance (par exemple un lancer de dés), ceci en vue d'évaluer les risques ou de mettre sur pied des stratégies pour faire face aux aléas. La théorie des probabilités ne va pas permettre de prédire quelle issue va se réaliser, mais quelle chance a chaque issue de se réaliser.

Ainsi, dans un premier temps, on va associer à chaque issue possible un nombre entre 0 et 1 qui traduit notre estimation des chances que cette issue a de se réaliser : on appelle ce nombre la probabilité de cette issue. On appelle "évènement" un ensemble d'issues. La probabilité qu'on associe à un évènement est la somme des probabilités de chaque issue de cet ensemble. Typiquement, la question est de déterminer la probabilité d'un évènement. La difficulté est d'une part de décrire l'expérience de façon commode afin d'énumérer toutes les issues possibles et leur

probabilité respective et d'autre part de décomposer l'évènement considéré en issues.

On peut donner une première définition naïve d'une probabilité :

**Définition 1.1**

Une probabilité est un nombre compris entre 0 et 1, associé à un évènement et indiquant les chances de réalisation de cet évènement au cours d'une expérience aléatoire.

## 1.2. Espace des Issues

Avant de calculer les probabilités d'évènements, il faut définir l'espace des issues de façon commode et complète. Cet espace comprendra toutes les issues possibles du jeu ou de l'expérience aléatoire que l'on considère, même éventuellement celles qui ne nous intéressent pas a priori. Dans chaque situation, l'espace des issues sera noté  $\Omega$ , alors que les issues seront notées  $\omega$ .

Autrement dit, pour modéliser une expérience aléatoire, on commence par préciser un **univers**  $\Omega$  des états possibles (choix pas nécessairement unique). Un **évènement** lié à l'expérience correspond à une "question pertinente" qui peut recevoir une réponse OUI ou NON à l'issue de l'expérience. On assimile l'évènement au sous-ensemble  $A$  de  $\Omega$  des états qui donnent la réponse OUI.

**Exemple 1.1**

**(Lancer de dé)**

- Univers :  $\Omega = \{1, 2, 3, 4, 5, 6\}$
- obtenir un chiffre pair :  $A = \{2, 4, 6\}$
- obtenir le chiffre 1 :  $A = \{1\}$
- obtenir un chiffre plus grand ou égal à 5 :  $A = \{5, 6\}$

Ces notions sont résumées dans la définition suivante :

**Définition 1.2**

Une expérience aléatoire est caractérisée par l'ensemble  $\Omega$  des résultats possibles, que l'on suppose dans premier temps fini ou dénombrable.  $\Omega$  est aussi appelé espace ou univers probabiliste. Un évènement  $A$  est une partie quelconque de  $\Omega$ . Un évènement  $A$  est réalisé si le résultat de l'expérience est dans  $\Omega$ .

## 1.3. Probabilités

On considère des expériences avec un nombre fini ou dénombrable d'issues. On note  $\Omega$  cet ensemble d'issues et  $\mathcal{F}$  l'ensemble des évènements.  $\mathcal{F}$  est l'ensemble

des parties de  $\Omega$ . Si  $\Omega = \{1, 2, 3\}$ , alors

$$\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\}.$$

Si  $A$  et  $B$  sont deux sous-ensembles de  $\Omega$ , on note

- $A \cup B = \{\omega; \omega \in A \text{ ou } \omega \in B\}$  :  $A$  ou  $B$  se réalise
- $A \cap B = \{\omega; \omega \in A \text{ et } \omega \in B\}$  :  $A$  et  $B$  se réalise
- $A \setminus B = \{\omega; \omega \in A \text{ et } \omega \notin B\}$  :  $A$  privé de  $B$ . L'évènement  $A$  se réalise et l'évènement  $B$  ne se réalise pas (se note aussi  $A - B$ ).
- $A^c = \Omega \setminus A$  : complémentaire de  $A$ . L'évènement  $A$  ne se réalise pas

Deux évènements  $A$  et  $B$  de  $\mathcal{F}$  sont **disjoints** s'ils n'ont aucune issue en commun, c'est à dire que  $A \cap B = \emptyset$ . Par exemple,  $A$  et  $A^c$  sont disjoints, ainsi que  $\emptyset$  et  $A$ . On peut maintenant définir une mesure de probabilité dans ce cadre d'expériences avec des issues dénombrables.

### Définition 1.3 (Probabilité)

Une probabilité  $P$  sur  $\Omega$  est une application

$$P : \mathcal{P}(\Omega) \rightarrow [0, 1]$$

$$A \mapsto P(A)$$

telle que

- $P(\emptyset) = 0$  (évènement impossible)
- $P(\Omega) = 1$  (évènement certain)
- si  $A \cap B = \emptyset$ , alors  $P(A \cup B) = P(A) + P(B)$

De cette définition découle directement les propriétés suivantes

### Proposition 1.1

1. Pour tout évènement  $A$ ,  $P(A) + P(A^c) = 1$
2. Si  $A \subset B$ , alors  $P(A) \leq P(B)$
3. Si  $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$  et que  $\forall i, P(\omega_i) = p_i$ , alors

$$\sum_{i=1}^{\infty} p_i = 1$$

4. Si  $(A_i)_{i \in I}$  est une famille dénombrable d'évènements disjoints, (i.e.  $A_i \cap A_j = \emptyset$  si  $i \neq j$ ), alors

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i)$$

5. Pour deux évènements quelconques  $A$  et  $B$

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

6. Si  $(A_n)_{n \geq 1}$  est une suite croissante d'évènements (i.e.  $A_n \subset A_{n+1}$ ), alors  $P(\cup_{n \geq 1} A_n) = \lim_{n \rightarrow \infty} P(A_n)$

*Démonstration.* Dans l'ordre :

- $A \cap A^c = \emptyset$ , donc  $P(A \cup A^c) = P(A) + P(A^c)$ . De plus  $A \cup A^c = \Omega$ . Donc  $P(A \cup A^c) = P(A) + P(A^c) = P(\Omega) = 1$ .
- Comme  $A \subset B$ , on peut écrire  $B = A \cup (B - A)$ , avec  $A \cap (B - A) = \emptyset$ . Donc  $P(B) = P(A) + P(B - A)$ , et comme  $P$  est à valeur positive on a  $P(A) < P(B)$
- On applique récursivement la propriété "si  $A \cap B = \emptyset$ , alors  $P(A \cup B) = P(A) + P(B)$ ".
- $P(A \cup B) = P(\{A \cap B^c\} \cup B) = P(A \cap B^c) + P(B)$  car  $\{A \cap B^c\} \cap B = \emptyset$ . De plus, on peut écrire  $P(A) = P(A \cap \{B \cup B^c\}) = P(A \cap B) + P(A \cap B^c)$ . Donc  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

■

$\Omega$  étant un ensemble fini ou dénombrable, se donner une probabilité  $P$  sur l'ensemble des parties de  $\Omega$  n'est rien d'autre que se donner la fonction sur  $\Omega : \omega \in \Omega \mapsto p(\omega) = P(\{\omega\})$ . En effet, pour tout évènement  $A$ , on peut écrire  $A$  comme la réunion disjointe et au plus dénombrable des singletons des éléments qui le composent. De plus, on a  $P(A) = \sum_{\omega \in A} p(\omega)$ . Cette fonction vérifie les propriétés suivantes :

- $p : \Omega \mapsto [0, 1]$ ;
- $\sum_{\omega \in \Omega} p(\omega) = 1$

Une telle fonction est une distribution de probabilité.

## 1.4. Variable aléatoires

La notion de variable aléatoire joue un rôle central en probabilité. La majorité des expériences peuvent se modéliser simplement à l'aide d'une variable aléatoire.

### 1.4.1. Définition

**Définition 1.4**

**(Variable aléatoire)**

Une variable aléatoire est une fonction de  $\Omega$  à valeur dans  $\mathbb{R}$ .

Une variable aléatoire est donc une **fonction** tout à fait classique. Pourquoi donc une définition de plus? En analyse, il est fondamental de savoir quel est le **domaine de définition** d'une fonction comme  $\sin(x)$ , ou  $\log(x)$ . On doit savoir où se trouve  $x$  : sur  $]0, \pi[$ , ou sur  $]0, \infty[$ , etc. En probabilité, ce qui est important est **la**

**valeur** de la fonction. En effet, le domaine  $\Omega$  n'est pas unique, puisqu'il dépend de notre façon de décrire le jeu ou l'expérience qu'on modélise; en revanche, ce qui est très important est la probabilité que l'on associe aux réalisations de la fonction.

Intuitivement, il s'agit donc d'un nombre dont la valeur dépend du résultat d'une expérience aléatoire. Souvent, on ignorera l'expérience en question : on ne fera aucune allusion à  $\Omega$  ou à l'un de ses éléments. La loi d'une variable aléatoire sera alors la probabilité qu'elle prenne une certaine valeur.

Les variables aléatoires sont en général notées par des lettres majuscules  $X, Y, \dots$ . Les valeurs qu'elles prennent lorsque l'issue  $\omega$  se réalise sont notées par  $X(\omega), Y(\omega), \dots$

### Exemple 1.2

Dans l'exemple d'un lancer de dé, le résultat est une variable aléatoire.

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

$$X : \Omega \rightarrow \mathbb{N}$$

$$\omega \mapsto \omega$$

Si l'on s'intéresse au résultat de la somme de deux dés, on a une autre variable aléatoire

$$\Omega = \{i + j, i, j \in \{1, 2, 3, 4, 5, 6\}\}$$

$$X : \Omega \rightarrow \mathbb{N}$$

$$(i, j) \mapsto i + j$$

### 1.4.2. Loi d'une variable aléatoire

Soit  $\Omega$  un espace d'issues,  $P$  une probabilité sur  $\Omega$ , et  $X$  une variable aléatoire définie sur  $\Omega$ . Soit  $X(\Omega) = \{x_1, \dots, x_n, \dots\} \subset \mathbb{R}$  l'ensemble des valeurs distinctes prises par  $X$ . Par convention, on note  $\{X = x_i\} \stackrel{\text{def}}{=} \{\omega; X(\omega) = x_i\}$ , et  $P(X = x_i) \stackrel{\text{def}}{=} P(\{\omega; X(\omega) = x_i\})$ . De même, si  $A$  est un sous-ensemble de  $X(\Omega)$ ,  $\{X \in A\} \stackrel{\text{def}}{=} \{\omega; X(\omega) \in A\}$ , et  $P(X \in A) = P(\{\omega; X(\omega) \in A\})$ .

### Définition 1.5 (Loi d'une V.A.)

La loi de  $X$  est la probabilité  $P_X$  sur  $X(\Omega)$ , donnée par :

$$\forall A \subset X(\Omega), P_X(A) = P(X \in A) .$$

La distribution de la loi de  $X$  est donc

$$p_X : X(\Omega) \rightarrow [0, 1]$$

$$x \mapsto p_X(x) = P(X = x)$$

En d'autres termes, c'est la probabilité des événements  $\omega$  qui vérifient  $X(\omega) = x$ ,

## 1. Issues, Probabilités, Variables Aléatoires

donc la probabilité que  $X$  prenne la valeur  $x$  lors d'une expérience aléatoire. Nécessairement, on a :

$$p(x) = 0 \text{ si } x \notin X(\omega)$$
$$\sum_x p(x) = 1 \text{ car } X \text{ prend forcément une valeur.}$$

Ainsi, pour connaître la loi d'une variable, il faut connaître l'ensemble de ses valeurs possibles, et la probabilité avec laquelle elle réalise chaque valeur.

### Exemple 1.3

Deux cas simples utilisant des jetés de dés.

