

PROBABILITES POUR L'INGENIEUR

COURS DE MASTER 1

Lionel Truquet

2011-2012

Table des matières

1	Espaces probabilisés, variables aléatoires et espérance mathématique	5
1.1	Espace probabilisable	5
1.2	Mesures de probabilités	7
1.2.1	Les mesures de probabilité discrètes.	8
1.2.2	Les mesures de probabilité à densité sur un intervalle de \mathbb{R}	10
1.2.3	Cas mixte. Mélange	11
1.2.4	Propriétés des mesures de probabilité	12
1.2.5	Indépendance d'événements	13
1.3	Les variables aléatoires	14
1.4	L'espérance mathématique	17
1.4.1	Espérance d'une variable aléatoire positive	17
1.4.2	Espérance d'une variable aléatoire de signe quelconque et propriétés	20
2	Indépendance. Variables aléatoires discrètes	25
2.1	Indépendance entre variables aléatoires	25
2.2	Les variables aléatoires discrètes	28
2.2.1	Les lois discrètes usuelles en modélisation	28
2.3	Loi d'un couple. Loi marginale	30
2.4	Le théorème de transfert	31
2.5	Les lois conditionnelles	32
3	Mesures et intégration	35
3.1	Mesure sur une tribu	35
3.1.1	Les mesures de Lebesgue-Stieltjes	36
3.1.2	Un exemple qui heurte l'intuition : l'escalier de Cantor	38
3.2	Intégrale d'une fonction mesurable par rapport à une mesure	39
3.2.1	Intégration par rapport à une mesure discrète.	42
3.2.2	L'intégrale de Lebesgue et l'intégrale de Riemann	43
3.2.3	Le presque partout	45
3.3	Mesure produit et théorème de Fubini	46
3.4	Mesure à densité et théorème de transfert	48
3.5	La formule du changement de variables	52
4	Les variables aléatoires à densité	55
4.1	Exemples de lois à densité	55
4.2	Densités marginales. Indépendance	56
4.3	Les calculs de lois en pratique.	59
4.4	Densités conditionnelles.	62

5	Les outils analytiques classiques en probabilité	65
5.1	La fonction de répartition	65
5.1.1	Propriétés générales des fonctions de répartition	65
5.1.2	Fonction de répartition inverse et simulation	66
5.2	Covariance et moments	68
5.2.1	Moments d'une variable aléatoires réelle	68
5.2.2	Covariance et corrélation	69
5.3	Fonction caractéristique d'une variable aléatoire	70
6	Loi et espérance conditionnelle	73
6.1	Généralisation des lois conditionnelles	73
6.2	Espérance conditionnelle	75
6.2.1	Variable aléatoire mesurable par rapport à une sous-tribu	75
6.2.2	Espérance conditionnelle par rapport à une sous-tribu	76
6.2.3	Propriétés générales de l'espérance conditionnelle	80
7	Les lois gaussiennes	85
7.1	Lois gaussiennes sur \mathbb{R}	85
7.2	Les matrices de variance-covariance	86
7.2.1	Quelques rappels sur les matrices	86
7.2.2	Vecteurs aléatoires, variance et covariance	86
7.3	Les vecteurs gaussiens	88
7.4	Quelques lois fondamentales pour la statistique	92
8	Convergence des suites de variables aléatoires	93
8.1	Comportement asymptotique d'une suite d'événements	93
8.1.1	Limite inférieure et supérieure	93
8.1.2	Lemme de Borel-Cantelli	94
8.2	Les modes de convergence	95
8.2.1	Convergence presque sûre et en probabilité	95
8.2.2	La convergence en loi	98
8.3	Deux théorèmes fondamentaux	101
8.3.1	La loi des grands nombres	101
8.3.2	Le théorème central limite	102

Chapitre 1

Espaces probabilisés, variables aléatoires et espérance mathématique

1.1 Espace probabilisable

Définition 1 On appelle espace probabilisable un couple (Ω, \mathcal{A}) où

- Ω est un ensemble appelé univers. Ses éléments ω sont appelés des épreuves.
- \mathcal{A} est une tribu sur Ω , c'est à dire que \mathcal{A} est un sous-ensemble de l'ensemble des parties de Ω vérifiant les trois propriétés suivantes :
 1. $\Omega \in \mathcal{A}$,
 2. Si $A \in \mathcal{A}$ alors A^c (l'événement complémentaire de A) appartient aussi à \mathcal{A} .
 3. Pour toute suite $(A_i)_{i \in \mathbb{N}}$ d'éléments de \mathcal{A} , on a $\cup_{i \in \mathbb{N}} A_i \in \mathcal{A}$.

Si $A \in \mathcal{A}$, alors A est appelé un événement.

Exemple. On peut munir n'importe quel ensemble Ω d'une tribu en considérant l'ensemble $\mathcal{P}(\Omega)$ de toutes ses parties. En effet l'ensemble $\mathcal{P}(\Omega)$ vérifie trivialement les trois points de la définition précédente. Lorsque l'univers est fini (par exemple $\Omega = \{1, 2, 3, 4, 5, 6\}$ pour décrire l'ensemble des résultats possibles lors d'un lancer de dé) ou alors infini mais dénombrable (par exemple $\Omega = \mathbb{N}$ pour décrire le nombre de connexions à un serveur entre deux instants donnés) la tribu $\mathcal{P}(\Omega)$ est assez naturelle, puisque on pourra très facilement construire des mesure de probabilité dessus. En revanche, si on souhaite choisir un univers non dénombrable (par exemple $\Omega = [0, +\infty[$ pour décrire les durées de vie possibles d'un appareil électrique), il est en fait impossible de construire mathématiquement certaines mesures de probabilités naturelles sur la tribu des événements $\mathcal{P}(\Omega)$. Dans ce cas, il faut se restreindre à une tribu plus petite (la tribu des Boréliens que nous verrons un peu plus loin).

Remarques

- Si \mathcal{A} est une tribu, on a toujours $\emptyset = \Omega^c \in \mathcal{A}$ (\emptyset désigne l'ensemble vide).
- Ainsi, une conséquence de la définition précédente est qu'une tribu d'événements est stable par réunion finie. En effet, si $A_1, \dots, A_N \in \mathcal{A}$, on a quitte à compléter par $A_i = \emptyset$ si $i = 0$ ou $i \geq N + 1$,

$$A = \cup_{i=1}^N A_i = \cup_{i \in \mathbb{N}} A_i \in \mathcal{A}.$$

- On en déduit aussi la stabilité de \mathcal{A} par intersection finie ou infinie dénombrable. En effet, lorsque $I = \{1, \dots, N\}$ ou $I = \mathbb{N}$, considérons une suite $(A_i)_{i \in I}$ d'éléments de \mathcal{A} . Posons $A = \cap_{i \in I} A_i$. En écrivant

$$A = (\cup_{i \in I} A_i^c)^c$$

et en utilisant la définition et la remarque précédente, on peut voir que $A \in \mathcal{A}$. En effet, on a $A_i^c \in \mathcal{A}$ pour $i \in I$ et la Définition 1 garantit que $\cup_{i \in I} A_i^c \in \mathcal{A}$ puis que $A \in \mathcal{A}$ (stabilité par passage au complémentaire).

Ainsi si A et B sont deux éléments de \mathcal{A} alors les quantités $A \setminus B = A \cap B^c$ (se lit A privé de B) et $B \setminus A = B \cap A^c$ sont encore des éléments de \mathcal{A} .

Les exemples intéressants de tribus peuvent se décrire à l'aide de la notion de tribu engendrée.

Définition 2 Soit $\mathcal{F} \subset \mathcal{P}(\Omega)$ un sous-ensemble de l'ensemble des parties de Ω . La tribu engendrée par \mathcal{F} est définie comme l'intersection de toutes les tribus contenant \mathcal{F} . Cette tribu est notée $\sigma(\mathcal{F})$.

Remarque. La définition précédente a bien un sens car on peut montrer qu'une intersection quelconque de tribus sur Ω est encore une tribu sur Ω . La tribu $\sigma(\mathcal{F})$ est donc la plus petite tribu contenant \mathcal{F} au sens de l'inclusion : toute tribu contenant \mathcal{F} contient également $\sigma(\mathcal{F})$.

Exemples

- Si $A \subset \Omega$, $\sigma(\{A\}) = \{\emptyset, \Omega, A, A^c\}$. On parle de tribu engendrée par A .
- Si A et B sont deux sous-ensembles disjoints de Ω et $\mathcal{F} = \{A, B\}$, on trouve

$$\sigma(\mathcal{F}) = \{\emptyset, \Omega, A, B, A \cup B, A^c, B^c, A^c \cap B^c = (A \cup B)^c\}.$$

- Si A et B sont deux sous-ensembles non disjoints de Ω et $\mathcal{F} = \{A, B\}$, alors la tribu $\sigma(\mathcal{F})$ est plus difficile à décrire. On peut arriver à décrire cette tribu à l'aide des événements $B \setminus A$, $A \setminus B$, $A \cap B$ et $(A \cup B)^c$ qui sont disjoints deux à deux et dont la réunion est Ω . En effet la tribu sera alors composée de toutes les réunions possibles entre ces éléments et le nombre d'éléments composant la tribu est $\sum_{k=0}^4 \binom{4}{k} = 16$. Ces 16 éléments sont

$$\emptyset, A \cap B^c, B \cap A^c, A^c \cap B^c, A \cap B, A, B, A \Delta B, (A \Delta B)^c, A^c, B^c,$$

$$A^c \cup B, B^c \cup A, A^c \cup B^c, A \cup B, \Omega,$$

avec $A \Delta B = (A \setminus B) \cup (B \setminus A)$ qui est appelé la différence symétrique de A et de B .

Exemple fondamental : la tribu borélienne sur \mathbb{R} . Posons $\mathcal{F} = \{[a, b] / a, b \in \mathbb{R}\}$. Autrement dit, \mathcal{F} est l'ensemble de tous les intervalles fermés et bornés de \mathbb{R} . Alors la tribu $\sigma(\mathcal{F})$ est aussi notée $\mathcal{B}(\mathbb{R})$ et est appelée tribu Borélienne sur \mathbb{R} . Les éléments de $\mathcal{B}(\mathbb{R})$ (qui sont des parties de \mathbb{R}) sont appelés des boréliens. La tribu $\mathcal{B}(\mathbb{R})$ est très riche : la définition d'une tribu entraîne que tous les intervalles de \mathbb{R} sont contenus dans $\mathcal{B}(\mathbb{R})$, mais aussi tous les ensembles dénombrables $\{x_1, x_2, \dots\}$ (en particulier \mathbb{N} , \mathbb{Z} ou encore \mathbb{Q}) et donc ensuite n'importe quelle intersection ou réunion finie ou infinie dénombrable constituée à l'aide de ces ensembles. On peut quand même montrer (bien que cela soit délicat) qu'il existe des parties de \mathbb{R} qui n'appartiennent pas à $\mathcal{B}(\mathbb{R})$. Sur la tribu d'événements $\mathcal{B}(\mathbb{R})$, il est possible de construire les probabilités dites à densité. Ces probabilités à densité ne peuvent être définies sur $\mathcal{P}(\mathbb{R})$, d'où l'importance de la tribu Borélienne.

On définit aussi une tribu Borélienne sur un intervalle I de \mathbb{R} : $\mathcal{B}(I) = \{A \cap I / A \in \mathcal{B}(\mathbb{R})\}$. On peut montrer que pour un intervalle I de \mathbb{R} , la tribu $\mathcal{B}(I)$ coïncide avec la tribu engendrée par l'ensemble des intervalles fermés bornés inclus dans I .

Produit d'espaces probabilisables. Considérons pour $i = 1, \dots, d$, un espace probabilisable $(\Omega_i, \mathcal{A}_i)$. Sur l'univers $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_d$, on définit la tribu produit par

$$\mathcal{A} = \sigma(\{A_1 \times A_2 \times \dots \times A_d / A_1 \in \mathcal{A}_1, \dots, A_d \in \mathcal{A}_d\}).$$

La tribu produit est donc la tribu engendrée par tous les produits cartésiens de la forme $A_1 \times \dots \times A_d$ avec $A_i \in \mathcal{A}_i$, $1 \leq i \leq d$.

Un exemple fondamental est la tribu borélienne sur \mathbb{R}^d : en considérant $\Omega_i = \mathbb{R}$ et $\mathcal{A}_i = \mathcal{B}(\mathbb{R})$ pour $1 \leq i \leq d$, la tribu produit est appelée la tribu borélienne sur \mathbb{R}^d et est notée $\mathcal{B}(\mathbb{R}^d)$. On peut montrer que $\mathcal{B}(\mathbb{R}^d)$ est aussi la tribu engendrée par les pavés fermés bornés (c'est à dire les ensembles de la forme $[a_1, b_1] \times [a_2, b_2] \times \dots \times [a_d, b_d]$). La tribu $\mathcal{B}(\mathbb{R}^d)$ est aussi très riche car elle contient tout un tas de parties de \mathbb{R}^d à cause des propriétés de stabilité de la Définition 1. On peut y trouver tous les pavés de \mathbb{R}^d (dont les intervalles sont fermés ou pas en leurs extrémités), tous les ensembles finis ou infini dénombrables, tous les ensembles définis par des équations du type $g(x) = 0$ ou $g(x) > 0$ avec $g : \mathbb{R}^d \rightarrow \mathbb{R}$ continue (en particulier un cercle ou un disque de \mathbb{R}^2 appartiennent à $\mathcal{B}(\mathbb{R}^2)$) et des ensembles beaucoup plus irréguliers. C'est sur cette tribu d'événements que sont définies les mesures de probabilités à densité (sur \mathbb{R}^d) que nous verrons en détail au Chapitre 3.

Produit infini d'espaces. Lorsqu'on souhaite par exemple étudier la suite (finie) des prix d'un actif financier, il arrive souvent que l'on ne puisse obtenir des résultats sur le comportement statistique de cette suite que lorsque son nombre d'éléments tend vers l'infini. Du point de vue de la modélisation probabiliste, il est alors nécessaire de définir un univers Ω comme un sous-ensemble de l'ensemble des suites de nombres réels. Ceci conduit à la notion de produit infini d'espaces mesurables $(\Omega_i, \mathcal{A}_i)$, $i \in \mathbb{N}$. L'univers $\Omega = \Omega_0 \times \Omega_1 \times \dots$ est alors muni de la tribu engendrée par l'ensemble \mathcal{C} des cylindres : un cylindre C est un sous-ensemble de Ω de la forme

$$C = \{\omega \in \Omega / \omega_0 \in A_0, \omega_1 \in A_1, \dots, \omega_n \in A_n\},$$

où $n \in \mathbb{N}$ et $A_i \in \mathcal{A}_i$ si $1 \leq i \leq n$. Lorsque pour $i \in \mathbb{N}$, $\Omega_i = \mathbb{R}$ et $\mathcal{A}_i = \mathcal{B}(\mathbb{R})$, la tribu $\sigma(\mathcal{C})$ contient par exemple les événements :

- ensemble des suites qui franchissent le niveau a :

$$\bigcup_{n=0}^{+\infty} \{x / x_n \geq a\}.$$

- ensemble des suites qui ont pour limite le réel a :

$$\bigcap_{k=1}^{+\infty} \left(\bigcup_{N=1}^{+\infty} \left(\bigcap_{n=N}^{+\infty} \left\{ x \in \Omega / |x_n - a| \leq \frac{1}{k} \right\} \right) \right).$$

Cette tribu, appelée tribu cylindrique, est en particulier utilisée pour la construction des suites de variables aléatoires indépendantes.

1.2 Mesures de probabilités

Définition 3 Soit (Ω, \mathcal{A}) un espace probabilisable. On dit que $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ est une mesure de probabilité si

1. $\mathbb{P}(\Omega) = 1$.
2. Si $(A_i)_{i \in \mathbb{N}}$ est une suite d'événements de \mathcal{A} deux à deux disjoints (c'est à dire $A_i \cap A_j = \emptyset$ si $i \neq j$) alors

$$\mathbb{P}(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i=0}^{\infty} \mathbb{P}(A_i).$$

Cette deuxième propriété est appelée la propriété de σ -additivité.

Si $A \in \mathcal{A}$, le nombre $\mathbb{P}(A)$ est appelé probabilité de l'événement A .

Remarques

- Soit $(A_i)_{i \in \mathbb{N}}$ une famille d'éléments de \mathcal{A} disjoints deux à deux. Posons $A = \cup_{i \in \mathbb{N}} A_i$. Si $\pi : \mathbb{N} \rightarrow \mathbb{N}$ est une permutation de \mathbb{N} , alors on a aussi l'égalité

$$A = \cup_{i \in \mathbb{N}} A_{\pi(i)}.$$

Pour que la définition précédente ait un sens, il faut que

$$\sum_{i=0}^{\infty} \mathbb{P}(A_i) = \sum_{i=0}^{\infty} \mathbb{P}(A_{\pi(i)}). \quad (1.1)$$

Mais cette égalité est automatiquement vérifiée car si \mathbb{P} est une mesure de probabilité, la série $\sum_{i=0}^{\infty} \mathbb{P}(A_i)$ est absolument convergente et de somme $\mathbb{P}(A)$. Une série absolument convergente étant commutativement convergente (on rappelle qu'une série $\sum_{i \in \mathbb{N}} x_i$ est dite commutativement convergente si elle est convergente et si sa somme coïncide avec celle de la série $\sum_{i=0}^{+\infty} x_{\pi(i)}$ pour toute permutation π de \mathbb{N}), l'égalité (1.1) est bien vérifiée.

Mentionnons également que pour toute série de terme général x_i positif, la limite $\lim_{N \rightarrow \infty} \sum_{i=1}^N x_i$ existe et est soit finie soit égale à $+\infty$. Cette limite est encore appelée la somme de la série. Une série à termes positifs est aussi commutativement convergente.

- Remarquons que cette définition entraîne automatiquement l'égalité $\mathbb{P}(\emptyset) = 0$. En effet, il suffit d'appliquer le point 2. de la définition en posant $A_i = \emptyset$ pour tout $i \in \mathbb{N}$: l'égalité

$$\mathbb{P}(\emptyset) = \sum_{i=0}^{\infty} \mathbb{P}(\emptyset) \text{ entraîne nécessairement l'égalité } \mathbb{P}(\emptyset) = 0.$$

- Soient A_1, \dots, A_n des éléments de \mathcal{A} disjoints deux à deux. En posant $A_i = \emptyset$ si $i > n$ ou si $i = 0$ et en utilisant la remarque précédente, la propriété de σ -additivité entraîne l'égalité

$$\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i).$$

On dit alors que \mathbb{P} vérifie la propriété additivité.

Définition 4 On appelle espace probabilisé tout triplet $(\Omega, \mathcal{A}, \mathbb{P})$ où (Ω, \mathcal{A}) est un espace probabilisable et \mathbb{P} est une mesure de probabilité sur \mathcal{A} .

Sur tout espace probabilisable (Ω, \mathcal{A}) , on peut construire des mesures de probabilité élémentaires : les masses de Dirac. Plus précisément, si $\omega_0 \in \Omega$, considérons δ_{ω_0} l'application définie sur \mathcal{A} par

$$\delta_{\omega_0}(A) = \begin{cases} 1 & \text{si } \omega_0 \in A, \\ 0 & \text{sinon} \end{cases}, \quad A \in \mathcal{A}.$$

Alors δ_{ω_0} est une mesure de probabilité appelée masse de Dirac au point ω_0 (vérifier la définition à titre d'exercice). Remarquons que si $\{\omega_0\} \in \mathcal{A}$ alors $\delta_{\omega_0}(\{\omega_0\}) = 1$. Cette mesure de probabilité a peu d'intérêt en modélisation (l'événement $\{\omega_0\}$ est certain) et permettra surtout d'écrire d'autres mesures de probabilité plus intéressantes.

1.2.1 Les mesures de probabilité discrètes.

Proposition 1 Soient $(x_i)_{i \in \mathbb{N}}$ une suite de points de \mathbb{R}^d et $(p_i)_{i \in \mathbb{N}}$ une suite de nombres réels positifs de somme 1. Posons $\Omega = \{x_i : i \in \mathbb{N}\}$ et $\mathcal{A} = \mathcal{P}(\Omega)$. Alors sur l'espace probabilisable (Ω, \mathcal{A}) , il existe une unique mesure de probabilité \mathbb{P} telle que

$$\mathbb{P}(\{x_i\}) = p_i, \quad i \in \mathbb{N}.$$

Pour tout $A \in \mathcal{A}$, on a alors

$$\mathbb{P}(A) = \sum_{i=0}^{\infty} \mathbb{1}_A(x_i) p_i,$$

où $\mathbb{1}_A : \mathbb{R}^d \rightarrow \{0,1\}$ est la fonction dite indicatrice de A et qui est définie par $\mathbb{1}_A(x) = 1$ si $x \in A$ et $\mathbb{1}_A(x) = 0$ si $x \in A^c$.

Avec les notations de la proposition précédente, on peut noter que

$$\sum_{i=0}^{\infty} \mathbb{1}_A(x_i) p_i = \sum_{i=0}^{\infty} \delta_{x_i}(A) p_i,$$

ce qui justifie la notation $\mathbb{P} = \sum_{i=0}^{\infty} p_i \delta_i$.

Preuve. Si \mathbb{P} est une mesure de probabilité telle que $\mathbb{P}(\{x_i\}) = p_i$ pour tout $i \in \mathbb{N}$, alors en posant $A_i = \{x_i\}$ si $x_i \in A$ et $A_i = \emptyset$ si $i \notin A$, on a nécessairement

$$\mathbb{P}(A) = \mathbb{P}(\cup_{i \in \mathbb{N}} A_i) = \sum_{i=0}^{\infty} \mathbb{P}(A_i) = \sum_{i=0}^{\infty} \mathbb{1}_A(x_i) p_i. \quad (1.2)$$

L'unicité d'une telle mesure de probabilité est donc claire. Montrons alors si $\mathbb{P} : \mathcal{A} \rightarrow [0,1]$ est définie pour tout $A \in \mathcal{A}$ par (1.2) alors \mathbb{P} est bien une mesure de probabilité. \mathbb{P} prend des valeurs dans $[0,1]$ puisque $0 \leq \mathbb{P}(A) \leq \sum_{i=0}^{\infty} p_i = 1$ pour tout $A \in \mathcal{A}$. On a ensuite d'abord $\mathbb{P}(\Omega) = \sum_{i=0}^{\infty} 1 \times p_i = 1$. Vérifions la propriété de σ -additivité. Soit $(A_n)_{n \in \mathbb{N}}$ est une suite d'éléments de \mathcal{A} deux à deux disjoints. Posons $A = \cup_{n \in \mathbb{N}} A_n$. On peut remarquer l'égalité

$$\mathbb{1}_A = \sum_{n=0}^{\infty} \mathbb{1}_{A_n}.$$

Ainsi on obtient

$$\mathbb{P}(A) = \sum_{i=0}^{\infty} \sum_{n=0}^{\infty} \mathbb{1}_{A_n}(x_i) p_i = \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} \mathbb{1}_{A_n}(x_i) p_i = \sum_{n=0}^{\infty} \mathbb{P}(A_n). \quad (1.3)$$

L'inversion des sommes dans (1.3) est justifiée grâce au Théorème de Fubini (cf Théorème 2 énoncé au Chapitre 2). \square

Exemples

- Lorsque $\Omega = \{1, \dots, n\}$, la loi uniforme sur Ω est définie par $\mathbb{P}(\{i\}) = \frac{1}{n}$, $1 \leq i \leq n$.
- Lorsque $\Omega = \{0, \dots, n\}$, la loi binomiale de paramètre (n, p) ($0 < p < 1$) est définie par

$$\mathbb{P}(\{i\}) = \binom{n}{p} p^i (1-p)^{n-i}, \quad 0 \leq i \leq n.$$

- Lorsque $\Omega = \mathbb{N}$, la loi de Poisson de paramètre $\lambda > 0$ est définie par

$$\mathbb{P}(\{i\}) = \exp(-\lambda) \frac{\lambda^i}{i!}, \quad i \in \mathbb{N}.$$

Remarquons que toute mesure de probabilité discrète \tilde{P} sur un sous-ensemble fini ou infini dénombrable Ω_0 de \mathbb{R}^d (e.g $\Omega_0 = \mathbb{N}$) peut être vue comme une mesure de probabilité sur $\Omega = \mathbb{R}^d$ muni de la tribu des boréliens. En effet, notons que $\mathcal{P}(\Omega_0) \subset \mathcal{B}(\mathbb{R}^d)$, et en posant $\mathbb{P}(A) = \tilde{P}(A \cap \Omega_0)$ pour $A \in \mathcal{B}(\mathbb{R}^d)$, on définit une mesure de probabilité \mathbb{P} sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ qui prolonge \tilde{P} . Cette observation justifie alors la définition suivante.

Définition 5 On dira qu'une mesure de probabilité \mathbb{P} sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est discrète s'il existe un sous-ensemble E de \mathbb{R}^d fini ou infini dénombrable tel que $\mathbb{P}(E) = 1$.

Le chapitre 2 sera consacré plus en détail aux mesures de probabilité discrètes.

1.2.2 Les mesures de probabilité à densité sur un intervalle de \mathbb{R}

Pour construire les mesures de probabilité à densité sur \mathbb{R} , la tâche est plus délicate. Certaines difficultés mathématiques rendent impossible la définition de ce type de probabilité sur la tribu d'événement $\mathcal{P}(\mathbb{R})$ (ensemble de toutes les parties de \mathbb{R}). On est obligé de se restreindre à la tribu des boréliens. Mais même avec cette restriction, il faut un bagage mathématique supplémentaire (l'intégrale de Lebesgue que nous discuterons un peu au Chapitre 3) pour pouvoir définir correctement ce type de mesure de probabilité. Cependant l'énoncé du théorème suivant met en lumière un principe assez général concernant la définition des mesures de probabilité : on définit une application \mathbb{P} sur l'ensemble des intervalles fermés bornés et on montre ensuite (à l'aide de théorèmes appropriés) que \mathbb{P} se prolonge de manière unique à la tribu engendrée par les intervalles, c'est à dire la tribu des boréliens.

Théorème 1 Si I est un intervalle de \mathbb{R} et $f : I \rightarrow \mathbb{R}_+$ une fonction intégrable. Alors sur $\Omega = I$ muni de la tribu des Boréliens $\mathcal{A} = \mathcal{B}(I)$, il existe une unique mesure de probabilité \mathbb{P} définie sur \mathcal{A} et telle que

$$\mathbb{P}([a, b]) = \int_a^b f(x)dx, \quad a, b \in I. \quad (1.4)$$

On dit que \mathbb{P} est une mesure de probabilité de densité f sur (Ω, \mathcal{A}) .

Contrairement aux mesures de probabilité discrètes, toute mesure \mathbb{P} à densité vérifie $\mathbb{P}(\{x\}) = 0$ pour tout x . L'utilisation de ce type de mesures de probabilité se justifie en pratique par la nature du phénomène observé : par exemple, si on observe des durées de vie d'appareils électriques du même type, les fréquences d'apparition d'une durée de vie donnée seront quasi nulles. On préfère plutôt préciser les probabilités qu'une durée de vie se situe dans un intervalle à l'aide d'une densité f : la valeur $\mathbb{P}([a, b])$ s'interprète alors comme l'aire de la partie située sous la courbe représentative de f et entre les droites d'équations $x = a$ et $x = b$ (voir Figure 1.1).

Cas particulier. Lorsque c et d sont deux réels tels que $c < d$, et si $f : [c, d] \rightarrow \mathbb{R}_+$ est définie par $f(x) = \frac{1}{d-c}$ pour $x \in [c, d]$ alors on parle de mesure de probabilité uniforme sur l'intervalle $[c, d]$. On a alors

$$\mathbb{P}([a, b]) = \frac{b-a}{d-c}, \quad c \leq a \leq b \leq d.$$

D'autres exemples bien connus : lorsque $I = \mathbb{R}$ et f est définie par $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ on parle de distribution gaussienne centrée réduite, lorsque $I = \mathbb{R}_+$ et f est définie par $f(x) = \lambda \exp(-\lambda x)$ on parle de distribution exponentielle de paramètre λ (λ désigne un réel strictement positif).

Remarque. Une mesure de probabilité à densité sur un intervalle I de \mathbb{R} peut être vue comme une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$: il suffit de prolonger la densité f en dehors de I en posant $f(x) = 0$ si $x \notin I$. On utilisera alors la notation (abusive) $f\mathbb{1}_I$ pour désigner la densité.

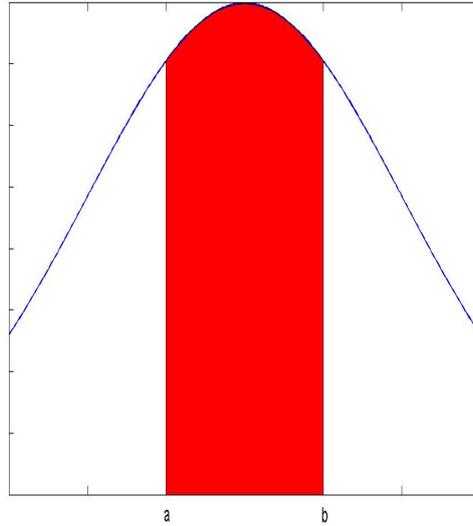


FIGURE 1.1: L'aire en rouge représente la probabilité $\mathbb{P}([a, b])$

1.2.3 Cas mixte. Mélange

On peut considérer d'autres exemples de mesures de probabilités sur \mathbb{R} en considérant des combinaisons convexes de mesures de probabilité discrètes et/ou à densité. La preuve de la proposition suivante est immédiate.

Proposition 2 *Toute combinaison convexe de mesures de probabilités sur (Ω, \mathcal{A}) est encore une mesure de probabilité. Autrement dit si $p \in [0, 1]$ et Q_1, Q_2 sont deux mesures de probabilités sur (Ω, \mathcal{A}) alors l'application $\mathbb{P} = pQ_1 + (1 - p)Q_2$ définie par $\mathbb{P}(A) = pQ_1(A) + (1 - p)Q_2(A)$ pour $A \in \mathcal{A}$ est une mesure de probabilité sur \mathcal{A} .*

Exemples

- Si Q_1 et Q_2 sont deux mesures de probabilité de densités respectives $f_1 : \mathbb{R} \rightarrow \mathbb{R}$ et $f_2 : \mathbb{R} \rightarrow \mathbb{R}$ et $p \in [0, 1]$, alors la probabilité $\mathbb{P} = pQ_1 + (1 - p)Q_2$ est une mesure de probabilité à densité. La densité f est alors donnée par la combinaison convexe correspondante des densités : $f = pf_1 + (1 - p)f_2$. On parle de densité mélange. Un mélange de loi permet de construire de nouvelles distributions empruntant les traits de plusieurs distributions élémentaires.
- On peut construire des mesures de probabilité dites mixtes qui ont une partie discrète et une partie à densité. Par exemple, considérons le temps d'attente d'un client qui arrive à un guichet. Il se peut que ce temps soit nul ou strictement positif. Une solution pour étudier cette expérience est de considérer l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ avec $\Omega = [0, +\infty[$, $\mathcal{A} = \mathcal{B}([0, +\infty[)$ et $\mathbb{P} = p\delta_0 + (1 - p)Q$ où $p \in]0, 1[$ et Q est la distribution exponentielle de paramètre $\lambda > 0$. Ainsi $\mathbb{P}(\{0\}) = p\delta(\{0\}) + (1 - p)Q(\{0\}) = p$ et si $[a, b]$ est un intervalle de \mathbb{R}_+ avec $a > 0$, on a

$$\mathbb{P}([a, b]) = (1 - p) \int_a^b \lambda \exp(-\lambda x) dx.$$

Cet exemple est un cas particulier de mesure de probabilité de la forme $\mathbb{P} = pQ_1 + (1 - p)Q_2$ où Q_1 est une mesure de probabilité discrète et Q_2 une mesure de probabilité à densité.

Remarque. La Définition 3 permet en fait la construction de mesure de probabilités qui ne possède ni une partie discrète ni une partie à densité. Pour ce type d'exemple, on aura $\mathbb{P}(\{x\}) = 0$ pour tout $x \in \mathbb{R}^d$ sans qu'il existe une densité. Dans le Chapitre 3, nous présenterons un exemple de ce type pour illustrer cette remarque.

1.2.4 Propriétés des mesures de probabilité

Voici maintenant des propriétés fondamentales des mesures de probabilité.

Proposition 3 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probablisé. On a

1. Si $A \in \mathcal{A}$, on a $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$.
2. Si A et B sont deux événements tels que $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$ et $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.
3. Si A et B sont deux événements, alors

$$\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B).$$

4. Pour toute suite $(A_n)_{n \in I}$ d'événements ($I = \mathbb{N}$ ou $\llbracket 1, N \rrbracket$) on a l'inégalité

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

5. Si $I = \mathbb{N}$ ou $\llbracket 1, N \rrbracket$ et $(A_n)_{n \in I}$ est un système complet d'événements (i.e disjoints deux à deux et de réunion Ω) alors

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap A_i).$$

Cette dernière formule est connue sous le nom de formule des probabilités totales.

Preuve

1. Les deux événements A et $B \setminus A$ sont disjoints et de réunion B . On a donc

$$\mathbb{P}(B) = \mathbb{P}(A \cup (B \setminus A)) = \mathbb{P}(A) + \mathbb{P}(B \setminus A).$$

On en déduit que $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$. Comme une probabilité est toujours positive, on en déduit également que $\mathbb{P}(B) \geq \mathbb{P}(A)$.

2. Les trois événements $A \setminus B$, $B \setminus A$ et $A \cap B$ sont disjoints deux à deux et leur réunion est $A \cup B$. On en déduit que

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B).$$

Vu que les événements $A \setminus B$ et $A \cap B$ d'une part et les événements $B \setminus A$ et $A \cap B$ d'autre part sont disjoints et de réunion A et B respectivement, on en déduit

$$\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B).$$

3. Posons $B_0 = A_0$ et pour $n \in \mathbb{N}^*$, $B_n = A_n \setminus \bigcup_{i=0}^{n-1} A_i$. Alors les événements B_n sont disjoints deux à deux et on peut vérifier l'égalité

$$\bigcup_{n \in \mathbb{N}} B_n = \bigcup_{n \in \mathbb{N}} A_n.$$

De plus comme $B_n \subset A_n$, on a $\mathbb{P}(B_n) \leq \mathbb{P}(A_n)$. Si $A = \bigcup_{n \in \mathbb{N}} A_n$, on conclut que

$$\mathbb{P}(A) = \sum_{n \in \mathbb{N}} \mathbb{P}(B_n) \leq \sum_{n \in \mathbb{N}} \mathbb{P}(A_n),$$

ce qui prouve l'inégalité annoncée.

4. Il suffit d'appliquer la propriété d'additivité ou de σ -additivité aux événements $B_n = A \cap A_n$, $n \in I$ qui sont disjoints deux à deux et de réunion A . \square

Proposition 4 Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements.

1. Si la suite est croissante au sens de l'inclusion, c'est à dire $A_n \subset A_{n+1}$ pour tout n , alors en posant $A = \cup_{n \in \mathbb{N}} A_n$, on a

$$\mathbb{P}(A) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

2. Si la suite est décroissante au sens de l'inclusion, c'est à dire $A_{n+1} \subset A_n$ pour tout n , alors en posant $A = \cap_{n \in \mathbb{N}} A_n$, on a également

$$\mathbb{P}(A) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

Remarque. On peut voir ces résultats comme des résultats de continuité : la propriété 1. est d'ailleurs appelée propriété de continuité supérieure et la propriété 2. propriété de continuité inférieure. Par exemple, pour toute mesure de probabilité \mathbb{P} sur \mathbb{R} muni de la tribu $\mathcal{B}(\mathbb{R})$, la Proposition 4 permet d'écrire :

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\left[-1 + \frac{1}{n}, 1 - \frac{1}{n} \right] \right) = \mathbb{P} (] - 1, 1 [),$$

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\left[-1 - \frac{1}{n}, 1 + \frac{1}{n} \right] \right) = \mathbb{P} ([-1, 1]).$$

La Proposition 4 sera démontrée en TD.

1.2.5 Indépendance d'événements

La définition de l'indépendance entre événements est fondamentale dans la théorie des probabilités.

Définition 6 – On dit que A et B sont deux événements indépendants si

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

- On dit qu'une famille d'événements $\{A_i / i \in I\}$ (I est un ensemble quelconque) est indépendante si pour tout $k \in \mathbb{N}^*$ et pour tout k -uplet (j_1, \dots, j_k) d'éléments distincts de I :

$$\mathbb{P}(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = \prod_{i=1}^k \mathbb{P}(A_{j_i}).$$

Exemple. Considérons l'exemple du lancer de deux dés en posant $\Omega = \llbracket 1, 6 \rrbracket^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ et \mathbb{P} la mesure de probabilité discrète définie par $\mathbb{P}(\{\omega_1, \omega_2\}) = \frac{1}{36}$. Alors pour toute partie A de Ω , on a $\mathbb{P}(A) = \frac{|A|}{36}$. Si A désigne l'événement "le premier dé donne 6" et B désigne l'événement "le deuxième dé donne 6", alors $A = \{6\} \times \llbracket 1, 6 \rrbracket$, $B = \llbracket 1, 6 \rrbracket \times \{6\}$ et $A \cap B = \{6\} \times \{6\}$. On a

$$\frac{1}{36} = \mathbb{P}(A \cap B) = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(A)\mathbb{P}(B).$$

Les événements A et B sont indépendants.

1.3 Les variables aléatoires

On se donne un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 7 1. On dit qu'une application $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire si pour tout couple (a, b) de nombres réels tels que $a \leq b$:

$$\{\omega \in \Omega / a \leq X(\omega) \leq b\} \in \mathcal{A}, \quad a \leq b.$$

2. On dit qu'une application $X : \Omega \rightarrow \mathbb{R}^d$ définie par $X(\omega) = (X_1(\omega), \dots, X_d(\omega))$ est une variable aléatoire si pour $1 \leq i \leq n$, X_i est une variable aléatoire réelle.

Cette définition semble naturelle car pour calculer la probabilité de l'ensemble $\{\omega \in \Omega / a \leq X(\omega) \leq b\}$ (si X désigne une variable aléatoire réelle), il faut que ce dernier soit bien un événement. En pratique, cette définition est assez générale : les fonctions rencontrées seront toujours des variables aléatoires lorsque la tribu \mathcal{A} est correctement choisie.

On pourra en fait calculer la probabilité qu'une variable aléatoire appartienne à n'importe quel Borélien, comme le montre la proposition suivante.

Proposition 5 Si $X : \Omega \rightarrow \mathbb{R}^d$ est une variable aléatoire, alors pour tout $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\{\omega \in \Omega / X(\omega) \in B\} \in \mathcal{A}.$$

Notation. Dans la suite, un événement du type $\{\omega \in \Omega / X(\omega) \in B\}$ sera noté en abrégé $\{X \in B\}$. Une autre notation que nous n'utiliserons pas souvent est $X^{-1}(B)$ (image réciproque de B par X).

Remarque. Il est alors facile de montrer qu'une fonction $X : \Omega \rightarrow \mathbb{R}^d$ est une variable aléatoire si et seulement si $\{X \in B\} \in \mathcal{A}$ pour tout $B \in \mathcal{B}(\mathbb{R}^d)$ (vérifier la condition suffisante en prenant pour B des produits cartésiens bien choisis).

Preuve de la proposition. Posons

$$\mathcal{M} = \left\{ B \in \mathcal{B}(\mathbb{R}^d) / \{X \in B\} \in \mathcal{A} \right\}.$$

On peut vérifier que \mathcal{M} est une tribu (exercice). Cette tribu contient les pavés fermés de \mathbb{R}^d : en effet comme X est une variable aléatoire, on a si $P = [a_1, b_1] \times \dots \times [a_d, b_d]$:

$$\{X \in P\} = \bigcap_{i=1}^d \{X_i \in [a_i, b_i]\} \in \mathcal{A},$$

en utilisant la stabilité de la tribu par intersection finie. Comme $\mathcal{B}(\mathbb{R}^d)$ est la plus petite tribu contenant les pavés, on conclut que $\mathcal{B}(\mathbb{R}^d) \subset \mathcal{M}$ (on a en fait l'égalité car $\mathcal{M} \subset \mathcal{B}(\mathbb{R}^d)$). D'où le résultat par définition de \mathcal{M} . \square

Exemples de variables aléatoires :

- Si $A \in \mathcal{A}$, on note $\mathbb{1}_A$ la fonction indicatrice de l'événement A . Cette fonction est définie par $\mathbb{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbb{1}_A(\omega) = 0$ sinon. Cette fonction est une variable aléatoire à valeurs dans $\{0, 1\}$. Pour le justifier, il suffit d'observer que $\{a \leq \mathbb{1}_A \leq b\}$ est égal à Ω , \emptyset , A ou A^c suivant le choix du couple (a, b) .

- Soit $X : \Omega \rightarrow \mathbb{R}^d$ une fonction telle que $X(\Omega)$ soit un ensemble fini ou infini dénombrable de \mathbb{R}^d : $X(\Omega) = \{x_1, x_2, \dots\}$ pour une suite x_1, x_2, \dots de points de \mathbb{R}^d . Alors X est une variable aléatoire si et seulement si pour tout i ,

$$\{X = x_i\} \in \mathcal{A}.$$

En effet, si a et b sont deux réels tels que $a \leq b$ et $J = \{i \in \mathbb{N} / x_i \in [a, b]\}$, alors J est un ensemble fini ou infini dénombrable et on a

$$\{a \leq X(\omega) \leq b\} = \cup_{i \in J} \{X(\omega) = x_i\}.$$

Dans ce cas, on dit que X est une variable aléatoire discrète. Remarquons qu'en posant $A_i = \{X = x_i\}$ pour $i \in \mathbb{N}^*$, on a la décomposition

$$X(\omega) = \sum_{i=1}^{\infty} x_i \mathbb{1}_{A_i}(\omega), \quad \omega \in \Omega.$$

- Soit un espace probabilisé produit $(\Omega, \mathcal{A}, \mathbb{P})$ avec $\Omega = \Omega_1 \times \dots \times \Omega_n$ et où $\Omega_i = \mathbb{R}^{d_i}$ pour $1 \leq i \leq n$. Ω est muni de la tribu produit

$$\mathcal{A} = \mathcal{B}(\mathbb{R}^{d_1}) \otimes \dots \otimes \mathcal{B}(\mathbb{R}^{d_n}),$$

tribu qui s'identifie en fait à la tribu borélienne sur $\mathbb{R}^{d_1 + \dots + d_n}$. Alors les applications coordonnées X_1, \dots, X_n définies par

$$X_i(\omega) = \omega_i, \quad \omega = (\omega_1, \dots, \omega_n) \in \Omega, \quad 1 \leq i \leq n,$$

sont des variables aléatoires. En effet si $B_i \in \mathcal{B}(\mathbb{R}^{d_i})$, on a, en posant $E_i = B_i$ et $E_j = \Omega_j$ pour $j \neq i$:

$$\{X_i \in B_i\} = \prod_{j=1}^n E_j \in \mathcal{A}.$$

- Il est parfois intéressant de considérer des variables aléatoires réelles pouvant prendre la valeur $+\infty$ (la définition reste inchangée). Donnons un exemple en considérant $\Omega = \mathbb{R}^{\mathbb{N}}$ que l'on munit de la tribu cylindrique et définissons pour $i \in \mathbb{N}$ l'application coordonnée X_i par $X_i(\omega) = \omega_i$, $\omega \in \Omega$. X_i est une variable aléatoire. Soit alors

$$T(\omega) = \inf \{i \geq 0 : X_i \geq a\},$$

avec la convention $T(\omega) = +\infty$ si $\{i \geq 0 : X_i \geq a\} = \emptyset$. Alors T est une variable aléatoire discrète. En effet, on peut écrire si $n \in \mathbb{N}$:

$$\{T = n\} = \cap_{i=0}^{n-1} \{X_i < a\} \cap \{X_n \geq a\}.$$

Ainsi $\{T = n\} \in \mathcal{A}$ car il s'agit d'une intersection finie d'événements.

