

# Probabilités

Olivier Roques

2016-2017

## Table des matières

<b>1</b>	<b>Chaînes de Markov</b>	<b>2</b>
<b>2</b>	<b>Mesures et intégration</b>	<b>2</b>
2.1	Mesures . . . . .	2
2.2	Intégration . . . . .	3
<b>3</b>	<b>Variables et vecteurs aléatoires réels</b>	<b>4</b>
3.1	Variables aléatoires réelles . . . . .	4
3.2	Vecteurs aléatoires réels . . . . .	5
<b>4</b>	<b>Inégalités</b>	<b>6</b>
<b>5</b>	<b>Loi conditionnelle</b>	<b>7</b>
<b>6</b>	<b>Transformées et fonction caractéristique</b>	<b>7</b>
6.1	Transformée de Laplace . . . . .	7
6.2	Fonction caractéristique et transformée de Fourier . . . . .	8
<b>7</b>	<b>Vecteurs gaussiens</b>	<b>8</b>
<b>8</b>	<b>Convergences</b>	<b>9</b>

# 1 Chaînes de Markov

**Définition 1.1** (Chaîne de Markov). Une *chaîne de Markov*  $(X_n) \in \mathbb{N}^{\mathbb{N}}$  vérifie  $\forall n \in \mathbb{N}, \forall x_0, \dots, x_{n+1} \in X_0(\Omega), \mathbf{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n) = \mathbf{P}(X_{n+1} = x_{n+1} \mid X_n = x_n)$ .

**Définition 1.2** (Matrice de transition). L'évolution d'une chaîne de Markov est entièrement déterminée par ses probabilités de transition  $\mathbf{P}(X_{n+1} = j \mid X_n = i) = \mathbf{P}(X_1 = j \mid X_0 = i) \forall i, j \in X_0(\Omega)$ . On définit alors la *matrice de transition*  $P$  tel que  $P_{i,j} = \mathbf{P}(X_1 = j \mid X_0 = i)$ . En notant la loi de  $X_n$  comme un vecteur ligne  $\pi(n)$ , on a  $\forall n \in \mathbb{N} \pi(n) = \pi(0)P^n$ .

**Définition 1.3** (Irréductibilité). Une chaîne de Markov est dite *irréductible* si pour tous sommets distincts  $(i, j)$  de son graphe de transition, il existe un chemin de  $i$  vers  $j$ .

**Définition 1.4** (Loi stationnaire). Un vecteur  $\pi$  vérifiant  $\pi P = \pi$  est appelée *loi stationnaire* de la chaîne de Markov associée. On a alors  $\forall n \in \mathbb{N}, \pi(n) = \pi$ .

**Théorème 1.1** (Perron-Frobenius). Une chaîne de Markov irréductible admet une unique loi stationnaire. On a alors  $\pi_i > 0$  pour tout  $i \in X_0(\Omega)$ .

**Définition 1.5** (Périodicité). La *période* d'un état du graphe de transition est le pgcd des longueurs des cycles passant par cet état. Tous les états d'une chaîne de Markov irréductible ont la même période. On dit qu'un état est *apériodique* lorsqu'il est de période 1.

**Théorème 1.2.** Soit  $\pi$  la loi stationnaire d'une chaîne de Markov irréductible et apériodique. Alors pour toute loi initiale  $\pi(0)$ ,  $\lim_{n \rightarrow +\infty} \pi(n) = \pi$ .

**Théorème 1.3** (Théorème ergodique). Soit  $\pi$  la loi d'une chaîne de Markov irréductible. Pour toute fonction  $\varphi$  sur  $X_0(\Omega)$ , on a  $\frac{1}{n} \sum_{k=1}^n \varphi(X_k) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \sum_{i \in X(\Omega)} \pi_i \varphi(i)$ .

## 2 Mesures et intégration

### 2.1 Mesures

**Définition 2.1** (Tribu). Une collection  $\mathcal{F}$  d'ensembles de  $\Omega$  est appelée *tribu* sur  $\Omega$  si elle vérifie :

- (i)  $\Omega \in \mathcal{F}$
- (ii)  $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$
- (iii)  $\forall (A_n) \in \mathcal{F}^{\mathbb{N}}, \bigcup_{i=0}^{+\infty} A_i \in \mathcal{F}$ .