- La loi du lancer de dé est donnée par

$$P(X = 1) = P(X = 2) = \dots = P(X = 6) = 1/6 .$$

- La loi de la somme du lancer de deux dés est donnée par

$$P(X = 2) = P(\{1, 1\}) = 1/36$$

$$P(X = 3) = P(\{1, 2\}, \{2, 1\}) = 2/36$$

$$P(X = 4) = P(\{1, 3\}, \{2, 2\}, \{3, 1\}) = 3/36$$

$$P(X = 5) = P(\{1, 4\}, \{2, 3\}, \{3, 2\}, \{4, 1\}) = 4/36$$

$$P(X = 6) = P(\{1, 5\}, \{2, 4\}, \{3, 3\}, \{4, 2\}, \{5, 1\}) = 5/36$$

$$P(X = 7) = P(\{1, 6\}, \{2, 5\}, \{3, 4\}, \{4, 3\}, \{5, 2\}, \{6, 1\},) = 6/36$$

$$P(X = 8) = P(\{2, 6\}, \{3, 5\}, \{4, 4\}, \{5, 3\}, \{6, 2\},) = 5/36$$

$$P(X = 9) = P(\{3, 6\}, \{4, 5\}, \{5, 4\}, \{6, 3\},) = 4/36$$

$$P(X = 10) = P(\{4, 6\}, \{5, 5\}, \{6, 4\},) = 3/36$$

$$P(X = 11) = P(\{5, 6\}, \{6, 5\},) = 2/36$$

$$P(X = 12) = P(\{6, 6\}) = 1/36$$

Dans les deux cas, on vérifie qu'on a bien une distribution de probabilité en vérifiant que la somme fait 1.

On rencontrera souvent des énoncés du type : "Soit  $X$  une variable aléatoire de loi [...]". Il faut comprendre ici que ce qui nous intéresse est la loi de  $X$  (les valeurs prises par  $X$  et les probabilités associées), et pas du tout l'espace  $\Omega$  sur lequel est définie  $X$ . Dans beaucoup de situations, l'espace  $\Omega$  ne sera même pas explicité.

La loi d'une variable aléatoire est parfaitement caractérisée par ce qu'on appelle la fonction de répartition, dont on donne la définition ci-après. La démonstration de la caractérisation de la loi par cette fonction viendra après avoir énoncé quelques propriétés élémentaires.

**Définition 1.6 (Fonction de répartition)**

La fonction de répartition  $F_X$  d'une variable aléatoire  $X$  est définie par

$$F_X : \mathbb{R} \rightarrow [0, 1]$$

$$x \mapsto P(X \leq x)$$

La fonction de répartition possède les propriétés élémentaires suivantes.

**Proposition 1.2**

Soit  $F_X$  la fonction de répartition de la variable aléatoire  $X$

1.  $F_X$  est une fonction croissante : si  $x \leq y$   $F_X(x) \leq F_X(y)$
2.  $F_X$  est continue à droite
3.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$
4.  $\lim_{x \rightarrow +\infty} F_X(x) = 1$
5.  $\forall x \in \mathbb{R}, P(X = x) = F_X(x) - \lim_{y \rightarrow x^-} F_X(y)$
6. Soit  $(\dots, x_{-1}, x_0, x_1, \dots)$  les valeurs distinctes prises par  $X$  rangées par ordre croissant. La fonction  $F_X$  est une fonction en escalier, constante sur les intervalles  $[x_i, x_{i+1}[$ , qui fait un saut au point  $x_i$  d'une amplitude  $P(X = x_i)$  (voir figure 1.1)

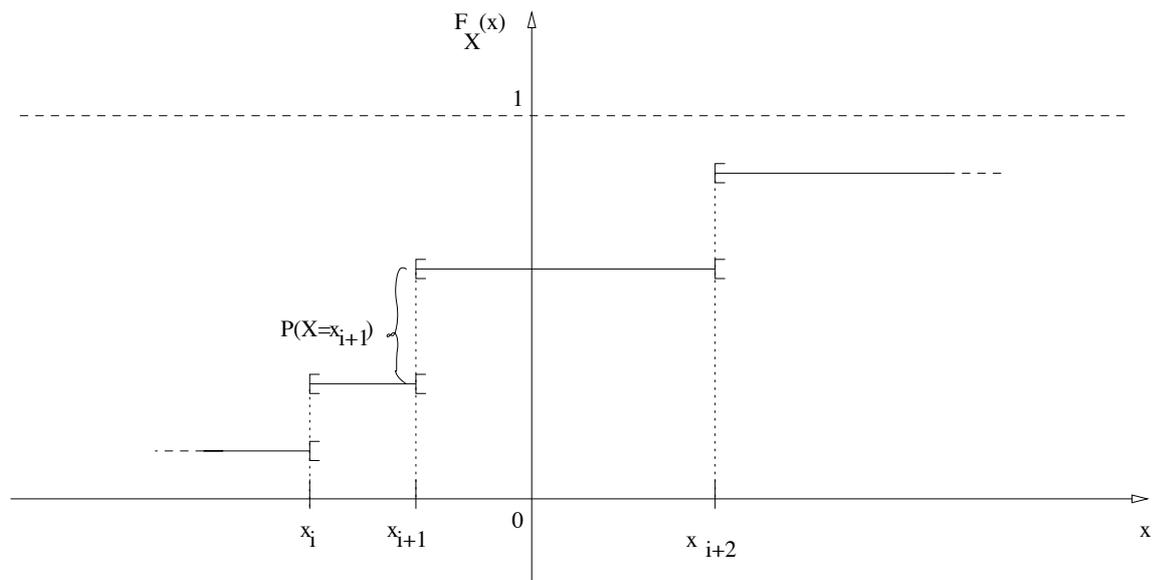


FIGURE 1.1. – Fonction de répartition

**Proposition 1.3**

La loi d'une variable aléatoire  $X$  est caractérisée par sa fonction de répartition  $F_X$ .

*Démonstration.* Pour démontrer cette proposition, il s'agit de voir qu'à partir de la donnée de la fonction  $F_X$ , on est capable de retrouver  $(\dots, x_{-1}, x_0, x_1, \dots, x_n, \dots)$  les valeurs distinctes prises par  $X$  (rangées par ordre croissant), et les nombres  $p_X(x_i) = P(X = x_i)$ . D'après la proposition précédente, les valeurs prises par  $X$  sont les points où  $F_X$  est discontinue, et les valeurs  $P(X = x_i)$  sont les amplitudes des sauts en ces points de discontinuité. ■

**1.4.3. Fonction indicatrice**

**Définition 1.7**

**(Fonction indicatrice)**

La fonction indicatrice  $\mathbb{1}_A$  d'un évènement  $A \subset \Omega$  est la variable aléatoire

$$\mathbb{1}_A : \Omega \rightarrow \{0, 1\}$$

$$\omega \mapsto \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$$

On rappelle qu'une partition de  $\Omega$  est une suite de sous-ensembles  $(A_i)_{i \in I}$  disjoints deux à deux, de réunion  $\Omega$ . En d'autres termes,

$$\forall i \neq j, A_i \cap A_j = \emptyset \text{ et } A_1 \cup A_2 \cup \dots \cup A_n = \Omega.$$

On donne alors quelques propriétés simples des fonctions indicatrices.

**Proposition 1.4**

1.  $\mathbb{1}_{A \cap B} = \mathbb{1}_A \mathbb{1}_B$
2. Si  $A \cap B = \emptyset$ , alors  $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B$
3.  $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$
4. Si  $(A_i)_{i \in I}$  est une partition de  $\Omega$ , alors  $\sum_{i \in I} \mathbb{1}_{A_i} = 1$

Les fonctions indicatrices permettent la définition suivante, liée aux variables aléatoires

**Définition 1.8**

Soit  $(A_i)_{i \in I}$  une partition de  $\Omega$ . On dit que  $X$  se décompose sur la partition  $(A_i)_{i \in I}$  ssi  $X$  peut s'écrire sous la forme (non unique)

$$X = \sum_{i \in I} x_i \mathbb{1}_{A_i};$$

ce qui signifie que  $X$  est constante sur les événements  $A_i$  et prend la valeur  $x_i$  sur  $A_i$ .

À toute variable aléatoire  $X$ , on peut associer une partition de  $\Omega$  de la façon suivante : soit  $x_1, \dots, x_n, \dots$  l'ensemble des valeurs distinctes prises par  $X$ . Posons

$$A_i = \{\omega : X(\omega) = x_i\} .$$

Les  $A_i$  forment bien une partition de  $\Omega$  :

$$A_i \cap A_j = \{\omega : X(\omega) = x_i \text{ et } X(\omega) = x_j\} = \emptyset, \text{ si } i \neq j, \text{ et } \cup_i A_i = \{\omega : \exists i \text{ tel que } X(\omega) = x_i\} = \Omega .$$

De plus, par définition des  $A_i$ , on a  $X = \sum_{i \in I} x_i \mathbb{1}_{A_i}$



# 2 ■ Indépendance, Probabilités conditionnelles

**Objectifs:**

- Connaître parfaitement la notion d'indépendance
- Manipuler les probabilités conditionnelles

**Contents**

---

<b>2.1. Indépendance</b> . . . . .	<b>17</b>
<b>2.2. De la loi de Bernoulli à la loi Binomiale</b> . . . . .	<b>18</b>
2.2.1. Loi de Bernoulli . . . . .	18
2.2.2. Loi géométrique . . . . .	19
2.2.3. Loi binomiale . . . . .	19
<b>2.3. Probabilités Conditionnelles</b> . . . . .	<b>21</b>
2.3.1. Définition . . . . .	21
2.3.2. Formules des probabilités totales et composées . . . . .	22

---

## 2.1. Indépendance

Soit  $\Omega$  un ensemble d'issues et  $P$  une probabilité sur  $\Omega$ .

**Définition 2.1** (Évènements indépendants)

Deux évènements  $A$  et  $B$  sont indépendants ssi

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

On peut alors définir aussi l'indépendance de deux variables aléatoires

## 2. Indépendance, Probabilités conditionnelles

### Définition 2.2 (V.A. indépendantes)

Deux variables aléatoires  $X : \Omega \rightarrow \{x_i, i \in I\}$  et  $Y : \Omega \rightarrow \{y_j, j \in J\}$  sont dites indépendantes ssi pour tout  $i, j$

$$P(\{X = x_i\} \cap \{Y = y_j\}) = P(X = x_i)P(Y = y_j)$$

On généralise ces deux définitions à un nombre fini ou dénombrable d'évènements  $\{A_i, i \in I\}$  et de variables aléatoires  $\{X_i, i \in I\}$

### Définition 2.3

Les évènements  $\{A_i, i \in I\}$  sont indépendants si pour tout ensemble d'indices  $J \subset I$

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j)$$

Par exemple, si l'on a trois évènements  $\{A, B, C\}$  il faut vérifier que

$$P(A \cap B \cap C) = P(A)P(B)P(C), P(A \cap B) = P(A)P(B), P(A \cap C) = P(A)P(C) \text{ et } P(C \cap B) = P(C)P(B)$$

### Définition 2.4

Les variables aléatoires  $\{X_i, i \in I\}$  sont indépendantes si pour tout ensemble d'indices  $J \subset I$  et tous réels  $x_j$  appartenant aux valeurs possibles de  $X_j, j \in J$

$$P\left(\bigcap_{j \in J} \{X_j = x_j\}\right) = \prod_{j \in J} P(X_j = x_j)$$

On a de plus la propriété suivante sur l'image de deux variables aléatoires

### Proposition 2.1

Soient deux variables aléatoires  $X, Y$  indépendantes. Si  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  sont deux fonctions quelconques, alors  $f(X)$  et  $g(Y)$  sont indépendantes.

## 2.2. De la loi de Bernoulli à la loi Binomiale

### 2.2.1. Loi de Bernoulli

On appelle pile-face toute expérience qui a deux issues possibles (pile/face; échec/succès; malade/sain; fumeur/non fumeur ...). L'espace des issues est ici un espace à deux éléments, notés 0 et 1 par commodité :

$$\Omega_1 = \{0, 1\}.$$

Se donner une probabilité  $P_1$  sur  $\Omega_1$ , revient donc à se donner un nombre  $p \in [0, 1]$  tel que

$$P_1(1) = p, P_1(0) = 1 - p.$$

## 2.2. De la loi de Bernoulli à la loi Binomiale

Par exemple, dans le cas du lancer d'une pièce équilibrée,  $p = 1/2$ . Dans l'exemple fumeur/non fumeur, en notant 1 l'issue "fumeur",  $p$  est la proportion de fumeurs dans la population considérée.