Si on s'intéresse à la suite des prix d'un actif financier, T représente le premier instant où le prix de cet actif dépasse le seuil a .

Définition 8 Soit $X : \Omega \rightarrow \mathbb{R}^d$ une variable aléatoire. Alors l'application $\mathbb{P}_X : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ définie par

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B), \quad B \in \mathcal{B}(\mathbb{R}^d),$$

est une mesure de probabilité appelée la loi de X sous \mathbb{P} .

On pourra à titre d'exercice vérifier que l'application \mathbb{P}_X de la définition précédente est bien une mesure de probabilité.

Ainsi définir la loi d'une variable aléatoire X reviendra à définir la mesure de probabilité \mathbb{P}_X . En particulier, nous utiliserons la terminologie suivante.

Définition 9 1. On dira qu'une variable aléatoire X est une variable aléatoire discrète lorsque \mathbb{P}_X est une mesure de probabilité discrète.

2. On dira qu'une variable aléatoire X suit une loi à densité et de densité f sur \mathbb{R} si \mathbb{P}_X est une mesure de probabilité de densité f (ainsi si $a \leq b$, on aura $\mathbb{P}_X([a, b]) = \mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$.)

Exemples

- Supposons que X soit une variable aléatoire constante (c'est à dire qu'il existe $c \in \mathbb{R}^d$ tel que $X(\omega) = c$ pour tout $\omega \in \Omega$). Alors on a $\mathbb{P}_X = \delta_c$ (masse de Dirac au point c).
- Pour le lancer de deux dés équilibrés, on pose $\Omega = \llbracket 1, 6 \rrbracket^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ et \mathbb{P} la mesure de probabilité discrète définie par $\mathbb{P}(\{\omega\}) = \frac{1}{36}$, $\omega \in \Omega$. Alors si $X_1(\omega) = \omega_1$ et $X_2(\omega) = \omega_2$, alors \mathbb{P}_{X_1} ou \mathbb{P}_{X_2} est la probabilité uniforme sur $\llbracket 1, 6 \rrbracket$ et $\mathbb{P}_{(X_1, X_2)} = \mathbb{P}$.
- Pour une variable aléatoire discrète, il n'est utile de préciser que les probabilités non nulles du type $\mathbb{P}(X = x)$. Par exemple soit $X = \sum_{i=1}^{\infty} x_i \mathbf{1}_{A_i}$ où $(x_i)_{i \geq 1}$ est une suite de points de \mathbb{R}^d et $(A_i)_{i \geq 1}$ une suite d'éléments de \mathcal{A} disjoints deux à deux. Alors en posant $p_i = \mathbb{P}(A_i)$, on a l'égalité

$$\mathbb{P}_X = \sum_{i=1}^{\infty} p_i \delta_{x_i}.$$

- Pour une variable aléatoire à densité, il suffit de préciser la densité. Supposons par exemple que $\Omega = \mathbb{R}$ et \mathbb{P} est la distribution exponentielle de paramètre 1. Posons $X(\omega) = \omega^2$, pour tout $\omega \in \Omega$. Soient alors $(a, b) \in \mathbb{R}^2$ tel que $0 \leq a \leq b$. Alors, on a

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}([\sqrt{a}, \sqrt{b}]) = \int_{\sqrt{a}}^{\sqrt{b}} \exp(-t) dt = \int_a^b 2u \exp(-u^2) du,$$

la dernière égalité se déduisant du changement de variable $u = t^2$ dans l'intégrale. Ainsi, X suit une loi à densité f_X donnée par

$$f_X(u) = 2u \exp(-u^2) \mathbf{1}_{\mathbb{R}_+}(u).$$

Remarque fondamentale concernant la description d'une expérience aléatoire. En pratique, il arrive souvent que l'on définisse des lois de variables aléatoires sans même définir explicitement le triplet $(\Omega, \mathcal{A}, \mathbb{P})$. En fait, on définira souvent \mathbb{P}_X sans définir \mathbb{P} car seule la probabilité \mathbb{P}_X nous intéressera. Ceci est lié au fait qu'on supposera le résultat $(x_1, \dots, x_d) \in \mathbb{R}^d$ d'une expérience aléatoire comme étant la réalisation $X(\omega)$ d'une variable aléatoire. On définit alors la loi Q de X , qui est celle permettant d'étudier cette expérience. Il existe toujours un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ naturel associé : on peut poser $\Omega = \mathbb{R}^d$, $X(\omega) = \omega$ et $\mathbb{P} = Q$ (ce qui entraîne $\mathbb{P}_X = \mathbb{P} = Q$). Toute loi de probabilité sur \mathbb{R}^d est donc la loi d'une variable aléatoire définie sur un espace probabilisé.

Adopter ce point de vue est souvent plus commode car plus économe en terme de description. Nous y reviendrons.

1.4 L'espérance mathématique

La définition de l'espérance mathématique d'une variable aléatoire, on parle aussi de moyenne d'une variable aléatoire, peut se faire uniquement à partir la définition donnée pour une variable aléatoire discrète à support fini.

Définition 10 Si X est une variable aléatoire à valeurs réelles avec $X(\Omega) = \{x_1, \dots, x_n\}$ et $\mathbb{P}(X = x_i) = p_i$ pour $i = 1, \dots, n$, on définit

$$\mathbb{E}(X) = \sum_{i=1}^n x_i p_i.$$

$\mathbb{E}(X)$ est appelée espérance de X .

Cette définition correspond bien à l'intuition de l'approche fréquentiste : s'il est possible d'observer une quantité à plusieurs reprises, il est naturel de définir empiriquement la moyenne en pondérant les valeurs possibles par leur fréquence d'apparition. Dans la modélisation probabiliste, ces fréquences sont remplacées par des probabilités.

Considérons maintenant une variable aléatoire positive quelconque. Si on reproduisait plusieurs fois la même expérience qui aboutirait à l'observation d'une valeur de cette variable, il est naturel de regrouper les valeurs possibles en différentes classes (par exemple $[0, h[$, $[h, 2h[$, ...) pour réitérer l'approche faite en discret : on regarde le nombre de fois où la valeur de la variable tombe dans une classe donnée. A la quantité $\sum_{k \geq 0} kh f_k$ où f_k est la fréquence d'apparition de la classe $[kh, (k+1)h[$, correspondrait le nombre $\sum_{k \geq 0} kh \mathbb{P}(kh \leq X < (k+1)h)$ qui devrait donner une idée de plus en plus précise de la valeur moyenne lorsque $h \rightarrow 0$. Il y a là l'idée d'approcher une variable aléatoire quelconque par une variable aléatoire discrète.

L'approche mathématique pour définir l'espérance d'une variable aléatoire est aussi basée sur l'approximation d'une variable aléatoire par une variable aléatoire discrète (suivant une approche qui permet également de répondre à d'autres exigences théoriques que nous n'évoquerons pas).

1.4.1 Espérance d'une variable aléatoire positive

Dans la suite, nous noterons \mathcal{D}_+ l'ensemble des variables aléatoires positives et qui ne prennent qu'un nombre fini de valeurs. Commençons par énoncer la proposition suivante.

Proposition 6 Soit X une variable aléatoire à valeurs réelles positives. Pour $n \in \mathbb{N}$, soit

$$X_n = \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} \mathbb{1}_{\{\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\}}.$$

Alors $(X_n)_{n \in \mathbb{N}}$ est une suite croissante d'éléments de \mathcal{D}_+ qui converge point par point vers la variable aléatoire X .

Preuve. Commençons par montrer que pour tout $\omega \in \Omega$, on a $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$. Soit $\omega \in \Omega$ et $\epsilon > 0$. Soit également n_0 tel que $n \geq n_0$ entraîne $X(\omega) < 2^n$. Alors si $n \geq n_0$, il existe $k \in \llbracket 0, 2^{2n} - 1 \rrbracket$ tel que $\frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n}$ et donc $X_n(\omega) = \frac{k}{2^n}$ (remarquer que k dépend de n). Ainsi

$$|X_n(\omega) - X(\omega)| \leq \frac{k}{2^n} < \epsilon, \quad n \geq n_0,$$

ce qui montre la convergence. L'autre point non trivial est la croissance de la suite que nous allons démontrer. Soit $\omega \in \Omega$ et $n \in \mathbb{N}$. Si $X(\omega) \geq 2^n$, alors $X_n(\omega) = 0$ et donc $X_n(\omega) \leq X_{n+1}(\omega)$. Si

maintenant il existe $k \in \llbracket 0, 2^{2n} - 1 \rrbracket$ tel que $\frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n}$, alors on a $\frac{2k}{2^{n+1}} \leq X(\omega) < \frac{2k+1}{2^n}$ ou $\frac{2k+1}{2^{n+1}} \leq X(\omega) < \frac{2k+2}{2^{n+1}}$. Dans ce cas on a $X_n(\omega) = \frac{k}{2^n}$ et ou bien $X_{n+1}(\omega) = \frac{2k}{2^{n+1}}$ ou bien $X_{n+1}(\omega) = \frac{2k+1}{2^{n+1}}$. Ceci prouve bien que l'on a toujours $X_n(\omega) \leq X_{n+1}(\omega)$. \square

Remarque. La variable aléatoire X_n définie dans la Proposition 6 prend la valeur $\frac{k}{2^n}$ lorsque $\frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n}$. Lorsque $\Omega = \mathbb{R}_+$, la figure 1.2 représente le graphe d'une fonction X ainsi que les graphes de X_1 , X_2 et X_3 (ces fonctions sont constantes par morceaux pour la fonction X considérée).

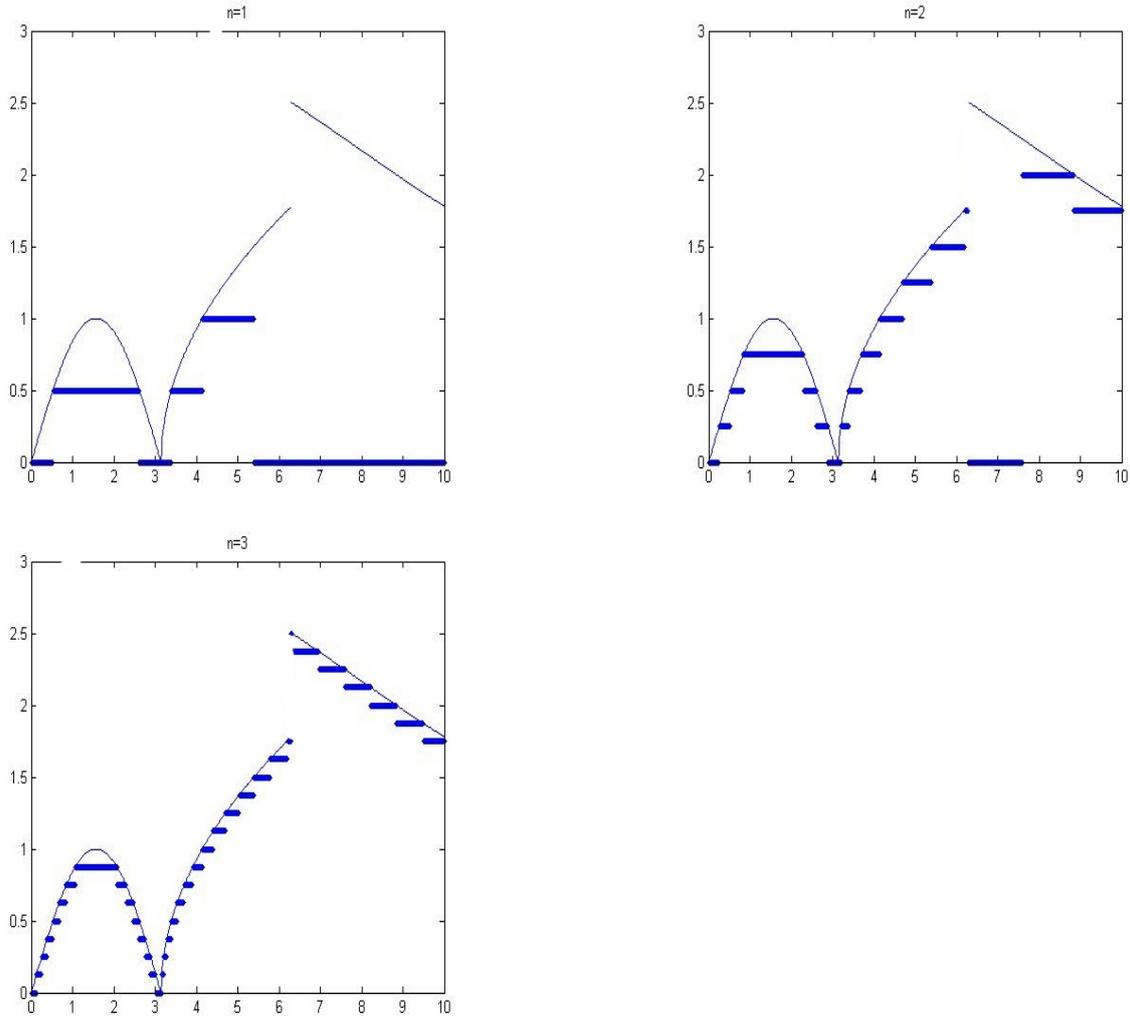


FIGURE 1.2:

Lemme 1 Soient Y et Z deux éléments de \mathcal{D}_+ . On suppose $X \leq Y$ (i.e pour tout $\omega \in \Omega$, $Y(\omega) \leq Z(\omega)$). Alors $\mathbb{E}(Y) \leq \mathbb{E}(Z)$.

Preuve. Quitte à rajouter des valeurs, on peut supposer que Y et Z prennent toutes les deux les valeurs $x_1 < x_2 < \dots < x_k$ (dont peut-être certaines avec probabilité 0). Comme $Y \leq Z$, alors pour $1 \leq i \leq n$, on a

$$\{Y = x_i\} = \cup_{j=i}^n \{Y = x_i, Z = x_j\}$$

et la réunion est formée d'événements disjoints deux à deux. En utilisant la propriété d'additivité de la mesure \mathbb{P} , on obtient

$$\mathbb{E}(Y) = \sum_{i=1}^n x_i \mathbb{P}(Y = x_i) = \sum_{i=1}^n x_i \sum_{j=i}^n \mathbb{P}(Y = x_i, Z = x_j).$$

Mais

$$\begin{aligned} \sum_{i=1}^n x_i \sum_{j=i}^n \mathbb{P}(Y = x_i, Z = x_j) &= \sum_{i=1}^n x_i \sum_{j=1}^n \mathbb{1}_{i \leq j} \mathbb{P}(Y = x_i, Z = x_j) \\ &= \sum_{j=1}^n \sum_{i=1}^j \mathbb{P}(Y = x_i, Z = x_j) \\ &\leq \sum_{j=1}^n x_j \sum_{i=1}^j \mathbb{P}(Y = x_i, Z = x_j) \\ &= \sum_{j=1}^n x_j \mathbb{P}(Z = x_j) \\ &= \mathbb{E}(Z). \end{aligned}$$

On a donc bien $\mathbb{E}(Y) \leq \mathbb{E}(Z)$. \square

Le lemme suivant sera capital pour justifier la définition de l'espérance.

Lemme 2 Soient $X \in \mathcal{D}_+$ et $(Y_n)_n$ une suite croissante d'éléments de \mathcal{D}_+ telle que $X \leq \lim_{n \rightarrow \infty} Y_n$. Alors

$$\mathbb{E}(X) \leq \lim_{n \rightarrow +\infty} \mathbb{E}(Y_n).$$

Preuve. Soit $0 \leq t < 1$. Posons si $n \in \mathbb{N}$, $B_n = \{Y_n \geq tX\}$. On peut facilement vérifier que sous les hypothèses du lemme, la suite d'événements $(B_n)_n$ est croissante pour l'inclusion et de réunion Ω . En utilisant le lemme précédent, on a les inégalités

$$\mathbb{E}(tX \mathbb{1}_{B_n}) \leq \mathbb{E}(Y_n \mathbb{1}_{B_n}) \leq \mathbb{E}(Y_n). \quad (1.5)$$

Remarquons ensuite que si $X = \sum_{i=1}^p x_i \mathbb{1}_{X=x_i}$, alors

$$\mathbb{E}(tX \mathbb{1}_{B_n}) = t \sum_{i=1}^p x_i \mathbb{P}(\{X = x_i\} \cap B_n).$$

En utilisant la propriété de continuité supérieure de \mathbb{P} , on obtient en faisant tendre n vers $+\infty$:

$$\lim_{n \rightarrow \infty} \mathbb{E}(tX \mathbb{1}_{B_n}) = t \sum_{i=1}^p x_i \mathbb{P}(X = x_i) = t\mathbb{E}(X).$$

En passant à la limite dans (1.5), nous obtenons

$$t\mathbb{E}(X) \leq \lim_{n \rightarrow \infty} \mathbb{E}(Y_n),$$

ce qui donne l'inégalité que nous devons prouver en faisant tendre t vers 1. \square

Nous sommes alors en mesure de définir l'espérance d'une variable aléatoire positive.

Proposition-Définition 1 Soit X une variable aléatoire positive et $(X_n)_{n \in \mathbb{N}}$ une suite croissante de variables aléatoires de \mathcal{D}_+ et convergente point par point vers X , alors la quantité $\lim_{n \rightarrow +\infty} \mathbb{E}(X_n)$ (qui est bien définie en tant que limite d'une suite croissante et qui peut valoir $+\infty$) ne dépend pas de la suite $(X_n)_{n \in \mathbb{N}}$. Ce nombre est appelé espérance mathématique de X et est noté $\mathbb{E}(X)$.

Preuve. Si $(X_n)_n$ et $(X'_n)_n$ sont deux suites de variables aléatoires qui satisfont les hypothèses de la proposition, alors le Lemme 2 assure que

$$\mathbb{E}(X_p) \leq \lim_{n \rightarrow \infty} \mathbb{E}(X'_n),$$

ce qui entraîne

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n) \leq \lim_{n \rightarrow \infty} \mathbb{E}(X'_n).$$

En inversant le rôle de la suite (X_n) et (X'_n) , on voit que l'inégalité inverse est également valable et donc que $\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \lim_{n \rightarrow \infty} \mathbb{E}(X'_n)$, ce qui justifie la définition donnée de l'intégrale. \square

Exemples

- Si X est une variable aléatoire discrète telle que $X(\Omega) \subset \mathbb{N}$. Plutôt que d'utiliser la suite croissante donnée dans la Proposition 6, on peut utiliser ici $X_n = \sum_{i=0}^n i \mathbb{1}_{\{X=i\}}$, ce qui donne :

$$\mathbb{E}(X) = \lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \lim_{n \rightarrow +\infty} \sum_{i=0}^n i \mathbb{P}(X=i) = \sum_{i \in \mathbb{N}} \mathbb{P}(X=i).$$

- Si X est une variable aléatoire à densité, de densité $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, on a en utilisant la Proposition 6

$$\mathbb{E}(X) = \lim_{n \rightarrow +\infty} \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} \mathbb{P}\left(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\right) = \lim_{n \rightarrow +\infty} \sum_{k=0}^{2^{2n}-1} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} \frac{k}{2^n} f(x) dx.$$

On pourra vérifier à titre d'exercice que la limite obtenue est $\int_0^{+\infty} f(x) dx$.

1.4.2 Espérance d'une variable aléatoire de signe quelconque et propriétés

Soit X une variable aléatoire prenant des valeurs réelles. La partie positive de X est la variable aléatoire notée X^+ définie par

$$X^+(\omega) = \begin{cases} X(\omega) & \text{si } X(\omega) \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

Remarquons que $X^+ = X \mathbb{1}_{X \geq 0}$. De même la partie négative X^- de X est définie par $X^- = -X \mathbb{1}_{X < 0}$. On peut voir que les variables aléatoires X^+ et X^- sont à valeurs positives et satisfont les égalités :

$$X = X^+ - X^-, \quad |X| = X^+ + X^-.$$

Définition 11 On dit qu'une variable aléatoire $X : \Omega \rightarrow \mathbb{R}$ est intégrable lorsque $\mathbb{E}(X^+) < +\infty$ et $\mathbb{E}(X^-) < +\infty$ (ce qui peut se résumer par $\mathbb{E}(|X|) < +\infty$, voir la proposition suivante). Dans ce cas, l'espérance de la variable aléatoire X est définie par

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

En utilisant la Proposition 6, on retrouve alors les formules déjà connues de la moyenne pour des variables aléatoires discrètes ou à densité (pouvant prendre des valeurs positives ou négatives).

- Soit X une variable aléatoire discrète à valeurs réelles et posons

$$\text{val}(X) = \{x \in \mathbb{R} : \mathbb{P}(X = x) \neq 0\} = \{x_1, x_2, \dots\},$$

qui est un sous-ensemble fini ou infini dénombrable de \mathbb{R} . Alors, on peut montrer que

$$\mathbb{E}(X^+) = \sum_{i=1}^{\infty} x_i \mathbb{1}_{x_i > 0} \mathbb{P}(X = x_i),$$

$$\mathbb{E}(X^-) = \sum_{i=1}^{\infty} -x_i \mathbb{1}_{x_i < 0} \mathbb{P}(X = x_i).$$

De plus, on a les égalités :

$$\mathbb{E}(X^+) + \mathbb{E}(X^-) = \sum_{i \geq 1} |x_i| \mathbb{P}(X = x_i),$$

et la somme de cette série correspond aussi à $\mathbb{E}(|X|)$. Lorsque cette somme est finie, l'espérance de X est donnée par la formule

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i).$$

Il est important de rappeler que pour une série à termes positifs ou absolument convergente, il est possible de sommer sans ambiguïté en utilisant une énumération quelconque.

- Si X est une variable aléatoire dont la loi a une densité f définie sur \mathbb{R} . Alors on trouve,

$$\mathbb{E}(X^+) = \int_0^{+\infty} x f(x) dx,$$

$$\mathbb{E}(X^-) = \int_{-\infty}^0 -x f(x) dx.$$

Lorsque ces deux quantités sont finies, leur somme vaut $\int_{-\infty}^{+\infty} |x| f(x) dx$ et coïncide avec $\mathbb{E}(|X|)$. Si cette dernière intégrale est finie, alors l'espérance de X est donnée par la formule

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx.$$

On définit également l'espérance mathématique des variables aléatoires à valeurs dans \mathbb{R}^d . Si $X = (X_1, \dots, X_d)$ et si les variables aléatoires X_1, \dots, X_d sont intégrables, on dit que X est intégrable et on définit

$$\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_d)).$$

On a alors les propriétés suivantes :

Proposition 7 Soient X et Y deux variables aléatoires à valeurs réelles ou bien toutes deux positives ou bien toutes deux intégrables. Soit également $\lambda \in \mathbb{R}$.

1. La variable aléatoire X est intégrable si et seulement si $\mathbb{E}(|X|) < +\infty$. De plus, on a

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

2. On a $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ et $\mathbb{E}(\lambda X) = \lambda \mathbb{E}(X)$ (propriétés de linéarité). En particulier $X + Y$ est intégrable si X et Y sont intégrables.
3. Si pour tout $\omega \in \Omega$, $X(\omega) \leq Y(\omega)$, alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.
4. Si $\mathbb{P}(X = Y) = 1$ alors $\mathbb{E}(X) = \mathbb{E}(Y)$.
5. Si X prend des valeurs positives, alors

$$\mathbb{E}(X) = 0 \quad \Leftrightarrow \quad \mathbb{P}(X = 0) = 1.$$

Preuve.

- Montrons d'abord que $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ lorsque X et Y sont à valeurs positives et ne prennent qu'un nombre fini de valeurs notées $\{x_1, \dots, x_m\}$ et $\{y_1, \dots, y_p\}$ respectivement. On peut alors écrire

$$X = \sum_{i=1}^m x_i \mathbb{1}_{X=x_i} = \sum_{i=1}^m \sum_{j=1}^p x_i \mathbb{1}_{\{X=x_i, Y=y_j\}}.$$

En décomposant aussi Y de cette façon, on peut écrire

$$X + Y = \sum_{i=1}^m \sum_{j=1}^p (x_i + y_j) \mathbb{1}_{\{X=x_i, Y=y_j\}}.$$

Par définition de l'espérance appliquée aux variables discrètes X , Y et $X + Y$, on a

$$\begin{aligned} \mathbb{E}(X + Y) &= \sum_{i=1}^m \sum_{j=1}^p (x_i + y_j) \mathbb{P}(X = x_i, Y = y_j) \\ &= \sum_{i=1}^m \sum_{j=1}^p x_i \mathbb{P}(X = x_i, Y = y_j) + \sum_{i=1}^m \sum_{j=1}^p y_j \mathbb{P}(X = x_i, Y = y_j) \\ &= \mathbb{E}(X) + \mathbb{E}(Y). \end{aligned}$$

Considérons maintenant le cas de deux variables aléatoires positives X et Y quelconques. Soient alors deux suites croissantes $(X_n)_n$ et $(Y_n)_n$ de variables aléatoires de \mathcal{D}_+ et telles que pour tout $\omega \in \Omega$,

$$\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega), \quad \lim_{n \rightarrow +\infty} Y_n(\omega) = Y(\omega).$$

On voit que la suite $Z_n = X_n + Y_n$ est une suite croissante d'éléments de \mathcal{D}_+ telle que

$$\lim_{n \rightarrow +\infty} Z_n(\omega) = X(\omega) + Y(\omega).$$

Ainsi, en utilisant l'égalité $\mathbb{E}(Z_n) = \mathbb{E}(X_n) + \mathbb{E}(Y_n)$ que nous avons prouvée juste avant, on a par définition de l'intégrale

$$\mathbb{E}(X + Y) = \lim_{n \rightarrow \infty} \mathbb{E}(Z_n) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n) + \lim_{n \rightarrow \infty} \mathbb{E}(Y_n) = \mathbb{E}(X) + \mathbb{E}(Y).$$

- Montrons ensuite le point 3.. Supposons d'abord $0 \leq X \leq Y$. Soit $(X_n)_n$ (resp. $(Y_n)_n$) une suite croissante d'éléments de \mathcal{D}_+ telle que $\lim_{n \rightarrow \infty} X_n = X$ (resp. $\lim_{n \rightarrow \infty} Y_n = Y$). Alors si $p \in \mathbb{N}$, on a $X_p \leq Y$ et le Lemme 2 garantit que

$$\mathbb{E}(X_p) \leq \lim_{n \rightarrow \infty} \mathbb{E}(Y_n) = \mathbb{E}(Y).$$

En passant à la limite sur p , on obtient $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

Si maintenant, X et Y sont de signe quelconque, alors $X^+ \leq Y^+$ et $Y^- \leq X^-$. D'après ce qui a été montré juste avant, on a

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) \leq \mathbb{E}(Y^+) - \mathbb{E}(Y^-) = \mathbb{E}(Y).$$

- Montrons ensuite que $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ pour des variables aléatoires X et Y intégrables. On a $(X + Y)^+ \leq X^+ + Y^+$ et $(X + Y)^- \leq X^- + Y^-$ ce qui entraîne l'intégrabilité de $X + Y$ en utilisant l'intégrabilité de X et de Y ainsi que la linéarité de l'espérance pour les variables aléatoires positives. On a alors les décompositions

$$X + Y = (X + Y)^+ - (X + Y)^- = X^+ + Y^+ - (X^- + Y^-),$$

ce qui permet d'avoir

$$(X + Y)^+ + X^- + Y^- = (X + Y)^- + X^+ + Y^+.$$

En prenant l'espérance dans cette égalité et en utilisant la linéarité de l'espérance pour les variables aléatoires positives, on trouve

$$\mathbb{E}((X + Y)^+) + \mathbb{E}(X^-) + \mathbb{E}(Y^-) = \mathbb{E}((X + Y)^-) + \mathbb{E}(X^+) + \mathbb{E}(Y^+).$$

On obtient alors $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

- Montrons la fin du point 2.. Si $X \in \mathcal{D}_+$ et $\lambda \in \mathbb{R}_+$, alors on a évidemment $\mathbb{E}(\lambda X) = \lambda \mathbb{E}(X)$. Si $(X_n)_n$ est une suite croissante d'éléments de \mathcal{D}_+ convergente point par point vers X alors $(\lambda X_n)_n$ est une suite croissante d'éléments de \mathcal{D}_+ convergente point par point vers λX . On en déduit

$$\mathbb{E}(\lambda X) = \lim_{n \rightarrow \infty} \mathbb{E}(\lambda X_n) = \lambda \lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \lambda \mathbb{E}(X).$$

Si maintenant $\lambda \geq 0$ et X est de signe quelconque (avec X intégrable) alors

$$\lambda X = \lambda X^+ - \lambda X^-$$

et on a d'après ce qui précède

$$\mathbb{E}(X) = \mathbb{E}(\lambda X^+) - \mathbb{E}(\lambda X^-) = \lambda \mathbb{E}(X).$$

On a le même résultat lorsque $\lambda < 0$ en écrivant $\lambda X = (-\lambda)X^- - (-\lambda)X^+ (= (\lambda X)^+ - (\lambda X)^-)$ et en appliquant les résultats déjà établis.

- Prouvons maintenant le point 1. Alors l'intégrabilité de X équivaut à celle de $|X|$ car $|X| = X^+ + X^-$ est intégrable si et seulement si X^+ et X^- sont intégrables. On a alors

$$|\mathbb{E}(X)| = |\mathbb{E}(X^+) - \mathbb{E}(X^-)| \leq \mathbb{E}(X^+) + \mathbb{E}(X^-) = \mathbb{E}(|X|).$$

- Prouvons ensuite le point 5. Supposons d'abord que $\mathbb{E}(X) = 0$. Alors pour tout entier $n \geq 1$,

$$X \geq X \mathbf{1}_{X \geq \frac{1}{n}} \geq \frac{1}{n} \mathbf{1}_{X \geq \frac{1}{n}}.$$

Le point 3. assure que $0 = \mathbb{E}(X) \geq \frac{1}{n} \mathbb{P}(X \geq \frac{1}{n})$. En utilisant la continuité supérieure de la mesure, on obtient

$$\mathbb{P}(X > 0) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(X \geq \frac{1}{n}\right) = 0$$

et donc $\mathbb{P}(X = 0) = 1 - \mathbb{P}(X > 0) = 1$.

Si maintenant $\mathbb{P}(X = 0) = 1$, alors toute variable aléatoire Y de \mathcal{D}_+ plus petite que X vérifie $\mathbb{P}(Y = 0) = 1$. Il suffit alors de prouver que $\mathbb{E}(Y) = 0$ (ce qui entraînera $\mathbb{E}(X) = 0$ en utilisant la définition de l'espérance). Si $Y = \sum_{i=1}^p y_i \mathbf{1}_{Y=y_i}$ alors on a $y_i = 0$

si $\mathbb{P}(Y = y_i) > 0$. On a donc bien $\mathbb{E}(Y) = \sum_{i=1}^p y_i \mathbb{P}(Y = y_i) = 0$.

- Prouvons enfin le point 4. Si $Y = 0$ et $X \in \mathcal{D}_+$ alors le résultat est une conséquence du point 5. Si maintenant $Y = 0$ et X est intégrable de signe quelconque, alors $\mathbb{P}(X^+ = 0) = \mathbb{P}(X^- = 0) = 1$. On a donc $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) = 0$. Enfin si X et Y sont deux variables aléatoires intégrables, on a en notant $A = \{X \neq Y\}$ l'égalité $X - Y = (X - Y)\mathbb{1}_A = Z$ avec $\mathbb{P}(Z = 0) = 1$. D'après ce qui précède, on a $\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = 0$ ce qui entraîne que $\mathbb{E}(X) = \mathbb{E}(Y)$. \square

Remarques.

- Soient m et M deux nombres réels. La proposition précédente montre en particulier que si X est une variable aléatoire intégrable telle que $X \leq M$ (resp. $X \geq m$) alors $\mathbb{E}(X) \leq M$ (resp. $\mathbb{E}(X) \geq m$).
- Commentons le point 4. Dire que $\mathbb{P}(X = Y) = 1$ ne signifie pas que $X(\omega) = Y(\omega)$ pour tout $\omega \in \Omega$. Par exemple, considérons $\Omega = [0, 1]$ muni de la probabilité uniforme. Posons $X(\omega) = \omega$ et $Y(\omega) = \omega$ si $\omega \neq 0.5$, $Y(0.5) = 0$. Dans ce cas $\mathbb{P}(X = Y) = \mathbb{P}([0, 1] \setminus \{0.5\}) = 1$ et pourtant $X(0.5) \neq Y(0.5)$. Souvent on dit que $X = Y$ presque sûrement (en abrégé p.s). Ainsi la valeur de l'espérance est inchangée si on remplace une variable aléatoire X par une variable Y qui lui est presque sûrement égale.

La variance d'une variable aléatoire X à valeurs réelles réelle mesure la dispersion de X autour de sa moyenne. Nous rappelons sa définition.

Définition 12 Lorsque une variable aléatoire X vérifie $\mathbb{E}(X^2) < +\infty$ alors la variance de X est le réel positif noté $\text{Var}(X)$ et défini par

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

On peut vérifier la deuxième égalité de la définition précédente en développant le carré et en utilisant les propriétés de linéarité de l'intégrale. Noter que $\mathbb{E}(X^2) < \infty$ entraîne que X est intégrable : en effet, on a l'inégalité $|X| \leq 1 + X^2$ et donc

$$\mathbb{E}(|X|) \leq 1 + \mathbb{E}(X^2) < \infty.$$

Mentionnons enfin que $\text{Var}(X) = 0$ entraîne $(X - \mathbb{E}(X))^2 = 0$ p.s en utilisant le point 5. de la proposition précédente. On en déduit facilement que $X = \mathbb{E}(X)$ presque sûrement.

Chapitre 2

Indépendance. Variables aléatoires discrètes

Dans tout ce chapitre, on suppose donné un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

2.1 Indépendance entre variables aléatoires

Définition 13 Des variables aléatoires X_1, X_2, \dots, X_n à valeurs dans $\mathbb{R}^{d_1}, \mathbb{R}^{d_2}, \dots, \mathbb{R}^{d_n}$ respectivement sont dites indépendantes si

$$\mathbb{P}(\cap_{i=1}^n \{X_i \in A_i\}) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i),$$

pour tout $(A_1, \dots, A_n) \in \mathcal{B}(\mathbb{R}^{d_1}) \times \dots \times \mathcal{B}(\mathbb{R}^{d_n})$.

Nous utiliserons souvent la notation

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(\cap_{i=1}^n \{X_i \in A_i\})$$

dans la suite. On définit également l'indépendance d'une famille quelconque de variables aléatoires (de cardinal infini) : une telle famille est dite indépendante si toute sous-famille finie est indépendante au sens de la définition ci-dessus.

Remarques

1. Rappelons que de façon générale, nous qualifions une variable aléatoire $X : \Omega \rightarrow \mathbb{R}^d$ de discrète lorsque $\mathbb{P}(X \in E) = 1$ où E est un sous-ensemble fini ou infini dénombrable de \mathbb{R}^d . Pour une variable aléatoire X discrète, rappelons la notation

$$\text{val}(X) = \left\{ x \in \mathbb{R}^d / \mathbb{P}(X = x) \neq 0 \right\},$$

qui est un sous-ensemble fini ou infini dénombrable de \mathbb{R}^d . Considérons alors n variables aléatoires discrètes X_1, \dots, X_n . Dans ce cas, on peut montrer que les variables aléatoires X_1, \dots, X_n sont indépendantes si et seulement si

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i), \quad \forall (x_1, \dots, x_n) \in \text{val}(X_1) \times \dots \times \text{val}(X_n).$$

2. Si n variables aléatoires sont indépendantes alors ces variables aléatoires sont indépendantes deux à deux. Pour le voir il suffit de poser $A_i = \mathbb{R}^{d_i}$ pour tous les indices i correspondants aux $n-2$ variables aléatoires restantes. En revanche, il faut bien garder à l'esprit que n variables aléatoires indépendantes deux à deux ne sont pas indépendantes au sens de la Définition 13. On pourra vérifier en exercice que si X_1 et X_2 sont deux variables aléatoires discrètes indépendantes et de même loi donnée par

$$\mathbb{P}(X_1 = 1) = 1 - \mathbb{P}(X_1 = -1) = \mathbb{P}(X_2 = 1) = 1 - \mathbb{P}(X_2 = -1) = \frac{1}{2}$$

et $X_3 = X_1 X_2$ alors les variables aléatoires X_1, X_2, X_3 sont indépendantes deux à deux mais pas au sens de la Définition 13.

Exemple. Posons $\Omega = \{1, 2, \dots, N\}^2$ muni de la probabilité uniforme, c'est à dire $\mathbb{P}(\{\omega\}) = \frac{1}{N^2}$. Alors les applications coordonnées $X_1 : \omega \mapsto \omega_1$ et $X_2 : \omega \mapsto \omega_2$ sont des variables aléatoires indépendantes et de même loi uniforme sur l'ensemble $\Omega_0 = \{1, 2, \dots, N\}$.

Existence d'un n -uplet de variables aléatoires indépendantes et de lois données. On peut justifier leur existence en utilisant la notion de mesure produit. Pour $1 \leq i \leq n$, soit $(\Omega_i, \mathcal{A}_i, Q_i)$ un espace probabilisable. Notons $\Omega = \Omega_1 \times \dots \times \Omega_d$ et munissons Ω de la tribu produit \mathcal{A} (voir Chapitre 1). Alors nous admettrons le résultat suivant

Proposition 8 *Sur l'espace probabilisable (Ω, \mathcal{A}) , il existe une unique mesure de probabilité \mathbb{P} telle que*

$$\mathbb{P}(A_1 \times A_2 \times \dots \times A_n) = Q_1(A_1) \times Q_2(A_2) \times \dots \times Q_n(A_n),$$

pour $A_i \in \mathcal{A}_i, 1 \leq i \leq n$. On dit que \mathbb{P} est une mesure produit.

Donnons un exemple. Il s'agit de la mesure de probabilité uniforme sur $\Omega = I_1 \times I_2$ où I_1 et I_2 sont deux intervalles fermés bornés de \mathbb{R} . Si pour $i = 1, 2$, $\mathcal{A}_i = \mathcal{B}(I_i)$ et Q_i est la mesure de probabilité uniforme sur (I_i, \mathcal{A}_i) , alors la mesure produit \mathbb{P} est appelée mesure de probabilité uniforme sur Ω . On peut alors montrer que si $A \in \mathcal{A}$ et $x \in \mathbb{R}^d$ sont tels que $x + A \in \mathcal{A}$, alors $\mathbb{P}(x + A) = \mathbb{P}(A)$ (invariance par translation, ce qui justifie le caractère uniforme). $\mathbb{P}(A)$ s'interprète comme le quotient entre l'aire de A et l'aire de Ω .

Voyons maintenant pourquoi ce résultat permet de construire n variables aléatoires X_1, \dots, X_n indépendantes et de lois respectives Q_1, \dots, Q_n . Si pour $1 \leq i \leq n$, Q_i est une mesure de probabilité sur \mathbb{R}^{d_i} , posons

$$\Omega = \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_n} \cong \mathbb{R}^{d_1 + \dots + d_n}.$$

Soit \mathcal{A} la tribu produit des tribus boréliennes $\mathcal{B}(\mathbb{R}^{d_1}), \dots, \mathcal{B}(\mathbb{R}^{d_1})$ (tribu qui coïncide en fait avec la tribu borélienne sur Ω). Soit alors \mathbb{P} est la mesure produit correspondante et posons $X_i(\omega) = \omega_i \in \mathbb{R}^{d_i}$ pour $(i, \omega) \in \{1, \dots, d\} \times \Omega$. Alors les variables aléatoires X_1, \dots, X_n sont indépendantes sous la probabilité \mathbb{P} et sont bien de lois respectives Q_1, \dots, Q_n . En effet si pour $1 \leq i \leq n$, $A_i \in \mathcal{B}(\mathbb{R}^{d_i})$ alors

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(A_1 \times \dots \times A_n) = \prod_{i=1}^n Q_i(A_i) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i).$$

Il existe un résultat similaire (mais un peu plus compliqué) qui garantit l'existence de suites de variables aléatoires indépendantes et de lois données (plus précisément il existe un espace probabilisé sur lequel sont définies une infinité de variables aléatoires indépendantes et de lois données). Nous admettrons l'existence de ce type de construction dans ce cours.

Notation. Pour une suite de variables aléatoires indépendantes et telle que toutes les variables aléatoires aient la même loi, on dira que la suite est **indépendante et identiquement distribuée**, ce qui sera noté en abrégé **i.i.d.**

Étudions maintenant les propriétés des variables aléatoires indépendantes.

Proposition 9 *Pour $1 \leq i \leq n$, soient X_i une variable aléatoire à valeurs dans \mathbb{R}^{d_i} et $f_i : \mathbb{R}^{d_i} \rightarrow \mathbb{R}^{e_i}$ une fonction qui définit également une variable aléatoire lorsque \mathbb{R}^{d_i} est muni de sa tribu borélienne. Supposons les variables aléatoires X_1, \dots, X_n indépendantes et posons $Y_i = f_i(X_i)$ pour $1 \leq i \leq n$. Alors les variables aléatoires Y_1, \dots, Y_n sont également indépendantes.*

Preuve. Il suffit de remarquer l'égalité

$$\{Y_i \in B_i\} = \{X_i \in f_i^{-1}(B_i)\},$$

pour $1 \leq i \leq n$ et pour tout borélien B_i . On applique ensuite la définition 13 aux boréliens $A_i = f_i^{-1}(B_i)$, $1 \leq i \leq n$. \square

Une conséquence fondamentale de l'indépendance entre variables aléatoires concerne le calcul de l'espérance du produit de variables aléatoires réelles indépendantes.

