



Université Abdelmalek Essaadi  
Ecole Nationale des Sciences Appliquées  
Al Hoceima, Maroc



## **Statistique Inférentielle pour les Sciences de données**

---

–Cours et exercices de Statistique Inférentielle–

---

**Mohamed ADDAM**

Professeur de Mathématiques

École Nationale des Sciences Appliquées d'Al Hoceima  
–ENSAH–

[addam.mohamed@gmail.com](mailto:addam.mohamed@gmail.com)

[m.addam@uae.ac.ma](mailto:m.addam@uae.ac.ma)

©Mohamed ADDAM.

Année Universitaire 2021/2022

04 Octobre 2021



# Table des matières

<b>1</b>	<b>Calcul des probabilités</b>	<b>5</b>
1.1	Espace de probabilité discret . . . . .	5
1.1.1	Espace probabilisable . . . . .	5
1.1.2	Atomes . . . . .	6
1.2	Espace probabilisés . . . . .	6
1.2.1	Propriétés élémentaires . . . . .	7
1.2.2	Donnée pratique d'une probabilité . . . . .	8
1.3	Probabilité conditionnelle . . . . .	8
1.4	Événements indépendants et indépendance mutuelle . . . . .	8
1.4.1	Événements indépendants . . . . .	8
1.4.2	Indépendance mutuelle . . . . .	9
1.5	Formule de Bayes et applications . . . . .	10
<b>2</b>	<b>Variables aléatoires : discrètes et continues</b>	<b>11</b>
2.1	Les variables aléatoires . . . . .	11
2.1.1	Applications mesurables . . . . .	11
2.1.2	Événements liés à une application mesurable . . . . .	11
2.1.3	Probabilité image . . . . .	12
2.1.4	Loi de probabilité . . . . .	12
2.1.5	Fonction de répartition . . . . .	13
2.1.6	Quantile $q_\alpha$ , Médiane et Mode . . . . .	13
2.1.7	Systèmes de variables aléatoires réelles . . . . .	14
2.1.8	Couple indépendants de lois marginales données . . . . .	14
2.1.9	Système de variables aléatoires réelles indépendantes . . . . .	15
2.2	Moments : espérance mathématique et variance d'une v.a. réelle . . . . .	15
2.2.1	Espérance mathématique d'un produit . . . . .	16
2.2.2	Moments d'ordre $k$ d'une variable aléatoire . . . . .	16
2.2.3	Variance et écart-type d'une variable aléatoire . . . . .	17
2.2.4	Coefficient d'asymétrie et coefficient d'aplatissement . . . . .	18
2.3	Loi et espérance conditionnelles . . . . .	19
2.4	Covariance, corrélation, vecteur-moyenne et matrice de covariance . . . . .	19
2.4.1	Notation matricielle . . . . .	19
2.4.2	Suite blanche . . . . .	20

<b>3</b>	<b>Variables aléatoires ayant une densité</b>	<b>21</b>
3.1	Densité d'une variable aléatoire et fonction de répartition . . . . .	21
3.2	Moments d'ordre $k$ d'une variable aléatoire ayant une densité . . . . .	22
3.3	Espérance mathématique et variance d'une variable aléatoire ayant densité . . . . .	22
3.4	Coefficient d'asymétrie et coefficient d'aplatissement . . . . .	23
<b>4</b>	<b>Lois de probabilité usuelles : discrètes et continues</b>	<b>25</b>
4.1	Loi conjointe et lois marginales de variables aléatoires réelles . . . . .	25
4.1.1	Exemple d'un couple de variables aléatoires réelles . . . . .	25
4.1.2	Déterminons maintenant les lois marginales . . . . .	26
4.2	Loi uniforme continue sur $(a, b)$ . . . . .	27
4.3	Loi de Poisson . . . . .	27
4.4	Loi normale $\mathcal{N}(\mu, \sigma^2)$ ou v.a. gaussienne . . . . .	28
4.5	Valeurs numériques de $\Pi(x)$ . . . . .	28
4.6	Loi Gamma et loi <b>Khi-deux</b> : $\chi^2$ . . . . .	29
4.6.1	Fonction Gamma $\Gamma$ : . . . . .	29
4.7	Loi Gamma . . . . .	30
4.8	Loi Beta . . . . .	31
4.9	Loi du <b>Khi-2</b> où $\chi^2$ . . . . .	32
4.10	Lois de Student et de Fisher-Snedecor . . . . .	32
4.11	Quelques lois de probabilités . . . . .	32
<b>5</b>	<b>Fonctions caractéristiques</b>	<b>35</b>
5.1	Fonctions génératrices de v. a. entières . . . . .	35
5.2	Fonction caractéristique d'une variable aléatoire réelle . . . . .	35
5.2.1	Définitions et notations . . . . .	35
5.2.2	Fonction caractéristique d'une v. a. réelle discrète . . . . .	36
5.2.3	Fonction caractéristique d'une v. a. réelle ayant une densité . . . . .	36
5.3	Mesure de probabilité sur $\mathbb{R}^d$ . . . . .	36
5.3.1	Notions d'intégration d'une fonction à valeurs complexes . . . . .	37
5.3.2	Fonction caractéristique d'une mesure de probabilité . . . . .	37
5.3.3	Étude du cas où $d = 1$ . . . . .	39
5.3.4	Formules de réciprocity de Perron-Stieltjes : cas $d = 1$ . . . . .	39
5.4	Moments d'une mesure de probabilité dans le cas $d = 1$ . . . . .	40
<b>6</b>	<b>Convergences et comportement asymptotique</b>	<b>43</b>
6.1	Quelques relations de grand intérêt en statistique . . . . .	43
6.1.1	Quantile $q_\alpha$ , Médiane et Mode . . . . .	43
6.2	Convergence stochastique . . . . .	44
6.2.1	Convergence en probabilité . . . . .	44
6.2.2	Convergence en loi (convergence faible) . . . . .	44
6.2.3	Convergence presque sûre . . . . .	44
6.2.4	Convergence en moyenne . . . . .	44
6.3	Loi des grands nombres . . . . .	45
6.4	Théorème central limite . . . . .	47
6.5	Loi normale : Théorème de Moivre-Laplace . . . . .	48

# Chapitre 1

## Calcul des probabilités

### 1.1 Espace de probabilité discret

#### 1.1.1 Espace probabilisable

**Définition 1.1.1** Soit  $\Omega$  un ensemble et  $\mathcal{P}(\Omega)$  l'ensemble des parties de  $\Omega$ .

On appelle algèbre de Boole de parties de  $\Omega$  tout sous-ensemble non vide  $\mathcal{B}$  de  $\mathcal{P}(\Omega)$  stable par l'intersection et la complémentation.

Autrement dit

1.  $(\forall (A, B) \in \mathcal{B}^2), A \cap B \in \mathcal{B};$
2.  $(\forall A \in \mathcal{B}), \overline{A} = C_{\Omega}^A \in \mathcal{B}.$

**Théorème 1.1.1** Une algèbre de Boole de parties d'un ensemble  $\Omega$  est stable pour l'opération  $(A, B) \mapsto A \cup B$ . De plus, il contient la partie pleine de  $\Omega$  et la partie vide  $\emptyset$ .

**Démonstration.**

1.  $\mathcal{B}$  étant non vide, contient un élément  $A$ , contient donc  $\overline{A}$ , et donc  $A \cap \overline{A} \in \mathcal{B}$ , c'est-à-dire que  $\emptyset \in \mathcal{B}$ .
2.  $\emptyset \in \mathcal{B}$ , alors  $\overline{\emptyset} = C_{\Omega}^{\emptyset} = \Omega \in \mathcal{B}$ .
3. Soit  $A$  et  $B$  deux éléments de  $\mathcal{B}$ .  $\mathcal{B}$  contient  $\overline{A}$  et  $\overline{B}$ , donc  $\overline{A} \cap \overline{B} \in \mathcal{B}$ , donc  $\overline{\overline{A} \cap \overline{B}} \in \mathcal{B}$ , qui n'est autre que  $\overline{\overline{A} \cap \overline{B}} = A \cup B$ , d'où  $A \cup B \in \mathcal{B}$  (Loi de De Morgan).

□

**Exercice 1.1.1** Montrer que l'intersection (et la réunion) d'une famille finie d'éléments d'une algèbre de Boole  $\mathcal{B}$  appartient à  $\mathcal{B}$ .

**Exemple 1.1.1**

1.  $\mathcal{P}(\Omega)$  est une algèbre de Boole de parties de  $\Omega$ . On l'appelle l'algèbre de Boole discrète.
2.  $\{\emptyset, \Omega\}$  est une algèbre de Boole de parties de  $\Omega$ . On l'appelle l'algèbre de Boole grossière.
3. Si  $\Omega$  a au moins deux éléments, soit  $A$  une partie propre (c'est-à-dire distincte de  $\Omega$  et de  $\emptyset$ ) de  $\Omega$ . Alors  $\{\emptyset, A, \overline{A}, \Omega\}$  est une algèbre de Boole de parties de  $\Omega$ .

**Définition 1.1.2** Soit  $\Omega$  un ensemble et soit  $\tau \subset \mathcal{P}(\Omega)$ .

$\tau$  est dite **tribu** ou bien  **$\sigma$ -algèbre** sur  $\Omega$  si elle vérifie les conditions suivantes :

- i)  $\Omega \in \tau$ ,
- ii) si  $A \in \tau$ , alors  $\overline{A} = C_{\Omega}^A \in \tau$ ,

iii) si  $(A_i)_{i \in \mathbb{N}} \subset \tau$ , alors  $(\bigcup_{i \in \mathbb{N}} A_i) \in \tau$ .

**Exemple 1.1.2** 1.  $\mathcal{P}(\Omega)$  est une  $\sigma$ -algèbre de parties de  $\Omega$ . On l'appelle la  **$\sigma$ -algèbre discrète**.

2.  $\{\emptyset, \Omega\}$  est une  $\sigma$ -algèbre de parties de  $\Omega$ . On l'appelle la **tribu grossière**.

3. Si  $\Omega$  a au moins deux éléments, soit  $A$  une partie propre (c'est-à-dire distincte de  $\Omega$  et de  $\emptyset$ ) de  $\Omega$ . Alors  $\tau = \{\emptyset, A, \bar{A}, \Omega\}$  est une tribu de parties de  $\Omega$ .

En particulier, la tribu grossière est la plus petite tribu et la tribu discrète est la plus grande tribu sur un ensemble  $\Omega$ . Et si  $\tau$  était une tribu sur  $\Omega$  alors

$$\{\emptyset, \Omega\} \subset \tau \subset \mathcal{P}(\Omega).$$

**Définition 1.1.3** Soit  $\Omega$  un ensemble et  $\tau$  une tribu sur  $\Omega$ .

On appelle **espace probabilisable** le couple  $(\Omega, \tau)$  formé par un ensemble  $\Omega$  et une  $\sigma$ -algèbre  $\tau$  de parties de  $\Omega$ .

- $\Omega$  est appelé **univers**.
- Les éléments de  $\Omega$  sont appelés **éventualités**.
- Les éléments de  $\tau$  sont appelés **événements**.
- L'événement  $\emptyset$  est appelé événement **impossible**.
- L'événement  $\Omega$  est appelé événement **certain**.
- Deux événements  $A$  et  $B$  tels que  $A \cap B = \emptyset$  sont dits **incompatibles**.

\*Un espace probabilisable  $(\Omega, \tau)$  est dit **fini** si  $\Omega$  est un **ensemble fini**.

## 1.1.2 Atomes

**Définition 1.1.4** Soit  $(\Omega, \tau)$  un espace probabilisable fini. Deux éventualités  $\omega$  et  $\omega'$  sont dites **inséparables par  $\tau$**  si et seulement si, tout événement qui contient l'une contient l'autre.

**Exemple 1.1.3** Soit  $\Omega = \{a, b, c\}$ ,  $\tau = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$  :  $\tau$  est une tribu sur  $\Omega$  et  $b$  et  $c$  sont inséparables par  $\tau$ .

**Théorème 1.1.2** Soit  $(\Omega, \tau)$  un espace probabilisable fini.

La relation définie sur  $\Omega$  par “ $\omega$  et  $\omega'$  sont inséparables par  $\tau$ ” est une relation d'équivalence.

Les classes d'équivalence sont des événements, qu'on appelle les **atomes** de  $\tau$ .

**Remarque 1.1.1** Les atomes de l'espace probabilisable  $(\Omega, \mathcal{P}(\Omega))$  ne sont autres que les événements  $\{\omega\}$  réduits à une éventualité.

## 1.2 Espace probabilisés

**Définition 1.2.1** Soit  $(\Omega, \tau)$  un espace probabilisable.

On appelle **probabilité** sur  $(\Omega, \tau)$  toute application  $P$  de  $\tau$  dans l'intervalle  $[0, 1]$  de  $\mathbb{R}$  vérifiant les axiomes suivants :

$$(A_1) \quad P(\Omega) = 1.$$

$$(A_2) \quad P \text{ est } \sigma\text{-additive} :$$

$$(A_n)_n \subset \tau, \quad A_i \cap A_j = \emptyset, \quad i \neq j \quad \Rightarrow \quad P\left(\bigcup_n A_n\right) = \sum_n P(A_n)$$

\*Si  $\Omega$  est un ensemble fini, alors le triplet  $(\Omega, \tau, P)$  est appelé un **espace probabilisé fini**.

### 1.2.1 Propriétés élémentaires

Les axiomes entraînent immédiatement un certain nombre de propriétés élémentaires, vérifiées dans tout espace probabilisé  $(\Omega, \tau, P)$  :

1. La probabilité de l'événement impossible est nulle, c'est-à-dire que

$$P(\Omega) = 1 = P(\emptyset) + P(C_\Omega^\emptyset) = P(\emptyset) + 1 \Rightarrow P(\emptyset) = 0$$

2. Soit  $\{A_1, A_2, \dots, A_n\}$  un ensemble fini d'événements **deux à deux incompatibles** ; on a

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

ceci traduit l'axiome  $(A_2)$  dans le cas d'un ensemble fini d'événements.

3. Si  $A$  et  $B$  sont deux événements contraires ( $B = \overline{A}$ ), alors la somme de leurs probabilités est égale à 1 :

$$P(\Omega) = 1 = P(A) + P(\overline{A})$$

car  $\Omega = A \cup \overline{A}$ .

4. Nous avons la propriété suivante :

**Théorème 1.2.1** Soient  $A$  et  $B$  deux événements quelconques d'un espace probabilisé  $(\Omega, \tau, P)$ , on a

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

**Démonstration.** Quels que soient les événements  $A$  et  $B$ , on peut écrire

$$A \cup B = (A \cap \overline{B}) \cup (A \cap B) \cup (\overline{A} \cap B)$$

et les trois événements  $A \cap \overline{B}$ ,  $A \cap B$  et  $\overline{A} \cap B$  sont deux à deux incompatibles. Donc

$$P(A \cup B) = P(A \cap \overline{B}) + P(A \cap B) + P(\overline{A} \cap B);$$

d'autre part,  $A$  est la réunion de deux événements incompatibles :

$$A = (A \cap \overline{B}) \cup (A \cap B),$$

donc  $P(A) = P(A \cap \overline{B}) + P(A \cap B)$ ,

ce qui s'écrit aussi

$$P(A \cap \overline{B}) = P(A) - P(A \cap B);$$

de la même façon, on établirait que

$$P(\overline{A} \cap B) = P(B) - P(A \cap B);$$

reportant ces deux valeurs dans l'expression de  $P(A \cup B)$ , on obtient bien

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

□

5. Si  $A \subset B$ , alors on a

$$P(B) = P(A) + P(B - A), \quad \text{d'où} \quad P(A) \leq P(B)$$

**Définition 1.2.2** 1. On appelle événement presque impossible tout événement  $A$  de probabilité nulle.

2. On appelle événement presque certain tout événement  $A$  de probabilité égale à 1.

## 1.2.2 Donnée pratique d'une probabilité

Soit  $(\Omega, \tau)$  un espace probabilisable fini et  $A_1, A_2, \dots, A_n$  les atomes de  $\tau$ .

