



الكلية متعددة التخصصات الناظور

ⵜⴰⵎⴻⵏⴰⵏⵜ ⵜⴰⵎⴻⵏⴰⵏⵜ ⵜⴰⵎⴻⵏⴰⵏⵜ | II.E:Q
Faculté Pluridisciplinaire de Nador

Département : Mathématiques

Filière : SMA - S6

Module : M33

Année universitaire : 2021/2022

Probabilités

Toufik Chaayra

t.chaayra@edu.umi.ac.ma

Table des matières

1	Rappels sur les probabilités élémentaires	5
1.1	Définitions	5
1.1.1	Probabilité sur un ensemble fini	6
1.2	Probabilités conditionnelles et indépendance	7
1.3	Probabilités conditionnelles	7
1.4	Événements indépendants	8
1.5	Formule de Bayes	9
1.5.1	Formule des probabilités totales	9
1.5.2	Formule de Bayes	9
2	Théorie de la mesure	11
2.1	Tribu sur un ensemble	11
2.2	Mesures et probabilités	12
3	Variables aléatoires	17
3.1	Définitions	17
3.2	Variable aléatoire	17
3.3	Fonctions de répartition	19
3.3.1	Fonctions de répartition d'une variable aléatoire réelle	19
3.4	Espérance mathématique	24
3.4.1	Rappels d'intégration	24
3.4.2	Espérance d'une variable aléatoire	25
3.5	Changement de variables	26
3.6	Fonctions caractéristiques	27
4	Vecteurs aléatoires	30
4.1	Loi d'un vecteur aléatoire	30
4.1.1	Loi marginale	31
4.2	Fonctions de répartition d'un vecteur aléatoire	34
4.3	Changement de variables	35
4.4	Fonctions caractéristiques	38
5	Indépendance	40
5.1	Définitions et critères d'indépendance	40
5.1.1	Indépendance de variables aléatoires et corrélation.	41
5.1.2	Indépendance, fonctions caractéristiques et de répartitions.	41
5.1.3	Corrélation et vecteur gaussien.	42
5.2	Loi conditionnelle	42

5.3	Noyaux et lois conditionnelles	43
5.4	Somme de variables aléatoires indépendantes.	46
6	Espérance conditionnelle	50
6.1	Introduction : variables aléatoires discrètes	50
6.2	Théorème d'existence et d'unicité	51
6.3	Variables à densité	52
6.4	Variables discrètes	53
6.5	Variables quelconques	54
7	Convergence de suites de variables aléatoires	55
7.1	Convergence presque sûre	55
7.2	Convergence en probabilité	58
7.3	Convergence dans L^p	60
7.4	Convergence en loi	60

Chapitre 1

Rappels sur les probabilités élémentaires

1.1 Définitions

Exemple 1.1.1 On lance une pièce de monnaie $n = 30$ fois et on note le résultat de chaque lancée. On considère l'événement $E = \{P\}$: le coté Pile apparait. Ensuite on calcul la fréquence relative de l'événement, c'est-à-dire : $f(E) = \frac{k}{n}$ ou k est le nombre de fois que le coté pile apparait et $n = 1, \dots, 30$ le nombre de lancée. Les résultats sont notés sur le tableau suivant :

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	F	P	P	P	F	P	P	F	F	P	P	F	F	F	P
$f(E)$	0/1	1/2	2/3	3/4	3/5	4/6	5/7	5/8	5/9	6/10	7/11	7/12	7/13	7/14	8/15
n	16	17	18	19	20	21	22	23	24	25	26	27			
	F	P	F	P	F	F	F	P	F	F	F	P			
$f(A)$	8/16	9/17	9/18	10/19	10/20	10/21	10/22	11/23	11/24	11/25	11/26	12/27			
n	28	29	30												
	F	F	P												
$f(A)$	12/28	12/29	13/30												

L'expérience montre qu'au fur et à mesure que n grandit, la fréquence relative $f(E)$ tend à se stabiliser autour de 0.5 .

Définition 1.1.1 On suppose qu'une expérience d'ensemble fondamental Ω est exécuté plusieurs fois sous les mêmes conditions. Pour chaque événement E de Ω on définit $n(E)$ comme le nombre de fois où l'événement E survient lors des n premières répétitions de l'expérience. La probabilité de l'événement E est définie par :

$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}. \quad (1.1)$$

$P(E)$ est définie comme la limite du pourcentage du nombre de fois où E survient par rapport au nombre total de répétition. On dit que la probabilité de l'événement E est tout simplement la valeur limite vers laquelle va tendre $f(E)$ quand n devient de plus en plus grand. C'est ce qu'on appelle probabilité empirique de l'événement E .

Blaise Pascal (né le 19 juin 1623 à Clermont-Ferrand) est l'un des premiers mathématiciens à définir le concept de probabilité de la manière suivante :

Définition 1.1.2 La probabilité d'un événement E , est

$$P(E) = \frac{\text{nombre de cas favorable à l'événement } E}{\text{nombre de cas possibles équivraisemblables dans } \Omega}$$

où équivraisemblables veut dire ayant des chances égales de se réaliser.

