

École nationale de la statistique et de l'analyse de l'information

Probabilités discrètes et calcul intégral

Jocelyn Julienne

Version de septembre 2014

Table des matières

Avant-propos	vii
I Probabilités discrètes	1
1 Modélisation des expériences aléatoires	5
1.1 Expérience et événement aléatoires	5
1.2 Notion de tribu	6
1.2.1 Définition et propriétés	7
1.2.2 Exemples de tribus importantes	8
1.3 Variable aléatoire	9
2 Probabilités	11
2.1 Définition et propriétés	11
2.1.1 Définition	11
2.1.2 Propriétés	11
2.2 Probabilité conditionnelle	13
2.2.1 Définition	13
2.2.2 Formule de Bayes	13
2.3 Indépendance d'événements	15
2.4 Probabilités sur un univers fini	16
2.5 Loi de probabilité d'une variable aléatoire	16
3 Variables aléatoires discrètes	19
3.1 Définition	19
3.2 Transformation d'une variable aléatoire discrète	20
3.3 Caractéristiques d'une variable aléatoire discrète	21
3.3.1 Fonction de répartition	21
3.3.2 Espérance mathématique	21
3.3.3 Moment d'ordre r	23
3.3.4 Variance et écart-type	23
3.3.5 Variable aléatoire centrée, réduite	25
3.3.6 Fonction génératrice	26

3.4	Couple de variables aléatoires discrètes	27
3.4.1	Définition	27
3.4.2	Lois marginales	28
3.4.3	Variables aléatoires indépendantes	28
3.4.4	Espérance, covariance et matrice de variances-covariances	28
3.4.5	Inégalité de Jensen	30
3.4.6	Fonction génératrice	31
4	Lois discrètes usuelles	33
4.1	La loi de Dirac	33
4.2	Caractérisation des variables aléatoires discrètes à valeurs dans $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.	33
4.3	Loi de Bernoulli $\mathcal{B}(p)$	34
4.4	Loi binomiale $\mathcal{B}(n, p)$	34
4.5	Loi de Poisson $\mathcal{P}(\lambda)$	35
4.6	Loi géométrique $\mathcal{G}(p)$	36
4.7	Loi de Pascal d'ordre n $\mathcal{Pa}(n, p)$	36
4.8	Loi binomiale négative $\mathcal{BN}(n, p)$	37
4.9	Loi hypergéométrique $\mathcal{H}(N, n, p)$	37
5	Exercice récapitulatif	39
II	Les fondements mathématiques des probabilités : une introduction aux théories de la mesure et de l'intégrale de Lebesgue	41
6	Motivation	43
7	Mesure positive	45
7.1	Définitions	45
7.2	Propriétés	45
7.3	Mesures sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$	48
7.3.1	La mesure de Lebesgue	48
7.3.2	La mesure de comptage	49
7.4	Mesure image	49
8	Mesurabilité	51
8.1	Application mesurable	51
8.2	Propriétés des applications mesurables	53
8.3	Produit d'espaces mesurables	54
9	Construction de l'intégrale	57
9.1	Intégrale d'une application étagée positive	57
9.2	Intégrale d'une application mesurable positive	58

9.3	Intégrale d'une application mesurable à valeurs réelles	59
9.4	Intégrale des applications à valeurs complexes	60
9.5	Remarque générale	60
10	Propriétés de l'intégrale et théorèmes de convergence	61
10.1	Notion de négligeabilité	61
10.2	Propriétés de l'intégrale	61
10.3	Théorèmes de convergence	62
10.4	Exemples d'application des théorèmes de convergence	63
11	Intégrale de Lebesgue-intégrale de Riemann	65
11.1	Rappels sur l'intégrale de Riemann	65
11.1.1	Cas d'une fonction bornée	65
11.1.2	Intégrales impropres	66
11.2	Riemann-intégrabilité et Lebesgue-intégrabilité	66
12	Compléments d'intégration	67
12.1	Mesure définie par une densité	67
12.2	Exemples de lois à densité	68
12.2.1	Loi uniforme $\mathcal{U}_{[a,b]}$	68
12.2.2	Loi exponentielle $\mathcal{E}(\theta)$	68
12.2.3	Loi normale $\mathcal{N}(\mu, \sigma^2)$	68
12.3	Théorème de Radon-Nikodym	68
12.4	Théorème de transfert	70
12.5	Intégrale par rapport à une mesure produit	71

Avant-propos

Le cours de Probabilités discrètes et calcul intégral (PDCI) est un cours fondamental et exigeant. Fondamental parce qu'il pose les bases mathématiques de la théorie des probabilités et exigeant car son degré de formalisation est relativement élevé. En théorie, il paraîtra difficile de suivre avec profit les enseignements de statistique inférentielle (Estimation et Tests) si les notions de probabilités ne sont pas maîtrisées, de même que les concepts de PDCI devront être acquis pour comprendre l'enseignement de Probabilités générales.

La première partie du cours devrait être un rappel pour les élèves, puisque la majorité d'entre eux aura déjà suivi un cours de probabilités. Un accent particulier sera mis sur les calculs de lois multivariées (bivariées en pratique), qui posent en général le plus de difficultés aux élèves. La seconde partie, la plus difficile, traitera de la théorie de la mesure et de l'intégrale de Lebesgue.

Par ailleurs, rappelons que la théorie des probabilités n'a pas pour but de confronter un modèle à la réalité. Ce travail incombe plutôt au statisticien (cf. cours de statistique inférentielle). En ce sens, ce cours s'apparenterait à un cours de mathématiques puisque toutes les notions probabilistes seront vues dans un cadre totalement axiomatisé.

Les objectifs de ce cours

Comme cela est indiqué dans la brique, l'objectif de cette matière est double :

- (1) la parfaite maîtrise du calcul des probabilités dans un univers discret au plus dénombrable (ce qui facilitera automatiquement le passage au cadre continu) ;
- (2) l'acquisition de notions élémentaires en théorie abstraite de la mesure et de l'intégration de Lebesgue.

Ce support de cours a été élaboré en s'inspirant de l'ouvrage de Philippe Tassi [5], ancien directeur de l'Ensaï du temps où celle-ci était encore située à Paris sous le nom de l'Ensaë-Cgsa.

Pour une bonne compréhension de cet enseignement (et donc la réussite à l'examen), il est impératif de revoir le cours le soir-même. Le travail requis doit être constant et régulier.

PARTIE I

Probabilités discrètes

« On ne connaît pas complètement une science tant qu'on n'en sait pas l'histoire. »

Auguste Comte, *Cours de philosophie positive*.

Les probabilités sont nées des problèmes de jeux de hasard. L'un des exemples les plus connus, relaté par Georg Cantor en 1873, est celui de l'écrivain Antoine Gombaud, chevalier de Méré (1607-1684) qui soumit les deux paris suivants à Blaise Pascal (1623-1662) :

- (1) si l'on jette 4 fois un dé à six faces, il y a plus de chances qu'on obtienne au moins un 6 plutôt qu'on n'en obtienne pas ¹.
- (2) si l'on jette 24 fois deux dés à six faces, il y a moins de chances qu'on obtienne un double six plutôt qu'on n'en obtienne pas ².

Gombaud avait bien appréhendé le résultat du 1^{er} pari en dépit d'un raisonnement faux, mais son résultat du 2^e s'est avéré erroné.

Même si des mathématiciens comme Pierre de Fermat (1601-1665), Christiaan Huygens (1629-1695), Jacques Bernoulli (1654-1705), Abraham de Moivre (1667-1754) et Pierre-Simon Laplace (1749-1827) ont contribué à l'essor des probabilités, ce n'est qu'en 1933 qu'elles ont été axiomatisées par Andreï Kolmogorov (1903-1987), faisant d'elles une branche des mathématiques à part entière, sans oublier bien sûr la contribution fondamentale de la théorie de Lebesgue (1875-1941).

Dans cette partie, on se propose d'exposer classiquement toutes les définitions et résultats relatifs aux probabilités sur des ensembles au plus dénombrables.

1. Réponse : $\mathbb{P}(\text{"Obtenir au moins un 6 en 4 lancers"}) = 1 - \left(\frac{5}{6}\right)^4 \simeq 0.518$
 2. Réponse : $\mathbb{P}(\text{"Obtenir un double 6 en 24 lancers"}) = \left(\frac{35}{36}\right)^{24} \simeq 0.508$

CHAPITRE 1

Modélisation des expériences aléatoires

La théorie des probabilités a pour but de modéliser des phénomènes aléatoires, c'est-à-dire des phénomènes dont l'issue n'est pas connue de manière certaine. Notons que l'antonyme de **aléatoire** est le mot **déterministe**. Un synonyme de aléatoire est **stochastique**.

Cette modélisation requiert plusieurs étapes dans la formalisation mathématique :

1. la connaissance du cadre dans lequel on observera la manifestation du hasard : il s'agit des **expériences aléatoires** ;
2. l'**événement aléatoire** associé à l'expérience aléatoire ;
3. l'utilisation des variables aléatoires, car il est fréquent que le modélisateur ne puisse pas expliciter l'événement aléatoire mais sa traduction « quantifiée », car il s'énonce souvent de manière numérique.

Cette formalisation est effectuée à l'aide de structures mathématiques *ad hoc*. Ainsi, nous aurons besoin d'une application nommée **probabilité** pour mesurer l'aléa de l'événement et définie sur un ensemble bien particulier.

1.1 Expérience et événement aléatoires

Une expérience aléatoire est caractérisée par l'ensemble Ω de tous les résultats possibles ω de l'expérience envisagée. Cet ensemble Ω est appelé généralement *univers*, *référentiel*, *espace fondamental* ou encore *ensemble des états de la nature*. Il est de cardinal fini ou infini. Sa détermination est totalement subordonnée à la complexité de l'expérience. Nous verrons plus loin qu'il est en général très difficile voire impossible d'expliciter Ω si bien que nous préférons travailler dans un espace plus « concret » : l'*espace image* via une variable aléatoire.

Un événement aléatoire A est un ensemble de résultats élémentaires ω . Ainsi, $A \subset \Omega$. Précisons que si l'un des événements élémentaires ω est réalisé alors l'événement complexe associé A l'est également.

Exemple 1.1.1 On considère l'expérience aléatoire *jeter un dé*. L'univers est alors $\Omega = \{\omega_1, \dots, \omega_6\}$ ¹ où ω_i représente l'événement élémentaire « le dé donne i » ($i \in \llbracket 1, 6 \rrbracket$). Un exemple d'événement aléatoire est « le dé donne un chiffre pair ». On le formalise par l'événement $A = \{\omega_2, \omega_4, \omega_6\}$. A est réalisé dès que l'un des trois événements élémentaires l'est.

Remarque 1.1.1 Le choix de Ω comporte une part d'arbitraire dépendant des idées *a priori* sur l'expérience aléatoire. Nous empruntons l'exemple suivant à Monfort [3] : on considère comme expérience aléatoire le jet d'une pièce de monnaie. On peut y associer l'un des univers suivants :

- $\Omega = \{\text{Pile}, \text{Face}\}$;
- $\Omega = \{\text{Pile}, \text{Face}, \text{Tranche}\}$;
- $\Omega = \mathbb{R}^3$ soit les coordonnées dans un repère donné du centre de gravité de la pièce de monnaie.

1.2 Notion de tribu

Nous venons de voir précédemment qu'on pouvait considérer un certain nombre d'événements aléatoires associés à une expérience aléatoire. Nous savons que tout événement aléatoire A est inclus dans Ω . La question que l'on peut se poser est la suivante : à quel ensemble appartient A ? Réponse : cet ensemble d'ensembles, respectant un certain nombre de propriétés mathématiques, est appelé **tribu**.

Notons \mathcal{A} l'ensemble de tous les événements aléatoires. En général, lorsqu'on n'a aucune idée sur les événements aléatoires à considérer, on prend $\mathcal{A} = \mathcal{P}(\Omega)$, c'est-à-dire l'ensemble des parties de Ω . C'est le cas notamment lorsque Ω est discret (*fini ou dénombrable*).

Si ce n'est pas le cas, $\mathcal{P}(\Omega)$ est trop grand, on lui préférera une classe de parties de Ω incluse dans $\mathcal{P}(\Omega)$ et que l'on dotera d'une structure d'algèbre pour des raisons de stabilité de telle sorte que les opérations ensemblistes ne fassent pas sortir de la classe.

Définition 1.2.1 Soit Ω un ensemble et \mathcal{A} une classe de parties de Ω . \mathcal{A} est une algèbre de Boole (ou algèbre) sur Ω si :

- $\Omega \in \mathcal{A}$;
- $\forall A, B \in \mathcal{A}, A \cup B \in \mathcal{A}$ (stabilité par union) ;
- $\forall A \in \mathcal{A}, \mathcal{C}_{\Omega}^A \in \mathcal{A}$ (stabilité par complémentation ou passage au complémentaire), où \mathcal{C}_{Ω}^A est le complémentaire à A dans Ω eu que l'on notera le plus souvent (lorsqu'il n'y aura pas d'ambiguïté) $\bar{A} = \Omega \setminus A$.

Remarque 1.2.1 Les deux dernières conditions impliquent évidemment la stabilité par intersection.

Toutefois, la structure d'algèbre de Boole n'est pas toujours suffisante car elle n'assure pas la stabilité par union ou intersection dénombrable (dénombrable signifie équipotent

1. Notez que l'écriture $\Omega = \{1, 2, 3, 4, 5, 6\}$ serait abusive car on aurait tendance à croire qu'un événement élémentaire de Ω est un chiffre, ce qui n'a pas de sens évidemment.

à \mathbb{N} , c'est-à-dire de même cardinal ou encore il existe une bijection entre \mathbb{N} et l'ensemble dénombrable). Ainsi, prenons une suite $(A_n)_{n \in \mathbb{N}^*}$ d'éléments de \mathcal{A} , croissante au sens de l'inclusion ($A_n \subset A_{n+1}$), alors rien ne garantit que $\bigcup_{n \in \mathbb{N}^*} A_n$ soit un élément de \mathcal{A} . On fait appel alors à une structure plus grosse : la *tribu* ou *σ -algèbre de Boole*.

1.2.1 Définition et propriétés

Définition 1.2.2 Soit Ω un ensemble. Une classe (ou une famille) de parties de Ω \mathcal{A} est une *tribu* sur Ω si :

- i) $\Omega \in \mathcal{A}$;
- ii) $\forall A \in \mathcal{A}, \bar{A} \in \mathcal{A}$ (stabilité par complémentation) ;
- iii) $\forall A_n \in \mathcal{A}, n \in \mathbb{N}^*, \bigcup_{n \in \mathbb{N}^*} A_n \in \mathcal{A}$ (stabilité par union dénombrable).

Exemple 1.2.1 Soit l'ensemble $\Omega = \{1, 2\}$. L'ensemble étant discret (il est même fini), on peut considérer comme tribu l'ensemble des parties de Ω :

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

Définition 1.2.3 Soit Ω un ensemble.

- On appelle *tribu grossière* ou tribu triviale, la tribu $\{\emptyset, \Omega\}$. Il s'agit de la plus petite tribu sur Ω .
- L'ensemble des parties de Ω , $\mathcal{P}(\Omega)$, est appelé *tribu discrète*. C'est la plus grosse tribu sur Ω .

En conséquence, pour toute tribu \mathcal{A} sur Ω , on a la relation suivante :

$$\{\emptyset, \Omega\} \subset \mathcal{A} \subset \mathcal{P}(\Omega) \quad (1.1)$$

Remarque 1.2.2 Afin de bien fixer les idées sur les objets traités précédemment, il nous paraît utile de souligner que :

- $\forall A \in \mathcal{A}$ on a $A \subset \Omega$; autrement dit, un événement aléatoire est un élément de la tribu, mais c'est un sous-ensemble de l'univers Ω (une tribu peut être considérée comme un ensemble d'ensembles).
- $\forall \omega \in \Omega \quad \{\omega\} \subset \Omega \quad \text{mais} \quad \{\omega\} \in \mathcal{A}$.

Définition 1.2.4 Soit Ω un ensemble et \mathcal{A} une tribu sur Ω . Alors le couple (Ω, \mathcal{A}) est appelé *espace mesurable* ou *espace probabilisable*.

Définition 1.2.5 On appelle *classe* (ou famille) de parties d'un ensemble Ω , un sous-ensemble de $\mathcal{P}(\Omega)$.

C'est un terme que nous retrouverons fréquemment par la suite.

Les principales propriétés concernant les tribus sont les suivantes.

Propriété 1.2.1

- (1) L'intersection dénombrable de tribus est une tribu (admis).
- (2) La réunion de tribus n'est pas une tribu en général (prendre par exemple les tribus engendrées — cf. définition ci-dessous — par les événements A et $B : \{\emptyset, \Omega, A, \bar{A}\}$ et $\{\emptyset, \Omega, B, \bar{B}\}$; l'union ne contient pas $A \cup B$ par exemple).

1.2.2 Exemples de tribus importantes

Nous donnons rapidement les définitions des tribus engendrée et borélienne. Elles ne sont pas forcément utiles lorsque Ω est discret. Nous y reviendrons dans la deuxième partie, lorsque nous traiterons de la théorie de Lebesgue.

Tribu engendrée

On se place dans le cas où, pour une expérience donnée, seules quelques parties de Ω sont intéressantes en tant qu'événements. On note cette classe de parties \mathcal{C} . On munit donc Ω non pas de la tribu discrète mais de la plus petite tribu contenant \mathcal{C} . Cette tribu est appelée *tribu engendrée* par \mathcal{C} .

Définition 1.2.6 Soit Ω un ensemble et \mathcal{C} une classe de parties de Ω . La *tribu engendrée* par \mathcal{C} est la plus petite tribu contenant \mathcal{C} ou encore l'intersection de toutes les tribus contenant \mathcal{C} . Cette tribu est notée $\sigma(\mathcal{C})$.

Théorème 1.2.1 (admis)

On peut également dire que $\sigma(\mathcal{C})$ est l'unique tribu vérifiant :

- (i) $\mathcal{C} \subset \sigma(\mathcal{C})$;
- (ii) si \mathcal{G} est une tribu sur Ω telle que $\mathcal{C} \subset \mathcal{G}$ alors $\sigma(\mathcal{C}) \subset \mathcal{G}$.