Proposition 10 *Soient X_1, \dots, X_n des variables aléatoires indépendantes, toutes à valeurs réelles. Alors le produit $X_1 \cdots X_n$ est intégrable si et seulement si les variables aléatoires X_1, \dots, X_n sont intégrables. Dans ce cas, on a la formule*

$$\mathbb{E}(X_1 X_2 \cdots X_n) = \prod_{i=1}^n \mathbb{E}(X_i).$$

Preuve. On peut le montrer pour deux variables aléatoires X et Y , le cas général s'en déduisant facilement par récurrence finie. Commençons par prouver ce résultat lorsque X et Y sont positives (dans ce cas, les espérances ont toujours un sens). Posons pour $n \in \mathbb{N}$,

$$X_n = \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} \mathbb{1}_{\{\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\}}, \quad Y_n = \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} \mathbb{1}_{\{\frac{k}{2^n} \leq Y < \frac{k+1}{2^n}\}}.$$

On a vu au chapitre précédent que la suite $(X_n)_{n \in \mathbb{N}}$ est une suite croissante de variables aléatoires positives ne prenant qu'un nombre fini de valeurs et convergente point par point vers X . Ainsi la suite $(Z_n)_{n \in \mathbb{N}}$ de variables aléatoires définie par $Z_n = X_n Y_n$ pour tout $n \in \mathbb{N}$ est aussi une suite croissante de variables aléatoires positives ne prenant qu'un nombre fini de valeurs et convergente point par point vers $Z = XY$. D'autre part, il est facile de vérifier que $\mathbb{E}(X_n Y_n) = \mathbb{E}(X_n) \mathbb{E}(Y_n)$ pour tout $n \in \mathbb{N}$, en utilisant l'indépendance de X et de Y . Ainsi la définition de l'espérance d'une variable aléatoire positive assure que

$$\mathbb{E}(Z) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n Y_n) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n) \mathbb{E}(Y_n) = \mathbb{E}(X) \mathbb{E}(Y).$$

Nous avons donc montré la proposition pour deux variables aléatoires indépendantes positives. Dans le cas général, si X et Y sont indépendantes alors $|X|$ et $|Y|$ le sont également et on a d'après ce qui précède

$$\mathbb{E}(|XY|) = \mathbb{E}(|X|) \mathbb{E}(|Y|).$$

Ceci prouve que la variable aléatoire XY est intégrable si et seulement si les variables aléatoires X et Y le sont. La formule pour le calcul de l'espérance de XY se déduit alors facilement, en utilisant les décompositions

$$X = X^+ - X^-, \quad Y = Y^+ - Y^-.$$

En effet la partie positive ou négative de X est indépendante de la partie positive ou négative de Y , d'après la Proposition 9. Ainsi en utilisant le résultat de la proposition pour les variables aléatoires positives, on obtient

$$\begin{aligned} \mathbb{E}(XY) &= \mathbb{E}(X^+Y^+) + \mathbb{E}(X^-Y^-) - \mathbb{E}(X^+Y^-) - \mathbb{E}(X^-Y^+) \\ &= \mathbb{E}(X^+) \mathbb{E}(Y^+) + \mathbb{E}(X^-) \mathbb{E}(Y^-) - \mathbb{E}(X^+) \mathbb{E}(Y^-) - \mathbb{E}(X^-) \mathbb{E}(Y^+) \\ &= (\mathbb{E}(X^+) - \mathbb{E}(X^-)) (\mathbb{E}(Y^+) - \mathbb{E}(Y^-)) \\ &= \mathbb{E}(X)\mathbb{E}(Y). \square \end{aligned}$$

On peut déduire du résultat précédent une formule bien connue pour le calcul de la variance de la somme de n variables aléatoires indépendantes réelles.

Corollaire 1 *Soient X_1, \dots, X_n des variables aléatoires indépendantes, toutes à valeurs réelles et de carré intégrable. Alors*

$$\text{Var} (X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var} (X_i).$$

Preuve. En posant $Y_i = X_i - \mathbb{E}(X_i)$ pour $1 \leq i \leq n$, on a

$$\begin{aligned} \text{Var} (X_1 + \dots + X_n) &= \mathbb{E}((Y_1 + \dots + Y_n)^2) \\ &= \sum_{i=1}^n \mathbb{E}(Y_i^2) + \sum_{i \neq j} \mathbb{E}(Y_i Y_j) \\ &= \sum_{i=1}^n \mathbb{E}(Y_i^2) + \sum_{i \neq j} \mathbb{E}(Y_i) \mathbb{E}(Y_j) \\ &= \sum_{i=1}^n \mathbb{E}(Y_i^2) \\ &= \sum_{i=1}^n \text{Var} (X_i). \end{aligned}$$

Ces égalités proviennent du fait que pour $1 \leq i \leq n$, la variables aléatoire Y_i est centré, de carré intégrable et indépendante Y_j lorsque $j \neq i$. \square

2.2 Les variables aléatoires discrètes

2.2.1 Les lois discrètes usuelles en modélisation

La loi de Bernoulli de paramètre p ($0 < p < 1$). On dit qu'une variable aléatoire X suit une loi de Bernoulli de paramètre p si

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p.$$

On utilise la notation $X \sim \mathcal{B}(p)$ pour indiquer que X suit cette loi. On peut remarquer que $p = \mathbb{E}(X)$ et $\text{Var}(X) = p(1-p)$. Cette loi intervient donc systématiquement lorsqu'il y a deux éventualités dans l'expérience considérée (jeu de pile ou face, individu sain ou malade,...). Remarquons également que pour tout événement $A \in \mathcal{A}$, la variable aléatoire $\mathbb{1}_A$ suit une loi de Bernoulli de paramètre $\mathbb{P}(A)$.

Loi binomiale. On dit que X suit une loi binomiale de paramètres n et p ($n \in \mathbb{N}^*$ et $p \in]0, 1[$) lorsque

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

où $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. La loi binomiale est aussi la loi d'une somme $X_1 + X_2 + \dots + X_n$ de n variables aléatoires indépendantes, toutes de loi de Bernoulli de paramètre p . Cette loi sert à modéliser le nombre de succès lors de la répétition de n expériences successives et identiques (e.g n lancers successifs d'une pièce de monnaie, on peut aussi y avoir le nombre de réponses à un sondage donné...). L'espérance de cette loi vaut de $\mathbb{E}(X_1 + \dots + X_n) = np$ et d'après le Corollaire 1, la variance de cette loi est $\text{Var}(X_1 + \dots + X_n) = np(1-p)$.

Loi de Poisson de paramètre $\lambda > 0$. On note $X \sim \mathcal{P}(\lambda)$ pour indiquer que la loi d'une variable aléatoire X est définie par

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

La loi de Poisson approxime bien les expériences impliquant des problèmes de comptage (e.g nombre de connexions à un serveur entre deux instants), en particulier pour des événements rares (e.g nombre de suicides par an). Si par exemple on compte le nombre de pièces défectueuses produites par une machine avec un nombre n de pièces produites importantes et une proportion p de pièces défectueuses petite alors la loi de Poisson approxime bien la loi binomiale. En effet, on peut montrer que lorsque $n \rightarrow +\infty$, $p \rightarrow 0$ et $np \rightarrow \lambda$, le nombre $\binom{n}{k} p^k (1-p)^{n-k}$ s'approche de $\exp(-\lambda) \frac{\lambda^k}{k!}$ pour $k \in \mathbb{N}$ fixé.

On a $\mathbb{E}(X) = \text{Var}(X) = \lambda$ si $X \sim \mathcal{P}(\lambda)$.

La loi uniforme. On dit qu'une variable aléatoire X suit une loi uniforme si $X(\Omega) = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ et

$$\mathbb{P}(X = x_k) = \frac{1}{n}, \quad k = 1, \dots, n.$$

On peut par exemple montrer que si on essaie d'ouvrir une porte à l'aide de n clefs en choisissant d'abord une clef au hasard puis une clef au hasard parmi les $n-1$ restantes (si la première n'ouvre pas la porte) et ainsi de suite jusqu'à ce que la porte soit ouverte, alors la loi du nombre de clefs utilisées suit la loi uniforme sur $\{1, \dots, n\}$.

L'espérance d'une variable aléatoire X de loi uniforme sur $\{1, \dots, n\}$ est $\mathbb{E}(X) = \frac{n+1}{2}$ et sa variance est $\text{Var}(X) = \frac{n^2-1}{12}$.

Loi géométrique de paramètre $p \in]0, 1[$. Il s'agit de la loi du temps de premier succès dans une suite d'expériences aléatoires indépendantes où la probabilité de succès est p . Plus précisément, considérons une suite $(X_i)_{i \geq 1}$ de variables aléatoires i.i.d suivant toutes la loi $\mathcal{B}(p)$. Alors la variable aléatoire

$$T = \inf \{i \geq 1 / X_i = 1\}$$

a sa loi définie par

$$\mathbb{P}(T = k) = \mathbb{P}(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) = (1-p)^{k-1}p, \quad k \in \mathbb{N}^*.$$

L'espérance vaut $\mathbb{E}(T) = \frac{1}{p}$ et la variance vaut $\text{Var}(X) = \frac{1-p}{p^2}$.

La loi hypergéométrique. Considérons une population composée de N individus parmi lesquels une proportion p possède un caractère donné. Si on tire au hasard $n \leq N$ individus dans la population, alors le nombre aléatoire X d'individus possédant le caractère envisagé vérifie

$$\mathbb{P}(X = k) = \frac{\binom{Np}{k} \binom{N-Np}{n-k}}{\binom{N}{n}},$$

pour $\max(0, n - N + Np) \leq k \leq \min(n, Np)$. Pour décrire rigoureusement cette expérience, on peut considérer l'univers Ω de tous les n -uplets contenant uniquement des 0 et des 1 avec la contrainte que le nombre de 0 soit plus petit que $N - Np$ et le nombre de 1 soit plus petit que Np . On munit alors Ω de l'ensemble de ses parties et de la probabilité uniforme et on pose $X(\omega) = \sum_{i=1}^n \mathbb{1}_{\omega_i=1}$ si $\omega \in \Omega$.

On peut montrer que $\mathbb{E}(X) = np$ et $\text{Var}(X) = \frac{N-n}{N-1}np(1-p)$ si $X \sim \mathcal{H}(N, n, p)$ (notation pour la loi hypergéométrique). Lorsque N devient grand, on peut aussi montrer que la loi de X est proche de la loi binomiale de paramètres n et p .

La loi binomiale négative. On dit qu'une variable aléatoire X suit une loi binomiale négative de paramètres $n \in \mathbb{N}^*$ et $p \in]0, 1[$ si

$$\mathbb{P}(X = k) \binom{k-1}{k-n} p^n (1-p)^{k-n}, \quad k = n, n+1, \dots$$

On remarquera que lorsque $n = 1$, on retrouve la loi géométrique de paramètre p . On peut aussi représenter cette loi à l'aide d'une suite $(X_i)_{i \geq 1}$ i.i.d de variables aléatoires toutes de loi de Bernoulli de paramètre p . On définit alors $T_0 = 0$ et les temps T_1, \dots, T_n par la relation

$$T_{i+1} = \inf \{j > T_i / X_j = 1\}, \quad i = 0, \dots, n-1.$$

On peut alors montrer que les variables aléatoires $T_1, T_2 - T_1, \dots, T_n - T_{n-1}$ sont indépendantes et toutes de loi géométrique de paramètre p et que leur somme T_n a la même loi que X (il s'agit donc de la loi du n -ième temps de succès). On en déduit facilement que $\mathbb{E}(X) = \sum_{i=0}^{n-1} \mathbb{E}(T_{i+1} - T_i) = \frac{n}{p}$ et

$$\text{Var}(X) = \sum_{i=0}^{n-1} \text{Var}(T_{i+1} - T_i) = \frac{n(1-p)}{p^2}.$$

2.3 Loi d'un couple. Loi marginale

- Si X et Y sont deux variables aléatoires discrètes à valeurs dans $E = \mathbb{R}^d$ et $F = \mathbb{R}^e$ respectivement, alors la variable aléatoire $U = (X, Y)$ est aussi une variable aléatoire discrète, à valeurs dans $E \times F$. En général, la loi de U n'est pas uniquement déterminée par les lois des variables aléatoires X et Y . Prenons un exemple. Si $X \sim \mathcal{B}(p)$, posons

$Y_1 = 1 - X$ et $Y_2 = X$. Alors Y_1 et Y_2 suivent des lois de Bernoulli de paramètre p . En revanche la loi du couple (X, Y_1) ne coïncide pas avec celle du couple (X, Y_2) car

$$\mathbb{P}(X = 0, Y_1 = 1) = \mathbb{P}(X = 0) = \frac{1}{2} \text{ et } \mathbb{P}(X = 0, Y_2 = 1) = 0.$$

Par contre les lois de X et de Y déterminent la loi du couple (X, Y) lorsque les deux variables sont indépendantes grâce aux égalités

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \quad (x, y) \in E \times F.$$

- Si $U = (X, Y)$ est une variable aléatoire discrète à valeurs dans $E \times F$ alors les lois des variables aléatoires X et Y sont appelées les lois marginales de U . On peut les calculer à partir de la loi du couple en utilisant les égalités :

$$\mathbb{P}(X = x) = \sum_{y \in \text{val}(Y)} \mathbb{P}(X = x, Y = y), \quad \mathbb{P}(Y = y) = \sum_{x \in \text{val}(X)} \mathbb{P}(X = x, Y = y).$$

2.4 Le théorème de transfert

Nous allons d'abord énoncer un résultat pratique lorsque on somme des termes indexés à l'aide de deux paramètres.

Théorème 2 (*Théorème de Fubini pour les suites*)

- Si $\{u_{i,j}/i, j \geq 1\}$ est une famille de nombres réels positifs, alors on a

$$\sum_{i=1}^{+\infty} \left(\sum_{j=1}^{+\infty} u_{i,j} \right) = \sum_{j=1}^{+\infty} \left(\sum_{i=1}^{+\infty} u_{i,j} \right). \quad (2.1)$$

Il se peut que ces deux sommes aient pour valeur $+\infty$.

- Si $\{u_{i,j}/i, j \geq 1\}$ est une famille de nombres réels de signe quelconque et si l'une des quantités $\sum_{i=1}^{+\infty} \left(\sum_{j=1}^{+\infty} |u_{i,j}| \right)$ ou $\sum_{j=1}^{+\infty} \left(\sum_{i=1}^{+\infty} |u_{i,j}| \right)$ est finie (ces deux quantités sont toujours égales d'après le premier point), alors l'égalité 2.1 est encore vraie.

Ce théorème est de portée assez générale. Toutefois, il peut arriver que l'inversion des signes somme ne soit pas possible : le cas de la suite u définie par $u_{i,j} = \frac{1}{i^2-j^2}$ si $i \neq j$ et $u_{i,i} = 0$ pour tous i et j entiers plus grands que 1 fournit un contre-exemple.

Remarque. L'espérance mathématique d'une variable aléatoire discrète X intégrable est donnée par la formule :

$$\mathbb{E}(X) = \sum_{x \in \text{val}(X)} x\mathbb{P}(X = x).$$

Lorsque $\text{val}(X)$ est infini dénombrable mais n'est pas l'ensemble des entiers naturels \mathbb{N} (par exemple \mathbb{Z}), on peut toujours lister les éléments de $\text{val}(X)$ à l'aide d'une suite quelconque, la valeur de la somme précédente ne dépendra pas de la suite choisie. En effet les séries positives ou absolument convergente peuvent être sommées indépendamment de l'énumération choisie (par exemple $0, 1, -1, 2, -2, 3, -3, \dots$ et $0, 1, 2, -1, -2, 3, 4, -3, -4, \dots$ sont deux énumérations différentes de \mathbb{Z}).

Le théorème de transfert (ici énoncé pour les variables discrètes) permet de calculer l'espérance d'une variable aléatoire du type $Y = f(X)$ en utilisant la loi de X uniquement (ce qui évite de calculer la loi de Y).

Théorème 3 Soient X une variable aléatoire discrète telle que $\text{val}(X) = \{x_1, x_2, \dots\}$ et $f : \text{val}(X) \rightarrow \mathbb{R}$ une fonction telle que $\mathbb{E}(|f(X)|) < +\infty$. Alors on a la formule :

$$\mathbb{E}(f(X)) = \sum_{i=1}^{+\infty} f(x_i) \mathbb{P}(X = x_i).$$

Preuve. Par hypothèse, la variable aléatoire $Y = f(X)$ est intégrable et vérifie $F = \text{val}(Y) = \{y_1, y_2, \dots\}$ de \mathbb{R} (F est donc composé des réels $f(x_j)$, $j \in \mathbb{N}^*$). Supposons d'abord f positive. D'après la formule de l'espérance pour une variable aléatoire discrète, on a :

$$\mathbb{E}(Y) = \sum_{j \geq 1} y_j \mathbb{P}(Y = y_j).$$

Mais on a l'égalité entre événements : $\{Y = y_j\} = \{X \in f^{-1}(\{y_j\})\}$ pour $j \geq 1$. Comme on a la formule

$$\mathbb{P}(X \in f^{-1}(\{y_j\})) = \sum_{i=1}^{+\infty} \mathbb{1}_{f^{-1}(\{y_j\})}(x_i) \mathbb{P}(X = x_i),$$

on obtient en reportant dans l'expression de $\mathbb{E}(Y)$:

$$\mathbb{E}(Y) = \sum_{j \geq 1} y_j \sum_{i=1}^{+\infty} \mathbb{1}_{f^{-1}(\{y_j\})}(x_i) \mathbb{P}(X = x_i).$$

On peut utiliser le théorème de Fubini pour inverser les sommes (les termes sont positifs), ce qui donne

$$\mathbb{E}(Y) = \sum_{i=1}^{+\infty} \mathbb{P}(X = x_i) \sum_{j \geq 1} y_j \mathbb{1}_{f^{-1}(\{y_j\})}(x_i).$$

Enfin, on remarque que $\sum_{j \geq 1} y_j \mathbb{1}_{f^{-1}(\{y_j\})}(x_i) = f(x_i)$, ce qui prouve le théorème lorsque f est positive.

Lorsque f est de signe quelconque, on reproduit exactement le calcul précédent jusqu'au passage de l'inversion des deux signes sommes. Cette inversion se justifie à l'aide du théorème de Fubini pour les termes de signe quelconques : en effet, si on considère la fonction $g = |f|$, la première partie de la preuve appliqué à $Y = g(X)$ montre que l'hypothèse de sommabilité

$$\sum_{i=1}^{+\infty} \mathbb{P}(X = x_i) \sum_{j \geq 1} y_j \mathbb{1}_{g^{-1}(\{y_j\})}(x_i) = \mathbb{E}(g(X)) = \mathbb{E}(|f(X)|) < +\infty$$

est bien vérifiée pour pouvoir appliquer le théorème de Fubini. \square

2.5 Les lois conditionnelles

Définition 14 Soit X et Y un couple de variables aléatoires discrètes à valeurs dans \mathbb{R}^d et \mathbb{R}^e respectivement. Si $y \in \text{val}(Y)$, on définit une loi de probabilité discrète Q_y sur les parties de $\text{val}(X)$ telle que

$$Q_y(\{x\}) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}, \quad x \in \text{val}(X).$$

Q_y est appelé loi conditionnelle de X sachant $Y = y$ (en abrégé loi de $X|Y = y$). On note $\mathbb{P}(X = x|Y = y)$ au lieu de $Q_y(\{x\})$.

$\mathbb{P}(X = x|Y = y)$ correspond à la probabilité conditionnelle $\mathbb{P}(A|B)$ des événements $A = \{X = x\}$ et $B = \{Y = y\}$:

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

Cette définition de la loi conditionnelle est en accord avec l'intuition de l'approche fréquentiste. Si on répétait plusieurs fois une expérience aléatoire ayant un nombre fini d'issues (u, v) et qu'on veuille étudier l'influence d'une valeur $v = y$ sur une valeur $u = x$, il est naturel de calculer la fréquence d'apparition de la valeur x lorsque y est également observée. Ceci revient alors à calculer le quotient $\frac{f_2}{f_1}$ où f_1 est la fréquence d'apparition de y et f_2 la fréquence d'apparition de (x, y) . D'où la définition de la probabilité conditionnelle à l'aide du quotient des probabilités correspondantes. On pourra remarquer que lorsque les variables aléatoires sont indépendantes alors la loi conditionnelle de $X|Y = y$ coïncide avec la loi de X .

Notation. La moyenne de la loi conditionnelle $X|Y = y$ sera notée $\mathbb{E}(X|Y = y)$. On a donc

$$\mathbb{E}(X|Y = y) = \sum_{x \in \text{val}(X)} x \mathbb{P}(X = x|Y = y),$$

qui sera bien définie lorsque

$$\sum_{x \in \text{val}(X)} |x| \mathbb{P}(X = x|Y = y) < +\infty.$$

Proposition 11 Si (X, Y) est un couple de variables aléatoires discrètes tel que X soit intégrable, alors on a la formule

$$\mathbb{E}(X) = \sum_{y \in \text{val}(Y)} \mathbb{E}(X|Y = y) \mathbb{P}(Y = y).$$

Preuve. A faire en exercice en utilisant la définition précédente et le théorème de Fubini. \square

Exemple de calcul. Entre deux instants donnés, des véhicules arrivent à un carrefour en nombre poissonien Y . Chaque véhicule prend alors sur sa droite ou sur sa gauche avec probabilité $\frac{1}{2}$, et chacun le fait indépendamment des autres. Soit X le nombre de véhicules qui on choisit de continuer sur leur gauche. On souhaite calculer $\mathbb{E}(X)$ (nombre moyen de véhicules ayant pris sur la gauche). Dans cet exemple, on considérera que

– La loi de X est le loi de Poisson de paramètre $\lambda > 0$.

– Le couple (X, Y) est tel que la loi de $X|Y = y$ est une loi binomiale de paramètres y et $\frac{1}{2}$.

On a alors $\mathbb{E}(X|Y = y) = \frac{y}{2}$ et

$$\mathbb{E}(X) = \sum_{y \in \mathbb{N}} \frac{y}{2} \exp(-\lambda) \frac{\lambda^y}{y!} = \frac{1}{2} \mathbb{E}(Y) = \frac{\lambda}{2}.$$

Chapitre 3

Mesures et intégration

Lorsque nous avons abordé les mesures de probabilité à densité, nous avons mentionné que pour une fonction $f : \mathbb{R} \rightarrow \mathbb{R}_+$ intégrable et d'intégrale 1, il existait une unique mesure de probabilité \mathbb{P} sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour tout couple de nombres réels (a, b) tel que $a \leq b$

$$\mathbb{P}([a, b]) = \int_a^b f(x)dx.$$

Il est alors naturel de se demander si cette mesure de probabilité peut être définie par l'égalité $\mathbb{P}(B) = \int_B f(x)dx$ pour tout borélien B . Malheureusement, l'intégrale classique (intégrale de Riemann) ne permet de donner un sens à ce type d'égalité pour tous les boréliens B . Il existe une autre manière de définir l'intégrale d'une fonction et qui permet d'écrire ce type d'égalité. Il s'agit de l'intégrale au sens de Lebesgue qui permet d'intégrer des fonctions beaucoup plus irrégulières que la méthode de Riemann. De plus lorsque on intégrera une fonction continue ou continue par morceaux sur un intervalle fermé borné, les deux méthodes d'intégration coïncideront ; l'intégrale de Lebesgue apparaîtra donc comme plus générale. Initialement, l'intégrale de Lebesgue a été introduite afin de faciliter certains des passages à la limite du type

$$\lim_{n \rightarrow \infty} \int f_n(x)dx = \int \lim_{n \rightarrow \infty} f_n(x)dx,$$

en particulier lorsque la suite de fonctions $(f_n)_n$ est une suite croissante de fonctions positives. Un problème dans l'utilisation de l'intégrale de Riemann réside dans le fait qu'une limite simple de fonctions intégrables n'est pas forcément intégrable (et ce même si toutes les fonctions sont définies sur $[0, 1]$ et à valeurs dans $[0, 1]$ par exemple). L'intégrale au sens de Lebesgue permet de corriger ce type de problème. Sa construction est basée sur la théorie de la mesure et la définition de l'espérance des variables aléatoires vue au Chapitre 1 est en fait un cas particulier de cette construction.

3.1 Mesure sur une tribu

Définition 15 Soit E un ensemble et \mathcal{E} une tribu sur E . Une application $\mu : \mathcal{E} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ est appelée une mesure si

1. $\mu(\emptyset) = 0$
2. Pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{E} , disjoints deux à deux, on a

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

(propriété de σ -additivité).

Toute mesure de probabilité est donc une mesure (la propriété d'additivité peut être vue comme un cas particulier de la propriété de σ -additivité en complétant une suite finie d'éléments d'une tribu par l'élément \emptyset). On peut remarquer que la propriété $\mathbb{P}(\Omega) = 1$ a été remplacé par la propriété $\mu(\emptyset) = 0$. Pour la notion générale de mesure, la valeur $\mu(E)$ peut être positive quelconque et éventuellement infinie. Une mesure de probabilité est donc simplement une mesure dont la masse totale $\mu(E)$ est égale à 1. On peut montrer que les propriétés de la Proposition 3 du Chapitre 1 restent vraies pour une mesure en général, sauf pour la première qui devient $\mu(A) + \mu(A^c) = \mu(E)$. De plus la Proposition 4 reste valable en rajoutant $\mu(A_0) < +\infty$ pour le deuxième point.

Exemple des mesures discrètes. Soit $(p_n)_{n \in \mathbb{N}}$ une suite de nombres réels positifs et $(x_n)_{n \in \mathbb{N}}$ une suite de points de \mathbb{R}^d . Alors l'application $\mu : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ définie par

$$\mu(A) = \sum_{n=0}^{+\infty} p_n \mathbf{1}_A(x_n), \quad A \in \mathcal{B}(\mathbb{R}^d),$$

est une mesure (la preuve est identique à celle donnée pour les mesures de probabilité discrète du Chapitre 1). Comme pour $x \in \mathbb{R}^d$ et $A \in \mathcal{P}(\mathbb{R}^d)$, on a $\mathbf{1}_A(x) = \mu_x(A)$, on note $\mu = \sum_{n=0}^{+\infty} p_n \delta_{x_n}$. Dans le cas particulier où $p_n = 1$ pour tout $n \in \mathbb{N}$, on parle de mesure de comptage sur l'ensemble $D = \{x_0, x_1, \dots\}$ car dans ce cas

$$\mu(A) = |A \cap D|, \quad A \in \mathcal{P}(\mathbb{R}^d).$$

La mesure de A est simplement le nombre d'éléments de D qui se trouvent aussi dans A . Un cas particulier important est celui de la mesure de comptage sur \mathbb{N} .

Lorsque la suite $(p_n)_{n \in \mathbb{N}}$ est sommable et de somme 1, on retrouve les mesures de probabilités discrètes.

La mesure de Lebesgue

Théorème 4 *Sur \mathbb{R} muni de la tribu des boréliens, il existe une unique mesure notée λ telle que*

$$\lambda([a, b]) = b - a, \quad a < b.$$

Cette mesure est appelée mesure de Lebesgue sur \mathbb{R} .

La mesure de Lebesgue correspond à une mesure de longueur et $\lambda(B)$ est souvent appelé la longueur du borélien B . Comme on le verra, on peut construire n'importe quelle mesure de probabilité à densité à partir de la mesure de Lebesgue uniquement. Notons que $\lambda(\{a\}) = 0$. On a aussi les égalités

$$\lambda([a, b]) = \lambda(]a, b]) = b - a.$$

De plus, en utilisant les propriétés de base d'une mesure, on peut voir que si I est un intervalle non borné, alors $\lambda(I) = +\infty$. En particulier $\lambda(\mathbb{R}) = +\infty$.

3.1.1 Les mesures de Lebesgue-Stieltjes

Cet exemple généralise la plupart des mesures sur \mathbb{R} vues jusqu'à présent. Considérons une fonction $F : \mathbb{R} \rightarrow \mathbb{R}_+$ croissante et continue à droite.

Théorème 5 *Il existe une unique mesure μ_F sur \mathbb{R} muni de la tribu des boréliens telle que*

$$\mu_F(]a, b]) = F(b) - F(a), \quad a < b.$$

Remarquons que si $F(x) = x$, on retrouve la mesure de Lebesgue. Remarquons que si $a < b$, on a en utilisant la continuité inférieure de la mesure

$$\mu_F([a, b]) = \lim_{n \rightarrow +\infty} \mu_F\left(]a - \frac{1}{n}, b]\right) = \lim_{n \rightarrow +\infty} F(b) - F\left(a - \frac{1}{n}\right) = F(b) - F(a^-),$$

où $F(a^-)$ désigne la limite à gauche de F au point a . Ainsi $\mu_F(\{a\}) = F(a) - F(a^-)$, quantité qui vaut 0 si F est continue. Donnons deux exemples fondamentaux.

- Pour une suite $(p_n)_n$ de réels positifs sommable et une suite $(x_n)_n$ de nombre réels distincts, posons

$$F(x) = \sum_{n=0}^{+\infty} \mathbf{1}_{x_n \leq x} p_n, \quad x \in \mathbb{R}.$$

On peut vérifier que F est continue à droite (et bien sûr croissante). L'unicité du Théorème 5 entraîne que la mesure μ_F et la mesure discrète $\sum_{n=0}^{+\infty} p_n \delta_{x_n}$ coïncident sur $\mathcal{B}(\mathbb{R})$.

- Pour une fonction $f : \mathbb{R} \rightarrow \mathbb{R}_+$ intégrable (au sens de Riemann), la fonction F définie par

$$F(x) = \int_{-\infty}^x f(z) dz, \quad x \in \mathbb{R}$$

est continue à droite. Lorsque f est d'intégrale 1, La mesure μ_F correspond alors à la mesure de probabilité à densité f .

En probabilité, on a un lien entre les mesures de Lebesgue-Stieljes et la notion de fonction de répartition d'une variable aléatoire.

Définition 16 *Soit X une variable aléatoire à valeurs réelles, définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. La fonction F définie par*

$$F(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R},$$

est une fonction continue à droite appelée fonction de répartition.

Notons que $F(x) = \mathbb{P}_X(]-\infty, x])$. Ainsi les propriétés des mesures (appliquées ici à la mesure \mathbb{P}_X) entraîne que F est bien croissante et continue à droite et que

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

Nous reparlerons de ces propriétés lors du Chapitre 5. Ainsi, si X est une variable aléatoire à valeurs réelles, alors \mathbb{P}_X est la mesure de Lebesgue-Stieljes associée à F . On a alors la propriété suivante.

Proposition 12 *1. Soit fonction F , croissante et continue à droite telle que $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$. Alors F est la fonction de répartition d'une mesure de probabilité (et donc d'une variable aléatoire).*

- 2. Deux variables aléatoires à valeurs réelles dont les fonctions de répartition sont égales ont la même loi. La fonction de répartition caractérise donc complètement la loi d'une variable aléatoire.*

Preuve.

1. Soit $\mathbb{P} = \mu_F$ la mesure de Lebesgue-Stieltjes associée à F . D'après la propriété de continuité supérieure de P , on a

$$\mathbb{P}(\mathbb{R}) = \lim_{n \rightarrow +\infty} (F(n) - F(-n)) = 1$$

et donc \mathbb{P} est une mesure de probabilité. En posant $X(\omega) = \omega$ pour $\omega \in \Omega$, on a

$$\mathbb{P}(X \leq x) = F(x) - \lim_{n \rightarrow +\infty} F(-n) = F(x), \quad x \in \mathbb{R}.$$

Ainsi F est la fonction de répartition de la variable aléatoire X .

2. Supposons que deux variables aléatoires X et Y aient la même fonction de répartition F . Alors $\mathbb{P}_X = \mathbb{P}_Y = \mu_F$ et donc X et Y ont la même loi. \square

3.1.2 Un exemple qui heurte l'intuition : l'escalier de Cantor

On pourrait penser qu'il n'est pas possible de rejoindre le point $(0, 0)$ au point $(1, 1)$ à l'aide d'une fonction croissante et continue qui n'est strictement croissante sur aucun sous-intervalle de $[0, 1]$, aussi petit soit-il. Et pourtant...

On construit une suite de fonctions $(F_n)_n$ définies sur $[0, 1]$ et à valeurs dans $[0, 1]$ de la façon suivante. On pose $F_0(x) = x$ pour tout $x \in [0, 1]$. On construit ensuite F_1 en divisant l'intervalle $[0, 1]$ en trois. Sur $[0, 1/3]$, F_1 est affine et vérifie $F_1(0) = 0$ et $F_1(1/3) = 1/2$. Sur $[1/3, 2/3]$, F_1 vaut $1/2$. Enfin F_1 est affine sur l'intervalle $[2/3, 1]$ (voir Figure 3.1). On itère ensuite ce procédé : on divise tout intervalle I où F_n est affine et la fonction F_{n+1} sera affine par morceaux sur cet intervalle, constante et égale à la valeur $\frac{\max_I(F_n) + \min_I(F_n)}{2}$ sur l'intervalle du milieu (voir la courbe de F_2 sur la Figure 3.2). Plus formellement, on a pour $x \in [0, 1]$ et $n \in \mathbb{N}$:

$$F_{n+1}(x) = \frac{1}{2}F_n(3x)\mathbb{1}_{[0, \frac{1}{3}]}(x) + \frac{1}{2}F_n(3x)\mathbb{1}_{[\frac{1}{3}, \frac{2}{3}]}(x) + \left(\frac{1}{2} + \frac{1}{2}F_n(3 - 2x)\right)\mathbb{1}_{[\frac{2}{3}, 1]}(x),$$

ce qui permet d'obtenir la borne

$$\sup_{x \in [0, 1]} |F_{n+1}(x) - F_n(x)| \leq \frac{1}{2^n}.$$

On peut alors montrer que la suite de fonctions $(F_n)_n$ converge uniformément sur $[0, 1]$ vers une fonction F qui sera ainsi croissante et continue sur $[0, 1]$. Par contre, il n'existe aucun sous-intervalle de $[0, 1]$ sur lequel F est strictement croissante. On peut voir sur la Figure 3.3 la courbe de l'approximation F_{10} de F . En prolongeant F par 0 sur les réels négatifs et par 1 sur $[1, +\infty[$, on obtient une fonction de répartition sur \mathbb{R} . Le complémentaire de la réunion des intervalles où F est constante est appelé l'ensemble de Cantor $C(3)$. On peut montrer que $C(3)$ est non dénombrable. De plus, $C(3)$ est un borélien de mesure de Lebesgue nulle (voir ci-après), ce qui fournit un exemple d'ensemble non dénombrable mais de longueur nulle (un ensemble dénombrable étant forcément de mesure de Lebesgue nulle par σ -additivité). Aussi la mesure de probabilité μ_F associée à F n'admet pas d'atome ($\mu_F(\{x\}) = 0$ pour tout x car F est continue) mais n'admet pas de densité non plus (voir un peu plus loin, on peut déjà constater que $F'(x) = 0$ si $x \in C(3)^c$!)

En modifiant l'ensemble de Cantor, on peut aussi construire un borélien de longueur non nulle mais d'intérieur vide (c'est à dire pour lequel il est impossible d'y inclure un intervalle ouvert, aussi petit soit-il). Voici un exemple de construction. Soit k un entier plus grand que 3. On part de l'intervalle $[0, 1]$ et on pose $E_1 = E_{1,1} =]\frac{1}{2} - \frac{1}{2k}, \frac{1}{2} + \frac{1}{2k}[$ (intervalle de longueur $\frac{1}{k}$ centré en $\frac{1}{2}$). L'ensemble $[0, 1] \setminus E_1$ est composé de deux intervalles disjoints I_1 et I_2 . On enlève alors deux

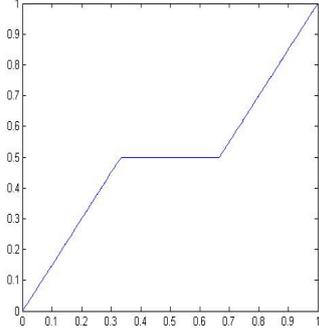


FIGURE 3.1: Graphe de F_0

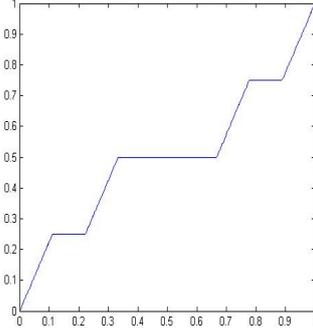


FIGURE 3.2: Graphe de F_1

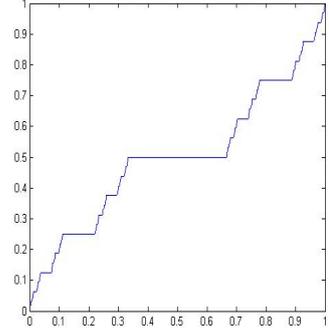


FIGURE 3.3: Graphe de F_{10}

intervalles ouverts E_{21} et E_{22} centré sur I_1 et I_2 et de longueur $\frac{1}{k^2}$. On pose alors $E_2 = E_{21} \cup E_{22}$. $[0, 1] \setminus (E_1 \cup E_2)$ est composé de 4 intervalles disjoints sur lesquels on enlève des intervalles de longueur $\frac{1}{k^3}$. Ainsi par récurrence, on peut définir pour tout $n \geq 1$ une suite d'intervalles ouverts disjoints deux à deux $(E_{ni})_{1 \leq i \leq 2^{n-1}}$ et tous de longueur $\frac{1}{k^n}$ et on pose $E_n = \cup_{i=1}^{2^{n-1}} E_{ni}$. Pour $k = 3$, E_n représente la réunion de tous les intervalles sur lesquels F_n devient constante sans que F_{n-1} le soit. Notons alors

$$C(k) = [0, 1] \setminus \cup_{n=1}^{\infty} E_n.$$

Cette notation est bien compatible avec le cas $k = 3$ car on retrouve bien l'ensemble de Cantor. Calculons la longueur de $C(k)$. Tout d'abord par additivité, on a

$$\lambda(E_n) = \sum_{i=1}^{2^{n-1}} \lambda(E_{ni}) = \frac{2^{n-1}}{k^n}.$$

Par σ -additivité, on trouve

$$\lambda(C(k)) = 1 - \sum_{n=1}^{+\infty} \frac{2^{n-1}}{k^n} = \frac{k-3}{k-2}.$$

Ainsi si $k = 3$, $C(k)$ est donc bien de longueur nulle, alors que pour $k > 3$, $C(k)$ a une longueur strictement positive. De plus $C(k)$ est d'intérieur vide (car la longueur des intervalles conservés d'une étape à la suivante est divisée par 2).

3.2 Intégrale d'une fonction mesurable par rapport à une mesure

Soit (E, \mathcal{E}, μ) un espace mesuré (c'est à dire un triplet composé d'un ensemble, d'une tribu sur cet ensemble et d'une mesure définie sur cette tribu). On considérera des fonctions $f : E \rightarrow \mathbb{R} \cup \{-\infty, +\infty\}$ dites mesurables, c'est à dire telles que

1. Pour tout couple (a, b) de nombres réels, $\{x \in E : a \leq f(x) \leq b\} \in \mathcal{E}$.
2. $\{x \in E : f(x) = -\infty\} \in \mathcal{E}$.

Une fonction mesurable est donc l'analogie d'une variable aléatoire dans le cadre la théorie des probabilités (on rajoute la possibilité pour ce type de fonctions de prendre des valeurs $+\infty$ ou $-\infty$, ce qui peut parfois s'avérer utile). Comme pour les variables aléatoires on peut montrer que pour tout borélien $B \in \mathcal{B}(\mathbb{R})$, l'ensemble $\{x \in E : f(x) \in B\}$ appartient à \mathcal{E} .

Dans la suite, on dira qu'une fonction $f : E \rightarrow \mathbb{R}$ est étagée si elle est mesurable et ne prend

qu'un nombre fini de valeurs. Si, on note y_1, \dots, y_N les valeurs distinctes d'une fonction étagée f , on convient d'écrire dans la suite

$$f = \sum_{i=1}^N y_i \mathbb{1}_{A_i},$$

où on a posé pour $1 \leq i \leq N$, $A_i = \{f = y_i\} = \{x \in E : f(x) = y_i\}$ qui est un élément de \mathcal{E} . On peut toujours approcher une fonction mesurable f à valeurs dans $\bar{\mathbb{R}}_+ = \mathbb{R} \cup \{+\infty\}$ par une suite croissante de fonctions positives étagées. Il suffit de modifier la suite définie au Chapitre 1 (cf Proposition 6) en posant pour $n \in \mathbb{N}$:

$$f_n(x) = \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} \mathbb{1}_{\frac{k}{2^n} \leq f < \frac{k+1}{2^n}} + 2^n \mathbb{1}_{f \geq 2^n}, \quad x \in E. \quad (3.1)$$

La preuve est alors quasiment identique à celle de la Proposition 6.

Pour pouvoir additionner ou multiplier des fonctions mesurables, il est nécessaire de prolonger l'addition et la multiplication à $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ en adoptant les conventions suivantes.

$$\begin{aligned} +\infty + a &= a + (+\infty) = +\infty, & a \in \mathbb{R} \cup \{+\infty\}, \\ -\infty + a &= a + (-\infty) = -\infty, & a \in \mathbb{R} \cup \{-\infty\}, \\ a \times (+\infty) &= (+\infty) \times a = +\infty, & a > 0, \\ a \times (+\infty) &= (+\infty) \times a = -\infty, & a < 0, \\ a \times (-\infty) &= (-\infty) \times a = -\infty, & a > 0, \\ a \times (-\infty) &= (-\infty) \times a = +\infty, & a < 0, \\ 0 \times (+\infty) &= (+\infty) \times 0 = 0 \times (-\infty) = (-\infty) \times 0 = 0. \end{aligned}$$

Seule la somme $+\infty + (-\infty)$ n'est pas définie.

Lorsque $E = \mathbb{R}$ et $\mathcal{E} = \mathcal{B}(\mathbb{R})$, la plupart des fonctions sont mesurables (les fonctions continues, continues par morceaux et des fonctions bien plus irrégulières encore). La notion de mesurabilité est stable par tout un tas d'opérations, comme le montre la proposition suivante.

Proposition 13 1. Soient $f, g : E \rightarrow \bar{\mathbb{R}}$ deux fonctions mesurables. Alors la fonction fg est encore mesurable. Si de plus $f + g$ est définie (c'est à dire lorsque la somme $+\infty + (-\infty)$ n'intervient pas) alors $f + g$ est mesurable.

2. Si $f : E \rightarrow \mathbb{R}^d$ est une fonction mesurable et $g : \mathbb{R}^d \rightarrow \mathbb{R}$ est une fonction mesurable lorsque \mathbb{R}^d est muni de la tribu des boréliens alors $g \circ f : E \rightarrow \mathbb{R}$ est mesurable.

3. Soit $f_n : E \rightarrow \bar{\mathbb{R}}$ une suite de fonctions mesurables. Alors les applications $\inf_{n \in \mathbb{N}} f_n$ et $\sup_{n \in \mathbb{N}} f_n$ sont mesurables. Si de plus la suite est convergente point par point vers une fonction f , alors f est également mesurable.

Remarquons que le point 2. permet d'affirmer que $g(X)$ est une variable aléatoire dès que X est une variable aléatoire et g est une fonction mesurable. La preuve de ce point est d'ailleurs triviale : si $A \in \mathcal{B}(\mathbb{R})$, alors

$$\{g(X) \in A\} = \{X \in g^{-1}(A)\}$$

qui est bien un élément de \mathcal{A} vu que la mesurabilité de g entraîne que $g^{-1}(A) \in \mathcal{B}(\mathbb{R}^d)$.

Cette stabilité de la notion de mesurabilité est remarquable. Dans l'intégrale de Riemann, les suites de fonctions Riemann-intégrables peuvent avoir une limite non Riemann-intégrable.