- Tout événement étant réunion d'un nombre fini d'atomes, deux probabilités  $P_1$  et  $P_2$  qui sont égales pour chaque atome sont égales pour chaque événement.
- Pour connaître une probabilité, il n'est donc pas nécessaire de connaître les probabilités de tous les événements : celles des atomes suffisent !!

**Inversement**, soit  $p_1, p_2, \dots, p_n, n$  nombres réels. A quelle condition existe-t-il une probabilité  $P$  (qui serait alors unique d'après la remarque qui précède) telle que

$$(\forall i \in \{1, 2, \dots, n\}) \quad P(A_i) = p_i?$$

Les conditions

(i)  $(\forall i \in \{1, 2, \dots, n\}), 0 \leq p_i \leq 1;$

(ii)  $1 = P(\Omega) = \sum_{i=1}^n P(A_i) = \sum_{i=1}^n p_i$

sont nécessaires.

**Définition 1.2.3** Un espace de probabilité  $(\Omega, \tau, P)$  est dit discret si  $\tau = \mathcal{P}(\Omega)$  est la tribu discrète.

## 1.3 Probabilité conditionnelle

**Définition 1.3.1** Soit  $(\Omega, \tau, P)$  un espace probabilisé et  $B$  un événement de probabilité non nulle. L'application

$$P_B : \tau \longrightarrow [0, 1], \quad A \longmapsto \frac{P(A \cap B)}{P(B)}$$

est une probabilité sur  $(\Omega, \tau)$  appelée probabilité conditionnelle par  $B$ .

**Propriété 1.3.1** Soit  $(\Omega, \tau, P)$  un espace de probabilisé et  $A$  et  $B$  deux événements de probabilités non nulles. Alors on a la formule des probabilités composées :

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

**Démonstration.**

□

## 1.4 Événements indépendants et indépendance mutuelle

### 1.4.1 Événements indépendants

**Définition 1.4.1** Soit  $(\Omega, \tau, P)$  un espace probabilisé. deux événements  $A$  et  $B$  tels que

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

sont appelés indépendants.



	$a$	$b$	$c$	$d$	$e$
$P_1$	1/6	1/6	1/6	1/6	1/3
$P_2$	1/10	2/10	3/10	1/10	3/10

**Exemple 1.4.1** Soit  $\Omega = \{a, b, c, d, e\}$  et  $\tau = \mathcal{P}(\Omega)$ . On définit deux probabilités  $P_1$  et  $P_2$  sur l'espace  $(\Omega, \mathcal{P}(\Omega))$  par le tableau suivant :

et les événements

$$a = \{a, b, c\}, \quad B = \{c, d\}, \quad A \cap B = \{c\};$$

alors

$$P_1(A) = \frac{3}{6}; \quad P_1(B) = \frac{2}{6}; \quad P_1(A \cap B) = \frac{1}{6}.$$

Dans  $(\Omega, \mathcal{P}(\Omega), P_1)$ ,  $A$  et  $B$  sont indépendants ( $\frac{3}{6} \times \frac{2}{6} = \frac{1}{6}$ ).

D'autre part

$$P_2(A) = \frac{6}{10}; \quad P_2(B) = \frac{4}{10}; \quad P_2(A \cap B) = \frac{3}{10}.$$

Dans  $(\Omega, \mathcal{P}(\Omega), P_2)$ ,  $A$  et  $B$  ne sont pas indépendants ( $\frac{6}{10} \times \frac{4}{10} \neq \frac{3}{10}$ ).

## 1.4.2 Indépendance mutuelle

**Définition 1.4.2** Soit  $(\Omega, \tau, P)$  un espace probabilisé. Des événements  $A_1, A_2, \dots, A_p$  sont dits mutuellement indépendants si, pour tout entier  $k$  et pour tout sous-ensemble  $\{A_{i_1}, \dots, A_{i_k}\}$  de  $\{A_1, A_2, \dots, A_p\}$  formé d'événements distincts, on a la relation

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$$

Il résulte immédiatement de cette définition que si des événements sont mutuellement indépendants, ils sont indépendants deux à deux. **La réciproque est fautive**, comme le montre l'exemple suivant

**Exemple 1.4.2** On tire une carte d'un jeu de 52 et on considère les événements suivants :

- $A$  : la carte tirée est rouge ;
- $B$  : la carte tirée est une majeure (cœur ou pique) ;
- $C$  : la carte tirée est un pique ou un carreau.

Alors les événements

- $A \cap B$  : la carte tirée est un cœur,
- $A \cap C$  : la carte tirée est un carreau,
- $B \cap C$  : la carte tirée est un pique.

ont chacun pour probabilité  $\frac{1}{4}$ , tandis que

$$P(A) = P(B) = P(C) = \frac{1}{2}.$$

Les événements  $A, B, C$  sont donc indépendants deux à deux.

Mais

$$A \cap B \cap C = \emptyset,$$

et donc  $P(A \cap B \cap C) = 0$ ,

tandis que

$$P(A)P(B)P(C) = \frac{1}{8}.$$

Les événements  $A, B, C$  ne sont donc pas mutuellement indépendants.

## 1.5 Formule de Bayes et applications

Soit  $(\Omega, \mathcal{P}(\Omega), P)$  un espace probabilisé et des événements  $C_1, C_2, \dots, C_n$  constituant une partition de  $\Omega$ , c'est-à-dire

$$(\forall i \in \{1, 2, \dots, n\}, \forall j \in \{1, 2, \dots, n\}) \quad (i \neq j) \Rightarrow (C_i \cap C_j = \emptyset) \text{ et } \Omega = C_1 \cup C_2 \cup \dots \cup C_n.$$

Nous supposons qu'aucun événement  $C_i$  n'est presque impossible.

Soit maintenant  $A$  un événement non presque impossible :

$$A = A \cap \Omega = (A \cap C_1) \cup (A \cap C_2) \cup \dots \cup (A \cap C_n).$$

Les événements  $A \cap C_i$  sont deux à deux incompatibles, et par suite :

$$P(A) = P(A \cap C_1) + P(A \cap C_2) + \dots + P(A \cap C_n).$$

D'autre part, pour un indice  $k$  donné :

$$P_A(C_k) = \frac{P(A \cap C_k)}{P(A)} = \frac{P(C_k)P_{C_k}(A)}{P(A)},$$

d'où l'on tire

$$P_A(C_k) = \frac{P(C_k)P_{C_k}(A)}{P(C_1)P_{C_1}(A) + P(C_2)P_{C_2}(A) + \dots + P(C_n)P_{C_n}(A)}.$$

Cette formule est connue sous le nom de **formule de Bayes**.

**Exemple 1.5.1** *Un examen se compose de questions auxquelles il faut répondre par **oui** ou **non**. Si un étudiant connaît la réponse, il répond correctement ; s'il l'ignore, il tire à pile ou face la réponse qu'il inscrira.*

*Un étudiant donné connaît 60% du programme : Quelle est la probabilité pour qu'une réponse juste soit due à ses connaissances plutôt qu'au hasard ?*

**Réponse :** Soit les événements :  $C_1$  : l'étudiant connaît la réponse ;

$C_2$  : l'étudiant ne connaît pas la réponse ;

$A$  : l'étudiant répond juste.

L'énoncé nous donne :  $P_{C_1}(A) = 1$ ,  $P_{C_2}(A) = 0.5$ ,  $P(C_1) = 0.6$  et  $P(C_2) = 0.4$ .

La formule de Bayes fournit alors la probabilité cherchée :

$$P_A(C_1) = \frac{P(C_1)P_{C_1}(A)}{P(C_1)P_{C_1}(A) + P(C_2)P_{C_2}(A)},$$

soit

$$P_A(C_1) = \frac{1 \times 0.6}{1 \times 0.6 + 0.5 \times 0.4} = \frac{3}{4} = 0.75$$

# Chapitre 2

## Variables aléatoires : discrètes et continues

### 2.1 Les variables aléatoires

#### 2.1.1 Applications mesurables

Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisable et  $X$  une application de  $\Omega$  dans un ensemble  $J$ . Si  $A$  est une partie de  $J$ , on notera  $X^{-1}(A)$  l'ensemble des éventualités dont l'image appartenant à  $A$ .

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

**Définition 2.1.1** Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisable fini. Une application  $X$  de  $\Omega$  dans un ensemble  $J$  est dite **mesurable** si, et seulement si, pour tout point  $x$  de  $J$ ,  $X^{-1}(\{x\})$  est un événement :

$$\forall x \in J \quad X^{-1}(\{x\}) \in \mathcal{B}.$$

Au lieu de “application mesurable”, on dit aussi “**variable aléatoire**”

**Remarque 2.1.1** 1. Si  $J = \mathbb{R}^n$  alors

(a) dans le cas  $n = 1$ , on dit “**variable aléatoire réelle**”,

(b) dans le cas  $n > 1$ , on dit “**variable aléatoire vectorielle**”.

2. Si  $\mathcal{B} = \mathcal{P}(\Omega)$ , toute application de  $\Omega$  dans  $J$  est mesurable.

#### 2.1.2 Événements liés à une application mesurable

**Théorème 2.1.1** Soit  $(\Omega, \mathcal{B})$  un espace probabilisable et  $X$  une application mesurable de  $\Omega$  dans un ensemble  $J$ . Pour toute partie  $A$  de  $J$ , on a  $X^{-1}(A)$  est un événement.

**Démonstration.**

1. Si  $\Omega$  est fini, alors  $X(\Omega) \cap A$  est fini.

Soit donc  $X(\Omega) \cap A = \{a_1, a_2, \dots, a_n\}$  (les  $a_i$  sont distincts).

$X^{-1}(A)$  est alors la réunion d'un nombre fini d'événements

$$X^{-1}(\{a_1\}), X^{-1}(\{a_2\}), \dots, X^{-1}(\{a_n\}).$$

$X^{-1}(A)$  est donc un événement

2. Si  $\Omega$  n'est pas fini, on a  $\Omega$  est la réunion d'atome de  $\mathcal{B}$  et dans ce cas  $X$  est constante sur tout atome. □

**Définition 2.1.2** Soit  $X : \Omega \longrightarrow J$  une variable aléatoire. Pour toute partie  $A$  de  $J$ ,  $X^{-1}(A)$  est appelé événement.

On appelle **indicatrice d'un événement**  $A$ , la fonction  $\varphi_A$  définie par

$$(\forall \omega \in \Omega) \quad \begin{cases} \varphi_A(\omega) = 1 & \text{si } \omega \in A, \\ \varphi_A(\omega) = 0 & \text{si } \omega \notin A, \end{cases}$$

### 2.1.3 Probabilité image

Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisé et  $X : \Omega \rightarrow J$  une variable aléatoire. Soit l'espace probabilisable  $(J, \mathcal{P}(J))$  où  $\mathcal{P}$  est l'ensemble des parties mesurables de  $J$ . On considère l'application

$$Q : \mathcal{P}(J) \rightarrow [0, 1], \quad A \mapsto P(X^{-1}(A)) \quad (2.1)$$

A toute partie  $A$  de  $J$  associons le nombre  $Q(A)$ , probabilité de l'événement  $X^{-1}(A)$ .

Le théorème suivant montre que  $X$  permet en quelque sorte de transporter une probabilité de  $(\Omega, \mathcal{B})$  vers  $(J, \mathcal{P}(J))$

**Théorème 2.1.2** Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisé et  $X : \Omega \rightarrow J$  une variable aléatoire. L'application  $Q$  définie en (2.1) est une probabilité sur  $(J, \mathcal{P}(J))$ .

**Démonstration.**  $Q$  est bien une application de  $\mathcal{P}(J)$  dans  $[0, 1]$ .

i)  $X^{-1}(J) = \Omega$  donc  $Q(J) = P(X^{-1}(J)) = P(\Omega) = 1$ .

ii) Soit  $K$  et  $L$  deux parties disjointes de  $J$  :

- $K \cap L = \emptyset$  et donc  $X^{-1}(K) \cap X^{-1}(L) = \emptyset$ ,
- $X^{-1}(K \cup L) = X^{-1}(K) \cup X^{-1}(L)$ ;

par suite

$$Q(K \cup L) = P(X^{-1}(K \cup L)) = P(X^{-1}(K) \cup X^{-1}(L));$$

or  $X^{-1}(K)$  et  $X^{-1}(L)$  sont deux événements disjoints et donc

$$P(X^{-1}(K \cup L)) = P(X^{-1}(K)) + P(X^{-1}(L)) = Q(K) + Q(L).$$

On a donc montré que

$$(K \cap L = \emptyset) \quad \Rightarrow \quad (Q(K \cup L) = Q(K) + Q(L)).$$

□

**Définition 2.1.3**  $Q$  est bien une probabilité sur  $(J, \mathcal{P}(J))$  : on l'appelle la **probabilité image** de  $P$  par  $X$ .

### 2.1.4 Loi de probabilité

**Définition 2.1.4** Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisé et  $X : \Omega \rightarrow J$  une variable aléatoire vectorielle. On appelle **loi de probabilité** de  $X$  la probabilité image de  $P$  par  $X$ , c'est à dire

$$Q(A) = P(X^{-1}(A)).$$

Une loi de probabilité est donc une probabilité sur l'espace probabilisable  $(J^n, \mathcal{P}(J^n))$  où  $J^n$  désigne une partie de  $\mathbb{R}^n$  contenant  $X(\Omega)$ . Nous utiliserons beaucoup cette notion dans le cas  $n = 1$  (variable aléatoire réelle).

**Définition 2.1.5** On appelle **variable aléatoire discrète** une application  $X : \Omega \rightarrow F$  où  $F$  est un ensemble fini ou dénombrable.

\*Pour  $x \in F$ , on note de façon concise  $[X = x]$  l'événement  $\{\omega \in \Omega / X(\omega) = x\}$ .

\*La famille des nombres réels  $(P([X = x]))_{x \in F}$  s'appelle la **loi** de  $X$ .

## 2.1.5 Fonction de répartition

Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisé et  $X : \Omega \rightarrow J$  une variable aléatoire. On appelle fonction de répartition de  $X$  la fonction réelle de variable réelle

$$x \mapsto F_X(x) = P([X < x])$$

où l'ensemble  $[X < x] = X^{-1}(]-\infty; x])$ . On a

$$F_X(x) = P(X^{-1}(]-\infty; x]))$$

**Propriété 2.1.1** Nous donnons ici quelques propriétés concernant la fonction de répartition : Soit  $X : \Omega \rightarrow J$  une variable aléatoire

1. Pour  $x \leq \min_{y \in \Omega} X(y)$  alors  $F_X(x) = P(X^{-1}(]-\infty; x])) = P(\emptyset) = 0$ ;
2. Pour  $x \geq \max_{y \in \Omega} X(y)$  alors  $F_X(x) = P(X^{-1}(]-\infty; x])) = P(\Omega) = 1$ ;
3. Une fonction de répartition est croissante ;
4. La probabilité pour que  $X$  prenne ses valeurs dans un intervalle  $[a, b[$  peut s'exprimer à l'aide de la fonction de répartition de  $X$ . En effet, l'événement contraire de  $[a \leq X < b]$  est la réunion des deux événements disjoints :

$$[X \geq b], \quad \text{de probabilité } 1 - F(b); \quad [X < a], \quad \text{de probabilité } F(a)$$

cette événement contraire a donc pour probabilité

$$P([X \geq b] \cup [X < a]) = 1 - F(b) + F(a),$$

et par suite

$$P([a \leq X < b]) = F(b) - F(a)$$

## 2.1.6 Quantile $q_\alpha$ , Médiane et Mode

### Définition 2.1.6 (Quantile)

Soit  $X$  une variable aléatoire continue. Le nombre  $q_\alpha$  tel que

$$P([X \leq q_\alpha]) = \alpha \in [0, 1]$$

est le **quantile** d'ordre  $\alpha$  de la loi de probabilité de  $X$ .