Exemple 1.2.2 Reprenons l'exemple du lancer de dé et A l'événement « obtenir un résultat pair ». La tribu engendrée par A n'est autre que l'ensemble suivant :

$$\sigma(\{A\}) = \{\emptyset, \Omega, A, \bar{A}\} \quad (1.2)$$

Tribu borélienne sur \mathbb{R}

La tribu borélienne est celle qui doit faire l'objet de toutes les attentions car c'est celle qui sera le plus souvent utilisée par la suite dans tous les cours de statistique mathématique et de probabilités.

Définition 1.2.7 On appelle *tribu borélienne* sur \mathbb{R} , notée $\mathcal{B}(\mathbb{R})$, la plus petite tribu contenant tous les intervalles de \mathbb{R} . Un élément de $\mathcal{B}(\mathbb{R})$ est appelé *borélien* de \mathbb{R} .

Remarque 1.2.3 Autrement dit, la tribu borélienne sur \mathbb{R} est la tribu engendrée par les intervalles ouverts de la forme $] - \infty, x[$ (avec $x \in \mathbb{R}$) de \mathbb{R} .

1.3 Variable aléatoire

L'utilisation des variables aléatoires n'est pas obligatoire lorsque l'univers est fini, voire dénombrable. En revanche, elle est incontournable lorsque Ω est infini. D'ailleurs, dans ce dernier cas, il est très fréquent qu'on ne sache pas déterminer les événements élémentaires de Ω , mais des « équivalents numériques ». Exemple : la durée de vie d'un composant électronique est inférieure à 2 ans.

En tout état de cause, il est toujours plus aisé de travailler sur des grandeurs quantifiées plutôt que sur les événements eux-mêmes. Cela est possible grâce aux variables aléatoires.

Définition 1.3.1 Soit (Ω, \mathcal{A}) et (E, \mathcal{E}) deux espaces probabilisables (par la suite, on prendra souvent $E = \mathbb{R}$ ou \mathbb{N} et $\mathcal{E} = \mathcal{B}(\mathbb{R})$ ou $\mathcal{P}(\mathbb{N})$ respectivement). On appelle *variable aléatoire* X , une application définie sur (Ω, \mathcal{A}) et à valeurs dans (E, \mathcal{E}) telle que l'image réciproque de tout élément B de \mathcal{E} appartienne à \mathcal{A} . Autrement dit :

$$\forall B \in \mathcal{E}, \quad X^{-1}(B) \in \mathcal{A} \quad (1.3)$$

On dit également que l'application X est *mesurable*.

Notation très importante On sera strict sur les notations suivantes :

$$X^{-1}(B) \triangleq \{X \in B\} \triangleq \{\omega \in \Omega : X(\omega) \in B\} \quad (1.4)$$

Le symbole \triangleq représente l'égalité par définition.

Le schéma de modélisation est donc le suivant :

$$\begin{array}{ccc} X : & (\Omega, \mathcal{A}) & \longrightarrow & (E, \mathcal{E}) \\ & \omega & \longmapsto & X(\omega) \end{array} \quad (1.5)$$

Définition 1.3.2 On appelle *support* de la variable aléatoire X l'ensemble $X(\Omega)$ inclus dans E . On donnera par la suite une autre définition du support en utilisant la notion de probabilité.

Exemple 1.3.1 On considère le lancer d'une pièce de monnaie non truquée (expérience aléatoire). Les événements élémentaires de l'expérience sont : $\omega_1 = \ll \text{la pièce tombe sur pile} \gg$ et $\omega_2 = \ll \text{la pièce tombe sur face} \gg$. L'univers est donc $\Omega = \{\omega_1, \omega_2\}$. La tribu associée à Ω sera la tribu discrète. On considère la variable aléatoire X donnant le résultat du lancer. Supposons qu'on veuille savoir si la pièce tombe sur pile. On peut considérer la modélisation suivante : $X(\omega_1) = 1$ et $X(\omega_2) = 0$ (il s'agit là d'un choix totalement arbitraire). On a donc le schéma :

$$\begin{array}{ccc} X : & (\Omega, \mathcal{A}) & \longrightarrow & (\mathbb{N}, \mathcal{P}(\mathbb{N})) \\ & \omega & \longmapsto & X(\omega) \end{array}$$

avec le support de X : $X(\Omega) = \{0, 1\}$.

Propriété 1.3.1 (admise)

Soit X une variable aléatoire réelle définie sur un espace probabilisable (Ω, \mathcal{A}) et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors $X^{-1}(\mathcal{B}(\mathbb{R}))$ est une tribu sur Ω : c'est la tribu des événements engendrée par X . On la note $\sigma(X)$.

Pour que le cadre de modélisation soit complet, il faut ajouter à l'espace probabilisable, un objet mathématique permettant de mesurer l'aléatoire : il s'agit de la (mesure de) probabilité.

CHAPITRE 2

Probabilités

2.1 Définition et propriétés

2.1.1 Définition

Définition 2.1.1 Soit (Ω, \mathcal{A}) un espace probabilisable. On appelle *probabilité* ou mesure de probabilité sur (Ω, \mathcal{A}) , toute application \mathbb{P} définie sur \mathcal{A} à valeurs dans $[0, 1]$ telle que :

- i) $\mathbb{P}(\Omega) = 1$ (axiome de normalisation) ;
- ii) Pour toute suite $(A_n)_{n \in \mathbb{N}^*}$ d'événements deux à deux disjoints :

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n) \quad (2.1)$$

Il s'agit de la propriété ou de l'axiome de σ -additivité.

Définition 2.1.2 Soit (Ω, \mathcal{A}) un espace probabilisable. Deux événements A et B sont disjoints si $A \cap B = \emptyset$.

Définition 2.1.3 Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé *espace probabilisé*.

2.1.2 Propriétés

Théorème 2.1.1 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Alors :

- i) $\mathbb{P}(\emptyset) = 0$;
- ii) Soient A et B deux événements disjoints. Alors :

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \quad (2.2)$$

- iii) Soient deux événements quelconques A et B . Alors :

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \quad (2.3)$$

Il s'agit de la *formule de Poincaré*. Sa généralisation à n événements A_1, \dots, A_n s'écrit de la manière suivante :

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n \left((-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}\left(\bigcap_{j=1}^k A_{i_j}\right) \right) \quad (2.4)$$

Ou écrit de manière plus explicite :

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \dots + (-1)^{n-1} \mathbb{P}(A_1 \cap \dots \cap A_n) \quad (2.5)$$

iv) Soient A et B deux événements quelconques tels que $A \subset B$. Alors :

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A) \quad \text{et} \quad \mathbb{P}(A) \leq \mathbb{P}(B) \quad (2.6)$$

v) $\forall A \in \mathcal{A} \quad \mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$.

vi) Soit $(A_n)_{n \in \mathbb{N}^*}$ une suite d'événements quelconque alors :

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) \leq \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n) \quad (2.7)$$

(propriété de σ -additivité)

vii) Soit $(A_n)_{n \in \mathbb{N}^*}$ une suite d'événements décroissante (au sens de l'inclusion) telle que $\bigcap_{n \in \mathbb{N}^*} A_n = \emptyset$ (on utilisera la notation suivante : $A_n \downarrow \emptyset$). Alors :

$$\mathbb{P}(A_n) \downarrow 0 \quad (2.8)$$

c'est-à-dire que la limite des $\mathbb{P}(A_n)$ tend vers 0 de manière décroissante. Il s'agit d'une propriété importante appelée *propriété de continuité monotone des probabilités*.

viii) Soit $(A_n)_{n \in \mathbb{N}^*}$ telle que $A_n \downarrow A$. Alors :

$$\mathbb{P}(A_n) \downarrow \mathbb{P}(A) \quad (2.9)$$

(propriété de continuité monotone des probabilités)

ix) Soit $(A_n)_{n \in \mathbb{N}^*}$ telle que $A_n \uparrow A$. Alors :

$$\mathbb{P}(A_n) \uparrow \mathbb{P}(A) \quad (2.10)$$

(propriété de continuité monotone des probabilités)

Remarque 2.1.1 Ces propriétés seront démontrées dans un cadre plus général (cf. théorème 7.2.1 page 45).

2.2 Probabilité conditionnelle

2.2.1 Définition

Définition 2.2.1 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et B un événement tel que $\mathbb{P}(B) > 0$. On appelle *probabilité d'un événement A conditionnellement à B* (ou probabilité de A sachant que B est réalisé) la quantité :

$$\mathbb{P}(A|B) \triangleq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (2.11)$$

Proposition 2.2.1 L'application $\mathbb{P}(\cdot|B)$ est une mesure de probabilité sur l'espace (Ω, \mathcal{A}) .

Démonstration 2.2.1

i) Montrons que l'axiome de normalisation est vérifié :

$$\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$$

ii) Montrons que la propriété de σ -additivité est également vérifiée. Soit $(A_n)_{n \in \mathbb{N}^*}$ une suite d'événements deux à deux disjoints. Alors :

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} A_n | B\right) &= \frac{\mathbb{P}((\bigcup_{n \in \mathbb{N}^*} A_n) \cap B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(\bigcup_{n \in \mathbb{N}^*} (A_n \cap B))}{\mathbb{P}(B)} \\ &= \frac{\sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} \quad \text{car l'union est disjointe} \\ &= \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n | B) \end{aligned}$$

Remarque 2.2.1 On peut alors définir un nouvel espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P}(\cdot|B))$ ou encore $(B, \mathcal{A} \cap B, \mathbb{P}(\cdot|B))$.

Remarque 2.2.2 Toutes les propriétés sur les probabilités sont évidemment valables pour les probabilités conditionnelles (puisque'il s'agit de mesures de probabilités).

2.2.2 Formule de Bayes

Théorème 2.2.1 (Formule ou règle de Bayes à deux événements)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et A et B deux événements de probabilités non nulles. Alors :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\bar{A})\mathbb{P}(\bar{A})} \quad (2.12)$$

Démonstration 2.2.2 Par définition, on a :

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

Or $\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A)$. Par ailleurs : $B = B \cap \Omega = B \cap (A \cup \bar{A}) = (B \cap A) \cup (B \cap \bar{A})$.

Par suite :

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}((B \cap A) \cup (B \cap \bar{A})) \\ &= \mathbb{P}(B \cap A) + \mathbb{P}(B \cap \bar{A}) \quad \text{car l'union est disjointe} \\ &= \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\bar{A})\mathbb{P}(\bar{A})\end{aligned}$$

Remarque 2.2.3 Soit $A \in \mathcal{A}$. On peut toujours partitionner Ω avec A et son complémentaire dans Ω \bar{A} : $\Omega = A \cup \bar{A}$.

Définition 2.2.2 Soit un espace probabilisable (Ω, \mathcal{A}) . On dit qu'une suite d'événements A_1, \dots, A_n forme une partition de Ω si :

$$(i) \bigcup_{i=1}^n A_i = \Omega;$$

(ii) les A_i sont deux à deux disjoints : $\forall i, j \in \llbracket 1, n \rrbracket^1, i \neq j \quad A_i \cap A_j = \emptyset$.

La définition vaut également pour une suite infinie de $(A_n)_{n \in \mathbb{N}^*}$.

Théorème 2.2.2 (Théorème des probabilités totales)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(A_n)_{n \in \mathbb{N}^*}$ une partition de Ω telle que $\forall n \in \mathbb{N}^*, \mathbb{P}(A_n) \neq 0$. Alors :

$$\forall B \in \mathcal{A} \quad \mathbb{P}(B) = \sum_{n \in \mathbb{N}^*} \mathbb{P}(B|A_n)\mathbb{P}(A_n) \quad (2.13)$$

Démonstration 2.2.3 Il suffit d'écrire $B = B \cap \Omega = B \cap (\bigcup_{n \in \mathbb{N}^*} A_n)$. □

Exemple 2.2.1

Une maladie affecte une proportion p de la population. Un test permet de dépister cette maladie avec la fiabilité suivante :

- le test est positif pour une proportion α de personnes affectées par la maladie ;
- le test est négatif pour une proportion β de personnes non affectées par la maladie.

Calculer la probabilité pour qu'un individu ayant un test positif soit atteint par la maladie.

Utilisons les notations suivantes :

- M = "l'individu est affecté par la maladie" ;
- P = "le test est positif".

Par hypothèse, nous avons : $\mathbb{P}(P|M) = \alpha$ et $\mathbb{P}(\bar{P}|\bar{M}) = \beta$. On nous demande de déterminer $\mathbb{P}(M|P)$. Il s'agit d'une application immédiate de la règle de Bayes, il vient alors :

$$\mathbb{P}(M|P) = \frac{\mathbb{P}(P|M)\mathbb{P}(M)}{\mathbb{P}(P|M)\mathbb{P}(M) + \mathbb{P}(P|\bar{M})\mathbb{P}(\bar{M})}$$

Or $\mathbb{P}(P|\bar{M}) = 1 - \beta$ et $\mathbb{P}(\bar{M}) = 1 - p$. Par suite :

$$\mathbb{P}(M|P) = \frac{\alpha p}{\alpha p + (1 - \beta)(1 - p)}$$

1. $\llbracket 1, n \rrbracket \triangleq \{1, \dots, n\}$. Il s'agit d'un intervalle discret.

2.3 Indépendance d'événements

Définition 2.3.1 (Indépendance de deux événements)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $A, B \in \mathcal{A}$. Les deux événements A et B sont dits *indépendants* pour la mesure de probabilité \mathbb{P} si :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \quad (2.14)$$

Remarque 2.3.1 L'écueil classique est de confondre l'indépendance et l'incompatibilité ($A \cap B = \emptyset$) entre deux événements. Ce sont deux notions qui n'ont rien à voir.

Proposition 2.3.1 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et A et B deux événements de probabilités non nulles. Si A et B sont indépendants alors :

$$\mathbb{P}(A|B) = \mathbb{P}(A) \quad \text{et} \quad \mathbb{P}(B|A) = \mathbb{P}(B) \quad (2.15)$$

Propriété 2.3.1 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et A et B deux événements indépendants. Alors :

- (a) \bar{A} et B sont indépendants ;
- (b) A et \bar{B} sont indépendants ;
- (c) \bar{A} et \bar{B} sont indépendants.

Démonstration 2.3.1 Nous ne démontrons que le résultat (a), les autres se traitant par le même raisonnement. B peut s'écrire : $B = (B \cap A) \cup (B \cap \bar{A})$ (union disjointe). Par suite :

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap \bar{A}) \quad (2.16)$$

Or, A et B sont indépendants par hypothèse. Donc $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. On remplace cette expression dans l'équation (2.16) et on en déduit le résultat. \square

Définition 2.3.2 (Indépendance de n événements)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. n événements A_1, \dots, A_n sont dits *indépendants* ou *mutuellement indépendants* si pour tout $k \in \llbracket 1, n \rrbracket$ et pour tout sous-ensemble $\{A_{i_1}, \dots, A_{i_k}\}$ de k événements choisis parmi l'ensemble $\{A_1, \dots, A_n\}$ on a :

$$\mathbb{P}\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k \mathbb{P}(A_{i_j}) \quad (2.17)$$

Remarque 2.3.2 La mutuelle indépendance est donc une propriété plus forte que l'indépendance deux à deux.

Proposition 2.3.2 n événements (mutuellement) indépendants sont indépendants deux à deux.

Démonstration 2.3.2 Elle est immédiate. Il suffit de prendre $k = 2$. \square

Définition 2.3.3 (Indépendance d'une suite d'événements)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(A_n)_{n \in \mathbb{N}^*}$ une suite d'éléments de \mathcal{A} . Les A_n forment une suite d'événements indépendants si pour tout $k \in \llbracket 1, n \rrbracket$ et pour tout sous-ensemble $\{A_{i_1}, \dots, A_{i_k}\}$ de k événements choisis parmi les $(A_n)_{n \in \mathbb{N}^*}$ on a :

$$\mathbb{P} \left(\bigcap_{j=1}^k A_{i_j} \right) = \prod_{j=1}^k \mathbb{P}(A_{i_j}) \quad (2.18)$$

2.4 Probabilités sur un univers fini

Dans ce chapitre, nous traitons à part des probabilités sur un univers fini car il s'agit d'une situation particulière où on peut admettre une définition spécifique de la probabilité (même si évidemment la définition générale de la mesure de probabilité est valable) en s'affranchissant des notions de tribus et d'espaces probabilisés (mais nous ne le ferons pas ici) puisque chaque événement élémentaire $\{\omega\}$ peut être explicité et probabilisé.

Soit Ω un ensemble fini. On le munit de la tribu discrète $\mathcal{P}(\Omega)$. L'espace $(\Omega, \mathcal{P}(\Omega))$ est probabilisable par la probabilité \mathbb{P} définie sur $\mathcal{P}(\Omega)$ à valeurs dans $[0, 1]$ et définie de la manière suivante :

$$\forall A \in \mathcal{P}(\Omega) \quad \mathbb{P}(A) = \frac{\text{Card} A}{\text{Card} \Omega} \quad (2.19)$$

Dans ce cadre, chaque événement élémentaire $\{\omega\}$ est *équiprobable*, c'est-à-dire que :

$$\mathbb{P}(\{\omega\}) = \frac{1}{\text{Card} \Omega} \quad (2.20)$$

Par suite, le calcul des probabilités sur l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ consiste à résoudre des problèmes de dénombrements.

Remarque 2.4.1 Il est clair que cette définition de la probabilité n'est pas valable dans le cas général, notamment lorsque Ω est dénombrable auquel cas $\text{Card} \Omega = +\infty$.

2.5 Loi de probabilité d'une variable aléatoire

Nous avons vu qu'hormis le cas où Ω est fini, on était souvent amené à modéliser des phénomènes à l'aide de variables aléatoires, ne serait-ce parce qu'il est impossible de déterminer Ω , même si son existence est avérée.