Nous allons intégrer les fonctions mesurables à partir des fonctions étagées (comme pour les variables aléatoires).

Pour une fonction étagée $f = \sum_{i=1}^N y_i \mathbf{1}_{f=y_i}$, on pose

$$\int f d\mu = \sum_{i=1}^N y_i \mu(f = y_i).$$

On retiendra en particulier que $\int \mathbf{1}_A d\mu = \mu(A)$ et que lorsque f est constante égale à m , on a $\int f d\mu = m \times \mu(E)$. La définition de l'intégrale pour une fonction mesurable générale se fait alors comme pour l'espérance des variables aléatoires.

Proposition 14 1. Soit $f : E \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ une fonction mesurable. L'intégrale de f par rapport à la mesure μ est notée $\int f d\mu$ ou aussi $\int f(x) d\mu(x)$ et est définie par

$$\int f d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu,$$

où $(f_n)_n$ est une suite croissante de fonctions étagées positives, convergente point par point vers f .

2. Soit $f : E \rightarrow \bar{\mathbb{R}}$ une fonction mesurable. Soient f^+ et f^- les parties positives et négatives de f (voir Chapitre 1). On dit que f est intégrable lorsque $\int f^+ d\mu < +\infty$ et $\int f^- d\mu < +\infty$ sont intégrables. Dans ce cas, on pose

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

Ainsi si $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probablisé et $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire intégrable, on a $\mathbb{E}(X) = \int X d\mathbb{P}$: l'espérance mathématique est donc un cas particulier d'intégrale par rapport à une mesure. On dispose alors des propriétés suivantes qui se démontrent de la même façon que les propriétés de l'espérance des variables aléatoires (voir Chapitre 1).

Proposition 15 Soient $f, g : E \rightarrow \bar{\mathbb{R}}$ deux fonctions mesurables ou bien toutes deux à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$ ou bien toutes deux intégrables.

1. On a $\int (f + g) d\mu = \int f d\mu + \int g d\mu$ lorsque la fonction somme $f + g$ est bien définie. En particulier, $f + g$ est intégrable lorsque f et g sont intégrables.
2. Si $a \in \mathbb{R}$, alors $\int a f d\mu = a \int f d\mu$.
3. Si $f \leq g$ alors $\int f d\mu \leq \int g d\mu$.
4. La fonction f est intégrable si et seulement si la fonction $|f|$ est intégrable (ce qui s'écrit $\int |f| d\mu < +\infty$). On a alors

$$\left| \int f d\mu \right| \leq \int |f| d\mu.$$

Si $A \in \mathcal{E}$ et $f : E \rightarrow \bar{\mathbb{R}}$, on définit

$$\int_A f d\mu = \int \mathbf{1}_A f d\mu.$$

Autrement dit, on définit l'intégrale de f sur A comme l'intégrale sur E de la fonction qui est égale à f sur A et qui vaut 0 sur A^c . Remarquons que lorsque f est intégrable, $f \mathbf{1}_A$ l'est également car sa valeur absolue est intégrable : en effet

$$|f \mathbf{1}_A| = \mathbf{1}_A |f| \leq |f|$$

et la proposition précédente assure que $\int |f\mathbf{1}_A|d\mu \leq \int |f|d\mu < +\infty$.

Une autre conséquence de la linéarité de l'intégrale est la suivante : si A et B sont deux éléments de \mathcal{E} disjoints et f une fonction mesurable positive ou intégrable alors

$$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$

En effet, en utilisant l'égalité $\mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B$, il est facile de vérifier que ces deux quantités valent $\int (\mathbf{1}_A + \mathbf{1}_B) f d\mu$.

3.2.1 Intégration par rapport à une mesure discrète.

Soit $\mu = \sum_{n=0}^{+\infty} p_n \delta_{x_n}$ où $(p_n)_n$ est suite de réels positifs et $(x_n)_n$ une suite de points distincts de $E = \mathbb{R}^d$ que l'on munit de l'ensemble de ses parties. Alors dans ce cas nous obtenons le résultat suivant.

Proposition 16 *Toute fonction $f : E \rightarrow \mathbb{R}$ est mesurable. Lorsque f prend ses valeurs dans \mathbb{R}_+ ou que f est intégrable, alors*

$$\int f d\mu = \sum_{n=0}^{+\infty} p_n f(x_n).$$

En particulier f est intégrable si et seulement si $\sum_{n=0}^{+\infty} p_n |f(x_n)| < +\infty$.

Preuve Le fait que toute fonction soit mesurable résulte du choix de la tribu $\mathcal{P}(\mathbb{R}^d)$.

- Commençons par le cas d'une fonction étagée positive prenant les valeurs y_1, \dots, y_N et posons $A_i = \{f = y_i\}$ pour $i = 1, \dots, N$. Remarquons alors que $x_n \in A_i \Leftrightarrow f(x_n) = y_i$. Alors on a

$$\begin{aligned} \int f d\mu &= \sum_{i=1}^N y_i \mu(A_i) \\ &= \sum_{i=1}^N y_i \sum_{n=0}^{+\infty} p_n \mathbf{1}_{A_i}(x_n) \\ &= \sum_{n=0}^{+\infty} p_n \sum_{i=1}^N y_i \mathbf{1}_{A_i}(x_n) \\ &= \sum_{n=0}^{+\infty} p_n \sum_{i=1}^N f(x_n) \mathbf{1}_{A_i}(x_n) \\ &= \sum_{n=0}^{+\infty} p_n f(x_n). \end{aligned}$$

La dernière égalité est due au fait que x_n ne peut appartenir qu'à un seul ensemble A_i .

- Passons maintenant au cas d'une fonction positive. Si f_k est une fonction étagée plus petite que f , alors on a

$$\int f_k d\mu = \sum_{n=0}^{+\infty} p_n f_k(x_n) \leq \sum_{n=0}^{+\infty} p_n f(x_n).$$

Ainsi, par définition de l'intégrale, on a nécessairement $\int f d\mu \leq \sum_{n=0}^{+\infty} p_n f(x_n)$. Pour montrer l'inégalité dans l'autre sens, on peut observer que $f \geq f \mathbf{1}_{\{x_0, \dots, x_n\}} = g$. La fonction g

ne prend qu'un nombre fini de valeurs et il est facile de prouver que $\int g d\mu = \sum_{n=0}^N p_n f(x_n)$ en utilisant le premier point. En utilisant le point 3. de la propriété 15, on a

$$\int f d\mu \geq \int g d\mu = \sum_{i=1}^N p_i f(x_i).$$

En passant à la limite sur N , on voit que

$$\int f d\mu \geq \sum_{n=0}^{+\infty} p_n f(x_n).$$

On a donc bien la formule annoncée.

- Considérons enfin le cas d'une fonction f de signe quelconque. f est intégrable si et seulement si $|f|$ l'est, ce qui signifie d'après le point précédent que $\sum_{n=0}^{+\infty} |f(x_n)| p_n < +\infty$. Dans ce cas, on a, en appliquant la définition de l'intégrale et le point précédent

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu = \sum_{n=0}^{+\infty} p_n (f^+(x_n) - f^-(x_n)) = \sum_{n=0}^{+\infty} p_n f(x_n),$$

ce qui achève la preuve. \square

Enfin, mentionnons le théorème fondamental suivant.

Théorème 6 (convergence monotone) *Pour toute suite croissante $f_n : E \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ de fonctions mesurables, on a*

$$\int \lim_{n \rightarrow +\infty} f_n d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu.$$

Évidemment lorsque une suite de fonctions positives est croissante, sa limite est bien définie. Cette limite f est de plus une fonction mesurable positive et l'intégrale est donc bien définie. Mentionnons une conséquence en probabilité. Supposons que $(X_n)_{n \in \mathbb{N}}$ soit une suite de variables aléatoires toutes positives. Alors on a

$$\mathbb{E} \left(\sum_{n=0}^{+\infty} X_n \right) = \sum_{n=0}^{+\infty} \mathbb{E}(X_n).$$

Pour le voir, il suffit de considérer $f_N = \sum_{n=0}^N X_n$ et d'appliquer le théorème de la convergence monotone.

3.2.2 L'intégrale de Lebesgue et l'intégrale de Riemann

Lorsque $\mu = \lambda$ (mesure de Lebesgue sur \mathbb{R}), on obtient une intégrale appelée intégrale de Lebesgue. Nous allons voir que cette intégrale coïncide le plus souvent avec l'intégrale de Riemann lorsque cette dernière a bien un sens. De plus il existe des fonctions intégrables au sens de Lebesgue mais pas au sens de Riemann. Rappelons la définition de l'intégrale de Riemann sur un intervalle fermé borné $[a, b]$ de \mathbb{R} . Une partie finie Δ de $[a, b]$ contenant les points a et b est appelée une subdivision de $[a, b]$ et sera notée

$$\Delta := a = x_1 < x_2 < \dots < x_n = b.$$

La finesse d'une subdivision est définie par $|\Delta| = \max_{1 \leq i \leq n} |x_{i+1} - x_i|$.

Soit $f : [a, b] \rightarrow \mathbb{R}$ une fonction bornée. Pour toute subdivision $\Delta : a = x_1 < x_2 < \dots < x_n$, on pose

$$m_i = \inf_{x_i \leq x \leq x_{i+1}} f(x), \quad M_i = \sup_{x_i \leq x \leq x_{i+1}} f(x).$$

On définit alors

$$s_\Delta = \sum_{i=1}^n m_i (x_{i+1} - x_i), \quad S_\Delta = \sum_{i=1}^n M_i (x_{i+1} - x_i).$$

L'idée est alors d'encadrer l'aire sous la courbe de f entre celle de $g_\Delta = \sum_{i=1}^n m_i \mathbf{1}_{]x_i, x_{i+1}]}$ et $h_\Delta = \sum_{i=1}^n M_i \mathbf{1}_{]x_i, x_{i+1}]}$. On peut remarquer que

$$\Delta \subset \Delta' \Rightarrow s_\Delta \leq s_{\Delta'} \leq S_{\Delta'} \leq S_\Delta.$$

On dit alors que f est Riemann-intégrable si $\sup_\Delta s_\Delta = \inf_\Delta S_\Delta$, valeur appelée intégrale de f sur $[a, b]$. On peut montrer que f est Riemann-intégrable si et seulement si il existe une suite croissante $(\Delta_k)_k$ de subdivisions dont la finesse tend vers 0 telle que

$$\lim_{k \rightarrow \infty} s_{\Delta_k} = \lim_{k \rightarrow \infty} S_{\Delta_k}.$$

Il existe des fonctions Lebesgue-intégrables qui ne sont pas Riemann-intégrables. C'est par exemple le cas pour $f = \mathbf{1}_\mathbb{Q}$ sur l'intervalle $[0, 1]$. En effet, pour toute subdivision Δ de $[0, 1]$, on a $M_i = 1$ et $m_i = 0$ pour tout i , ce qui entraîne $s_\Delta = 0$ et $S_\Delta = 1$ et f ne peut être Riemann-intégrable sur $[0, 1]$. En revanche f est Lebesgue-intégrable et

$$\int_0^1 f(x) d\lambda(x) = \lambda(\mathbb{Q} \cap [0, 1]) = 0,$$

car $\mathbb{Q} \cap [0, 1]$ est dénombrable. Ainsi, le théorème de convergence monotone n'est pas vrai pour les fonctions Riemann-intégrables : si $(r_n)_n$ est une énumération des nombres rationnels de $[0, 1]$, la suite de fonctions $f_n = \mathbf{1}_{\{r_0, \dots, r_n\}}$ est croissante vers $f = \mathbf{1}_{\mathbb{Q} \cap [0, 1]}$ qui n'est pas Riemann-intégrable alors que f_n l'est et vérifie $\int_0^1 f_n(x) dx = 0$.

Nous avons en fait le résultat suivant.

Proposition 17 *Soit $f : [a, b] \rightarrow \mathbb{R}$ une fonction bornée.*

1. *Si f est Riemann-intégrable, f est Lebesgue-intégrable et les valeurs des deux intégrales sont les mêmes.*
2. *La fonction f est Riemann-intégrable si et seulement si l'ensemble D des points de discontinuité de f vérifie $\lambda(D) = 0$.*

On peut vérifier que la fonction $\mathbf{1}_\mathbb{Q}$ n'est continue en aucun point, le point 2. de la proposition confirme la non-intégrabilité de cette fonction au sens de Riemann.

Au niveau des intégrales impropres, toute fonction positive admettant une intégrale impropre au sens de Riemann est intégrable au sens de Lebesgue et les intégrales sont égales. Prouvons-le pour une fonction $f : [0, +\infty[\rightarrow [0, +\infty[$. On a

$$\int_0^{+\infty} f(x) dx = \lim_{n \rightarrow +\infty} \int_0^n f(x) dx = \lim_{n \rightarrow +\infty} \int_0^n f(x) d\lambda(x) = \int_0^{+\infty} f(x) d\lambda(x),$$

où la deuxième égalité résulte de la proposition précédente et la troisième du théorème de la convergence monotone (car $f_n = f \mathbf{1}_{[0, n]}$ est une suite croissante de fonctions positives).

Il existe néanmoins un cas où une fonction peut admettre une intégrale impropre au sens de Riemann sans être intégrable au sens de Lebesgue. C'est le cas des fonctions oscillantes. Un exemple simple est donnée par la fonction $f = \sum_{n=0}^{+\infty} a_n \mathbf{1}_{[n, n+1[}$ qui admet une intégrale impropre au sens de Riemann lorsque la série $\sum_{n=0}^{+\infty} a_n$ est convergente mais qui n'est pas Lebesgue-intégrable sur $[0, +\infty[$ lorsque cette série n'est pas absolument convergente (l'intégrale de Lebesgue demande à ce que la valeur absolue de f ait une intégrale finie).

Notation. Au vu des résultats précédents, nous noterons souvent $\int f(x)dx$ (au lieu de $\int f(x)d\lambda(x)$) l'intégrale de la fonction f au sens de Lebesgue. De plus les résultats utilisés pour les intégrales de Riemann (intégration par parties, formule du changement de variable) peuvent aussi être utilisés pour l'intégrale de Lebesgue (nous énoncerons la formule du changement de variables dans \mathbb{R}^n un peu plus loin). On pourra noter la différence fondamentale concernant la construction de ces deux intégrales : l'intégration au sens de Riemann se base sur l'approximation par des fonctions dites en escalier construites à l'aide de subdivisions sur l'axe des abscisses alors que l'intégrale de Lebesgue utilise des fonctions étagées construites à l'aide de subdivisions de l'axe des ordonnées (voir la suite de fonctions 3.1).

3.2.3 Le presque partout

Soit (E, \mathcal{E}, μ) un espace mesuré. On dira que deux fonctions mesurables $f, g : E \rightarrow \bar{\mathbb{R}}$ sont égales presque partout et on note $f = g$ p.p, lorsque $\mu(f \neq g) = 0$. Lorsque μ est une mesure de probabilité, on dit plutôt que f et g sont égales presque sûrement (et on note $f = g$ p.s), ce qui revient aussi à avoir $\mu(f = g) = 1$.

Proposition 18 Soit $f, g : E \rightarrow \bar{\mathbb{R}}$ deux fonctions mesurables.

1. Supposons $f = g$ p.p. Alors f est intégrable si et seulement si g est intégrable. Dans ce cas on a $\int f d\mu = \int g d\mu$. En particulier, pour $A \in \mathcal{E}$, on a $\int_A f(x) d\mu(x) = 0$ lorsque $\mu(A) = 0$.
2. Supposons que f soit à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$. Alors $\int f d\mu = 0$ si et seulement si $f = 0$ p.p.
3. Supposons que f soit à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$. De plus si $\int f d\mu < +\infty$ alors $\mu(f = \infty) = 0$. Dans ce cas on dit que $f < +\infty$ p.p.

Remarques.

- Considérons le cas où $(E, \mathcal{E}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. Si $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction intégrable et $g : \mathbb{R} \rightarrow \mathbb{R}$ telle que $g(x) = f(x)$ pour $x \in D^c$. Si $\lambda(D) = 0$ (c'est par exemple le cas si D est dénombrable), alors le premier point de cette proposition assure que l'intégrale de g est égale à l'intégrale de f .
- Le deuxième point a déjà été utilisé au Chapitre 1 pour montrer qu'une variable aléatoire positive d'espérance nulle est une variable aléatoire nulle p.s et donc qu'une variable aléatoire de carré intégrable et de variance nulle était presque sûrement égale à sa moyenne.
- Illustrons le troisième point dans le cadre des probabilités. Considérons une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires telle que pour tout $n \in \mathbb{N}$, X_n suive une loi de Bernoulli de paramètre p_n . Alors si $\sum_{n=0}^{+\infty} p_n < +\infty$, la variable aléatoire $X = \sum_{n=0}^{+\infty} X_n$ (qui existe en tant que limite d'une suite croissante de variables aléatoires) est finie presque sûrement. En effet, on a par convergence monotone

$$\mathbb{E} \left(\sum_{n=0}^{+\infty} X_n \right) = \sum_{n=0}^{+\infty} \mathbb{E}(X_n) = \sum_{n=0}^{+\infty} p_n < +\infty,$$

et la variable aléatoire X est finie presque sûrement d'après le troisième point de la proposition précédente. La variable aléatoire X ne peut évidemment être finie en tout point ω : elle vaut $+\infty$ lorsque qu'une infinité de 1 apparaisse dans la suite ; par contre la probabilité que cela se produise est nulle.

3.3 Mesure produit et théorème de Fubini

Les produits de mesure généralisent ceux déjà vues pour les mesures de probabilité (cf Chapitre 1). Cette notion permet également de définir l'aire ou le volume d'un borélien de \mathbb{R}^2 ou de \mathbb{R}^3 , en effectuant un produit de mesure à l'aide de la mesure de Lebesgue. Rappelons (voir Chapitre 1) que si E et F sont deux ensembles, chacun muni d'une tribu notée respectivement \mathcal{E} et \mathcal{F} alors la tribu produit sur $E \times F$ est notée $\mathcal{E} \otimes \mathcal{F}$ et est définie par

$$\mathcal{E} \otimes \mathcal{F} = \sigma(\{B \times C : (B, C) \in \mathcal{E} \times \mathcal{F}\}).$$

On définit le produit de deux mesures μ et ν sur $E \times F$ lorsque les deux mesures sont σ -finies. On dit qu'une mesure μ est σ -finie lorsqu'on peut écrire $E = \cup_{n \in \mathbb{N}} E_n$ avec $(E_n)_{n \in \mathbb{N}}$ une suite d'éléments de \mathcal{E} , croissante pour l'inclusion, et telle que $\mu(E_n) < +\infty$ pour tout $n \in \mathbb{N}$. Toute mesure de probabilité est σ -finie (en prenant $E_n = E$ pour tout n). La mesure de Lebesgue est aussi σ -finie car $\mathbb{R} = \cup_{n \in \mathbb{N}} [-n, n]$ et $\lambda([-n, n]) = 2n < +\infty$. Plus généralement, toute mesure de Lebesgue-Stieljes est σ -finie.

Théorème 7 *Si (E, \mathcal{E}, μ) et (F, \mathcal{F}, ν) sont deux espaces mesurés avec des mesures μ et ν toutes deux σ -finies, alors il existe une unique mesure, notée $\mu \otimes \nu$, définie sur $\mathcal{E} \otimes \mathcal{F}$ et telle que*

$$\mu \otimes \nu(B \times C) = \mu(B) \times \nu(C), \quad (B, C) \in \mathcal{E} \times \mathcal{F}. \quad (3.2)$$

On dit que $\mu \otimes \nu$ est la mesure produit de μ et de ν .

Par récurrence, on peut aussi définir la mesure produit de n mesures σ -finies (comme cela a été directement énoncé au Chapitre 1).

Cas particulier fondamental. Supposons que $E = F = \mathbb{R}$ est muni de la tribu des boréliens et que $\mu = \nu = \lambda$ la mesure de Lebesgue. Dans ce cas $\lambda \otimes \lambda$ est notée λ_2 et est appelée mesure de Lebesgue sur \mathbb{R}^2 . Ainsi pour deux boréliens B et C de \mathbb{R} , on a

$$\lambda_2(B \times C) = \lambda(B)\lambda(C).$$

On dit alors que λ_2 est la mesure d'aire (pour $A \in \mathcal{B}(\mathbb{R}^2)$, on dit que $\lambda_2(A)$ est le volume du borélien A). Il est également possible de définir par récurrence le produit $\lambda \otimes \dots \otimes \lambda$ de n mesures de Lebesgue, mesure notée λ_n et qui est appelée la mesure de Lebesgue sur \mathbb{R}^n . Pour $n = 3, 4, \dots$, on parle de mesure de volume.

Revenons au cas général d'un produit $\mu \otimes \nu$. Évidemment un élément A de la tribu $\mathcal{E} \otimes \mathcal{F}$ ne peut pas toujours s'écrire $B \times C$ avec $(B, C) \in \mathcal{E} \times \mathcal{F}$ (par exemple si A est un disque de \mathbb{R}^2). Il est alors possible de calculer la mesure de A à partir de la mesure de ces "sections". Plus précisément, si $(x, y) \in E \times F$, on définit les sections $A(x)$ et $A(y)$ de A par

$$A(x) = \{y' \in F : (x, y') \in A\}, \quad A(y) = \{x' \in E : (x', y) \in A\}.$$

Il est possible de vérifier que $A(x) \in \mathcal{F}$ et $A(y) \in \mathcal{E}$. De plus on peut montrer que

$$(\mu \otimes \nu)(A) = \int \mu(A(y)) d\nu(y) = \int \nu(A(x)) d\mu(y). \quad (3.3)$$

On peut donc calculer la mesure de A en intégrant la mesure de ses sections. Lorsque $\mu = \nu = \lambda$, les longueurs de deux sections d'un borélien de \mathbb{R}^2 sont représentées sur la Figure 3.4. La

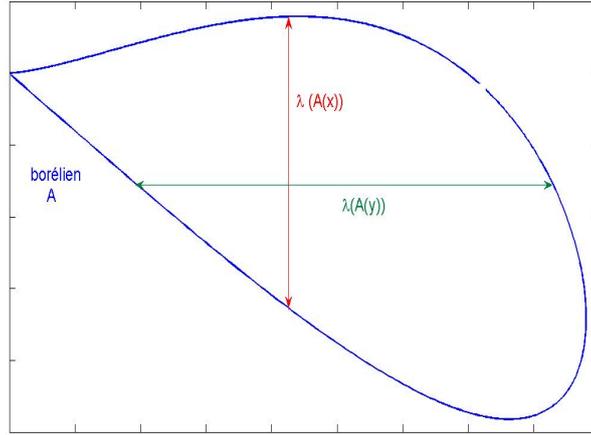


FIGURE 3.4: Un borélien et les longueurs de deux de ses sections

possibilité d'intégrer les sections en commençant par une ou l'autre des variables x ou y peut être justifiée : les deux applications $\kappa_1 : \mathcal{E} \otimes \mathcal{F} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ et $\kappa_2 : \mathcal{E} \otimes \mathcal{F} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ définies par

$$\kappa_1(A) = \int \mu(A(y)) d\nu(y), \quad \kappa_2(A) = \int \nu(A(x)) d\mu(x),$$

sont en fait deux mesures σ -finies qui satisfont (3.2) et ces mesures sont donc égales par unicité.

On a alors le théorème fondamental dit théorème de Fubini qui permet de calculer l'intégrale d'une fonction par rapport à la mesure produit en utilisant uniquement le calcul intégral sur les mesures μ et ν .

Théorème 8 Soit $f : E \times F \rightarrow \bar{\mathbb{R}}$ une application mesurable (par rapport à la tribu produit).

1. Si f prend des valeurs positives ou éventuellement $+\infty$ alors

$$\int f(x, y) d\mu \otimes \nu(x, y) = \int \left(\int f(x, y) d\mu(x) \right) d\nu(y) = \int \left(\int f(x, y) d\nu(y) \right) d\mu(x). \quad (3.4)$$

2. Supposons que f prenne des valeurs quelconques. Alors, f est intégrable si et seulement si l'une ou l'autre des intégrales

$\int \left(\int |f(x, y)| d\nu(y) \right) d\mu(x)$, $\int \left(\int |f(x, y)| d\mu(x) \right) d\nu(y)$ est finie (remarquer que ces deux intégrales sont forcément égales d'après le point 1.). Dans ce cas on a encore (3.4).

On pourra remarquer que lorsque $f = \mathbb{1}_A$, alors on retrouve bien les égalités (3.3).

Ce théorème est extrêmement utile. Voyons deux cas particuliers rencontrés couramment.

- Lorsque $E = F = \mathbb{R}$ et $\mu = \nu = \lambda$, on a sous les hypothèses du théorème

$$\int f(x, y) d\lambda_2(x, y) = \int \left(\int f(x, y) dx \right) dy = \int \left(\int f(x, y) dy \right) dx.$$

L'intégrale sur \mathbb{R}^2 fournit un calcul du volume situé sous le graphe de la fonction f . Ce volume peut être calculé, en commençant d'abord par calculer l'aire située sous le graphe des applications partielles $x \mapsto f(x, y)$ ou $y \mapsto f(x, y)$ (voir Figure 3.5). Mentionnons deux cas particuliers fondamentaux. Lorsque $f(x, y) = g(x)h(y)$ pour deux fonctions mesurables

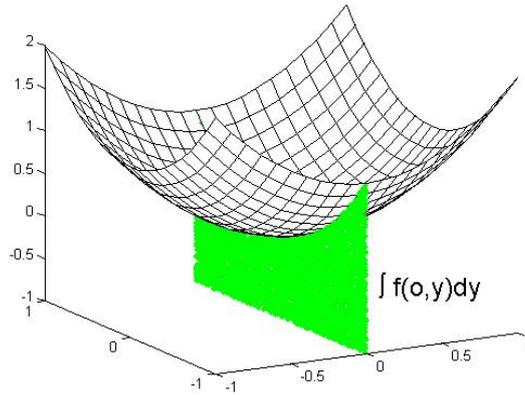


FIGURE 3.5: Illustration du calcul d'aire

g et h , alors f est intégrable (pour λ_2) si et seulement si g et h sont intégrables (pour λ). De plus, on a dans ce cas

$$\int f(x, y) d\lambda_2(x, y) = \int g(x) dx \times \int h(y) dy,$$

égalité qui est en fait toujours vérifiée si h et g sont à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$. Le deuxième cas concerne l'intégration sur un sous-ensemble de \mathbb{R}^2 . Rappelons que par définition

$$\int_A f(x, y) d\lambda_2(x, y) = \int \mathbf{1}_A(x, y) f(x, y) d\lambda_2(x, y).$$

Lorsque $A = B \times C$ et sous les conditions du théorème de Fubini, on a

$$\int_A f(x, y) d\lambda_2(x, y) = \int_B \left(\int_C f(x, y) dy \right) dx = \int_C \left(\int_B f(x, y) dx \right) dy.$$

– Toujours lorsque $E = F = \mathbb{R}$ mais que $\nu = \lambda$ et μ est la mesure de comptage sur \mathbb{N} , on a

$$\sum_{n=0}^{+\infty} \int f(n, y) dy = \int \left(\sum_{n=0}^{+\infty} f(n, y) \right) dy,$$

lorsque f est positive ou intégrable (ce qui signifie que $\sum_{n=0}^{+\infty} \int |f(n, y)| dy < +\infty$).

– Lorsque μ et ν sont toutes deux égales à la mesure de comptage sur \mathbb{N} , on retrouve le théorème de Fubini énoncé au Chapitre 2.

3.4 Mesure à densité et théorème de transfert

Maintenant que l'intégrale d'une fonction par rapport à une mesure a été définie, on peut construire d'autres mesures appelées mesures à densité.

Proposition-Définition 2 Soit (E, \mathcal{E}, μ) un espace mesuré et soit $f : E \rightarrow \mathbb{R}_+$ une fonction mesurable. Alors l'application $\nu : \mathcal{E} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ définie par

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

est une mesure appelée mesure de densité f par rapport à μ . On note $\nu = f \cdot \mu$.

Remarque. Il est important de remarquer que la densité f de la mesure ν n'est pas unique. Supposons qu'une fonction mesurable g soit telle que $\mu(\{x \in E : g(x) \neq f(x)\}) = 0$ (c'est à dire $f = g$ p.p). Alors g est encore une densité car

$$\begin{aligned} \nu(A) &= \int_A f(x) d\nu(x) \\ &= \int_A \mathbf{1}_{f(x)=g(x)} f(x) d\mu(x) + \int_A \mathbf{1}_{f(x) \neq g(x)} f(x) d\mu(x) \\ &= \int_A \mathbf{1}_{f(x)=g(x)} g(x) d\mu(x) + 0 \\ &= \int_A \mathbf{1}_{f(x)=g(x)} g(x) d\mu(x) \\ &= \int_A g(x) d\mu(x). \end{aligned}$$

Nous avons utilisé le premier point de la Proposition 18. On peut donc modifier une densité sur un ensemble de mesure nulle. Ainsi, si $\mu = \lambda$ la mesure de Lebesgue sur \mathbb{R} , toute fonction obtenue en modifiant la densité de départ en un nombre fini ou infini dénombrable de points est encore une densité. Lorsque $\mu = \lambda_2$ est la mesure de Lebesgue sur \mathbb{R}^2 , alors on peut modifier une densité le long d'un segment ou d'une droite de \mathbb{R}^2 : par exemple si $D = \{(x, y) : x = 0\}$, on a par le théorème de Fubini

$$\lambda_2(D) = \int \lambda(D_y) dy = \int \lambda(\{0\}) dy = 0.$$

Preuve. Vérifions que ν est bien une mesure.

On a $\nu(\emptyset) = 0$ car la fonction $\mathbf{1}_\emptyset f$ est nulle. De plus si $(A_n)_{n \in \mathbb{N}}$ est une suite d'éléments de \mathcal{E} disjoints deux à deux, alors en utilisant le fait que

$$\mathbf{1}_{\cup_{n=0}^{+\infty} A_n} = \sum_{n=0}^{+\infty} \mathbf{1}_{A_n}$$

et le théorème de la convergence monotone, on a

$$\begin{aligned} \nu(\cup_{n=0}^{+\infty} A_n) &= \int \mathbf{1}_{\cup_{n=0}^{+\infty} A_n} f d\mu \\ &= \int \sum_{n=0}^{+\infty} \mathbf{1}_{A_n} f d\mu \\ &= \sum_{n=0}^{+\infty} \int \mathbf{1}_{A_n} f d\mu \\ &= \sum_{n=0}^{+\infty} \nu(A_n). \end{aligned}$$

Ainsi ν est bien une mesure. \square

Exemples

- Lorsque $(E, \mathcal{E}, \mu) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ et $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ est une fonction intégrable et d'intégrale 1, on dit que ν est une mesure de probabilité à densité par rapport à la mesure de Lebesgue (lorsque il n'y aura pas d'ambiguïté, on parlera simplement de mesure de probabilité à densité).

- Supposons $(E, \mathcal{E}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ et soit $f : \mathbb{R} \rightarrow \mathbb{R}_+$ une fonction intégrable par rapport à λ . Alors la mesure $\nu = f \cdot \lambda$ coïncide avec la mesure de Lebesgue-Stieltjes μ_F associée à la fonction $F : \mathbb{R} \rightarrow \mathbb{R}_+$ définie par $F(x) = \int_{-\infty}^x f(z)dz$ pour tout $x \in \mathbb{R}$. Pour vérifier ceci, on vérifie que ν et μ_F sont égales sur les intervalles. Il est ensuite possible de montrer que l'ensemble des boréliens A pour lesquels ν et μ_F coïncident est une tribu. Comme cette tribu contient les intervalles, elle est forcément égale à $\mathcal{B}(\mathbb{R})$, ce qui montre l'égalité des deux mesures sur $\mathcal{B}(\mathbb{R})$.
- Lorsque $E = \mathbb{R}$, $\mathcal{E} = \mathcal{P}(\mathbb{R})$ et μ désigne la mesure de dénombrement sur \mathbb{N} , toute mesure discrète $\nu = \sum_{n=0}^{+\infty} p_n \delta_n$ possède une densité f par rapport à μ . Il suffit de définir $f : \mathbb{R} \rightarrow \mathbb{R}_+$ par $f(n) = p_n$ pour $n \in \mathbb{N}$ et $f(x) = 0$ si $x \notin \mathbb{N}$. En effet, en utilisant l'expression de intégrales pour les mesures discrètes, on a alors pour $A \in \mathcal{P}(\mathbb{R})$:

$$\int \mathbb{1}_A f d\mu = \sum_{n=0}^{+\infty} \mathbb{1}_A(n) f(n) = \sum_{n=0}^{+\infty} \mathbb{1}_A(n) p_n = \nu(A).$$

Proposition 19 Soient (E, \mathcal{E}, μ) un espace mesuré et $f : E \rightarrow \mathbb{R}_+$ une fonction mesurable. Posons $\nu = f \cdot \mu$. Soit enfin $g : E \rightarrow \mathbb{R}$ une fonction mesurable.

1. Si g prend ses valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$, alors

$$\int g d\nu = \int f g d\mu.$$

2. La fonction g est intégrable par rapport à ν si et seulement si $f g$ est intégrable par rapport à μ . Dans ce cas, on a encore

$$\int g d\nu = \int f g d\mu.$$

Preuve.

1. L'égalité est vérifiée pour une fonction étagée $g = \sum_{i=1}^N y_i \mathbb{1}_{A_i}$ positive. En effet, on a

$$\int g d\nu = \sum_{i=1}^N y_i \nu(A_i) = \sum_{i=1}^N y_i \int \mathbb{1}_{A_i} f d\mu = \int g f d\mu,$$

en utilisant la linéarité de l'intégrale.

Soit maintenant g une fonction mesurable à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$ et $(g_n)_n$ une suite croissante de fonctions étagées positives convergente point par point vers g . Alors $(g_n f)_n$ est une suite croissante de fonctions mesurables positives convergente point par point vers la fonction $g f$. Le théorème de la convergence monotone et la définition de l'intégrale entraîne alors

$$\int g d\nu = \lim_{n \rightarrow +\infty} \int g_n d\nu = \lim_{n \rightarrow +\infty} \int g_n f d\mu = \int g f d\mu.$$

2. Si maintenant g prend des valeurs quelconques, on sait que g est intégrable par rapport à ν si et seulement si $\int |g| d\nu < +\infty$. Or d'après le premier point $\int |g| d\nu = \int |g f| d\mu$. On a donc bien la condition nécessaire et suffisante annoncée pour l'intégrabilité. En appliquant les formules du point 1. à g^+ et g^- , on a bien la formule annoncée pour l'intégrale. \square

Exemple. Sur l'espace probabilisé $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ avec $\mathbb{P} = f \cdot \lambda$ ($f : \mathbb{R} \rightarrow \mathbb{R}_+$ étant une fonction mesurable telle que $\int f(x)dx = 1$), toute variable aléatoire X positive vérifie $\mathbb{E}(X) = \int X(x)f(x)dx$.

Passons maintenant au théorème de transfert dans un cadre général. Considérons un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et une variable aléatoire $X : \Omega \rightarrow \mathbb{R}^d$. Si on considère l'univers des valeurs possibles pour X , on a un autre espace probabilisé : $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X)$. Le théorème de transfert permet de calculer l'espérance d'une variable aléatoire fonction de X uniquement à partir de la loi de X .

Théorème 9 (théorème de transfert) Soit $g : \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$ une fonction mesurable.

1. Si g prend ses valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$, alors on a la formule

$$\mathbb{E}(g(X)) = \int g(x)d\mathbb{P}_X(x). \quad (3.5)$$

2. Dans le cas général, la variable aléatoire $Y = g(X)$ est intégrable si et seulement si g est \mathbb{P}_X intégrable. Dans ce cas, on a encore la formule (3.5).

Cas particuliers

- Lorsque $\mathbb{P}_X = f \cdot \lambda_d$ (on dit que la loi de X est à densité), on a pour toute fonction mesurable positive g :

$$\mathbb{E}(g(X)) = \int g(x)f(x)dx.$$

Cette égalité est due au théorème de transfert et à la Proposition 19. La même égalité vaut aussi lorsque $\int |g(x)|f(x)dx < +\infty$.

- Lorsque $\mathbb{P}_X = \sum_{n=0}^{+\infty} p_n \delta_{x_n}$ est une mesure de probabilité discrète (et donc que X est une variable aléatoire discrète) on retrouve le théorème de transfert déjà énoncé lors du Chapitre 2.

Preuve du théorème. Comme auparavant, on prouve ce résultat en suivant les étapes de la construction de l'intégrale.

1. Si $g = \sum_{i=1}^N y_i \mathbb{1}_{A_i}$ est une fonction étagée, on a

$$g(X) = \sum_{i=1}^N y_i \mathbb{1}_{X \in A_i}.$$

Par définition, on a

$$\mathbb{E}(g(X)) = \sum_{i=1}^N y_i \mathbb{P}(X \in A_i) = \sum_{i=1}^N y_i \int \mathbb{1}_{A_i}(x)d\mathbb{P}_X(x).$$

En utilisant la linéarité de l'intégrale, on a bien $\mathbb{E}(g(X)) = \int g(x)d\mathbb{P}_X(x)$.

Si maintenant $(g_n)_n$ est une suite croissante de fonctions étagées positives convergente point par point vers g alors $(g_n(X))_n$ est aussi une suite croissante de fonctions étagées positives convergente point par point vers $g(X)$. On a alors

$$\mathbb{E}(g(X)) = \lim_{n \rightarrow +\infty} \mathbb{E}(g_n(X)) = \lim_{n \rightarrow +\infty} \int g_n(x)d\mathbb{P}_X(x) = \int g(x)d\mathbb{P}_X(x).$$

2. Comme d'après le premier point $\mathbb{E}(|g(X)|) = \int |g(x)|d\mathbb{P}_X(x)$, la variable aléatoire $Y = g(X)$ est intégrable si et seulement si g est \mathbb{P}_X -intégrable. En décomposant $g = g^+ - g^-$, on obtient le résultat en utilisant le point précédent et la linéarité de l'intégrale. \square

3.5 La formule du changement de variables

Théorème 10 Soient A et B deux ensembles ouverts de \mathbb{R}^n et $\phi : A \rightarrow B$ une application continûment différentiable ainsi que sa réciproque et $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction mesurable. Pour $u \in A$, soit $J\phi(u)$ le Jacobien de ϕ au point u , c'est à dire le déterminant de la matrice $\left(\frac{\partial \phi_i}{\partial x_j}(u)\right)_{1 \leq i, j \leq n}$. Alors

$$f \text{ intégrable sur } B \Leftrightarrow u \mapsto f(\phi(u)) |J\phi(u)| \text{ intégrable sur } A.$$

Dans ce cas, on a

$$\int_B f(x) dx = \int_A f(\phi(u)) |J\phi(u)| du.$$

Remarques

- Lorsque $n = 1$, $A = [a, b]$ et $B = [c, d]$, le théorème précédent entraîne

$$\int_c^d f(x) dx = \int_a^b f(\phi(u)) |\phi'(u)| du.$$

Cette formule est bien compatible avec la formule

$$\int_c^d f(x) dx = \int_{\phi^{-1}(c)}^{\phi^{-1}(d)} f(\phi(u)) \phi'(u) du.$$

En effet si l'application ϕ est croissante, il en va de même de ϕ^{-1} et ϕ' est positive ce qui entraîne $\phi^{-1}(c) = a$, $\phi^{-1}(d) = b$ et $|\phi'(u)| = \phi'(u)$ pour tout $u \in A$.

Si maintenant l'application ϕ est décroissante, il en va de même de ϕ^{-1} et ϕ' est négative ce qui entraîne $\phi^{-1}(c) = b$, $\phi^{-1}(d) = a$ et $|\phi'(u)| = -\phi'(u)$ pour tout $u \in A$. Ainsi d'après la convention utilisée sur les bornes de l'intervalle d'intégration, on a

$$\int_a^b f(\phi(u)) |\phi'(u)| du = - \int_a^b f(\phi(u)) \phi'(u) du = \int_b^a f(\phi(u)) \phi'(u) du = \int_{\phi^{-1}(c)}^{\phi^{-1}(d)} f(\phi(u)) \phi'(u) du.$$

- Remarquons aussi que sous réserve des conditions d'intégrabilité, on peut aussi écrire la formule :

$$\int_A f(\phi(u)) du = \int_B f(x) |J\phi^{-1}(x)| dx.$$

On gardera à l'esprit que lorsqu'on change la variable u par la variable $x = \phi(u)$, on a $u = \phi^{-1}(x)$ et il faut remplacer du par $|J\phi^{-1}(x)| dx$ dans l'intégrale de départ. On peut voir en dérivant l'égalité $\phi \circ \phi^{-1}(x) = x$ que $J\phi^{-1}(x) = \frac{1}{J\phi(u)}$.

L'exemple des coordonnées polaires. Posons

$$B = \mathbb{R}^2 \setminus \{(x, 0) : x \leq 0\}, \quad A =]0, +\infty[\times]-\pi, \pi[$$

et soit $\phi : A \rightarrow B$ défini par

$$\phi(\rho, \theta) = (\rho \cos(\theta), \rho \sin(\theta)), \quad (\rho, \theta) \in A.$$

On peut montrer que ϕ vérifie les hypothèses du changement de variables et que

$$\phi^{-1}(x, y) = \left(\sqrt{x^2 + y^2}, 2 \arctan \left(\frac{y}{x + \sqrt{x^2 + y^2}} \right) \right), \quad (x, y) \in B.$$

De plus on peut vérifier que $|J\phi(\rho, \theta)| = \rho$.

Appliquons ce changement de variables pour le calcul de $I = \int e^{-\frac{x^2}{2}} dx$. Pour cela, nous considérons l'intégrale double

$$\int \exp\left(-\frac{x^2 + y^2}{2}\right) d\lambda_2(x, y) = I^2,$$

où l'égalité précédente est justifiée par le théorème de Fubini. On voit facilement que

$$\begin{aligned} \int \exp\left(-\frac{x^2 + y^2}{2}\right) d\lambda_2(x, y) &= \int \mathbb{1}_B(x, y) \exp\left(-\frac{x^2 + y^2}{2}\right) d\lambda_2(x, y) \\ &+ \int \mathbb{1}_{B^c}(x, y) \exp\left(-\frac{x^2 + y^2}{2}\right) d\lambda_2(x, y) \\ &= \int \mathbb{1}_B(x, y) \exp\left(-\frac{x^2 + y^2}{2}\right) d\lambda_2(x, y) \end{aligned}$$

en remarquant la nullité de la deuxième intégrale (appliquer le théorème de Fubini). En effectuant le changement de variables avec les coordonnées polaires, on obtient

$$\int \mathbb{1}_B(x, y) \exp\left(-\frac{x^2 + y^2}{2}\right) d\lambda_2(x, y) = \int_A \rho \exp\left(-\frac{\rho^2}{2}\right) d\rho d\theta = 2\pi,$$

en appliquant de nouveau le théorème de Fubini. On trouve donc $I = \sqrt{2\pi}$ ce qui justifie la définition de la loi gaussienne centrée réduite.