**Définition 2.2** (Tribu engendrée). Soit  $\mathcal{C}$  une collection d'ensembles de  $\Omega$ . On appelle *tribu engendrée* par  $\mathcal{C}$  sur  $\Omega$  l'intersection de toutes les tribus sur  $\Omega$  contenant  $\mathcal{C}$ . On la note  $\sigma(\mathcal{C})$ .

**Définition 2.3** (Tribu de Borel). La *tribu de Borel* sur  $\mathbb{R}^d$ , notée  $\mathcal{B}(\mathbb{R}^d)$  est la tribu définie par

$$\mathcal{B}(\mathbb{R}^d) = \sigma \left( \left\{ \prod_{i=1}^d ]a_i, b_i[ \mid (a_i, b_i) \in \mathbb{R}^2, a_i < b_i \right\} \right)$$

**Définition 2.4** (Application mesurable). Une application  $X : (\Omega, \mathcal{E}) \rightarrow (F, \mathcal{F})$  est dite *mesurable* si  $\forall H \in \mathcal{F}, X^{-1}(H) \in \mathcal{E}$ . En pratique, pour établir la mesurabilité d'une fonction, il suffit de montrer que  $\forall C \in \mathcal{C}, X^{-1}(C) \in \mathcal{E}$  où  $\mathcal{C}$  est une collection d'ensembles telle que  $\mathcal{F} = \sigma(\mathcal{C})$ .

**Propriété 2.1.** Une application  $f : \mathbb{R}^d \rightarrow \mathbb{R}^n$  continue est mesurable. Si  $X, Y$  sont des fonctions mesurables,  $X \circ Y, X + Y, XY, \max(X, Y), \min(X, Y)$  sont mesurables.

**Définition 2.5** (Mesure). Une *mesure* sur  $(E, \mathcal{E})$  est une fonction  $\mu$  telle que :

- (i)  $\mu : \mathcal{E} \rightarrow [0, +\infty]$
- (ii)  $\mu(\emptyset) = 0$
- (iii)  $\forall (A_n) \in \mathcal{E}^{\mathbb{N}}$  deux à deux disjoints,  $\mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$ .

Lorsque  $\mu(E) < \infty$  on dit que la mesure est finie. Si de plus  $\mu(E) = 1$ , la mesure est une *mesure de probabilité*.

**Propriété 2.2.** Soient  $A, B$  des éléments de  $\mathcal{E}$ .

- (i) Pour toute partition  $(B_n)$  d'ensembles mesurables de  $E$ ,  $\mu(A) = \sum_{n=0}^{+\infty} \mu(A \cap B_n)$ .
- (ii) Si  $\mu(A \cup B) < +\infty$  alors  $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$
- (iii) Si  $A \subset B$  alors  $\mu(A) \leq \mu(B)$ .
- (iv) Si  $(A_n) \in \mathcal{E}^{\mathbb{N}}$  est croissante telle que  $A_n \xrightarrow[n \rightarrow +\infty]{} A$  alors  $\lim_{n \rightarrow +\infty} \mu(A_n) = \mu(A)$ .
- (v) Si  $(A_n) \in \mathcal{E}^{\mathbb{N}}$  est décroissante telle que  $A_n \xrightarrow[n \rightarrow +\infty]{} A$  et  $\mu(A_0) < +\infty$  alors  $\lim_{n \rightarrow +\infty} \mu(A_n) = \mu(A)$ .
- (vi) On a  $\forall (A_n) \in \mathcal{E}^{\mathbb{N}}, \mu(\bigcup_{n=0}^{+\infty} A_n) \leq \sum_{n=0}^{+\infty} \mu(A_n)$ .

**Définition 2.6** (Mesure image). Soit  $f : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$  une fonction mesurable. La *mesure image* de  $\mu$  par  $f$ , notée  $\mu_f$  est définie par :  $\forall A \in \mathcal{F}, \mu_f(A) = \mu(f^{-1}(A))$ .