Définition 2.5.1 (Loi de probabilité d'une variable aléatoire)

Soit X une variable aléatoire définie sur l'espace probabilisable (Ω, \mathcal{A}) et à valeurs dans (E, \mathcal{E}) (dans la pratique E sera souvent \mathbb{R} ou \mathbb{N}). On appelle *loi de probabilité* de la variable aléatoire X , la mesure de probabilité \mathbb{P}_X de \mathcal{E} dans $[0, 1]$ définie de la manière suivante :

$$\forall B \in \mathcal{E} \quad \mathbb{P}_X(B) \triangleq \mathbb{P}(X^{-1}(B)) \triangleq \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}) \triangleq \mathbb{P}(X \in B) \quad (2.21)$$

Autrement dit, la probabilité \mathbb{P}_X est la composée des applications \mathbb{P} et X^{-1} :

$$\mathbb{P}_X \triangleq \mathbb{P} \circ X^{-1} \quad (2.22)$$

\mathbb{P}_X est également appelée *mesure de probabilité image* de \mathbb{P} par X .

Remarque 2.5.1 $(E, \mathcal{E}, \mathbb{P}_X)$ est un espace probabilisé, appelée *espace probabilisé image* de l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ par X . Toutes les propriétés de \mathbb{P} sont valables également pour \mathbb{P}_X .

CHAPITRE 3

Variables aléatoires discrètes

Dans ce chapitre, nous nous plaçons dans des espaces probabilisés $(\Omega, \mathcal{A}, \mathbb{P})$ où Ω est au plus dénombrable. En particulier, nous traiterons plus spécifiquement d'espaces probabilisés images par une variable aléatoire (v.a.) X $(E, \mathcal{E}, \mathbb{P}_X)$ avec $E \subset \mathbb{N}$, support des lois discrètes usuelles (Bernoulli, binomiale, Poisson, géométrique, etc.).

3.1 Définition

Définition 3.1.1 Soit un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Soit une variable aléatoire X à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$. On dit que X est une v.a. *discrète* s'il existe un sous-ensemble \mathcal{S}_X de \mathbb{R} , au plus dénombrable tel que :

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) \in \mathcal{S}_X\}) = \mathbb{P}(X \in \mathcal{S}_X) = 1 \quad (3.1)$$

Définition 3.1.2 Soit X une v.a. discrète. L'ensemble \mathcal{S}_X est appelé *support* de X ou de la loi de probabilité de X . Il est égal à :

$$\mathcal{S}_X = \{x \in \mathbb{R} : \mathbb{P}(X = x) > 0\} \quad (3.2)$$

avec $\mathbb{P}(X = x) \triangleq \mathbb{P}_X(\{x\})$.

Remarque 3.1.1 $\forall I \subset \mathbb{R}$, on a :

$$\mathbb{P}_X(I) = \mathbb{P}(X \in I) = \sum_{x \in I \cap \mathcal{S}_X} \mathbb{P}(X = x) = \sum_{x \in I \cap \mathcal{S}_X} \mathbb{P}_X(\{x\}) \quad (3.3)$$

Proposition 3.1.1 (TRÈS IMPORTANT)

La loi d'une v.a. discrète (v.a.d.) est entièrement définie par la connaissance des valeurs des probabilités ponctuelles, c'est-à-dire de chaque $\mathbb{P}(X = x) \quad \forall x \in \mathcal{S}_X$.

Définition 3.1.3 (Terminologie)

Soit X une v.a.d. On appelle *fonction de masse* de X l'application f_X définie sur \mathbb{R} à valeurs dans $[0, 1]$:

$$f_X(x) = \begin{cases} \mathbb{P}(X = x) & \text{si } x \in \mathcal{S}_X \\ 0 & \text{sinon} \end{cases} \quad (3.4)$$

Proposition 3.1.2 (admis)

Il est clair qu'on a le résultat suivant :

$$\sum_{x \in \mathcal{S}_X} \mathbb{P}(X = x) = 1 \quad (3.5)$$

3.2 Transformation d'une variable aléatoire discrète

Lorsqu'on modélise un phénomène aléatoire, il est fréquent qu'il existe des relations entre les différentes variables aléatoires d'intérêt.

Lemme 3.2.1 Soit X une v.a. définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et de support au plus dénombrable \mathcal{S}_X . Soit ψ une application quelconque de \mathcal{S}_X vers \mathcal{S} , ensemble au plus dénombrable. Alors $Y = \psi \circ X = \psi(X)$ est une v.a.d. de support $\mathcal{S}_Y = \mathcal{S}$ et de loi de probabilité \mathbb{P}_Y :

$$\forall y \in \mathcal{S}_Y \quad \mathbb{P}_Y(\{y\}) = \sum_{x \in \mathcal{S}_X : \psi(x)=y} \mathbb{P}_X(\{x\}) \quad (3.6)$$

ou écrit autrement :

$$\forall y \in \mathcal{S}_Y \quad \mathbb{P}(Y = y) = \sum_{x \in \mathcal{S}_X : \psi(x)=y} \mathbb{P}(X = x) \quad (3.7)$$

Démonstration 3.2.1 Soit $y \in \mathcal{S}_Y$. On a :

$$\{Y = y\} = \{\psi(X) = y\} = \bigcup_{x \in \mathcal{S}_X : \psi(x)=y} \{X = x\}$$

Or, l'union ci-dessus est clairement disjointe. Par suite, en utilisant la propriété de σ -additivité de la mesure de probabilité \mathbb{P} , on obtient le résultat requis. \square

Exemple 3.2.1 Soit X une v.a.d. de support $\mathcal{S}_X = \mathbb{Z}^*$ et de loi de probabilité :

$$\forall x \in \mathbb{N}^* \quad \mathbb{P}(X = x) = \mathbb{P}(X = -x) = \frac{1}{2x(x+1)} \quad (3.8)$$

Déterminons la loi de $Y = X^2$. La première chose à faire est de déterminer le support de Y . Or il est clair que $\mathcal{S}_Y = \{y \in \mathbb{N}^* : y = x^2, x \in \mathbb{N}^*\}$. Il reste donc à calculer la fonction de masse de Y . Soit $y \in \mathcal{S}_Y$. On a $\mathbb{P}(Y = y) = \mathbb{P}(X^2 = y)$. Or :

$$\{X^2 = y\} = \{X = -\sqrt{y}\} \cup \{X = \sqrt{y}\} \quad (\text{l'union est disjointe})$$

Par suite :

$$\forall y \in \mathcal{S}_Y \quad \mathbb{P}(Y = y) = \mathbb{P}(X = -\sqrt{y}) + \mathbb{P}(X = \sqrt{y}) = \frac{1}{\sqrt{y}(\sqrt{y}+1)}$$

3.3 Caractéristiques d'une variable aléatoire discrète

Les différentes notions exposées ci-dessous ne concernent pas que les v.a. discrètes. Toutefois, leurs expressions diffèrent selon la nature de la v.a., à savoir discrète ou continue. Nous verrons dans la deuxième partie de ce support de cours que la théorie de la mesure et de l'intégrale de Lebesgue permettra de donner une définition générale de ces caractéristiques, quelle que soit la nature de la v.a.

3.3.1 Fonction de répartition

La fonction de répartition n'est pas très utile dans le cas des variables aléatoires discrètes. Elles servent lorsque les variables aléatoires sont continues.

Définition 3.3.1 On appelle *fonction de répartition* d'une v.a.d. X une application F_X définie sur \mathbb{R} à valeurs dans $[0, 1]$ dont l'expression est :

$$\forall x \in \mathbb{R} \quad F_X(x) \triangleq \mathbb{P}(X \leq x) \quad (3.9)$$

et qui possède les propriétés suivantes :

- elle est non décroissante ;
- elle est continue à droite et admet une limite à gauche (cadlag) ;
- $\lim_{x \rightarrow +\infty} F_X(x) = 1$ et $\lim_{x \rightarrow -\infty} F_X(x) = 0$.

Remarque 3.3.1

- La définition ci-dessus correspond à la définition anglo-saxonne, largement utilisée. Il existe une définition « à la française » selon laquelle la fonction de répartition est continue à gauche et définie à droite avec $F_X(x) = \mathbb{P}(X < x)$.
- La courbe représentative de la fonction de répartition d'une v.a.d. est étagée.
- La loi d'une v.a.d. est entièrement déterminée soit par sa fonction de masse soit par sa fonction de répartition.

3.3.2 Espérance mathématique

Nous savons maintenant qu'une v.a. est une variable dont on ne connaît pas les valeurs *a priori*. Pour pallier cet inconvénient, une première idée est de remplacer la v.a. par sa valeur moyenne théorique (c'est-à-dire celle donnée par le modèle), appelée espérance mathématique.

Définition 3.3.2 Soit une v.a. discrète définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans l'espace probabilisé $(E, \mathcal{E}, \mathbb{P}_X)$ et de support \mathcal{S}_X . On appelle *espérance mathématique* de X la quantité déterministe suivante, sous réserve d'existence :

$$\mathbb{E}(X) \triangleq \sum_{x \in \mathcal{S}_X} x \mathbb{P}(X = x) \quad (3.10)$$

Remarque 3.3.2 Si le support \mathcal{S}_X est fini alors la somme est finie et l'espérance mathématique existe. Si le support est dénombrable, alors la somme n'est pas finie et rien ne garantit *a priori* la convergence de la série.

Proposition 3.3.1 Une v.a.d. X admet une espérance si $\sum_{x \in \mathcal{S}_X} |x| \mathbb{P}(X = x)$ converge.

Proposition 3.3.2 (admise)

Soient deux v.a.d. X et Y et ψ une application définie sur \mathcal{S}_X et à valeurs réelles telle que $Y = \psi(X)$ et $\sum_{x \in \mathcal{S}_X} |\psi(x)| \mathbb{P}(X = x)$ converge. Alors :

$$\mathbb{E}(Y) \triangleq \mathbb{E}(\psi(X)) \triangleq \sum_{x \in \mathcal{S}_X} \psi(x) \mathbb{P}(X = x) \quad (3.11)$$

Remarque 3.3.3 La série mentionnée dans la proposition 3.3.1 ci-dessus correspond en fait à $\mathbb{E}(|X|)$. Une v.a.d. X admet donc une espérance mathématique si $\mathbb{E}|X| < +\infty$ (on verra dans la partie 2 qu'il s'agit d'une condition d'intégrabilité de X).

Exemple 3.3.1

Soit X une variable aléatoire de loi :

$$\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \frac{1}{2} \quad (3.12)$$

Le support de X étant fini, l'espérance existe et vaut 0.

Exemple 3.3.2

Soit X de loi :

$$\forall n \in \mathbb{N}^* \quad \mathbb{P}(X = n) = \frac{1}{\zeta(3)} \frac{1}{n^3} \quad (3.13)$$

où $\zeta(3) = \sum_{n \in \mathbb{N}^*} \frac{1}{n^3}$. Calculons $\mathbb{E}(|X|)$ pour savoir si X admet une espérance.

$$\mathbb{E}(|X|) = \sum_{n \in \mathbb{N}^*} n \mathbb{P}(X = n) = \frac{1}{\zeta(3)} \sum_{n \in \mathbb{N}^*} \frac{1}{n^2} \quad (3.14)$$

Or, la série de terme général $\frac{1}{n^2}$ est une série de Riemann convergente. L'espérance de X est donc bien définie et on a :

$$\mathbb{E}(X) = \sum_{n \in \mathbb{N}^*} n \mathbb{P}(X = n) = \frac{1}{\zeta(3)} \sum_{n \in \mathbb{N}^*} \frac{1}{n^2} = \frac{1}{\zeta(3)} \frac{\pi^2}{6} \quad (3.15)$$

Propriété fondamentale de l'espérance mathématique

Propriété 3.3.1 Soit X une v.a.d. Alors, sous réserve d'existence, $\forall a, b \in \mathbb{R}$:

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b \quad (3.16)$$

Nous verrons plus loin que l'espérance mathématique est même un opérateur linéaire.

Démonstration 3.3.1 Elle est immédiate :

$$\begin{aligned} \mathbb{E}(aX + b) &\stackrel{\Delta}{=} \sum_{x \in \mathcal{S}_X} (ax + b) \mathbb{P}(X = x) \\ &= \sum_{x \in \mathcal{S}_X} (ax \mathbb{P}(X = x) + b \mathbb{P}(X = x)) \\ &= a \sum_{x \in \mathcal{S}_X} x \mathbb{P}(X = x) + b \underbrace{\sum_{x \in \mathcal{S}_X} \mathbb{P}(X = x)}_1 \\ &= a\mathbb{E}X + b \end{aligned}$$

3.3.3 Moment d'ordre r

Définition 3.3.3 Soit X une v.a.d. de support \mathcal{S}_X . On appelle *moment simple d'ordre r* ($r \in \mathbb{R}$) de X , la quantité déterministe suivante (sous réserve d'existence) :

$$\mathbb{E}(X^r) \stackrel{\Delta}{=} \sum_{x \in \mathcal{S}_X} x^r \mathbb{P}(X = x) \quad (3.17)$$

Remarque 3.3.4 L'espérance mathématique de X correspond donc à son moment simple d'ordre 1.

Définition 3.3.4 On appelle *moment centré d'ordre r* d'une v.a.d. X , la quantité déterministe (sous réserve d'existence) :

$$\mathbb{E}[(X - \mathbb{E}X)^r] \stackrel{\Delta}{=} \sum_{x \in \mathcal{S}_X} (x - \mathbb{E}X)^r \mathbb{P}(X = x) \quad (3.18)$$

3.3.4 Variance et écart-type

Caractériser une v.a. par son espérance mathématique est loin d'être suffisant.

Définition 3.3.5 Soit X une v.a.d. définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ de support \mathcal{S}_X . La *variance* de X est la quantité déterministe suivante (sous réserve d'existence) :

$$\mathbb{V}(X) \stackrel{\Delta}{=} \mathbb{E}[(X - \mathbb{E}X)^2] \quad (3.19)$$

La variance de X est donc son moment centré d'ordre 2. Il s'agit d'un indicateur qui mesure la dispersion (théorique) autour de l'espérance mathématique. Dans ce sens, on peut affirmer que la variance est un indicateur de risque.

Remarque 3.3.5 On verra dans la deuxième partie du cours que la variance n'est rien d'autre qu'une norme dans l'espace fonctionnel des variables aléatoires centrées.

Interprétation de la variance Plus la variance sera élevée, plus la dispersion autour de l'espérance sera importante. Le modélisateur souhaite toujours disposer d'une variance la plus faible possible.

Propriété 3.3.2

- a) Une variance est toujours positive.
 b) $\forall a, b \in \mathbb{R} \quad \mathbb{V}(aX + b) = a^2 \mathbb{V}(X)$ (on dit que la variance est un opérateur quadratique).

Démonstration 3.3.2

- a) Par définition :

$$\mathbb{V}(X) = \sum_{x \in \mathcal{S}_X} (x - \mathbb{E}X)^2 \mathbb{P}(X = x) \quad (3.20)$$

Il s'agit d'une série à termes positifs donc, la variance est positive.

- b) Soit $a, b \in \mathbb{R}$.

$$\begin{aligned} \mathbb{V}(aX + b) &\stackrel{\Delta}{=} \sum_{x \in \mathcal{S}_X} (ax + b - \mathbb{E}(aX + b))^2 \mathbb{P}(X = x) \\ &= \sum_{x \in \mathcal{S}_X} (a(x - \mathbb{E}X))^2 \mathbb{P}(X = x) \\ &= a^2 \sum_{x \in \mathcal{S}_X} (x - \mathbb{E}X)^2 \mathbb{P}(X = x) \\ &= a^2 \mathbb{V}(X) \end{aligned}$$

Décomposition de König-Huygens

Proposition 3.3.3 (FORMULE DE KÖNIG-HUYGENS)

Soit X une v.a. admettant une variance. Alors :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 \quad (3.21)$$

Cette décomposition est très utile en pratique pour calculer les variances.

Démonstration 3.3.3

$$\begin{aligned} \mathbb{V}(X) &\stackrel{\Delta}{=} \mathbb{E}[(X - \mathbb{E}X)^2] \quad \text{puis on développe le carré} \\ &= \mathbb{E}[X^2 - 2X\mathbb{E}(X) + (\mathbb{E}X)^2] \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X)) + \mathbb{E}((\mathbb{E}X)^2) \quad \text{par linéarité de } \mathbb{E} \\ &= \mathbb{E}(X^2) - 2(\mathbb{E}X)^2 + (\mathbb{E}X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}^2 X \end{aligned}$$

En effet : $\mathbb{E}(X\mathbb{E}X) = \mathbb{E}X \times \mathbb{E}X = \mathbb{E}^2 X$ (car $\mathbb{E}X$ est une constante) et $\mathbb{E}(\mathbb{E}^2 X) = \mathbb{E}^2 X$. \square

Proposition 3.3.4 Soit X une v.a.d. Pour que $\mathbb{V}(X)$ existe, il faut et il suffit que le moment d'ordre deux existe ($\mathbb{E}(X^2) < +\infty$).

Démonstration 3.3.4

⇐) Condition suffisante : on utilise l'inégalité suivante :

$$\forall x \in \mathbb{R} \quad |x| \leq x^2 + 1$$

Par suite :

$$\sum_{x \in \mathcal{S}_X} |x| \mathbb{P}(X = x) \leq \sum_{x \in \mathcal{S}_X} (x^2 + 1) \mathbb{P}(X = x)$$

c'est-à-dire :

$$\sum_{x \in \mathcal{S}_X} |x| \mathbb{P}(X = x) \leq \sum_{x \in \mathcal{S}_X} x^2 \mathbb{P}(X = x) + \underbrace{\sum_{x \in \mathcal{S}_X} \mathbb{P}(X = x)}_1$$

inégalité qui peut s'écrire encore :

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

En conséquence, si le moment d'ordre 2 existe, alors l'espérance mathématique existe également.

D'après la formule de König-Huygens, on en déduit l'existence de la variance.

⇒) La condition nécessaire se démontre en utilisant par exemple l'inégalité suivante :

$$\forall x, y \in \mathbb{R} \quad x^2 \leq (x - y)^2 + 2xy$$

On remplace y par $\mathbb{E}X$. On multiplie par $\mathbb{P}(X = x)$ pour tout $x \in \mathcal{S}_X$, puis on somme. \square

L'écart-type

Une variante de la variance est l'écart-type, très utilisé en pratique car il est de même dimension (de même unité) que la variable aléatoire.