Chapitre 4

Les variables aléatoires à densité

Dans ce chapitre, nous nous replaçons dans le cadre de la théorie des probabilités. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. En utilisant les résultats du chapitre précédent, on conviendra qu'une variable aléatoire $X : \Omega \rightarrow \mathbb{R}^d$ est à densité (sous-entendu par rapport à la mesure de Lebesgue sur \mathbb{R}^d notée λ_d) lorsque sa loi \mathbb{P}_X est une mesure à densité par rapport à λ_d . Ceci signifie qu'il existe une fonction mesurable $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ telle que $\int f(x)d\lambda_d(x) = \int f(x)dx = 1$ et

$$\mathbb{P}(X \in A) = \int_A f(x)dx, \quad A \in \mathcal{B}(\mathbb{R}^d).$$

4.1 Exemples de lois à densité

Voici quelques lois à densité classiques rencontrées en probabilités et en statistique.

Les lois Gaussiennes sur \mathbb{R} . On dit qu'une variable aléatoire X suit une loi gaussienne de paramètres $m \in \mathbb{R}$ et $\sigma > 0$ (on parle aussi de loi normale ou encore de loi de Laplace-Gauss) et on note $X \sim \mathcal{N}(m, \sigma^2)$ lorsque X est à densité par rapport à la mesure de Lebesgue et que la densité est donnée par :

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

On peut remarquer que si $U \sim \mathcal{N}(0, 1)$ (on parle de loi gaussienne centrée réduite) alors la variable aléatoire $m + \sigma U$ suit la loi $\mathcal{N}(m, \sigma^2)$ (pour le montrer, on peut soit utiliser le théorème de transfert et effectuer un changement de variable ou calculer la fonction de répartition, voir la section sur le calcul de lois). Il est facile de vérifier que $\mathbb{E}(X) = m$ (en utilisant les propriétés de symétrie de la densité) et que $\text{Var}(X) = \sigma^2$ (on peut le démontrer d'abord pour la loi gaussienne centrée réduite en effectuant une intégration par parties).

Les propriétés de symétrie de cette loi permettent de modéliser les variations de certaines grandeurs autour d'une valeur donnée (par exemple les erreurs de mesure). Cette loi est importante en statistique du fait de son rôle de loi limite (nous en reparlerons lorsque nous énoncerons le théorème central-limite).

Les lois uniformes. Lorsque $[c, d]$ est un intervalle de \mathbb{R} , la loi uniforme sur $[c, d]$ (déjà rencontrée au Chapitre 1) est la loi dont la densité f est définie par

$$f(x) = \frac{1}{d-c} \mathbf{1}_{[c,d]}(x), \quad x \in \mathbb{R}.$$

On peut en fait définir une loi uniforme sur n'importe quel borélien de mesure de Lebesgue non nulle. Par exemple sur \mathbb{R}^2 , la loi uniforme sur le disque unité possède une densité $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ définie par

$$f(x, y) = \frac{1}{\pi} \mathbb{1}_{x^2+y^2 \leq 1}, \quad (x, y) \in \mathbb{R}^2.$$

Lois gamma. On dit qu'une variable aléatoire X suit une loi gamma de paramètres $k > 0$ et $\theta > 0$ (notation $X \sim \gamma(k, \theta)$), si la densité de X est donnée par

$$f(x) = \frac{x^{k-1} \exp\left(-\frac{x}{\theta}\right)}{\Gamma(k)\theta^k} \mathbb{1}_{x>0},$$

où on a

$$\Gamma(k) = \int_0^{+\infty} x^{k-1} \exp(-x) dx.$$

On a alors $\mathbb{E}(X) = k\theta$ et $\text{Var}(X) = k\theta^2$. On peut remarquer que pour $k = 1$ on a une loi exponentielle de paramètre $\frac{1}{\theta}$. Lorsque k est entier, on peut montrer que la loi $\gamma(k, \theta)$ correspond également à la loi de la somme de k variables aléatoires indépendantes toutes de loi exponentielle de paramètre $\frac{1}{\theta}$ (voir plus loin lorsque nous étudierons la convolution). La loi du χ^2 , très utilisée en statistique est un cas particulier de la loi gamma (nous reparlerons de la loi du χ^2 lors du chapitre sur les vecteurs gaussiens).

Lois de Weibull. On dit qu'une variable aléatoire X suit une loi de Weibull de paramètres α, β, γ si la densité de X est donnée par

$$f(x) = \frac{\beta}{\alpha} \left(\frac{x-\gamma}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{x-\gamma}{\alpha}\right)^\beta\right) \mathbb{1}_{x>\gamma}.$$

Cette loi est fréquemment utilisée en fiabilité pour modéliser le taux de défaillance d'un matériel (voir le TD pour des précisions). Lorsque $\gamma = 0$ et $\beta = 1$, on retrouve la loi exponentielle de paramètre 1. Plus généralement, on peut montrer que $Y = \left(\frac{X-\gamma}{\alpha}\right)^\beta$ suit une loi exponentielle de paramètre 1.

Loi de Cauchy. La densité f est définie par

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

Cette loi a la particularité de ne pas avoir de moyenne, car $\int |x|f(x)dx = +\infty$. On peut montrer que la loi de Cauchy est la loi du quotient de deux variables aléatoires gaussiennes, centrées réduites et indépendantes.

4.2 Densités marginales. Indépendance

Définition 17 Soit $Z = (X, Y)$ une variable aléatoire à valeurs dans \mathbb{R}^2 et dont la loi possède une densité f par rapport à la mesure de Lebesgue sur \mathbb{R}^2 . Soient $f_1 : \mathbb{R} \rightarrow \mathbb{R}_+$ et $f_2 : \mathbb{R} \rightarrow \mathbb{R}_+$ les fonctions définies par

$$f_1(x) = \int f(x, y) dy, \quad x \in \mathbb{R},$$

$$f_2(y) = \int f(x, y) dx, \quad y \in \mathbb{R}.$$

Alors les fonctions f_1 et f_2 sont des densités de probabilité appelées les densités marginales de Z et correspondent aux densités des mesures \mathbb{P}_X et \mathbb{P}_Y respectivement.

La définition précédente se généralisent au cas d'une variable aléatoire Z à valeurs dans \mathbb{R}^d . Dans ce cas, la i -ième densité marginale est simplement la fonction obtenue en intégrant la densité de Z par rapport aux $d - 1$ variables d'indices $j \neq i$. On notera que deux variables aléatoires réelles peuvent avoir une densité sans que ce soit le cas pour le couple : par exemple si $Y = X$ alors le couple $Z = (X, X)$ ne peut avoir de densité car la diagonale de \mathbb{R}^2 , $D = \{(x, y) : x = y\}$, est de mesure nulle (pour λ_2).

Proposition 20 Soient des variables aléatoires X_1, \dots, X_n à valeurs respectives dans $\mathbb{R}^{d_1}, \dots, \mathbb{R}^{d_n}$ et admettant des densités notées respectivement f_1, \dots, f_d . Alors X_1, \dots, X_d sont indépendantes si et seulement si la variable aléatoire $Z = (X_1, \dots, X_n)$ (à valeurs dans \mathbb{R}^d , $d = d_1 + \dots + d_n$) admet une loi à densité f par rapport λ_d , donnée par

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n), \quad (x_1, \dots, x_n) \in \mathbb{R}^d.$$

Preuve. Supposons d'abord les variables aléatoires X_1, \dots, X_n indépendantes. Alors si $A_i \in \mathcal{B}(\mathbb{R}^{d_i})$ pour $1 \leq i \leq n$, on a

$$\begin{aligned} \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) &= \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n) \\ &= \int_{A_1} f(x_1) dx_1 \cdots \int_{A_n} f(x_n) dx_n \\ &= \int_{A_1 \times \dots \times A_n} f_1(x_1) \cdots f_n(x_n) dx_1 \cdots dx_n, \end{aligned}$$

la dernière égalité résultant du théorème de Fubini. D'après l'unicité de la mesure produit, on en déduit que la loi de Z vérifie

$$\mathbb{P}(Z \in A) = \int_A \prod_{i=1}^n f_i(x_i) d\lambda_d(x_1, \dots, x_n), \quad A \in \mathcal{B}(\mathbb{R}^d).$$

Inversement, supposons que la densité de Z soit donnée par le produit des densités f_1, \dots, f_n . Alors l'application du théorème de Fubini entraîne automatiquement l'indépendance des variables X_1, \dots, X_n (reprendre l'enchaînement des égalités ci-dessus mais en sens inverse). \square

Remarque. Supposons que la densité f d'un couple (X, Y) s'écrivent sous la forme d'un produit, c'est à dire

$$f(x, y) = g(x)h(y), \quad (x, y) \in \mathbb{R}^2,$$

pour deux fonctions g et h à valeurs positives (mais sans savoir à priori qu'il s'agit de densités de probabilité). Alors le théorème de Fubini permet d'affirmer que les fonctions g et h sont, à une constante près, les densités de X et de Y respectivement. Par exemple, la densité de X est $x \rightarrow \int h(y) dy \times g(x)$. De plus les variables aléatoires X et Y sont indépendantes. Cette remarque se généralise à un nombre $n \geq 2$ de variables aléatoires. On retiendra qu'il suffit d'arriver à séparer les variables dans la densité jointe pour conclure à l'indépendance.

Proposition 21 Soient X et Y deux variables aléatoires indépendantes à valeurs dans \mathbb{R}^d et dont les lois admettent des densités f_X et f_Y respectivement. Alors la variable aléatoire $Z = X + Y$ admet une densité $f_Z : \mathbb{R}^d \rightarrow \mathbb{R}_+$ donnée par

$$f_Z(z) = \int f_X(z - y)f_Y(y)dy = \int f_Y(z - x)f_X(x)dx, \quad z \in \mathbb{R}^d.$$

On dit que f_Z est le produit de convolution de f_X et de f_Y .

Preuve. Soit $h = \mathbf{1}_A$, $A \in \mathcal{B}(\mathbb{R}^d)$. Nous avons

$$\begin{aligned} \mathbb{P}(Z \in A) &= \mathbb{E}(h(X + Y)) \\ &= \int h(x + y) f_X(x) f_Y(y) d\lambda_{2d}(x, y) \\ &= \int \left(\int h(x + y) f_X(x) d\lambda_d(x) \right) f_Y(y) d\lambda_d(y) \\ &= \int \left(\int h(z) f_X(z - y) d\lambda_d(z) \right) f_Y(y) d\lambda_d(y) \\ &= \int h(z) \left(\int f_X(z - y) f_Y(y) d\lambda_d(y) \right) d\lambda_d(z). \end{aligned}$$

La deuxième égalité résulte du théorème de transfert et la troisième du théorème de Fubini. La quatrième égalité est obtenue en effectuant le changement de variables $z = x + y$ (de Jacobien 1). La dernière égalité est obtenue en utilisant le théorème de Fubini. \square

Remarque. Si on ne suppose pas l'indépendance de X et de Y , alors la somme Z n'a pas nécessairement une densité (c'est le cas si $Y = -X$ par exemple).

Exemple de calcul. Il est utile de savoir refaire la démonstration précédente car celle-ci permet également de trouver des densités de variables du type $Z = X - Y$ ou encore $Z = XY$ lorsque les variables X et Y sont indépendantes. Appliquons directement la formule obtenue lorsque les variables aléatoires X et Y suivent toutes les deux la loi exponentielle de paramètre $\lambda > 0$. Dans ce cas

$$\begin{aligned} f_Z(z) &= \int \mathbf{1}_{\mathbb{R}_+}(z - y) \mathbf{1}_{\mathbb{R}_+}(y) \lambda^2 \exp(-\lambda z) dy \\ &= \begin{cases} \int_0^z \lambda^2 \exp(-\lambda z) dy = \lambda^2 z \exp(-\lambda z), & \text{si } z \geq 0 \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

Ainsi on trouve $f_Z(z) = \mathbf{1}_{\mathbb{R}_+}(z) \lambda^2 z \exp(-\lambda z)$ pour $z \in \mathbb{R}$. Z suit donc une loi $\gamma(2, \lambda^{-1})$. On peut aussi montrer la somme de deux variables aléatoires indépendantes et de lois respectives $\gamma(a, \theta)$ et $\gamma(b, \theta)$ suit une loi $\gamma(a + b, \theta)$. On voit alors immédiatement par récurrence que la loi d'une somme de n variables aléatoires indépendantes toutes de loi exponentielle de paramètre λ est une loi $\gamma(n, \lambda^{-1})$.

Exemple de calcul : la densité des statistiques d'ordre. Soient X_1, \dots, X_n des variables aléatoires i.i.d toutes de loi à densité f sur \mathbb{R} . Alors on peut montrer que pour $i \neq j$ ($1 \leq i, j \leq n$), $\mathbb{P}(X_i = X_j) = 0$. Pour montrer ceci, on commence par remarquer que le couple (X_i, X_j) est une variable aléatoire à densité qui est donnée par $(x, y) \mapsto f(x)f(y)$. De plus on a l'égalité $\mathbb{P}(X_i = X_j) = \mathbb{E}(\mathbf{1}_D(X_i, X_j))$ où $D = \{(x, y) \in \mathbb{R}^2 : x = y\}$. Avec l'abus de notation déjà utilisé auparavant, on pourrait aussi écrire $\mathbf{1}_D = \mathbf{1}_{x=y}$. Appliquons maintenant le théorème de transfert. On a

$$\mathbb{E}(\mathbf{1}_D(X_i, X_j)) = \int \mathbf{1}_D(x, y) f(x) f(y) dx dy = \int f(y) \left(\int_y^y f(x) dx \right) dy,$$

la deuxième égalité résultant du théorème de Fubini. Comme pour tout y , $\int_y^y f(x) dx = 0$, on obtient bien $\mathbb{P}(X_i = X_j) = 0$. On peut même affirmer que

$$\mathbb{P}(\cup_{i \neq j} \{X_i = X_j\}) \leq \sum_{i \neq j} \mathbb{P}(X_i = X_j) = 0.$$

Posons $B = \cup_{i \neq j} \{X_i = X_j\}$. Pour $\omega \in B^c$, les valeurs $X_1(\omega), \dots, X_n(\omega)$ sont toutes distinctes. Pour tout $\omega \in B$, soit alors $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ une permutation telle que $X_{\sigma(1)}(\omega) < \dots < X_{\sigma(n)}(\omega)$ (σ dépend de ω) et posons

$$X_{(i)}(\omega) = X_{\sigma(i)}(\omega), \quad 1 \leq i \leq n.$$

Si $\omega \in B$, on pose $X_{(i)}(\omega) = 0$ pour $1 \leq i \leq n$. Nous allons déterminer la densité du vecteur aléatoire $(X_{(1)}, \dots, X_{(n)})$. Remarquons au passage que pour $\omega \in B^c$, on a $X_{(1)}(\omega) = \min_{1 \leq i \leq n} X_i(\omega)$ et $X_{(n)}(\omega) = \max_{1 \leq i \leq n} X_i(\omega)$. Soit $A \in \mathcal{B}(\mathbb{R}^n)$. Posons $X^{(o)} = (X_{(1)}, \dots, X_{(n)})$ et $X = (X_1, \dots, X_n)$. On a alors

$$\begin{aligned} \mathbb{P}(X^{(o)} \in A) &= \mathbb{P}(\{X^{(o)} \in A\} \cap B^c) \\ &= \sum_{\sigma \in \mathcal{S}_n} \mathbb{P}(\{X_{\sigma(1)} < \dots < X_{\sigma(n)}\} \cap \{(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in A\}) \\ &= n! \int_A \mathbb{1}_{x_1 < \dots < x_n} f(x_1) \cdots f(x_n) d\lambda_n(x_1, \dots, x_n). \end{aligned}$$

Commentons ces égalités. La première égalité est due au fait que $\mathbb{P}(C) = \mathbb{P}(C \cap B^c)$ pour tout $C \in \mathcal{A}$ car $\mathbb{P}(B^c) = 1$. La deuxième égalité est due au fait que

$$B^c = \cup_{\sigma \in \mathcal{S}_n} \{X_{\sigma(1)} < \dots < X_{\sigma(n)}\},$$

où on a noté \mathcal{S}_n comme étant l'ensemble de toutes les permutations de $\{1, \dots, n\}$. De plus cette réunion est disjointe. La dernière égalité est due au théorème de transfert et au fait que pour tout $\sigma \in \mathcal{S}_n$ donné, le n -uplet $(X_{\sigma(1)}, \dots, X_{\sigma(n)})$ a la même loi que X .

Finalement, on voit que $X^{(o)}$ est à densité. Cette densité est donnée par

$$(x_1, \dots, x_n) \mapsto n! \mathbb{1}_{x_1 < \dots < x_n} f(x_1) \cdots f(x_n).$$

4.3 Les calculs de lois en pratique.

Dans cette section, nous donnons quelques exemples de calculs de lois dans diverses situations.

- **Variable discrète.** Pour connaître la loi, on calcule la probabilité de chaque valeur possible. Prenons l'exemple de $Y = [X]$ (partie entière de X) lorsque X suit une loi exponentielle de paramètre λ . L'ensemble des valeurs possibles pour Y est \mathbb{N} . Si $n \in \mathbb{N}$, on a

$$\mathbb{P}(Y = n) = \mathbb{P}(n \leq X < n + 1) = \int_n^{n+1} \lambda \exp(-\lambda x) dx = \exp(-\lambda n) - \exp(-\lambda(n + 1)).$$

Ainsi Y suit une loi géométrique de paramètre $p = 1 - \exp(-\lambda)$.

- **Variable réelle à densité.** Si X est une variable aléatoire réelle à densité et que g est une fonction bijective suffisamment régulière, la loi de $Y = g(X)$ sera une loi à densité que l'on peut déterminer en utilisant un changement de variables dans l'intégrale. Prenons un exemple. On a mentionné sans le prouver qu'une variable aléatoire Y suit une loi de Weibull de paramètres α, β, γ si $Y = \alpha X^{1/\beta} + \gamma = g(X)$ où X suit une loi exponentielle de paramètre 1. Calculons alors la densité de Y . Pour cela, on pose $h = \mathbb{1}_A$ pour un borélien $A \in \mathcal{B}(\mathbb{R})$. Le théorème de transfert donne

$$\mathbb{E}(h(Y)) = \mathbb{E}(h \circ g(X)) = \int h(g(x)) \mathbb{1}_{x>0} \exp(-x) dx.$$

Effectuons ensuite le changement de variables $y = g(x)$ (g est continûment dérivable ainsi que sa réciproque de $]0, \infty[$ sur $] \gamma, +\infty[$) ce qui donne $x = g^{-1}(y) = \left(\frac{y-\gamma}{\alpha}\right)^\beta$. On obtient alors

$$\mathbb{E}(h(Y)) = \int h(y) \mathbb{1}_{y>\gamma} \exp\left(-\left(\frac{y-\gamma}{\alpha}\right)^\beta\right) \frac{\beta}{\alpha} \left(\frac{y-\gamma}{\alpha}\right)^{\beta-1} dy.$$

On retrouve l'expression de la densité d'une loi de Weibull.

Il est également possible de retrouver ce résultat en calculant la fonction de répartition $F_Y : \mathbb{R} \rightarrow [0, 1]$ définie par $F_Y(y) = \mathbb{P}(Y \leq y)$ à partir de la fonction de répartition de X . Ici, nous avons pour $y > \gamma$,

$$F_Y(y) = \mathbb{P}\left(\alpha X^{1/\beta} + \gamma \leq y\right) = \mathbb{P}\left(X \leq \left(\frac{y-\gamma}{\alpha}\right)^\beta\right) = 1 - \exp\left(-\left(\frac{y-\gamma}{\alpha}\right)^\beta\right).$$

Ainsi, si $y > \gamma$, F_Y est dérivable en y et on a

$$F'_Y(y) = \exp\left(-\left(\frac{y-\gamma}{\alpha}\right)^\beta\right) \frac{\beta}{\alpha} \left(\frac{y-\gamma}{\alpha}\right)^{\beta-1}.$$

De plus si $y < \gamma$, on a $F_Y(y) = 0$ donc $F'_Y(y) = 0$. Lorsque une fonction de répartition F est **continue**, continûment dérivable sur un intervalle I ouvert et nulle à gauche de I (si la borne inférieure de I est $> -\infty$) et valant 1 à droite de I (si la borne supérieure de I est $< +\infty$), on a automatiquement $F(z) = \int_{-\infty}^z F'(u) du$ et F' est une densité de la loi. On retrouve donc bien le résultat précédent.

- **Calcul d'espérance.** On utilise le théorème de transfert et le théorème de Fubini (lorsque plusieurs variables sont en jeu). Par exemple, si X est une variable aléatoire admettant une densité f_X et h une fonction mesurable à valeurs réelles, on a (sous les hypothèses du théorème de transfert)

$$\mathbb{E}(h(X)) = \int h(x) f_X(x) dx.$$

- **Image d'un couple de variables aléatoires à densité par une application.** Lorsque l'application est bijective, on peut utiliser la formule du changement de variables. Voici un exemple de calcul. Soit (X, Y) un couple de variables aléatoires suivant la loi uniforme sur le disque unité de \mathbb{R}^2 , c'est à dire de densité définie par

$$f_{X,Y}(x, y) = \frac{1}{\pi} \mathbb{1}_{x^2+y^2 < 1}.$$

Soit R et Θ le rayon et l'angle associés au point (X, Y) (coordonnées polaires du point (X, Y)). On a

$$(X, Y) = (R \cos(\Theta), R \sin(\Theta)) = \phi(R, \Theta).$$

Cherchons la loi du couple (R, Θ) . Pour une fonction $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable bornée, on a en utilisant le théorème de transfert

$$\begin{aligned} \mathbb{E}(h(R, \Theta)) &= \mathbb{E}(h \circ \phi^{-1}(X, Y)) \\ &= \int h \circ \phi^{-1}(x, y) \frac{1}{\pi} \mathbb{1}_{x^2+y^2 < 1} dx dy \\ &= \int_B h \circ \phi^{-1}(x, y) \frac{1}{\pi} \mathbb{1}_{x^2+y^2 < 1} dx dy \end{aligned}$$

où $B = \mathbb{R}^2 \setminus (\mathbb{R}_- \times \{0\})$. On pose $(r, \theta) = \phi^{-1}(x, y)$ ce qui donne $(x, y) = \phi(r, \theta)$. On a déjà évoqué au chapitre précédent que $\phi : A \rightarrow B$ était continûment différentiable que sa

réciproque avec $A = \mathbb{R}_+^* \times]-\pi, \pi[$ et $J\phi(r, \theta) = r$. On remplace alors (x, y) par (r, θ) dans l'intégrale, le domaine d'intégration B par A et $dx dy$ par $|J\phi(r, \theta)| dr d\theta$. On obtient

$$\mathbb{E}(h(R, \Theta)) = \int_A h(r, \theta) \mathbf{1}_{r^2 < 1} r dr d\theta = \int h(r, \theta) r \mathbf{1}_{0 < r < 1} \mathbf{1}_{-\pi < \theta < \pi} dr d\theta.$$

On voit alors que R et Θ sont des variables aléatoires indépendantes et en renormalisant, on voit que $f_\Theta = \frac{1}{2\pi} \mathbf{1}_{]-\pi, \pi[}$ ce qui signifie que Θ suit la loi uniforme sur l'intervalle $[-\pi, \pi]$. De plus la densité de R est donnée par $f_R(r) = 2r \mathbf{1}_{0 < r < 1}$. Ces densités marginales pouvaient être retrouvées directement en utilisant la définition intuitive de la loi uniforme. En effet la probabilité d'un borélien A inclus dans le disque est $\frac{\lambda_2(A)}{\pi}$. Ainsi pour $0 < r < 1$, la probabilité $\mathbb{P}(R \leq r) = r^2$ (aire du disque de centre 0 et de rayon r) correspond bien à la fonction de répartition au point r de la loi trouvée. De plus si $\theta \in]-\pi, \pi[$, on a $\mathbb{P}(\Theta \leq \theta) = \frac{\theta + \pi}{2\pi}$ (aire du secteur angulaire entre $-\pi$ et θ) ce qui correspond bien à la fonction de répartition de la loi uniforme sur l'intervalle $[-\pi, \pi]$.

- **Loi d'une fonction d'une variable à densité et d'une variable discrète.** Pour calculer la loi d'une variable aléatoire $Z = F(X, Y)$ à valeurs dans \mathbb{R}^d où X est discrète et Y est une variable à densité indépendante de X , on se donne une fonction h mesurable bornée arbitraire (par exemple une indicatrice) et on peut évaluer l'espérance de $h(Z)$ de la façon suivante.

$$\begin{aligned} \mathbb{E}(h(Z)) &= \mathbb{E}\left(\sum_{x \in \text{val}(X)} h(F(x, Y)) \mathbf{1}_{X=x}\right) \\ &= \sum_{x \in \text{val}(X)} \mathbb{E}(h(F(x, Y)) \mathbf{1}_{X=x}) \\ &= \sum_{x \in \text{val}(X)} \mathbb{E}(h(F(x, Y))) \mathbb{P}(X = x). \end{aligned}$$

La deuxième égalité se justifie à l'aide du théorème de Fubini car h est bornée et la troisième égalité est due à l'indépendance de X et de Y . On poursuit alors le calcul de la loi de $F(x, Y)$ en utilisant éventuellement un changement de variable.

Supposons par exemple que $Z = (1-X)Y_1 + XY_2$ où $Y = (Y_1, Y_2)$ est un couple de variables aléatoires réelles indépendantes de $X \sim \mathcal{B}(p)$ et tel que $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. En notant f_1 (resp. f_2) la densité de Y_1 (resp. Y_2), on a alors

$$\begin{aligned} \mathbb{E}(h(Z)) &= (1-p)\mathbb{E}(h(Y_1)) + p\mathbb{E}(h(Y_2)) \\ &= (1-p) \int h(z) f_1(z) dz + p \int h(z) f_2(z) dz \\ &= \int h(z) ((1-p)f_1(z) + pf_2(z)) dz. \end{aligned}$$

La loi de Z est donc une loi à densité f_Z donnée par un mélange des deux densités f_1 et f_2 (voir Figure 4.1). Remarquons que si $f_1 = f_2$ alors on a $f_Z = f_1$ et $Z \sim \mathcal{N}(m_1, \sigma_1^2)$.

On peut aussi obtenir des lois mixtes : par exemple la variable aléatoire $X = \epsilon Y$ où Y est une variable aléatoire de densité f sur \mathbb{R} , indépendante de $\epsilon \sim \mathcal{B}(p)$, vérifie pour une fonction h mesurable bornée,

$$\begin{aligned} \mathbb{E}(h(X)) &= \mathbb{E}(h(0)\mathbf{1}_{\epsilon=0}) + \mathbb{E}(h(Y)\mathbf{1}_{\epsilon=1}) \\ &= h(0)(1-p) + p \int h(y) f(y) dy \\ &= \int h(y) ((1-p)d\delta_0(y) + pf(y)dy). \end{aligned}$$

On a donc $\mathbb{P}_X = (1-p)d\delta_0 + pf \cdot \lambda$.

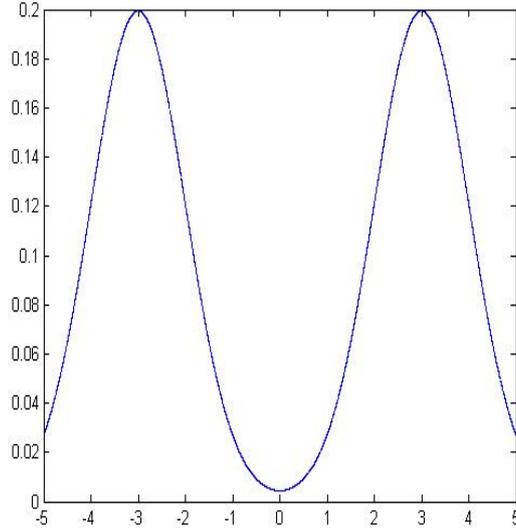


FIGURE 4.1: Densité mélange de deux densités gaussienne de moyennes respectives -3 et 3 et de variance 1

4.4 Densités conditionnelles.

Définition 18 Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{R}^d \times \mathbb{R}^k$ et dont la loi admet la densité $f_{X,Y}$. Soit $y \in \mathbb{R}^k$ tel que $f_Y(y) \neq 0$. On appelle loi conditionnelle de X sachant $Y = y$ (et on note loi de $X|Y = y$) la loi de probabilité qui possède la densité notée $f_{X|Y}(\cdot|y)$ et qui est définie par

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad x \in \mathbb{R}^d.$$

On dit que $f_{X|Y}(\cdot|y)$ est la densité conditionnelle de X sachant Y . Lorsque $y \in \mathbb{R}^k$ vérifie $f_Y(y) = 0$, on convient de définir la loi de $X|Y = y$ par une mesure de probabilité de densité arbitraire g sur \mathbb{R}^d et on pose $f_{X|Y}(\cdot|y) = g$.

On définit de façon analogue la loi de $Y|X = x$ pour $x \in \mathbb{R}^d$. On peut motiver la définition précédente à partir des probabilités conditionnelles entre événements. Pour simplifier, supposons $d = k = 1$. On ne peut pas définir $\mathbb{P}(X \in A|Y = y)$ à partir du quotient $\frac{\mathbb{P}(X \in A, Y = y)}{\mathbb{P}(Y = y)}$, le numérateur et le dénominateur étant tous deux nuls. Un bon moyen pour mesurer l'influence d'une valeur y de Y sur celles de X consiste à regarder la limite des probabilités conditionnelles

$$\mathbb{P}(X \in A|y - \epsilon < Y < y + \epsilon) = \frac{\mathbb{P}(X \in A, y - \epsilon < Y < y + \epsilon)}{\mathbb{P}(y - \epsilon < Y < y + \epsilon)}, \quad (4.1)$$

lorsque $\epsilon \rightarrow 0$. Si la densité est non nulle dans un voisinage arbitrairement petit de y , ces quotients sont bien définis. On a alors

$$\begin{aligned} \mathbb{P}(X \in A|y - \epsilon < Y < y + \epsilon) &= \frac{\int_{y-\epsilon}^{y+\epsilon} (\int_A f_{X,Y}(x, v) dx) dv}{\int_{y-\epsilon}^{y+\epsilon} f_Y(v) dv} \\ &= \frac{G(y + \epsilon) - G(y - \epsilon)}{F(x + \epsilon) - F(x - \epsilon)}, \end{aligned}$$

où on a posé $G(z) = \int_{-\infty}^z (\int_A f_{X,Y}(x, v) dx) dv$ et où F désigne la fonction de répartition de X . En divisant par 2ϵ , on a sous réserve de dérivabilité

$$\lim_{\epsilon \rightarrow 0} \mathbb{P}(X \in A | y - \epsilon < Y < y + \epsilon) = \frac{G'(y)}{F'(y)} = \frac{\int_A f_{X,Y}(x, y) dx}{f_Y(y)}.$$

On obtient donc l'intégrale sur A de la densité conditionnelle définie précédemment.

Notations. Vu la remarque précédente, on notera $\mathbb{P}(X \in A | Y = y) = \int_A f_{X|Y}(x|y) dx$. La fonction $x \mapsto \mathbb{P}(X \leq x | Y = y)$ est appelée fonction de répartition conditionnelle de $X | Y = y$. De plus, lorsque elle est bien définie, on notera $\mathbb{E}(X | Y = y) = \int x f_{X|Y}(x|y) dx$ la moyenne de la loi conditionnelle de $X | Y = y$.

Proposition 22 *Si X une variable aléatoire intégrable et à valeurs réelles alors la moyenne conditionnelle de la loi de $X | Y = y$ est bien définie, et on a*

$$\mathbb{E}(X) = \int \mathbb{E}(X | Y = y) f_Y(y) dy.$$

Preuve. Si X est intégrable, on a en utilisant le théorème de Fubini

$$\begin{aligned} \mathbb{E}(X) &= \int x f_X(x) dx \\ &= \int x f_{X,Y}(x, y) dx dy \\ &= \int x f_{X|Y}(x|y) f_Y(y) dx dy \\ &= \int \mathbb{E}(X | Y = y) f_Y(y) dy. \square \end{aligned}$$

Relation avec l'indépendance. Lorsque X et Y sont indépendantes, en prenant $f_{X,Y}(x, y) = f_X(x) f_Y(y)$, on voit que la densité conditionnelle de la loi de $X | Y = y$ ne dépend pas de y .

Exemple de calcul. Soient X et Y deux variables aléatoires indépendantes toutes deux de loi exponentielle de paramètre $\lambda > 0$. Posons $Z = X + Y$ et calculons une densité de la loi conditionnelles de $X | S = s$. Cherchons d'abord une densité pour le couple (X, S) . Pour cela, on se donne une fonction $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable bornée (du type $h = \mathbf{1}_A$ pour $A \in \mathcal{B}(\mathbb{R}^2)$). On a alors

$$\begin{aligned} \mathbb{E}(h(X, S)) &= \mathbb{E}(h(X, X + Y)) \\ &= \int h(x, x + y) \lambda^2 \exp(-\lambda(x + y)) \mathbf{1}_{\mathbb{R}_+}(x) \mathbf{1}_{\mathbb{R}_+}(y) dx dy \\ &= \int \left(\int h(x, x + y) \lambda^2 \exp(-\lambda(x + y)) \mathbf{1}_{\mathbb{R}_+}(y) dy \right) \mathbf{1}_{\mathbb{R}_+}(x) dx \\ &= \int \left(\int h(x, s) \lambda^2 \exp(-\lambda s) \mathbf{1}_{s > x} ds \right) \mathbf{1}_{\mathbb{R}_+}(x) dx. \end{aligned}$$

La troisième égalité est due à un changement de variables ($y \leftrightarrow s = x + y$ pour x fixé); on aurait aussi effectué directement un changement de deux variables $(x, y) \leftrightarrow (u, s) = \phi(x, y) = (x, x + y)$

qui donnerait $J\phi^{-1}(u, s) = 1$.

On voit que la densité $f_{X,S}$ du couple (X, S) est donnée par

$$f_{X,S}(x, s) = \lambda^2 \exp(-\lambda s) \mathbf{1}_{s \geq x \geq 0}.$$

Calculons maintenant la densité conditionnelle $f_{X|S}(\cdot|s)$. On a déjà vu dans ce chapitre que

$$f_S(s) = \lambda^2 \exp(-\lambda s) s \mathbf{1}_{\mathbb{R}_+}(s).$$

Ainsi pour $s > 0$ donné, on a $f_{X,S}(x, s) = \lambda^2 \exp(-\lambda s) \mathbf{1}_{[0,s]}(x)$ et $f_S(s) = \lambda^2 \exp(-\lambda s) s$, ce qui donne

$$f_{X|S}(x|s) = \frac{1}{s} \mathbf{1}_{[0,s]}(x).$$

On reconnaît la loi uniforme sur l'intervalle $[0, s]$. En particulier $\mathbb{E}(X|S = s) = \frac{s}{2}$.

Chapitre 5

Les outils analytiques classiques en probabilité

5.1 La fonction de répartition

Pour une variable aléatoire X prenant des valeurs réelles, la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par $F_X(x) = \mathbb{P}(X \leq x)$ est appelée la fonction de répartition de X . Donnons deux exemples de fonction de répartition.

- Lorsque X suit une loi exponentielle de paramètre λ , alors $F_X(x) = (1 - \exp(-\lambda x)) \mathbb{1}_{x>0}$.
- Lorsque X suit une loi de Bernoulli de paramètre p , alors $F_X(x) = (1 - p) \mathbb{1}_{0 \leq x < 1} + \mathbb{1}_{x \geq 1}$.

Le dernier exemple montre qu'une fonction de répartition n'est pas toujours une fonction continue. Quelques-unes des propriétés des fonctions de répartition avaient été mentionnées sans démonstration lors du chapitre Mesure et Intégration.

5.1.1 Propriétés générales des fonctions de répartition

Proposition 23 *Soit F la fonction de répartition d'une variable aléatoire X . On a les propriétés suivantes.*

1. F est une fonction croissante qui vérifie $0 \leq F(x) \leq 1$, $\lim_{x \rightarrow +\infty} F(x) = 1$ et $\lim_{x \rightarrow -\infty} F(x) = 0$.
2. F est continue à droite. De plus $\lim_{y \nearrow x} F(y) = \mathbb{P}(X < x)$. Par conséquent F est continue lorsque pour tout $x \in \mathbb{R}$, le saut de F au point x , $s(x) = \mathbb{P}(X = x)$, est nul.

Preuve.

1. Si $x < y$, on a $\{X \leq x\} \subset \{X \leq y\}$, ce qui entraîne $F(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F(y)$. F est donc bien croissante. Soit ensuite $(x_n)_n$ une suite croissante de nombres réels positifs et de limite $+\infty$. Posons pour $n \in \mathbb{N}$, $A_n =]-\infty, x_n]$. Alors $(A_n)_n$ est une suite croissante d'intervalles dont la réunion est \mathbb{R} . D'après la propriété de continuité supérieure de la mesure \mathbb{P}_X , on a

$$\lim_{n \rightarrow +\infty} F(x_n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X(\mathbb{R}) = 1.$$

On aurait pu tout aussi bien utiliser la continuité supérieure de \mathbb{P} en posant $A_n = \{X \leq x_n\}$. En posant $B_n =]-\infty, -x_n]$ et en utilisant la continuité inférieure de \mathbb{P}_X , on obtient

$$\lim_{n \rightarrow +\infty} F(-x_n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(B_n) = \mathbb{P}_X(\emptyset) = 0.$$

Ceci prouve bien que les limites annoncées pour F .

2. La continuité à droite est une conséquence de la continuité inférieure, car si $(x_n)_n$ est une suite décroissante de limite x , alors

$$\lim_{n \rightarrow +\infty} F(x_n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X (] - \infty, x_n]) = \mathbb{P}_X (] - \infty, x]) = F(x).$$

On a utilisé le fait que $] - \infty, x] = \bigcap_{n=0}^{+\infty}] - \infty, x_n]$.

Si maintenant $(x_n)_n$ est une suite strictement croissante de nombres réels convergente vers x , alors

$$\bigcup_{n=0}^{+\infty}] - \infty, x_n] =] - \infty, x[.$$

La continuité supérieure de \mathbb{P}_X entraîne que

$$\lim_{n \rightarrow +\infty} F(x_n) = \mathbb{P}_X (] - \infty, x[) = \mathbb{P}(X < x).$$

Ainsi le saut $s(x) = \mathbb{P}(X = x)$ de F au point x vaut aussi

$$s(x) = \mathbb{P}(X \leq x) - \mathbb{P}(X < x) = F(x) - \lim_{y \nearrow x} F(y).$$

Ainsi F est continue si et seulement si $s(x) = 0$ pour tout $x \in \mathbb{R}$. \square

On voit donc que la fonction de répartition d'une variable aléatoire à densité est toujours continue. En revanche, la fonction de répartition d'une variable aléatoire X discrète, telle que $\mathbb{P}_X = \sum_{n=0}^{+\infty} p_n \delta_{x_n}$ et où $(x_n)_n$ est une suite strictement croissante de nombres réels, est définie par $F_X(x) = \sum_{n: x_n \leq x} p_n$: F_X est donc constante entre les x_n et le saut au point x est nul sauf si $x = x_n$ auquel cas $s(x_n) = p_n$.

On a déjà évoqué que la fonction de répartition permettait d'obtenir la densité en calculant sa dérivée. Il faut cependant que la variable aléatoire ait une densité (en particulier F doit être continue) : dériver la fonction de répartition d'une variable aléatoire discrète donne une fonction nulle entre les valeurs x_n telles que $\mathbb{P}(X = x_n) > 0$, ce qui ne correspond évidemment pas à une densité de probabilité par rapport à la mesure de Lebesgue. En revanche cette approche peut être utilisée pour toute fonction de répartition F continue, continûment dérivable sur un intervalle $I =]a, b[$ et égale à 1 (resp. 0) sur $[b, +\infty[$ (resp. $] - \infty, a]$). Par exemple, si X suit une loi $\mathcal{N}(0, 1)$, quelle est la loi de X^2 ? En notant Φ la fonction de répartition de X , on a $F_{X^2}(x) = 0$ si $x \leq 0$ et si $x > 0$:

$$F_{X^2}(x) = \mathbb{P}(X^2 \leq x) = \mathbb{P}(-\sqrt{x} \leq X \leq \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = 2\Phi(\sqrt{x}) - 1.$$

En utilisant que $\Phi'(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$, on obtient alors

$$F'_{X^2}(x) = \frac{1}{\sqrt{2\pi x}} \exp(-x/2).$$

Ainsi X^2 suit une loi $\gamma(1/2, 2)$.

5.1.2 Fonction de répartition inverse et simulation

Soit $F : \mathbb{R} \rightarrow [0, 1]$ une fonction de répartition. On définit alors une fonction $G :]0, 1[\rightarrow \mathbb{R}$ par

$$G(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}, \quad t \in]0, 1[.$$

G est appelée la fonction de répartition inverse de F . Voici les propriétés de base de G .

Proposition 24 1. G est une fonction croissante. Lorsque F est strictement croissante et continue alors $G = F^{-1}$.

2. Si F est continue alors on a toujours $F(G(t)) = t$ pour $t \in]0, 1[$.
3. Si F est strictement croissante, alors $G(F(x)) = x$ pour tout $x \in \mathbb{R}$.
4. Dans le cas général, on a toujours l'équivalence

$$F(x) \geq t \Leftrightarrow x \geq G(t).$$

Pour voir que G n'est pas toujours l'inverse de F , on peut observer le comportement de la fonction de répartition dans deux cas pathologiques : lorsque F est constante juste après $G(t)$ (voir Figure 5.1) ou lorsque F est discontinue en x (Figure 5.2).

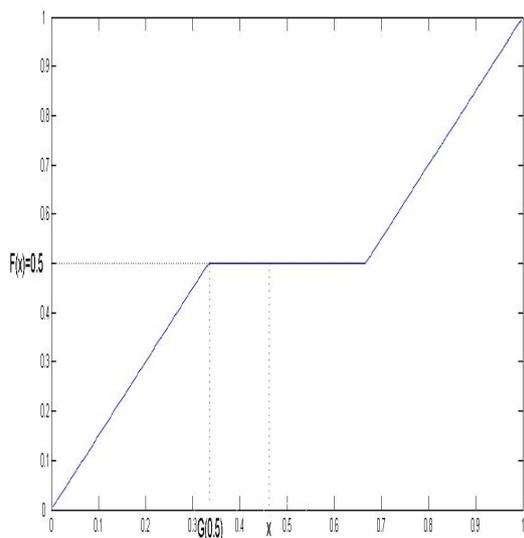


FIGURE 5.1: F constante au voisinage de x

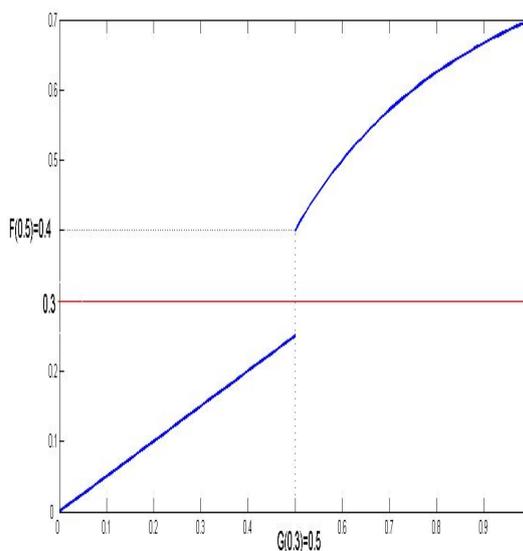


FIGURE 5.2: F discontinue en x

Preuve.