**Théorème 2.1** (Mesure de Lebesgue). Il existe une unique mesure, appelée *mesure de Lebesgue* et notée  $\lambda$ , sur  $\mathbb{R}^d$  muni de la tribu des boréliens telle que  $\lambda\left(\prod_{i=1}^d ]a_i, b_i[ \right) = \prod_{i=1}^d (b_i - a_i)$ . Elle est de plus invariante par translation :  $\forall x \in \mathbb{R}^d, \forall A$  borélien,  $\lambda(A + x) = \lambda(A)$ .

**Théorème 2.2** (Mesure produit). Il existe une unique mesure, notée  $\mu \otimes \nu$  et appelée *mesure produit* de  $\mu$  et  $\nu$  sur  $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$  telle que  $\mu \otimes \nu(A \times B) = \mu(A)\nu(B)$  pour tout  $A \in \mathcal{E}_1, B \in \mathcal{E}_2$ .

## 2.2 Intégration

**Définition 2.7** (Fonction étagée). Soit  $f : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  une fonction mesurable.  $f$  est dite *étagée* lorsqu'elle prend un nombre fini de valeurs :  $\exists n \in \mathbb{N}, (A_1, \dots, A_n) \in \mathcal{E}^n$  deux à deux disjoints tels que  $f = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}$ .

**Définition 2.8** ( $\mu$ -intégrale). Pour une fonction étagée positive  $f$ , on définit sa  *$\mu$ -intégrale* par 
$$\int f d\mu = \sum_{k=1}^n \alpha_k \mu(A_k).$$

**Définition 2.9.** Pour  $f$  mesurable positive et  $(f_n)$  une suite de fonctions étagées convergeant vers  $f$ ,  $\int f \, d\mu = \lim_{n \rightarrow +\infty} \int f_n \, d\mu$ . Si  $f$  n'est plus seulement positive et que  $\int |f| \, d\mu$  est finie,  $f$  est  $\mu$ -intégrable et  $\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu$ . Enfin pour la mesure de Lebesgue,  $\int f \, d\lambda = \int f(x) \, dx$ .

**Définition 2.10.** On dit d'une propriété qu'elle est vraie  $\mu$ -presque-partout lorsque son complémentaire est de mesure nulle.

**Théorème 2.3** (Convergence dominée). Soit  $(f_n)$  une suite de fonctions mesurables qui converge  $\mu$ .p.p vers  $f$ . S'il existe  $g$  telle que  $|f_n(x)| \leq g(x)$  pour presque tout  $x$  et  $\int g \, d\mu < +\infty$ , alors

$$\lim_{n \rightarrow +\infty} \int f_n \, d\mu = \int \left( \lim_{n \rightarrow +\infty} f_n \right) d\mu = \int f \, d\mu.$$

**Théorème 2.4** (Fubini). Soit  $(E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \otimes \nu)$  un espace mesuré produit et  $f : E \times F \rightarrow \mathbb{R}$  une fonction mesurable. Si

(i)  $\iint |f(x, y)| \, d\mu(x) \, d\nu(y) < +\infty$

(ii)  $x \mapsto \int f(x, y) \, d\nu(y)$  mesurable

(iii)  $y \mapsto \int f(x, y) \, d\mu(x)$  mesurable

Alors on a  $\int f(x, y) \, d(\mu \otimes \nu)(x, y) = \int \left( \int f(x, y) \, d\nu(y) \right) d\mu(x) = \int \left( \int f(x, y) \, d\mu(x) \right) d\nu(y)$ .

### 3 Variables et vecteurs aléatoires réels

On se place sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbf{P})$ . Soit  $(E, \mathcal{E})$  un espace mesurable.

#### 3.1 Variables aléatoires réelles

**Définition 3.1** (Variable aléatoire). Une *variable aléatoire*  $X$  sur  $E$  une application mesurable de  $\Omega$  dans  $E$ .

**Définition 3.2** (Fonction de répartition). Soit  $X$  une v.a.r. La *fonction de répartition* de  $X$  est définie par :

$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto \mathbf{P}(X \leq x) \end{aligned}$$

Elle caractérise totalement la loi de  $X$ .