Définition 3.3.6 Soit X une v.a. On appelle *écart-type* de X la racine carrée de la variance de X (sous réserve d'existence) :

$$\sigma(X) = \sqrt{\mathbb{V}(X)} \quad (3.22)$$

Remarque 3.3.6 On veillera à ne pas confondre la notation de l'écart-type de X avec celle de la tribu engendrée par X . En général, le contexte empêchera toute ambiguïté.

3.3.5 Variable aléatoire centrée, réduite

Définition 3.3.7 Une variable aléatoire X est dite *centrée* si son espérance est nulle (sous réserve d'existence).

Remarque 3.3.7 En conséquence, pour centrer une v.a., il suffit de lui soustraire son espérance mathématique.

Définition 3.3.8 Une variable aléatoire X est dite *réduite* si sa variance égale 1 (sous réserve d'existence).

Remarque 3.3.8 Pour réduire une v.a., il suffit de la normaliser par son écart-type :

$$\mathbb{V}\left(\frac{X}{\sigma(X)}\right) = \frac{\mathbb{V}(X)}{\sigma^2(X)} = 1$$

3.3.6 Fonction génératrice

Dans cette sous-section, nous considérerons des v.a.d. de support \mathbb{N} (ou bien inclus dans \mathbb{N} *mutatis mutandis*¹).

A ce stade, on peut dire que les fonctions génératrices servent à calculer les moments des v.a., lorsque les expressions ne sont pas triviales. Leur utilisation requiert les connaissances de base sur les séries entières.

En outre, il existe un outil plus général que la fonction génératrice : la fonction caractéristique appelée également transformée de Fourier (qui sera vue en statistique 2).

Définition 3.3.9 On appelle *fonction génératrice* ou *transformée en z* d'une v.a.d. X de support \mathcal{S}_X , la fonction G_X définie par :

$$\forall z \in \mathbb{C}, |z| \leq 1 \quad G_X(z) = \mathbb{E}(z^X) \quad (3.23)$$

Donc :

$$G_X(z) = \sum_{x \in \mathcal{S}_X} z^x \mathbb{P}(X = x) \quad (3.24)$$

Proposition 3.3.5 (admise)

G_X est continue à l'intérieur du disque unité, c'est-à-dire pour tout $z \in \mathbb{C}$ tel que $|z| \leq 1$ et indéfiniment dérivable sur le domaine $\{z \in \mathbb{C} : |z| < 1\}$.

Proposition 3.3.6 La loi d'une v.a.d. X est entièrement déterminée par sa fonction génératrice.

Démonstration 3.3.5 En effet, on montre facilement que :

$$\forall x \in \mathbb{N} \quad \mathbb{P}(X = x) = \frac{G_X^{(x)}(0)}{x!} \quad (3.25)$$

avec $G_X^{(0)} = G_X$. □

Proposition 3.3.7 (admise)

Si G_X est au moins deux fois dérivable en $z = 1$, alors :

$$\begin{aligned} \mathbb{E}(X) &= G'_X(1) \\ \mathbb{E}(X^2) &= G''_X(1) + G'_X(1) \end{aligned} \quad (3.26)$$

et plus généralement, si G_X est dérivable à l'ordre r en 1, alors la dérivée r^e en 1 vaut :

$$G_X^{(r)}(1) = \sum_{x=r}^{+\infty} x(x-1)\dots(x-r+1)\mathbb{P}(X = x) \quad (3.27)$$

Exemple 3.3.3

1. En changeant ce qui doit être changé.

Soit une variable aléatoire discrète X de loi :

$$\forall x \in \llbracket 0, n \rrbracket, \quad \mathbb{P}(X = x) = C_n^x p^x (1-p)^{n-x} \quad (3.28)$$

La fonction génératrice de X vaut :

$$G_X(z) = (1 - p + pz)^n \quad (3.29)$$

Il vient alors :

$$\begin{aligned} G'_X(z) &= np(1 - p + pz)^{n-1} \\ G''_X(z) &= n(n-1)p^2(1 - p + pz)^{n-2} \end{aligned}$$

Par suite :

$$\begin{aligned} G'_X(1) &= np \\ G''_X(1) &= n(n-1)p^2 \end{aligned}$$

On en déduit donc $\mathbb{E}(X) = np$ et $\mathbb{V}(X) = np(1-p)$.

3.4 Couple de variables aléatoires discrètes

Dans cette section, nous ne traiterons que des couples de v.a., mais toutes les définitions et résultats sont valables pour des n -uplets de v.a. *mutatis mutandis*.

3.4.1 Définition

Définition 3.4.1 Soient deux v.a. définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ et $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_Y)$ respectivement. On appelle *loi de probabilité du couple* (X, Y) (ou loi jointe ou loi conjointe) la probabilité :

$$\forall B \in \mathcal{B}(\mathbb{R}^2) \quad \mathbb{P}_{X,Y}(B) = \mathbb{P}(\{(X, Y) \in B\}) \quad (3.30)$$

L'espace probabilisé image du couple (X, Y) est $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \mathbb{P}_{X,Y})$. L'ensemble B peut être le produit cartésien de deux intervalles sur \mathbb{R} de la forme $I_1 \times I_2$.

Définition 3.4.2 Le couple de v.a. (X, Y) est de loi discrète s'il existe un sous-ensemble $\mathcal{S}_{X,Y}$ de \mathbb{R}^2 au plus dénombrable tel que :

$$\mathbb{P}_{X,Y}(\mathcal{S}_{X,Y}) = 1 \quad (3.31)$$

$\mathcal{S}_{X,Y}$ est appelé support de la loi de (X, Y) :

$$\mathcal{S}_{X,Y} = \{(x, y) \in \mathbb{R}^2 : \mathbb{P}_{X,Y}(\{(x, y)\}) > 0\} \quad (3.32)$$

Propriété 3.4.1

$\mathcal{S}_{X,Y} \subset \mathcal{S}_X \times \mathcal{S}_Y$. Il y a égalité en cas d'indépendance de X et Y (cf. plus bas).

Proposition 3.4.1 La loi d'un couple de v.a. (X, Y) est entièrement déterminée par la connaissance de toutes les probabilités $\mathbb{P}_{X,Y}(\{(x, y)\})$ pour tout $(x, y) \in \mathcal{S}_{X,Y}$.

Notations

$$\mathbb{P}_{X,Y}(\{(x,y)\}) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x, Y = y) \quad (3.33)$$

3.4.2 Lois marginales

La proposition suivante permet de déterminer les lois de X et de Y (dites lois marginales) à partir de la loi du couple (X, Y) .

Proposition 3.4.2 (admis)

Soit (X, Y) un couple de v.a.d. de loi $\mathbb{P}_{X,Y}$. Alors les lois marginales de X et de Y sont respectivement les suivantes :

$$\begin{aligned} \mathbb{P}(X = x) &= \sum_{y \in \mathcal{S}_Y} \mathbb{P}(X = x, Y = y) \\ \mathbb{P}(Y = y) &= \sum_{x \in \mathcal{S}_X} \mathbb{P}(X = x, Y = y) \end{aligned} \quad (3.34)$$

Cette proposition est généralisable aux dimensions supérieures à 2. Soit (X_1, \dots, X_n) un n -uplet de v.a.d. de loi jointe $\mathbb{P}_{X_1, \dots, X_n}$. La loi marginale du sous-vecteur $(X_{i_1}, \dots, X_{i_k})$ ($k < n$) est donnée par la formule suivante :

$$\mathbb{P}(X_{i_1} = x_{i_1}, \dots, X_{i_k} = x_{i_k}) = \sum_{\substack{x_i \in \mathcal{S}_{X_i} \\ i \in \llbracket 1, n \rrbracket \setminus \{i_1, \dots, i_k\}}} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \quad (3.35)$$

3.4.3 Variables aléatoires indépendantes

Nous donnons la définition de l'indépendance directement pour n v.a.

Définition 3.4.3 n v.a.d. X_1, \dots, X_n de support respectif $\mathcal{S}_{X_1}, \dots, \mathcal{S}_{X_n}$ sont dites indépendantes si pour tout n -uplet $(x_1, \dots, x_n) \in \mathcal{S}_{X_1} \times \dots \times \mathcal{S}_{X_n}$, les événements $\{X_1 = x_1\}, \dots, \{X_n = x_n\}$ le sont.

On en déduit la définition de l'indépendance de deux v.a.d. X et Y .

Définition 3.4.4 Deux v.a.d. X et Y de support respectif \mathcal{S}_X et \mathcal{S}_Y sont indépendantes si :

$$\forall (x, y) \in \mathcal{S}_X \times \mathcal{S}_Y \quad \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \quad (3.36)$$

3.4.4 Espérance, covariance et matrice de variances-covariances

Espérance

Définition 3.4.5 Soit un n -uplet de v.a. $X = (X_1, \dots, X_n)$. L'espérance mathématique de X est le vecteur déterministe (sous réserve d'existence) :

$$\mathbb{E}(X) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_n) \end{pmatrix} \quad (3.37)$$

Propriété 3.4.2 Soit X_1 et X_2 deux v.a.d. admettant une espérance et $\lambda \in \mathbb{R}$. Alors :

$$\mathbb{E}(\lambda X_1 + X_2) = \lambda \mathbb{E}(X_1) + \mathbb{E}(X_2) \quad (3.38)$$

L'espérance mathématique est un opérateur linéaire. Ce résultat est valable quelle que soit la nature de la v.a.

Covariance

Définition 3.4.6 Soit deux v.a. X_i et X_j admettant un moment d'ordre deux. On appelle *covariance* de X_i et X_j la quantité déterministe :

$$\text{cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)] \quad (3.39)$$

Comme son nom l'indique, la covariance de deux v.a. informe sur la dispersion de l'une par rapport à l'autre et plus précisément la dispersion du couple (X_i, X_j) par rapport au couple $(\mathbb{E}X_i, \mathbb{E}X_j)$.

Proposition 3.4.3 (Décomposition de König-Huygens)

$$\text{cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \quad (3.40)$$

Démonstration 3.4.1 Il suffit de développer la formule de l'équation (3.39) et utiliser les propriétés de l'espérance. \square

Remarque 3.4.1 La variance d'une v.a. X n'est rien d'autre que la covariance de X avec elle-même : $\mathbb{V}(X) = \text{cov}(X, X)$.

Remarque 3.4.2 Précisons pourquoi l'existence de la covariance de deux v.a. X et Y est subordonnée à l'existence de leurs moments d'ordre deux. Il suffit d'utiliser l'inégalité de Cauchy-Schwarz :

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)} \quad (3.41)$$

Démonstration 3.4.2 (Inégalité de Cauchy-Schwarz)

Soit $\lambda \in \mathbb{R}$. Étudions la quantité $\mathbb{E}(|X| - \lambda|Y|)^2$.

$$\mathbb{E}(|X| - \lambda|Y|)^2 = \mathbb{E}(X^2) - 2\lambda\mathbb{E}(|XY|) + \lambda^2\mathbb{E}(Y^2)$$

Il s'agit d'un polynôme de degré 2 en λ . On cherche à annuler cette équation en λ . Or cette quantité est toujours positive, ce qui signifie que le discriminant est négatif. Ce dernier vaut :

$$\Delta' = \lambda^2[\mathbb{E}^2(|XY|) - \mathbb{E}(X^2)\mathbb{E}(Y^2)]$$

D'où le résultat. \square

Propriété 3.4.3 La covariance est un opérateur symétrique bilinéaire :

- $\text{cov}(X, Y) = \text{cov}(Y, X)$ pour toutes v.a. X et Y ;

- $\text{cov}(\lambda X + Y, Z) = \lambda \text{cov}(X, Z) + \text{cov}(Y, Z)$ pour tout $\lambda \in \mathbb{R}$ et toutes v.a. X, Y et Z .

Proposition 3.4.4 (admis)

Deux v.a. X et Y sont indépendantes si pour toutes fonctions mesurable ϕ et ψ quelconques on a (sous réserve d'existence) :

$$\mathbb{E}\phi(X)\psi(Y) = \mathbb{E}\phi(X)\mathbb{E}\psi(Y) \quad (3.42)$$

Proposition 3.4.5 Si deux v.a. X et Y sont indépendantes alors leur covariance est nulle.

LA RÉCIPROQUE EST FAUSSE EN GÉNÉRAL.

Démonstration 3.4.3 Il suffit d'utiliser la décomposition de König-Huygens et la proposition ci-dessus. \square

Variance

Définition 3.4.7 Soit un n -uplet de v.a. $X = (X_1, \dots, X_n)$. On appelle variance de X la matrice déterministe d'ordre n (sous réserve d'existence) :

$$\mathbb{V}(X) = (\text{cov}(X_i, X_j))_{(i,j) \in \llbracket 1, n \rrbracket^2} \quad (3.43)$$

Cette matrice est appelée communément *matrice de covariances* ou *matrice de variances-covariances*.

Proposition 3.4.6 (admis)

Une matrice de covariances est définie positive.

Proposition 3.4.7 Soit X et Y deux v.a. admettant une variance et $\lambda \in \mathbb{R}$. Alors :

- $\mathbb{V}(\lambda X) = \lambda^2 \mathbb{V}(X)$;
- $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\text{cov}(X, Y)$. Si X et Y sont *indépendantes* alors

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

3.4.5 Inégalité de Jensen

Théorème 3.4.1 (admis)

Soit X une v.a. définie sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et admettant une espérance. Soit ψ une application \mathcal{A} -mesurable convexe telle que $\mathbb{E}\psi(X)$ existe. Alors on a :

$$\mathbb{E}\psi(X) \geq \psi(\mathbb{E}(X)) \quad (3.44)$$

L'inégalité est stricte si la convexité l'est et elle est inversée en cas de concavité.

3.4.6 Fonction génératrice

N.B. : dans cette sous-section, on supposera que les v.a.d. sont à valeurs dans l'espace mesurable $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.

Définition 3.4.8 Soit X_1 et X_2 deux v.a.d. La fonction génératrice du couple (X_1, X_2) vaut :

$$G_{X_1, X_2}(z_1, z_2) = \sum_{x_1 \in \mathbb{N}} \sum_{x_2 \in \mathbb{N}} z_1^{x_1} z_2^{x_2} \mathbb{P}(X_1 = x_1, X_2 = x_2) \quad (3.45)$$

où z_1 et z_2 sont des complexes de module inférieur à 1.

On peut généraliser cette définition à n v.a.d.

Proposition 3.4.8 Soient n v.a.d. *indépendantes* X_1, \dots, X_n . Alors :

$$G_{X_1, \dots, X_n}(z_1, \dots, z_n) = \prod_{i=1}^n G_{X_i}(z_i) \quad (3.46)$$

où les z_i sont des complexes de module inférieur à 1.

CHAPITRE 4

Lois discrètes usuelles

4.1 La loi de Dirac

Définition 4.1.1 Une v.a. X suit une loi de Dirac en un point $a \in \mathbb{R}$ (on note $X \sim \delta_a$) si elle est constante et vaut a . X est appelée *variable aléatoire dégénérée*. Son support est le singleton $\mathcal{S}_X = \{a\}$.

Remarque 4.1.1 La tribu associée à une v.a. dégénérée est la tribu grossière $\{\emptyset, \Omega\}$.

Fonction de masse

$$\forall x \in \mathbb{R} \quad \mathbb{P}(X = x) = \begin{cases} 0 & \text{si } x \neq a \\ 1 & \text{si } x = a \end{cases} \triangleq \delta_a(x) \quad (4.1)$$

δ_a est appelée *masse de Dirac* ou *mesure de Dirac* en a .

Espérance $\mathbb{E}(X) = a$

Variance $\mathbb{V}(X) = 0$

Fonction génératrice $\forall z \in \mathbb{C}, |z| \leq 1 \quad G_X(z) = z^a$

4.2 Caractérisation des variables aléatoires discrètes à valeurs dans $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$

Proposition 4.2.1 (admis)

Soit une v.a.d. X définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans l'espace probabilisable $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. Alors il existe un sous-ensemble I de \mathbb{N} et une suite $(p_x)_{x \in \mathbb{N}}$ telle que la loi de probabilité de X s'écrive sous la forme :

$$\mathbb{P}_X = \sum_{x \in I} p_x \delta_x \quad (4.2)$$

Évidemment, $p_x = \mathbb{P}(X = x)$ et I est le support de X .

Cette écriture qui semble artificielle, sera très utile pour appréhender la deuxième partie de ce cours.

4.3 Loi de Bernoulli $\mathcal{B}(p)$

Définition 4.3.1 X suit de loi de Bernoulli de paramètre p ($X \sim \mathcal{B}(p)$), $p \in [0, 1]$, si sa fonction de masse est égale à :

$$\forall x \in \{0, 1\} \quad \mathbb{P}(X = x) = \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0 \end{cases} \quad (4.3)$$

On peut écrire également :

$$\forall x \in \{0, 1\} \quad \mathbb{P}(X = x) = p^x(1 - p)^{(1-x)} \quad (4.4)$$

ou encore :

$$\forall x \in \mathbb{R} \quad \mathbb{P}(X = x) = p^x(1 - p)^{1-x} \mathbf{1}_{\{0,1\}}(x) \quad (4.5)$$

où $\forall B \in \mathcal{B}(\mathbb{R})$:

$$\mathbf{1}_B(x) = \begin{cases} 1 & \text{si } x \in B \\ 0 & \text{sinon} \end{cases} \quad (4.6)$$

est appelée *fonction indicatrice*.

ou encore :

$$\forall x \in \mathbb{R} \quad \mathbb{P}(X = x) = p\delta_1(x) + (1 - p)\delta_0(x) \quad (4.7)$$

Remarque 4.3.1 Le support d'une Bernoulli est $\{0, 1\}$.

Modélisation La loi de Bernoulli est utilisée pour modéliser des phénomènes dont l'issue est binaire : pile ou face, vrai ou faux, l'individu est-il chômeur ou non, l'élève a-t-il validé son année ou non, etc. C'est le statisticien qui attribue la valeur 1 ou 0 selon le résultat d'intérêt. Exemple : on souhaite modéliser le phénomène du chômage. La valeur 1 sera affectée de la probabilité $p = \mathbb{P}(\text{"être chômeur"})$, 0 sinon.