$P_1$  précédemment définie s'appelle **la loi de Bernoulli** de paramètre  $p$ . Ainsi, on dira qu'une variable  $X$  est une variable de Bernoulli de paramètre  $p$  (et on note  $X \sim \mathcal{B}(p)$ ), si  $X$  prend ses valeurs dans  $\{0, 1\}$  et  $P(X = 1) = p = 1 - P(X = 0)$ . On peut alors écrire la définition suivante

### Définition 2.5

#### (Loi de Bernoulli)

La loi de Bernoulli de paramètre  $p \in [0, 1]$ , notée  $\mathcal{B}(p)$  est définie sur un univers à deux éléments  $\Omega = \{0, 1\}$  par

$$P(0) = 1 - p \quad P(1) = p$$

La loi de Bernoulli correspond à la probabilité d'obtenir un succès lors d'un lancer.

### 2.2.2. Loi géométrique

En reprenant l'exemple du lancer de pièce, on peut s'intéresser au nombre de lancers nécessaire pour obtenir "face". Si on note  $T$  ce nombre de lancers, on a alors obtenue  $T - 1$  fois pile avec la probabilité  $(1 - p)$  puis une fois face (au  $T$ -ème lancer) avec la probabilité  $p$ . Ceci définit la loi géométrique de paramètre  $p$ .

### Définition 2.6

#### (Loi géométrique)

La loi géométrique de paramètre  $p \in [0, 1]$  est donnée par :

$$P(T = k) = (1 - p)^{k-1} p.$$

On note  $T \sim \mathcal{G}(p)$ .

La loi géométrique correspond à la probabilité d'obtenir un succès au bout de  $k$  lancers.

Dans le cas d'un lancer de pièce équilibré ( $p = 1/2$ ), on a  $P(T = k) = \frac{1}{2^k}$ .

### 2.2.3. Loi binomiale

Considérons maintenant le cas où l'on lance  $n$  fois la même pièce de paramètre  $p$ . Une issue consiste donc à spécifier l'état 1 ou 0 de chaque lancer.

$$\Omega_n = \{(\omega_1, \dots, \omega_n), \text{ où } \omega_i \in \{0, 1\}\}$$

$\Omega_n$  a  $2^n$  éléments. Les  $n$  lancers sont semblables et indépendants, donc

$$\forall \omega = (\omega_1, \dots, \omega_n) \in \Omega_n, P_n(\omega_1, \dots, \omega_n) = P_1(\omega_1)P_1(\omega_2) \dots P_1(\omega_n).$$

## 2. Indépendance, Probabilités conditionnelles

On peut définir sur  $\Omega_n$   $n$  variables aléatoires de Bernoulli  $X_1, \dots, X_n$  où

$$\forall i = 1, \dots, n \quad \forall \omega \in \Omega_n \quad X_i(\omega) = \omega_i$$

$X_i \sim \mathcal{B}(p)$  donne ainsi le résultat du  $i$ -ème lancer.

Sur  $\Omega_n$ , on s'intéresse alors au nombre de fois où "face" a été tirés sur les  $n$  lancers. On définit pour cela la variable

$$S_n(\omega) = X_1(\omega) + X_2(\omega) + \dots + X_n(\omega) = \omega_1 + \omega_2 + \dots + \omega_n$$

$S_n$  correspond alors au nombre de 1 (ou de face) dans l'issue  $\omega$ . Dans l'exemple d'un sondage de  $n$  personnes sur la question fumeur/non fumeur,  $S_n$  représente le nombre de fumeurs sur les  $n$  personnes interrogées, si on a choisi de noter par 1 l'issue "fumeur".

Calculons la loi de  $S_n$ . Tout d'abord, l'ensemble des valeurs possibles pour  $S_n$  est l'ensemble  $\{0, 1, \dots, n\}$ . Il s'agit donc de calculer pour tout  $k \in \{0, \dots, n\}$ ,  $P_n(S_n = k)$ . Si  $\omega$  est tel que  $S_n(\omega) = k$ , le  $n$ -uplet  $\omega$  est constitué de  $k$  1 et  $n - k$  0. Par conséquent

$$P_n(\omega) = P_1(\omega_1)P_1(\omega_n) = p^k(1-p)^{n-k}$$

on a donc

$$P_n(S_n = k) = \sum_{\omega; S_n(\omega) = k} P_n(\omega) = \text{card}(\{\omega; S_n(\omega) = k\}) p^k(1-p)^{n-k}$$

$\text{card}(\{\omega; S_n(\omega) = k\})$  correspond au nombre de choix des  $k$  lancers où pile sort, parmi les  $n$  lancers, i.e. de  $k$  positions parmi  $n$  positions possibles. La première position peut se choisir parmi  $n$ ; La seconde doit correspondre à une position différente de celle précédemment choisie, et donc  $n - 1$  choix restent possibles... et ainsi de suite jusqu'à la  $k$ -ème position qui peut être choisie parmi  $n - k$ . Cependant, avec cet algorithme, un même choix de  $k$  positions apparaîtra  $k!$  fois. On a donc  $n(n-1)(n-2)\dots(n-k+1)/k!$  choix possibles. C'est le nombre de combinaison de  $k$  parmi  $n$  :  $\binom{n}{k}$ . On a donc

$$\text{card}(\{\omega; S_n(\omega) = k\}) = \binom{n}{k} = \frac{n!}{(n-k)!k!}$$

et

$$P_n(S_n = k) = \binom{n}{k} p^k(1-p)^{n-k}$$

### Définition 2.7 (Loi binomiale)

Soit  $X_k \sim \mathcal{B}(p)$ ,  $k \in \{1, \dots, n\}$   $n$  variables aléatoires de Bernoulli indépendantes.

Soit

$$S_n = \sum_{k=1}^n X_k$$

On dit que  $S_n$  suit la loi binomiale de paramètres  $n > 0$  et  $p \in [0, 1]$ , qu'on note

$S_n \sim \mathcal{B}(n, p)$ . De plus, on a

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

On note  $S_n \sim \mathcal{B}(n, p)$ .

La loi binomiale correspond à la probabilité d'obtenir  $k$  succès parmi  $n$  lancers.

## 2.3. Probabilités Conditionnelles

### 2.3.1. Définition

#### Définition 2.8

Soit  $A$  et  $B$  deux évènements sur  $(\Omega, \mathcal{F}, P)$  avec  $P(B) > 0$ . La probabilité conditionnelle de  $A$  sachant  $B$  est

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

On dit aussi que  $P(A|B)$  est la probabilité de  $A$  sachant  $B$ .

La propriété suivante est une conséquence directe et intuitive.

#### Proposition 2.2

$P(\cdot|B)$  est une probabilité sur  $\Omega$  telle que  $P(B|B) = 1$ .

Intuitivement, si deux évènements sont indépendants, savoir que  $B$  est réalisé ne nous apprend rien sur  $A$ .

#### Proposition 2.3

Si  $A$  et  $B$  sont deux évènements indépendants avec  $P(B) > 0$  alors

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

*Démonstration.*  $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$  ■

#### Théorème 2.1 (Loi de Bayes)

Soit  $A$  et  $B$  deux évènements sur  $(\Omega, \mathcal{F}, P)$  avec  $P(B) > 0$ . Alors

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

*Démonstration.* On a  $P(B|A) = \frac{P(B \cap A)}{P(A)}$  et donc  $P(A \cap B) = P(B|A)P(A)$ , qu'on injecte dans la définition de  $P(A|B)$ . ■

### 2.3.2. Formules des probabilités totales et composées

**Théorème 2.2 (Formule des probabilités totales)**

Soit  $\{A_i\}_{i \in I}$  une partition de  $\Omega$ , telle que pour tout  $i$ ,  $P(A_i) > 0$ . Pour tout évènement  $B \subset \Omega$ ,

$$P(B) = \sum_{i \in I} P(B|A_i)P(A_i)$$

*Démonstration.* Les  $\{A_i\}_{i \in I}$  formant une partition de  $\Omega$  on  $B = B \cup (\cap A_i) = \cup (B \cap A_i)$ . Comme les  $A_i$  sont disjoints, les  $A_i \cap B$  le sont aussi. Donc,

$$P(B) = \sum P(B \cap A_i) = \sum P(B|A_i)P(A_i)$$

■

**Proposition 2.4 (Formule des probabilités composées)**

Soient  $A_1, A_2, \dots, A_n$  des évènements tels que  $P(A_1 \cap A_2 \cap \dots \cap A_n) > 0$

$$P(A_1 \cap \dots \cap A_n) = P(A_n|A_1 \cap \dots \cap A_{n-1})P(A_{n-1}|A_1 \cap \dots \cap A_{n-2}) \dots P(A_2|A_1)P(A_1).$$

*Démonstration.* La démonstration peut se faire par récurrence. Pour  $n = 2$ , il s'agit simplement de la définition de la probabilité conditionnelle. Supposons la formule vraie à l'ordre  $n$  et montrons la à l'ordre  $n + 1$ .

$$\begin{aligned} P(A_1 \cap \dots \cap A_{n+1}) &= P(A_{n+1}|A_1 \cap \dots \cap A_n)P(A_1 \cap \dots \cap A_n) \text{ par définition} \\ &= P(A_{n+1}|A_1 \cap \dots \cap A_n)P(A_n|A_1 \cap \dots \cap A_{n-1}) \dots P(A_2|A_1)P(A_1) \\ &\text{par hypothèse de récurrence.} \end{aligned}$$

■

# 3 ■ Espérance, Variance

## Objectifs:

- Moments d'une VA
- Comprendre ce qu'est une variance
- Connaître quelques lois discrètes classiques

## Contents

---

<b>3.1. Espérance</b> . . . . .	<b>23</b>
3.1.1. Définition . . . . .	23
3.1.2. Propriétés . . . . .	25
<b>3.2. Variance</b> . . . . .	<b>26</b>
<b>3.3. Lois classiques et leurs moments</b> . . . . .	<b>28</b>
3.3.1. Loi de Bernoulli . . . . .	28
3.3.2. Loi géométrique . . . . .	28
3.3.3. Loi binomiale . . . . .	29
3.3.4. Loi de Poisson . . . . .	30

---

## 3.1. Espérance

### 3.1.1. Définition

#### Définition 3.1

Soit  $\Omega$  un espace fini ou dénombrable,  $P$  une probabilité sur  $\Omega$  et  $X$  une variable aléatoire. On appelle espérance de  $X$ , ou moyenne de  $X$ , la quantité

$$E[X] = \sum_{\omega \in \Omega} P(\omega)X(\omega)$$

Dans le cas où  $\Omega$  est un ensemble dénombrable,  $E[X]$  n'est définie que si la série est absolument convergente. On supposera que cette condition est vérifiée par la suite lorsqu'on parle d'espérance d'une variable.