Ces quantile sont notés de différentes façons : par  $u_\alpha$  pour la loi normale, par  $t_\alpha^n$  pour la loi de Student à  $n$  degrés de liberté et par  $\chi_\alpha^n$  pour la loi du  $\chi_n^2$ .

### Définition 2.1.7 (Médiane)

Soit  $X$  une variable aléatoire continue. La valeur  $q_{0.5}$  est la médiane de la variable aléatoire si

$$P([X \leq q_{0.5}]) = 0.5.$$

La médiane est le quantile  $\alpha = 0.5$  de la variable aléatoire  $X$ .

### Définition 2.1.8 (Mode)

Soit  $X$  une variable aléatoire continue de densité de probabilité  $f$ . Un point  $x_0$  du domaine de définition  $\mathcal{D}$  de  $f$  tel que  $f(x_0) = \max_{x \in \mathcal{D}} f(x)$  est appelé un **mode** de la variable aléatoire  $X$ .

## 2.1.7 Systèmes de variables aléatoires réelles

**Théorème 2.1.3** soit  $(\Omega, \mathcal{B})$  un espace probabilisable et une base  $\mathcal{A}$  de  $\mathbb{R}^n$ .

Une application  $X$  de  $\Omega$  dans  $\mathbb{R}^n$  est une variable aléatoire vectorielle si, et seulement si, les  $n$  applications coordonnées,  $X_1, X_2, \dots, X_n$  de  $X$  dans  $\mathcal{A}$  sont des variables aléatoires réelles.

**Démonstration.**  $\Rightarrow$  Soit  $S_n = (X_1, X_2, \dots, X_n)$  une suite de  $n$  variables aléatoires réelles.

Soit  $\{e_1, e_2, \dots, e_n\}$  une base de l'espace vectoriel  $\mathbb{R}^n$ . Posons

$$X = X_1.e_1 + X_2.e_2 + \dots + X_n.e_n.$$

$X$  est une application de  $\Omega$  dans  $\mathbb{R}^n$ . Est-ce une variable aléatoire vectorielle ?

Il faudrait pour cela que, quel que soit le vecteur

$$x = x_1.e_1 + x_2.e_2 + \dots + x_n.e_n$$

de  $\mathbb{R}^n$ ,  $X^{-1}(\{x\})$  soit un événement. Or  $X^{-1}(\{x\})$  est l'intersection des  $n$  événements

$$[X_1 = x_1], [X_2 = x_2], \dots, [X_n = x_n],$$

c'est donc un événement.

$\Leftarrow$  soit  $x$  une variable aléatoire vectorielle (à valeurs dans  $\mathbb{R}^n$ ) et  $\{e_1, e_2, \dots, e_n\}$  une base de l'espace vectoriel  $\mathbb{R}^n$ . Les coordonnées  $(X_1, X_2, \dots, X_n)$  de  $X$ , définies par

$$X = X_1.e_1 + X_2.e_2 + \dots + X_n.e_n$$

sont des applications de  $\Omega$  dans  $\mathbb{R}$ . Sont-ce des variables aléatoires réelles ?

Raisonnons pour  $X_1$  ; il faudrait pour cela que, quel que soit le réel  $x_1$ ,  $X_1^{-1}(\{x_1\})$  soit un événement . Or  $X_1^{-1}(\{x_1\})$  est la réunion des événements

$$X = (x_1, x_2, \dots, x_n)$$

où  $x_1, x_2, \dots, x_n$  prennent toutes les valeurs possibles.

$X_1^{-1}(\{x_1\})$  est donc un événement et  $X_1$  est bien une variable aléatoire. □

## 2.1.8 Couple indépendants de lois marginales données

La probabilité de l'intersection de deux événements ne peut se calculer à partir des probabilités de ces événements : les lois marginales d'un couple ne permettent pas de déterminer la loi conjointe.

Il en va différemment si l'on sait qu'il s'agit d'un couple indépendant. En effet, pour  $(i, j) \in X(\Omega)$

$$P([X = (i, j)]) = P([X_1 = i])P([X_2 = j]).$$

**Exemple 2.1.1** Déterminer la loi conjointe d'un couple de variables aléatoires réelles indépendantes, de même loi On prépare le tableau, avec ses marges et on écrit dans chaque case le produit des probabilités figurant dans

$i$	0	1	2
$P([X = i])$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

les marges correspondantes.

$X_2 \backslash X_1$	0	1	2	
0	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$	$\frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$	$\frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$	$\rightarrow \frac{1}{2}$
1	$\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$	$\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$	$\frac{1}{3} \times \frac{1}{6} = \frac{1}{18}$	$\rightarrow \frac{1}{3}$
2	$\frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$	$\frac{1}{3} \times \frac{1}{6} = \frac{1}{18}$	$\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$	$\rightarrow \frac{1}{6}$
	$\downarrow$ $\frac{1}{2}$	$\downarrow$ $\frac{1}{3}$	$\downarrow$ $\frac{1}{6}$	

## 2.1.9 Système de variables aléatoires réelles indépendantes

On peut généraliser la définition donnée pour un couple.

**Définition 2.1.9** Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisé. Un système de  $n$  variables aléatoires réelles  $X = (X_1, X_2, \dots, X_n)$  tels que

$$(\forall (a_1, a_2, \dots, a_n) \in \mathbb{R}^n \quad [X_1 = a_1], [X_2 = a_2], \dots [X_n = a_n])$$

sont des événements mutuellement indépendants est appelé système de  $n$  variables aléatoires réelles indépendantes.

Comme pour un couple, on peut se restreindre aux éléments  $(a_1, a_2, \dots, a_n)$  qui appartiennent à  $X(\Omega)$ .

De même que pour un couple, si  $(X_1, X_2, \dots, x_n)$  est un système indépendant, quelless que soient la parties  $A_1, A_2, \dots, A_n$  de  $\mathbb{R}$ , les événements  $X_1^{-1}(A_1), X_2^{-1}(A_2), \dots, X_n^{-1}(A_n)$  sont indépendants.

## 2.2 Moments : espérance mathématique et variance d'une v.a. réelle

**Définition 2.2.1** Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisé fini et  $(A_1, A_2, \dots, A_n)$  une suite formée de sous-ensembles de  $\mathcal{B}$ . On appelle espérance mathématique la forme linéaire  $E$  définie l'espace vectoriel des variables aléatoires réelles  $V$  sur  $(\Omega, \mathcal{B})$  par ses valeurs pour les indicatrices constituant la base :  $(\varphi_{A_1}, \dots, \varphi_{A_n})$  de la manière suivante

$$E(\varphi_{A_i}) = P(A_i).$$

Si  $X$  est une variable aléatoire réelle prenant la valeur  $x_i$  sur  $A_i$  ( $i = 1, \dots, n$ ).

On a

$$X = \sum_{i=1}^n x_i \varphi_{A_i}$$

et donc on a  $E(X) = \sum_{i=1}^n x_i P(A_i)$ . On peut établir que si  $X(\Omega) = \{y_1, y_2, \dots, y_r\}$  où les  $y_i$  sont distincts alors

l'expression de l'espérance mathématique  $E$  devient

$$E(X) = \sum_{i=1}^r y_i P(X = y_i).$$

**Définition 2.2.2** Une variable aléatoire réelle  $X$  est dite centrée si son espérance mathématique  $E(X)$  est nulle.

**Exercice 2.2.1** Montrer que la variable aléatoire  $X - E(X)$  est centrée.

**Exemple 2.2.1** (Espérance d'une variable binomiale)

Une variable binomiale de paramètres  $n$  et  $p$  ( $n \in \mathbb{N}^*$  et  $0 < p < 1$ ) est une variable aléatoire telle que

$$X(\Omega) = \{0, 1, 2, \dots, n\} \quad \text{et} \quad P(X = i) = C_n^i p^i (1-p)^{n-i}.$$

D'après son expression, l'espérance mathématique est

$$\begin{aligned} E(X) &= \sum_{i=1}^n i C_n^i p^i (1-p)^{n-i} \\ &= \sum_{i=1}^n i \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\ &= np \sum_{i=1}^n \frac{(n-1)!}{i!(n-i)!} p^{i-1} (1-p)^{n-i} \\ &= np \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} p^j (1-p)^{n-1-j} \\ &= np(p + (1-p))^{n-1} \\ &= np \end{aligned}$$

**2.2.1 Espérance mathématique d'un produit**

Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisé fini. Un couple de variables aléatoires réelles  $(X, Y)$  définies sur  $(\Omega, \mathcal{B})$ .  
Soit

$$\begin{aligned} X(\Omega) &= \{x_1, x_2, \dots, x_n\} \\ Y(\Omega) &= \{y_1, y_2, \dots, y_p\}. \end{aligned}$$

où les  $x_i$  étant distincts et  $y_i$  étant distincts.

On a  $(X, Y)(\Omega) = \{(x_i, y_j)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}\}$  les  $(x_i, y_j)$  étant distincts,

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^p x_i y_j P((X, Y) = (x_i, y_j)).$$

Si on ne connaît pas les lois marginales de  $X$  et  $Y$ , il n'est pas possible de poursuivre puisqu'on ne peut pas calculer

$$P((X, Y) = (x_i, y_j)).$$

♣ Si  $X$  et  $Y$  sont indépendantes, alors

$$P((X, Y) = (x_i, y_j)) = P(X = x_i)P(Y = y_j),$$

et on a

$$E(XY) = E(X)E(Y).$$

**2.2.2 Moments d'ordre  $k$  d'une variable aléatoire**

Soit  $(\Omega, \tau, P)$  un espace probabilisé et  $X$  une variable aléatoire réelle.



**Définition 2.2.3** On appelle **moment d'ordre  $k$**  ( $k \geq 0$ ) de  $X$  et on note  $M_k(X)$  l'espérance mathématique de la variable aléatoire  $X^k$ . On écrit

$$M_k(X) = E[X^k].$$

\*On s'intéresse plus volontiers au moment d'ordre deux par rapport à un réel  $a$  :

$$M_2(X - a) = E[(X - a)^2].$$

En Physique, ce moment mesure la **dispersion** de la variable aléatoire  $X$  par rapport au nombre  $a$ .

### **Théorème 2.2.1 (Changement d'origine)**

Soit  $X$  une variable aléatoire réelle. De toutes les variables  $X - a$  ( $a \in \mathbb{R}$ ), c'est la variable centrée qui a le plus petit moment d'ordre deux.

**Démonstration.** Considérons deux réels  $a$  et  $b$  et comparons les moments

$$M_2(X - a) \quad \text{et} \quad M_2(X - b).$$

On a

$$\begin{aligned} (X - a)^2 &= (X - b + (b - a))^2 \\ &= (X - b)^2 + 2(b - a)(X - b) + (b - a)^2 \end{aligned}$$

d'où l'on déduit

$$E[(X - a)^2] = E[(X - b)^2] + 2(b - a)(E[X] - b) + (b - a)^2;$$

or, l'espérance d'une constante est égale à cette constante. On a donc :

$$M_2(X - a) = M_2(X - b) + 2(b - a)(E[X] - b) + (b - a)^2$$

en particulier, si on choisit  $b = E[X]$  :

$$M_2(X - a) = M_2(X - E[X]) + (E[X] - a)^2.$$

□

## **2.2.3 Variance et écart-type d'une variable aléatoire**

**Définition 2.2.4** On appelle **variance** (ou **indicateur de dispersion**) d'une variable aléatoire réelle  $X$ , la quantité  $\sigma^2(X) := \text{Var}(X)$  définie par

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

**Propriété 2.2.1** Soit  $k$  un réel. Alors on a les propriétés suivantes :

1.  $\text{Var}(k.X) = k^2.\text{Var}(X)$ ,
2.  $\text{Var}(X + k) = \text{Var}(X)$ ,
3.  $\text{Var}(X) = E(X^2) - [E(X)]^2$ .

**Définition 2.2.5** On appelle l'**écart-type** d'une variable aléatoire réelle  $X$ , la quantité  $\sigma(X)$  définie par

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

**Exemple 2.2.2** Soit  $X$  une variable aléatoire binomiale de paramètres  $n$  et  $p$  ( $n \in \mathbb{N}^*$  et  $0 < p < 1$ ). On a

$$X(\Omega) = \{0, 1, 2, \dots, n\} \quad \text{et} \quad P(X = i) = C_n^i p^i (1-p)^{n-i}.$$

Calculons d'abord  $E(X^2)$ , ou mieux encore, pour un calcul plus facile calqué sur celui de l'espérance,  $E(X(X-1))$  :

$$\begin{aligned} E(X(X-1)) &= \sum_{i=0}^n i(i-1) \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\ &= np \sum_{i=0}^n \frac{n!}{(i-2)!(n-i)!} p^{i-2} (1-p)^{n-i} \\ &= n(n-1)p^2 \sum_{i=2}^n \frac{(n-2)!}{(i-2)!(n-i)!} p^{i-2} (1-p)^{n-i} \\ &= n(n-1)p^2 \end{aligned}$$

et donc on a

$$E(X^2) = E(X(X-1)) + E(X) = n(n-1)p^2 + np.$$

En fin

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

**Exemple 2.2.3** Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisé fini et  $X : \Omega \rightarrow X(\Omega)$  la variable aléatoire de Bernoulli, c'est-à-dire que  $X(\Omega) = \{0, 1\}$  et  $P(X = 0) = q$ ,  $P(X = 1) = p$  où  $0 < q, p < 1$   $p + q = 1$ . L'espérance de la variable aléatoire de Bernoulli est

$$\begin{aligned} E(X) &= \sum_{i=0}^1 iP(X = i) = 0P(X = 0) + 1P(X = 1) = 0(1-p) + 1p = p. \\ E(X^2) &= 0^2 \cdot p + 1^2 \cdot q = q = 1-p. \\ E(\cos(\pi X)) &= \cos(0) \cdot p + \cos(\pi) \cdot q = p - q = p - (1-p) = 2p - 1. \\ \text{Var}(X) &= E(X^2) - [E(X)]^2 = p - p^2 = p(1-p). \end{aligned}$$

## 2.2.4 Coefficient d'asymétrie et coefficient d'aplatissement

**Définition 2.2.6** Soit  $X$  une variable aléatoire d'espérance mathématique  $\mu$  et de variance  $\text{Var}(X)$ .