Espérance $\mathbb{E}(X) = p$

Variance $\mathbb{V}(X) = p(1 - p)$

Fonction génératrice $\forall z \in \mathbb{C}, |z| \leq 1 \quad G_X(z) = 1 - p(1 - z)$

4.4 Loi binomiale $\mathcal{B}(n, p)$

Définition 4.4.1 $X \sim \mathcal{B}(n, p)$, $n \in \mathbb{N}^*$ et $p \in [0, 1]$, si sa fonction de masse est égale à :

$$\forall x \in \llbracket 0, n \rrbracket \quad \mathbb{P}(X = x) = C_n^x p^x (1 - p)^{n-x} \quad (4.8)$$

Modélisation La loi binomiale est l'une des plus utilisées lorsqu'il s'agit de modéliser un nombre d'individus (au sens statistique) parmi n vérifiant une certaine propriété : nombre d'élèves parmi 130 ayant validé l'année, nombre de "6" obtenus en n lancers d'un dé, etc.

Espérance $\mathbb{E}(X) = np$

Variance $\mathbb{V}(X) = np(1 - p)$

Fonction génératrice $\forall z \in \mathbb{C}, |z| \leq 1 \quad G_X(z) = (1 - p(1 - z))^n$

Lien entre la loi de Bernoulli et la loi binomiale

Proposition 4.4.1 Soit X_1, \dots, X_n n v.a. *indépendantes* et identiquement distribuées (iid) de loi $\mathcal{B}(p)$. Alors :

$$\sum_{i=1}^n X_i \sim \mathcal{B}(n, p) \quad (4.9)$$

Remarque 4.4.1 $\mathcal{B}(p) = \mathcal{B}(1, p)$

4.5 Loi de Poisson $\mathcal{P}(\lambda)$

Définition 4.5.1 $X \sim \mathcal{P}(\lambda)$, $\lambda \in \mathbb{R}^{*+}$, si sa fonction de masse est égale à :

$$\forall x \in \mathbb{N} \quad \mathbb{P}(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad (4.10)$$

Modélisation La loi de Poisson est souvent utilisée pour modéliser un nombre d'événements rares (suicides, pannes, etc.) et les phénomènes liés aux files d'attente (fréquentation d'un guichet, etc.).

Espérance $\mathbb{E}(X) = \lambda$

Variance $\mathbb{V}(X) = \lambda$

Fonction génératrice $\forall z \in \mathbb{C}, |z| \leq 1 \quad G_X(z) = e^{\lambda(z-1)}$

Proposition 4.5.1 Soit $X \sim \mathcal{B}(n, p)$. On suppose que p dépend de n de telle sorte que :

$$\lim_{n \rightarrow +\infty} np_n = \lambda > 0$$

Alors à partir d'un certain rang, la loi binomiale $\mathcal{B}(n, p)$ peut être approchée par une loi de Poisson $\mathcal{P}(\lambda)$.

En pratique, lorsque $n > 30$ et $np < 5$, l'approximation de la loi binomiale par la loi de Poisson est permise.

4.6 Loi géométrique $\mathcal{G}(p)$

Définition 4.6.1 $X \sim \mathcal{G}(p)$, $p \in]0, 1]$, si sa fonction de masse est égale à :

$$\forall x \in \mathbb{N}^* \quad \mathbb{P}(X = x) = p(1 - p)^{x-1} \quad (4.11)$$

Modélisation La loi géométrique est utilisée pour modéliser le nombre d'essais nécessaires (les essais étant indépendants) pour observer la première fois un événement d'intérêt, celui-ci se réalisant avec une probabilité p . Exemple : nombre de lancers nécessaires pour obtenir un "6".

Espérance $\mathbb{E}(X) = \frac{1}{p}$

Variance $\mathbb{V}(X) = \frac{1-p}{p^2}$

Fonction génératrice $\forall z \in \mathbb{C}, |z| \leq 1 \quad G_X(z) = \frac{pz}{1 - z(1 - p)}$

4.7 Loi de Pascal d'ordre n $\mathcal{Pa}(n, p)$

Définition 4.7.1 $X \sim \mathcal{Pa}(n, p)$, $n \in \mathbb{N}^*$ et $p \in [0, 1]$, si sa fonction de masse est égale à :

$$\forall x \in \llbracket n, +\infty \llbracket \quad \mathbb{P}(X = x) = C_{x-1}^{n-1} p^n (1 - p)^{x-n} \quad (4.12)$$

Heuristique de la loi $\forall x \in \mathbb{R} \quad \mathbb{P}(X = x) = p C_{x-1}^{n-1} p^{n-1} (1 - p)^{(x-1)-(n-1)} \mathbf{1}_{\llbracket n, +\infty \llbracket}(x)$

Modélisation Elle sert à modéliser le nombre d'essais nécessaires pour observer n événements de probabilité p .

Espérance $\mathbb{E}(X) = \frac{n}{p}$

Variance $\mathbb{V}(X) = \frac{n(1-p)}{p^2}$

Fonction génératrice $\forall z \in \mathbb{C}, |z| \leq 1 \quad G_X(z) = \left(\frac{pz}{1 - z(1 - p)} \right)^n$

Relation entre la loi de Pascal d'ordre n $\mathcal{Pa}(n, p)$ et la loi géométrique $\mathcal{G}(p)$

Proposition 4.7.1 Soient n v.a. X_1, \dots, X_n indépendantes et de même loi $\mathcal{G}(p)$. Alors :

$$\sum_{i=1}^n X_i \sim \mathcal{Pa}(n, p) \quad (4.13)$$

4.8 Loi binomiale négative $\mathcal{BN}(n, p)$

Définition 4.8.1 $X \sim \mathcal{BN}(n, p)$ si sa fonction de masse est égale à :

$$\forall x \in \mathbb{N} \quad \mathbb{P}(X = x) = C_{n+x-1}^{n-1} p^n (1-p)^x \quad (4.14)$$

La loi binomiale négative $\mathcal{BN}(n, p)$ n'est rien d'autre que la loi de Pascal d'ordre n $\mathcal{Pa}(n, p)$ normalisée à 0.

Proposition 4.8.1 Soit $X \sim \mathcal{Pa}(n, p)$. Alors $X - n \sim \mathcal{BN}(n, p)$.

Proposition 4.8.2 Soit $X \sim \mathcal{BN}(n, p)$. Posons $P = \frac{1-p}{p}$. Alors :

$$\begin{aligned} \mathbb{E}(X) &= nP \\ \mathbb{V}(X) &= nP(1+P) \end{aligned} \quad (4.15)$$

4.9 Loi hypergéométrique $\mathcal{H}(N, n, p)$

Définition 4.9.1 $X \sim \mathcal{H}(N, n, p)$ si sa fonction de masse est égale à :

$$\forall x \in [\max(0, n - N(1-p)); \min(n, Np)] \quad \mathbb{P}(X = x) = \frac{C_{Np}^x C_{N-Np}^{n-x}}{C_N^n} \quad (4.16)$$

Np doit être entier.

Modélisation Soit une population de N individus vérifiant un certain caractère avec une probabilité p . On tire un échantillon de n individus parmi N . Une v.a. de loi hypergéométrique représente le nombre d'individus parmi n présentant le caractère en question.

Remarque 4.9.1

- Le ratio $\frac{n}{N}$ est appelé *taux de sondage*.
- C_N^n est le nombre d'échantillons possibles de taille n parmi une population de N individus.
- C_{Np}^x est le nombre de cas favorables à ce que x individus parmi Np vérifient le caractère d'intérêt.

Espérance $\mathbb{E}(X) = np$

Variance $\mathbb{V}(X) = \frac{N-n}{N-1} np(1-p)$

Fonction génératrice

$$\forall z \in \mathbb{C}, |z| \leq 1 \quad G_X(z) = \frac{C_{N-Np}^n}{C_N^n} {}_2F_1(-n, -Np; N - Np - n + 1; z) \quad (4.17)$$

où :

$${}_2F_1(-n, -Np; N - Np - n + 1; z) = \sum_{x \in \mathbb{N}} c_x z^x \quad (\text{série hypergéométrique})$$

avec $c_0 = 1$ et :

$$\forall x \in \mathbb{N} \quad \frac{c_{x+1}}{c_x} = \frac{(x-n)(x-Np)}{x+N(1-p)-n} \cdot \frac{1}{x+1}$$

Relation entre la loi $\mathcal{B}(n, p)$ et la loi $\mathcal{H}(N, n, p)$

Proposition 4.9.1 Lorsque N tend vers l'infini, alors la loi hypergéométrique $\mathcal{H}(N, n, p)$ est approximativement équivalente à la loi binomiale $\mathcal{B}(n, p)$.

Proposition 4.9.2 Soit n v.a. X_1, \dots, X_n (parmi N) *non indépendantes* mais de même loi de Bernoulli $\mathcal{B}(p)$. Alors :

$$\sum_{i=1}^n X_i \sim \mathcal{H}(N, n, p) \quad (4.18)$$

CHAPITRE 5

Exercice récapitulatif

Soit un entier strictement positif n . Soit un couple de variables aléatoires discrètes (X, Y) de loi jointe :

$$\forall (x, y) \in \llbracket 1, n \rrbracket^2, \quad \mathbb{P}(X = x, Y = y) = K \mathbf{1}_{\{x \leq y\}} \quad (5.1)$$

où K est un réel positif.

1. Déterminer la constante K .
2. Déterminer les lois marginales de X et Y .
3. Déterminer leurs espérances et variances respectives.
4. Calculer, pour tout $k \in \mathbb{N}$, $\mathbb{P}(X = Y - k)$.
5. Calculer pour tout $l \in \mathbb{N}^*$, $\mathbb{P}(X = Y + l)$.
6. Calculer $\text{cov}(X, Y)$.

PARTIE II

Les fondements mathématiques des
probabilités : une introduction aux théories
de la mesure et de l'intégrale de Lebesgue

CHAPITRE 6

Motivation

Nous abordons la partie « ardue » de ce cours. Pour tous ceux qui douteraient, à tort ou à raison, de l'utilité d'une telle abstraction, voici dans un premier temps quelques citations (source : Tassi [5]) :

- Jean Dieudonné : « (...) le cercle vicieux de la définition [d'une probabilité] en termes de cas équipossibles, l'apparition de variables aléatoires discrètes mais à un nombre infini de valeurs possibles, telles que celles de Poisson, et des variables aléatoires à loi absolument continue telles que des variables aléatoires normales, ont amené graduellement à une idéalisation mathématique qui engloberait toutes ces possibilités » ;
- John Maynard Keynes à propos des probabilités : « les savants y décèlent un relent d'astrologie ou d'alchimie » (1921) ;
- Richard Edler von Mises : « le calcul des probabilités n'est pas une discipline mathématique » (1919).

Depuis Keynes et Von Mises, la théorie des probabilités s'est « mathématisée » avec les contributions fondamentales de Kolmogorov, de Borel¹ (théorie de la mesure) et Lebesgue (intégrale du même nom) et elle est actuellement utilisée dans tous les domaines de la vie scientifique, économique et sociale.

Dans cette partie, nous n'avons pas pour ambition de faire des élèves des caciques de la théorie abstraite de la mesure (il faudrait trois fois plus d'heures de cours) mais plutôt de les initier aux différents concepts mathématiques sur lesquels s'appuie la théorie des probabilités. Et il s'agit là de connaissances minimales devant être acquises par un élève de l'Ensaï. Toutefois, ceux parmi les élèves ingénieurs qui souhaitent s'orienter vers les filières « gestion des risques et ingénierie financière » et « génie statistique » devront faire preuve d'une attention supplémentaire en vue du cours de calcul stochastique (théorie de l'intégration par rapport à un processus stochastique bien particulier) qu'ils devront suivre en 3^e année.

L'un des grands apports de la théorie de la mesure (une probabilité étant une mesure de masse 1) est qu'elle institue un cadre général d'analyse d'une variable aléatoire, qu'elle soit

1. Fondateur de l'Institut statistique de l'université de Paris (Isup) en 1922.

discrète ou continue. Ainsi, dans les cours traditionnels d'introduction aux probabilités, on est souvent amené à croire que la définition de l'espérance mathématique (par exemple) est différente selon la nature de la variable aléatoire. Elle vaudra :

$$\mathbb{E}(X) = \sum_{x \in \mathcal{S}_X} x f_X(x) \quad (6.1)$$

si X est discrète (f_X étant la fonction de masse) et :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx \quad (6.2)$$

si X est continue où f_X est la fonction de densité.

En réalité, ces deux espérances ont la même définition originelle :

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\{\omega\}) = \int_{\mathbb{R}} x d\mathbb{P}_X(\{x\}) \quad (6.3)$$

Lorsque X est discrète (à valeurs dans \mathbb{N} par exemple) :

$$\mathbb{P}_X = \sum_{x \in \mathbb{N}} p_x \delta_x \quad (6.4)$$

et si X est continue :

$$\mathbb{P}_X = f_X \cdot \lambda \quad (6.5)$$

où λ est la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Par ailleurs l'intégrale de Riemann est insuffisante pour ce cadre d'analyse. En effet, en théorie des probabilités, on peut être amené à intégrer des fonctions très irrégulières. L'un des exemples de base les plus connus est la fonction de Dirichlet sur $[0, 1]$ ² : $\mathbb{1}_{\mathbb{Q}}(x)$, qui n'est pas Riemann-intégrable. Il existe également des phénomènes qui peuvent être modélisés par un mélange de lois discrètes et continues.

2. L'intégrale de Lebesgue répond à la question : « quelle est la probabilité qu'un nombre tiré entre 0 et 1 soit rationnel ? »

CHAPITRE 7

Mesure positive

On reprendra les définitions vues (tribus, espace mesurable, etc.) dans le chapitre 1 de ce cours. Par abus de langage, on nommera **mesure** une mesure positive.

7.1 Définitions

Définition 7.1.1 On appelle *mesure positive* sur un espace mesurable (Ω, \mathcal{A}) , toute application μ définie sur \mathcal{A} à valeurs dans $\overline{\mathbb{R}}^+ = \mathbb{R}^+ \cup \{+\infty\}$ et σ -additive, c'est-à-dire que pour toute suite $(A_n)_{n \in \mathbb{N}^*}$ d'éléments de \mathcal{A} deux à deux *disjoints* :

$$\mu\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_{n \in \mathbb{N}^*} \mu(A_n) \quad (7.1)$$

Définition 7.1.2

- On appelle *masse* d'une mesure μ le nombre positif (fini ou non) $\mu(\Omega)$.
- Une mesure de masse 1 est appelée *probabilité* ou *mesure de probabilité*.
- μ est dite *finie* si $\forall A \in \mathcal{A} \quad \mu(A) < +\infty$ ou ce qui est équivalent, $\mu(\Omega) < +\infty$.
- μ est dite σ -*finie* s'il existe une suite $(A_n)_{n \in \mathbb{N}^*}$ d'éléments de \mathcal{A} , croissante (au sens de l'inclusion), telle que $\Omega = \bigcup_{n \in \mathbb{N}^*} A_n$ et $\forall n \in \mathbb{N}^* \quad \mu(A_n) < +\infty$.
- Le triplet $(\Omega, \mathcal{A}, \mu)$ est appelé *espace mesuré* ou espace probabilisé si μ est une probabilité.

Exemple 7.1.1

La masse de Dirac est une mesure de probabilité, donc une mesure finie.

7.2 Propriétés

On retrouve toutes les propriétés vérifiées par une probabilité.

Théorème 7.2.1 Soit μ une mesure sur un espace mesurable (Ω, \mathcal{A}) . Alors :

- (i) $\mu(\emptyset) = 0$;

(ii) Si A_1, \dots, A_n sont des éléments de \mathcal{A} deux à deux disjoints, alors μ est additive :

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i) \quad (7.2)$$

(iii) Si $A, B \in \mathcal{A}$ avec $A \subset B$ alors $\mu(A) \leq \mu(B)$.

(iv) μ est sous- σ -additive : pour toute suite $(A_n)_{n \in \mathbb{N}^*}$ d'éléments quelconques de \mathcal{A}

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

Démonstration 7.2.1

(i) Soit $A \in \mathcal{A}$ tel que $\mu(A) < +\infty$. Soit la suite $(A_n)_{n \in \mathbb{N}^*}$ d'éléments deux à deux disjoints définie de la manière suivante : $A_1 = A$ et $\forall n \geq 2, A_n = \emptyset$. D'après la σ -additivité, il vient :

$$\mu(A) = \mu(A) + \sum_{n \geq 2} \mu(A_n)$$

Or $\sum_{n \geq 2} \mu(A_n) = \mu\left(\bigcup_{n \geq 2} A_n\right) = \mu(\emptyset)$. D'où le résultat requis.

(ii) On pose $B_i = A_i \forall i \in \llbracket 1, n \rrbracket$ et $B_i = \emptyset \forall i > n$. On conclut en utilisant la σ -additivité de μ et le résultat précédent.

(iii) Si $A \subset B$ alors on peut écrire $B = A \cup (B \setminus A)$; l'union étant disjointe, on utilise l'additivité de μ . Par suite : $\mu(B) = \mu(A) + \mu(B \setminus A)$. D'où l'inégalité requise.

(iv) Soit une suite d'éléments $(B_n)_{n \in \mathbb{N}^*}$ de \mathcal{A} construite de la manière suivante : $B_1 = A_1$, $B_2 = A_2 \cap \bar{A}_1$, ..., $B_n = A_n \cap \bar{A}_{n-1} \cap \dots \cap \bar{A}_1$. Les B_n sont deux à deux disjoints. On vérifiera que $\forall n \in \mathbb{N}^*, B_n \subset A_n$ et :

$$\bigcup_{n \in \mathbb{N}^*} B_n = \bigcup_{n \in \mathbb{N}^*} A_n \quad (\text{démonstration de l'égalité laissée en exercice})$$

Par suite :

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

Théorème 7.2.2 Soit (Ω, \mathcal{A}) un espace mesurable. Soit μ une application additive définie sur \mathcal{A} à valeurs dans $\overline{\mathbb{R}^+}$ et non constante en la valeur $+\infty$. Les conditions suivantes sont équivalentes :

- (i) μ est une mesure sur \mathcal{A} .
- (ii) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'éléments de \mathcal{A} telle que $A_n \uparrow A$, avec $A \in \mathcal{A}$, alors $\mu(A_n) \uparrow \mu(A)$. Il s'agit de la *propriété de continuité monotone* d'une mesure.