### 3. Espérance, Variance

#### Proposition 3.1

Si  $(A_i)_{i \in I}$  est une partition de  $\Omega$  et si  $X$  se décompose sur cette partition, c'est-à-dire  $X = \sum_{i \in I} x_i \mathbb{1}_{A_i}$ , alors

$$E[X] = \sum_{i \in I} x_i P(A_i)$$

*Démonstration.* On part de la définition de l'espérance d'une V.A.

$$\begin{aligned} E[X] &= \sum_{\omega \in \Omega} P(\omega) X(\omega) \\ &= \sum_{i \in I} \sum_{\omega \in A_i} P(\omega) X(\omega) \text{ car les } A_i \text{ forment une partition de } \Omega \\ &= \sum_{i \in I} \sum_{\omega \in A_i} x_i P(\omega) \\ &= \sum_{i \in I} x_i \sum_{\omega \in A_i} P(\omega) \\ &= \sum_{i \in I} x_i P(A_i) \end{aligned}$$

■

La définition 3.1 a l'avantage d'être indépendante de tout choix de partition de  $\Omega$ , et donc de décomposition de  $X$  sur  $\Omega$ . Cependant, en pratique, on utilisera très souvent la formule de la proposition 3.1. En effet, pour calculer  $E[X]$ , il suffit de connaître la loi de  $X$ , c'est-à-dire les valeurs distinctes  $(x_i)_{i \in I}$  prises par  $X$  et les probabilités  $P(X = x_i)$ . Ainsi, si l'on note  $A_i = \{X = x_i\}$ , on a bien une partition de l'univers telle que  $X$  se décompose sur cette partition, et donc

$$E[X] = \sum_{i \in I} x_i P(X = x_i)$$

#### Théorème 3.1 (V.A. à valeur dans $\mathbb{N}$ )

Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{N}$ , on a

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

Une conséquence directe de la proposition 3.1 est

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

#### Proposition 3.2 (Image par une fonction)

Soit  $f$  une fonction, alors

$$E[f(X)] = \sum_{\omega \in \Omega} f(X(\omega)) P(\omega)$$

Si  $(A_i)_{i \in I}$  est une partition de  $\Omega$  tel que  $X$  se décompose sur cette partition.  
Alors

$$E[f(X)] = \sum_{i \in I} f(x_i)P(A_i)$$

Avec  $A_i = \{X = x_i\}$ , on a

$$E[f(X)] = \sum_{i \in I} f(x_i)P(X = x_i)$$

**Théorème 3.2 (Image par une fonction d'une V.A. à valeur dans  $\mathbb{N}$ )**

Soient  $f$  une fonction et  $X$  une variable aléatoire à valeurs dans  $\mathbb{N}$ . Alors

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

### 3.1.2. Propriétés

**Proposition 3.3 (Linéarité de l'espérance)**

Soit  $X$  et  $Y$  deux variables aléatoires, et soit  $\lambda \in \mathbb{R}$ . Alors

1.  $E[\lambda X] = \lambda E[X]$
2.  $E[X + Y] = E[X] + E[Y]$

*Démonstration.* En revenant à la définition, on a

1.

$$E[\lambda X] = \sum_{\omega \in \Omega} \lambda X(\omega)P(\omega) = \lambda \sum_{\omega \in \Omega} X(\omega)P(\omega) = \lambda E[X]$$

2.

$$E[X+Y] = \sum_{\omega \in \Omega} (X(\omega)+Y(\omega))P(\omega) = \sum_{\omega \in \Omega} X(\omega)P(\omega) + \sum_{\omega \in \Omega} Y(\omega)P(\omega) = E[X]+E[Y]$$

■

De plus,

**Proposition 3.4 (Propriétés)**

Soient  $X, Y$  deux variables aléatoires sur  $(\Omega, P)$ , et  $\lambda \in \mathbb{R}$

1.  $E[\lambda] = \lambda$
2.  $X \geq 0 \Rightarrow E[X] \geq 0$
3.  $X \geq Y \Rightarrow E[X] \geq E[Y]$
4.  $|E[X]| \leq E[|X|]$

### 3. Espérance, Variance

#### Théorème 3.3 (Produit de deux V.A. indépendantes)

Soient  $X, Y$  deux variables aléatoires **indépendantes**. Alors

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

*Démonstration.* On note  $(x_i)_{i \in I}$  et  $(y_j)_{j \in J}$  les valeurs distinctes prises par  $X$  et  $Y$  respectivement. On note les partitions associées  $A_i = \{X = x_i\}$  et  $B_j = \{Y = y_j\}$ . Ainsi, les  $A_i \cap B_j$  forment aussi une partition de  $\Omega$ , et l'on a

$$X = \sum_{i,j} x_i \mathbb{1}_{A_i \cap B_j}, \quad Y = \sum_{i,j} y_j \mathbb{1}_{A_i \cap B_j} \quad \text{et} \quad XY = \sum_{i,j} x_i y_j \mathbb{1}_{A_i \cap B_j}.$$

Donc

$$E[XY] = \sum_{i,j} x_i y_j P(A_i \cap B_j).$$

$X$  et  $Y$  étant indépendantes, on a  $P(A_i \cap B_j) = P(A_i)P(B_j)$ . Ainsi,

$$\begin{aligned} E[XY] &= \sum_{i,j} x_i y_j P(A_i)P(B_j) \\ &= \left( \sum_i x_i P(A_i) \right) \left( \sum_j y_j P(B_j) \right) \\ &= E[X]E[Y] \end{aligned}$$

■

## 3.2. Variance

#### Définition 3.2

La variance d'une variable aléatoire  $X$  est le nombre positif donné par

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

La variance est définie à l'aide de l'espérance, c'est donc la moyenne de la dispersion des valeurs prises par  $X$  autour de sa moyenne.

#### Proposition 3.5 (Propriétés)

Soit  $X$  une variable aléatoire et  $\lambda \in \mathbb{R}$

1.  $\text{Var}\{\lambda X\} = \lambda^2 \text{Var}\{X\}$
2.  $\text{Var}\{X + \lambda\} = \text{Var}\{X\}$
3.  $\text{Var}\{\lambda\} = 0$
4.  $\text{Var}\{X\} = 0$  ssi  $\forall \omega \in \Omega$  tel que  $P(\Omega) > 0$ ,  $X(\omega) = E[X]$ , c'est à dire que  $X$  est constante sur les issues possibles.

$$\blacksquare \quad 5. \text{Var}\{X\} = E[X^2] - (E[X])^2$$

*Démonstration.* 1. En revenant à la définition

$$\begin{aligned} \text{Var}\{\lambda X\} &= E[(\lambda X - E[\lambda X])^2] \\ &= E[\lambda^2(X - E[X])^2] \text{ par linéarité de l'espérance} \\ &= \lambda^2 \text{Var}\{X\} \text{ par linéarité de l'espérance} \end{aligned}$$

2.  $\text{Var}\{X + \lambda\} = E[(X + \lambda - E[X + \lambda])^2] = E[(X + \lambda - E[X] - \lambda)^2] = \text{Var}\{X\}$
3.  $\text{Var}\{\lambda\} = E[(\lambda - E[\lambda])^2] = E[(\lambda - \lambda)^2] = E[0] = 0$
4.  $\text{Var}\{X\}$  est une somme de terme positifs, donc  $\text{Var}\{X\} = 0$  ssi tous les termes sont nuls.
5. En utilisant la linéarité de l'espérance avec  $(X - E[X])^2 = X^2 - 2XE[X] + E[X]^2$ , on a

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

■

**Proposition 3.6** (Somme de deux V.A. indépendantes)

Soit  $X$  et  $Y$  deux variables aléatoires indépendantes. Alors

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

*Démonstration.* Il suffit de revenir à la définition et de développer

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

Notons  $f$  et  $g$  les deux fonctions  $f(x) = x - E[X]$  et  $g(x) = x - E[Y]$ . On a  $E[f(X)] = 0 = E[g(Y)]$ . De plus, comme  $X$  et  $Y$  sont indépendantes, alors  $f(X)$  et  $g(Y)$  sont indépendantes. donc

$$E[(X - E[X])(Y - E[Y])] = E[X - E[X]]E[Y - E[Y]] = 0$$

■

### 3.3. Lois classiques et leurs moments

#### 3.3.1. Loi de Bernoulli

**Définition 3.3**

**(Loi de Bernoulli)**

Soit  $X \sim \mathcal{B}(p)$ , alors  $X$  prend les valeurs 0 ou 1, et  $P(X = 1) = p = 1 - P(X = 0)$ . (cf. 2.2.1)

On donne l'espérance et la variance d'une V.A. de Bernoulli

**Proposition 3.7**

Soit  $X \sim \mathcal{B}(p)$ , alors

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

*Démonstration.* Par définition,  $X \in \{0, 1\}$  et  $P(X = 1) = p = 1 - P(X = 0)$ . Donc

$$E[X] = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = p$$

De plus,

$$E[X^2] = 1^2 \cdot P(X = 1) + 0^2 \cdot P(X = 0) = p$$

et

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

■

#### 3.3.2. Loi géométrique

**Définition 3.4**

**(Loi géométrique)**

Soit  $X \sim \mathcal{G}(p)$ , alors  $X \in \mathbb{N}^*$  et  $P(X = k) = p(1-p)^{k-1}$ . (cf. 2.2.2)

On donne l'espérance et la variance d'une V.A. de loi géométrique

**Proposition 3.8**

Soit  $X \sim \mathcal{G}(p)$ , alors

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

*Démonstration.*

$$\begin{aligned} E[X] &= \sum_{k=1}^{+\infty} kP(X = k) = p \sum_{k=1}^{+\infty} k(1-p)^{k-1} = -p \frac{d}{dp} \left( \sum_{k=0}^{+\infty} (1-p)^k \right) \\ &= -p \frac{d}{dp} \left( \frac{1}{p} \right) = \frac{p}{p^2} = \frac{1}{p} \end{aligned}$$

$$\begin{aligned}
 E[X^2] &= \sum_{k=1}^{+\infty} k^2 P(X=k) = p \sum_{k=1}^{+\infty} k^2 (1-p)^{k-1} \\
 &= p \sum_{k=1}^{+\infty} k(k-1)(1-p)^{k-1} + p \sum_{k=1}^{+\infty} k(1-p)^{k-1} \\
 &= p(1-p) \sum_{k=2}^{+\infty} k(k-1)(1-p)^{k-2} + E[X] \\
 &= p(1-p) \frac{d^2}{dp^2} \left( \sum_{k=0}^{+\infty} (1-p)^k \right) + E[X] \\
 &= p(1-p) \frac{d^2}{dp^2} \left( \frac{1}{p} \right) + \frac{1}{p} \\
 &= -p(1-p) \frac{d}{dp} \left( \frac{1}{p^2} \right) + \frac{1}{p} \\
 &= \frac{2p(1-p)}{p^3} + \frac{1}{p} \\
 &= \frac{2}{p^2} - \frac{1}{p}
 \end{aligned}$$

Ainsi,

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

■

### 3.3.3. Loi binomiale

#### Définition 3.5 (Loi binomiale)

Soit  $S_n \sim \mathcal{B}(n, p)$ , alors  $S_n \in \mathbb{N}$  avec  $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ . (cf. 2.2.3)

On donne l'espérance et la variance d'une V.A. Binômiale

#### Proposition 3.9

Soit  $S_n \sim \mathcal{B}(n, p)$ , alors

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

*Démonstration.* On a vu que la variable aléatoire  $S_n$  est la somme de  $n$  variable aléatoire de Bernoulli de paramètre  $p$  indépendantes :

$$S_n = \sum_{k=1}^n X_k, \quad X_k \sim \mathcal{B}(p), \text{ indépendantes}$$

Ainsi, par linéarité de l'espérance

$$E[S_n] = \sum_{k=1}^n E[X_k] = np$$

### 3. Espérance, Variance

De plus, comme les  $X_k$  sont indépendantes,

$$\text{Var}\{S_n\} = \sum_{k=1}^n \text{Var}\{X_k\} = np(1-p)$$

■

#### 3.3.4. Loi de Poisson

##### Définition 3.6

##### (Loi de Poisson)

Soit  $\lambda > 0$ . On dit que  $X$  suit une loi de Poisson de paramètre  $\lambda$ , que l'on note  $X \sim \mathcal{P}(\lambda)$ , si  $X$  prend ses valeurs dans  $\mathbb{N}$  et

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

Cette définition donne bien une probabilité sur  $\mathbb{N}$ , étant donné que

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

On donne l'espérance et la variance d'une V.A. de Poisson

##### Proposition 3.10

Soit  $X \sim \mathcal{P}(\lambda)$ , alors

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

*Démonstration.*

$$\begin{aligned} E[X] &= \sum_{k=0}^{+\infty} k P(X = k) = \sum_{k=0}^{+\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda \end{aligned}$$

$$\begin{aligned} E[X^2] &= \sum_{k=0}^{+\infty} k^2 P(X = k) = \sum_{k=0}^{+\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{+\infty} k \frac{\lambda^k}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{k=1}^{+\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \lambda e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda^2 e^{-\lambda} \sum_{k=2}^{+\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \\ &= \lambda^2 + \lambda \end{aligned}$$

### 3.3. Lois classiques et leurs moments

et donc

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

■



Deuxième partie

Probabilités continues



# 4 ■ Probabilités et variables aléatoires continues

## Objectifs:

- Transposer les notions de probabilités au continue
- Savoir manipuler les V.A. continue à densité
- Connaître la loi normale et ses propriétés
- Manipuler les fonctions caractéristiques

## Contents

---

<b>4.1. Axiomatique des probabilités</b> . . . . .	<b>35</b>
<b>4.2. Variable aléatoires continues</b> . . . . .	<b>38</b>
4.2.1. Variable absolument continue . . . . .	39
4.2.2. Indépendance . . . . .	40
4.2.3. Moments . . . . .	41
<b>4.3. Lois continues classiques</b> . . . . .	<b>42</b>
4.3.1. Loi uniforme sur un intervalle $[a,b]$ . . . . .	42
4.3.2. Loi exponentielle de paramètre $\lambda$ . . . . .	43
4.3.3. Loi normale de moyenne $\mu$ et variance $\sigma^2$ . . . . .	44
<b>4.4. Fonction caractéristique</b> . . . . .	<b>44</b>

---

## 4.1. Axiomatique des probabilités

Jusqu'à présent, nous n'avons considéré que des phénomènes aléatoires prenant un nombre fini ou dénombrable de valeurs. Or, ce n'est pas toujours le cas : si l'on prend pour exemple un jeu de fléchettes, celle-ci peut se planter n'importe où sur la cible, qui constitue alors un espace d'issues continu. Ne pouvant pas dénombrer toutes les issues possibles (la cible étant un espace continu), on va s'intéresser aux zones dans lesquels la fléchette peut atterrir. Intuitivement, plus l'aire

#### 4. Probabilités et variables aléatoires continues

de la zone considérée est grande, plus on a de chance que la fléchette se plante dans cette zone.