1. On appelle **coefficient d'asymétrie** (skewness) de la v.a.  $X$ , noté  $\gamma_1(X)$ , le coefficient donné par l'expression

$$\gamma_1(X) = \frac{\mathbb{E}[(X - \mu)^3]}{(\text{Var}(X))^3}.$$

2. On appelle **coefficient d'aplatissement** (kurtosis) de la v.a.  $X$ , noté  $\gamma_2(X)$ , le coefficient donné par l'expression

$$\gamma_2(X) = \frac{\mathbb{E}[(X - \mu)^4]}{(\text{Var}(X))^4}.$$

## 2.3 Loi et espérance conditionnelles

**Définition 2.3.1** Soient  $X$  et  $Y$  deux variables aléatoires discrètes à valeurs respectives dans  $F$  et  $G$ . Pour  $y \in G$ , on appelle **loi conditionnelle** de  $X$  sachant  $Y = y$  la famille des nombres

$$(P_{[Y=y]}([X = x]))_{x \in X}$$

**Remarque 2.3.1** On a

$$P_{[Y=y]}([X = x]) = \begin{cases} \frac{P([X = x] \cap [Y = y])}{P([Y = y])}, & \text{si } P([Y = y]) > 0, \\ P([X = x]), & \text{sinon,} \end{cases}$$

**Proposition 2.3.1** Les variables  $X$  et  $Y$  sont indépendantes si, et seulement si la loi conditionnelle de  $X$  sachant  $Y = y$  ne dépend pas de  $y \in G$ . C'est-à-dire

$$P_{[Y=y]}([X = x]) = P([X = x]).$$

## 2.4 Covariance, corrélation, vecteur-moyenne et matrice de covariance

**Définition 2.4.1** Soit  $(X, Y)$  deux variables aléatoires,

1. on appelle **covariance** entre  $X$  et  $Y$  la quantité définie par :

$$\begin{aligned} \text{cov}(X, Y) &= E((X - E(X))(Y^* - E(Y^*))), \\ &= E(XY^*) - E(X)E(Y^*). \end{aligned}$$

2. On dit que  $X$  et  $Y$  ne sont pas corrélées si  $\text{cov}(X, Y) = 0$  soit encore  $E(XY^*) = E(X)E(Y^*)$ .

3. On appelle **coefficient de corrélation** la quantité définie par :

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}.$$

On a  $-1 \leq \rho(X, Y) \leq 1$ . Soit  $(X_1, \dots, X_n)$   $n$  variables aléatoires de moyennes respectives  $E(X_i)$ .

1. On appelle **vecteur-moyenne** le vecteur de dimension  $n$  dont les composantes sont les moyennes  $E(X_i)$ .

2. On appelle **matrice de covariance** la matrice  $C$  de dimension  $(n \times n)$  dont l'élément générateur est  $C_{ij} = \text{cov}(X_i, X_j)$  pour  $1 \leq i \leq n$  et  $1 \leq j \leq n$ .

### 2.4.1 Notation matricielle

Si on note :

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

le vecteur aléatoire dont les  $n$  composantes sont les variables aléatoires  $X_k$ , le vecteur-moyenne s'écrit :

$$E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$$

et la matrice de covariance :

$$\mathbf{C} = E((\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))^T).$$

Notons que les éléments diagonaux de la matrice de covariance représentent les variances respectives des  $n$  variables aléatoires ; ils sont donc positifs.

Si les  $n$  variables aléatoires ne sont pas corrélées, leur matrice de covariance est diagonale.

**Propriété 2.4.1** Toute matrice de covariance est positive. Ce qui signifie que, quel que soit le vecteur  $\mathbf{x} \in \mathbb{R}^n$  on a  $\mathbf{x}^T \mathbf{C} \mathbf{x} \geq 0$ .

**Démonstration.** Considérons la v.a.  $Y = \sum_{k=1}^N a_k (X_k - E(X_k))$ . On a  $E(|Y|^2) \geq 0$  car l'espérance mathématique d'une variable aléatoire positive est positif. Exprimons maintenant  $E(|Y|^2)$ , il vient :

$$\begin{aligned} E(|Y|^2) &= E \left( \sum_{k=1}^N a_k (X_k - E(X_k)) \sum_{\ell=1}^N a_\ell^* (X_\ell - E(X_\ell))^* \right) \\ &= \sum_{k=1}^N \sum_{m=1}^N a_k E((X_k - E(X_k))(X_m - E(X_m))^*) a_m^* \\ &= \sum_{k=1}^N \sum_{m=1}^N a_k c_{km} a_m^* = \mathbf{x}^T \mathbf{C} \mathbf{x} \end{aligned}$$

comme  $E(|Y|^2) \geq 0$  alors  $\mathbf{x}^T \mathbf{C} \mathbf{x} \geq 0$  pour tout  $\mathbf{x} \in \mathbb{R}^n$ . □

## 2.4.2 Suite blanche

**Définition 2.4.2** Soit  $X_1, \dots, X_n$   $n$  variables aléatoires. On dit qu'elles forment une suite blanche si on a à la fois

- i)  $\text{var}(X_i) = \sigma^2 = \text{constante}$ ,
- ii) et  $\text{cov}(X_i, X_j) = 0$  pour  $i \neq j$ .

Leur matrice de covariance s'écrit donc :

$$\mathbf{C} = \sigma^2 I_n$$

où  $I_n$  est la matrice identité de taille  $n \times n$ .

# Chapitre 3

## Variable aléatoire ayant une densité

### 3.1 Densité d'une variable aléatoire et fonction de répartition

Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisé et  $X : \Omega \rightarrow J$  une variable aléatoire.

**Définition 3.1.1** Une fonction  $f : \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R}$  est la densité de la variable  $X$  si et seulement si

1. la fonction  $f$  est positive,
2.  $\int_{\mathcal{D}} f(t)dt = 1$ .

**Définition 3.1.2** Soit  $(\Omega, \mathcal{B}, P)$  un espace probabilisé et  $X : \Omega \rightarrow J$  une variable aléatoire réelle ayant une densité  $f_X$ . On appelle fonction de répartition de  $X$  la fonction réelle de variable réelle définie par

$$x \mapsto F_X(x) = P([X < x]) = \int_{-\infty}^x f_X(t)dt.$$

**Remarque 3.1.1** La mesure de probabilité  $dP$  et la mesure de Lebesgue  $dt$  sont liées par la relation suivante :  $dP(t) = f_X(t)dt$  où  $f_X$  est la densité d'une variable aléatoire  $X$ .

**Propriété 3.1.1** Soit  $X : \Omega \rightarrow J$  une variable aléatoire réelle ayant une densité de probabilité  $f_X$ , alors on a les propriétés suivantes

1. La densité de la variable aléatoire  $X$  est donnée par :

$$f_X(x) = \lim_{h \rightarrow 0} \frac{F_X(x+h) - F_X(x)}{h} = F'_X(x).$$

2. Si la variable  $X$  a une densité  $f_X$  alors la fonction de répartition est

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

En général, on a

$$P(a \leq X \leq b) = \int_a^b f_X(t)dt = F_X(b) - F_X(a).$$

### 3.2 Moments d'ordre $k$ d'une variable aléatoire ayant une densité

Soit  $(\Omega, \tau, P)$  un espace probabilisé et  $X$  une variable aléatoire réelle ayant une densité  $f_X$  de domaine de définition  $\mathcal{D}$ .

**Définition 3.2.1** On appelle **moment** d'ordre  $k$  ( $k \geq 0$ ) de la variable aléatoire  $X$ , noté  $M_k(X)$ , l'intégrale définie par

$$M_k(X) = \int_{\mathcal{D}} t^k f_X(t) dt.$$

Par la suite, on s'intéresse plus volontiers aux moments d'ordres 1 et 2 d'une variable aléatoire par rapport à un réel  $a$  :

$$M_2(X - a) = \int_{\mathcal{D}} (t - a)^2 f_X(t) dt.$$

Et, plus particulièrement lorsque le réel  $a$  est le centre de gravité de la variable aléatoire  $X$ .

### 3.3 Espérance mathématique et variance d'une variable aléatoire ayant densité

**Définition 3.3.1** Soit  $X$  une variable  $X$  à densité de probabilité  $f_X$  de domaine de définition  $\mathcal{D}$ , on a

1. l'espérance mathématique ou la moyenne de la variable aléatoire  $X$ , notée  $\mathbb{E}(X)$ , est le moment d'ordre 1 de  $X$  par rapport à 0 donné par

$$\mathbb{E}(X) = M_1(X) = \int_{\mathcal{D}} t f_X(t) dt.$$

2. la variance de la variable aléatoire  $X$ , notée  $\sigma^2(X) := \text{Var}(X)$ , est le moment d'ordre 2 de  $X$  par rapport à sa moyenne  $\mathbb{E}(X)$  donnée par

$$\sigma^2(X) = M_2(X - \mathbb{E}(X)) = \int_{\mathcal{D}} (t - \mathbb{E}(X))^2 f_X(t) dt.$$

**Propriété 3.3.1** Soit  $X$  une variable  $X$  à densité de probabilité  $f_X$  de domaine de définition  $\mathcal{D}$ , alors on a

$$\sigma^2(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2,$$

où  $\mathbb{E}(X^2) = \int_{\mathcal{D}} t^2 f_X(t) dt.$

**Démonstration.** Pour tout  $t \in \mathcal{D}$ , on a

$$(t - \mathbb{E}(X))^2 = t^2 - 2\mathbb{E}(X)t + (\mathbb{E}(X))^2,$$

alors  $(t - \mathbb{E}(X))^2 f_X(t) = t^2 - 2\mathbb{E}(X)t + (\mathbb{E}(X))^2 f_X(t),$

Par passage à l'intégrale, on obtient

$$\begin{aligned} \int_{\mathcal{D}} (t - \mathbb{E}(X))^2 f_X(t) dt &= \int_{\mathcal{D}} (t^2 f_X(t) - 2\mathbb{E}(X)t f_X(t) + (\mathbb{E}(X))^2 f_X(t)) dt \\ &= \int_{\mathcal{D}} t^2 f_X(t) dt - 2\mathbb{E}(X) \int_{\mathcal{D}} t f_X(t) dt + (\mathbb{E}(X))^2 \int_{\mathcal{D}} f_X(t) dt \\ &= \int_{\mathcal{D}} t^2 f_X(t) dt - 2\mathbb{E}(X)\mathbb{E}(X) + (\mathbb{E}(X))^2 \end{aligned}$$

car  $\mathbb{E}(X) = \int_{\mathcal{D}} t f_X(t) dt$  et  $1 = \int_{\mathcal{D}} f_X(t) dt.$

Finalement, on obtient

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

□

**Exemple 3.3.1** On considère la variable aléatoire  $X$  de densité de probabilité  $f_X$  donnée par

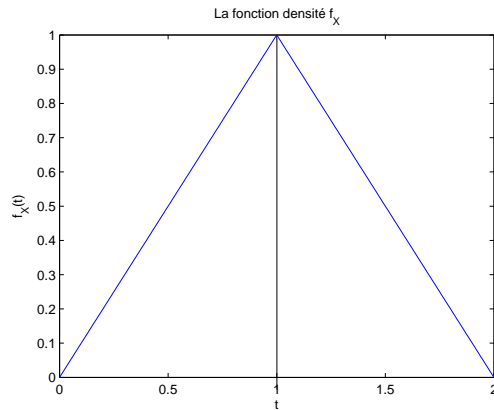
$$f_X(t) \begin{cases} t & \text{si } t \in [0, 1], \\ -t + 2 & \text{si } t \in ]1, 2], \end{cases}$$

$$\mathbb{E}(X) = \int_0^2 t f_X(t) dt = \int_0^1 t^2 dt + \int_1^2 t(2-t) dt = \frac{1}{3} + \left(3 - \frac{7}{3}\right) = 1.$$

$$\mathbb{E}(X^2) = \int_0^2 t^2 f_X(t) dt = \int_0^1 t^3 dt + \int_1^2 t^2(2-t) dt = \frac{1}{4} + \left(\frac{14}{3} - \frac{15}{4}\right) = \frac{7}{6}.$$

$$\text{Var}(X) = \int_{-\infty}^{+\infty} t^2 f_X(t) dt - (\mathbb{E}(X))^2 = \frac{7}{6} - 1^2 = \frac{1}{6} = 0.1667.$$

$L'$  écart-type de la variable aléatoire réelle  $X$  dont la densité  $f_X$  est donnée par



$$\sigma(X) = \sqrt{\text{Var}(X)} = \sqrt{\frac{1}{6}} = 0.4082.$$

### 3.4 Coefficient d'asymétrie et coefficient d'aplatissement

**Définition 3.4.1** Soit  $X$  une variable aléatoire réelle ayant une densité d'espérance  $\xi = \mathbb{E}[X]$  et de variance  $\zeta = \text{Var}(X)$ .

1. On appelle **coefficient d'asymétrie** (skewness) de la variable aléatoire  $X$ , noté  $\gamma_1(X)$ , le coefficient donné par l'expression

$$\gamma_1(X) = \frac{1}{\zeta^3} \int_{\mathcal{D}} (t - \xi)^3 f_X(t) dt.$$

2. On appelle **coefficient d'aplatissement** (kurtosis) de la variable aléatoire  $X$ , noté  $\gamma_2(X)$ , le coefficient donné par l'expression

$$\gamma_2(X) = \frac{1}{\zeta^4} \int_{\mathcal{D}} (t - \xi)^4 f_X(t) dt.$$





# Chapitre 4

## Lois de probabilité usuelles : discrètes et continues

### 4.1 Loi conjointe et lois marginales de variables aléatoires réelles

**Définition 4.1.1** 1. Nous choisirons systématiquement comme base  $\mathcal{A}$  la base canonique de  $\mathbb{R}^n$  et nous écrivons

$$X = (X_1, X_2, \dots, X_n)$$

2. La loi de probabilité de  $X$  est appelée la **loi conjointe du système**  $(X_1, X_2, \dots, X_n)$ . Les lois des variables aléatoires  $X_i$  sont appelées les **lois marginales du système**.

#### 4.1.1 Exemple d'un couple de variables aléatoires réelles

Le cas d'un système de deux variables aléatoires réelles est particulièrement simple. Illustrons-le par un exemple.

Soit  $\Omega = \{a, b, c, d, e\}$  et soit  $\mathcal{B}$  une tribu de parties de  $\Omega$  dont les atomes sont  $\{a, b\}$ ,  $\{c\}$ ,  $\{d\}$ ,  $\{e\}$ . On donne deux variables aléatoires réelles  $X_1$  et  $X_2$  par le tableau suivant :

$\omega$	$a$	$b$	$c$	$d$	$e$
$X_1(\omega)$	0	0	1	1	2
$X_2(\omega)$	0	0	1	2	1

et on considère le couple de variables aléatoires réelles  $X = (X_1, X_2)$ .

Donnons-nous une probabilité  $P$  en posant

$$P(\{a, b\}) = \frac{1}{2}, \quad P(\{c\}) = P(\{d\}) = P(\{e\}) = \frac{1}{6}.$$

Il est commode de présenter  $X(\Omega)$  en un tableau dont chaque case représente un élément de  $X(\Omega)$ .