**Définition 3.3** (Densité). Lorsque  $F_X$  est dérivable, on définit la *densité*  $f$  de la loi de  $X$  par  $f(x) = \mathbf{P}'_X(x)$ . On a alors  $\mathbf{P}(X \in H) = \int_H f$ .

**Théorème 3.1.** Les trois propositions sont équivalentes :

- (i)  $X$  admet une densité ;
- (ii)  $F_X$  est absolument continue (donc en particulier continue) ;

(iii)  $\mathbf{P}_X(A) = 0$  pour tout ensemble négligeable  $A \in \mathcal{F}$ .

**Définition 3.4** (Espérance). L'espérance de la v.a.r.  $X$  est donnée par  $\mathbf{E}[X] = \int_{\Omega} X(\omega) d\mathbf{P}$ . Pour  $A \in \mathcal{F}$  tel que  $\mathbf{P}(A) > 0$ , on définit de plus l'espérance conditionnelle  $\mathbf{E}[X|A] = \int_{\Omega} X(\omega) d\mathbf{P}(\omega|A) = \int_{\Omega} X(\omega)\mathbf{1}_A(\omega) \frac{d\mathbf{P}(\omega)}{\mathbf{P}(A)} = \frac{\mathbf{E}[X\mathbf{1}_A]}{\mathbf{P}(A)}$ .

**Théorème 3.2** (Théorème de transfert). Soit  $X : \Omega \rightarrow \mathbb{R}$  une v.a.r. de densité  $f_X$  et  $g : E \rightarrow \mathbb{R}$  une fonction mesurable telle que  $\mathbf{E}[g(X)]$  soit définie. Alors on a  $\mathbf{E}[g(X)] = \int_{X(\Omega)} g(x)f_X(x) dx$ .

**Théorème 3.3** (Fonction test). Si pour toute fonction  $\varphi$  mesurable et positive  $\mathbf{E}[\varphi(X)] = \int_{X(\Omega)} \varphi(x)f(x) dx$  alors  $X$  a pour densité  $f$ .

**Définition 3.5** (Covariance). Soient  $X$  et  $Y$  deux v.a.r. d'ordre 2. Leur covariance est définie par

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

Le coefficient de corrélation est définie par  $\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y}$ . Lorsque  $\rho_{X,Y} = 0$ , on dit que  $X$  et  $Y$  sont *décorrélées*.

**Propriété 3.1.** Soient  $X$  et  $Y$  deux v.a.r. d'ordre 2 et  $\alpha, \beta \in \mathbb{R}^2$ . On a alors :

- $\text{Cov}(X + \alpha, Y + \beta) = \text{Cov}(X, Y)$
- $\text{Cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$ , donc  $X \perp Y \implies \text{Cov}(X, Y) = 0$ .
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$

**Définition 3.6** (Produit scalaire).  $\langle X, Y \rangle = \text{Cov}(X, Y)$  définit un produit scalaire. On a alors  $\|X\| = \sqrt{\text{Cov}(X, X)} = \sigma$ .

### 3.2 Vecteurs aléatoires réels

**Définition 3.7** (Loi jointe, marginale). La loi du vecteur aléatoire  $X = (X_1, \dots, X_d)$  est appelée *loi jointe* des variables aléatoires  $X_1, \dots, X_d$  dont les lois sont appelées *lois marginales*. La *fonction de répartition* de  $X$  est l'application  $F_X : \mathbb{R}^d \rightarrow \mathbb{R}_+$  définie pour tout  $(x_1, \dots, x_d) \in \mathbb{R}^d$  par :

$$F_X(x_1, \dots, x_d) = \mathbf{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f_X(x_1, \dots, x_d) dx_1 \dots dx_d$$

On a alors  $F_{X_i}(x) = \lim_{\substack{x_i=x \\ x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d \rightarrow +\infty}} F_X(x_1, \dots, x_d)$

**Théorème 3.4.** Les propositions suivantes sont équivalentes, et on peut de plus les généraliser à une collection de vecteurs aléatoires :

(i)  $X_1, \dots, X_d$  sont des variables aléatoires indépendantes.

(ii) La loi jointe est égale au produit des lois marginales :  $\mathbf{P}_X = \bigotimes_{i=1}^d P_{X_i}$ .