1. Le fait que G soit une fonction croissante est immédiat car si $0 < t_1 < t_2 < 1$,

$$\{x \in \mathbb{R} : F(x) \geq t_2\} \subset \{x \in \mathbb{R} : F(x) \geq t_1\}.$$

La deuxième assertion résulte des points 2. et 3..

2. Remarquons que l'on a toujours $F(G(t)) \geq t$ que F soit continue ou pas car si $(x_n)_n$ est une suite de points strictement décroissante vers $G(t)$, alors $F(x_n) \geq t$ par croissance de F et en utilisant la définition de G , ce qui donne

$$\lim_{n \rightarrow \infty} F(x_n) = F(G(t)) \geq t,$$

par continuité à droite de F .

Ensuite, si $x < G(t)$ alors $F(x) \leq t$. Si F est continue en $G(t)$, on en déduit que $F(G(t)) \leq t$ en faisant tendre x vers $G(t)$. On a donc bien $F(G(t)) = t$.

3. Remarquons que l'on a toujours $G(F(x)) \leq x$ pour tout x tel que $0 < F(x) < 1$. Ceci est dû à la définition de la borne inférieure car x appartient à $\{u \in \mathbb{R} : F(u) \geq F(x)\}$. Supposons de plus F strictement croissante. Alors si $x \in \mathbb{R}$, $F(x) \in]0, 1[$. Si $x_0 < x$ alors $F(x_0) < F(x)$ et donc $x_0 \leq G(F(x))$. En faisant tendre x_0 vers x , nous obtenons $x \leq G(F(x))$. On a donc bien $G(F(x)) = x$.

4. Si $F(x) \geq t$ alors par définition de la borne inférieure, on a $G(t) \leq x$. Si maintenant $G(t) \leq x$, la croissance de F assure que

$$F(x) \geq F(G(t)) \geq t,$$

où la deuxième inégalité a été prouvée lors de la preuve du point 2. \square

La fonction de répartition inverse permet de simuler une variable aléatoire réelle X de loi donnée à partir d'une variable aléatoire de loi uniforme sur $[0, 1]$.

Proposition 25 *Soit U une variable aléatoire de loi uniforme sur $[0, 1]$ et F une fonction de répartition donnée. Alors la variable aléatoire $G(U)$ a pour fonction de répartition F .*

Preuve. D'après la proposition précédente, on a pour $x \in \mathbb{R}$

$$\mathbb{P}(G(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x),$$

ce qui prouve le résultat. \square

Exemple. Pour simuler une variable aléatoire de loi exponentielle de paramètre λ à partir d'une variable aléatoire U de loi uniforme sur $[0, 1]$, on peut remarquer que si $F(x) = 1 - \exp(-\lambda x)$ alors pour $t \in]0, 1[$, on a

$$G(t) = \inf\{x : F(x) = t\} = -\frac{\ln(1-t)}{\lambda}.$$

La variable aléatoire $G(U)$ suit alors la loi exponentielle de paramètre λ . Remarquons que $1 - U$ suit aussi une loi uniforme sur $[0, 1]$ et on peut tout aussi bien considérer la variable aléatoire $-\ln(U)/\lambda$.

La fonction de répartition inverse n'est pas toujours facile à calculer et parfois d'autres méthodes de simulation sont préférables.

Réciproque partielle de la Proposition 25. Soit X une variable aléatoire de fonction de répartition F continue sur \mathbb{R} . Alors la variable aléatoire $F(X)$ suit une loi uniforme sur $[0, 1]$. En effet, on a pour $0 < t < 1$:

$$\mathbb{P}(F(X) \geq t) = \mathbb{P}(X \geq G(t)) = 1 - F(G(t)) = 1 - t.$$

En notant H la fonction de répartition de $F(X)$, on voit alors que

$$1 - H(t) = \mathbb{P}(F(X) > t) = \lim_{n \rightarrow \infty} \mathbb{P}\left(F(X) \geq t + \frac{1}{n}\right) = 1 - t,$$

en utilisant la continuité supérieure de la mesure. Ainsi $H(t) = t$ pour $0 < t < 1$, ce qui prouve que $F(X)$ a la même loi qu'une variable aléatoire de loi uniforme sur $[0, 1]$.

5.2 Covariance et moments

5.2.1 Moments d'une variable aléatoires réelle

Soit X une variable aléatoire à valeurs réelles. Les moments de X sont les nombres $m_k = \mathbb{E}(X^k)$ pour $k \in \mathbb{N}^*$. Le nombre m_k est bien défini si $\mathbb{E}(|X|^k) < +\infty$. Il est possible qu'aucun moment n'existe (c'est par exemple le cas si X suit une loi de Cauchy). D'autres variables

aléatoires peuvent au contraire admettre des moments à tout ordre k ; c'est le cas si X est une variable aléatoire bornée par $M > 0$ car dans ce cas $\mathbb{E}(|X|^k) \leq M^k$.

Si $X \mathcal{N}(0, 1)$, alors X admet également des moments à tout ordre car si $k \in \mathbb{N}$ et on peut même les calculer : les moments d'ordre k impair sont nuls car la fonction à intégrer est impaire, tandis que si $k = 2p$ pour $p \geq 1$, une intégration par parties montre que $m_{2p} = (2p - 1)m_{2p-2}$ ce qui donne au final $m_{2p} = \frac{(2p)!}{2^p p!}$ pour $p \in \mathbb{N}^*$ (en particulier on a $m_4 = 3$).

Voici quelques propriétés importantes concernant les moments.

Proposition 26 Soient X et Y deux variables aléatoires toutes deux à valeurs réelles.

1. Si X admet un moment d'ordre k alors X admet un moment à tout ordre $j \leq k$.
2. Si X et Y admettent un moment d'ordre k , alors $Z = X + Y$ admet également un moment d'ordre k .
3. Si X et Y admettent un moment d'ordre 2, alors XY admet un moment d'ordre 1 et on a l'inégalité suivante appelée inégalité de Cauchy-Schwarz :

$$\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}.$$

4. Si X et Y sont indépendantes et admettent toutes deux un moment d'ordre k , alors le produit XY aussi et on a

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

5. Si X est une variable aléatoire intégrable et $\phi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe telle que $\phi(X)$ soit intégrable, alors on a l'inégalité dite de Jensen :

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

Voici enfin deux inégalités classiques faisant intervenir les moments.

Proposition 27 Soit X une variable aléatoire à valeurs réelles et $k \in \mathbb{N}^*$. Alors pour tout $t > 0$,

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

Lorsque $p = 1$ on parle d'inégalité de Markov. Lorsque $p = 2$ et X est de carré intégrable, on en déduit l'inégalité dite de Tchebychev :

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

Preuve. On a l'inégalité $|X|^p \geq t^p \mathbf{1}_{|X| \geq t}$ et l'inégalité est obtenue en prenant l'espérance de part et d'autre de cette inégalité. L'inégalité de Tchebychev est une conséquence de cette inégalité appliquée à $X - \mathbb{E}(X)$ au lieu de X et pour $p = 2$. \square

5.2.2 Covariance et corrélation

La covariance entre deux variables aléatoires X et Y à valeurs réelles et toutes deux de carré intégrable est défini par

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

Cette expression a bien un sens en vertu de l'inégalité de Cauchy-Schwarz. Remarquons que $\text{Cov}(X, Y) = \text{Cov}(Y, X)$. On a également les propriétés de bilinéarité

$$\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y), \quad \text{Cov}(X, Y_1 + Y_2) = \text{Cov}(X, Y_1) + \text{Cov}(X, Y_2),$$

valable pour des variables aléatoires X_1, X_2, Y_1, Y_2 également de carré intégrable. De plus pour tous réels a, b, c, d , on a

$$\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y).$$

En particulier, les translations laissent invariante la covariance. La covariance permet de quantifier la liaison entre deux variables aléatoires avec le coefficient de corrélation.

Proposition-Définition 3 *Le coefficient de corrélation linéaire de deux variables aléatoires X et Y de carré intégrable est défini par*

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)},$$

où $\sigma(X) = \sqrt{\text{Var}(X)}$ désigne l'écart-type de X . De plus, $|r(X, Y)| \leq 1$ et $r(X, Y) = \pm 1$ si et seulement si $\frac{X - \mathbb{E}(X)}{\sigma(X)} = \pm \frac{Y - \mathbb{E}(Y)}{\sigma(Y)}$ presque sûrement.

Preuve. On pose $T = \frac{X - \mathbb{E}(X)}{\sigma(X)}$ et $S = \frac{Y - \mathbb{E}(Y)}{\sigma(Y)}$. Alors T et S sont des variables aléatoires de moyenne nulle et de variance 1. De plus $\text{Cov}(X, Y) = \mathbb{E}(ST)$. L'inégalité de Cauchy-Schwarz garantit que

$$|r(X, Y)| \leq \mathbb{E}(|ST|) \leq 1.$$

Supposons que $r(X, Y) = 1$. Alors $\mathbb{E}(ST) = 1$ et donc

$$\mathbb{E}((S - T)^2) = \mathbb{E}(S^2) + \mathbb{E}(T^2) - 2\mathbb{E}(ST) = 0.$$

On en déduit alors que $\mathbb{P}(S = T) = 1$ en utilisant les résultats du premier chapitre. Si $r(X, Y) = -1$, on a de même $\mathbb{E}((S + T)^2) = 0$ ce qui donne $\mathbb{P}(S = -T) = 1$. Le sens réciproque de l'équivalence est évident. \square

Remarque. Lorsque $r(X, Y) = 0$, on parle de variables aléatoires décorrélées. L'indépendance entraîne bien sûr la décorrélation mais le contraire est faux : les variables aléatoires X et $Y = \epsilon X$ où $X \sim \mathcal{N}(0, 1)$ et $\mathbb{P}(\epsilon = 1) = 1 - \mathbb{P}(\epsilon = -1) = \frac{1}{2}$ sont décorrélées mais pas indépendantes vu que $X^2 = Y^2$.

Le coefficient de corrélation sera plutôt positif si les deux variables ont tendance à s'écarter de leur moyenne dans le même sens et négatif si elles ont tendance à s'écarter de leur moyenne en sens contraire. En revanche, un coefficient de corrélation nul dénote une absence claire de liaison linéaire : avec les notations de la preuve précédente l'écart quadratique minimal $\mathbb{E}((T - aS)^2)$ est atteint pour $a = r(X, Y)$ et vaut $\mathbb{E}(T^2) - r(X, Y)^2$ qui est clairement maximale lorsque $r(X, Y) = \text{Cov}(X, Y) = 0$. Cependant il peut exister une liaison non linéaire comme dans le contre-exemple ci dessus.

5.3 Fonction caractéristique d'une variable aléatoire

Pour deux vecteurs u et v dans \mathbb{R}^d représentés à l'aide de matrices colonnes, nous noterons u^T la transposée de u (qui est donc une matrice ligne) et le produit $u^T v = \sum_{i=1}^d u_i v_i$ désignera le produit scalaire canonique de u et de v .

Définition 19 Si X est une variable aléatoire à valeurs dans \mathbb{R}^d , sa fonction caractéristique $\phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ est définie par

$$\phi_X(t) = \mathbb{E}(\exp(it^T X)), \quad t \in \mathbb{R}^d.$$

L'espérance d'une variable aléatoire Z pouvant prendre des valeurs complexes est définie par

$$\mathbb{E}(\operatorname{Re}(Z)) + i\mathbb{E}(\operatorname{Im}(Z)).$$

Ainsi

$$\phi_X(t) = \mathbb{E}(\cos(t^T X)) + i\mathbb{E}(\sin(t^T X))$$

est toujours bien définie car les fonctions cosinus et sinus sont bornées.

Exemples

– Si X suit une loi de Poisson de paramètre $\lambda > 0$. Alors

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

On rappelle que pour un nombre complexe z , on a $\exp(z) = \sum_{k=0}^{+\infty} \frac{z^k}{k!}$.

– Supposons maintenant que X suive une loi gaussienne centrée réduite. Le calcul direct de la fonction caractéristique est plus délicat. On peut faire le calcul en dérivant ϕ_X . Nous admettrons la possibilité de dériver sous le signe espérance, ce qui donne

$$\phi'_X(t) = \mathbb{E}(iX \exp(itX)).$$

On calcule alors cette dernière expression à l'aide d'une intégration par parties (les calculs pour des fonctions à valeurs complexes sont les mêmes que pour des fonctions à valeurs réelles). On a alors

$$\begin{aligned} \phi'_X(t) &= \int \frac{ix}{\sqrt{2\pi}} \exp(itx) \exp(-x^2/2) dx \\ &= \left[-\exp(itx) \exp(-x^2/2) \frac{i}{\sqrt{2\pi}} \right]_{-\infty}^{+\infty} - \int t \exp(itx) \exp(-x^2/2) dx. \end{aligned}$$

On trouve donc $\phi'_X(t) = -t\phi_X(t)$ ce qui donne $\phi_X(t) = \exp(-t^2/2)$.

On peut en déduire la fonction caractéristique de la variable $Y = m + \sigma X \sim \mathcal{N}(m, \sigma^2)$:

$$\phi_Y(t) = \exp(itm)\phi_X(\sigma t) = \exp(itm) \exp\left(-\frac{\sigma^2 t^2}{2}\right).$$

La fonction caractéristique et les moments. Il est possible de montrer (mais nous l'admettrons) que si une variable aléatoire à valeurs réelles X admet un moment d'ordre $n \in \mathbb{N}^*$, alors la fonction caractéristique est n fois dérivable et les dérivées peuvent être obtenues en dérivant sous le signe \mathbb{E} . Ainsi pour $k = 1, \dots, n$, on a

$$\phi_X^{(k)}(t) = i^k \mathbb{E}(X^k \exp(itX)),$$

ce qui donne $m_k = \phi_X^{(k)}(0)$. On pourra vérifier à titre d'exercice qu'on retrouve par exemple la moyenne et la variance d'une loi de Poisson en dérivant deux fois la fonction caractéristique.

Proposition 28 *La fonction caractéristique d'une variable aléatoire à valeurs dans \mathbb{R}^d est une fonction continue telle que $\phi_X(0) = 1$. De plus si $Y : \Omega \rightarrow \mathbb{R}^d$ est une autre variable aléatoire telle que $\phi_Y(t) = \phi_X(t)$ pour tout $t \in \mathbb{R}^d$, alors X et Y ont la même loi. La fonction caractéristique caractérise donc complètement la loi.*

Évidemment deux variables aléatoires ayant même loi ont même fonction caractéristique d'après le théorème de transfert. La preuve de l'implication inverse donnée dans cette proposition est un peu technique et sera admise. On retiendra que si les intégrales des fonctions $x \mapsto \cos(t^T x)$ et $x \mapsto \sin(t^T x)$ coïncident pour deux lois alors ces lois sont égales. Aussi on peut en déduire la propriété suivante : si $\mathbb{E}(h(X)) = \mathbb{E}(h(Y))$ pour toute fonction $h : \mathbb{R}^d \rightarrow \mathbb{R}$ continue bornée, alors $\mathbb{P}_X = \mathbb{P}_Y$ (les variables X et Y ont la même loi).

La fonction génératrice des moments. Pour une variable aléatoire X discrète à valeurs dans \mathbb{N} , on utilise plutôt la fonction génératrice G_X définie par

$$G_X(s) = \mathbb{E}(s^X) = \sum_{k=0}^{+\infty} s^k \mathbb{P}(X = k), \quad 0 < s < 1.$$

Une généralisation de cette notion est aux variables aléatoires à valeurs dans \mathbb{R}^d est donnée par

$$L_X(t) = \mathbb{E}(\exp(t^T X)), \quad t \in \mathbb{R}^d.$$

Ces deux notions évitent le recours aux nombres complexes et les moments peuvent être aussi obtenus par dérivation. Le problème est que contrairement à la fonction caractéristique qui est toujours définie, la fonction génératrice des moments ne l'est pas toujours si la variable aléatoire prend des valeurs négatives et/ou n'est pas bornée.

Proposition 29 *Deux variables aléatoires X et Y à valeurs respectives dans \mathbb{R}^d et \mathbb{R}^p sont indépendantes si et seulement si*

$$\phi_{(X,Y)}(u, v) = \mathbb{E} \left(\exp \left(\sum_{i=1}^d u_i X_i + \sum_{j=1}^p v_j Y_j \right) \right) = \phi_X(u) \phi_Y(v), \quad (u, v) \in \mathbb{R}^d \times \mathbb{R}^p.$$

Évidemment la condition nécessaire est évidente car si X et Y sont indépendantes alors il en va de même des variables $\exp(iu^T X)$ et $\exp(iv^T Y)$ et l'espérance de leur produit coïncide avec le produit de leur espérance. La preuve de la condition suffisante est un peu technique et ne sera pas abordée.

Chapitre 6

Loi et espérance conditionnelle

6.1 Généralisation des lois conditionnelles

Dans cette section, nous généralisons la notion de loi conditionnelle pour un couple de variables aléatoires dont la loi admet une densité par rapport à une mesure produit. Soient (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{R}^d \times \mathbb{R}^p$ et μ et ν deux mesures σ -finies sur \mathbb{R}^d et \mathbb{R}^p respectivement. On suppose que la loi de (X, Y) admet une densité notée $f_{X,Y}$ par rapport à la mesure produit $\mu \otimes \nu$, autrement dit

$$\mathbb{P}(X \in A, Y \in B) = \int_{A \times B} f_{X,Y}(x, y) d(\mu \otimes \nu)(x, y), \quad A \times B \in \mathcal{B}(\mathbb{R}^d) \times \mathcal{B}(\mathbb{R}^p). \quad (6.1)$$

Rappelons que le théorème de Fubini garantit que

$$\mathbb{P}(X \in A, Y \in B) = \int_A \left(\int_B f_{X,Y}(x, y) d\nu(y) \right) d\mu(x) = \int_B \left(\int_A f_{X,Y}(x, y) d\mu(x) \right) d\nu(y)$$

et nous remplacerons souvent la notation $d(\mu \otimes \nu)(x, y)$ par $d\mu(x)d\nu(y)$. Dans ce cas, les lois de X et de Y sont données par

$$\mathbb{P}(X \in A) = \int_A \left(\int f_{X,Y}(x, y) d\nu(y) \right) d\mu(x), \quad A \in \mathcal{B}(\mathbb{R}^d).$$

$$\mathbb{P}(Y \in B) = \int_B \left(\int f_{X,Y}(x, y) d\mu(x) \right) d\nu(y), \quad B \in \mathcal{B}(\mathbb{R}^p).$$

Ainsi la loi de X admet une densité f_X par rapport à μ qui est donnée par

$$f_X(x) = \int f_{X,Y}(x, y) d\nu(y), \quad x \in \mathbb{R}^d.$$

De même, la loi de Y admet une densité f_Y par rapport à ν qui est donnée par

$$f_Y(y) = \int f_{X,Y}(x, y) d\mu(x), \quad y \in \mathbb{R}^p.$$

Lorsque μ et ν sont les mesures de Lebesgue, on retrouve les lois à densité précédemment étudiées. Lorsque μ et ν sont les mesures de comptage sur \mathbb{N} , on retrouve les couples de variables aléatoires discrètes : la densité $f_{X,Y}$ peut alors être définie par $f_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$ pour $(x, y) \in \mathbb{N}^2$. Un exemple de cas non étudié précédemment est celui où μ désigne la mesure de Lebesgue sur \mathbb{R} et ν désigne la mesure de comptage sur \mathbb{N} . Pour ce dernier cas, on pourra remarquer que

$$\mathbb{P}(X \in A, Y = y) = \int_A f_{X,Y}(x, y) dx, \quad A \in \mathcal{B}(\mathbb{R})$$

et que $f_Y(y) = \mathbb{P}(Y = y)$ pour tout $y \in \mathbb{N}$.

Définition 20 Soit (X, Y) un couple de variables aléatoires dont la loi vérifie (6.1). Soit également g une densité de probabilité quelconque pour la mesure μ . On définit alors la loi conditionnelle de $X|Y = y$ comme étant la mesure de probabilité de densité notée $f_{X|Y}(\cdot|y)$ par rapport à la mesure μ , densité qui est définie par

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad x \in \mathbb{R}^d,$$

lorsque $f_Y(y)$ est différent de 0. Si $y \in \mathbb{R}^p$ vérifie $f_Y(y) = 0$, on définit $f_{X|Y}(\cdot|y) = g$.

Lorsque μ et ν sont ou bien toutes deux des mesures de Lebesgue ou bien toutes deux la mesure de comptage sur \mathbb{N} , on retrouve la définition des lois conditionnelles donnée pour un couple de variables aléatoires à densité ou pour un couple de variables discrètes. Donnons un exemple de situation où on est amené à considérer des lois conditionnelles pour un couple (X, Y) à densité par rapport au produit de la mesure de Lebesgue et de la mesure de comptage. Supposons que l'on dispose de n machines pour effectuer un tâche donnée, les machines ont des durées de vie X_1, \dots, X_n de lois de densités respectives f_1, \dots, f_n . La probabilité que la machine i soit choisie pour effectuer cette tâche est p_i . On observe uniquement une durée de vie X sans savoir quelle machine a été choisie. Quelle est alors la probabilité que la machine i ait été choisie? On peut remarquer qu'il est possible de représenter X sous la forme

$$X = \sum_{i=1}^n X_i \mathbb{1}_{Y=i},$$

où Y une variable aléatoire indépendante de (X_1, \dots, X_n) et telle que $\mathbb{P}(Y = i) = p_i$ pour $i = 1, \dots, n$. Si $A \in \mathcal{B}(\mathbb{R})$ et $i \in \{1, \dots, n\}$, on a

$$\mathbb{P}(X \in A, Y = i) = \mathbb{P}(X_i \in A, Y = i) = \mathbb{P}(X_i \in A)p_i = \int_A p_i f_i(x) dx.$$

En posant $f_{X,Y}(x,y) = f_y(x)p_y \mathbb{1}_{\{1, \dots, n\}}(y)$, on voit que (X, Y) a une densité par rapport à $\lambda \otimes \nu$, où λ désigne la mesure de Lebesgue sur \mathbb{R} et ν la mesure de comptage sur \mathbb{N} . De plus la loi de $X|Y = i$ a la densité f_i . On cherche ici la loi de $Y|X = x$. Nous avons

$$f_X(x) = \int p_y f_y(x) d\nu(y) = \sum_{j=1}^n p_j f_j(x)$$

(la densité de X est donc donnée par un mélange) et d'après la définition de la loi conditionnelle, on a si $i = 1, \dots, n$:

$$\mathbb{P}(Y = i|X = x) = \frac{p_i f_i(x)}{\sum_{j=1}^n p_j f_j(x)}.$$

Moyenne conditionnelle En gardant les notations précédentes, pour un couple de variable aléatoire (X, Y) admettant une densité $f_{X,Y}$ par rapport à une mesure produit $\mu \otimes \nu$ et lorsque X est une variable aléatoire intégrable et à valeurs réelles, la moyenne conditionnelle est définie par

$$\mathbb{E}(X|Y = y) = \int x f_{X|Y}(x|y) dx.$$

Lorsque ν est la mesure de comptage sur \mathbb{N} , on pourra remarquer que si $\mathbb{P}(Y = y) > 0$ alors

$$\mathbb{E}(X|Y = y) = \frac{\mathbb{E}(X \mathbb{1}_{Y=y})}{\mathbb{P}(Y = y)}.$$

6.2 Espérance conditionnelle

L'objectif de la notion d'espérance conditionnelle peut être résumé grossièrement ainsi. Lorsque une partie de l'information est disponible (des événements ou encore une variable aléatoire, ce qui bien souvent revient au même), on souhaite attribuer une valeur moyenne à une variable aléatoire X non observée mais qui tient compte de cette information (par exemple, lors du lancer de deux dés, on sait juste que la somme est paire ou pas mais on ne connaît pas les numéros sur les dés). L'espérance conditionnelle à une sous-tribu est une variable aléatoire qui pourra être construite à partir de l'information disponible et qui sous certaines hypothèses sera la meilleure approximation de X au sens de la perte quadratique. Lorsque l'information disponible se résumera à une variable aléatoire Y , la moyenne de la loi conditionnelle de $X|Y = y$ sera directement à l'origine de cette variable aléatoire.

6.2.1 Variable aléatoire mesurable par rapport à une sous-tribu

Soit \mathcal{B} une sous-tribu de \mathcal{A} . On dira qu'une variable aléatoire $Z : \Omega \rightarrow \mathbb{R}$ est \mathcal{B} -mesurable si pour tous $a \leq b$, l'événement $\{a \leq Z \leq b\}$ appartient à \mathcal{B} . Ainsi la variable aléatoire Z peut être construite uniquement à l'aide des événements de \mathcal{B} , en utilisant l'approximation par une variable aléatoire discrète (on renvoie au Chapitre 1 pour cette approximation).

De façon imagée, \mathcal{B} peut être vue comme une quantité d'information disponible et les variables aléatoires \mathcal{B} -mesurables comme toutes les fonctions qu'il est possible de construire à partir de cette information.

Donnons un premier exemple. Si \mathcal{B} est la tribu engendrée par un événement A de \mathcal{A} , alors

$$\mathcal{B} = \{\emptyset, A, A^c, \Omega\}.$$

Si Z est une variable aléatoire \mathcal{B} -mesurable alors il existe deux nombres réels α et β tels que

$$Z = \alpha \mathbb{1}_A + \beta \mathbb{1}_{A^c}.$$

En effet, si Z est constante et égale à c , on peut prendre $\alpha = \beta = c$. Si Z prend deux valeurs distinctes a et b , on a la décomposition

$$Z = a \mathbb{1}_{Z=a} + b \mathbb{1}_{Z=b}$$

et comme $\{Z = a\}, \{Z = b\} \in \mathcal{B}$ sont deux événements non vides et de réunion Ω , l'un vaut A l'autre A^c et on pose $\alpha, \beta = a, b$. Ensuite il n'est pas possible que Z prenne trois valeurs distinctes car il n'y a pas trois éléments distincts non vides et différents de l'univers dans \mathcal{B} . Remarquons que lorsque $A = \emptyset$ ou $A = \Omega$, alors les variables aléatoires \mathcal{B} -mesurables sont simplement les constantes.

Cet exemple se généralise au cas d'une tribu $\mathcal{B} = \sigma(\{A_1, A_2, \dots, A_n\})$ engendrée par un nombre fini d'événements A_1, \dots, A_n disjoints deux à deux et de réunion Ω . Dans ce cas \mathcal{B} est la tribu constituée de toutes les réunions formées à partir des A_i et on peut montrer que qu'une variable aléatoire \mathcal{B} -mesurable est de la forme

$$Z = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}.$$

Passons maintenant au cas d'une tribu engendrée par une variable aléatoire.

Définition 21 Si $Y : \Omega \rightarrow \mathbb{R}^d$ est une variable aléatoire, la tribu engendrée par Y est la tribu notée $\sigma(Y)$ et qui est définie par

$$\sigma(Y) = \left\{ \{Y \in A\} : A \in \mathcal{B}(\mathbb{R}^d) \right\}.$$

On pourra vérifier à titre d'exercice que la famille d'événements entre accolades donnée dans la définition ci-dessus vérifie bien la définition d'une tribu. La tribu $\mathcal{B} = \sigma(\{A_1, A_2, \dots, A_n\})$ engendrée par un nombre fini d'événements A_1, \dots, A_n disjoints deux à deux et de réunion Ω , est un exemple de ce type de tribu si on pose par exemple

$$Y = \sum_{i=1}^n (i-1) \mathbf{1}_{A_i}.$$

En particulier la tribu engendrée par l'événement A coïncide avec la tribu engendrée par la variable aléatoire $\mathbf{1}_A$.

Remarque. En fait, on peut toujours voir une tribu comme engendrée par une variable aléatoire mais ceci demande une définition plus générale de la notion de variable aléatoire qu'il n'est pas question d'aborder dans ce cours. Par exemple en mathématiques financières, il existe la notion d'une tribu engendrée par des variables aléatoires $Z_s, 0 \leq s \leq t$ qui représente l'ensemble des prix de l'instant initial jusqu'au temps $t > 0$: dans ce cas la variable aléatoire sous-jacente est à valeurs dans un espace de fonction et pas dans \mathbb{R}^d .

Lorsque une tribu \mathcal{B} est générée par une variable aléatoire, on a une description plus parlante des variables aléatoires à valeurs réelles et \mathcal{B} -mesurables.

Lemme 3 (lemme de Doob) *Soit $Z : \Omega \rightarrow \mathbb{R}$ une variable aléatoire mesurable par rapport à la tribu $\mathcal{B} = \sigma(Y)$ où $Y : \Omega \rightarrow \mathbb{R}^d$ est une variable aléatoire. Alors, il existe une fonction $g : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable et telle que $Z = g(Y)$.*

En d'autres termes, les variables aléatoires $\sigma(Y)$ -mesurables coïncident avec les fonctionnelles (mesurables) de Y .

6.2.2 Espérance conditionnelle par rapport à une sous-tribu

Le théorème suivant est à la base de la définition de l'espérance conditionnelle. Pour simplifier les énoncés nous introduisons l'ensemble $\mathcal{L}_{\mathcal{B}}^1$ (resp. $\mathcal{L}_{\mathcal{B}}^2$) constitué des variables aléatoires $U : \Omega \rightarrow \mathbb{R}$, \mathcal{B} -mesurables et intégrables (resp. dont le carré est intégrable, i.e $\mathbb{E}(U^2) < +\infty$).

Théorème 11 *Soit X une variable aléatoire à valeurs réelles et intégrable. Soit également \mathcal{B} une sous-tribu de \mathcal{A} . Alors il existe une variable aléatoire $Z \in \mathcal{L}_{\mathcal{B}}^1$ telle que pour tout $A \in \mathcal{B}$, on ait*

$$\mathbb{E}(\mathbf{1}_A X) = \mathbb{E}(\mathbf{1}_A Z). \quad (6.2)$$

Cette variable aléatoire est unique à l'égalité presque sûre près : si $Z' \in \mathcal{L}_{\mathcal{B}}^1$ qui vérifie (6.2) alors $\mathbb{P}(Z = Z') = 1$.

En outre, si $\mathbb{E}(X^2) < +\infty$ alors $Z \in \mathcal{L}_{\mathcal{B}}^2$ et dans ce cas on a pour toute variable aléatoire $U \in \mathcal{L}_{\mathcal{B}}^2$

$$\mathbb{E}(UX) = \mathbb{E}(UZ), \quad (6.3)$$

ainsi que l'inégalité

$$\mathbb{E}((X - Z)^2) \leq \mathbb{E}((X - U)^2). \quad (6.4)$$

Réciproquement, si $\mathbb{E}(X^2) < +\infty$ et $Z \in \mathcal{L}_{\mathcal{B}}^2$ vérifie (6.4) pour tout $U \in \mathcal{L}_{\mathcal{B}}^2$ alors Z vérifie (6.3) (et donc également (6.2)).

Définition 22 *Conservons les hypothèses du théorème précédent. Alors toute variable aléatoire $Z \in \mathcal{L}_{\mathcal{B}}^1$ satisfaisant (6.2) est appelée une version de l'espérance conditionnelle de X sachant \mathcal{B} . Nous noterons alors abusivement $\mathbb{E}(X|\mathcal{B})$ toute version de l'espérance conditionnelle.*

Lorsque la variable aléatoire X est de carré intégrable, l'espérance conditionnelle minimise l'écart quadratique $\mathbb{E}((X - U)^2)$ sur $U \in \mathcal{L}_{\mathcal{B}}^2$ et de plus, d'après le théorème précédent, ce minimum est unique à l'égalité presque sûre près. C'est pour cela que cette notion est souvent utilisée pour faire de la prévision en modélisation lorsqu'on souhaite prévoir une variable aléatoire X non observée à partir de certains événements (ou variables aléatoires) connus. Lorsque la variable aléatoire est simplement intégrable, cette propriété de minimisation n'est plus forcément valable : l'espérance conditionnelle apparaît alors comme une variable aléatoire de $\mathcal{L}_{\mathcal{B}}^1$ qui a la même valeur moyenne que X sur les événements de \mathcal{B} (voir (6.2)). Reste à savoir comment calculer l'espérance conditionnelle en pratique. C'est ce que nous verrons un peu plus loin.

Quelques éléments de preuve pour le Théorème 11.

- L'existence de la variable aléatoire Z est difficile à justifier et est admise. Montrons l'unicité. Si Z et Z' sont deux éléments de $\mathcal{L}_{\mathcal{B}}^1$ satisfaisant (6.2), alors en posant $A = \{Z > Z'\}$ qui est bien un événement de \mathcal{B} , on a $\mathbb{E}(\mathbf{1}_A(Z - Z')) = 0$ et la variable aléatoire $V = \mathbf{1}_A(Z - Z')$ prend des valeurs positives. D'après les propriétés de l'espérance, on a $\mathbb{P}(V = 0) = 1$. Mais $\{V = 0\} = \{Z \leq Z'\}$ et donc $\mathbb{P}(Z \leq Z') = 1$. En échangeant le rôle de Z et Z' , on a aussi $\mathbb{P}(Z' \leq Z) = 1$. L'intersection de deux événements de probabilité 1 étant de probabilité 1, on conclut que $\mathbb{P}(Z = Z') = 1$.
- Montrons ensuite que la variable aléatoire Z du théorème vérifie (6.3) lorsque X est de carré intégrable. Il faut auparavant montrer que Z est de carré intégrable. Remarquons qu'il est clair que par linéarité, les égalités (6.2) entraîne que $\mathbb{E}(XU) = \mathbb{E}(ZU)$ pour toute variable aléatoire U , \mathcal{B} -mesurable et ne prenant qu'un nombre fini de valeurs. Considérons alors une suite croissante $(V_n)_n$ d'éléments de \mathcal{D}_+ (voir le Chapitre 1), \mathcal{B} -mesurables et convergente point par point vers $|Z|$. Posons $U_n = (\mathbf{1}_{Z \geq 0} - \mathbf{1}_{Z < 0})V_n$. Alors on a

$$\mathbb{E}(V_n|Z) = \mathbb{E}(U_n Z) = \mathbb{E}(U_n X) \leq \sqrt{\mathbb{E}(V_n^2)} \sqrt{\mathbb{E}(X^2)},$$

où on a utilisé l'inégalité de Cauchy-Schwarz. Comme $V_n|Z| \geq V_n^2$, on en déduit $\mathbb{E}(V_n^2) \leq \mathbb{E}(X^2)$ et comme (V_n^2) est une suite croissante de \mathcal{D}_+ convergente point par point vers Z^2 , il résulte de la définition de l'espérance que :

$$\mathbb{E}(Z^2) = \lim_{n \rightarrow \infty} \mathbb{E}(V_n^2) \leq \mathbb{E}(X^2), \quad (6.5)$$

ce qui prouve bien que $\mathbb{E}(Z^2) < +\infty$. Remarquons ensuite que si les égalités (6.3) sont valables pour U ne prenant qu'un nombre fini de valeurs alors, elles sont aussi valables pour U bornée : en effet, il est facile d'approcher U bornée par une suite (U_n) de variables aléatoires ne prenant qu'un nombre fini de valeurs de sorte que

$$\lim_{n \rightarrow +\infty} \sup_{\omega \in \Omega} |U_n(\omega) - U(\omega)| = 0.$$

Ceci entraîne que

$$\begin{aligned} |\mathbb{E}(UX) - \mathbb{E}(UZ)| &\leq |\mathbb{E}(UX) - \mathbb{E}(U_n X)| + |\mathbb{E}(U_n Z) - \mathbb{E}(UZ)| \\ &\leq 2 \sup_{\omega \in \Omega} |U_n(\omega) - U(\omega)| \times (\mathbb{E}(|X|) + \mathbb{E}(|Z|)). \end{aligned}$$

En faisant tendre n vers $+\infty$ on obtient les égalités 6.3 pour des variables aléatoires U bornées. Ces égalités s'étendent à toutes les variables aléatoires $U \in \mathcal{L}_{\mathcal{B}}^2$. En effet si $U \in \mathcal{L}_{\mathcal{B}}^2$ et $n \in \mathbb{N}^*$, la variable aléatoire $U_n = U \mathbf{1}_{|U| \leq n}$ est bornée et on a $\mathbb{E}(U_n X) = \mathbb{E}(U_n Z)$.

Montrons alors que $\lim_{n \rightarrow \infty} \mathbb{E}(U_n X) = \mathbb{E}(UX)$ (le même raisonnement vaudra pour Z). Les propriétés de l'espérance et l'inégalité de Cauchy-Schwarz assure que

$$|\mathbb{E}(U_n X) - \mathbb{E}(UX)| \leq \mathbb{E}(|X| \cdot |U_n - U|) \leq \sqrt{\mathbb{E}(X^2)} \sqrt{\mathbb{E}((U - U_n)^2)}.$$

Montrons pour conclure que $\lim_{n \rightarrow \infty} \mathbb{E}((U - U_n)^2) = 0$. On a $S_n = (U - U_n)^2 = U^2 \mathbf{1}_{|U| > n}$ et en posant $T_n = U^2 \mathbf{1}_{|U| \leq n}$, on a

$$\mathbb{E}(U^2) = \mathbb{E}(T_n) + \mathbb{E}(S_n).$$

Par convergence monotone (voir le chapitre Mesure et intégration), on a $\lim_{n \rightarrow \infty} \mathbb{E}(T_n) = \mathbb{E}(U^2)$. On conclut alors que $\lim_{n \rightarrow \infty} \mathbb{E}(S_n) = 0$. On a de même $\lim_{n \rightarrow \infty} \mathbb{E}(U_n Z) = \mathbb{E}(UZ)$ et on prouve alors facilement (6.3).

- Pour finir montrons qu'une variable aléatoire $Z \in \mathcal{L}_{\mathcal{B}}^2$ satisfait (6.3) pour tout $U \in \mathcal{L}_{\mathcal{B}}^2$ si et seulement si elle satisfait à 6.4 pour tout $U \in \mathcal{L}_{\mathcal{B}}^2$.

Supposons d'abord que Z satisfait à 6.3 pour tout $U \in \mathcal{L}_{\mathcal{B}}^2$ (donc en particulier pour $U = Z$). Alors pour U donné dans $\mathcal{L}_{\mathcal{B}}^2$, nous avons

$$\begin{aligned} \mathbb{E}((X - U)^2) - \mathbb{E}((X - Z)^2) &= \mathbb{E}(U^2) - 2\mathbb{E}(XU) + 2\mathbb{E}(XZ) - \mathbb{E}(Z^2) \\ &= \mathbb{E}(U^2) - 2\mathbb{E}(ZU) + 2\mathbb{E}(Z^2) - \mathbb{E}(Z^2) \\ &= \mathbb{E}((U - Z)^2) \\ &\geq 0, \end{aligned}$$

ce qui prouve (6.4).

Supposons maintenant que Z satisfait à 6.4 pour tout $U \in \mathcal{L}_{\mathcal{B}}^2$. Soit alors $t > 0$. On a $U + tZ \in \mathcal{L}_{\mathcal{B}}^2$ et (6.4) entraîne que

$$\mathbb{E}((X - Z - tU)^2) \geq \mathbb{E}((X - Z)^2).$$

En développant les carré, on voit alors que

$$t^2 \mathbb{E}(U^2) - 2t \mathbb{E}((X - Z)U) \geq 0.$$

En divisant l'inégalité précédente par t et en faisant tendre t vers 0, on obtient $\mathbb{E}((X - Z)U) \leq 0$. En remplaçant U par $-U$ dans cette dernière inégalité, on obtient également $\mathbb{E}((X - Z)U) \geq 0$. Finalement on conclut que $\mathbb{E}((X - Z)U) = 0$ ce qui prouve (6.3). \square

Deux exemples simples. Lorsque $\mathcal{B} = \{\emptyset, \Omega\}$, les variables aléatoires \mathcal{B} -mesurables sont les constantes. Ce n'est pas surprenant, cette sous-tribu n'apporte pas d'information réelle. Dans ce cas la constante qui vérifie (6.2) est $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X)$ (on a le choix uniquement entre $A = \emptyset$ et $A = \Omega$). Du point de vue de la minimisation de l'écart quadratique, ce n'est pas surprenant : on sait sans utiliser ce qui précède que la constante a qui minimise $\mathbb{E}((X - a)^2)$ est $a = \mathbb{E}(X)$ et la valeur minimale est la variance (pour le prouver directement, développer le carré et étudier le polynôme du second degré en a).

Supposons maintenant que la sous-tribu \mathcal{B} soit engendrée par un seul événement A non vide et différent de Ω . Alors l'espérance conditionnelle est de la forme

$$Z = \alpha \mathbf{1}_A + \beta \mathbf{1}_{A^c},$$

pour des nombres réels α et β à déterminer. On utilise les égalités (6.2). On a

$$\mathbb{E}(X \mathbf{1}_A) = \mathbb{E}(Z \mathbf{1}_A) = \alpha \mathbb{P}(A) + \beta \times 0,$$

ce qui donne $\alpha = \frac{\mathbb{E}(X\mathbf{1}_A)}{\mathbb{P}(A)}$. De même,

$$\mathbb{E}(X\mathbf{1}_{A^c}) = \mathbb{E}(Z\mathbf{1}_{A^c}) = \beta\mathbb{P}(A^c),$$

et donc $\beta = \frac{\mathbb{E}(X\mathbf{1}_{A^c})}{\mathbb{P}(A^c)}$. On en déduit

$$\mathbb{E}(X|\mathcal{B}) = \frac{\mathbb{E}(X\mathbf{1}_A)}{\mathbb{P}(A)}\mathbf{1}_A + \frac{\mathbb{E}(X\mathbf{1}_{A^c})}{\mathbb{P}(A^c)}\mathbf{1}_{A^c}.$$

Par exemple si on lance deux dés équilibrés, on peut considérer A l'événement "la somme est un nombre pair" et X le numéro du premier dé. Si on observe qu'effectivement la somme est paire, on prévoira X par la valeur $\frac{\mathbb{E}(X\mathbf{1}_A)}{\mathbb{P}(A)}$. Si en revanche la somme est impaire, on utilisera la valeur $\frac{\mathbb{E}(X\mathbf{1}_{A^c})}{\mathbb{P}(A^c)}$.