On peut vérifier que les deux Définitions 1.1.1 et 1.1.1 vérifient la définition générale d'une probabilité qu'on présentera au chapitre 2.

1.1.1 Probabilité sur un ensemble fini

On suppose que l'espace fondamental est fini

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}.$$

On peut construire une probabilité P en se donnant des nombres $p_i = P(\omega_i)$ tels que

$$0 \leq p_i \leq 1, \quad \forall i \in \{1, 2, \dots, n\}$$

et

$$p_1 + p_2 + \dots + p_n = 1.$$

Cas d'équiprobabilité ($p_1 = p_2 = \dots = p_n = 1/n$). Si tous les événements élémentaires ont la même probabilité, on dit qu'il y a équiprobabilité. Cette probabilité vaut alors $1/n$ et dans ce cas la probabilité d'un événement A contenant k événements élémentaires vaut k/n . Plus généralement, on écrit :

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}.$$

On retrouve alors la définition d'une probabilité comme étant le quotient :

$$P(A) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Exemple 1.1.2 On lance un dé. Soit $\Omega = \{1, 2, \dots, 6\}$.

• (dé équilibré) $P(\{i\}) = 1/6$. Dans ce cas,

$$P(\{1, 2, 3\}) = \frac{1}{2} = \frac{\text{card}(\{1, 2, 3\})}{\text{card}(\{1, 2, \dots, 6\})}.$$

• (dé pipé) $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = 1/7$ et $P(\{6\}) = 2/7$. Dans ce cas

$$P(\{1, 2, 6\}) = \frac{4}{7} \neq \frac{\text{card}(\{1, 2, 3\})}{\text{card}(\{1, 2, \dots, 6\})}.$$

Exemple 1.1.3 Un sac contenant 5 jetons dont 3 noirs et 2 blancs.

1.) On y tire au hasard un jeton. Quelle est la probabilité que le jeton soit de couleur noir ?

2.) Supposons que l'on tire maintenant deux jetons en même temps (= à la fois = simultanément), quelle est la probabilité d'obtenir 2 jetons noirs ?

Réponse : 1.) L'ensemble fondamental Ω contient au total 5 jetons. La probabilité de l'événement "E : obtenir un jeton noir" est donnée par

$$P(E) = \frac{\text{card}(E)}{\text{card}(\Omega)} = \frac{C_3^1}{C_5^1} = 3/5.$$

2.) L'ensemble fondamental Ω contient au total $C_5^2 = 10$ couple de deux jetons. Soit l'événement "E : obtenir 2 jetons noirs". Or il y a $C_3^2 \times C_2^0$ façons de tirer 2 jetons noirs et 0 blanc. Donc

$$P(E) = \frac{\text{card}(E)}{\text{card}(\Omega)} = \frac{C_3^2 \times C_2^0}{C_5^2} = \frac{3}{10}$$

Exemple 1.1.4 Un comité de 5 personnes doit être choisi parmi les 6 hommes et 9 femmes d'un groupe. Si le choix est le résultat du hasard, quelle est la probabilité que le comité soit composé de 3 hommes et 2 femmes ?

Réponse : L'ensemble fondamental Ω contient au total C_{15}^5 combinaisons possibles qui ont les mêmes chances d'apparaître. Soit l'événement "E : le comité est composé de 3 hommes et 2 femmes". Il y a C_6^3 façons de choisir 3 hommes parmi 6 et C_9^2 façons de choisir 2 femmes parmi 9. La probabilité cherchée est $P(E) = \frac{C_6^3 C_9^2}{C_{15}^5} = \frac{240}{1001}$.

1.2 Probabilités conditionnelles et indépendance

1.3 Probabilités conditionnelles

Exemple 1.3.1 Dans un groupe de 30 élèves, il y a 18 filles et 12 garçons ; parmi les filles, il y en a 6 qui pratiquent le jogging régulièrement, alors que chez les garçons on en compte 5. Si une personne est choisie au hasard dans ce groupe :

- Quelle est la probabilité que l'individu choisi pratique le jogging ?
- Quelle est la probabilité que l'individu choisi pratique le jogging sachant que c'est une fille ?

Notons E l'événement "la personne choisie pratique le jogging" et F l'événement "la personne choisie est une fille".

a. Il est clair que $P(E) = (6 + 5)/30 = 11/30$.

b. il est clair que $P(E \text{ sachant } F) = 6/18 = 1/3$.

pour répondre à la deuxième question, il suffit de chercher les cas favorables à l'événement E (il y en a 6) en se restreignant aux 18 filles, c'est à dire en tenant compte de la réalisation de E.

C'est ce qu'on appelle probabilité conditionnelle de E par rapport à F.