**Ainsi pourra-t-on présenter la loi conjointe en indiquant dans la case  $(i, j)$  la probabilité**

$$P([X = (i, j)]).$$

Déterminons cette loi

D'où la loi conjointe, présentée en tableau

$X_2 \backslash X_1$	0	1	2
0			
1			
2			

Valeurs de $X$ ( $i, j$ )	éventualités formant $X = (i, j)$	Probabilité de $X = \{i, j\}$
(0, 0)	$a, b$	$\frac{1}{2}$
(0, 1)	$\emptyset$	0
(0, 2)	$\emptyset$	0
(1, 0)	$\emptyset$	0
(1, 1)	$c$	$\frac{1}{2}$
(1, 2)	$d$	$\frac{1}{2}$
(2, 0)	$\emptyset$	0
(2, 1)	$e$	$\frac{1}{6}$
(2, 2)	$\emptyset$	0

#### 4.1.2 Déterminons maintenant les lois marginales

$$\begin{aligned}
P([X_1 = 0]) &= P(\{a, b\}) = \frac{1}{2} \\
P([X_1 = 1]) &= P(\{c, d\}) = \frac{1}{3} \\
P([X_1 = 2]) &= P(\{e\}) = \frac{1}{6} \\
P([X_2 = 0]) &= P(\{a, b\}) = \frac{1}{2} \\
P([X_2 = 1]) &= P(\{c, e\}) = \frac{1}{3} \\
P([X_2 = 2]) &= P(\{d\}) = \frac{1}{6}
\end{aligned}$$

On peut remarquer que ces probabilités se trouvent sur le tableau, si on effectue, **dans les marges**, les sommes des probabilités inscrites dans une même ligne ou colonne (d'où le nom **loi marginale**).

Quoi d'étonnant à cela puisque, par exemple

$$[X_1 = 1] = [X = (1, 0)] \cup [X = (1, 1)] \cup [X = (1, 2)]$$

et que les trois événements ainsi réunis sont deux à deux disjoints.

$X_2 \backslash X_1$	0	1	2	
0	$\frac{1}{2}$	0	0	$\rightarrow$ total $\frac{1}{2}$
1	0	$\frac{1}{6}$	$\frac{1}{6}$	$\rightarrow$ total $\frac{1}{3}$
2	0	$\frac{1}{6}$	0	$\rightarrow$ total $\frac{1}{6}$
	$\downarrow$ total $\frac{1}{2}$	$\downarrow$ total $\frac{1}{3}$	$\downarrow$ total $\frac{1}{6}$	

## 4.2 Loi uniforme continue sur $(a, b)$

**Définition 4.2.1** Une variable aléatoire  $X$  est dite **uniforme** ou **équirépartie** sur  $(a, b)$  avec  $b > a$ , si sa densité de probabilité a pour expression :

$$f_X(t) = \begin{cases} \frac{1}{b-a}, & \text{si } t \in (a, b), \\ 0, & \text{sinon.} \end{cases}$$

On peut calculer l'espérance mathématique ou la moyenne de cette variable aléatoire, sa variance et son écart-type :

1. L'espérance mathématique :

$$E(X) = \int_a^b t f_X(t) dt = \int_a^b \frac{t}{b-a} dt = \frac{a+b}{2}.$$

2. La variance :

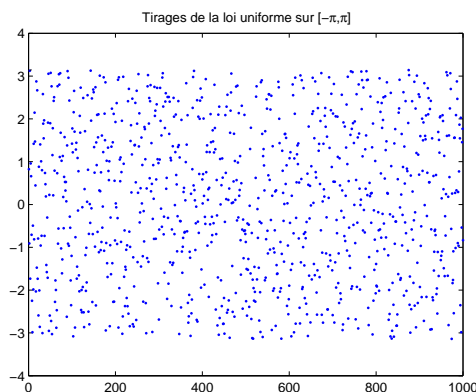
$$E(X^2) = \int_a^b t^2 f_X(t) dt = \int_a^b \frac{t^2}{b-a} dt = \frac{a^2 + ab + b^2}{3}.$$

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

3. L'écart-type :

$$\sigma(X) = \sqrt{\text{Var}(X)} = \sqrt{\frac{(b-a)^2}{12}} = \frac{b-a}{2\sqrt{3}}.$$

4. Distribution d'un nuage de points par une loi uniforme :



## 4.3 Loi de Poisson

**Définition 4.3.1** Soit  $(\Omega, \tau, P)$  un espace probabilisé.

Une variable aléatoire  $X$  a une loi de Poisson de paramètre  $\lambda$  réel positif si les valeurs de cette variable sont les entiers positifs et si les probabilités associées sont :

$$P(k; \lambda) = P([X = k]) = e^{-\lambda} \frac{\lambda^k}{k!}$$

C'est-à-dire que

$$i) X : \Omega \longrightarrow \mathbb{N}$$

ii) Pour tout  $k \geq 0$ , la probabilité de l'événement  $[X = k]$  est

$$P(k; \lambda) = P([X = k]) = e^{-\lambda} \frac{\lambda^k}{k!}$$

**Propriété 4.3.1** Soit  $(\Omega, \tau, P)$  un e.p. et  $X$  une variable aléatoire ayant la loi de Poisson de paramètre  $\lambda$  réel positif. Alors on a

1.  $\mu = E[X] = \lambda$ ,
2.  $\text{Var}(X) = \sigma^2(X) = \lambda$ .

## 4.4 Loi normale $\mathcal{N}(\mu, \sigma^2)$ ou v.a. gaussienne

**Définition 4.4.1** 1. Une variable aléatoire  $X$  a une loi normale réduite si elle a une densité  $\varphi$  définie par :

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

On dit que c'est la loi normale  $\mathcal{N}(0, 1)$ . La fonction de répartition  $\Pi$  associée à la variable  $X$  est définie par

$$\Pi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt,$$

où  $E[X] = 0$  et  $\sigma(X) = 1$  est son écart-type.

2. Une variable aléatoire  $Y$  a une loi normale non-réduite si elle a une densité  $\psi$  définie par :

$$\psi(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu)^2/2\sigma^2}.$$

On dit que c'est la loi normale  $\mathcal{N}(\mu, \sigma^2)$ . La fonction de répartition  $\Pi$  associée à la variable  $X$  est définie par

$$\Pi(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-(t-\mu)^2/2\sigma^2} dt,$$

où  $\mu = E[Y]$  et  $\sigma$  est son écart-type.

## 4.5 Valeurs numériques de $\Pi(x)$

$x$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\Pi(x)$	0.5	0.54	0.579	0.618	0.655	0.692	0.726	0.758	0.788	0.816
$x$	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
$\Pi(x)$	0.841	0.864	0.885	0.903	0.919	0.933	0.945	0.955	0.964	0.971
$x$	2.0	2.2	2.4	2.6	2.8	3.0	3.3	3.6	4.0	4.5
$\Pi(x)$	0.977	0.986	0.992	0.995	0.997	0.9986	0.9995	0.9998	0.99996	0.999997

**Exercice 4.5.1** 1. Montrer que  $\Pi(x) + \Pi(-x) = 1$ .

2. Soit  $Y$  la loi normale  $\mathcal{N}(\mu, \sigma^2)$ . Montrer que  $X = \frac{(Y-\mu)}{\sigma}$  est une variable aléatoire réduite.
3. Montrer que la fonction de densité  $f$  et la fonction de répartition  $F$  associées sont données par :

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}, \quad F(y) = \Pi((y-\mu)/\sigma).$$

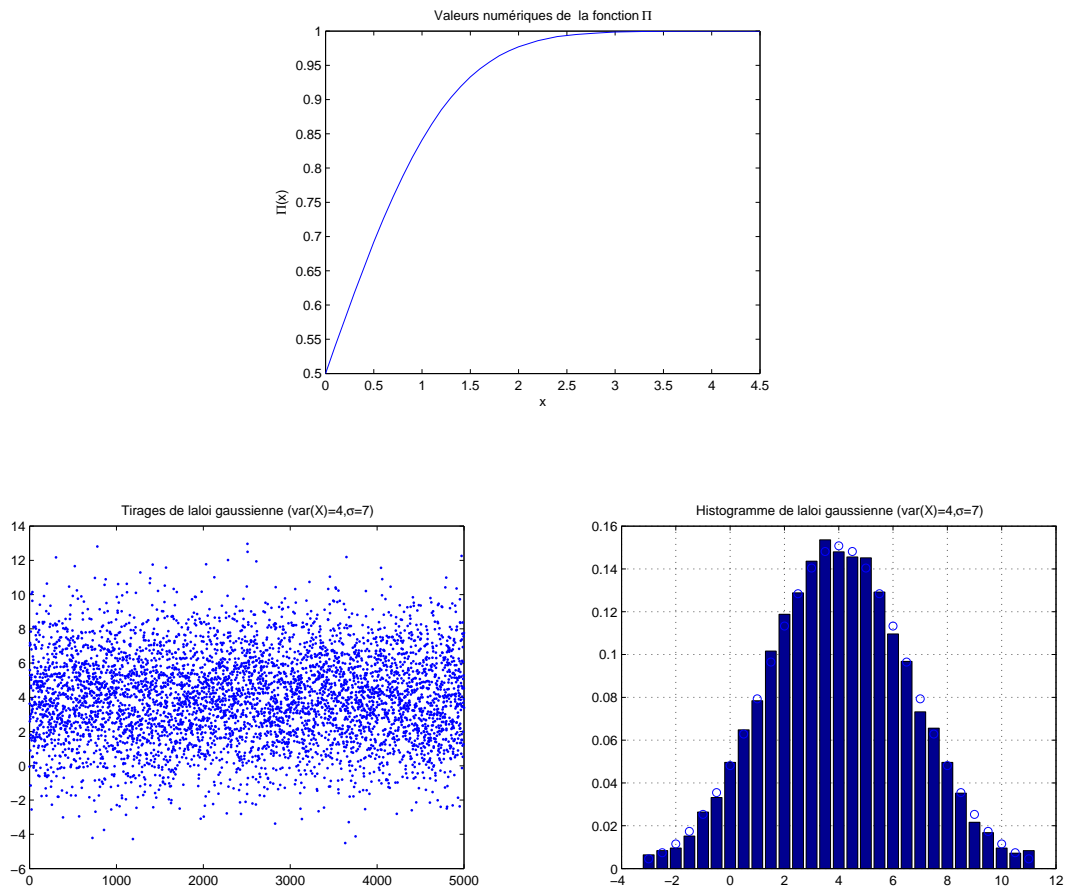


FIGURE 4.1 – Ici  $\mu = E(X)$  et  $\sigma = \sqrt{\sigma^2}$ .

## 4.6 Loi Gamma et loi Khi-deux : $\chi^2$

### 4.6.1 Fonction Gamma $\Gamma$ :

Soit  $z \in \mathbb{C}$  un nombre complexe, on définit pour tout réel positif  $x$ , la fonction puissance  $x \mapsto x^z$  par

$$x^z = e^{z \ln(x)}$$

**Lemme 4.6.1** Soit  $z$  un nombre complexe. La fonction  $t \mapsto t^{z-1}e^{-t}$  est intégrable sur  $]0, +\infty[$  si et seulement si  $\text{Re}(z) > 0$ .

**Démonstration.** La fonction  $t \mapsto t^{z-1}e^{-t}$  est continue sur  $]0, +\infty[$  pour nombre complexe  $z$ . Le lemme se démontre à l'aide des théorèmes de comparaisons du calcul intégral et des formules suivantes

$$|t^{z-1}e^{-t}| = t^{\text{Re}(z)-1}e^{-t} \sim_{t \rightarrow 0^+} t^{\text{Re}(z)-1} \text{ et } |t^{z-1}e^{-t}| = t^{\text{Re}(z)-1}e^{-t} =_{t \rightarrow +\infty} o(e^{-t/2}).$$

□

**Définition 4.6.1** On appelle la fonction **Gamma**, notée  $\Gamma$ , la fonction  $\Gamma : \{z \in \mathbb{C} / \text{Re}(z) > 0\} \rightarrow \mathbb{C}$  définie par l'expression intégrale suivante :

$$\Gamma(z) = \int_0^{+\infty} t^{z-1}e^{-t} dt.$$

D'une façon analogue, on définit la fonction  $\Gamma_a$  par l'expression suivante :

$$\Gamma_a(z) = \int_0^\alpha t^{z-1} e^{-t} dt.$$

**Exemple 4.6.1** Pour  $\alpha > 0$  on a

$$\Gamma_a \alpha(1) = \int_0^\alpha t^{1-1} e^{-t} dt = [-e^{-t}]_0^\alpha = 1 - e^{-\alpha} \rightarrow 1$$

lorsque  $\alpha \rightarrow +\infty$ . D'où  $\Gamma(1) = 1$ .

**Proposition 4.6.1** on a  $\Gamma(z+1) = z\Gamma(z)$ .

**Démonstration.** On a  $\Gamma_\alpha(z+1) = \int_0^\alpha t^z e^{-t} dt$ . Par une intégration par partie, on a

$$\Gamma_\alpha(z+1) = \int_0^\alpha t^{z-1} e^{-t} dt = -\alpha^z e^{-\alpha} + z\Gamma_\alpha(z).$$

Ce qui implique

$$|\Gamma_\alpha(z+1) - z\Gamma_\alpha(z)| \leq \alpha^{\operatorname{Re}(z)} e^{-\alpha} \rightarrow 0$$

lorsque  $\alpha \rightarrow +\infty$ . D'où

$$\Gamma(z+1) = z\Gamma(z)$$

□

**Corollaire 4.6.1** Pour  $n \in \mathbb{N}^*$ , on a  $\Gamma(n+1) = n!$ .

**Démonstration.** Pour  $n \in \mathbb{N}^*$ , on a  $\Gamma(n+1) = n\Gamma(n)$ ,  $\Gamma(n) = (n-1)\Gamma(n-1)$ ,  $\dots$  et par récurrence on obtient  $\Gamma(2) = 1$ . D'où

$$\Gamma(n) = n \times (n-1) \dots 2 \times 1 = n!$$

□

## 4.7 Loi Gamma

**Proposition 4.7.1** Soit  $a$  et  $\theta$  des réels strictement positifs, alors la fonction réelle  $f$  de variable réelle  $x$  définie par

$$f(x) = \begin{cases} \frac{1}{\Gamma(a)} \theta^a x^{a-1} e^{-\theta x}, & \text{si } x > 0, \\ 0, & \text{sinon.} \end{cases}$$

est une densité de probabilité d'une variable aléatoire  $X$ .

**Démonstration.** Pour tout  $a$  et  $\theta > 0$ , pour  $\xi = \theta x$  on a

$$\int_{-\infty}^{+\infty} \theta^a x^{a-1} e^{-\theta x} 1_{\{x>0\}} dx = \int_0^{+\infty} \theta^a x^{a-1} e^{-\theta x} dx = \int_0^{+\infty} y^{a-1} e^{-y} dy = \Gamma(a),$$

d'où le résultat.

□

**Définition 4.7.1** On dit qu'une variable aléatoire  $X$  suit la **loi Gamma de paramètres**  $a$  et  $\theta$  où  $a$  et  $\theta > 0$  et on note  $X \sim \Gamma(a, \theta)$  si  $X$  possède la densité de probabilité

$$f(x) = \begin{cases} \frac{1}{\Gamma(a)} \theta^a x^{a-1} e^{-\theta x}, & \text{si } x > 0, \\ 0, & \text{sinon.} \end{cases}$$

En particulier, pour  $a = 1$  alors la variable aléatoire  $X \sim \Gamma(1, \theta)$  n'est que la loi exponentielle de paramètre  $\theta$ .