C'est la notion de "tribu" qui permet de généraliser la notion d'évènements dans un espace continu.

##### Définition 4.1 (Tribu ou $\sigma$ -algèbre – espace probabilisable)

Soit  $\Omega$  un espace d'issues. On appelle tribu d'évènements (ou  $\sigma$ -algèbre), toute famille  $\mathcal{A}$  de sous-ensemble de  $\Omega$  vérifiant

1.  $\Omega \in \mathcal{A}$  et  $\emptyset \in \mathcal{A}$
2. Si  $A \in \mathcal{A}$ ,  $A^c \in \mathcal{A}$  (stabilité par passage au complémentaire)
3. Pour toute famille finie ou dénombrable  $(A_i)_{i \in I}$  d'éléments de  $\mathcal{A}$ ,  $\bigcup_{i \in I} A_i \in \mathcal{A}$  (stabilité par union)

Le couple  $(\Omega, \mathcal{A})$  est appelé espace probabilisable.

Les points 2 et 3 donne directement la proposition suivante

##### Proposition 4.1

Toute intersection dénombrable d'éléments de  $\mathcal{A}$  est encore dans  $\mathcal{A}$

##### Exemple 4.1

1. L'ensemble des parties de  $\Omega$  est une tribu, c'est la plus grande. C'est celle que nous avons toujours considéré de façon implicite dans le cas où  $\Omega$  est fini ou dénombrable.
2.  $\{\emptyset, \Omega\}$  est une tribu, c'est la plus petite.
3. Si  $A \subset \Omega$ ,  $\{\emptyset, A, A^c, \Omega\}$  est une tribu. C'est la plus petite tribu contenant  $A$ . On dit que c'est la tribu engendrée par  $A$ .

##### Définition 4.2 (Probabilité – espace probabilisé)

Soit  $\Omega$  un espace d'issues, et  $\mathcal{A}$  une tribu sur  $\Omega$ . Une probabilité sur  $(\Omega, \mathcal{A})$  est une application  $P : \mathcal{A} \rightarrow [0, 1]$  telle que

1.  $P(\Omega) = 1$
2. Pour toute famille finie ou dénombrable  $(A_i)_{i \in I}$  d'éléments de  $\mathcal{A}$  deux à deux disjoints,

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i)$$

Le triplet  $(\Omega, \mathcal{A}, P)$  est appelé espace probabilisé.

Dans le cas fini ou dénombrable, on prendra toujours  $\mathcal{A} = \mathcal{P}(\Omega)$ , et l'on retrouve toutes les notions vues dans les cours précédents.

On peut vérifier que toutes les propriétés d'une probabilité vues dans le cas fini ou dénombrable restent vraies dans le cadre d'un espace probabilisé général. On peut aussi redéfinir dans ce cadre la notion d'évènements indépendants, de probabilité conditionnelle etc.

**Proposition 4.2 (Rappels de calcul de probabilité)**

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé,  $A, B \in \mathcal{A}$ .

- $P(\emptyset) = 0$
- $P(\Omega) = 1$
- $P(A^c) = 1 - P(A)$
- $B \subset A \Rightarrow P(B) \leq P(A)$
- $B \subset A \Rightarrow P(A - B) = P(A) - P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Continuité croissante : si  $(A_n)_{n \in \mathbb{N}}$  est une suite croissante d'évènements ( $A_n \subset A_{n+1}$ ), alors

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$$

- Continuité décroissante : si  $(A_n)_{n \in \mathbb{N}}$  est une suite décroissante d'évènements ( $A_{n+1} \subset A_n$ ), alors

$$P\left(\bigcap_{n=0}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$$

Tout comme dans le cas discret, on peut définir la notion de probabilité conditionnelle.

**Définition 4.3 (Probabilité conditionnelle)**

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé, et soit  $B \in \mathcal{A}$  tel que  $P(B) > 0$ . La probabilité conditionnelle de  $A$  sachant  $B$  est

$$\forall A \in \mathcal{A} \quad P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}$$

On retrouve toutes les formules déjà vues

**Proposition 4.3 (Loi de Bayes)**

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé, et soient  $A$  et  $B$  deux évènements de  $\mathcal{A}$  tels que  $P(A) > 0$  et  $P(B) > 0$ . Alors

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

#### 4. Probabilités et variables aléatoires continues

##### Proposition 4.4 (Formule des probabilités composées)

Soient  $A_1, A_2, \dots, A_n$  des évènements tels que  $P(A_1 \cap A_2 \cap \dots \cap A_n) > 0$

$$P(A_1 \cap \dots \cap A_n) = P(A_n | A_1 \cap \dots \cap A_{n-1}) P(A_{n-1} | A_1 \cap \dots \cap A_{n-2}) \dots P(A_2 | A_1) P(A_1).$$

##### Proposition 4.5 (Formule des probabilités totales)

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Soit  $\{B_n\}_{n \in \mathbb{N}}$  une suite complète d'évènements (i.e telle que  $B_i \cap B_j = \emptyset$  si  $i \neq j$  et  $\Omega = \bigcup_{n \in \mathbb{N}} B_n$ ). Alors pour tout évènements  $A \in \mathcal{A}$

$$P(A) = \sum_{n \in \mathbb{N}} P(B_n) P(A | B_n)$$

Et par conséquent, en combinant cette propriété avec la loi de Bayes, on a la formule de probabilité des causes (aussi appelé formule de Bayes)

##### Proposition 4.6 (Formule des probabilités des causes)

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Soit  $\{B_n\}_{n \in \mathbb{N}}$  une suite complète d'évènements (i.e telle que  $B_i \cap B_j = \emptyset$  si  $i \neq j$  et  $\Omega = \bigcup_{n \in \mathbb{N}} B_n$ ). Alors pour tout évènements  $A \in \mathcal{A}$ , pour tout  $i \in \mathbb{N}$

$$P(B_i | A) = \frac{P(B_i) P(A | B_i)}{\sum_{n \in \mathbb{N}} P(B_n) P(A | B_n)}$$

Et l'on redéfinit la notion d'indépendance

##### Définition 4.4 (Indépendance)

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé, et  $A, B \in \mathcal{A}$ ,  $A$  et  $B$  sont indépendants relativement à  $P$  (ou  $P$ -indépendants) si et seulement si

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

Par conséquent, si  $A$  et  $B$  sont indépendants relativement à  $P$ , on a  $P(A|B) = P(A)$  et  $P(B|A) = P(B)$ . De même, on dit qu'une suite  $(A_i)_{i \in I}$  est  $P$ -mutuellement indépendantes si pour tout sous-ensemble fini  $J$  de  $I$  on a

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

## 4.2. Variable aléatoires continues

Maintenant que la notion de probabilité est généralisée au cas continu, on peut aussi généraliser la notion de variable aléatoire

**Définition 4.5 (Variable aléatoire)**

Une variable aléatoire  $X$  sur l'espace probablisable  $(\Omega, \mathcal{A})$  est une application de  $\Omega$  dans  $\mathbb{R}$  telle que pour tout  $x \in \mathbb{R}$ ,  $\{X \leq x\} \in \mathcal{A}$ .

Dans le cas discret, on retrouve la même notion de variable aléatoire, puisque  $\mathcal{A} = \mathcal{P}(\Omega)$ . On peut aussi définir la fonction de répartition.

**Définition 4.6 (Fonction de répartition)**

La fonction de répartition  $F_X$  d'une variable aléatoire  $X$  sur l'espace probablisé  $(\Omega, \mathcal{A}, P)$  est définie par

$$F_X : \mathbb{R} \rightarrow [0, 1] \\ x \mapsto P(X \leq x)$$

La proposition suivante caractérise la courbe d'une fonction de répartition

**Proposition 4.7**

Soit  $X$  une variable aléatoire sur l'espace probablisé  $(\Omega, \mathcal{A}, P)$  et  $F_X$  sa fonction de répartition. Alors  $F_X$  est une fonction croissante de  $\mathbb{R}$  dans  $[0, 1]$ , continue à droite et telle que  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  et  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .

On peut noter que  $\{y < X \leq x\} = \{X \leq x\} \cap \{X \leq y\}^c$  est un élément de la tribu  $\mathcal{A}$ . On peut donc parler de  $P(y < X \leq x)$  et par les axiomes de probabilités, on a

$$P(y < X \leq x) = F_X(x) - F_X(y)$$

Et l'on définit la loi de probabilité

**Définition 4.7 (Loi de probabilité)**

Soit  $X$  une variable aléatoire définie sur l'espace probablisé  $(\Omega, \mathcal{A}, P)$ . On appelle loi de probabilité de  $X$  la fonction qui à toute réunion finie ou dénombrable  $B$  d'intervalles, associe le nombre  $P(X \in B)$ .

Pour connaître la loi d'une variable aléatoire  $X$ , il suffit de connaître  $P(X \in I)$  pour tout intervalle  $I$ . Ces probabilités peuvent s'exprimer en fonction de la fonction de répartition. Ainsi, on obtient la proposition suivante

**Proposition 4.8**

La fonction de répartition de  $X$  détermine la loi de  $X$ .

**4.2.1. Variable absolument continue****Définition 4.8**

Soit  $X$  une variable aléatoire sur l'espace probablisé  $(\Omega, \mathcal{A}, P)$ .  $X$  est dite absolument continue ssi il existe une fonction  $f_X$  continue (sauf, éventuellement, en

#### 4. Probabilités et variables aléatoires continues

un nombre dénombrable de points), telle que pour tout  $x \in \mathbb{R}$

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

On dit que  $X$  est de densité  $f_X$ .

##### Remarque 4.1

Il existe des variables aléatoires qui ne sont ni discrètes, ni à densité.

##### Définition 4.9 (Densité de probabilité)

Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$ .  $f$  est une densité de probabilité si c'est une fonction positive d'intégrale 1. C'est-à-dire

$$\forall x \in \mathbb{R} f(x) \geq 0 \quad \text{et} \quad \int_{\mathbb{R}} f(t) dt = 1$$

La densité de probabilité permet de caractériser la mesure de probabilité

##### Proposition 4.9

. Soit  $X$  une variable aléatoire de densité  $f_X$ . Alors

1. pour tout  $x \in \mathbb{R} P(X = x) = 0$
2. pour tout  $x, y \in \mathbb{R}$  tels que  $x \leq y$ ,  $P(X \in (x, y)) = \int_x^y f_X(t) dt$
3.  $f_X(x) = F'_X(x)$

Ainsi, quel que soit le type d'intervalle considéré (ouvert, fermé, semi-ouvert...) on a

$$P(X \in (x, y)) = F_X(y) - F_X(x) = \int_x^y f_X(t) dt .$$

#### 4.2.2. Indépendance

##### Définition 4.10 (Variables aléatoires indépendantes)

Soient  $X$  et  $Y$  deux variables aléatoires sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ .  $X$  et  $Y$  sont indépendantes ssi pour tout  $B$  et  $C$  réunions dénombrables d'intervalles

$$P(\{X \in B\} \cap \{Y \in C\}) = P(X \in B)P(Y \in C)$$

Comme pour le cas discret, on a la proposition suivante

##### Proposition 4.10

Soient  $X$  et  $Y$  deux variables aléatoires indépendantes, et soient  $f$  et  $g$  deux fonctions continues. Alors les variables  $f(X)$  et  $f(Y)$  sont indépendantes.