(iii)  $\forall (x_1, \dots, x_d) \in \mathbb{R}^d, F_X(x_1, \dots, x_d) = F_{X_1}(x_1) \times \dots \times F_{X_d}(x_d)$ .

(iv) Pour toutes fonctions mesurables  $h_1, \dots, h_d : \mathbb{R} \rightarrow \mathbb{R}$  telles que les v.a.  $h_i(X_i)$  soient toutes positives ou intégrables,

$$\mathbf{E}[h_1(X_1) \times \dots \times h_d(X_d)] = \mathbf{E}[h_1(X_1)] \times \dots \times \mathbf{E}[h_d(X_d)]$$

(v) Si de plus chaque v.a.r.  $X_i$  admet une densité  $f_{X_i}$ , alors

$$f_X(x_1, \dots, x_d) = f_{X_1}(x_1) \times \dots \times f_{X_d}(x_d)$$

De plus, une famille de variables aléatoires (de cardinal non nécessairement fini) est dite indépendante si toute sous-famille finie est indépendante.

**Définition 3.8** (Espérance). L'*espérance* de  $X$  est définie comme le vecteur des espérances de ses composantes. Elle est bien définie si et seulement si toutes les composantes admettent une espérance. Un vecteur aléatoire  $X$  est dit d'*ordre*  $p$  si toutes ses composantes sont d'ordre  $p$ .

**Définition 3.9** (Covariance). On appelle *matrice de covariance* d'un vecteur aléatoire  $X$  d'ordre 2 la matrice  $\text{Cov}(X) = (\text{Cov}(X_i, X_j))_{i,j}$ . La diagonale de  $\text{Cov}(X)$  est égale au vecteur des variances. Si  $X_1, \dots, X_d$  sont décorrélées,  $\text{Cov}(X)$  est une matrice diagonale. On définit de plus la *covariance de deux vecteurs* par  $\text{Cov}(X, Y) = (\text{Cov}(X_i, Y_j))_{i,j}$ .

**Propriété 3.2.** Soit  $X, Y$  des vecteurs aléatoires d'ordre 2 de taille  $d$ ,  $A \in \mathcal{M}_{n,d}(\mathbb{R})$  et  $b$  un vecteur de taille  $n$ . On a alors :

- $\text{Cov}(AX + b) = A\text{Cov}(X)A^T$ .
- $\text{Cov}(X) = \mathbf{E}[(X - \mathbf{E}[X])(X - \mathbf{E}[X])^T]$ .
- $\forall i, (\text{Cov}(X))_{i,i} = \text{Cov}(X_i, X_i) = \text{V}(X_i)$ .
- $\text{Cov}(X)$  est positive.
- $\text{Cov}(X) = 0$  si et seulement si  $X = \mathbf{E}[X]$ .
- $\text{Cov}(Y, X) = \text{Cov}(X, Y)^T$ .
- $\text{Cov}(X + Y) = \text{Cov}(X) + \text{Cov}(Y) + \text{Cov}(X, Y) + \text{Cov}(Y, X)$ .

**Théorème 3.5** (Formule du changement de variable). Soient  $U$  et  $V$  deux ouverts de  $\mathbb{R}^d$  et  $\phi : U \rightarrow V$  un difféomorphisme. Alors si  $f$  est une fonction définie sur  $V$  à valeurs positives,  $\int_U f \circ \phi = \int_V \frac{f}{|\det J_\phi \circ \phi^{-1}|} = \int_V f |\det J_{\phi^{-1}}|$ .

## 4 Inégalités

**Théorème 4.1** (Inégalité de Hölder). Pour tout  $p, q \geq 0$  tels que  $\frac{1}{p} + \frac{1}{q} = 1$ , on a :

$$\mathbf{E}[|XY|] \leq \mathbf{E}[|X|^p]^{\frac{1}{p}} \cdot \mathbf{E}[|Y|^q]^{\frac{1}{q}}$$

Lorsque  $p = q = 2$ , on l'appelle *inégalité de Cauchy-Schwarz*.

**Théorème 4.2** (Inégalité de Markov). Pour tout  $\varepsilon > 0, p \geq 1, \mathbf{P}(|X| \geq \varepsilon) \leq \frac{\mathbf{E}[|X|^p]}{\varepsilon^p}$ .