## 4.8 Loi Beta

**Définition 4.8.1** Soient  $z$  et  $w$  deux variables complexes telles que  $\operatorname{Re}(z) > 0$  et  $\operatorname{Re}(w) > 0$ . On appelle fonction **bêta**, la fonction notée  $\beta$ , définie par

$$\beta(z, w) = \int_0^1 t^{z-1} (1-t)^{w-1} dt$$

**Proposition 4.8.1** Pour  $p > 0$  et  $q > 0$ , on a  $\beta(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ .

**Démonstration.** On note  $V = \{(x, y) \in \mathbb{R}^2 / x > 0, y > 0\}$ . On a

$$\Gamma(p)\Gamma(q) = \int_V x^{p-1} y^{q-1} e^{-x-y} dx dy$$

Soit  $(\rho, t) \in ]0, +\infty[ \times ]0, 1[ = U$  et soit  $\Phi : U \rightarrow V$ ,  $\Phi(\rho, t) = (x, y)$  où  $x = \rho t$  et  $y = \rho(1-t)$ , alors  $U$  est la boule ouverte. Le jacobien de  $\Phi$  est :  $\operatorname{Jac}(\Phi) = -\rho \neq 0$ .

Ce qui montre que  $\Phi$  est un difféomorphisme de  $U$  dans  $V$ . Donc,

$$\begin{aligned} \int_V x^{p-1} y^{q-1} e^{-x-y} dx dy &= \int_U \rho^{p+q-2} t^{p-1} (1-t)^{q-1} e^{-\rho} \times |-\rho| d\rho dt \\ &= \int_0^1 t^{p-1} (1-t)^{q-1} dt \int_0^{+\infty} \rho^{p+q-1} e^{-\rho} d\rho \end{aligned}$$

D'où

$$\Gamma(p)\Gamma(q) = \Gamma(p+q)\beta(p, q), \quad p > 0, \quad q > 0$$

□

**Proposition 4.8.2** Soient  $p$  et  $q$  des réels strictement positifs, alors la fonction réelle  $f$  de variable réelle  $x$  définie par

$$f(x) = \begin{cases} \frac{1}{\beta(p, q)} x^{p-1} (1-x)^{q-1}, & \text{si } 0 \leq x \leq 1, \\ 0, & \text{sinon.} \end{cases}$$

est une densité de probabilité d'une variable aléatoire  $X$ .

**Démonstration.** Pour tout  $a$  et  $\theta > 0$ , pour  $\xi = \theta x$  on a

$$\int_{-\infty}^{+\infty} f(x) dx = \frac{1}{\beta(p, q)} \int_0^1 x^{p-1} (1-x)^{q-1} dx = \frac{1}{\beta(p, q)} \beta(p, q) = 1,$$

d'où le résultat.

□

**Définition 4.8.2** On dit qu'une variable aléatoire  $X$  suit la **loi Beta** de paramètre  $p > 0$  et  $q > 0$  et on note  $X \sim \beta(p, q)$  si  $X$  possède la densité de probabilité

$$f(x) = \begin{cases} \frac{1}{\beta(p, q)} x^{p-1} (1-x)^{q-1}, & \text{si } 0 \leq x \leq 1, \\ 0, & \text{sinon.} \end{cases}$$

## 4.9 Loi du Khi-2 où $\chi^2$

**Définition 4.9.1** Soient  $X_1, \dots, X_n$  une suite de  $n$  variables aléatoires normales centrées réduites indépendantes.

On appelle loi du **Khi-2** à  $n$  degrés de liberté, notée  $\chi^2(n)$  la loi de la variables aléatoire  $X = \sum_{i=1}^n X_i^2$ .

**Proposition 4.9.1** Soit  $X$  une variable aléatoire qui suit la loi du  $\chi^2(n)$  à  $n$  degrés de libertés. Alors

1. la loi de la variable aléatoire  $\chi^2(n)$  est la loi  $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$  de densité  $f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} 1_{]0, +\infty[}(x)$ .
2. On a  $\mathbb{E}[\chi^2(n)] = n$  et  $\sigma^2(\chi^2(n)) = 2n$ .

## 4.10 Lois de Student et de Fisher-Snedecor

**Définition 4.10.1** Soient  $X$  une variable aléatoire normale centrée réduite et  $Y$  une variable aléatoire de loi  $\chi^2(n)$  indépendante de  $X$ . On appelle loi de **Student** à  $n$  degrés de libertés, notée  $t(n)$ , la loi de la variable aléatoire  $T_n = \frac{X}{\sqrt{\frac{Y}{n}}}$ .

**Proposition 4.10.1** 1. la loi de la variable aléatoire  $t(n)$  est la loi de densité  $f(t) = \frac{\Gamma\left(\frac{(n+1)}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{n\pi}} \frac{1}{\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}$ .

2. L'espérance et la variance d'une variable aléatoire de Student sont

$$\mathbb{E}[T_n] = 0 \quad \text{pour tout } n > 1 \quad \text{et} \quad \sigma^2(T_n) = \frac{n}{n-2} \quad \text{pour tout } n > 2.$$

**Définition 4.10.2** Soient  $X$  et  $Y$  deux variables aléatoires indépendantes de lois  $\chi_n^2$  et  $\chi_p^2$ . On appelle loi de **Fisher** de paramètres  $n$  et  $p$ , notée  $F_{n,p}$ , la loi de la variable  $F = \frac{\frac{X}{n}}{\frac{Y}{p}}$ .

L'espérance et la variance d'une variable aléatoire de loi  $F_{n,p}$  sont

$$E(F) = \frac{p}{p-2} \quad \text{pour tout } p > 2,$$

$$\sigma^2(F) = \frac{2p^2(n+p-2)}{n(p-2)^2(p-4)} \quad \text{pour tout } p > 4.$$

## 4.11 Quelques lois de probabilités



Loi	Type	Probabilité ou Densité de probabilité(ddd)	Moyenne	Variance
Bernoulli	D	$P([X = 0]) = 1 - p$ et $P([X = 1]) = p$	$p$	$p(1 - p)$
Uniforme	D	$P([X = i]) = \frac{1}{n}$ pour $1 \leq i \leq n$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Binomiale	D	$P([X = i]) = C_n^i p^i (1-p)^{n-i}$ pour $1 \leq i \leq n$	$np$	$np(1-p)$
Géométrique	D	$P([X = i]) = p(1-p)^{i-1}$ pour $i = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Pascal	D	$P([X = i]) = C_{i-1}^{n-1} p^n (1-p)^{i-n}$	$\frac{n}{p}$	$\frac{n(1-p)}{p^2}$
Poisson	D	$P([X = i]) = \frac{\lambda^i}{i!} e^{-\lambda}$ pour $\lambda > 0$ et $i = 1, 2, \dots$	$\lambda$	$\lambda$
Uniforme	C	$f(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x)$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Gauss	C	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ pour $x \in \mathbb{R}$	$\mu$	$\sigma^2$
Cauchy	C	$f(x) = \frac{a}{\pi(a^2+x^2)}$ pour $x \in \mathbb{R}$	non défini	non défini
Gamma	C	$f(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}$ pour $x > 0$	$\frac{k}{\lambda}$	$\frac{k}{\lambda^2}$
Exponentielle	C	$f(x) = \frac{1}{a} x^{k-1} e^{-\frac{x}{a}}$ pour $x > 0$ et $a > 0$	$a$	$a^2$
Rayleigh	C	$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ pour $x > 0$	$\sigma \sqrt{\frac{\pi}{2}}$	$\sigma^2(1 - \frac{\pi}{2})$
Laplace	C	$f(x) = \frac{a}{2} e^{-a x }$	0	$\frac{2}{a^2}$
$\chi^2$	C	$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbb{I}_{]0,+\infty[}(x)$	$n$	$2n$
Student	C	$f(x) = \frac{\frac{n+1}{2}}{\sqrt{n\pi} \Gamma(\frac{n}{2}) \left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}$	0	$\frac{n}{n-2}$ pour $n > 2$
Weibull	C	$f(x) = \beta x^{\beta-1} e^{-x^\beta}$	$\Gamma(1 + \frac{1}{\beta})$	$\Gamma(1 + \frac{2}{\beta}) - (E(X))^2$

TABLE 4.1 – Type :  $D \equiv$  loi discrète ;  $C \equiv$  loi continue



# Chapitre 5

## Fonctions caractéristiques de variables aléatoires

Dans ce chapitre, on désigne par  $\mu$  une mesure de probabilité sur un espace probabilisable  $(\Omega, \tau)$ .

### 5.1 Fonctions génératrices de v. a. entières

Nous allons nous concentrer pour un moment sur les variables à valeurs dans l'ensemble  $\mathbb{N}$  des entiers  $\geq 0$  : Dans ce cas les moments seront plus faciles à calculer grâce à l'introduction de la notion de fonction génératrice de  $X$  :

**Définition 5.1.1** Soit  $(\Omega, \tau, \mu)$  un espace probabilisé et  $X : \Omega \longrightarrow \mathbb{N}$  une variable aléatoire. On appelle la **fonction génératrice** de la variable aléatoire  $X$ , la fonction  $g$  définie par

$$g(x) = \sum_{i=0}^{+\infty} p_i x^i,$$

où  $p_i = \mu(A_i)$  est la probabilité de l'événement  $A_i = [X = i]$  pour tout  $i \in \mathbb{N}$ . La donnée de la fonction  $g$  est équivalente à la donnée des  $p_i$  et définit la loi de  $X$ .

**Proposition 5.1.1** La série de terme général  $u_i = p_i x^i$  est absolument et uniformément convergente dans  $[-1, 1]$ .

**Démonstration.** Les probabilités  $p_i$  sont des nombres positifs ; par suite, la série de terme général :

$$\|p_i x^i\| = \sup_{-1 \leq x \leq 1} |p_i x^i| = p_i$$

est convergente. D'où la série de terme général  $p_i x^i$  est alors uniformément et absolument convergente dans  $[-1; 1]$ .

□

### 5.2 Fonction caractéristique d'une variable aléatoire réelle

#### 5.2.1 Définitions et notations

Ici, nous donnons une définition générale pour tous les types de variables aléatoires.

**Définition 5.2.1** Soit  $(\Omega, \tau, \mu)$  un espace probabilisé et  $X : \Omega \longrightarrow X(\Omega)$  une variable aléatoire. On appelle **fonction caractéristique** de  $X$ , notée  $\varphi_X$ , toute fonction de variable réelle  $t \in \mathbb{R}$  définie par

$$\varphi_X(t) = \mathbb{E}(e^{itX}).$$

### 5.2.2 Fonction caractéristique d'une v. a. réelle discrète

Soit  $(\Omega, \tau, \mu)$  un espace probabilisé et  $X : \Omega \longrightarrow \{x_k, k \in \mathbb{N}\}$  une variable aléatoire à valeurs discrètes. Soit  $(p_k = \mu([X = x_k]))_{k \in \mathbb{N}}$  la loi de la variable aléatoire (v.a.)  $X$ .

**Définition 5.2.2** On appelle **fonction caractéristique** de la v.a.  $X$  la fonction  $\varphi_X$  définie par :

$$\varphi_X : \mathbb{R} \longrightarrow \mathbb{C}, \quad t \longmapsto \varphi(t) = \sum_{k=0}^{+\infty} p_k e^{itx_k}.$$

**Propriété 5.2.1** Nous avons les propriétés suivantes :

1.  $\varphi$  est définie pour tout réel  $t$  et :

$$\varphi(0) = 1 \quad \text{et} \quad |\varphi(t)| \leq 1.$$

2.  $\varphi$  est uniformément continue sur  $\mathbb{R}$ .

3. Si  $\mathbb{E}[X] < +\infty$  alors la fonction  $\varphi$  est dérivable et :

$$\varphi'(0) = i\mathbb{E}[X].$$

4. Si de plus  $x_k$  sont les entiers positifs alors on a

$$\varphi(t) = g(e^{it})$$

où  $g$  est la fonction génératrice de  $X$ .

### 5.2.3 Fonction caractéristique d'une v. a. réelle ayant une densité

**Définition 5.2.3** Soit  $(\Omega, \tau, \mu)$  un espace probabilisé et  $X : \Omega \longrightarrow \mathbb{R}$  une variable aléatoire continue ayant une densité de probabilité  $f$ . La fonction caractéristique  $\varphi_X$  de la v.a.  $X$  est définie par

$$\varphi_X(t) = \int_{\mathcal{D}} e^{itx} d\mu(x) = \int_{\mathcal{D}} e^{itx} f(x) dx,$$

où  $\mathcal{D}$  est le domaine de définition de  $f$ .

**Exemple 5.2.1** On considère la variable aléatoire  $X$  dont la densité  $f$  donnée par

$$f(x) = \begin{cases} x, & \text{si } x \in [0, 1], \\ -x + 2, & \text{si } x \in ]1, 2], \end{cases}$$

$$\varphi(t) = \int_0^2 e^{itx} f(x) dx = \int_0^1 x e^{itx} dx + \int_1^2 e^{itx} (2 - x) dx.$$

## 5.3 Mesure de probabilité sur $\mathbb{R}^d$

Dans cette section, on s'intéresse à la mesure de probabilité, encore notée  $\mu$ , sur un espace probabilisable  $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d))$  où  $d \geq 1$  et  $\mathbb{B}(\mathbb{R}^d)$  est la tribu borélienne sur  $\mathbb{R}^d$ .

**Définition 5.3.1** La mesure  $\mu$  est dite une probabilité sur  $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d))$  si et seulement si elle vérifie les conditions suivantes

i)  $\mu$  est une mesure finie,

ii)  $\mu(\mathbb{R}^d) = 1$ .

Pour tout  $x$  et  $y$  dans  $\mathbb{R}^d$  on note par

$$(x, y) = \sum_{i=1}^d x_i y_i$$

le produit scalaire euclidien sur  $\mathbb{R}^d$ .

### 5.3.1 Notions d'intégration d'une fonction à valeurs complexes

Soit  $f : \mathbb{R}^d \rightarrow \mathbb{C}$  une application. Pour tout  $x \in \mathbb{R}$ , on note par  $\operatorname{Re} f(x)$  la partie réelle et  $\operatorname{Im} f(x)$  la partie imaginaire de  $f(x)$  qui sont définies par

$$\operatorname{Re} f(x) = \frac{1}{2} \left( f(x) + \overline{f(x)} \right) \quad \text{et} \quad \operatorname{Im} f(x) = \frac{1}{2i} \left( f(x) - \overline{f(x)} \right)$$

respectivement.

**Définition 5.3.2** Soit  $f : \mathbb{R}^d \rightarrow \mathbb{C}$  une application. On dit que  $f$  est  $\mu$ -intégrable sur  $\mathbb{R}^d$  si l'application  $|f| : \mathbb{R}^d \rightarrow \mathbb{R}$  est intégrable par rapport à la mesure  $\mu$ .