## 4.2.3. Moments

**Définition 4.11 (Espérance)**

Soit  $X$  une variable aléatoire de densité  $f_X$ . Soit  $h$  une fonction continue. On définit l'espérance de  $h(X)$  par

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

dès que cette intégrale est bien définie.

Ainsi, l'espérance de  $X$  est

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

**Définition 4.12 (Moment d'ordre  $n$ )**

Soit  $X$  une variable aléatoire de densité  $f_X$ . Le moment d'ordre  $n$  de  $X$  est donné par

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

Les propriétés de l'espérance vue dans le cas discret restent vrai dans le cas continue, on en rappelle quelques unes

**Proposition 4.11 (Propriétés de l'espérance)**

. Soit  $X$  une variable aléatoire admettant une espérance, sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ .

1. Soit  $\lambda \in \mathbb{R}$ ,  $E[\lambda X] = \lambda E[X]$
2.  $E[X + Y] = E[X] + E[Y]$
3. Si  $X \geq 0$ ,  $E[X] \geq 0$

De même qu'en discret, on définit la variance d'une variable aléatoire

**Définition 4.13 (Variance)**

Soit  $X$  une variable aléatoire sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ . La variance de  $X$ , si elle existe, est définie par

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

Avec

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

**Définition 4.14 (Espaces  $L^n$ )**

Soit  $X$  une variable aléatoire à densité. Si le moment d'ordre  $n$  de  $X$  existe, on dit que  $X \in L^n$ . En particulier, si  $X \in L^1$  alors l'espérance de  $X$  existe, et si  $X \in L^2$  alors sont moment d'ordre 2, et donc sa variance, existe.

### 4.3. Lois continues classiques

#### 4.3.1. Loi uniforme sur un intervalle $[a, b]$

##### 4.3.1.1. Définition et propriétés

Soit  $a, b \in \mathbb{R}$ , avec  $a < b$ . La variable aléatoire  $X$  suit une loi uniforme sur l'intervalle  $[a, b]$ , notée  $X \sim \mathcal{U}([a, b])$ , si elle a pour densité de probabilité

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$$

Elle admet pour espérance

$$E[X] = \frac{b+a}{2}$$

et pour variance

$$\text{Var}\{X\} = \frac{(b-a)^2}{12}$$

##### 4.3.1.2. Preuves

On vérifie qu'on a bien une densité de probabilité :  $f(x) \geq 0 \forall x$  et

$$\int_{\mathbb{R}} f_X(x) dx = \int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} [x]_a^b = 1$$

Calcul de la moyenne

$$\begin{aligned} E[x] &= \int_{\mathbb{R}} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[ \frac{1}{2} x^2 \right]_a^b \\ &= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \end{aligned}$$

Calcul du moment d'ordre deux

$$\begin{aligned} E[x] &= \int_{\mathbb{R}} x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \left[ \frac{1}{3} x^3 \right]_a^b \\ &= \frac{b^3 - a^3}{2(b-a)} = \frac{(b-a)(a^2 + b^2 + ab)}{3(b-a)} \\ &= \frac{a^2 + b^2 + ab}{3} \end{aligned}$$

donc

$$\begin{aligned}\text{Var}\{x\} &= E[x^2] - E[x]^2 = \frac{a^2 + b^2 + ab}{3} - \frac{(a+b)^2}{4} \\ &= \frac{a^2 + b^2 - 2ab}{12} = \frac{(b-a)^2}{12}\end{aligned}$$

### 4.3.2. Loi exponentielle de paramètre $\lambda$

#### 4.3.2.1. Définition et propriétés

Soit  $\theta > 0$ . La variable aléatoire  $X$  suit une loi exponentielle de paramètre  $\lambda$ , notée  $X \sim \mathcal{E}(\lambda)$ , si elle a pour densité de probabilité

$$f_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \lambda e^{-\lambda x} & \text{si } x \geq 0 \end{cases}$$

Elle admet pour espérance

$$E[X] = \frac{1}{\lambda}$$

et pour variance

$$\text{Var}\{X\} = \frac{1}{\lambda^2}$$

#### 4.3.2.2. Preuves

On vérifie qu'on a bien une densité de probabilité.  $f(x) \geq 0 \forall x$  et

$$\int_{\mathbb{R}} f_X(x) dx = \int_0^{+\infty} e^{-\lambda x} dx = \lambda \left[ -\frac{1}{\lambda} e^{-\lambda x} \right]_0^{+\infty} = 1$$

Calcul de la moyenne

$$\begin{aligned}E[x] &= \int_{\mathbb{R}} x f(x) dx = \int_0^{+\infty} x \lambda e^{-\lambda x} dx \\ &= \lambda \left( \left[ -\frac{x e^{-\lambda x}}{\lambda} \right]_0^{+\infty} + \int_0^{+\infty} \frac{e^{-\lambda x}}{\lambda} dx \right) = \int_0^{+\infty} e^{-\lambda x} dx \\ &= \left[ -\frac{e^{-\lambda x}}{\lambda} \right]_0^{+\infty} = \frac{1}{\lambda}\end{aligned}$$

Calcul du moment d'ordre 2 : faire une double IPP.

### 4.3.3. Loi normale de moyenne $\mu$ et variance $\sigma^2$

#### 4.3.3.1. Définition et propriétés

Soit  $\mu \in \mathbb{R}$  et  $\sigma > 0$ . La variable aléatoire  $X$  suit une loi normale de moyenne  $\mu$  et de variance  $\sigma^2$ , notée  $X \sim \mathcal{N}(\mu, \sigma^2)$ , si elle a pour densité de probabilité

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

Elle admet pour espérance

$$E[X] = \mu$$

et pour variance

$$\text{Var}\{X\} = \sigma^2$$

## 4.4. Fonction caractéristique

La fonction caractéristique d'une variable aléatoire à densité est la transformée de Fourier de sa densité. C'est un outil pratique pour caractériser les lois.

#### Définition 4.15 (Fonction caractéristique)

Soit  $X$  une variable aléatoire sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ . La fonction caractéristique de  $X$  est la quantité

$$\varphi_X : \mathbb{R} \rightarrow \mathbb{C} \tag{4.1}$$

$$t \mapsto \varphi_X(t) = E[e^{itX}] \tag{4.2}$$

- Si  $X$  est une variable aléatoire discrète, on a

$$\varphi_X(t) = \sum_{k \in \mathbb{N}} P(X = k) e^{itk}$$

- Si  $X$  admet une densité  $f_X$ , alors pour tout  $t \in \mathbb{R}$

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

La fonction caractéristique caractérise (!) parfaitement la loi d'une variable aléatoire :

#### Proposition 4.12

Soit  $X$  et  $Y$  deux variables aléatoires sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ , de loi respective  $P_X$  et  $P_Y$ , alors

$$P_X = P_Y \Leftrightarrow \varphi_X = \varphi_Y$$

**Proposition 4.13 (Propriétés générales)**

Soit  $X$  une variable aléatoire sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ , de loi  $P$  (discrete ou continue).

1.  $\varphi_X(0) = 1$
2. Une fonction caractéristique est uniformément continue et bornée

$$\forall t \in \mathbb{R} |\varphi_X(t)| \leq 1$$

3. Une fonction caractéristique possède la symétrie hermitienne

$$\forall t \in \mathbb{R} \varphi_X(-t) = \overline{\varphi_X(t)}$$

4.  $P_X = P_{-X} \Leftrightarrow \varphi_X$  est paire  $\Leftrightarrow \varphi_X$  est réelle
5. Soit  $a, b \in \mathbb{R}$ , alors  $\forall t \in \mathbb{R} \varphi_{aX+b}(t) = e^{ibt} \varphi_X(at)$

*Démonstration.* voir le cours de traitement du signal sur les transformées de Fourier! ■

On a bien entendu le théorème d'inversion

**Théorème 4.1 (Inversion de la fonction caractéristique)**

Soit  $X$  une variable aléatoire sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Si  $\varphi_X \in L^1(\mathbb{R})$ , alors  $X$  admet une densité  $f_X$  et

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(t) e^{-itx} dt \quad p.p.$$

La fonction caractéristique permet de calculer les moments d'une variable aléatoire.

**Théorème 4.2**

Soit  $X$  une variable aléatoire sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Si  $X$ , de densité  $f_X$ , admet des moments jusqu'à l'ordre  $p \geq 1$ , alors  $\varphi_X$  est de classe  $C^p$  au moins et

$$\varphi_X^{(k)}(t) = \int_{\mathbb{R}} (ix)^k e^{itx} f_X(x) dx$$

De plus, on a

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}[X^k].$$

En particulier, on notera la moyenne  $\mathbb{E}[X] = -i\varphi_X'(0)$  et le moment d'ordre 2  $\mathbb{E}[X^2] = \varphi_X''(0)$ . On obtient de plus un développement limité de  $\varphi$

$$\varphi_X(t) = 1 + itm_1 + \frac{(it)^2}{2} m_2 + \dots + \frac{(it)^p}{p!} m_p + o(|t|^p)$$

où  $m_k$  désigne le moment d'ordre  $k$  :  $m_k = \mathbb{E}[X^k]$

#### 4. Probabilités et variables aléatoires continues

Une des intérêts de la fonction caractéristique, est qu'elle permet de calculer la loi d'une somme de variables aléatoires.

##### **Théorème 4.3 (Fonction caractéristique d'une somme de variable aléatoire)**

Soient  $X_1, X_2, \dots, X_n$  des variables aléatoires indépendantes sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Soit  $S_n = X_1 + X_2 + \dots + X_n$ , la variable aléatoire définie comme la somme des  $n$  variables aléatoires  $X_i$ . Alors

$$\varphi_{S_n}(t) = \prod_{k=1}^n \varphi_{X_k}(t)$$

Ce théorème est valable aussi bien dans le cas continue que dans le cas discret. Pour le montrer, il faut admettre le résultat suivant

##### **Théorème 4.4 (Loi d'une somme de deux variables aléatoires indépendantes)**

Soit  $X$  et  $Y$  deux variables aléatoires indépendantes sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ .

1. Si  $X$  et  $Y$  sont discrètes de loi  $P_X = (p_k = P(X = k))_{k \in \mathbb{N}}$  et  $P_Y = (q_k = P(Y = k))_{k \in \mathbb{N}}$ . Alors

$$P_{X+Y}(X + Y = n) = \sum_{k \in \mathbb{N}} p_k q_{n-k}$$

ie

$$P_{X+Y} = P_X \star P_Y$$

2. Si  $X$  et  $Y$  admettent pour densité respectives  $f_X$  et  $f_Y$ , alors

$$f_{X+Y}(t) = \int_{\mathbb{R}} f_X(x) f_Y(t-x) dx$$

ie

$$f_{X+Y} = f_X \star f_Y$$

# 5 ■ Vecteurs de variables aléatoires

**Objectifs:**

- Transposer les notions de probabilités au continue
- Savoir manipuler les V.A. continue à densité
- Connaître la loi normale et ses propriétés
- Manipuler les fonctions caractéristiques

**Contents**

---

<b>5.1. Couple de variables aléatoires</b>	<b>47</b>
5.1.1. Loi jointe, lois marginales	47
5.1.2. Variables aléatoires indépendantes	48
5.1.3. Moments d'un couple	49
5.1.4. Fonction caractéristique d'un couple	51
<b>5.2. Vecteur de variables aléatoires</b>	<b>52</b>

---

## 5.1. Couple de variables aléatoires

### 5.1.1. Loi jointe, lois marginales

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Dans le cas discret, on sait manipuler les couples, voire les n-uplets, de variables aléatoires. En effet, on appelle loi du couple  $(X, Y)$ , ou loi jointe, la probabilité  $P_{(X,Y)}(x, y) = P[(X, Y) = (x, y)] = P(X = x, Y = y)$ .