Si de plus μ est finie, les deux propositions précédentes sont équivalentes aux conditions suivantes :

(iii) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'éléments de \mathcal{A} telle que $A_n \downarrow A$, avec $A \in \mathcal{A}$, alors $\mu(A_n) \downarrow \mu(A)$.

(iv) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'éléments de \mathcal{A} telle que $A_n \downarrow \emptyset$, alors $\mu(A_n) \downarrow 0$.

Démonstration 7.2.2 Montrons tout d'abord l'équivalence de (i) et (ii).

• (i) \Rightarrow (ii) :

Soit donc $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'éléments de \mathcal{A} telle que $A_n \uparrow A$, avec $A \in \mathcal{A}$. La suite $(A_n)_{n \in \mathbb{N}^*}$ est donc croissante au sens de l'inclusion. Posons :

$$B_1 = A_1 \quad \text{et} \quad \forall n \geq 2 \quad B_n = A_n \setminus A_{n-1}$$

Les B_n sont deux à deux disjoints et on a :

$$\forall n \in \mathbb{N}^* \quad \bigcup_{k=1}^n B_k = A_n \quad \text{et} \quad \bigcup_{n \in \mathbb{N}^*} B_n = A$$

Par suite :

$$\begin{aligned} \mu(A) &= \mu\left(\bigcup_{n \in \mathbb{N}^*} B_n\right) \\ &= \sum_{n \in \mathbb{N}^*} \mu(B_n) \\ &\stackrel{\Delta}{=} \lim_{n \rightarrow +\infty} \sum_{k=1}^n \mu(B_k) \\ &= \lim_{n \rightarrow +\infty} \mu\left(\bigcup_{k=1}^n B_k\right) \quad \text{par additivité} \\ &= \lim_{n \rightarrow +\infty} \mu(A_n) \end{aligned}$$

On sait également que la suite $(\mu(A_n))_{n \in \mathbb{N}^*}$ est croissante. D'où :

$$\mu(A_n) \uparrow \mu(A)$$

• (ii) \Rightarrow (i) :

Soit $(B_n)_{n \in \mathbb{N}^*}$ une suite d'éléments de \mathcal{A} deux à deux disjoints tels que $B = \bigcup_{n \in \mathbb{N}^*} B_n$. Nous

allons montrer la σ -additivité de μ . Posons $A_n = \bigcup_{k=1}^n B_k$. On a $A_n \uparrow B$. Par suite :

$$\sum_{k=1}^n \mu(B_k) = \mu\left(\bigcup_{k=1}^n B_k\right) = \mu(A_n) \uparrow \mu(B)$$

Il vient alors :

$$\lim_{n \rightarrow +\infty} \sum_{k=1}^n \mu(B_k) = \mu\left(\bigcup_{n \in \mathbb{N}^*} B_n\right) \quad \text{par continuité monotone de } \mu$$

autrement dit :

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

Pour prouver l'équivalence entre les 4 conditions lorsque μ est finie, on se reportera à Montfort [3]. \square

7.3 Mesures sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

Les deux mesures suivantes sont fondamentales puisque ce sont les seules quasiment qui serviront en probabilités.

7.3.1 La mesure de Lebesgue

Définition 7.3.1 On appelle *mesure de Lebesgue* sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, la mesure λ telle que pour tout intervalle $]a, b]$, $a, b \in \mathbb{R}$ avec $a < b$:

$$\lambda(]a, b]) = b - a \quad (7.3)$$

Remarque 7.3.1 Par cette définition, on admettra l'existence et l'unicité de la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Propriété 7.3.1

- (i) λ est σ -finie mais non finie.
- (ii) $\forall a, b \in \mathbb{R} \quad \lambda(]a, b]) = \lambda(]a, b[) = \lambda([a, b]) = \lambda([a, b[)$.
- (iii) $\forall a \in \mathbb{R} \quad \lambda(\{a\}) = 0$; on dit que λ est une mesure *continue*.
- (iv) La mesure de Lebesgue est invariante par symétrie et par translation, c'est-à-dire que pour tout $x \in \mathbb{R}$, $A \in \mathcal{B}(\mathbb{R})$, on a :

$$\lambda(A^{-1}) = \lambda(x + A) = \lambda(A) \quad (7.4)$$

où :

$$\begin{aligned} A^{-1} &= \{y \in \mathbb{R} : -y \in A\} \\ x + A &= \{y \in \mathbb{R} : y = x + z, z \in A\} \end{aligned} \quad (7.5)$$

Démonstration 7.3.1

- (i) λ n'est pas finie car par exemple pour $x \in \mathbb{R}$, on a $\lambda(]-\infty, x]) = +\infty$. En revanche, elle est σ -finie car \mathbb{R} peut s'écrire de la manière suivante : $\mathbb{R} = \bigcup_{n \in \mathbb{N}}]-n, n]$ et $\lambda(]-n, n]) = 2n < +\infty, \forall n \in \mathbb{N}$.
- (ii) Nous montrerons seulement l'égalité $\lambda(]a, b]) = \lambda(]a, b[)$, les autres égalités se démontrant de la même manière. Posons $\forall n \in \mathbb{N}^*, A_n = \left]a, b - \frac{1}{n}\right]$. Il est clair que la suite $(A_n)_{n \in \mathbb{N}^*}$ est une suite croissante de $\mathcal{B}(\mathbb{R})$. On a de plus : $A_n \uparrow A$ avec $A =]a, b[\in \mathcal{B}(\mathbb{R})$. D'après la propriété de continuité monotone de λ , il vient $\lambda(A_n) \uparrow \lambda(A)$.

Remarque 7.3.2 On veillera à utiliser la propriété de continuité croissante. En effet, λ n'étant pas finie, la propriété de continuité décroissante ne peut être invoquée.

Remarque 7.3.3 On définit de la même façon la mesure de Lebesgue sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ ($d \in \mathbb{N}^*$), la mesure notée λ_d , telle que pour tout pavé $A = \prod_{i=1}^d]a_i, b_i]$ on ait :

$$\lambda_d(A) = \prod_{i=1}^d (b_i - a_i) \quad (7.6)$$

7.3.2 La mesure de comptage

Définition 7.3.2 On appelle *mesure de comptage* ou *mesure de dénombrement* sur l'espace mesurable $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, la mesure :

$$\mu_c = \sum_{x \in \mathbb{N}} \delta_x \quad (7.7)$$

où δ_x est la masse de Dirac en x . La mesure de comptage est une mesure discrète sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Propriété 7.3.2 A l'instar de la mesure de Lebesgue, la mesure de comptage est σ -finie mais non finie.

Démonstration 7.3.2

- μ_c est σ -finie. Il suffit d'écrire $\mathbb{N} = \bigcup_{n \in \mathbb{N}}]0, n]$.

- Mais $\mu_c(\mathbb{N}) = \sum_{x \in \mathbb{N}} \delta_x(\mathbb{N}) = \sum_{x \in \mathbb{N}} 1 = +\infty$. □

7.4 Mesure image

Définition 7.4.1 (mesure image)

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, f une application mesurable de (Ω, \mathcal{A}) dans (E, \mathcal{E}) . On appelle *mesure image* de μ par f , toute application définie sur \mathcal{E} $\mu_f = \mu \circ f^{-1}$:

$$\forall B \in \mathcal{E} \quad \mu_f(B) = \mu(f^{-1}(B)) \quad (7.8)$$

Exemple 7.4.1 Soit X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs (E, \mathcal{E}) . La loi image de \mathbb{P} par X est $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$.

CHAPITRE 8

Mesurabilité

Nous complétons dans ce chapitre la notion de mesurabilité déjà donnée dans la définition 1.3.1 page 9 pour définir une variable aléatoire.

8.1 Application mesurable

On rappelle que si f est une application de Ω dans E , f^{-1} est appelée *application réciproque* de f . C'est un *homomorphisme* pour les opérations ensemblistes et en particulier de réunion et intersection quelconques et de complémentation :

- $f^{-1}\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f^{-1}(A_i)$;
- $f^{-1}\left(\bigcap_{i \in I} A_i\right) = \bigcap_{i \in I} f^{-1}(A_i)$;
- $f^{-1}(\bar{A}) = \overline{f^{-1}(A)}$.

où I est un ensemble quelconque.

Définition 8.1.1

Soit (Ω, \mathcal{A}) et (E, \mathcal{E}) deux espaces mesurables. Une application f définie sur Ω à valeurs dans E est dite mesurable si :

$$\forall B \in \mathcal{E} \quad f^{-1}(B) \in \mathcal{A} \quad \text{ou bien} \quad f^{-1}(\mathcal{E}) \subset \mathcal{A} \quad (8.1)$$

f dite souvent \mathcal{A} -mesurable. On notera souvent $f^{-1}(B) = \{f \in B\} = \{\omega \in \Omega : f(\omega) \in B\}$.

Remarque 8.1.1

Si (E, \mathcal{E}) est un espace topologique (par exemple $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$), f est dite *borélienne*. On rappelle qu'un *espace topologique* est un couple (E, \mathcal{T}) où E est un ensemble et \mathcal{T} une famille de parties de E , vérifiant les propriétés suivantes :

- (i) l'ensemble vide et E appartiennent à \mathcal{T} ;
- (ii) \mathcal{T} est stable par réunion quelconque ;

(iii) \mathcal{T} est stable par intersection finie.

Les éléments de \mathcal{T} sont appelés des ouverts.

Exemple 8.1.1 On a déjà vu qu'une v.a. était une application mesurable.

Proposition 8.1.1

Soit (Ω, \mathcal{A}) un espace mesurable. Un ensemble $A \subset \Omega$ est mesurable (c'est-à-dire $A \in \mathcal{A}$) si et seulement si $\mathbb{1}_A$ est mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Démonstration 8.1.1

La condition suffisante est évidente car $\mathbb{1}_A$ supposée mesurable et en particulier : $\mathbb{1}_A^{-1}(\{1\}) = A \in \mathcal{A}$. Montrons la condition nécessaire. Pour cela, nous distinguons quatre cas :

- Si B contient 0 et 1 alors $\mathbb{1}_A^{-1}(B) = \Omega \in \mathcal{A}$;
- Si B ne contient ni 0 ni 1 alors $\mathbb{1}_A^{-1}(B) = \emptyset \in \mathcal{A}$;
- Si B contient 1 mais pas 0 alors $\mathbb{1}_A^{-1}(B) = A \in \mathcal{A}$;
- Si B contient 0 mais pas 1 alors $\mathbb{1}_A^{-1}(B) = \bar{A} \in \mathcal{A}$. □

Exemple 8.1.2 Un exemple classique d'application non mesurable sur un espace mesurable (Ω, \mathcal{A}) est l'application $\mathbb{1}_A$ avec $A \notin \mathcal{A}$.

Théorème 8.1.1 (admis)

Soit Ω un ensemble et (E, \mathcal{E}) un espace mesurable et soit f une application de Ω dans E . Alors :

- (i) $f^{-1}(\mathcal{E}) = \{f^{-1}(B) : B \in \mathcal{E}\}$ est une tribu sur Ω , appelée *tribu engendrée* par f , c'est-à-dire la plus petite tribu \mathcal{T} sur Ω rendant mesurable l'application f .
- (ii) Pour toute classe de parties \mathcal{C} de E engendrant \mathcal{E} ($\sigma(\mathcal{C}) = \mathcal{E}$), on a :

$$f^{-1}(\sigma(\mathcal{C})) = \sigma(f^{-1}(\mathcal{C})) \quad (8.2)$$

Corollaire 8.1.1 (Critère de mesurabilité, admis)

Soit (Ω, \mathcal{A}) et (E, \mathcal{E}) deux espaces mesurables. Soit \mathcal{C} une classe de parties de E engendrant \mathcal{E} (c'est-à-dire $\sigma(\mathcal{C}) = \mathcal{E}$). Alors f est mesurable si et seulement si $f^{-1}(\mathcal{C}) \subset \mathcal{A}$.

Corollaire 8.1.2

Soit (Ω, \mathcal{A}) et (E, \mathcal{E}) deux espaces topologiques munis de leurs tribus boréliennes respectives. Toute application continue f de Ω dans E est mesurable.

Démonstration 8.1.2

Soit \mathcal{O}_1 (resp. \mathcal{O}_2), la famille des ouverts de Ω (resp. E). f étant continue, on a $f^{-1}(\mathcal{O}_2) \subset \mathcal{O}_1$ (car l'image réciproque continue d'un ouvert est un ouvert). D'après le théorème 8.1.1, on a :

$$f^{-1}(\mathcal{E}) = f^{-1}(\sigma(\mathcal{O}_2)) = \sigma(f^{-1}(\mathcal{O}_2)) \subset \sigma(\mathcal{O}_1) = \mathcal{A}$$

Corollaire 8.1.3

Soit f une application de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. f est mesurable si et seulement si :

$$f^{-1}(]-\infty, x]) \in \mathcal{A} \quad \forall x \in \mathbb{R} \quad (8.3)$$

Démonstration 8.1.3

Il s'agit de l'application directe du corollaire 8.1.1 en posant $\mathcal{C} = \{]-\infty, x] : x \in \mathbb{R}\}$. \square

Théorème 8.1.2 Soient $(\Omega_i, \mathcal{A}_i)_{i=1,2,3}$ trois espaces mesurables. Soit f (resp. g) une application mesurable de Ω_1 (resp. Ω_2) dans Ω_2 (resp. Ω_3). Alors $g \circ f$ est une application mesurable de Ω_1 dans Ω_3 .

Démonstration 8.1.4 Nous devons montrer que $(g \circ f)^{-1}(\mathcal{A}_3) \subset \mathcal{A}_1$. Nous savons que $(g \circ f)^{-1}(\mathcal{A}_3) = f^{-1}(g^{-1}(\mathcal{A}_3))$. Or g étant \mathcal{A}_2 -mesurable, $g^{-1}(\mathcal{A}_3) \subset \mathcal{A}_2$. Par suite :

$$f^{-1}(g^{-1}(\mathcal{A}_3)) \subset f^{-1}(\mathcal{A}_2)$$

Et comme f est \mathcal{A}_1 -mesurable, $f^{-1}(\mathcal{A}_2) \subset \mathcal{A}_1$. D'où le résultat requis. \square

8.2 Propriétés des applications mesurables

Théorème 8.2.1 (admis)

Soient f et g deux applications mesurables définies sur un même espace mesurable (Ω, \mathcal{A}) et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. L'application

$$\begin{aligned} h : \Omega &\longrightarrow \mathbb{R}^2 \\ \omega &\longmapsto (f(\omega), g(\omega)) \end{aligned} \quad (8.4)$$

est mesurable si \mathbb{R}^2 est muni de sa tribu borélienne $\mathcal{B}(\mathbb{R}^2)$.

Théorème 8.2.2 (admis)

Soient f et g deux applications mesurables définies sur un même espace mesurable (Ω, \mathcal{A}) et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors les fonctions suivantes sont mesurables : $-f$, $f+g$, αf ($\alpha \in \mathbb{R}$), fg , $\inf(f, g)$, $\sup(f, g)$, $f^+ = \sup(0, f)$ (appelée *partie positive* de f), $f^- = \sup(0, -f)$ (appelée *partie négative* de f) et $|f|$.

Théorème 8.2.3 (admis)

Soit $(f_n)_{n \in \mathbb{N}^*}$ une suite d'applications mesurables définies sur (Ω, \mathcal{A}) et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors les applications suivantes sont mesurables à conditions qu'elles soient finies : $\sup_{n \in \mathbb{N}^*} f_n$, $\inf_{n \in \mathbb{N}^*} f_n$. En particulier si $\lim_{n \rightarrow +\infty} f_n = f$ est finie, alors f est mesurable.

Nous allons à présent nous focaliser sur une famille de fonctions mesurables d'une très grande importance, puisqu'elles vont contribuer à la construction de l'intégrale de Lebesgue : il s'agit des *fonctions étagées*.

Définition 8.2.1 Soit f borélienne définie sur un espace mesurable (Ω, \mathcal{A}) . f est appelée *fonction \mathcal{A} -étagée* (ou étagée) sur Ω si elle est combinaison linéaire *finie* de fonctions indicatrices de parties \mathcal{A} -mesurables de Ω . Autrement dit f s'écrit de la façon suivante :

$$f = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i} \quad (8.5)$$

où les α_i sont réels, et les A_i des éléments de \mathcal{A} .

Théorème 8.2.4 (fondamental, admis)

Soit f une application définie sur (Ω, \mathcal{A}) et à valeurs dans $(\overline{\mathbb{R}}, \mathcal{B}_{\overline{\mathbb{R}}})$.

- (i) Si f est positive, elle est mesurable si et seulement si elle est limite d'une suite croissante d'applications mesurables étagées positives.
- (ii) f est mesurable si et seulement si elle est limite d'une suite de fonctions mesurables étagées.

8.3 Produit d'espaces mesurables

Nous allons donner deux définitions des tribus produits, l'une utilisant les projecteurs et l'autre les pavés mesurables. La notion de produit d'espaces mesurables sera utile en probabilités lorsqu'on étudie les lois de n -uples de variables aléatoires.

Soient $(\Omega_1, \mathcal{A}_1)$ et $(\Omega_2, \mathcal{A}_2)$ deux espaces mesurables. On note p_i ($i = 1, 2$), l'application projection de $\Omega_1 \times \Omega_2$ dans Ω_i . Nous en avons besoin pour définir un espace produit mesurable.

Définition 8.3.1 On appelle *tribu produit* $\mathcal{A}_1 \otimes \mathcal{A}_2$ sur $\Omega_1 \times \Omega_2$ la tribu engendrée par les applications projections p_1 et p_2 , c'est-à-dire la plus petite tribu rendant mesurables ces applications. L'espace $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ est appelé *espace produit mesurable*.