Tribu engendrée par une variable aléatoire discrète Il s'agit de généraliser l'exemple précédent (qui concernait en fait le cas d'une tribu engendrée par une variable aléatoire de loi de Bernoulli). Supposons maintenant que $\mathcal{B} = \sigma(Y)$ où $Y : \Omega \rightarrow \mathbb{R}^d$ est une variable aléatoire discrète. Dans ce cas, on peut chercher l'espérance conditionnelle sous la forme $Z = g(Y)$ pour une fonction $g : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable et telle que

$$\mathbb{E}(|g(Y)|) = \sum_{y \in \text{val}(Y)} |g(y)| \times \mathbb{P}(Y = y) < +\infty.$$

Si $y \in \text{val}(Y)$, on pose $A = \{Y = y\}$. On a alors en utilisant (6.2)

$$\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(Z\mathbf{1}_A) = \sum_{w \in \text{val}(Y)} g(w)\mathbb{E}(\mathbf{1}_{Y=w}\mathbf{1}_{Y=y}) = g(y)\mathbb{P}(Y = y).$$

On obtient alors $g(y) = \frac{\mathbb{E}(X\mathbf{1}_{Y=y})}{\mathbb{P}(Y=y)}$ pour tout $y \in \text{val}(Y)$. On a

$$\sum_{y \in \text{val}(Y)} |g(y)| \times \mathbb{P}(Y = y) \leq \mathbb{E}(|X|) < +\infty$$

et on peut poser

$$\mathbb{E}(X|\mathcal{B}) = g(Y) = \sum_{y \in \text{val}(Y)} \frac{\mathbb{E}(X\mathbf{1}_{Y=y})}{\mathbb{P}(Y = y)}\mathbf{1}_{Y=y}.$$

Espérance conditionnelle et moyenne conditionnelle. Voici le lien fondamental entre la moyenne de la loi conditionnelle et l'espérance conditionnelle lorsque \mathcal{B} est une tribu engendrée par une variable aléatoire $Y : \Omega \rightarrow \mathbb{R}^d$ et lorsque le couple (X, Y) possède une densité $f_{X,Y}$ par rapport une mesure produit $\mu \otimes \nu$.

Proposition 30 *On suppose X intégrable. Soit $g(y) = \mathbb{E}(X|Y = y)$. Alors $\mathbb{E}(X|\sigma(Y)) = g(Y)$.*

Preuve. On vérifie l'égalité (6.2). Tout événement de $\mathcal{B} = \sigma(Y)$ est de la forme $\{Y \in A\}$ pour $A \in \mathcal{B}(\mathbb{R}^d)$. On a alors

$$\begin{aligned} \mathbb{E}(X\mathbf{1}_{\{Y \in A\}}) &= \int x\mathbf{1}_A(y)f_{X,Y}(x,y)d\mu(x)d\nu(y) \\ &= \int x\mathbf{1}_A(y)f_{X|Y}(x|y)f_Y(y)d\mu(x)d\nu(y) \\ &= \int \mathbf{1}_A(y)\left(\int f_{X|Y}(x|y)d\mu(x)\right)f_Y(y)d\nu(y) \\ &= \int_A g(y)f_Y(y)d\nu(y) \\ &= \mathbb{E}(\mathbf{1}_A(Y)g(Y)) \end{aligned}$$

D'après la caractérisation (6.2) de l'espérance conditionnelle, la proposition est prouvée. \square

Notation. On utilisera plutôt la notation $\mathbb{E}(X|Y)$ au lieu de $\mathbb{E}(X|\sigma(Y))$.

Remarque. Reprenons l'exemple de la tribu engendrée par une variable aléatoire discrète Y à valeurs dans $\mathbb{N} = \text{val}(Y)$ et examinons la cas où le couple (X, Y) admet une densité $f_{X,Y}$ par rapport à une mesure produit $\mu \otimes \nu$ avec ν la mesure de comptage sur \mathbb{N} . Alors, on a prouvé que

$$\mathbb{E}(X|Y) = \sum_{n=0}^{+\infty} \frac{\mathbb{E}(X\mathbf{1}_{Y=n})}{\mathbb{P}(Y=n)} \mathbf{1}_{Y=n}.$$

On a vu aussi (voir le début de ce chapitre) que

$$\mathbb{E}(X|Y=n) = \frac{\mathbb{E}(X\mathbf{1}_{Y=n})}{\mathbb{P}(Y=n)}.$$

Cet exemple confirme donc bien le résultat de la Proposition 30.

6.2.3 Propriétés générales de l'espérance conditionnelle

Voici les propriétés générales de l'espérance conditionnelle.

Proposition 31 Soient X et X' deux variables aléatoires intégrables, \mathcal{B} une sous-tribu de \mathcal{A} et $U \in \mathcal{L}_{\mathcal{B}}^1$.

1. On a $\mathbb{E}(\mathbb{E}(X|\mathcal{B})) = \mathbb{E}(X)$.
2. On a $\mathbb{E}(U|\mathcal{B}) = U$.
3. On a les propriétés de linéarité suivantes (les égalités étant valables presque sûrement)

$$\mathbb{E}(X + X'|\mathcal{B}) = \mathbb{E}(X|\mathcal{B}) + \mathbb{E}(X'|\mathcal{B}),$$

$$\mathbb{E}(UX|\mathcal{B}) = U\mathbb{E}(X|\mathcal{B}),$$

lorsque U est bornée. Si de plus, $\mathbb{E}(X^2) < +\infty$, alors la deuxième égalité est aussi valable pour $U \in \mathcal{L}_{\mathcal{B}}^2$.

4. Si X est à valeurs positives, alors $\mathbb{E}(X|\mathcal{B})$ est aussi à valeurs positives. De plus si $X \leq X'$, alors $\mathbb{E}(X|\mathcal{B}) \leq \mathbb{E}(X'|\mathcal{B})$.
5. Si X est indépendante de \mathcal{B} , alors $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X)$.

6. Si \mathcal{C} est une tribu telle que $\mathcal{C} \subset \mathcal{B}$ alors

$$\mathbb{E}(\mathbb{E}(X|\mathcal{B})|\mathcal{C}) = \mathbb{E}(X|\mathcal{C}).$$

On retiendra aussi les deux cas particuliers suivants du point 2. : $\mathbb{E}(\lambda X|\mathcal{B}) = \lambda \mathbb{E}(X|\mathcal{B})$ et $\mathbb{E}(\lambda|\mathcal{B}) = \lambda$ pour tout réel λ . On peut remarquer que les trois premiers points montrent que l'espérance conditionnelle se comporte en gros comme une espérance mais pour laquelle les variables aléatoires \mathcal{B} -mesurables jouent le rôle de constantes.

Le point 6. est souvent utilisé dans le cas suivant : $\mathcal{C} = \sigma(Y_1)$ et $\mathcal{B} = \sigma((Y_1, Y_2))$ où Y_1 et Y_2 sont deux variables aléatoires, l'inclusion résultant du fait que toute fonction de Y_1 est une fonction de (Y_1, Y_2) (voir l'exemple du modèle autorégressif).

Preuve. Pour la plupart de ces propriétés, il suffit de montrer que la variable aléatoire annoncée vérifie la caractérisation (6.2) : pour montrer une égalité du type $\mathbb{E}(S|\mathcal{B}) = Z$, on montre que $Z \in \mathcal{L}_\mathcal{B}^1$ puis que pour tout $A \in \mathcal{B}$, on a

$$\mathbb{E}(\mathbf{1}_A S) = \mathbb{E}(\mathbf{1}_A Z). \quad (6.6)$$

1. Le premier point se déduit de la relation (6.2) en prenant $A = \Omega$.
2. Si $A \in \mathcal{B}$, alors en posant $S = U$, on voit que la variable aléatoire $Z = U$, qui appartient à $\mathcal{L}_\mathcal{B}^1$, vérifie trivialement (6.6). Par unicité de l'espérance conditionnelle (à l'égalité presque sûre près), on a $\mathbb{E}(U|\mathcal{B}) = U$.
3. La variable aléatoire $S = X + X'$ est intégrable ainsi que la variable aléatoire \mathcal{B} -mesurable

$$Z = V_1 + V_2 = \mathbb{E}(X|\mathcal{B}) + \mathbb{E}(X'|\mathcal{B}).$$

De plus Z vérifie (6.6) car si $A \in \mathcal{B}$,

$$\begin{aligned} \mathbb{E}(\mathbf{1}_A Z) &= \mathbb{E}(\mathbf{1}_A V_1) + \mathbb{E}(\mathbf{1}_A V_2) \\ &= \mathbb{E}(\mathbf{1}_A X) + \mathbb{E}(\mathbf{1}_A X') \\ &= \mathbb{E}(\mathbf{1}_A S). \end{aligned}$$

Toujours par unicité, on conclut

$$\mathbb{E}(X + X'|\mathcal{B}) = \mathbb{E}(X|\mathcal{B}) + \mathbb{E}(X'|\mathcal{B}).$$

Pour prouver que $\mathbb{E}(UX|\mathcal{B}) = U\mathbb{E}(X|\mathcal{B})$ lorsque X est de carré intégrable et $U \in \mathcal{L}_\mathcal{B}^2$, on commence par remarquer que la variable aléatoire candidate est bien intégrable en tant que produit de variable aléatoire de carré intégrable et qu'elle est de plus \mathcal{B} -mesurable. Il suffit ensuite de voir que si $A \in \mathcal{B}$, on a (en posant $Z = \mathbb{E}(X|\mathcal{B})$)

$$\mathbb{E}(\mathbf{1}_A UZ) = \mathbb{E}(\mathbf{1}_A UX),$$

en utilisant (6.3) que l'on a appliqué à la variable aléatoire $\mathbf{1}_A U \in \mathcal{L}_\mathcal{B}^2$ au lieu de U et à $S = UX$ au lieu de X . Ainsi (6.6) est vérifiée pour tout $A \in \mathcal{B}$. Lorsque X est seulement intégrable, il faut d'abord que la caractérisation (6.2) s'étend en remplaçant les indicatrices $\mathbf{1}_A$ par des variables aléatoires \mathcal{B} -mesurables bornées (ceci a en fait été déjà montré dans la preuve du Théorème 11) et on procède comme pour le cas X de carré intégrable (cas précédent) pour finir la preuve.

4. Posons $Z = \mathbb{E}(X|\mathcal{B})$. En choisissant $A = \{Z < 0\}$, on a $Z\mathbf{1}_A \leq 0$ et vu que $0 \leq \mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(Z\mathbf{1}_A)$, on en déduit que $\mathbb{E}(Z\mathbf{1}_A) = 0$ et donc également que l'espérance de la variable aléatoire positive $-Z\mathbf{1}_A$ est nulle. Les propriétés de l'espérance assure que $\mathbb{P}(A) = 0$. On en déduit $\mathbb{P}(Z \geq 0) = 1$.
La deuxième assertion est une conséquence directe de la première car $X' - X \geq 0$ entraîne $\mathbb{E}(X' - X|\mathcal{B}) \geq 0$. La linéarité de l'espérance conditionnelle (point 3.) permet alors de conclure.
5. Lorsque X est indépendante de \mathcal{B} , alors $Z = \mathbb{E}(X)$ appartient à $\mathcal{L}_{\mathcal{B}}^1$ (comme toute constante). De plus pour $A \in \mathcal{B}$, (6.6) est vérifiée car les deux membres sont tous les deux égaux à $\mathbb{E}(X)\mathbb{P}(A)$.
6. Posons $S = \mathbb{E}(X|\mathcal{B})$. La variable aléatoire $Z = \mathbb{E}(S|\mathcal{C})$ appartient bien à $\mathcal{L}_{\mathcal{C}}^1$ (car S est intégrable). De plus si $A \in \mathcal{C}$, on a

$$\mathbb{E}(\mathbf{1}_A Z) = \mathbb{E}(\mathbf{1}_A S) = \mathbb{E}(\mathbf{1}_A X).$$

La première égalité utilise la définition de l'espérance conditionnelle et la deuxième également car $A \in \mathcal{B}$. On a conclut alors à l'égalité annoncée. \square

Exemple d'application pour un modèle autorégressif. On considère une suite i.i.d U de variables aléatoires à valeurs réelles, intégrables et de toutes de moyenne nulle. Posons alors $X_0 = x$ (variable aléatoire constante réelle) et définissons X_1, x_2, \dots de façon récursive par

$$X_{n+1} = aX_n + b + U_{n+1}, \quad n \in \mathbb{N}.$$

Ce type de dynamique est fréquemment utilisé pour modéliser l'évolution temporelle de certains phénomènes (évolution d'actifs d'un jour au suivant, évolution du niveau d'eau dans un lac...). On pourra remarquer en itérant l'équation que X_n peut s'écrire comme une combinaison linéaire des variables U_1, \dots, U_n . La variable aléatoire U_{n+1} est donc indépendante de X_n et plus généralement de $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Calculons alors l'espérance conditionnelle $\mathbb{E}(X_{n+1}|\mathcal{F}_n)$. On a la série d'égalités

$$\begin{aligned} \mathbb{E}(X_{n+1}|\mathcal{F}_n) &= \mathbb{E}(aX_n + b + U_{n+1}|\mathcal{F}_n) \\ &= a\mathbb{E}(X_n|\mathcal{F}_n) + b + \mathbb{E}(U_{n+1}|\mathcal{F}_n) \\ &= aX_n + b + \mathbb{E}(U_{n+1}) \\ &= aX_n + b. \end{aligned}$$

Pour la première égalité, on a utilisé la linéarité de l'espérance conditionnelle et le fait qu'une constante est \mathcal{F}_n -mesurable. Pour la deuxième, on a utilisé que X_n était \mathcal{F}_n -mesurable (et donc le point 2. de la Proposition 31) et le fait que U_{n+1} était indépendante de \mathcal{F}_n (et donc le point 5. de la Proposition 31). On peut alors faire des prévisions à horizon $h \geq 2$. Par exemple, on a en utilisant ce qui précède et le point 6. de la Proposition 31

$$\begin{aligned} \mathbb{E}(X_{n+2}|\mathcal{F}_n) &= \mathbb{E}(\mathbb{E}(X_{n+2}|\mathcal{F}_{n+1})|\mathcal{F}_n) \\ &= \mathbb{E}(aX_{n+1} + b|\mathcal{F}_n) \\ &= a\mathbb{E}(X_{n+1}|\mathcal{F}_n) + b \\ &= a^2X_n + ab + b. \end{aligned}$$

On voit alors par récurrence que pour $h \geq 1$,

$$\mathbb{E}(X_{n+h}|\mathcal{F}_n) = a^h X_n + a^{h-1}b + \dots + ab + b.$$

Au passage on pourra remarquer l'égalité $\mathcal{F}_n = \sigma(U_1, \dots, U_n)$: en effet, pour $i = 1, \dots, n$, X_i est une fonction de U_i, \dots, U_1 et inversement on a $U_i = X_i - aX_{i-1} - b$.

Mentionnons enfin deux dernières propriétés utiles en pratique aussi bien pour le calcul de l'espérance conditionnelle que pour celui des lois conditionnelles.

Proposition 32 *On considère un couple (X, Y) de variables aléatoires pour lequel nous avons défini la notion de loi conditionnelle.*

1. *On suppose que $X = F(U, Y)$ où F est mesurable et U est une variable aléatoire indépendante de Y . Posons $g(y) = \mathbb{E}(F(U, y))$. Alors $g(Y) = \mathbb{E}(X|Y)$. On a même plus : la loi conditionnelle de $X|Y = y$ est aussi la loi de $F(U, y)$.*
2. *Si ϕ est une fonction mesurable et telle que $\phi(X)$ soit intégrable alors*

$$\mathbb{E}(\phi(X)|Y = y) = \int \phi(x) f_{X|Y}(x|y) d\mu(x).$$

Le point 2. est une sorte de théorème de transfert conditionnel : si on connaît la loi conditionnelle de $X|Y = y$ il n'est pas nécessaire de connaître celle de $\phi(X)|Y = y$ pour calculer l'espérance conditionnelle correspondante, il suffit d'intégrer ϕ par rapport à la loi conditionnelle de $X|Y = y$.

Exemples de calculs

- Considérons un modèle autorégressif de la forme

$$X_{n+1} = f(X_n, U_{n+1}), \quad n \in \mathbb{N}$$

avec U une suite i.i.d, f une fonction connue et $X_0 = x_0$ est une constante. On voit que X_n s'écrit comme une fonction des variables aléatoires U_n, U_{n-1}, \dots, U_1 . U_{n+1} est donc une variable aléatoire indépendante de X_n et même indépendante de $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$. D'après la proposition précédente la loi de $X_{n+1}|X_n = x$ coïncident avec la loi de $f(x, U_{n+1})$ (remarquer qu'il s'agit aussi de la loi de $X_{n+1}|X_n, \dots, X_1$). Prenons le cas où $f(x, y) = y\sqrt{1 + ax^2}$ et la suite U est une suite de gaussiennes $\mathcal{N}(0, 1)$: il s'agit d'un exemple de modèle ARCH (autorégressif et conditionnellement hétéroscédastique). Dans ce cas, la loi de $X_{n+1}|X_n = x$ est la loi $\mathcal{N}(0, 1 + ax^2)$. On remarquera également que $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = 0$ ici. En revanche, on peut regarder $\mathbb{E}(X_{n+1}^2|\mathcal{F}_n)$: un calcul à l'aide des propriétés de la loi conditionnelle montre que

$$\mathbb{E}(X_{n+1}^2|\mathcal{F}_n) = (1 + aX_n^2) \mathbb{E}(U_{n+1}^2|\mathcal{F}_n) = (1 + aX_n^2) \mathbb{E}(U_{n+1}) = 1 + aX_n^2.$$

On aurait aussi pu obtenir directement ce résultat en utilisant le point 2. de la proposition précédente avec $\phi(z) = z^2$ et en utilisant la loi conditionnelle $\mathcal{N}(0, 1 + ax^2)$.

- Supposons par exemple qu'une variable aléatoire X s'écrit sous la forme $X = \exp(YU)$ avec Y et U deux variables aléatoires indépendantes et à valeurs réelles et à densité avec par exemple U de loi uniforme sur $[0, 1]$. Calculons $\mathbb{E}(X|Y)$. La proposition précédente assure que $\mathbb{E}(X|Y) = g(Y)$ où g est la fonction définie par

$$g(y) = \mathbb{E}(\exp(yU)) = \int_0^1 \exp(yu) du = \frac{\exp(y) - 1}{y},$$

en convenant que le quotient vaille 1 si $y = 0$. On obtient alors la formule

$$\mathbb{E}(X|Y) = \frac{\exp(Y) - 1}{Y}.$$

- Retrouvons par le conditionnement un résultat déjà obtenu. Si $X = YX_1 + (1 - Y)X_2$ avec X_1, X_2 et Y indépendantes, $Y \sim \mathcal{B}(p)$ et X_i à densité f_i sur \mathbb{R} pour $i = 1, 2$, alors on voit que d'après la proposition précédente, la loi de $X|Y = 1$ est la loi de X_1 et que la loi de $X|Y = 0$ est la loi de X_2 .

Preuve partielle du point 1 de la Proposition 32. Prouvons l'assertion sur l'espérance conditionnelle. Encore une fois, il suffit de vérifier que la variable aléatoire $g(Y)$ satisfait (6.2). Remarquons déjà que $g(Y)$ est intégrable : en effet le théorème de transfert et le théorème de Fubini assure que

$$\mathbb{E}(|X|) = \int |F(u, y)| d\mathbb{P}_Y(y) d\mathbb{P}_U(u) = \int \mathbb{E}(|F(U, y)|) d\mathbb{P}_Y(y)$$

et est un nombre fini par hypothèse sur X . Ceci montre que $g(Y)$ est intégrable. Aussi pour un événement $\{Y \in A\} \in \sigma(Y)$, nous avons

$$\begin{aligned} \mathbb{E}(X \mathbf{1}_A(Y)) &= \int F(u, y) \mathbf{1}_A(y) d\mathbb{P}_U(u) d\mathbb{P}_Y(y) \\ &= \int_A g(y) d\mathbb{P}_Y(y) \\ &= \mathbb{E}(\mathbf{1}_A(Y) g(Y)). \end{aligned}$$

L'égalité (6.2) est bien vérifiée. \square

Chapitre 7

Les lois gaussiennes

7.1 Lois gaussiennes sur \mathbb{R}

Nous avons déjà défini les lois gaussiennes sur \mathbb{R} lors du chapitre sur les variables aléatoires à densité. On rappelle qu'une variable aléatoire X à valeurs réelles suit une loi gaussienne de moyenne $m \in \mathbb{R}$ et de variance $\sigma^2 > 0$ si la densité de X est donnée par

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

On inclut aussi dans la définition le cas dégénérée où $\sigma = 0$: dans ce cas on convient que la variable aléatoire X est presque sûrement égale à sa moyenne m (i.e la loi de X est la masse de Dirac δ_m).

Dans tous les cas, on notera $X \sim \mathcal{N}(m, \sigma^2)$. On rappelle que lorsque $m = 0$ et $\sigma = 1$, on parle de loi gaussienne centrée réduite.

Une propriété importante de ces lois est leur stabilité vis à vis des transformations affines. La proposition suivante se prouve aisément en effectuant un changement de variable.

Proposition 33 Soient $m \in \mathbb{R}$ et σ un réel positif.

Si $X \sim \mathcal{N}(0, 1)$ alors $Y = m + \sigma X \sim \mathcal{N}(m, \sigma^2)$.

Inversement, si $\sigma > 0$ et $Y \sim \mathcal{N}(m, \sigma^2)$, alors $\frac{Y-m}{\sigma} \sim \mathcal{N}(0, 1)$.

Lorsque nous avons introduit les fonctions caractéristiques, nous avons calculé celle de la loi $\mathcal{N}(m, \sigma^2)$. Si $X \sim \mathcal{N}(m, \sigma^2)$, alors

$$\phi_X(t) = \exp(itm) \exp\left(-\frac{\sigma^2 t^2}{2}\right), \quad t \in \mathbb{R},$$

cette expression étant aussi valable lorsque $\sigma = 0$. L'utilisation des fonctions caractéristiques et de leurs propriétés permet de prouver facilement la proposition suivante (cf TD).

Proposition 34 Une somme de variables aléatoires indépendantes et toutes de loi gaussienne sur \mathbb{R} suit encore une loi gaussienne sur \mathbb{R} .

Ainsi si $X = X_1 + \dots + X_n$ est une somme de variables aléatoires indépendantes telle que $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$ pour $1 \leq i \leq n$ alors $X \sim \mathcal{N}(m, \sigma^2)$ avec nécessairement $m = m_1 + \dots + m_n$ et $\sigma^2 = \sigma_1^2 + \dots + \sigma_n^2$.

7.2 Les matrices de variance-covariance

7.2.1 Quelques rappels sur les matrices

Soit A une matrice carrée de taille $d \times d$ dont les entrées sont des nombres réels. On notera $\det(A)$ son déterminant. Le noyau de A est le sous-espace vectoriel de \mathbb{R}^d

$$\text{Ker}(A) = \left\{ x \in \mathbb{R}^d : Ax = 0 \right\}$$

(on conviendra de noter les vecteurs de \mathbb{R}^d par des matrices colonnes). Alors on les équivalences fondamentales suivantes :

A inversible $\Leftrightarrow \det(A) \neq 0 \Leftrightarrow \text{Ker}(A) = \{0\}$.

Ainsi dire que A n'est pas inversible signifie qu'il existe $x \in \mathbb{R}^d \setminus \{0\}$ tel que $Ax = 0$.

On dit que A est matrice symétrique lorsque $A_{ij} = A_{ji}$ pour $1 \leq i, j \leq d$.

Dans la suite nous noterons \mathcal{S}_d^+ l'ensemble des matrices carrées A de taille $d \times d$ qui sont semi-définies positives c'est à dire telles que $x^T Ax \geq 0$ pour tout $x \in \mathbb{R}^d$. On notera également \mathcal{S}_d^{++} l'ensemble des matrices carrées A de taille $d \times d$ qui sont symétriques et définies positives (c'est-à-dire telles que $x^T Ax > 0$ pour $x \in \mathbb{R}^d \setminus \{0\}$).

On rappelle que toute matrice symétrique est diagonalisable dans une base orthonormée. Ceci signifie qu'il existe une matrice P de taille $d \times d$ telle que $P^T P = P P^T = I_d$ (matrice identité) et $A = P D P^T$ où D est une matrice diagonale dont les éléments diagonaux sont des nombres réels appelés les valeurs propres de A . Dans ce cas, on a $A \in \mathcal{S}_d^+$ (resp. $A \in \mathcal{S}_d^{++}$) si et seulement si les valeurs propres notées $\lambda_1, \dots, \lambda_d$ (avec éventuellement des répétitions) sont toutes positives (resp. strictement positives). Ainsi on a $A \in \mathcal{S}_d^{++}$ si et seulement si $A \in \mathcal{S}_d^+$ et $\det(A) \neq 0$.

Toute matrice A de \mathcal{S}_d^+ admet une unique racine carrée dans \mathcal{S}_d^+ (c'est-à-dire une matrice B telle que $B^2 = A$) : cette racine carrée sera notée $A^{1/2}$. Lorsqu'on parlera de la racine carrée de A , il s'agira de $A^{1/2}$. Lorsque $A \in \mathcal{S}_d^{++}$, $A^{1/2}$ appartient aussi à \mathcal{S}_d^{++} . Si $A = P D P^T$ on a $A^{1/2} = P D^{1/2} P^T$ et $D^{1/2}$ est la matrice diagonale dont les éléments diagonaux sont donnés par les racines carrées des valeurs propres $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}$. Enfin lorsque $A \in \mathcal{S}_d^{++}$ alors $A^{-1} \in \mathcal{S}_d^{++}$ et de plus la racine carrée de A^{-1} est l'inverse de la racine carrée de A et sera notée $A^{-1/2}$: on

$$A^{-1/2} = P \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \dots \\ 0 & \ddots & 0 \\ 0 & \dots & \frac{1}{\sqrt{\lambda_d}} \end{pmatrix} \text{ lorsque } A = P D P^T.$$

7.2.2 Vecteurs aléatoires, variance et covariance

Soit X une variable aléatoire à valeurs dans \mathbb{R}^d (on parle aussi de vecteur aléatoire) pour $d \in \mathbb{N}^*$. Si $X = (X_1, \dots, X_d)^T$ est tel que $\mathbb{E}(X_i^2) < +\infty$ pour $i = 1, \dots, d$, alors on définit la matrice de covariance de X par

$$\text{Var}(X) = [\text{Cov}(X_i, X_j)]_{1 \leq i, j \leq d}.$$

Remarquer que si les coordonnées de X sont indépendantes (ou même décorréélées deux à deux ce qui est plus faible) alors $\text{Var}(X)$ est une matrice diagonale. Dans la suite, nous conviendrons que si K est une matrice $p \times n$ dont les entrées définissent des variables aléatoires $K_{i,j}$ alors l'espérance $\mathbb{E}(K)$ est la matrice $p \times n$ telle que $\mathbb{E}(K)_{ij} = \mathbb{E}(K_{ij})$ (on prend l'espérance de chacune des entrées de la matrice). Nous avons alors les propriétés fondamentales suivantes.

Proposition 35 1. On a les expressions

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

2. Si A est une matrice de nombres réels de taille $n \times d$, alors le vecteur aléatoire $Y = AX$ vérifie

$$\mathbb{E}(Y) = A\mathbb{E}(X), \quad \text{Var}(Y) = A \text{Var}(X)A^T.$$

3. si le déterminant de $\text{Var}(X)$ vérifie $\det(\text{Var}(X)) = 0$ si et seulement si il existe $x \in \mathbb{R}^d \setminus \{0\}$ et un réel λ tel que $\sum_{i=1}^d x_i X_i = \lambda$ presque sûrement.
4. Toute matrice de covariance est symétrique et semi-définie positive (i.e $x^T \text{Var}(X)x \geq 0$ pour tout $x \in \mathbb{R}^d$). Inversement pour toute matrice Γ symétrique et semi-définie positive de taille d , il existe un vecteur aléatoire X tel que $\text{Var}(X) = \Gamma$.

Preuve.

- Il s'agit d'une simple réécriture matricielle.
- Soit $i \in \{1, \dots, n\}$. Alors, on a

$$Y_i = \sum_{j=1}^d A_{ij} X_j.$$

Par linéarité de l'espérance on obtient

$$\mathbb{E}(Y_i) = \sum_{j=1}^d A_{ij} \mathbb{E}(X_j).$$

On a donc bien $\mathbb{E}(Y) = A\mathbb{E}(X)$. De plus, le calcul de la variance montre que

$$\text{Var}(Y_i) = \sum_{j,j'=1}^d A_{ij} A_{ij'} \text{Cov}(X_j, X_{j'}) = \sum_{j,j'=1}^d A_{ij} \text{Var}(X)_{jj'} A_{j'i}^T.$$

$\text{Var}(Y_i)$ coïncide bien avec la i -ième coordonnée de $A \text{Var}(X)A^T$.

3. Supposons d'abord que $\det(\text{Var}(X)) = 0$. Dans ce cas, il existe $x \in \mathbb{R}^d \setminus \{0\}$ tel que $\text{Var}(X)x = 0$ (le noyau de la matrice n'est pas réduit à 0) donc tel que $x^T \text{Var}(X)x = 0$. Mais comme $x^T \text{Var}(X)x = \text{Var}(x^T X)$ d'après la point précédent, on en déduit que la variable aléatoire $x^T X = \sum_{i=1}^d x_i X_i$ est presque sûrement égale à sa moyenne notée λ . Inversement, supposons qu'il existe $x \in \mathbb{R}^d \setminus \{0\}$ et un réel λ tel que $\sum_{i=1}^d x_i X_i = \lambda$ presque sûrement. Alors si $1 \leq j \leq d$,

$$\sum_{i=1}^d x_i \text{Var}(X)_{ij} = \text{Cov}\left(\sum_{i=1}^d x_i X_i, X_j\right) = \text{Cov}(\lambda, X_j) = 0.$$

Comme $\text{Var}(X)_{ij} = \text{Var}(X)_{ji}$, on obtient $(\text{Var}(X)x)_j = 0$. Ainsi $x \neq 0$ appartient au noyau de $\text{Var}(X)$ ce qui entraîne que $\det(\text{Var}(X)) = 0$.

4. Le fait que $\text{Var}(X)$ soit symétrique résulte de sa définition. De plus cette matrice est semi-définie positive car $x^T \text{Var}(X)x = \text{Var}(x^T X)$ d'après ce qui précède. Inversement toute matrice symétrique étant diagonalisable dans une base orthonormée, on a

$$\Gamma = P \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{pmatrix} P^T = PDP^T,$$

où $P^T P = P P^T = I_d$ (matrice identité de taille d). De plus les valeurs propres sont positives (car la matrice est semi-définie positive) et en posant $Q = P D^{1/2}$ (où $D^{1/2}$ est

la matrice diagonale formée par $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}$, on obtient $\Gamma = QQ^T$. Ainsi si Z est un vecteur aléatoire de taille d et dont les entrées sont indépendantes et de variance 1 (en particulier $\text{Var}(Z) = I_d$), alors $X = QZ$ vérifie $\text{Var}(X) = \Gamma$. \square

On définit également la covariance entre deux vecteurs aléatoires X et Y à valeurs respectives dans \mathbb{R}^d et \mathbb{R}^n : il s'agit de la matrice notée (X, Y) de taille (p, n) et dont le terme (i, j) est donnée par $\text{Cov}(X_i, Y_j)$. On a alors les expressions matricielles :

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

Les propriétés suivantes sont immédiates.

$$\text{Cov}(Y, X) = \text{Cov}(X, Y)^T, \quad \text{Cov}(AX, Y) = A \text{Cov}(X, Y), \quad \text{Cov}(X, BY) = \text{Cov}(X, Y)B^T,$$

si A et B sont deux matrices à d colonnes et n colonnes respectivement. Enfin, si $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$ (vecteur aléatoire de dimension $d + n$), on remarquera que la matrice de covariance de Z peut être définie par blocs :

$$\text{Var}(Z) = \begin{pmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Var}(Y) \end{pmatrix}.$$

7.3 Les vecteurs gaussiens

Définition 23 *Un vecteur aléatoire X de \mathbb{R}^d est appelé un vecteur gaussien si pour tout $u \in \mathbb{R}^d$, la variable aléatoire $u^T X$ suit une loi gaussienne sur \mathbb{R} .*

En d'autres termes, un vecteur aléatoire X est gaussien si et seulement si toute combinaison linéaire de ses coordonnées est une variable aléatoire gaussienne dans \mathbb{R} . En particulier les variables aléatoires coordonnées X_1, \dots, X_d suivent des lois gaussiennes sur \mathbb{R} . Comme pour les lois gaussiennes sur \mathbb{R} , les transformations affines préserve le caractère gaussien.

Proposition 36 *Si X est un vecteur gaussien à valeurs dans \mathbb{R}^d , A une matrice de taille $n \times d$ à entrées réelles et $b \in \mathbb{R}^n$ alors $Y = AX + b$ est un vecteur gaussien de \mathbb{R}^n .*

Preuve. Si $u \in \mathbb{R}^n$, on a $u^T Y = v^T X + \lambda$ où on a posé $v = A^T u$ et $\lambda = u^T b$. Par hypothèse $v^T X$ suit une loi gaussienne sur \mathbb{R} et donc $u^T Y$ aussi d'après les propriétés d'invariance des lois gaussiennes sur \mathbb{R} . \square

Proposition 37 *Soit X un vecteur gaussien de \mathbb{R}^d . Notons $\mathbb{E}(X) = m$ et $\text{Var}(X) = \Gamma$. Alors la fonction caractéristique de X a l'expression suivante :*

$$\phi_X(t) = \exp(im^T t) \exp\left(-\frac{1}{2}t^T \Gamma t\right), \quad t \in \mathbb{R}^d.$$

Étant donné que la fonction caractéristique caractérise complètement la loi, on voit que la loi d'un vecteur gaussien dépend uniquement de sa moyenne et de sa matrice de covariance. Ainsi la loi d'un vecteur gaussien de \mathbb{R}^d de moyenne m et de variance Γ sera notée $\mathcal{N}_d(m, \Gamma)$.

Preuve. Il suffit d'observer que pour $t \in \mathbb{R}^d$, on a $\phi_X(t) = \phi_{t^T X}(1)$. Comme la variable aléatoire $t^T X$ suit une loi gaussienne sur \mathbb{R} , dont la moyenne est donnée par $t^T m$ et dont la matrice de covariance vaut $\text{Var}(X) = t^T \Gamma t$, le résultat est obtenu en utilisant l'expression de la fonction caractéristique des lois gaussiennes sur \mathbb{R} . \square

Proposition 38 *Soit X un vecteur gaussien de \mathbb{R}^d , de moyenne m et de variance Γ inversible (on dit alors que le vecteur est non dégénéré). Alors X admet une densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d et qui est donnée par*

$$f_X(x) = \frac{\exp\left(-\frac{1}{2}(x-m)^T \Gamma^{-1}(x-m)\right)}{(2\pi)^{\frac{d}{2}} \sqrt{\det(\Gamma)}}, \quad x \in \mathbb{R}^d.$$

Preuve. Commençons pour le cas $m = 0$ et $\Gamma = I_d$. Dans ce cas la fonction caractéristique de X est donnée par $\phi_X(t) = \exp\left(-\frac{t^T t}{2}\right)$, pour $t \in \mathbb{R}^d$, ce qui correspond à la fonction caractéristique de d variables aléatoires indépendantes toutes de loi $\mathcal{N}(0, 1)$. On en déduit que X admet une densité sur \mathbb{R}^d qui est définie par

$$f_X(x) = \frac{\exp\left(-\frac{1}{2}x^T x\right)}{(2\pi)^{\frac{d}{2}}}, \quad x \in \mathbb{R}^d.$$

ce qui est bien l'expression annoncée.

Passons maintenant au cas général. La matrice symétrique Γ est inversible, elle est alors définie positive et admet une unique racine carrée notée $\Gamma^{1/2}$ symétrique et définie positive. Ainsi si $X \sim \mathcal{N}_d(m, \Gamma)$ alors X a la même loi que le vecteur $Y = m + \Gamma^{1/2}Z$ où $Z \sim \mathcal{N}_d(0, I_d)$ (car ce sont deux vecteurs gaussiens qui ont même moyenne et même variance). Il suffit donc de calculer la densité du vecteur Y . Si $h : \mathbb{R}^d \rightarrow \mathbb{R}$ est une fonction mesurable bornée, on a

$$\mathbb{E}(h(Y)) = \int h\left(m + \Gamma^{1/2}x\right) \frac{\exp\left(-\frac{1}{2}x^T x\right)}{(2\pi)^{\frac{d}{2}}} dx,$$

en utilisant le théorème de transfert et la première partie de la preuve. Pour montrer que l'on obtient bien la densité annoncée, il suffit d'effectuer le changement de variable $y = m + \Gamma^{1/2}x$. On a alors $x = \Gamma^{-1/2}(y - m) = \phi(y)$ et le jacobien est donné par

$$J\phi(y) = \det\left(\Gamma^{-1/2}\right) = \frac{1}{\sqrt{\det(\Gamma)}}$$

. Des calculs élémentaires conduisent alors à l'expression de la densité. \square

Remarque sur les vecteurs gaussiens dégénérés. Reprenons les notations de la proposition précédente. Lorsque Γ n'est pas inversible, on a vu que le vecteur X appartenait à un hyperplan affine de \mathbb{R}^d avec probabilité 1 (on parle de vecteur gaussien dégénéré). Dans ce cas, le vecteur X ne peut pas avoir de densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d car un hyperplan affine est de mesure nulle (pour λ_d).

Indépendance et décorrélation. Supposons que les entrées d'un vecteur gaussien X soient décorrélées (i.e. $\text{Cov}(X_i, X_j) = 0$ pour $i \neq j$) et que $\text{Var}(X_i) = \sigma_i^2 \geq 0$ pour $i = 1, \dots, d$. Dans ce cas, la fonction caractéristique est donnée par

$$\phi_X(t) = \prod_{j=1}^d \exp(it_j m_j) \exp\left(-\frac{\sigma_j^2 t_j^2}{2}\right) = \prod_{j=1}^d \phi_{X_j}(t_j).$$

D'après les propriétés des fonctions caractéristiques, on voit que les variables aléatoires X_1, \dots, X_d sont indépendantes. Cette équivalence entre décorrélation et indépendance est une propriété typique des vecteurs gaussiens et sera généralisée un peu plus loin dans ce chapitre.

De la loi $\mathcal{N}_d(0, I_d)$ à la loi $\mathcal{N}_d(m, \Gamma)$. Soit $m \in \mathbb{R}^d$ et Γ une matrice symétrique semi-définie positive. Alors on peut voir qu'un vecteur $Y \sim \mathcal{N}_d(m, \Gamma)$ a la même loi que le vecteur $Z = m + \Gamma^{1/2}X$ où $X \sim \mathcal{N}_d(0, I_d)$. Cette propriété a déjà été évoquée dans la preuve de la proposition précédente. Lorsque Γ est inversible, on peut aussi voir que le vecteur $X = \Gamma^{-1/2}(Y - m)$ suit la loi $\mathcal{N}_d(0, I_d)$.

Des entrées gaussiennes ne forment pas toujours un vecteur gaussien. On a vu que si X était un vecteur gaussien alors les variables aléatoires coordonnées X_1, \dots, X_d étaient gaussiennes. En revanche, un vecteur aléatoire dont toutes les entrées suivent des lois gaussiennes sur \mathbb{R} n'est pas forcément un vecteur gaussien (**sauf si X_1, \dots, X_d sont indépendantes** auquel cas la fonction caractéristique permet de conclure). Par exemple, si $X_1 \sim \mathcal{N}(0, 1)$ et V est une variable aléatoire indépendante de X et telle que $\mathbb{P}(V = 1) = \mathbb{P}(V = -1) = \frac{1}{2}$ alors en posant $X_2 = VX_1$, on peut vérifier que $X_2 \sim \mathcal{N}(0, 1)$. Mais $X = (X_1, X_2)^T$ ne peut être un vecteur gaussien car la somme $X_1 + X_2 = (1 + V)X_1$ peut valoir 0 avec probabilité $\frac{1}{2}$ (la somme ne suit donc pas une loi gaussienne).

Proposition 39 Soit $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$ un vecteur gaussien avec X et Y deux vecteurs aléatoires (forcément gaussiens) à valeurs respectives dans \mathbb{R}^d et \mathbb{R}^n . Alors les vecteurs aléatoires X et Y sont indépendants si et seulement si $\text{Cov}(X, Y) = 0$. En d'autres termes, pour un vecteur gaussien, l'indépendance de deux sous-vecteurs équivaut à leur décorrélation.

Preuve. Posons $m = d + n$.

Supposons d'abord que X et Y vérifie $\text{Cov}(X, Y) = 0$. Pour montrer leur indépendance, il suffit d'utiliser le critère d'indépendance basé sur les fonctions caractéristiques. Remarquons que la variance Γ de Z est donnée par blocs par $\text{Var}(Z) = \begin{pmatrix} \text{Var}(X) & 0_{dn} \\ 0_{nd} & \text{Var}(Y) \end{pmatrix}$. Soit $t = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^m$. Alors

$$t^T \text{Var}(Z)t = u^T \text{Var}(X)u + v^T \text{Var}(Y)v.$$

La fonction caractéristique de Z est alors donnée par

$$\phi_Z(t) = \exp(i(u^T \mathbb{E}(X) + v^T \mathbb{E}(Y))) \exp\left(-\frac{1}{2}(u^T \text{Var}(X)u + v^T \text{Var}(Y)v)\right) = \phi_X(u)\phi_Y(v).$$

Les vecteurs aléatoires X et Y sont donc bien indépendants.

Inversement si X et Y sont indépendants, alors les variables aléatoires X_i et Y_j sont indépendantes pour $1 \leq i \leq d$ et $1 \leq j \leq n$ et donc $\langle X_i, Y_j \rangle = 0$. On a donc $\langle X, Y \rangle = 0$. \square

Une propriété fondamentale des lois gaussiennes sur \mathbb{R}^d est leur stabilité vis à vis du conditionnement. La proposition suivante est fondamentale pour le calcul des lois conditionnelles dans un vecteur gaussien. On retiendra surtout l'idée de la preuve car elle permet de retrouver facilement des formules un peu complexes.

Proposition 40 Soit $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$ un vecteur gaussien avec X un vecteur aléatoire à valeurs dans \mathbb{R}^d et Y un vecteur aléatoire à valeurs dans \mathbb{R}^n . On suppose que Y est non dégénérée (i.e

$\text{Var}(Y)$ est inversible). Alors la loi conditionnelle de $X|Y = y$ est une loi gaussienne $\mathcal{N}_d(m, \Gamma)$ dont les paramètres sont donnés par

$$\begin{cases} m = \mathbb{E}(X) + \text{Cov}(X, Y) \text{Var}(Y)^{-1} (y - \mathbb{E}(Y)) \\ \Gamma = \text{Var}(X) - \text{Cov}(X, Y) \text{Var}(Y)^{-1} \text{Cov}(Y, X) \end{cases}$$

Preuve. La preuve peut se décomposer en trois étapes.

- On commence par chercher une matrice C de taille $d \times n$ telle que le vecteur $U = X - CY$ soit décorrélé avec Y (et donc indépendant car le vecteur formé à l'aide de U et de Y est un vecteur gaussien en tant qu'image du vecteur Z par une application linéaire). On a

$$0 = \text{Cov}(U, Y) = \text{Cov}(X, Y) - \text{Cov}(CY, Y) = \text{Cov}(X, Y) - C \text{Var}(Y).$$

On trouve alors $C = \text{Cov}(X, Y) \text{Var}(Y)^{-1}$.