## 5. Vecteurs de variables aléatoires

### Définition 5.1 (Loi d'un couple de V.A. continues)

Dans le cas continu, la loi du couple de variable aléatoire  $X, Y$  est donnée par la fonction de répartition jointe

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

et, si les variables sont absolument continues, par la densité de probabilité jointe

$$P((X, Y) \in (x_1, x_2) \times (y_1, y_2)) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{(X,Y)}(x, y) dx dy$$

avec évidemment

$$f_{(X,Y)}(x, y) \geq 0 \quad \forall x, y \quad \int_{\mathbb{R}} \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx dy = 1$$

Connaissant la loi du couple de variables aléatoires  $(X, Y)$ , on peut retrouver les loi de probabilité individuelles  $X$  et  $Y$ . On appelle ces lois, les lois marginales

### Définition 5.2 (lois marginales discrètes)

Soit  $(X, Y)$  un couple de variables aléatoires discrètes. On appelle loi marginale de  $X$  (resp.  $Y$ ) la loi donnée par

$$P_X(X = x) = \sum_{y \in \mathbb{N}} P(X = x, Y = y) \quad (\text{resp. } P_Y(Y = y) = \sum_{x \in \mathbb{N}} P(X = x, Y = y))$$

En appliquant le théorème de Bayes, on retrouve la formule bien connue

$$P_X(X = x) = \sum_{y \in \mathbb{N}} P(X = x | Y = y) P(Y = y) .$$

On généralise cette définition aux lois continues

### Définition 5.3 (lois marginales continues)

Soit  $(X, Y)$  un couple de variables aléatoires continues à densité  $f_{(X,Y)}$ . On appelle densité marginale de  $X$  (resp.  $Y$ ) la densité de probabilité donnée par

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

La réciproque est cependant fautive : on ne peut pas, en général, retrouver la loi du couple à partir des lois marginales!

## 5.1.2. Variables aléatoires indépendantes

On généralise maintenant les notions d'indépendance et de probabilité conditionnelles au variable continues à densité.

**Définition 5.4 (variable aléatoire indépendantes à densité)**

Soit  $X$  et  $Y$  deux variables aléatoires de densité  $f_X$  et  $f_Y$ .  $X$  et  $Y$  sont indépendantes ssi

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$$

**Définition 5.5 (Loi de  $X$  sachant  $Y$ )**

Soit  $X$  et  $Y$  deux variables aléatoires de densité  $f_X$  et  $f_Y$ . On définit la nouvelle variable aléatoire  $X|Y$  ( $X$  sachant  $Y$ ) de densité

$$f_{X|Y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}$$

On retrouve la loi de Bayes :

$$f_{X|Y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)} = \frac{f_{Y|X}(y)f_X(x)}{f_Y(y)}$$

Ainsi, si  $X$  et  $Y$  sont indépendantes, on retrouve bien

$$f_{X|Y}(x) = f_X(x) \quad \text{et} \quad f_{Y|X}(y) = f_Y(y)$$

**5.1.3. Moments d'un couple**

On étend la définition d'espérance et de variance d'un couple, et l'on introduit une nouvelle grandeur : la covariance.

**Définition 5.6 (Espérance d'un couple)**

Soit  $X, Y$  deux variables aléatoires. L'espérance du couple  $(X, Y)$ , notée  $E[(X, Y)]$  est le couple  $(E[X], E[Y])$ . De plus, pour toute fonction réelle  $h$  du couple, l'espérance, si elle existe, est donnée par

- Si les variables aléatoires sont discrètes

$$E[h(X, Y)] = \sum_{x \in \mathbb{N}, y \in \mathbb{N}} h(x, y) P_{X,Y}(X = x, Y = y)$$

- Si le couple admet une densité  $f_{X,Y}$

$$E[h(X, Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) f_{X,Y}(x, y) dx dy$$

En particulier, on peut calculer  $E[XY]$

- Si les variables aléatoires sont discrètes, on a

$$E[XY] = \sum_{x \in \mathbb{N}, y \in \mathbb{N}} xy P(X = x, Y = y)$$

## 5. Vecteurs de variables aléatoires

- Si le couple admet une densité  $f_{X,Y}$ , on a

$$E[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xy f(x, y) dx dy$$

L'espérance de la variable aléatoire  $X$  se calcule aussi bien à l'aide de la loi du couple que de la loi de variable. Il suffit pour cela de revenir à la définition des loi marginales. Dans le cas discret

$$\begin{aligned} E[h(X)] &= \sum_{x \in \mathbb{N}} h(x) P_X(X = x) \\ &= \sum_{x, y \in \mathbb{N}} h(x) P_{X,Y}(X = x, Y = y) \end{aligned}$$

et dans le cas de variable à densité

$$\begin{aligned} E[h(X)] &= \int_{\mathbb{R}} h(x) f_X(x) dx \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x) f_{X,Y}(x, y) dx dy \end{aligned}$$

On retrouve les propriétés déjà vues dans le cas discret

### Proposition 5.1 (Espérance du produit de deux variables aléatoires indépendantes)

Soit  $X, Y$  deux variables aléatoires indépendantes (discrètes ou continues). Alors

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

La notion de variance d'une variable aléatoire à été vue dans le cours précédent. Lorsqu'on considère un couple de variable aléatoire, il faut introduire la notion de covariance, afin de quantifier l'écart conjoint par rapport à leur espérance.

### Définition 5.7 (Covariance)

Soit  $X$  et  $Y$  deux variables aléatoires (discrètes ou continues), d'espérance respective  $E[X]$ ,  $E[Y]$  et de variance respective  $\text{Var}\{X\}$ ,  $\text{Var}\{Y\}$ . La covariance du couple est la quantité

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

Cette quantité est symétrique  $\text{Cov}[X, Y] = \text{Cov}[Y, X]$ , et permet de définir la matrice de covariance (ou de dispersion)

$$\Gamma_{(X,Y)} = \begin{bmatrix} \text{Var}\{X\} & \text{Cov}[X, Y] \\ \text{Cov}[Y, X] & \text{Var}\{Y\} \end{bmatrix}$$

**Proposition 5.2 (Covariance de deux variables aléatoires indépendantes)**

Soit  $X, Y$  deux variables aléatoires indépendantes. Alors

$$\text{Cov}[X, Y] = 0$$

*Démonstration.* Il suffit de revenir à la définition et d'utiliser le fait que  $E[XY] = E[X]E[Y]$  car les variables aléatoires sont indépendantes. ■

La réciproque est bien entendue fautive en générale!

Si l'on rapporte la covariance au produit des variances des deux variables aléatoire, on trouve un coefficient compris entre  $-1$  et  $1$ , appelé coefficient de corrélation linéaire (ou de Bravais-Pearson)

**Définition 5.8 (Coefficient de corrélation linéaire)**

Soit  $X$  et  $Y$  deux variables aléatoires (discrètes ou continues), d'espérance respective  $E[X]$ ,  $E[Y]$  et de variance respective  $\text{Var}\{X\}$ ,  $\text{Var}\{Y\}$ . Le coefficient de corrélation linéaire est la quantité

$$\rho = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}\{X\}\text{Var}\{Y\}}}$$

On a  $-1 \leq \rho \leq 1$

Comme son nom l'indique, cette quantité mesure à quel point on peut établir une relation linéaire entre les deux variables  $X$  et  $Y$ . Si  $\rho = 0$  alors les variables ne sont pas corrélées **linéairement**, mais ne sont pas indépendantes pour autant. Si  $|\rho| \approx 1$ , alors les variables sont très fortement corrélées linéairement, ie  $X \approx aY + b$ ,  $a, b \in \mathbb{R}$ , avec  $a > 0$  si  $\rho \approx 1$  et  $a < 0$  si  $\rho \approx -1$

**5.1.4. Fonction caractéristique d'un couple**

On peut étendre la notion de fonction de caractéristique à un couple de variable aléatoire.

**Définition 5.9 (Fonction caractéristique d'un couple)**

Soit  $(X, Y)$  un couple de variables aléatoire sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ . La fonction caractéristique du couple est donnée par

$$\varphi_{(X, Y)}(a, b) = E \left[ e^{i(ax+by)} \right]$$

En particulier, si  $(X, Y)$  admet pour densité jointe la fonction  $f_{(X, Y)}$ , alors

$$\varphi_{(X, Y)}(a, b) = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{i(ax+by)} f_{(X, Y)}(x, y) dx dy$$

On peut alors caractériser l'indépendance des deux variables à l'aide de la fonction caractéristique

## 5. Vecteurs de variables aléatoires

### Proposition 5.3

Soit  $(X, Y)$  un couple de variables aléatoires sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ .  
 $X$  et  $Y$  sont indépendantes ssi

$$\varphi_{(X,Y)}(a, b) = \varphi_X(a)\varphi_Y(b)$$

## 5.2. Vecteur de variables aléatoires

Les notions de loi jointe, d'espérance et de matrice de covariance se généralisent directement au cas où l'on a un vecteur de variable aléatoire  $(X_1, X_2, \dots, X_n)$

### Définition 5.10 (densité)

une fonction  $f_{(X_1, \dots, X_n)} : \mathbb{R}^n \rightarrow \mathbb{R}$  est une densité si

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \geq 0 \forall (x_1, \dots, x_n)$$

et

$$\int_{\mathbb{R}^n} f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) dx_1 \dots dx_n = 1$$

### Définition 5.11 (Espérance d'un vecteur de variables aléatoires)

Soit  $(X_1, \dots, X_n)$  un vecteur de variables aléatoires admettant une densité  $f_{(X_1, \dots, X_n)}$ , tel que pour tout  $i \in \{1, \dots, n\}$ ,  $X_i \in L^1(\mathbb{R})$ . Alors

$$E[(X_1, \dots, X_n)] = (E[X_1], \dots, E[X_n])$$

**Définition 5.12 (Matrice de covariance d'un vecteur de variables aléatoires)**

Soit  $(X_1, \dots, X_n)$  un vecteur de variables aléatoires admettant une densité  $f_{(X_1, \dots, X_n)}$ , tel que pour tout  $i \in \{1, \dots, n\}$ ,  $X_i \in L^1$  et  $X_i \in L^2$ . Alors on peut définir la matrice de covariance  $\Sigma = (\sigma_{i,j})$  tel que

- $\Sigma$  est une matrice symétrique, définie positive
- Avec pour éléments diagonaux  $\sigma_{i,i} = \text{Var}\{X_i\}$
- Pour éléments en dehors de la diagonale  $\sigma_{i,j} = \sigma_{j,i} = \text{Cov}[X_i, X_j]$

On étend aussi directement la définition de fonction caractéristique et ses propriétés.

**Définition 5.13 (Fonction caractéristique d'un vecteur de variables aléatoires)**

Soit  $\underline{X} = (X_1, \dots, X_n)$  un vecteur de  $n$  variables aléatoires sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ . La fonction caractéristique du vecteur est la fonction

$$\varphi_{\underline{X}} : \mathbb{R}^n \rightarrow \mathbb{C} \tag{5.1}$$

$$\underline{u} \mapsto \underline{X}(\underline{u}) = \mathbb{E} \left[ e^{i\langle \underline{u}, \underline{X} \rangle} \right] \tag{5.2}$$

En particulier, si  $\underline{X}$  admet pour densité jointe la fonction  $f_{\underline{X}}$ , alors

$$\varphi_{\underline{X}}(\underline{u}) = \int_{\mathbb{R}^n} e^{i\langle \underline{u}, \underline{x} \rangle} f_{\underline{X}}(\underline{x}) \, d\underline{x}$$

**Proposition 5.4**

Soit  $\underline{X} = (X_1, \dots, X_n)$  un vecteur de  $n$  variables aléatoires sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  et  $\varphi_{\underline{X}}$  sa fonction caractéristique. Alors

1.  $\varphi_{\underline{X}}(\underline{0}) = 1$
2.  $\|\varphi_{\underline{X}}\| \leq 1$
3.  $\varphi_{\underline{X}}$  est continue.