**Théorème 4.3** (Inégalité de Chebyshev). Pour tout  $\varepsilon > 0$ ,  $\mathbf{P}(|X - \mathbf{E}[X]| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}$ .

**Théorème 4.4** (Loi faible des grands nombres). Soit  $(X_k)$  une famille de variables aléatoires i.i.d. discrètes. On suppose que  $\mathbf{E}[X_0^2] < +\infty$ . Alors, en notant  $S_n = \sum_{k=0}^n X_k$ , on a :

$$\forall \varepsilon > 0, \mathbf{P}\left(\left|\frac{1}{n}S_n - \mathbf{E}[X_0]\right| \geq \varepsilon\right) \leq \frac{\text{Var}(X)}{n\varepsilon^2}$$

**Théorème 4.5** (Inégalité de Jensen). Soit  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction convexe. Soit  $X$  une v.a.r. telle que  $\mathbf{E}[|X|] < +\infty$  et  $\mathbf{E}[|\varphi(X)|] < +\infty$ . Alors on a  $\varphi(\mathbf{E}[X]) \leq \mathbf{E}[\varphi(X)]$ .

## 5 Loi conditionnelle

Soient  $X : \Omega \rightarrow \mathbb{R}^n$  et  $Y : \Omega \rightarrow \mathbb{R}^p$  deux vecteurs aléatoires.

**Définition 5.1** (Support). On définit le *support d'une probabilité*  $\mathbf{P}$ , noté  $\text{supp}(\mathbf{P})$  comme le complémentaire du plus grand ouvert (pour l'inclusion) de probabilité nulle.

**Définition 5.2.** La loi conditionnelle de  $Y$  sachant  $\{X = x\}$  avec  $x \in \text{supp}(\mathbf{P}_X)$  est définie telle que pour toutes fonction  $\theta$  et  $\varphi$  mesurables et bornées :

$$\mathbf{E}[\varphi(X)\theta(Y)] = \int_{\mathbb{R}^n} \varphi(x) \int_{\mathbb{R}^p} \theta(y) d\mathbf{P}_{Y|X=x}(y) d\mathbf{P}_X(x)$$

Si  $Y$  et  $X$  sont indépendantes,  $P_{Y|X=x} = P_Y$ . On a de plus, pour toute fonction  $\psi : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$  mesurable bornée,  $\mathbf{E}[\psi(X, Y)] = \int_{\mathbb{R}^n \times \mathbb{R}^p} \psi(x, y) d\mathbf{P}_{Y|X=x}(y) d\mathbf{P}_X(x)$ .

**Théorème 5.1** (Bayes). Soit  $f_{X,Y}$  la densité jointe de  $(X, Y)$ . La loi conditionnelle de  $Y$  sachant  $\{X = x\}$ , où  $x \in \text{supp}(\mathbf{P}_X)$ , a pour densité  $f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X|Y=y}(x)f_Y(y)}{f_X(x)}$ .

**Définition 5.3** (Espérance conditionnelle). On définit l'*espérance conditionnelle* de  $Y$  sachant  $\{X = x\}$  par  $\mathbf{E}[Y | X = x] = \int y d\mathbf{P}_{Y|X=x}(y)$ . On a alors, pour toute fonction mesurable bornée  $\theta$ ,

$$\mathbf{E}[\theta(Y)] = \int_{\mathbb{R}^n} \mathbf{E}[\theta(Y) | X = x] d\mathbf{P}_X(x)$$

## 6 Transformées et fonction caractéristique

### 6.1 Transformée de Laplace

**Définition 6.1** (Transformée de Laplace). Soit  $X$  une v.a.r. à valeurs dans  $\mathbb{R}_+$ . On définit la *transformée de Laplace* de  $X$  par :  $\forall t \geq 0$ ,  $L_X(t) = \mathbf{E}[e^{-tX}] = \int_{\mathbb{R}_+} e^{-tx} d\mathbf{P}_X(x)$ .