Théorème 8.3.1 Soit g une application définie sur un espace mesurable (F, \mathcal{F}) à valeurs dans un espace produit mesurable $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$. g est mesurable si et seulement si les applications $p_i \circ g$ ($i = 1, 2$) définies sur (F, \mathcal{F}) dans $(\Omega_i, \mathcal{A}_i)$ le sont.

On peut également définir les espaces produits à partir des pavés mesurables.

Définition 8.3.2 On appelle *pavé mesurable* de $\Omega_1 \times \Omega_2$, tout sous-ensemble de la forme $A_1 \times A_2$ où $A_i \in \mathcal{A}_i$ ($i = 1, 2$). Il sera noté $\mathcal{A}_1 * \mathcal{A}_2$.

Exemple 8.3.1 $\Omega_i = \mathbb{R}$. Un pavé mesurable de \mathbb{R}^2 est le produit cartésien de deux \mathbb{R} -boréliens. En particulier, un pavé du type $]a, b[\times]c, d[$ est un pavé mesurable.

Remarque 8.3.1 La famille des pavés mesurables $\mathcal{A}_1 * \mathcal{A}_2$ n'est pas une tribu. En revanche, elle est stable par :

- intersection : $(A_1 \times A_2) \cap (B_1 \times B_2) = (A_1 \cap B_1) \times (A_2 \cap B_2)$ est un pavé mesurable ;

– complémentation : $(A_1 \times A_2)^c = (A_1^c \times A_2) \cup (A_1 \times A_2^c)$.

Par ailleurs, $\Omega_1 \times \Omega_2$ est un pavé mesurable.

Théorème 8.3.2 La tribu produit $\mathcal{A}_1 \otimes \mathcal{A}_2$ est la tribu engendrée par la famille des pavés mesurables $\mathcal{A}_1 * \mathcal{A}_2$.

Exemple 8.3.2 La tribu borélienne $\mathcal{B}(\mathbb{R}^2)$ est la tribu produit $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

On peut généraliser le produit de deux espaces mesurables au produit d'une famille d'espaces mesurables $(\Omega_i, \mathcal{A}_i)_{i \in I}$ (I étant un ensemble dénombrable).

Définition 8.3.3 Soit $\Omega = \prod_{i \in I} \Omega_i$. La tribu engendrée par les parties de Ω est appelée *tribu produit* des tribus \mathcal{A}_i et notée $\bigotimes_{i \in I} \mathcal{A}_i$.

Le couple $\left(\Omega, \bigotimes_{i \in I} \mathcal{A}_i \right)$ est appelé espace mesurable produit des espaces mesurables $(\Omega_i, \mathcal{A}_i)_{i \in I}$.

CHAPITRE 9

Construction de l'intégrale

Dans ce chapitre, nous allons construire l'intégrale par rapport à une mesure positive μ . L'application la plus importante pour nous concerne le cas où μ est la mesure de Lebesgue afin de définir en particulier les lois à densités de probabilité.

Dans tout le chapitre, on se placera dans un espace mesuré $(\Omega, \mathcal{A}, \mu)$.

9.1 Intégrale d'une application étagée positive

Soient A_1, \dots, A_n n éléments de \mathcal{A} formant une partition de Ω . On note \mathcal{E}^+ l'ensemble des applications mesurables étagées positives :

$$\mathcal{E}^+ = \left\{ \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i} : \alpha_i \in \mathbb{R}^+ \right\} \quad (9.1)$$

Définition 9.1.1 Soit une application $f \in \mathcal{E}^+$. On appelle *intégrale de f par rapport à μ* , l'application définie sur \mathcal{E}^+ à valeurs dans $\overline{\mathbb{R}}^+$ définie de la manière suivante :

$$I(f) = \int f \, d\mu \triangleq \sum_{i=1}^n \alpha_i \mu(A_i) \quad (9.2)$$

L'intégrale ne dépend pas de la partition choisie. $I(f)$ est une quantité éventuellement infinie. On conviendra par la suite que $0 \times (+\infty) = 0$.

Notation

L'intégrale de f par rapport à μ se note plus précisément :

$$I(f) = \int_{\Omega} f(\omega) \, d\mu(\omega) \quad (9.3)$$

Propriété 9.1.1

- a) $\forall f \in \mathcal{E}^+, I(f) \geq 0$.
- b) $\forall A \in \mathcal{A}, \mu(A) = \int \mathbb{1}_A \, d\mu = \int_A d\mu(\omega)$ (ceci montre que l'intégrale est une extension de la notion de mesure aux applications mesurables).

- c) $\forall f, g \in \mathcal{E}^+, f + g \in \mathcal{E}^+$ et $I(f + g) = I(f) + I(g)$.
d) $\forall f \in \mathcal{E}^+, \forall a \in \mathbb{R}, af \in \mathcal{E}^+$ et $I(af) = aI(f)$.
e) $\forall f, g \in \mathcal{E}^+, \text{ si } f \leq g \text{ alors } I(f) \leq I(g)$.

Démonstration 9.1.1 Elle est laissée au lecteur à titre d'exercice. \square

Exemple 9.1.1 Prenons le cas où $\mu = \mathbb{P}$ (mesure de probabilité) et $f = X$, une variable aléatoire discrète positive et finie de support $\mathcal{S}_X = \{x_1, \dots, x_n\}$. X peut s'écrire de la manière suivante :

$$X = \sum_{i=1}^n x_i \mathbb{1}_{A_i} \quad \text{où} \quad A_i = \{X = x_i\} \quad (9.4)$$

Par suite :

$$I(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \sum_{i=1}^n x_i \mathbb{P}(A_i) = \sum_{i=1}^n x_i \mathbb{P}(X = x_i) \stackrel{\Delta}{=} \mathbb{E}(X) \quad (9.5)$$

Proposition 9.1.1 (très importante)

Soit $(f_n)_{n \in \mathbb{N}^*}$ une suite croissante d'éléments de \mathcal{E}^+ convergeant point par point vers une application $f \in \mathcal{E}^+$. Alors :

$$\int f_n d\mu \uparrow \int f d\mu \quad \text{quand} \quad n \longrightarrow +\infty \quad (9.6)$$

Remarque 9.1.1 Il s'agit d'une première version de la propriété dite de Beppo-Lévi ou de convergence monotone. C'est une extension à \mathcal{E}^+ de la propriété de continuité monotone d'une mesure.

9.2 Intégrale d'une application mesurable positive

On note \mathcal{M}^+ l'ensemble des applications mesurables positives définies sur (Ω, \mathcal{A}) à valeurs dans $\overline{\mathbb{R}^+}$. Dans cette section, nous allons étendre la notion d'intégrale sur \mathcal{E}^+ à \mathcal{M}^+ .

Définition 9.2.1 Soit μ une mesure sur (Ω, \mathcal{A}) . Soit $f \in \mathcal{M}^+$. On appelle intégrale de f la quantité :

$$I(f) = \int f d\mu = \sup \left\{ \int h d\mu : h \in \mathcal{E}^+, h \leq f \right\} \quad (9.7)$$

Proposition 9.2.1 Soit $f \in \mathcal{M}^+$ et $(h_n)_{n \in \mathbb{N}^*}$ une suite croissante d'éléments de \mathcal{E}^+ avec $f = \lim_{n \rightarrow +\infty} h_n$. Alors :

$$\int f d\mu = \lim_{n \rightarrow +\infty} \int h_n d\mu \quad (9.8)$$

Nous retrouvons les mêmes propriétés que celles de l'intégrale des fonctions étagées.

Proposition 9.2.2

- (i) $\forall f, g \in \mathcal{M}^+, I(f + g) = I(f) + I(g)$;

- (ii) $\forall f \in \mathcal{M}^+, \forall \alpha \in \mathbb{R}^+, I(\alpha f) = \alpha I(f)$;
- (iii) $\forall f, g \in \mathcal{M}^+, \text{ si } f \leq g, \text{ alors } I(f) \leq I(g)$;
- (iv) Soit $(f_n)_{n \in \mathbb{N}^*}$ une suite croissante d'éléments de \mathcal{M}^+ et $f = \lim_{n \rightarrow +\infty} f_n$. Alors :

$$\lim_{n \rightarrow +\infty} \int f_n d\mu = \int f d\mu \quad (9.9)$$

Il s'agit de la propriété de Beppo-Lévi appliquée aux applications de \mathcal{M}^+ .

9.3 Intégrale d'une application mesurable à valeurs réelles

On note \mathcal{M} l'ensemble des applications mesurables définies sur $(\bar{\mathbb{R}}, \mathbb{B}_{\bar{\mathbb{R}}})$.

Définition 9.3.1 Soit $f \in \mathcal{M}$. On définit $f^+ = \max(0, f)$ et $f^- = \max(0, -f)$, éléments de \mathcal{M}^+ et on a l'égalité $f = f^+ - f^-$. f est dite μ -intégrable si :

$$\int f^+ d\mu < +\infty \quad \text{et} \quad \int f^- d\mu < +\infty \quad (9.10)$$

L'intégrale de f par rapport à μ est la quantité :

$$I(f) = \int f d\mu = \int f^+ d\mu - \int f^- d\mu \quad (9.11)$$

Pour établir la μ -intégrabilité d'une application f , on utilise en pratique le critère suivant :

Proposition 9.3.1 f est μ -intégrable si et seulement si $\int |f| d\mu < +\infty$.

Démonstration 9.3.1 Il suffit de voir que $|f| = f^+ + f^-$. □

Théorème 9.3.1 (admis)

On note $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ l'ensemble des applications de \mathcal{M} μ -intégrables. $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ est un espace vectoriel sur \mathbb{R} et $f \mapsto \int f d\mu$ est une forme linéaire.

Proposition 9.3.2 Soient f et g deux éléments de \mathcal{M} et \mathcal{M}^+ respectivement avec g μ -intégrable et $|f| \leq g$. Alors f est μ -intégrable.

Proposition 9.3.3 Si f est μ -intégrable alors $|\int f d\mu| \leq \int |f| d\mu$.

Exemple 9.3.1 (Espérance mathématique)

Soit X une v.a. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On définit l'espérance mathématique de X , l'intégrale de X par rapport à \mathbb{P} :

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) \quad (9.12)$$

sous réserve que $\mathbb{E}(|X|) < +\infty$.

Exemple 9.3.2 (Intégrale par rapport à une mesure discrète)

Une mesure μ est discrète si elle est de la forme :

$$\mu = \sum_{n=1}^{+\infty} p_n \delta_{\omega_n} \quad (9.13)$$

où les p_n sont positifs et $\omega_n \in \Omega$ pour tout $n \in \mathbb{N}^*$. Alors $\forall f \in \mathcal{M}$:

$$\int_{\Omega} f(\omega) d\mu(\omega) = \sum_{n \in \mathbb{N}^*} p_n f(\omega_n) \quad (9.14)$$

f est donc intégrable si et seulement si la série de terme général $p_n f(\omega_n)$ est absolument convergente.

9.4 Intégrale des applications à valeurs complexes

Nous obtenons le même genre de résultat que dans la section précédente. Il suffit de raisonner en module. L'intégrabilité de f est alors assurée par celles des parties réelle et imaginaire de f .

9.5 Remarque générale

Nous venons de construire l'intégrale par rapport à une mesure positive μ . La plupart des démonstrations a été négligée car un certain nombre d'entre elles sont relativement difficiles et non obligatoires en première lecture (en ce qui nous concerne). Toutefois, il est utile de bien comprendre le cheminement souvent utilisé pour démontrer les différents résultats vérifiés par l'intégrale : on démontre la propriété pour une fonction positive étagée, ensuite on étend le résultat aux fonctions mesurables positives par Beppo-Lévi et on généralise aux fonctions mesurables de signe quelconque en écrivant : $f = f^+ - f^-$.

CHAPITRE 10

Propriétés de l'intégrale et théorèmes de convergence

10.1 Notion de négligeabilité

Définition 10.1.1 Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Un élément A de \mathcal{A} est dit μ -négligeable s'il existe $B \in \mathcal{A}$ tel que $A \subset B$ et $\mu(B) = 0$ (ce qui entraîne bien sûr $\mu(A) = 0$).

Définition 10.1.2 Soit Π une propriété définie sur Ω . On dit que Π est vraie μ -presque partout (ou presque partout — p.p.) si l'ensemble $\{\omega \in \Omega : \Pi(\omega) \text{ est fausse}\}$ est μ -négligeable.

Remarque 10.1.1 Si μ est une probabilité, alors on parle de propriété vraie *presque sûrement* (p.s.).

Définition 10.1.3 Une fonction borélienne f est μ -négligeable si l'ensemble :

$$\{\omega \in \Omega : f(\omega) \neq 0\} \quad (10.1)$$

est μ -négligeable.

Propriété 10.1.1 Toute réunion dénombrable d'ensembles de mesure nulle est de mesure nulle.

Démonstration 10.1.1 En TD. □

10.2 Propriétés de l'intégrale

Théorème 10.2.1 (admis)

Soit f une application borélienne à valeurs dans $\overline{\mathbb{R}^+}$. Alors :

$$\int f \, d\mu = 0 \iff f = 0 \quad \mu - \text{p.p.} \quad (10.2)$$

Proposition 10.2.1 Soit f une application mesurable et $A \in \mathcal{A}$. Si A est μ -négligeable alors :

$$\int_A f \, d\mu \triangleq \int f \mathbb{1}_A \, d\mu = 0 \quad (10.3)$$

Démonstration 10.2.1 On a $\{f \mathbb{1}_A \neq 0\} \subset A$ et $\mu(A) = 0$. Par suite, $f \mathbb{1}_A = 0$ μ -p.p. d'où $|f| \mathbb{1}_A = 0$ μ -p.p. Donc, d'après le théorème 10.2.1 :

$$\int |f| \mathbb{1}_A \, d\mu = 0 \Leftrightarrow \int f^+ \mathbb{1}_A \, d\mu = 0 \quad \text{et} \quad \int f^- \mathbb{1}_A \, d\mu = 0$$

Il vient alors :

$$\int f \mathbb{1}_A \, d\mu = \int f^+ \mathbb{1}_A \, d\mu - \int f^- \mathbb{1}_A \, d\mu = 0$$

Théorème 10.2.2 (admis)

Soient f et g des applications mesurables à valeurs dans $\overline{\mathbb{R}}$ égales μ -p.p.

- (i) Si f et g sont positives alors $\int f \, d\mu = \int g \, d\mu$.
- (ii) Si f est intégrable alors g est intégrable et $\int f \, d\mu = \int g \, d\mu$.

Théorème 10.2.3 (admis)

Toute application μ -intégrable à valeurs dans $\overline{\mathbb{R}}$ est finie μ -p.p.

Exemple 10.2.1 (fonctions égales p.p.)

Soit l'espace mesuré $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. On considère deux applications f et g définies sur \mathbb{R} à valeurs dans \mathbb{R} telles que $f(x) = x$ et $g(x) = x \mathbb{1}_{\mathbb{R} \setminus \mathbb{Q}}(x)$. Alors f et g sont égales λ -p.p.

10.3 Théorèmes de convergence

Il s'agit en particulier des théorèmes d'interversion de limites (plus précisément limite et intégrale). Nous avons déjà vu le théorème le plus important pour démontrer la plupart des résultats de la théorie de l'intégration : le théorème de Beppo-Lévi appelé encore *théorème de convergence monotone*. Nous énonçons pour information le lemme de Fatou (qui sert essentiellement à démontrer des propriétés). Quant au *théorème de convergence dominée* (ou encore théorème de Lebesgue), il s'agit de l'un des théorèmes les plus utilisés en pratique (avec celui de Beppo-Lévi).

En tout état de cause, il s'agit de résultats très utiles requérant des hypothèses faibles contrairement au cadre de l'intégrale de Riemann qui nécessite souvent la convergence uniforme pour pouvoir intervertir les limites.

Théorème 10.3.1 (lemme de Fatou)

Soit $(f_n)_{n \in \mathbb{N}^*}$ une suite d'applications mesurables à valeurs dans $\overline{\mathbb{R}}^+$. Alors :

$$\lim_{n \rightarrow +\infty} \inf \int f_n \, d\mu \geq \int \lim_{n \rightarrow +\infty} \inf f_n \, d\mu \quad (10.4)$$

Remarque 10.3.1

$$\lim_{n \rightarrow +\infty} \inf f_n = \sup_{n \in \mathbb{N}^*} \inf_{k \geq n} f_k$$

Théorème 10.3.2 (Théorème de convergence dominée ou de Lebesgue)

Soit g une application μ -intégrable et $(f_n)_{n \in \mathbb{N}^*}$ une suite d'applications mesurables telles que :

$$\forall n \in \mathbb{N}^* \quad |f_n| \leq g \quad \mu - p.p. \quad (10.5)$$

On suppose en outre que $(f_n)_{n \in \mathbb{N}^*}$ converge μ -p.p. vers une application f mesurable. Alors :

(i) f est μ -intégrable ;

(ii)

$$\lim_{n \rightarrow +\infty} \int |f_n - f| d\mu = 0 \quad (10.6)$$

ou encore :

$$\lim_{n \rightarrow +\infty} \int f_n d\mu = \int f d\mu \quad (10.7)$$

Remarque 10.3.2 Le théorème de Lebesgue s'étend sans difficulté aux applications complexes en passant par les modules.

10.4 Exemples d'application des théorèmes de convergence

Théorème 10.4.1 Soit $(f_n)_{n \in \mathbb{N}^*}$ une suite d'applications mesurables positives. Alors :

$$\int \sum_{n \in \mathbb{N}^*} f_n d\mu = \sum_{n \in \mathbb{N}^*} \int f_n d\mu \quad (10.8)$$

Démonstration 10.4.1 On utilise le théorème de Beppo-Lévi à la suite des sommes partielles des f_n . □

Théorème 10.4.2 (admis)

Si $(f_n)_{n \in \mathbb{N}^*}$ est une suite d'applications intégrables vérifiant :

$$\sum_{n \in \mathbb{N}^*} \int |f_n| d\mu < +\infty \quad (10.9)$$

Alors la série $\sum_{n \in \mathbb{N}^*} f_n$ converge μ -p.p., sa somme est intégrable et :

$$\int \sum_{n \in \mathbb{N}^*} f_n d\mu = \sum_{n \in \mathbb{N}^*} \int f_n d\mu \quad (10.10)$$

Théorème 10.4.3 (admis)

Soit $(f_n)_{n \in \mathbb{N}^*}$ une suite d'applications mesurables dont la série converge μ -p.p., on a l'inégalité (dite de convexité) suivante :

$$\int \left| \sum_{n \in \mathbb{N}^*} f_n \right| d\mu \leq \sum_{n \in \mathbb{N}^*} \int |f_n| d\mu \quad (10.11)$$

CHAPITRE 11

Intégrale de Lebesgue-intégrale de Riemann

Dans tout ce chapitre, on considère l'espace mesuré $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ où λ est la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On considère une application f borélienne.