**Théorème 5.1 (Caractérisation des moments)**

Soit  $\underline{X} = (X_1, \dots, X_n)$  un vecteur de  $n$  variables aléatoires sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ , tel que  $\underline{X}$  admettent  $m$  moments. Alors sa fonction caractéristique  $\varphi_{\underline{X}}$  est de classe  $C^m$  (au moins). De plus, pour tout choix d'indices  $i_1, \dots, i_m$

$$\frac{\partial^m}{\partial u_{i_1} \partial u_{i_2} \dots \partial u_{i_m}} \varphi_{\underline{X}}(\underline{u}) = i^m \mathbb{E} \left[ e^{i\langle \underline{u}, \underline{X} \rangle} X_{i_1} X_{i_2} \dots X_{i_m} \right]$$

## 5. Vecteurs de variables aléatoires

### **Théorème 5.2 (Transformation affine)**

Soit  $\underline{X} = (X_1, \dots, X_n)$  un vecteur de  $n$  variables aléatoires sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  et  $\varphi_{\underline{X}}$  sa fonction caractéristique. Soit  $a \in \mathbb{R}^m$  et  $A \in \mathbb{R}^{m \times n}$ . Alors

$$\varphi_{A\underline{X}+a}(\underline{u}) = e^{i\langle a, \underline{u} \rangle} \varphi_{\underline{X}}(A^T \underline{u})$$

**Théorème 5.3 (Variables aléatoires indépendantes)**

Soit  $\underline{X} = (X_1, \dots, X_n)$  un vecteur de  $n$  variables aléatoires sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  et  $\varphi_{\underline{X}}$  sa fonction caractéristique. Les éléments  $X_i$  sont indépendants ssi

$$\varphi_{\underline{X}}(u_1, \dots, u_n) = \prod_{i=1}^n \varphi_{X_i}(u_i)$$



# 6

## ■ Suite de variables aléatoires

### Objectifs:

- Connaître les inégalités classiques
- Connaître les différents modes de convergences
- Connaître les grands théorèmes des probabilités (Lois des grands nombres et de la limite centrale)

### Contents

---

<b>6.1. Inégalités</b> . . . . .	<b>57</b>
<b>6.2. Modes de convergence</b> . . . . .	<b>58</b>
6.2.1. Limite d'une suite de variable aléatoire . . . . .	58
6.2.2. Convergence en loi . . . . .	59
<b>6.3. Loi des grands nombres</b> . . . . .	<b>60</b>
<b>6.4. Théorème de la limite centrale</b> . . . . .	<b>61</b>

---

## 6.1. Inégalités

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et soit  $X : \Omega \rightarrow \mathbb{R}$  et  $Y : \Omega \rightarrow \mathbb{R}$  deux variables aléatoires.

### Proposition 6.1

si  $E[X^2] < +\infty$ , alors

$$|E[X]| \leq E[|X|] \leq \sqrt{E[X^2]}$$

*Démonstration.* La première inégalité n'est rien d'autre que l'inégalité triangulaire.

Pour la seconde inégalité, on a  $E[(|X| - E[|X|])^2] = E[|X|^2] - E[|X|]^2 \geq 0$ . ■

## 6. Suite de variables aléatoires

### Proposition 6.2 (Cauchy-Schwarz)

Si  $E[X^2] < +\infty$  et  $E[Y^2] < +\infty$ , alors

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

*Démonstration.* C'est exactement l'inégalité de Cauchy-Schwarz, qu'on réécrit avec les espérances. ■

### Proposition 6.3 (Inégalité de Markov)

Si  $E[|X|] < +\infty$  alors

$$\forall a > 0 \quad P(|X| > a) \leq \frac{E[|X|]}{a}$$

*Démonstration.* On a  $|X| \geq a \mathbb{1}_{|X| \geq a}$  et donc  $E[|X|] \geq E[a \mathbb{1}_{|X| \geq a}] = aE[\mathbb{1}_{|X| \geq a}] = aP(|X| > a)$  ■

### Proposition 6.4 (Inégalité de Bienaymé-Tchebychev)

Si  $E[X^2] < +\infty$  alors

$$\forall a > 0 \quad P((X - E[X])^2 > a) \leq \frac{\text{Var}\{X\}}{a^2}$$

*Démonstration.* On applique l'inégalité de Markov à la variable aléatoire  $(X - E[X])^2$  pour un réel  $a^2$ . ■

## 6.2. Modes de convergence

### 6.2.1. Limite d'une suite de variable aléatoire

On considère une suite  $(X_n)_{n \in \mathbb{N}}$  de variables aléatoires sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On s'intéresse ici aux différents mode de convergence de la suite de variables aléatoires  $(X_n)_{n \in \mathbb{N}}$  vers une variable aléatoire limite  $X$ .

#### Définition 6.1 (Convergence presque sûre (p.s.))

On dit que la suite  $(X_n)_{n \in \mathbb{N}}$  converge presque sûrement vers  $X$  ssi

$$P\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1$$

On notera

$$X_n \xrightarrow[p.s.]{} X$$

La convergence presque sûre est l'analogue de la convergence "presque partout" en analyse.

**Définition 6.2 (Convergence en moyenne ou dans  $L^1$ )**

Si  $X_n \in L^1$  et  $X \in L^1$ , on dit que la suite  $(X_n)_{n \in \mathbb{N}}$  converge en moyenne vers  $X$  ssi

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

On notera

$$X_n \xrightarrow[L^1]{} X$$

**Définition 6.3 (Convergence en moyenne quadratique ou dans  $L^2$ )**

Si  $X_n \in L^2$  et  $X \in L^2$ , on dit que la suite  $(X_n)_{n \in \mathbb{N}}$  converge en moyenne quadratique vers  $X$  ssi

$$\lim_{n \rightarrow +\infty} E[|X_n - X|^2] = 0$$

On notera

$$X_n \xrightarrow[L^2]{} X$$

**Définition 6.4 (Convergence en probabilité)**

On dit que la suite  $(X_n)_{n \in \mathbb{N}}$  converge en probabilité vers  $X$  ssi

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

On notera

$$X_n \xrightarrow[P]{} X$$

On peut aussi manipuler les images d'une suite de variable aléatoire par une fonction continue.

**Théorème 6.1**

Soit  $h : \mathbb{R} \rightarrow \mathbb{R}$  une fonction continue. Alors

- $X_n \xrightarrow[P]{} X \Rightarrow h(X_n) \xrightarrow[P]{} h(X)$
- $X_n \xrightarrow[p.s.]{} X \Rightarrow h(X_n) \xrightarrow[p.s.]{} h(X)$

**6.2.2. Convergence en loi**

Ce mode de convergence est différent en nature des modes de convergence précédents : on ne regarde plus les variables aléatoires elles-mêmes, mais leurs lois de probabilité. Ainsi, la suite  $(X_n)_{n \in \mathbb{N}}$  de variables aléatoires et la limite  $X$ , toujours à valeur dans  $\mathbb{R}$ , peuvent être définis sur des espaces différents.

## 6. Suite de variables aléatoires

### Définition 6.5 (Convergence en loi)

La suite  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers  $X$  ssi

$$\lim_{n \rightarrow \infty} P_{X_n} = P_X$$

Ce qui signifie que pour toute fonction continue bornée  $h$  sur  $\mathbb{R}$ , on a

$$\lim_{n \rightarrow \infty} E[h(X_n)] = E[h(X)].$$

On notera

$$X_n \xrightarrow{\mathcal{L}} X$$

On utilisera en pratique le théorème suivant

### Théorème 6.2

Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoire admettant pour fonction de répartition  $F_{X_n}$ .  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers  $X$  de fonction de répartition  $F_X$  ssi

$$\lim_{n \rightarrow \infty} F_{X_n} = F$$

Ou bien le théorème de continuité de Levy

### Théorème 6.3 (Levy)

La suite  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers  $X$  ssi la suite des fonctions caractéristiques  $\varphi_{X_n}$  converge en chaque point vers la fonction caractéristique de  $X$ .

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \forall t \in \mathbb{R} \varphi_{X_n}(t) \rightarrow \varphi_X(t)$$

On peut établir une hiérarchie des différentes convergences

$$X_n \xrightarrow{L^2} X \Rightarrow X_n \xrightarrow{L^1} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X$$

et

$$X_n \xrightarrow{p.s.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X$$

La convergence en loi est la plus faible des convergences envisagées.

## 6.3. Loi des grands nombres

La loi des grands nombres permet de faire le lien "naturel" entre les probabilités et les statistiques sur un grand nombre d'expérience (comme calculer de manière empirique une moyenne sur un grand nombre de jeté de pièce ou de dé). On se place sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$

**Théorème 6.4 (Loi faible des grands nombres)**

Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes et identiquement distribuées, de carré intégrables. Soit  $S_n = X_1 + X_2 + \dots + X_n$  et  $\bar{X}_n = \frac{S_n}{n}$  la moyenne empirique sur  $n$  variable aléatoires. Alors la suite  $(\bar{X}_n)_{n \in \mathbb{N}}$  converge dans  $L^2$  (et donc aussi dans  $L^1$ ) vers la constante  $m = E[X_n]$  qui est l'espérance commune des  $X_n$ .

$$\forall \varepsilon > 0 \lim_{n \rightarrow +\infty} P\left(\left|\bar{X}_n - m\right| \leq \varepsilon\right) = 1$$

*Démonstration.* On a

$$\begin{aligned} E\left[(\bar{X}_n - m)^2\right] &= E\left[\left(\frac{1}{n} \sum_i (X_i - m)\right)^2\right] = E\left[\frac{1}{n^2} \sum_i \sum_j (X_i - m)(X_j - m)\right] \\ &= \frac{1}{n^2} \sum_i \sum_j E[(X_i - m)(X_j - m)] \\ &= \frac{1}{n^2} \sum_i E[(X_i - m)^2] \text{ car les } X_i \text{ sont indépendantes} \\ &= \frac{\text{Var}\{X_i\}}{n} \rightarrow 0 \end{aligned}$$

■

**Théorème 6.5 (Loi forte des grands nombres)**

Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes et identiquement distribuées et intégrables. Soit  $S_n = X_1 + X_2 + \dots + X_n$  et  $\bar{X}_n = \frac{S_n}{n}$  la moyenne empirique sur  $n$  variable aléatoires. Alors la suite  $(\bar{X}_n)_{n \in \mathbb{N}}$  converge presque sûrement vers la constante  $m = E[X_n]$  qui est l'espérance commune des  $X_n$ .

$$P\left(\lim_{n \rightarrow +\infty} \bar{X}_n = m\right) = 1$$

## 6.4. Théorème de la limite centrale

Ce théorème est une justification de l'utilité de la loi normale. En effet, cette loi apparaît comme la limite de la somme de variables aléatoires indépendantes de carré intégrable, dont la variance est petite devant la somme.

**Théorème 6.6 (De Moivre-Laplace)**

Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes de Bernoulli :  $\forall i,$

6. Suite de variables aléatoires

$X_i \sim \mathcal{B}(p)$ . Soit  $S_n = X_1 + X_2 + \dots + X_n$ . Alors, pour tout  $-\infty \leq a < b \leq +\infty$

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \frac{1}{2\pi} \int_a^b e^{-x^2/2} dx$$

Ce résultat historique est généralisé par le théorème suivant

**Théorème 6.7 (Limite centrale)**

Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires i.i.d. de carré intégrable, d'espérance commune  $m$  et de variance commune  $\sigma^2$ . Soit  $S_n = X_1 + X_2 + \dots + X_n$ . Alors

$$\begin{aligned} \frac{S_n - nm}{\sqrt{n}} &\xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2) \\ \frac{S_n - nm}{\sqrt{n\sigma^2}} &\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \end{aligned}$$

*Démonstration.* Soit  $Z_n = \frac{\sqrt{(n)}}{\sigma} (\frac{1}{n} S_n - m)$ . En posant  $Y_i = \frac{X_i - m}{\sigma}$ , on a

$$Z_n = \sum_{i=1}^n \frac{Y_i}{\sqrt{n}}$$

où les  $Y_i$  sont des variables aléatoires centrées réduites (ie de moyenne de nulle et de variance 1). Le  $X_i$  étant i.i.d., les  $Y_i$  sont aussi i.i.d. Ainsi, on a

$$\varphi_{Z_n}(t) = \prod_{i=1}^n \varphi_{Y_i}\left(\frac{t}{\sqrt{n}}\right) = \left(\varphi_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

En faisant un développement limité de  $\varphi_{Y_i}$  on obtient donc

$$\varphi_{Z_n}(t) = \left(1 - \frac{1}{2n} t^2 + o\left(\frac{t^2}{n}\right)\right)^n$$

et donc

$$\lim_{n \rightarrow \infty} \varphi_{Z_n}(t) = e^{-\frac{t^2}{2}}$$

qui est la fonction caractéristique d'une variable aléatoire normale centrée réduite. Le théorème de continuité de Levy permet de conclure. ■