**Propriété 6.1.** On a les propriétés suivantes :

- $L_X(0) = 1$ ,  $L'_X(0) = -\mathbf{E}[X]$ ,  $L''_X(0) = \mathbf{E}[X^2]$ .
- $X$  et  $Y$  sont indépendantes si et seulement si  $L_{X+Y} = L_X L_Y$ .

## 6.2 Fonction caractéristique et transformée de Fourier

**Définition 6.2** (Fonction caractéristique). On appelle *fonction caractéristique* de  $X$  la fonction  $\Phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$  définie par

$$\Phi_X(t) = \mathbf{E}[e^{i\langle t, X \rangle}] = \int_{\mathbb{R}^d} e^{i\langle t, X \rangle} d\mathbf{P}_X(x) = \mathbf{E} \left[ \prod_{k=1}^d e^{it_k X_k} \right]$$

**Définition 6.3** (Transformée de Fourier). Soit  $\mu$  une mesure finie sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . La *transformée de Fourier* de  $\mu$  est la fonction  $\widehat{\mu} : \mathbb{R}^d \rightarrow \mathbb{C}$  définie par :  $\widehat{\mu} = \int e^{i\langle t, x \rangle} d\mu(x)$ . On a alors  $\Phi_X = \widehat{\mathbf{P}_X}$ .

**Propriété 6.2.** On a les propriétés suivantes :

- $\forall t \in \mathbb{R}^d, |\Phi_X(t)| \leq 1$ .
- $\Phi_X$  est continue.
- $\Phi_{AX+b}(t) = e^{i\langle t, b \rangle} \Phi_{AX}(t) = e^{i\langle t, b \rangle} \Phi_X(A^T t)$ .
- $\Phi_X = 0$  si et seulement si  $X \stackrel{\text{loi}}{=} -X$ .

**Théorème 6.1** (Caractérisation de la loi). On a équivalence entre

- (i)  $\Phi_X = \Phi_Y$
- (ii)  $X$  et  $Y$  ont même loi.

**Théorème 6.2** (Caractérisation de l'indépendance).  $X_1, \dots, X_p$  sont indépendants si et seulement si  $\Phi_{X_1+\dots+X_p} = \prod_{k=1}^p \Phi_{X_k}$ .

**Théorème 6.3.** Soit  $X$  une v.a.r. de moment d'ordre  $p > 0$ . Alors  $\Phi_X$  est de classe  $\mathcal{C}^p$  et on a :

$$\frac{d^p \Phi_X}{dt^p}(t) = i^p \mathbf{E}[X^p e^{itX}]$$

En particulier, on a  $\mathbf{E}[X^p] = (-1)^p i^p \frac{d^p \Phi_X}{dt^p}(0)$ .

## 7 Vecteurs gaussiens

Soit  $X$  un vecteur aléatoire de  $\mathbb{R}^d$  possédant des moments d'ordre 2.

**Définition 7.1** (Loi gaussienne). La *loi gaussienne* (ou *loi normale*) de paramètres  $\mu \in \mathbb{R}, \sigma^2 \geq 0$ , notée  $\mathcal{N}(\mu, \sigma^2)$ , est la loi définie par :

- Si  $\sigma = 0$ , c'est la *mesure de Dirac* en  $\mu$  ;
- Si  $\sigma > 0$ , sa densité est :  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ .

**Définition 7.2** (Vecteur gaussien).  $X$  est un *vecteur gaussien* si et seulement si  $\forall a \in \mathbb{R}^d$  la loi de  $\langle a, X \rangle$  est une loi gaussienne (de paramètres  $a^T \mu, a^T \Gamma a$ ). Alors  $\forall k, X_k \sim \mathcal{N}(\mu_k, \Gamma_{k,k})$ .

**Théorème 7.1.** On a équivalence entre :

- (i)  $X$  est un vecteur gaussien d'espérance  $m \in \mathbb{R}^d$  et de matrice de covariance  $\Gamma$  ;
- (ii) La fonction caractéristique du vecteur aléatoire  $X$  est  $\Phi_X(t) = e^{i\langle t, m \rangle - \frac{1}{2}t^T \Gamma t}$ .