- Comme $X = U + CY$ est une fonction de deux variables aléatoires indépendantes, les propriétés sur les lois conditionnelles vues au chapitre précédent montre que la loi de $X|Y = y$ coïncide avec la loi de la variable aléatoire $U + Cy = X + C(y - Y)$. Il s'agit donc bien d'une loi gaussienne.
- On détermine la moyenne et la variance de la loi gaussienne obtenue. On a

$$\mathbb{E}(U + Cy) = \mathbb{E}(X) + C(y - \mathbb{E}(Y)) = \mathbb{E}(X) + \text{Cov}(X, Y) \text{Var}(Y)^{-1} (y - \mathbb{E}(Y)),$$

ce qui est bien l'expression attendue.

Pour la variance, on a

$$\begin{aligned} & \text{Var}(U + Cy) \\ &= \text{Var}(U) \\ &= \text{Var}(X - CY) \\ &= \text{Var}(X) + \text{Var}(CY) - \text{Cov}(X, CY) - \text{Cov}(CY, X) \\ &= \text{Var}(X) + C \text{Var}(Y) C^T - \text{Cov}(X, Y) C^T - C \text{Cov}(Y, X) \\ &= \text{Var}(X) + \text{Cov}(X, Y) \text{Var}(Y)^{-1} \text{Cov}(Y, X) - 2 \text{Cov}(X, Y) \text{Var}(Y)^{-1} \text{Cov}(Y, X) \\ &= \text{Var}(X) - \text{Cov}(X, Y) \text{Var}(Y)^{-1} \text{Cov}(Y, X). \end{aligned}$$

La proposition est prouvée. \square

Remarque fondamentale. Sous les hypothèses de la proposition précédente avec $d = 1$, on voit que l'espérance conditionnelle $\mathbb{E}(X|Y)$ est une combinaison linéaire du type $\sum_{i=1}^n \alpha_i Y_i + \beta$ pour des nombres réels $\alpha_1, \dots, \alpha_n, \beta$.

Un exemple. Supposons que $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$ soit un vecteur gaussien de moyenne $m = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ et de matrice de covariance $\Gamma = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$. Le vecteur gaussien est non dégénéré. A priori, on peut aussi calculer la loi conditionnelle de $X|Y = y$ en utilisant le quotient des densités. Cependant, il faut déjà inverser la matrice Γ pour récupérer la densité de Z et effectuer ensuite quelques calculs. Appliquons la méthode de la preuve de la proposition précédente. L'égalité $\text{Cov}(U, Y) = \text{Cov}(X - cY, Y) = 0$ est réalisée pour $c = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} = -0.5$. Comme

$X = U - 0.5Y$, la loi de $X|Y = y$ est aussi la loi de $U - 0.5y = X - 0.5(y - Y)$. La moyenne de cette loi est donc $\mathbb{E}(U - 0.5y) = -1 - 0.5(y + 1) = -1.5 - 0.5y$. La variance est donnée par

$$\text{Var}(U - 0.5y) = \text{Var}(X + 0.5Y) = \text{Var}(X) + 0.25 \text{Var}(Y) + \text{Cov}(X, Y) = 1.25 - 0.5 = 0.75.$$

On obtient donc la loi $\mathcal{N}(-1.5 - 0.5y, 0.75)$. Ainsi $\mathbb{E}(X|Y) = -1.5 - 0.5Y$. On pourra retrouver à titre d'exercice ces résultats en utilisant la formule de la densité conditionnelle.

7.4 Quelques lois fondamentales pour la statistique

Nous donnons ici trois exemples de loi qui interviennent assez fréquemment en statistique et qui sont construites à partir de variables aléatoires X_1, X_2, \dots indépendantes et toutes de loi $\mathcal{N}(0, 1)$.

La loi du χ^2 . La loi de X_1^2 est une loi $\gamma(\frac{1}{2}, 2)$ (voir TD). Ainsi la somme $X_1^2 + X_2^2 + \dots + X_n^2$ est la convolée de n variables aléatoires toutes de loi γ et elle suit donc une loi $\gamma(\frac{n}{2}, 2)$. Cette loi est aussi appelée loi du χ^2 (prononcer khi-deux) à n degrés de liberté.

La loi de Student. Il s'agit de la loi de la variable aléatoire

$$T = \frac{X_{n+1}}{\sqrt{(X_1^2 + \dots + X_n^2)/n}}.$$

La loi de T correspond donc à la loi de $\frac{Z}{\sqrt{Y/n}}$ où les variables aléatoires Z et Y sont indépendantes et de lois respectives $\mathcal{N}(0, 1)$ et χ^2 à n degrés de liberté. On peut montrer que la loi de T a pour densité

$$t \mapsto \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

La loi de Fisher. La loi de $\frac{X_{n+1}^2 + \dots + X_{n+m}^2}{X_1^2 + \dots + X_n^2}$ est appelée loi de Fisher à m et n degrés de liberté.

À une renormalisation près, cette loi est aussi celle du quotient de deux variables aléatoires indépendantes suivant toutes les deux une loi du χ^2 , à m et n degrés de liberté respectivement. On peut alors montrer que la densité est donnée par

$$z \mapsto \mathbb{1}_{\mathbb{R}_+}(z) \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})} \frac{m^{\frac{m}{2}} n^{\frac{n}{2}} z^{\frac{m}{2}-1}}{(n + mz)^{\frac{m+n}{2}}}.$$

Chapitre 8

Convergence des suites de variables aléatoires

On se fixe donc un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. De plus nous utiliserons la notation $|\cdot|$ aussi bien pour la valeur absolue sur \mathbb{R} que pour la norme euclidienne sur \mathbb{R}^d , $d \geq 2$.

8.1 Comportement asymptotique d'une suite d'événements

8.1.1 Limite inférieure et supérieure

Pour étudier les problèmes asymptotiques liés à des suites d'événements, la notion de limite inférieure et de limite supérieure intervient fréquemment. Si (A_n) est suite d'éléments de \mathcal{A} , la limite inférieure est l'événement noté $\underline{\lim} A_n$ qui est constitué de l'ensemble des épreuves ω qui appartiennent à tous les A_n à partir d'un certain indice p (qui peut dépendre de ω). Par exemple, lorsque $\Omega = \mathbb{R}$, en posant $A_n = [\frac{1}{n}, 1]$ si $n \geq 1$, on voit que $\underline{\lim} A_n =]0, 1]$. Il existe une écriture ensembliste : on a

$$\liminf_n A_n = \bigcup_{p \in \mathbb{N}} \left(\bigcap_{n=p}^{+\infty} A_n \right).$$

En effet dire que $\omega \in \underline{\lim} A_n$ signifie qu'il existe un entier p tel que pour tout $n \geq p$, on ait $\omega \in A_n$.

La limite supérieure de la suite (A_n) est l'événement noté $\limsup_n A_n$ qui est composé des épreuves ω qui appartiennent à une infinité d'événements A_n . Par exemple, si

$$A_1 = [0, 1], \quad A_2 = [1, 2], \quad A_3 = [0, 1], \quad A_4 = [1, 2], \dots,$$

on voit que $\limsup_n A_n = [0, 2]$ alors que $\liminf_n A_n = \{1\}$. On a l'écriture ensembliste

$$\limsup_n A_n = \bigcap_{p \in \mathbb{N}} \left(\bigcup_{n=p}^{+\infty} A_n \right).$$

En effet dire que $\omega \in \bigcap_{p \in \mathbb{N}} \left(\bigcup_{n=p}^{+\infty} A_n \right)$ signifie que pour tout entier p , il existe un entier $n \geq p$ tel que $\omega \in A_n$.

De par la définition de ces ensembles, on voit aisément que $\liminf_n A_n \subset \limsup_n A_n$. De plus, on a les relations

$$\left(\liminf_n A_n \right)^c = \limsup_n A_n^c, \quad \left(\limsup_n A_n \right)^c = \liminf_n A_n^c,$$

qui peuvent se démontrer en utilisant les définitions ensemblistes de la limite inférieure ou supérieure (ou par le bon sens, par exemple ne pas être dans tous les A_n à partir d'un certain rang revient à être une infinité de fois dans leur complémentaire).

8.1.2 Lemme de Borel-Cantelli

Ce lemme est d'un usage courant lorsque on s'intéresse au comportement asymptotique d'une suite d'événements. Pour la suite, on rappelle qu'une suite d'événements indépendants est une suite $(A_n)_{n \in \mathbb{N}}$ telle que pour tout $k \in \mathbb{N}^*$ et tout k -uplet $(n_1, \dots, n_k) \in \mathbb{N}^k$, les événements A_{n_1}, \dots, A_{n_k} sont indépendants.

Lemme 4 Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements.

1. On a l'implication :

$$\sum_{n=0}^{+\infty} \mathbb{P}(A_n) < +\infty \Rightarrow \mathbb{P}\left(\limsup_n A_n\right) = 0.$$

2. Si on suppose que les événements A_n sont indépendants :

$$\sum_{n=0}^{+\infty} \mathbb{P}(A_n) = +\infty \Rightarrow \mathbb{P}\left(\limsup_n A_n\right) = 1.$$

Remarque. Ainsi pour des événements indépendants, on a

$$\sum_{n=0}^{+\infty} \mathbb{P}(A_n) < +\infty \Leftrightarrow \mathbb{P}\left(\limsup_n A_n\right) = 0.$$

Exemples. Donnons deux illustrations de ce lemme.

- Considérons une suite infinie de pile ou face, c'est-à-dire une suite i.i.d (X_n) de variables aléatoires toutes de loi de Bernoulli de paramètre $\frac{1}{2}$ ainsi qu'un entier $N \geq 1$. En utilisant le lemme de Borel-Cantelli, montrons qu'avec probabilité 1, on obtiendra une infinité de fois N piles consécutifs (on pourrait même montrer que l'on obtiendrait une infinité de fois toute séquence finie donnée). Pour cela posons pour $n \in \mathbb{N}$, $A_n = \{X_{nN} = 1, \dots, X_{(n+1)N-1} = 1\}$. La suite (A_n) est une suite d'événements indépendants

chacun étant de probabilité $\frac{1}{2^N}$. Donc $\sum_{n=0}^{+\infty} \mathbb{P}(A_n) = +\infty$. Le deuxième point du lemme de Borel-Cantelli assure qu'avec probabilité 1, toute réalisation appartiendra à une infinité d'événements A_n . Il y aura donc une infinité de fois N piles consécutifs.

- Maintenant supposons que $(X_n)_n$ soit une suite de variables aléatoires telles que pour tout $n \in \mathbb{N}$, X_n suive la loi de Bernoulli de paramètre 2^{-n} , alors les réalisations de cette suite ne comportent qu'un nombre fini de 1 presque sûrement. En effet posons $A_n = \{X_n = 1\}$. Alors

$$\sum_{n=0}^{+\infty} \mathbb{P}(A_n) = \sum_{n=0}^{+\infty} \frac{1}{2^n} < +\infty.$$

Le premier point du lemme de Borel-Cantelli assure qu'avec probabilité 1, une réalisation appartient au complémentaire de $\limsup_n A_n$ donc à $\liminf_n A_n^c$. Il n'y a donc que des 0 au bout d'un moment.

Preuve. Posons $A = \limsup_n A_n$.

1. Pour le premier point, remarquons que pour tout $n \in \mathbb{N}$, $A \subset \cup_{p \geq n} A_p$, ce qui donne les majorations :

$$\mathbb{P}(A) \leq \mathbb{P}\left(\cup_{p \geq n} A_p\right) \leq \sum_{p \geq n} \mathbb{P}(A_p),$$

ce qui donne le résultat car le membre de droite est le reste d'une série convergente et tend donc vers 0 lorsque n tend vers $+\infty$.

2. Pour le deuxième point, on peut se rappeler que $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \mathbb{P}(\liminf_n A_n^c)$. Il suffit donc de prouver que $\mathbb{P}(\liminf_n A_n^c) = 0$. En posant $B_n = \cap_{p \geq n} A_p^c$, on a

$$A^c = \left\{ \liminf_n A_n^c \right\} = \cup_{n \geq 0} B_n.$$

Les événements B_n forment une suite croissante d'événements pour l'inclusion, on a donc $\mathbb{P}(A^c) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n)$. Nous allons montrer que $\mathbb{P}(B_n) = 0$ ce qui permettra de conclure. Remarquons que en posant pour $q \geq n$, $C_q = \cap_{p=n}^q A_p^c$, on a $\mathbb{P}(B_n) = \lim_{q \rightarrow +\infty} \mathbb{P}(C_q)$, car les événements C_q forment une suite décroissante d'événements pour l'inclusion. En utilisant l'indépendance des événements A_n^c ainsi que l'inégalité $e^{-x} \geq 1 - x$, on obtient

$$\mathbb{P}(C_q) = \sum_{p=n}^q (1 - \mathbb{P}(A_p)) \leq e^{-\sum_{p=n}^q \mathbb{P}(A_p)}.$$

Ainsi $\lim_{q \rightarrow +\infty} \mathbb{P}(C_q) = 0$ et donc $\mathbb{P}(B_n) = 0$. \square

8.2 Les modes de convergence

8.2.1 Convergence presque sûre et en probabilité

Définition 24 On dit qu'une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires converge presque sûrement vers une variable aléatoire X si il existe $\tilde{\Omega} \in \mathcal{A}$ vérifiant $\mathbb{P}(\tilde{\Omega}) = 1$ et $\forall \omega \in \tilde{\Omega}$, $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$. Autrement dit, la convergence a lieu point par point en dehors d'un ensemble de mesure nulle.

En abrégé nous noterons $X_n \rightarrow X$ p.s. Il existe un mode de convergence qui est plus faible, la convergence en probabilité.

Définition 25 On dit qu'une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires à valeurs dans \mathbb{R}^d converge en probabilité vers une variable aléatoire X si pour tout $\epsilon > 0$,

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

Remarques

1. La convergence presque sûre entraîne la convergence en probabilité. En effet la convergence presque sûre entraîne que pour $\epsilon > 0$, $\mathbb{P}(\limsup_n A_n) = 0$ avec $A_n = \{|X_n - X| > \epsilon\}$. Comme

$$\mathbb{P}(A_n) \leq \mathbb{P}(\cup_{p \geq n} A_p)$$

et le membre de droite converge vers $\mathbb{P}(\limsup_q A_q) = 0$ lorsque n tend vers $+\infty$, on en déduit que $\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = 0$.

2. La convergence en probabilité n'entraîne pas la convergence presque sûre comme le montre l'exemple suivant. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes telles que $\mathbb{P}(X_n = 1) = \frac{1}{n} = 1 - \mathbb{P}(X_n = 0)$. Cette suite converge en probabilité vers 0 car si $0 < \epsilon < 1$,

$$\mathbb{P}(|X_n| > \epsilon) = \mathbb{P}(X_n = 1) = \frac{1}{n} \rightarrow_{n \rightarrow +\infty} 0.$$

Remarquons au passage que pour montrer la convergence en probabilité, on peut se restreindre à des valeurs de ϵ plus petite qu'un certain seuil dans la définition (car si $\epsilon < \epsilon'$, $\mathbb{P}(|X_n| > \epsilon') \leq \mathbb{P}(|X_n| > \epsilon)$). Remarquons ensuite que

$$\sum_{n=1}^{+\infty} \mathbb{P}(X_n = 1) = \sum_{n=1}^{+\infty} \frac{1}{n} = +\infty$$

et donc que $\mathbb{P}(\limsup_n \{X_n = 1\}) = 1$ en utilisant le deuxième point du lemme de Borel-Cantelli. Ainsi p.s, la suite $(X_n)_n$ prend une infinité de fois la valeur 1, elle ne peut donc converger vers 0 presque sûrement.

La convergence presque sûre ou en probabilité est stable par composition par une fonction continue.

Proposition 41 *Si $X_n \rightarrow X$ p.s (resp. en probabilité) et f une fonction continue, $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$, alors $f(X_n) \rightarrow f(X)$ p.s (resp. en probabilité).*

Preuve. C'est clair pour la convergence presque sûre. Pour la convergence en probabilité, soit $\epsilon > 0$ et $\delta > 0$. On va montrer qu'il existe $n_0 \in \mathbb{N}$ tel que $\forall n \geq n_0$, $\mathbb{P}(|f(X_n) - f(X)| > \epsilon) < \delta$. Choisissons d'abord $r > 0$ tel que $\mathbb{P}(|X| > r) \leq \delta/2$, ce qui est toujours possible. Sur $\{|x| \leq 2r\}$, f est uniformément continue. Choisissons alors $\eta > 0$ tel que

$$|x|, |y| \leq 2r, \quad |x - y| < \eta \quad \Rightarrow \quad |f(x) - f(y)| \leq \epsilon.$$

Remarquons alors que $|x| \leq r$ et $|x - y| \leq \min(r, \eta)$ entraîne $|f(x) - f(y)| \leq \epsilon$. Donc $|f(x) - f(y)| \leq \epsilon$ entraîne ou bien $|x| > r$ ou bien $|x - y| > \min(r, \eta)$. En notant $A_n = \{|f(X_n) - f(X)| > \epsilon\}$, nous obtenons

$$\mathbb{P}(A_n) \leq \mathbb{P}(|X| > r) + \mathbb{P}(|X_n - X| > \min(r, \eta)).$$

Le premier terme de droite est $< \delta/2$ et la convergence en probabilité de la suite (X_n) permet de trouver $n_0 \in \mathbb{N}$ tel que le deuxième terme soit $< \delta/2$ si $n \geq n_0$. Ainsi $\mathbb{P}(A_n) \leq \delta$ si $n \geq n_0$. Comme δ est arbitraire, on en déduit $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 0$. La convergence en probabilité de la suite $(f(X_n))_n$ vers $f(X)$ en découle. \square

Une condition suffisante pour la convergence presque sûre est donnée par la proposition suivante.

Proposition 42 *(Critère de convergence p.s)*

Si pour tout $\epsilon > 0$, $\sum_{n \geq 0} \mathbb{P}(|X_n - X| > \epsilon) < +\infty$ alors $X_n \rightarrow X$ p.s.

En particulier, si il existe $p > 0$ tel que $\sum_{n \geq 0} \mathbb{E}(|X_n - X|^p) < +\infty$, alors $X_n \rightarrow X$ p.s.

Exemple. Un exemple simple de suite qui converge p.s vers 0 est par exemple $X_n = \frac{Y}{n}$, $n \geq 1$ avec Y une variable aléatoire donnée. Ce critère permet de dire un peu plus en considérant la suite définie par $X_n = \frac{Y_n}{n}$ avec $(Y_n)_n$ une suite de variables aléatoires toutes de même loi. En effet si $\mathbb{E}(Y_1^2) < +\infty$, on a

$$\sum_{n=1}^{+\infty} \mathbb{P}(|Y_n| > n\epsilon) = \sum_{n=1}^{+\infty} \mathbb{P}(|Y_1| > n\epsilon) \leq \sum_{n=1}^{+\infty} \frac{\mathbb{E}(Y_1^2)}{\epsilon^2 n^2} < +\infty,$$

en utilisant l'inégalité de Markov, ce qui permet de conclure $X_n \rightarrow 0$ p.s. En fait ce résultat est valable si $\mathbb{E}(|Y_1|) < +\infty$ seulement en utilisant des majorations plus précises. On peut d'ailleurs prouver que dans le cas de variables aléatoires indépendantes :

$$X_n \rightarrow 0 \text{ p.s.} \Leftrightarrow \mathbb{E}(|Y_1|) < +\infty.$$

On trouvera une démonstration de ces résultats dans [1] (Exercice 9.12, p. 85).

Remarque. La proposition précédente fournit une condition suffisante de convergence presque sûre mais non nécessaire. Prenons $\Omega = [0, 1]$ muni de la tribu des boréliens et \mathbb{P} la probabilité uniforme sur Ω . Si on pose pour $n \in \mathbb{N} \setminus \{0\}$, $X_n = \mathbf{1}_{[0, \frac{1}{n}]}$ alors on voit que pour $\epsilon > 0$,

$$\mathbb{P}(|X_n| > \epsilon) = \mathbb{P}\left([0, \frac{1}{n}]\right) = \frac{1}{n}.$$

Ainsi $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > \epsilon) = +\infty$ alors que $(X_n)_{n \geq 1}$ converge presque sûrement vers 0 (prendre $\tilde{\Omega} =]0, 1]$ dans la définition de la convergence presque sûre).

Preuve de la proposition 42. Pour tout $\epsilon > 0$, posons $A_{n,\epsilon} = \{|X_n - X| > \epsilon\}$ et $A_\epsilon = \limsup_n A_{n,\epsilon}$. D'après le premier point du lemme de Borel-Cantelli, l'hypothèse $\sum_{n \geq 0} \mathbb{P}(|X_n - X| > \epsilon) < +\infty$ entraîne que $\mathbb{P}(A_\epsilon) = 0$. Ainsi $\mathbb{P}(\cup_{k \in \mathbb{N}^*} A_{1/k}) = 0$ par σ -sous-additivité. Par passage au complémentaire, on a

$$\mathbb{P}\left(\cap_{k \in \mathbb{N}^*} A_{1/k}^c\right) = \mathbb{P}\left(\cap_{k \in \mathbb{N}^*} \liminf_n A_{n,1/k}^c\right) = 1,$$

ce qui se lit

$$(\forall k \in \mathbb{N}^*, \exists p \text{ tel que } \forall n \geq p, |X_n - X| \leq 1/k), \quad \text{p.s.}$$

La convergence presque sûre de la suite (X_n) en découle.

Remarquons de plus que la condition $\sum_{n \geq 0} \mathbb{E}(|X_n - X|^p) < +\infty$ entraîne que pour tout $\epsilon > 0$,

$$\sum_{n \geq 0} \mathbb{P}(|X_n - X| > \epsilon) = \sum_{n \geq 0} \mathbb{P}(|X_n - X|^p > \epsilon^p) \leq \sum_{n \geq 0} \epsilon^{-p} \mathbb{E}(|X_n - X|^p) < \infty,$$

par application de l'inégalité de Markov. Ceci justifie le deuxième point. \square

Finissons ce paragraphe en donnant un dernier lien entre la convergence en probabilité et la convergence presque sûre.

Proposition 43 *Si $X_n \rightarrow X$ en probabilité, alors il existe une sous-suite $(X_{n_j})_j$ qui converge presque sûrement vers X .*

Preuve. Remarquons tout d'abord que pour $\epsilon > 0$ et $\delta > 0$, la convergence en probabilité entraîne l'existence d'un entier n_0 tel que $\forall n \geq n_0$, $\mathbb{P}(|X_n - X| > \epsilon/2) \leq \delta/2$. Remarquons également que si deux réels y, z vérifient $|y - z| > \epsilon$ alors si x est un réel, on a ou bien $|y - x| > \epsilon/2$ ou bien $|z - x| > \epsilon/2$. Ainsi si $p, q \geq n_0$, on a l'inclusion

$$\{|X_p - X_q| > \epsilon\} \subset \{|X_p - X| > \epsilon/2\} \cup \{|X_q - X| > \epsilon/2\}.$$

On en déduit

$$\mathbb{P}(|X_p - X_q| > \epsilon) \geq \mathbb{P}(|X_p - X| > \epsilon/2) + \mathbb{P}(|X_q - X| > \epsilon/2) \leq \delta.$$

Ainsi pour tout $\epsilon, \delta > 0$, il existe un entier n_0 tel que

$$\forall p, q \geq n_0, \quad \mathbb{P}(|X_p - X_q| > \epsilon) \leq \delta.$$

Ainsi, il est possible de construire une suite de nombre entiers $(n_j)_j$ avec $n_0 = 0$ et pour $j \in \mathbb{N}^*$,

$$n_j = \inf \left\{ n > n_{j-1} / \forall p, q \geq n, \quad \mathbb{P} \left(|X_p - X_q| > \frac{1}{2^j} \right) < \frac{1}{2^j} \right\}.$$

De plus cette suite tend en croissant vers $+\infty$. Ainsi

$$\sum_{j=0}^{+\infty} \mathbb{P} \left(|X_{n_{j+1}} - X_{n_j}| > \frac{1}{2^j} \right) \leq \sum_{j=0}^{+\infty} \frac{1}{2^j} < +\infty.$$

Le lemme de Borel-Cantelli assure alors que l'événement $\limsup_j \left\{ |X_{n_{j+1}} - X_{n_j}| > \frac{1}{2^j} \right\}$ est de probabilité 0 et donc que son complémentaire $B = \liminf_j \left\{ |X_{n_{j+1}} - X_{n_j}| \leq \frac{1}{2^j} \right\}$ est de probabilité 1. Mais si $\omega \in B$, la série de terme général $|X_{n_{j+1}}(\omega) - X_{n_j}(\omega)|$ est convergente donc la suite $(X_{n_j}(\omega))_j$ converge. La suite de variables aléatoires $(X_{n_j})_j$ converge p.s et la limite est forcément X car $(X_{n_j})_j$ converge en probabilité vers X . \square

8.2.2 La convergence en loi

Définition 26 On dit qu'une suite de variables aléatoires $(X_n)_n$ converge en loi vers une variable aléatoire X si pour toute fonction continue bornée $h : \mathbb{R}^d \rightarrow \mathbb{R}$,

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

Remarques.

1. Ce type de convergence ne concerne que la suite des lois des variables aléatoires et pas le comportement des trajectoires $n \rightarrow X_n(\omega)$. Par exemple toute suite de variables aléatoires de même loi converge en loi puisque la loi des variables est constante.
2. Contrairement à la convergence p.s ou en probabilité, la convergence en loi de $(X_n)_n$ vers X n'est pas équivalente à la convergence en loi de $(X_n - X)_n$ vers 0, comme le montre l'exemple $X_n = -X = Y$ où $Y \sim \mathcal{N}(0, 1)$.
3. Cette définition entraîne automatiquement que si $(X_n)_n$ converge en loi vers X et $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ est une application continue, alors la suite $(f(X_n))_n$ converge en loi vers $f(X)$.

Dans la suite on notera $X_n \xrightarrow{\mathcal{L}} X$ pour exprimer la convergence en loi. On peut alors montrer que la convergence en loi est la plus faible des trois convergences introduites jusqu'ici.

Proposition 44 La convergence en probabilité entraîne la convergence en loi.

Preuve. Soit une suite $(X_n)_n$ telle que $X_n \xrightarrow{\mathcal{L}} X$. Soit $h : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction continue bornée. Soit $\epsilon > 0$. Remarquons que pour $x, y \in \mathbb{R}^d$:

$$|h(x) - h(y)| \leq \epsilon + 2 \|h\|_\infty \mathbb{1}_{|h(x) - h(y)| > \epsilon}.$$

Ainsi

$$|\mathbb{E}(h(X_n)) - \mathbb{E}(h(X))| \leq \mathbb{E}(|h(X_n) - h(X)|) \leq \epsilon + 2 \|h\|_\infty \mathbb{P}(|h(X_n) - h(X)| > \epsilon).$$

La suite $(h(X_n))_n$ convergeant vers $h(X)$ en probabilité (d'après la proposition 41), on en déduit que pour n suffisamment grand

$$|\mathbb{E}(h(X_n)) - \mathbb{E}(h(X))| \leq \epsilon(1 + 2 \|h\|_\infty).$$

Ainsi $\lim_{n \rightarrow +\infty} \mathbb{E}(h(X_n)) = \mathbb{E}(h(X))$. D'où la convergence en loi. \square

Remarque : La réciproque de cette propriété est fautive, comme le montre l'exemple suivant. Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$ et $X_n = -X$ alors $(X_n)_n$ converge en loi vers X (toutes les variables suivent la loi $\mathcal{N}(0, 1)$) mais pas en probabilité vers X .

Pour les variables aléatoires à valeurs dans \mathbb{Z} , on a la caractérisation suivante :

Proposition 45 Si $(X_n)_n$ est une suite de variables aléatoires discrètes alors

$$X_n \xrightarrow{\mathcal{L}} X \iff \forall k \in \mathbb{Z}, \lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k).$$

Preuve. La démonstration de la condition nécessaire est laissée à titre d'exercice (on pourra considérer des fonctions continues h_k telles que $h_k(k) = 1$ et $h_k(x) = 0$ si $|x - k| > \frac{1}{2}$). Démontrons la condition suffisante. On suppose que pour tout $k \in \mathbb{Z}$, $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$. Soit $h : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue et bornée. Il faut montrer que

$$\lim_{n \rightarrow +\infty} \sum_{k \in \mathbb{Z}} h(k) \mathbb{P}(X_n = k) = \sum_{k \in \mathbb{Z}} h(k) \mathbb{P}(X = k). \quad (8.1)$$

Cette convergence est immédiate si on suppose l'existence de $N \in \mathbb{N}$ tel que $\mathbb{P}(X_n = k) = \mathbb{P}(X = k) = 0$ pour $|k| > N$ (convergence d'une somme finie de suites). Pour le cas général, on commence par choisir $N \in \mathbb{N}$ tel que $\|h\|_\infty \sum_{|k| > N} \mathbb{P}(X = k) < \epsilon$ (ce qui est toujours possible car on considère le reste d'une série convergente). On a alors

$$\left| \sum_{k \in \mathbb{Z}} h(k) \mathbb{P}(X_n = k) - \sum_{k \in \mathbb{Z}} h(k) \mathbb{P}(X = k) \right| \leq A_n + B_n,$$

avec

$$A_n = \left| \sum_{|k| \leq N} h(k) \mathbb{P}(X_n = k) - \sum_{|k| \leq N} h(k) \mathbb{P}(X = k) \right|,$$

et

$$B_n = \|h\|_\infty \sum_{|k| > N} \mathbb{P}(X_n = k) + \|h\|_\infty \sum_{|k| > N} \mathbb{P}(X = k).$$

Il est évident que $\lim_{n \rightarrow \infty} A_n = 0$. De plus en écrivant

$$\sum_{|k| > N} \mathbb{P}(X_n = k) = 1 - \sum_{|k| \leq N} \mathbb{P}(X_n = k),$$

on voit que

$$\lim_{n \rightarrow \infty} \sum_{|k| > N} \mathbb{P}(X_n = k) = \sum_{|k| > N} \mathbb{P}(X = k),$$

et donc que

$$\lim_{n \rightarrow +\infty} B_n = 2 \|h\|_\infty \sum_{|k| > N} \mathbb{P}(X = k) < 2\epsilon.$$

On en déduit l'existence d'un entier n_0 tel que si $n \geq n_0$, $A_n + B_n < 3\epsilon$. Comme ϵ peut être arbitrairement petit, on voit que

$$\lim_{n \rightarrow \infty} (A_n + B_n) = 0.$$

On a bien $\lim_{n \rightarrow \infty} \mathbb{E}(h(X_n)) = \mathbb{E}(h(X))$, ce qui montre la convergence en loi annoncée. \square

Remarque. Concernant les variables aléatoires X_n à densité f_{X_n} (par rapport à la mesure de Lebesgue sur \mathbb{R}^d), on peut montrer que si $\lim_{n \rightarrow +\infty} f_{X_n}(x) = f(x)$ en tout point $x \in \mathbb{R}^d$ avec f densité de probabilité sur \mathbb{R}^d , alors la suite $(X_n)_n$ converge en loi vers la loi de densité f . Ce résultat constitue le Lemme de Scheffé. Cependant la réciproque de ce lemme est fautive : $(X_n)_n$ peut converger en loi sans que la suite des densités converge point par point (un exemple de ce cas de figure est donnée dans [1], p. 319).

La convergence en loi peut s'exprimer à l'aide de la convergence des fonctions caractéristiques seulement. Nous admettrons le théorème suivant.

Théorème 12 $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si $\forall t \in \mathbb{R}^d$,

$$\lim_{n \rightarrow +\infty} \phi_{X_n}(t) = \phi_X(t).$$

Remarque. Il est facile de montrer que dans le cas de vecteurs aléatoires la convergence en loi entraîne la convergence en loi des marginales. Cependant la réciproque est fautive en général (sauf dans le cas où les marginales sont indépendantes). En fait, le théorème précédent permet de montrer l'équivalence

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \forall t \in \mathbb{R}^d, \quad t^T X_n \xrightarrow{\mathcal{L}} t^T X.$$

Mentionnons un dernier critère de convergence en loi très utile pour le cas des variables aléatoires à valeurs réelles. Nous admettrons le résultat suivant.

Proposition 46 Pour une suite de variables aléatoires réelles, on a $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si la suite de leur fonctions de répartition F_{X_n} satisfait $\lim_{n \rightarrow +\infty} F_{X_n}(t) = F_X(t)$ en tout point t de continuité de F_X .

Remarque. La convergence des fonctions de répartition a lieu en tout point si la loi limite est à densité. D'ailleurs si la fonction de répartition limite est continue, il est même possible de prouver que $\sup_{t \in \mathbb{R}} |F_{X_n}(t) - F_X(t)|$ tend vers 0 (c'est-à-dire que la convergence est uniforme). Cependant en général, on ne peut espérer avoir la convergence en tout point comme le montre l'exemple suivant. Si $X_n = \frac{1}{n}$, alors $X_n \rightarrow 0$ p.s. donc $X_n \xrightarrow{\mathcal{L}} 0$, alors que $F_{X_n}(0) = 0$, pour tout n et $F_X(0) = 1$.

8.3 Deux théorèmes fondamentaux

8.3.1 La loi des grands nombres

Théorème 13 Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles, indépendantes et identiquement distribuées telle que $\mathbb{E}(|X_1|) < +\infty$. Alors

$$\lim_{n \rightarrow +\infty} \frac{X_1 + X_2 + \dots + X_n}{n} = \mathbb{E}(X_1), \quad p.s.$$

Remarques

1. En considérant par exemple le jeu de pile ou face avec des lancers indépendants, on comprend pourquoi la convergence ne peut avoir lieu en tout point $\omega \in \Omega = \{0, 1\}^{\mathbb{N}}$, puisque pour toute suite constante à partir d'un certain rang, la limite ci dessus vaut 0 ou 1 alors que la moyenne d'un lancer est 1/2.
2. On peut aussi obtenir ce type de convergence pour certaines suites de variables aléatoires dépendantes (e.g certains processus de type autorégressifs vus dans ce cours).

Preuve. On va démontrer le résultat seulement lorsque $\mathbb{E}(X_1^4) < +\infty$. Pour cela posons

$$S_n = (X_1 + \dots + X_n) - n\mathbb{E}(X_1) = \sum_{i=1}^n (X_i - \mathbb{E}(X_1)).$$

Les variables $Y_i = X_i - \mathbb{E}(X_1)$ possèdent aussi un moment d'ordre 4 et elles sont indépendantes et centrées. Pour montrer que $\lim_{n \rightarrow +\infty} \frac{S_n}{n} = 0$ p.s. (qui est le résultat attendu), on va montrer que

$\sum_{n=1}^{+\infty} n^{-4} \mathbb{E}(S_n^4) < +\infty$, ce qui conduira au résultat, d'après le critère donné par la proposition

42. Nous avons si $n \geq 1$ (en convenant que $S_0 = 0$),

$$S_n^4 = S_{n-1}^4 + Y_n^4 + 4S_{n-1}^3 Y_n + 6S_{n-1}^2 Y_n^2 + 4S_{n-1} Y_n^3.$$

En prenant l'espérance dans cette dernière égalité, on obtient, vu que les variables aléatoires Y_n et S_{n-1} sont indépendantes et centrées :

$$\mathbb{E}(S_n^4) = \mathbb{E}(S_{n-1}^4) + \mathbb{E}(Y_n^4) + 6\mathbb{E}(Y_n^2) \mathbb{E}(S_{n-1}^2).$$

Remarquons que par indépendance des variables Y_i , $\mathbb{E}(S_{n-1}^2) = (n-1)\mathbb{E}(Y_1^2)$. Ainsi en posant $a = 6(\mathbb{E}(Y_1^2))^2$ et $b = \mathbb{E}(Y_1^4) - a$, nous avons :

$$\mathbb{E}(S_n^4) = \mathbb{E}(S_{n-1}^4) + an + b.$$

En itérant cette égalité, on obtient $\mathbb{E}(S_n^4) = \frac{n^2}{2}a + n(b + \frac{a}{2})$. Il est alors immédiat que

$$\sum_{n=1}^{+\infty} n^{-4} \mathbb{E}(S_n^4) < +\infty,$$

ce qui termine la preuve. \square

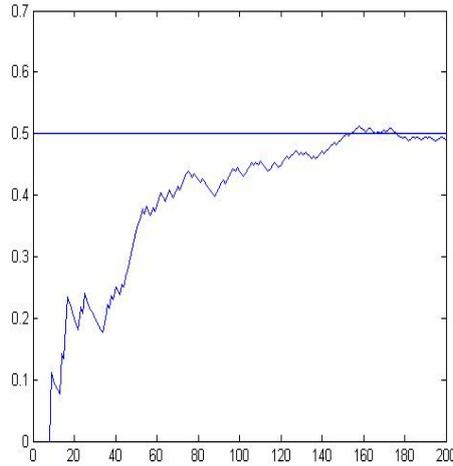


FIGURE 8.1: Convergence de $\frac{1}{n} \sum_{j=1}^n X_j$ pour une suite i.i.d de loi de Bernoulli ($p = \frac{1}{2}$).

8.3.2 Le théorème central limite

Théorème 14 Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes à valeurs dans \mathbb{R}^d , de même loi et de carré intégrable. Soit $\Gamma = \text{Var}(X_1)$. Alors

$$\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1) \right) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, \Gamma).$$

Remarque. Ce type de convergence vers une loi Gaussienne peut avoir lieu dans certains cas pour des variables non identiquement distribuées voire dépendantes. Ce théorème exprime qu'une somme importante de phénomènes indépendants et de faible amplitude suit approximativement une loi gaussienne, ce qui justifie et rend pertinent l'utilisation de cette loi en pratique.

Preuve. Comme pour la loi des grands nombres, quitte à poser $Y_j = X_j - \mathbb{E}(X_j)$, on peut supposer les variables centrées, ce que nous ferons. Notons qu'il suffit de montrer le résultat pour des variables aléatoires à valeurs dans \mathbb{R} . En effet pour le cas vectoriel, en posant $T_n = \frac{1}{\sqrt{n}} X_i$, il suffit de montrer que pour tout u , $u^T T_n$ converge en loi vers $u^T X$ avec $X \sim \mathcal{N}_d(0, \Gamma)$ (d'après la remarque juste après le théorème 12). Mais ceci résulte du cas réel appliqué aux variables $u^T X_i$, $i \in \mathbb{N}^*$.

Considérons donc le cas réel. Nous allons utiliser le théorème 12. Pour cela, en notant pour une variable aléatoire Y , ϕ_Y sa fonction caractéristique, nous avons pour $t \in \mathbb{R}$ fixé l'égalité suivante :

$$\phi_{T_n}(t) = \prod_{j=1}^n \phi_{X_j} \left(\frac{t}{\sqrt{n}} \right) = \left(\phi_{X_1} \left(\frac{t}{\sqrt{n}} \right) \right)^n.$$

Les deux égalités précédentes résultent respectivement de l'indépendance et de l'équidistribution des variables X_j , $j \in \mathbb{N}^*$. Remarquons ensuite que l'existence du moment d'ordre 2 pour X_1 justifie le développement limité à l'ordre 2 et en 0 de sa fonction caractéristique

$$\phi_{X_1} \left(\frac{t}{\sqrt{n}} \right) = 1 + i \frac{t}{\sqrt{n}} \mathbb{E}(X_1) - \frac{t^2}{2n} \mathbb{E}(X_1^2) + o\left(\frac{1}{n}\right).$$

Ainsi nous avons

$$\phi_{T_n}(t) = \left(1 + \frac{z_n}{n}\right)^n,$$

où la suite de nombres complexes $(z_n)_n$ converge vers $-\frac{t^2 \mathbb{E}(X_1^2)}{2}$. Pour montrer que $\lim_{n \rightarrow +\infty} \phi_{T_n}(t) = e^{-\frac{1}{2}t^2 \mathbb{E}(X_1^2)}$ (qui est bien la fonction caractéristique au point t de la loi $\mathcal{N}(0, \text{Var}(X_1))$), nous admettrons l'inégalité

$$\left|e^z - \left(1 + \frac{z}{n}\right)^n\right| \leq e^{|z|} - \left(1 + \frac{|z|}{n}\right)^n, \quad \forall (z, n) \in \text{Cov} \times \mathbb{N}.$$

Cette inégalité qui se prouve en développant l'exponentielle en série entière évite l'utilisation du logarithme complexe. Elle permet d'obtenir les majorations :

$$\begin{aligned} \left|e^{-\frac{1}{2}t^2 \mathbb{E}(X_1^2)} - \phi_{T_n}(t)\right| &\leq \left|e^{-\frac{1}{2}t^2 \mathbb{E}(X_1^2)} - e^{z_n}\right| + \left|e^{z_n} - \left(1 + \frac{z_n}{n}\right)^n\right| \\ &\leq \left|e^{-\frac{1}{2}t^2 \mathbb{E}(X_1^2)} - e^{z_n}\right| + \left|e^{|z_n|} - \left(1 + \frac{|z_n|}{n}\right)^n\right|, \end{aligned}$$

ce qui montre la convergence désirée en utilisant la continuité de la fonction exponentielle ainsi que les égalités

$$\left(1 + \frac{|z_n|}{n}\right)^n = \exp\left(n \ln\left(1 + \frac{|z_n|}{n}\right)\right) = \exp(|z_n| + o(1)).$$

Ainsi le théorème 12 permet de conclure que $T_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{Var}(X_1))$. \square

Exemple d'utilisation de l'approximation Gaussienne. Une graine a une probabilité $p = 0.75$ de donner naissance à une plante. Combien doit-on semer de graines pour être sûr à 99% d'obtenir au moins 50 plantes (on utilisera la valeur approchée 0.01 pour $\mathbb{P}(\mathcal{N}(0, 1) > 2.3)$) ? Pour répondre, on doit chercher n tel que

$$\mathbb{P}\left(\sum_{i=1}^n X_i > 49\right) \geq 0.99,$$

où (X_i) est une suite i.i.d de variables aléatoires toutes de loi de Bernoulli de paramètre p . On pourrait alors chercher la valeur minimale de n pour laquelle une binomiale de paramètres n et p vérifie cette inégalité. Afin de limiter les calculs, on peut aussi se servir de l'approximation gaussienne car

$$\mathbb{P}\left(\sum_{i=1}^n X_i > 49\right) = \mathbb{P}\left(\frac{\sum_{i=1}^n (X_i - p)}{\sqrt{np(1-p)}} > \frac{49 - np}{\sqrt{np(1-p)}}\right).$$

Comme la fonction de répartition de $T_n = \frac{\sum_{i=1}^n (X_i - p)}{\sqrt{np(1-p)}}$ converge uniformément vers celle de la loi $\mathcal{N}(0, 1)$, on peut alors résoudre le problème de façon approchée en cherchant n tel que

$$\mathbb{P}\left(\mathcal{N}(0, 1) \leq \frac{49 - np}{\sqrt{np(1-p)}}\right) \leq 0.01.$$

Comme $\mathbb{P}(\mathcal{N}(0, 1) \leq -2.3) \approx 0.01$, on peut chercher le plus petit entier n tel que

$$\frac{49 - np}{\sqrt{np(1-p)}} \leq -2.3,$$

ce qui donne $n \geq 77$.

Bibliographie

- [1] Ouvrard, J.Y, *Probabilités 2*. Cassini.