11.1 Rappels sur l'intégrale de Riemann

11.1.1 Cas d'une fonction bornée

Soit un intervalle $[a, b]$, et π une partition de $[a, b]$. Il s'agit donc de l'ensemble des réels x_0, x_1, \dots, x_n tels que $a = x_0 < x_1 < \dots < x_n = b$.

Définition 11.1.1 On appelle *sommes de Darboux* de f relativement à la partition π , les deux sommes :

$$s_f(\pi) = \sum_{i=1}^n (x_i - x_{i-1}) m_i \quad \text{et} \quad S_f(\pi) = \sum_{i=1}^n (x_i - x_{i-1}) M_i \quad (11.1)$$

où $\forall i = 1, \dots, n : m_i = \inf_{x \in [x_{i-1}, x_i]} f(x)$ et $M_i = \sup_{x \in [x_{i-1}, x_i]} f(x)$

Définition 11.1.2 On dit que f est *Riemann-intégrable* sur $[a, b]$ si pour tout $\varepsilon > 0$, il existe une partition π de $[a, b]$ telle que :

$$S_f(\pi) - s_f(\pi) < \varepsilon \quad (11.2)$$

Théorème 11.1.1 Si f est Riemann-intégrable alors on a :

$$\inf_{\pi} S_f(\pi) = \sup_{\pi} s_f(\pi) = I \quad (11.3)$$

où π appartient à l'ensemble des partitions de $[a, b]$ et I est notée :

$$I = \int_a^b f(x) \, dx \quad (11.4)$$

Proposition 11.1.1 Les applications en escalier, continues, réglées (c'est-à-dire limites de suites uniformes convergentes d'applications en escalier) sont Riemann-intégrables sur $[a, b]$.

Proposition 11.1.2 L'application $x \mapsto \int_a^x f(t) dt$ est continue sur $[a, b]$. Si f est continue sur $[a, b]$, alors F est une primitive de f .

11.1.2 Intégrales impropres

Définition 11.1.3 Soit f une application définie sur $[a, b[$ avec $b \leq +\infty$. On dit que l'intégrale $\int_a^b f(x) dx$ est convergente si pour tout c tel que $c < b$, l'intégrale de Riemann $\int_a^c f(x) dx$ existe et si $\lim_{c \rightarrow b} \int_a^c f(x) dx$ existe. On l'appelle l'*intégrale impropre* de Riemann.

Remarque 11.1.1 On procède de la même façon lorsque l'intégrale est impropre au niveau de la borne inférieure.

Théorème 11.1.2 Si $\int_a^b |f(x)| dx$ converge alors $\int_a^b f(x) dx$ converge et on dit que f est absolument convergente.

11.2 Riemann-intégrabilité et Lebesgue-intégrabilité

Une application Lebesgue-intégrable n'est en général pas Riemann-intégrable. L'un des exemples les plus connus est celui de la fonction de Dirichlet $f(x) = \mathbb{1}_{\mathbb{Q} \cap [0, 1]}(x)$.

f est bien intégrable au sens de Lebesgue. En effet :

$$\int_{\mathbb{R}} \mathbb{1}_{\mathbb{Q} \cap [0, 1]}(x) d\lambda(x) = \lambda(\mathbb{Q} \cap [0, 1]) = 0 \quad (11.5)$$

Cependant f n'est pas Riemann-intégrable car pour toute partition π , on a $s_f(\pi) = 0$ et $S_f(\pi) = 1$ (prendre par exemple une partition de pas $1/10$).

Théorème 11.2.1 Si f est une application Riemann-intégrable sur $]a, b[$, elle est Lebesgue-intégrable sur $]a, b[$ et les deux intégrales coïncident :

$$\int_{]a, b[} f(x) d\lambda(x) = \int_a^b f(x) dx \quad (11.6)$$

Remarque 11.2.1 Les notations sont conventionnelles, toutefois nous veillerons à les respecter strictement. On remarquera également que :

$$\int_{]a, b[} f(x) d\lambda(x) = \int_{\mathbb{R}} f(x) \mathbb{1}_{]a, b[} d\lambda(x) \quad (11.7)$$

CHAPITRE 12

Compléments d'intégration

12.1 Mesure définie par une densité

Définition 12.1.1 Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $f \in \mathcal{M}^+$. Alors l'application $\nu : \mathcal{A} \longrightarrow \overline{\mathbb{R}}^+$ définie par :

$$\forall A \in \mathcal{A} \quad \nu(A) = \int_A f \, d\mu \quad (12.1)$$

est une nouvelle mesure sur (Ω, \mathcal{A}) , appelée mesure de densité f par rapport à μ .

Propriété 12.1.1

- Si f est intégrable, alors ν est finie.
- Si $\nu(\Omega) = 1$ alors ν est une mesure de probabilité. La mesure ν est dite à densité.
- Si f et g sont égales μ -p.p., ν est également une mesure de densité g par rapport à μ . f et g sont deux versions de la densité.

Notation

On note $\frac{d\nu}{d\mu}$ la classe d'équivalence des différentes versions de la densité de ν par rapport à μ . Cependant, en pratique, nous confondrons la classe d'équivalence et une version particulière de la densité.

On notera également $\nu = f \cdot \mu$. On dit que ν est le produit de l'application f par μ .

Remarque 12.1.1 Si f est μ -intégrable, on peut construire une probabilité sur l'espace mesurable (Ω, \mathcal{A}) :

$$\forall A \in \mathcal{A} \quad \mathbb{P}(A) = \frac{\nu(A)}{\nu(\Omega)} \quad (12.2)$$

Théorème 12.1.1 Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, ν une mesure de densité f par rapport à μ et g une application mesurable définie sur (Ω, \mathcal{A}) . Alors g est ν -intégrable si et seulement si gf est μ -intégrable et on a :

$$\int_{\Omega} g \, d\nu = \int_{\Omega} gf \, d\mu \quad (12.3)$$

12.2 Exemples de lois à densité

Nous présentons ci-après quelques lois à densité usuelles.

Définition 12.2.1 Soit X une v.a. définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$. La loi de X est dite à densité s'il existe une application borélienne positive f telle que la probabilité image \mathbb{P}_X s'écrive de la manière suivante :

$$\forall B \in \mathcal{B}(\mathbb{R}) \quad \mathbb{P}_X(B) = \int_B f \, d\lambda = \int_B f(x) \, d\lambda(x) \quad (12.4)$$

On a donc, par identification, $\nu = \mathbb{P}_X$ et $\mu = \lambda$ (mais les Ω et \mathcal{A} ne sont pas les mêmes).

12.2.1 Loi uniforme $\mathcal{U}_{[a,b]}$

Définition 12.2.2 Une v.a. X suit une loi uniforme $\mathcal{U}_{[a,b]}$ ($a < b$) si sa densité de probabilité s'écrit :

$$f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) \quad (12.5)$$

La loi uniforme (notamment $\mathcal{U}_{[0,1]}$) est utilisée dans tous les logiciels statistiques pour générer des valeurs de lois à densité.

12.2.2 Loi exponentielle $\mathcal{E}(\theta)$

Définition 12.2.3 Une v.a. X suit une loi exponentielle de paramètre $\theta > 0$ si sa densité s'écrit :

$$f(x) = \theta e^{-\theta x} \mathbb{1}_{\mathbb{R}^+}(x) \quad (12.6)$$

La loi exponentielle est très utilisée pour modéliser les durées de vie.

12.2.3 Loi normale $\mathcal{N}(\mu, \sigma^2)$

Définition 12.2.4 Une v.a. X suit une loi normale (ou de Gauss-Laplace ou de Gauss) $\mathcal{N}(\mu, \sigma^2)$ ($\mu \in \mathbb{R}, \sigma > 0$) si sa densité s'écrit :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \mathbb{1}_{\mathbb{R}}(x) \quad (12.7)$$

12.3 Théorème de Radon-Nikodym

Théorème 12.3.1 Soit ν une mesure de densité par rapport à μ . Alors :

$$\forall A \in \mathcal{A} \quad \mu(A) = 0 \implies \nu(A) = 0 \quad (12.8)$$

Démonstration 12.3.1 Il s'agit d'une application de la proposition [10.2.1](#) page [62](#). □

Définition 12.3.1 (Absolue continuité)

Soient μ et ν deux mesures sur (Ω, \mathcal{A}) . ν est dite *absolument continue* par rapport à μ (on dit également que μ domine ν et on note $\nu \ll \mu$) si :

$$\forall A \in \mathcal{A} \quad \mu(A) = 0 \implies \nu(A) = 0 \quad (12.9)$$

Proposition 12.3.1 Si ν est une mesure de densité par rapport à μ , alors elle lui est absolument continue.

Exemple 12.3.1 Soit une v.a. X de loi à densité. Alors, \mathbb{P}_X est absolument continue par rapport à la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Autrement dit, λ domine \mathbb{P}_X . Par abus de langage, lorsque la loi de X est à densité, on dit parfois que X est absolument continue.

Le théorème fondamental suivant est la réciproque (sous certaines conditions) du théorème 12.3.1.

Théorème 12.3.2 (Radon-Nikodym)

Soit μ une mesure σ -finie et ν une mesure sur (Ω, \mathcal{A}) absolument continue par rapport à μ . Alors il existe une application f positive, unique à une μ -équivalence près telle que :

$$\forall A \in \mathcal{A} \quad \nu(A) = \int_A f \, d\mu \quad (12.10)$$

Le résultat demeure si ν est σ -finie. Et si ν est finie alors f est μ -intégrable. f est appelée dérivée de Radon-Nikodym de ν par rapport à μ et sera notée :

$$f = \frac{d\nu}{d\mu} \quad \text{ou} \quad d\nu = f \, d\mu \quad (12.11)$$

Exemple 12.3.2 (Application aux lois absolument continues)

Soit une v.a. X de loi \mathbb{P}_X . Cette mesure de probabilité est finie. La mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est σ -finie. On sait alors que \mathbb{P}_X est absolument continue par rapport à λ . La densité de X est notée f . D'après le théorème de Radon-Nikodym, f existe bien et est unique à une λ -équivalence près. Elle est λ -intégrable. On a :

$$\forall B \in \mathcal{B}(\mathbb{R}) \quad \mathbb{P}_X(B) \triangleq \int_B d\mathbb{P}_X = \int_B f \, d\lambda \quad (12.12)$$

où $d\mathbb{P}_X = f \, d\lambda$.

Exemple 12.3.3 (Application aux lois discrètes)

Soit une v.a.d. X à valeurs dans $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. \mathbb{P}_X est finie, la mesure de comptage sur \mathbb{N} $\mu_c = \sum_{n \in \mathbb{N}} \delta_n$ est σ -finie, le théorème de Radon-Nikodym s'applique encore une fois. Il existe une densité μ_c -intégrable valant pour tout $k \in \mathbb{N}$:

$$\frac{d\mathbb{P}_X}{d\mu_c}(\{k\}) = \mathbb{P}(X = k) \quad (12.13)$$

On retrouve bien formellement :

$$\forall k \in \mathbb{N} \quad \mathbb{P}_X(\{k\}) = \mathbb{P}(X = k) \sum_{n \in \mathbb{N}} \delta_n(\{k\}) = \mathbb{P}(X = k) \quad (12.14)$$

Définition 12.3.2 On dit que deux mesures ν et μ définies sur (Ω, \mathcal{A}) sont étrangères (ou ν est étrangère à μ) s'il existe $N \in \mathcal{A}$ tel que $\mu(N) = 0$ et $\nu(\overline{N}) = 0$.

Exemple 12.3.4 La mesure de Lebesgue λ et la mesure de comptage μ_c définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ou encore sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ sont étrangères.

Théorème 12.3.3 (Décomposition de Lebesgue)

Soient μ et ν deux mesures σ -finies sur un espace mesurable (Ω, \mathcal{A}) . Il existe alors une application mesurable f , unique à une μ -équivalence près et une mesure γ définie sur le même espace mesurable, unique et étrangère à μ telles que :

$$\nu = f \cdot \mu + \gamma \quad \text{ou} \quad \forall A \in \mathcal{A} \quad \nu(A) = \int_A f d\mu + \gamma(A) \quad (12.15)$$

12.4 Théorème de transfert

Nous avons vu qu'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ est souvent abstrait. En général, on ne peut pas expliciter Ω ni \mathcal{A} , même si on sait que ces objets existent. Il est alors plus pratique d'effectuer les calculs sur l'espace image via une v.a. X . Le théorème de transfert justifie ce passage de l'espace abstrait à l'espace image.

Nous rappelons la définition d'une mesure image.

Définition 12.4.1 Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, (E, \mathcal{E}) un espace mesurable et f une application mesurable de (Ω, \mathcal{A}) dans (E, \mathcal{E}) . On appelle *mesure image* de μ par f , la mesure ν sur (E, \mathcal{E}) définie par :

$$\forall B \in \mathcal{E} \quad \nu(B) = \mu(f^{-1}(B)) \quad (12.16)$$

Théorème 12.4.1 (Théorème de transfert)

On reprend les notations précédentes. Soit h une application mesurable définie sur $(E, \mathcal{B}(\mathbb{R}))$. Alors h est ν -intégrable si et seulement si $h \circ f$ est μ -intégrable et en outre :

$$\int_E h d\nu = \int_\Omega h \circ f d\mu \quad (12.17)$$

Corollaire 12.4.1 Soit X une v.a. définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mu)$ et à valeurs dans un espace mesurable (E, \mathcal{E}) et de loi \mathbb{P}_X . Soit h une application mesurable de (E, \mathcal{E}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors l'espérance mathématique de $h \circ X$ existe si et seulement si h est \mathbb{P}_X intégrable et on a :

$$\mathbb{E}(h \circ X) = \mathbb{E}h(X) = \int_E h d\mathbb{P}_X \quad (12.18)$$

En particulier, si $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on a :

$$\mathbb{E}h(X) = \int_{\mathbb{R}} h(x) d\mathbb{P}_X(x) \quad (12.19)$$

Application aux lois discrètes

Si $\mathbb{P}_X \ll \mu_c$ sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. $\mathbb{E}h(X)$ existe si et seulement si $\sum_{n \in \mathbb{N}} p_n |h(n)| < +\infty$ et :

$$\mathbb{E}h(X) = \sum_{n \in \mathbb{N}} h(n) p_n \quad (12.20)$$

Application aux lois continues

Si $\mathbb{P}_X \ll \lambda$ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La loi de X est à densité f . $\mathbb{E}h(X)$ existe si et seulement si $h \cdot f$ est λ -intégrable et on a :

$$\mathbb{E}h(X) = \int_{\mathbb{R}} h(x) f(x) d\lambda(x) \quad (12.21)$$

12.5 Intégrale par rapport à une mesure produit

Théorème 12.5.1 Soient n espaces mesurés $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, \dots, n$ avec les μ_i σ -finies. Soit $\Omega = \Omega_1 \times \dots \times \Omega_n$ et $\mathcal{A} = \bigotimes_{i=1}^n \mathcal{A}_i$. Alors, il existe une unique mesure μ appelée *mesure produit* sur l'espace mesurable (Ω, \mathcal{A}) telle que :

$$\forall A_i \in \mathcal{A}_i \quad \mu(A_1 \times \dots \times A_n) = \prod_{i=1}^n \mu(A_i) \quad (12.22)$$

et on note $\mu = \bigotimes_{i=1}^n \mu_i$.

Théorème 12.5.2 Soit $n = 2$. La mesure produit μ est définie par :

$$\forall A \in \mathcal{A}_1 \otimes \mathcal{A}_2 \quad \mu(A) = \int_{\Omega_2} \mu_1(A_{\omega_2}) d\mu(\omega_2) = \int_{\Omega_1} \mu_2(A_{\omega_1}) d\mu(\omega_1) \quad (12.23)$$

avec

$$A_{\omega_1} = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\}$$

appelée section de A en ω_1 et \mathcal{A}_2 -mesurable et

$$A_{\omega_2} = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\}$$

appelée section de A en ω_2 et \mathcal{A}_1 -mesurable.

Théorème 12.5.3 (Théorème de Fubini)

Soit deux espaces mesurés $(\Omega_1, \mathcal{A}_1, \mu_1)$ et $(\Omega_2, \mathcal{A}_2, \mu_2)$, μ_1 et μ_2 étant σ -finies. Soit $\mu = \mu_1 \otimes \mu_2$ sur $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$. Soit f une application mesurable positive ou μ -intégrable. Alors :

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f d\mu &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2) \end{aligned} \quad (12.24)$$

En pratique, on se trouvera très souvent dans la situation où $\Omega_1 = \Omega_2 = \mathbb{R}$, $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{B}(\mathbb{R})$ et $\mu_1 = \mu_2 = \lambda$. Pour montrer la λ_2 -intégrabilité de f ($\lambda_2 = \lambda \otimes \lambda$), il suffit de montrer que l'une des deux intégrales suivantes est finie :

$$\int_{\Omega_1} \int_{\Omega_2} |f| \, d\mu_2 \, d\mu_1$$

ou

$$\int_{\Omega_2} \int_{\Omega_1} |f| \, d\mu_1 \, d\mu_2$$

Bibliographie

- [1] BILLINGSLEY P., *Probability and Measure*, Wiley, 1995.
- [2] BRÉMEAUD Pierre, *Introduction aux probabilités*, Springer, 1997.
- [3] MONFORT A., *Cours de probabilités*, Economica, 1996.
- [4] NEVEU J., *Bases mathématiques du calcul des probabilités*, Masson et Cie, 1970.
- [5] TASSI P., LEGAIT S., *Théorie des probabilités en vue des applications statistiques*, Technip, 1990.