Dans ce cas, les fonctions  $\operatorname{Re} f : \mathbb{R}^d \rightarrow \mathbb{R}$  et  $\operatorname{Im} f : \mathbb{R}^d \rightarrow \mathbb{R}$  sont  $\mu$ -intégrables et on a

$$\int_{\mathbb{R}^d} f(x) d\mu(x) = \int_{\mathbb{R}^d} \operatorname{Re} f(x) d\mu(x) + i \int_{\mathbb{R}^d} \operatorname{Im} f(x) d\mu(x).$$

**Proposition 5.3.1** Soit  $f : \mathbb{R}^d \rightarrow \mathbb{C}$  et  $g : \mathbb{R}^d \rightarrow \mathbb{C}$  deux applications  $\mu$ -intégrables. Alors, pour tout  $(\alpha, \beta) \in \mathbb{C}^2$  on a

$$\begin{aligned} i) \quad & \int_{\mathbb{R}^d} (\alpha f(x) + \beta g(x)) d\mu(x) = \alpha \int_{\mathbb{R}^d} f(x) d\mu(x) + \beta \int_{\mathbb{R}^d} g(x) d\mu(x), \\ ii) \quad & \left| \int_{\mathbb{R}^d} f(x) d\mu(x) \right| \leq \int_{\mathbb{R}^d} |f(x)| d\mu(x). \end{aligned}$$

**Démonstration.**

□

### 5.3.2 Fonction caractéristique d'une mesure de probabilité

**Définition 5.3.3** Soit  $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d), \mu)$  un espace probabilisé. On appelle **fonction caractéristique** de  $\mu$ , notée  $\varphi_\mu$ , l'application  $\varphi_\mu : \mathbb{R}^d \rightarrow \mathbb{C}$  définie pour tout  $x \in \mathbb{R}^d$  par

$$\varphi_\mu(x) = \int_{\mathbb{R}^d} e^{i(x,\xi)} d\mu(\xi).$$

La fonction  $\varphi_\mu$  est appelée aussi la **Transformée de Fourier** de  $\mu$ . Et, elle est bien définie car

$$|\varphi_\mu(x)| \leq \int_{\mathbb{R}^d} |e^{i(x,\xi)}| d\mu(\xi) = \int_{\mathbb{R}^d} d\mu(\xi) = \mu(\mathbb{R}^d) = 1$$

puisque  $|e^{i(x,\xi)}| = 1$ .

**Définition 5.3.4** Soit  $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d), \mu)$  un espace probabilisé. On appelle **transformée de Fourier** de  $\mu$ , notée  $\phi_\mu$ , l'application  $\phi_\mu : \mathbb{R}^d \rightarrow \mathbb{C}$  définie pour tout  $x \in \mathbb{R}^d$  par

$$\phi_\mu(x) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{i(x,\xi)} d\mu(\xi).$$

Lorsque  $\mu$  possède une densité de probabilité  $f$  par rapport à la mesure de Lebesgue  $dx$ , on dit que  $\phi_\mu$  est l'image de Fourier de  $f$  notée

$$\hat{f}(x) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{i(x,\xi)} f(\xi) d\xi.$$

**Remarque 5.3.1** On remarque que l'on a la relation suivante  $(2\pi)^{d/2} \phi_\mu(x) = \varphi_\mu(x)$ , pour tout  $x \in \mathbb{R}^d$ .

**Propriété 5.3.1** Soit  $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d), \mu)$  un espace probabilisé et  $\varphi_\mu : \mathbb{R}^d \rightarrow \mathbb{C}$  la fonction caractéristique de  $\mu$ . Alors

i)  $\varphi_\mu(0) = 1$  et  $|\varphi_\mu(x)| \leq 1$ ,

ii) pour tout  $x \in \mathbb{R}^d$  on a  $\varphi_\mu(-x) = \overline{\varphi_\mu(x)}$

iii) si  $\mu = \bigotimes_{k=1}^d \mu_k$  où  $(\mu_k)_{1 \leq k \leq d}$  sont des mesures de probabilités sur  $\mathbb{R}$ , alors on a pour tout  $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$  on a

$$\varphi_\mu(x) = \varphi_{\mu_1}(x_1) \otimes \varphi_{\mu_2}(x_2) \otimes \dots \otimes \varphi_{\mu_d}(x_d).$$

**Démonstration.**

□

**Proposition 5.3.2** L'application  $\varphi_\mu : \mathbb{R}^d \rightarrow \mathbb{C}$  est uniformément continue sur  $\mathbb{R}^d$ .

**Démonstration.**

□

**Proposition 5.3.3** Soit  $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d), \mu)$  un espace probabilisé, soit  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  une transformation affine,  $T(x) = A(x) + b$  où  $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$  est un endomorphisme et  $b \in \mathbb{R}^d$ . Si  $\mu_T$  désigne la mesure image de  $\mu$  par  $T$ , alors pour tout  $x \in \mathbb{R}^d$  on a

$$\varphi_{\mu_T}(x) = e^{i(x,b)} \varphi_\mu(A^T(x))$$

où  $A^T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  est l'endomorphisme transposé de  $A$ .

**Démonstration.**

□

**Proposition 5.3.4** Soit  $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d), \mu)$  un espace probabilisé et  $n \in \mathbb{N}$  un entier naturel tel que  $\int_{\mathbb{R}^d} \|x\|^n d\mu(x) < +\infty$ , alors

1. la fonction caractéristique  $\varphi_\mu$  de  $\mu$  est de classe  $C^n$  sur  $\mathbb{R}^d$  et pour tout  $\alpha \in \mathbb{N}^d$  tel que  $|\alpha| \leq n$  on a

$$(\mathcal{D}^\alpha \varphi_\mu)(0) = i^{|\alpha|} M_\alpha$$

$$\text{où } M_\alpha = \int_{\mathbb{R}^d} x^\alpha d\mu(x).$$

2. La fonction caractéristique  $\varphi_\mu$  admet un développement limité de Mac-Laurin en 0 à l'ordre  $n$  donné par

$$\varphi_\mu(x) = \sum_{|\alpha| \leq n} \frac{i^\alpha M_\alpha}{\alpha!} x^\alpha + \|x\|^n \varepsilon_n(x),$$

$$\text{où } \alpha! = \alpha_1! \dots \alpha_d! \text{ et } \lim_{\|x\| \rightarrow 0} \varepsilon_n(x) = 0.$$

**Démonstration.**

□

**Proposition 5.3.5** Soit  $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d), \mu)$  un espace probabilisé et  $\nu$  une autre mesure de probabilité sur  $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d))$ , alors on a

$$\int_{\mathbb{R}^d} \varphi_\mu(x) d\nu(\xi) = \int_{\mathbb{R}^d} \varphi_\nu(x) d\mu(\xi).$$

**Démonstration.**

□

### 5.3.3 Étude du cas où $d = 1$

Dans le cas  $d = 1$ , la fonction caractéristique  $\varphi_\mu$  prend une expression simple

$$\varphi_\mu(x) = \int_{\mathbb{R}} e^{i\xi x} d\mu(\xi),$$

où  $\mu$  est une probabilité sur  $((\mathbb{R}, \mathbb{B}(\mathbb{R})))$ .

**Définition 5.3.5** Soit  $\mu$  une probabilité sur  $((\mathbb{R}, \mathbb{B}(\mathbb{R})))$ .

1. On dit que  $\mu$  est la loi normale réduite et centrée si et seulement si elle admet la fonction  $f(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$  comme densité par rapport à la mesure de Lebesgue  $dt$  sur  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ , c'est à dire que

$$d\mu(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

2. On dit que  $\mu$  est la loi exponentielle de paramètre 1 si et seulement si elle admet la fonction  $f(t) = e^{-t} \chi_{]0, +\infty[}(t)$  comme densité par rapport à la mesure de Lebesgue  $dt$  sur  $((\mathbb{R}, \mathbb{B}(\mathbb{R})))$ , c'est à dire que

$$d\mu(t) = e^{-t} \chi_{]0, +\infty[}(t) dt.$$

**Proposition 5.3.6** 1. si  $\mu$  est une loi normale réduite et centrée (ou de Gauss-Laplace)  $\mathcal{N}(0, 1)$ , alors pour tout  $x \in \mathbb{R}$  on a

$$\varphi_\mu(x) = \exp\left(-\frac{x^2}{2}\right) \in \mathbb{R}.$$

2. si  $\mu$  est la mesure de Dirac au point  $a$  sur  $\mathbb{R}$ , alors pour tout  $x \in \mathbb{R}$  on a

$$\varphi_\mu(x) = \exp(iax) \in \mathbb{C}.$$

3. si  $\mu$  est la loi exponentielle de paramètre 1, alors pour tout  $x \in \mathbb{R}$  on a

$$\varphi_\mu(x) = \frac{1}{1 - ix} \in \mathbb{C}.$$

**Démonstration.**

□

### 5.3.4 Formules de réciprocity de Perron-Stieltjes : cas $d = 1$

**Proposition 5.3.7** Soient  $a$  et  $b$  des réels avec  $a < b$ , alors

1. La suite d'intégrales au sens de Riemann  $\left(u_n = \frac{1}{2\pi} \int_{-n}^n \frac{e^{-ia t} - e^{-ib t}}{it} \varphi_\mu(t) dt\right)_{n \in \mathbb{N}}$  est convergente dans  $\mathbb{R}$ , et on a

$$\lim_{n \rightarrow +\infty} u_n = \mu(]a, b]) + \frac{1}{2} \mu(\{a, b\})$$

2. La suite d'intégrales au sens de Riemann  $\left(v_n = \frac{1}{2n} \int_{-n}^n e^{-ia t} \varphi_\mu(t) dt\right)_{n \in \mathbb{N}}$  est convergente dans  $\mathbb{R}$ , et on a

$$\lim_{n \rightarrow +\infty} v_n = \mu(\{a\})$$

**Démonstration.**

□

### **Théorème 5.3.1 (Théorème d'injectivité)**

Deux mesures de probabilités sur  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$  ont la même fonction caractéristique si et seulement si elles sont égales.

**Démonstration.**

□

Dans le cas où la fonction caractéristique est intégrable au sens de Lebesgue sur  $\mathbb{R}$ , on peut préciser la connaissance de  $\mu$  :

**Proposition 5.3.8** Si  $\varphi_\mu$  est intégrable au sens de Lebesgue sur  $\mathbb{R}$ , alors  $\mu$  admet une densité de probabilité  $f$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ .

De plus,  $f$  est positive, bornée et continue sur  $\mathbb{R}$ , et que pour tout  $x \in \mathbb{R}$  on a

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi_\mu(t) dt.$$

**Démonstration.**

□

**Remarque 5.3.2** La réciproque de la proposition est fausse : la loi exponentielle de paramètre 1 fournit un contre-exemple.

**Proposition 5.3.9** Supposons qu'une mesure de probabilité  $\mu$  admette une densité  $f$  de classe  $\mathcal{C}^n$  telle que pour tout entier  $0 \leq k \leq n$ , la dérivée  $f^{(k)}$  d'ordre  $k$  de  $f$  soit intégrable au sens de Lebesgue sur  $\mathbb{R}$ , alors

$$\lim_{x \rightarrow +\infty} x^{n-1} \varphi_\mu(x) = 0.$$

**Démonstration.**

□

## **5.4 Moments d'une mesure de probabilité dans le cas $d = 1$**

Soit  $\mu$  une mesure de probabilité sur  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ .

**Définition 5.4.1** Soit  $n \in \mathbb{N}$  un entier naturel tel que :  $\int_{-\infty}^{+\infty} |x^n| dx < +\infty$ .

On appelle **moment** d'ordre  $n$  de la probabilité  $\mu$  le nombre réel  $M_n = \int_{-\infty}^{+\infty} x^n d\mu(x)$ .

En particulier, pour  $n = 0$ , on a  $M_0 = \mu(\mathbb{R}) = 1$ .

**Proposition 5.4.1** Soit  $n \in \mathbb{N}$  un entier naturel tel que :  $\int_{-\infty}^{+\infty} |x^n| d\mu(x) < +\infty$ , alors la fonction caractéristique  $\varphi_\mu$  de  $\mu$  est de classe  $\mathcal{C}^n$  sur  $\mathbb{R}$  et pour tout  $x \in \mathbb{R}$ , sa dérivée d'ordre  $n$  en  $x$  est donnée par

$$\varphi_\mu^{(n)}(x) = \int_{-\infty}^{+\infty} (it)^n e^{itx} d\mu(t).$$

De plus, on a  $\varphi_\mu^{(n)}(0) = \int_{-\infty}^{+\infty} (it)^n d\mu(t) = i^n M_n$ .

La fonction caractéristique  $\varphi_\mu$  admet un développement limité de Taylor en 0 à l'ordre  $n$  donné par

$$\varphi_\mu(x) = \sum_{k=0}^n \frac{i^k M_k}{k!} x^k + o(x^n).$$



**Démonstration.**

□

**Remarque 5.4.1** *La proposition montre qu'il y a un lien entre l'intégrabilité de la fonction puissance et la régularité de la fonction caractéristique. C'est à dire que si  $\phi(x) = x^n$  est  $\mu$ -intégrable, alors  $\phi$  est de classe  $C^n$ .*

**Mais**, la réciproque est fausse, en effet, il est possible de trouver une mesure de probabilité  $\mu$  dont la fonction caractéristique est de classe  $C^1$  et la fonction  $\phi(x) = x$  n'est pas  $\mu$ -intégrable.

On peut tout de même en donner une réciproque à la proposition précédente sous la forme suivante :

**Proposition 5.4.2** *Soit  $n \in \mathbb{N}$  un entier naturel tel que la fonction caractéristique  $\varphi_\mu$  soit dérivable en 0 à l'ordre  $n$ , alors  $M_{2p} < +\infty$  où  $2p$  est le plus grand entier pair inférieur à  $n$ . En particulier,  $\mu$  admet des moments jusqu'à l'ordre  $2p$ .*

*De plus, si  $\varphi_\mu$  admet un développement limité de Taylor en 0 à l'ordre  $n$  s'écrivant*

$$\varphi_\mu(x) = 1 + \sum_{k=1}^n a_k x^k + o(x^n),$$

*alors pour tout  $1 \leq k \leq 2p$  on a  $M_k = (-i)^k a_k k!$ .*

**Démonstration.**

□



# Chapitre 6

## Convergences et comportement asymptotique

### 6.1 Quelques relations de grand intérêt en statistique

Soit  $(\Omega, \mathcal{B}, P)$  un e.p. et soit  $X_1, \dots, X_n$  des variables aléatoires réelles sur  $\Omega$ . En statistique, on est souvent amené à construire les variables aléatoires suivantes :

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i \\ \mathcal{S}^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \\ T &= \sqrt{n} \frac{\bar{X} - \mu}{\mathcal{S}}.\end{aligned}$$

Dans le cas, fréquent, où l'on admet ou vérifie, que les  $X_i$  sont des lois normales de même paramétrage  $(\mu, \sigma^2)$ , alors

- i)  $\bar{X}$  suit une loi normale  $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ .
- ii)  $\mathcal{S}^2$  suit une loi du  $\chi^2$  à  $(n-1)$  degrés de liberté.
- iii)  $T$  suit une loi de Student à  $(n-1)$  degrés de liberté.

#### 6.1.1 Quantile $q_\alpha$ , Médiane et Mode

##### Définition 6.1.1 (Quantile)

Soit  $X$  une variable aléatoire continue. Le nombre  $q_\alpha$  tel que

$$P([X \leq q_\alpha]) = \alpha \in [0, 1]$$

est le **quantile** d'ordre  $\alpha$  de la loi de probabilité de  $X$ .

Ces quantile sont notés de différentes façons :

- 1.  $u_\alpha$  : pour la loi normale,
- 2.  $t_\alpha^n$  : pour la loi de Student à  $n$  degrés de liberté,
- 3.  $\chi_\alpha^n$  : pour la loi du  $\chi_n^2$ .

##### Définition 6.1.2 (Médiane)

Soit  $X$  une variable aléatoire continue. La valeur  $q_{0.5}$  est la médiane de la variable aléatoire si

$$P([X \leq q_{0.5}]) = 0.5.$$

La médiane est le quantile  $\alpha = 0.5$  de la variable aléatoire  $X$ .

##### Définition 6.1.3 (Mode)

Soit  $X$  une variable aléatoire continue de densité de probabilité  $f$ . Un point  $x_0$  du domaine de définition  $\mathcal{D}$  de  $f$  tel que  $f(x_0) = \max_{x \in \mathcal{D}} f(x)$  est appelé un **mode** de la variable aléatoire  $X$ .