**Théorème 7.2.** Soient  $(X_k)_{1 \leq k \leq d}$  des v.a. réelles. On a équivalence entre :

- (i) Le vecteur aléatoire  $(X_1, \dots, X_d)$  est un vecteur gaussien et  $\forall i \neq j, \text{Cov}(X_i, X_j) = 0$ .
- (ii) Les v.a.  $X_1, \dots, X_d$  sont des v.a. gaussiennes indépendantes.

**Propriété 7.1.** Soient  $A \in \mathcal{M}_{p,d}(\mathbb{R})$  et  $d \in \mathbb{R}^d$ . Soit  $X \sim \mathcal{N}_d(m, \Gamma)$ . Alors  $AX + b$  est un vecteur gaussien d'espérance  $Am + b$  et de matrice de covariance  $A\Gamma A^T$ .

**Propriété 7.2.** Soient  $X_1, \dots, X_d$  des vecteurs aléatoires gaussiens indépendants tels que  $X_i \sim \mathcal{N}_d(m_i, \Gamma_i)$ . Alors on a :  $X_1 + \dots + X_d \sim \mathcal{N}_d\left(\sum_{k=1}^d m_k, \sum_{k=1}^d \Gamma_k\right)$

**Théorème 7.3 (Densité).** Soit  $X \sim \mathcal{N}_d(m, \Gamma)$ . On distingue 2 cas :

- Lorsque  $\Gamma$  n'est pas inversible,  $X$  n'admet pas de densité.
- Lorsque  $\Gamma$  est inversible, la densité est :  $f(x) = \frac{1}{\sqrt{2\pi}^d} \frac{1}{\sqrt{\det(\Gamma)}} e^{-\frac{1}{2}(x-m)^T \Gamma^{-1}(x-m)}$

## 8 Convergences

**Définition 8.1** (Convergence presque-partout). On dit qu'une suite de v.a.  $(X_n)$  converge **P**-presque-partout vers une v.a.  $X$  lorsqu'il existe un ensemble  $A$  tel que  $\mathbf{P}(A^c) = 0$  et qu'on a, pour tout  $\omega \in A$ ,  $X_n(\omega) \xrightarrow[n \rightarrow +\infty]{} X(\omega)$ .

**Définition 8.2** (Convergence en loi). On dit qu'une suite de v.a.  $(X_n) \in (\mathbb{R}^k)^\mathbb{N}$  converge en loi vers  $X$  lorsque l'une des propriétés équivalentes suivantes est vérifiée :

- Pour toute fonction continue bornée  $f : \mathbb{R}^k \rightarrow \mathbb{R}$ ,  $\mathbf{E}[f(X_n)] \xrightarrow[n \rightarrow +\infty]{} \mathbf{E}[f(X)]$  ;
- Pour tout  $x \in \mathbb{R}^k$  tel que  $\mathbf{P}(X = x) = 0$ ,  $\mathbf{P}(X_n \leq x) \xrightarrow[n \rightarrow +\infty]{} \mathbf{P}(X \leq x)$  ;
- Pour tout  $t \in \mathbb{R}^k$ ,  $\Phi_{X_n}(t) \xrightarrow[n \rightarrow +\infty]{} \Phi_X(t)$ .

**Théorème 8.1.** La convergence presque-partout implique la convergence en loi mais la réciproque est fausse.

**Théorème 8.2** (Loi forte des grands nombres). Soit  $(X_n)$  une suite de v.a. indépendantes, identiquement distribuées telle que  $\mathbf{E}[|X_0|] < +\infty$ . Alors :

$$\frac{1}{n} \sum_{k=0}^n X_k \xrightarrow[n \rightarrow +\infty]{} \mathbf{E}[X_0], \quad \mathbf{P} - p.p.$$

**Théorème 8.3** (Théorème central limite). Soit  $(X_n)$  une suite de v.a. indépendantes, identiquement distribuées telle que  $\mathbf{E}[|X_0|^2] < +\infty$ . Alors, en notant  $\sigma^2 = \text{Var}(X_0)$  :

$$\frac{\sqrt{n}}{\sigma} \left( \frac{1}{n} \sum_{k=0}^n (X_k - \mathbf{E}[X_0]) \right) \xrightarrow[n \rightarrow +\infty]{\text{en loi}} \mathcal{N}(0, 1)$$