## 6.2 Convergence stochastique

On s'intéresse à la loi d'une suite de v.a. identiques, et plus particulièrement à la convergence à l'infini. Pour étudier cette convergence, il existe de nombreux outils dont nous donnons ici les principaux.

### 6.2.1 Convergence en probabilité

**Définition 6.2.1** Soit  $(X_n)_n$  une suite de variables aléatoires réelles. On dit que la suite  $(X_n)$  converge en **probabilité** vers la variable aléatoire  $X$  si et seulement si  $\forall \eta, \varepsilon$  (donnés arbitrairement petits),  $\exists n_0 \in \mathbb{N}$  tel que  $n > n_0 \Rightarrow P(|X - X_n| > \varepsilon) < \eta$ .

C'est-à-dire

$$X_n \xrightarrow{P} X \Leftrightarrow \forall \varepsilon > 0, \lim_{n \rightarrow +\infty} P(|X_n - X| > \varepsilon) = 0.$$

### 6.2.2 Convergence en loi (convergence faible)

**Définition 6.2.2** Soit  $(X_n)$  une suite de variables aléatoires réelles de fonction de répartition  $x \mapsto F_n(x)$ , et soit  $X$  une variable aléatoire de fonction de répartition  $x \mapsto F(x)$ . On dit que la suite  $(X_n)$  converge en **loi** vers la variable aléatoire  $X$  si et seulement si  $(F_n(x))$  converge vers  $F(x)$ .

C'est-à-dire

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \forall t \in \mathbb{R}, \lim_{n \rightarrow +\infty} \mathbb{E}(e^{itX_n} - e^{itX}) = 0.$$

### 6.2.3 Convergence presque sûre

**Définition 6.2.3** Soit  $(X_n)_n$  une suite de variables aléatoires réelles, et  $X$  une variable aléatoire. On dit que la suite  $(X_n)$  converge **presque sûre** vers la variable aléatoire  $X$ , noté  $X_n \xrightarrow{p.s.} X$ , si et seulement si

$$P\left(\limsup_{n \rightarrow +\infty} |X_n - X| = 0\right) = 1.$$

### 6.2.4 Convergence en moyenne

**Définition 6.2.4** Soit  $(X_n)_n$  une suite de variables aléatoires réelles. On dit que la suite  $(X_n)$  converge en **moyenne d'ordre  $p$**  vers la variable aléatoire  $X$  si et seulement si

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^p) = 0.$$

La plus utilisée de ces convergence est la convergence en moyenne quadratique ( $p = 2$ ).

**Théorème 6.2.1** Soient  $(X_n)_n$  et  $(Y_n)_n$  deux suites de variables aléatoires réelles.

1.  $X_n \xrightarrow{p.s.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X$ .
2.  $X_n \xrightarrow{P} c \Leftrightarrow X_n \xrightarrow{\mathcal{L}} c$  si  $c$  est une constante.
3. Si  $X_n \xrightarrow{\mathcal{L}} X$  et  $X_n - Y_n \xrightarrow{P} 0$  alors  $Y_n \xrightarrow{\mathcal{L}} X$ .
4. Si  $X_n \xrightarrow{P} X$  et  $Y_n \xrightarrow{P} Y$  alors  $(X_n, Y_n) \xrightarrow{P} (X, Y)$ .
5. Si  $X_n \xrightarrow{\mathcal{L}} X$  et  $Y_n \xrightarrow{P} c$  alors  $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$ .

**Lemme 6.2.1 (Slutsky)**

Si  $X_n \xrightarrow{\mathcal{L}} X$  et  $Y_n \xrightarrow{\mathcal{L}} c$  où  $c$  est une constante, alors

1.  $X_n + Y_n \xrightarrow{P} X + c.$
2.  $X_n Y_n \xrightarrow{\mathcal{L}} c.X.$
3.  $\frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{X}{c}$  si  $c \neq 0.$

**Théorème 6.2.2** Soit  $g : x \in A \rightarrow g(x) \in \mathbb{R}$  une fonction continue en tout point de  $A$  tel que  $P(X \in C) = 1$ , alors si  $X_n \xrightarrow{\mathcal{L}} X$  (resp.  $\xrightarrow{P} X$ ,  $\xrightarrow{p.s.} X$ ) alors  $g(X_n) \xrightarrow{\mathcal{L}} g(X)$  (resp.  $\xrightarrow{P} g(X)$ ,  $\xrightarrow{p.s.} g(X)$ ).

**Remarque 6.2.1** Si  $g$  est bornée, alors  $\mathbb{E}(g(X_n)) \rightarrow \mathbb{E}(g(X))$  ce qui ne permet pas, par exemple d'obtenir la convergence des moments car la fonction  $x \mapsto x^k$  est non bornée.

**Définition 6.2.5 (Uniforme intégrabilité)**

Une suite de variables aléatoires réelles  $(X_n)_n$  est dite **uniformément intégrable** si

$$\lim_{M \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{E}(|X_n| 1_{[|X_n| > M]}) = 0,$$

où  $1_{[|X_n| > M]}$  est la fonction caractéristique de l'événement  $[|X_n| > M]$ .

**Théorème 6.2.3 (Convergence des moments)**

Soit  $g : x \in A \rightarrow g(x) \in \mathbb{R}$  une fonction continue en tout point de  $A$  tel que  $P(X \in C) = 1$  et  $X_n \xrightarrow{\mathcal{L}} X$ . Alors  $\mathbb{E}(g(X_n)) \rightarrow \mathbb{E}(g(X))$  si et seulement si  $(g(X_n))$  est uniformément intégrable.

## 6.3 Loi des grands nombres

**Proposition 6.3.1 (Inégalité de Markov)**

Si  $Y$  est une variable aléatoire positive ou nulle dont l'espérance existe, alors pour tout  $x > 0$  on a

$$P([Y \geq x]) \leq \frac{1}{x} \mathbb{E}(Y).$$

**Démonstration.** Soit  $x$  un réel strictement positif. On a

$$X = X \cdot \mathbb{I}_{[X < x]} + X \cdot \mathbb{I}_{[X \geq x]}$$

alors

$$\mathbb{E}(X) = \mathbb{E}(X \cdot \mathbb{I}_{[X < x]}) + \mathbb{E}(X \cdot \mathbb{I}_{[X \geq x]}) \geq \mathbb{E}(X \cdot \mathbb{I}_{[X \geq x]}) \geq \mathbb{E}(x \cdot \mathbb{I}_{[X \geq x]}) \geq x \mathbb{E}(\mathbb{I}_{[X \geq x]})$$

or  $\mathbb{E}(\mathbb{I}_{[X \geq x]}) = P([X \geq x])$ , alors

$$\mathbb{E}(X) \geq x P([X \geq x])$$

d'où l'inégalité de Markov. □

**Proposition 6.3.2 (Inégalité de Tchebychev)**

Si  $X$  est une variable aléatoire ayant un second moment, alors pour tout  $t > 0$  on a

$$P([X - \mathbb{E}(X) \geq t]) \leq \frac{1}{t^2} \sigma^2(X).$$

**Démonstration.** Soit  $t$  un réel strictement positif. On applique l'inégalité de Markov à  $Y = (X - \mathbb{E}(X))^2$  et à  $x = t^2$ . Comme

$$P([X - \mathbb{E}(X)] \geq t) = P([(X - \mathbb{E}(X))^2 \geq t^2])$$

et

$$P([(X - \mathbb{E}(X))^2 \geq t^2]) \leq \frac{1}{t^2} \mathbb{E}((X - \mathbb{E}(X))^2) = \frac{1}{t^2} \sigma^2(X)$$

l'inégalité de Tchebychev est aussi démontrée.  $\square$

Finalement, la variance d'une somme de variables aléatoires indépendantes est la somme des variances. Plus précisément :

**Proposition 6.3.3** Si  $X_1, X_2, \dots, X_N$  sont des variables aléatoires indépendantes ayant un second moment, alors

$$\sigma^2 \left( \sum_{i=1}^N X_i \right) = \sum_{i=1}^N \sigma^2(X_i).$$

En corollaire, on a donc la loi faible des grands nombres qui dit que en un certain sens, si des variables aléatoires sont indépendantes et de même loi, alors leur moyenne arithmétique tend vers leur espérance commune. Plus précisément :

**Théorème 6.3.1 (Loi faible des grands nombres)**

Soit  $X_1, X_2, \dots, X_N, \dots$  une suite infinie de v.a. indépendantes et de même loi, et possédant un second moment. Alors, pour tout nombre  $\varepsilon > 0$  fixé on a

$$\lim_{n \rightarrow +\infty} P \left( \left[ \left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1) \right| \geq \varepsilon \right] \right) = 0$$

**Démonstration.** Notons  $S_n = \sum_{i=1}^n X_i$ . Alors  $\mathbb{E}(S_n/n) = \mathbb{E}(X_1)$  et

$$\sigma^2(S_n/n) = \sigma^2(S_n)/n = (\sigma^2(X_1) + \dots + \sigma^2(X_n))/n^2 = \sigma^2(X_1)/n.$$

Appliquons alors l'inégalité de Tchebychev à  $X = S_n/n$  et à  $t = \varepsilon$  ; on obtient

$$P \left( \left| \frac{S_n}{n} - \mathbb{E}(X_1) \right| \geq \varepsilon \right) \leq \frac{1}{n\varepsilon^2} \sigma^2(X_1).$$

Or pour  $\varepsilon$  fixé, on a  $\lim_{n \rightarrow +\infty} \frac{1}{n\varepsilon^2} \sigma^2(X_1) = 0$ , alors

$$\lim_{n \rightarrow +\infty} P \left( \left| \frac{S_n}{n} - \mathbb{E}(X_1) \right| \geq \varepsilon \right) = 0$$

ce qui achève la démonstration.  $\square$

**Commentaires :** l'importance philosophique de la loi des grands nombres est non négligeable : elle justifie la démarche que nous avons adoptée pour modéliser le calcul des probabilités. L'idée d'expérience décrite au début de ce cours est la sélection d'un point  $\omega$  dans un espace d'observables  $\Omega$ , mais par un procédé susceptible d'être répété ad libitum et dans les mêmes conditions. Soit  $S$  une partie de  $\Omega$  comptons le nombre de fois où  $S$  est réalisé en  $n$  essais, divisons ce nombre par  $n$  et notons par  $f_n$  la fraction, ou la fréquence, ainsi obtenue. L'idée de probabilité est basée sur la constatation physique que la suite des  $f_n$  converge vers un nombre  $P(S)$  qu'on appellera probabilité de  $S$ . Si la théorie est bien faite, c'est-à-dire si les axiomes sont bien choisis, on doit retrouver cette constatation physique quelque part à l'état de théorème dans la théorie développée à partir de ces axiomes. C'est le cas. En effet, le  $\Omega$  initial décrivant une expérience est remplacé par un produit infini  $\prod_{i=1}^{\infty} \Omega_i$  où les  $\Omega_i$  sont identiques à l' $\Omega$  initial, et sont les résultats possibles de l'expérience répétée à l'instant  $j$  : Les points de

ce produit sont donc des suites infinies  $\omega = (\omega_i)_{i=1}^\infty$ . Quant à la probabilité sur le produit, elle est telle que toutes les fonctions  $f_j(\omega) = \omega_j$  soient indépendantes. Ceci fait, notons  $X_j(\omega) = 1$  si  $j \in S$  et  $X_j(\omega) = 0$  sinon. On a une suite de v.a. de Bernoulli indépendantes et de même loi d'espérance  $p = P(S)$ , et de même loi. La loi faible des grands nombres dit que  $f_n = \frac{1}{n}(X_1 + \dots + X_n)$  converge vers  $P(S)$ ; dans le sens décrit au théorème de la loi faible des grands nombres. Il existe un théorème avec une conclusion plus précise, appelé loi forte des grands nombres, que nous exposons maintenant.

### **Théorème 6.3.2 (Loi forte des grands nombres)**

Soit  $X_1, X_2, \dots, X_N, \dots$  une suite infinie de v.a. Bernoulli indépendantes et de même loi  $q\delta_0 + p\delta_1$ ; avec  $0 < p = 1 - q < 1$ . Alors,

$$P\left(\left[\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n X_i = p\right]\right) = 1$$

## **6.4 Théorème central limite**

Le théorème central limite est l'un des résultats les plus importants de la théorie des probabilités. De façon informelle, ce théorème donne une estimation très précise de l'erreur que l'on commet en approchant l'espérance mathématique par la moyenne arithmétique. Ce phénomène a d'abord été observé par Gauss qui l'appelait loi des erreurs; mais ce dernier n'en a pas donné de démonstration rigoureuse. La preuve du théorème a été apportée par Moivre et Laplace; le théorème porte donc parfois leurs noms.

Ce théorème est fondamental car il justifie toutes les approximations par la loi normale.

### **Théorème 6.4.1 (Théorème central limite)**

Soit  $X_1, X_2, \dots, X_N, \dots$  une suite de v.a. indépendantes de même loi d'espérance mathématique  $\mu$  et de variance  $\sigma^2$ . Alors la suite des v. a.  $Y_n = \frac{1}{\sqrt{n}} \left( \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma} \right)$  converge en loi vers la v. a. normale centrée réduite  $\mathcal{N}(0, 1)$ .

### **Corollaire 6.4.1 (Théorème central limite)**

Soit  $X_1, X_2, \dots, X_N, \dots$  une suite de v.a. indépendantes de même loi d'espérance mathématique  $\mu$  et de variance  $\sigma^2$ . Alors la suite des v. a.

$$Y_1 = X_1 - \mu, Y_2 = \frac{1}{\sqrt{2}}(X_1 + X_2 - 2\mu), \dots, Y_n = \frac{1}{\sqrt{n}}(X_1 + X_2 + \dots + X_n - n\mu)$$

converge en loi vers la v. a. normale  $\mathcal{N}(0, \sigma^2)$ .

### **Corollaire 6.4.2 (Théorème central limite)**

Soit  $X_1, X_2, \dots, X_N, \dots$  une suite de v.a. indépendantes de même loi d'espérance mathématique  $\mu$  et de variance  $\sigma^2$ . Alors la suite des v. a.

$$Y_1 = X_1, Y_2 = \frac{1}{\sqrt{2}}(X_1 + X_2), \dots, Y_n = \frac{1}{\sqrt{n}}(X_1 + X_2 + \dots + X_n)$$

converge en loi vers la v. a. normale  $\mathcal{N}(\mu, \sigma^2)$ .

## 6.5 Loi normale : Théorème de Moivre-Laplace

Soit  $Y$  la loi normale  $\mathcal{N}(\mu, \sigma^2)$  alors  $X = \frac{(Y-\mu)}{\sigma}$  est une variable aléatoire normale centrée réduite. La fonction de répartition  $\Pi$  associée à la variable  $X$  est définie par

$$\Pi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt,$$

### **Théorème 6.5.1 (Théorème de Moivre-Laplace)**

Soit  $B(n, p)$  une v.a. binomiale de paramètres  $n$  et  $p$  (avec  $0 < p < 1$  est constante donnée), la variable :

$$T_n = \frac{B(n, p) - np}{\sqrt{np(1-p)}}$$

converge en loi vers la v. a. normale centrée réduite  $\mathcal{N}(0, 1)$ .

Cet énoncé signifie que la fonction de répartition de  $T_n$  tend vers la fonction de répartition  $\Pi$  de la loi normale centrée réduite, c'est-à-dire :

$$\lim_{n \rightarrow +\infty} P[a \leq T_n \leq b] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt$$