Définition 1.3.1 Soient E et F deux événements dans le cadre d'une expérience aléatoire ; alors la probabilité de E par rapport à F (ou sachant F), notée $P(E|F)$ est donnée par :

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \tag{1.2}$$

où $P(F) \neq 0$.

Exemple 1.3.2 Supposons que nous jetons deux dés. Quelle est la probabilité que la somme des deux dés donne 8 sachant que le premier dé donne 3.

Si nous désignons par :

E = événement somme des deux dés est 8.

F = événement le premier dé donne 3.

$E \cap F = \{(3, 5)\}$, $P(E \cap F) = 1/36$, $P(F) = 1/6$, donc la probabilité cherché est $P(E|F) = P(E \cap F)/P(F) = 1/6$.

Théorème 1.3.1 si E_1, E_2, \dots, E_n sont des des événements attachés à une même expérience, alors :

$$\begin{aligned} P(E_1 \cap E_2 \cap \dots \cap E_n) &= & (1.3) \\ &= P(E_1)P(E_2|E_1)P(E_3|(E_1 \cap E_2))P(E_4|(E_1 \cap E_2 \cap E_3))\dots P(E_n|((E_1 \cap E_2 \cap \dots \cap E_{n-1})) \end{aligned}$$

1.4 Événements indépendants

D'après la section précédente, on sait que la probabilité de E sachant que F est réalisé n'est en général pas égale à la probabilité non conditionnelle de E : $P(E)$. En d'autres termes, le fait de savoir que F est réalisé influence la probabilité de de F . Dans le cas où $P(E|F)$ est bien égal à $P(E)$, l'événement E est dit indépendant de F . Plus précisément, E est indépendant de F si le fait de savoir que F est réalisé ne change pas la probabilité de E . On a donc $P(E \cap F) = P(E|F)P(F)$ si E et F sont indépendants alors, $P(E \cap F) = P(E)P(F)$. Ainsi on a la définition suivante :

Définition 1.4.1 Deux événements E et F sont dits indépendants si

$$P(E \cap F) = P(E)P(F).$$

Deux événements sont dépendants s'ils ne sont pas indépendant.

Exemple 1.4.1 On jette deux pièces et on suppose que les 4 résultats possibles sont équiprobables.

On désigne par E «la première pièce montre pile» et par F «la seconde pièce montre face».

$E = \{(P, F), (P, P)\}$, $F = \{(P, F), (F, F)\}$, il est claire que $P(E) = 2/4 = 1/2$ et $P(F) = 1/2$. On a $E \cap F = \{(P, F)\}$ et $P(E \cap F) = 1/4 = P(E)P(F)$, donc E et F sont indépendants.

Exemple 1.4.2 On jette deux dés équilibrés. E_1 est l'événement «la somme des dés est 6» et F l'événement «le premier dé donne 4». Dans ce cas : $E_1 = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$, $P(E_1) = 5/36$ et $P(F) = 1/6$ et $P(E_1 \cap F) = P(\{(4, 2)\}) = 1/36$, il est claire que $P(E_1 \cap F) \neq P(E_1)P(F)$ et donc E et F ne sont pas indépendants.

La probabilité d'obtenir 6 sur deux dés dépend clairement du résultat apparu sur le premier dé. E_1 et F ne peuvent donc être indépendants.

Si par contre E_2 est l'événement «somme des dés est 7», E_2 est il indépendant de F ?

Dans ce cas, $E_2 = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, $P(E_2) = 6/36 = 1/6$, $P(F) = 1/6$. $E_2 \cap F = \{(4, 3)\}$, $P(E_2 \cap F) = 1/36 = P(E_2)P(F)$, donc E_2 et F sont indépendants.

Définition 1.4.2 Trois événements E, F et G sont dits totalement indépendants si

$$P(E \cap F \cap G) = P(E)P(F)P(G)$$

$$P(E \cap F) = P(E)P(F)$$

$$P(E \cap G) = P(E)P(G)$$

$$P(F \cap G) = P(F)P(G)$$

Plus généralement on a la définition suivante,

Définition 1.4.3 *Un ensemble d'événements E_1, E_2, \dots, E_n est dit totalement indépendant si pour tout sous ensemble $E_{k_1}, E_{k_2}, \dots, E_{k_m}$, $m \leq n$ on a :*

$$P(E_{k_1} \cap E_{k_2} \cap \dots \cap E_{k_m}) = P(E_{k_1})P(E_{k_2}) \dots P(E_{k_m}).$$

Conséquence : Si E, F et G sont totalement indépendants, E sera indépendant de tout événement formé à partir de F et G . Montrons par exemple que E et $F \cup G$ sont indépendants. En effets,

$$\begin{aligned} P(E \cap (F \cup G)) &= P((E \cap F) \cup (E \cap G)) \\ &= P(E \cap F) + P(E \cap G) - P(E \cap F \cap G) \\ &= P(E)P(F) + P(E)P(G) - P(E)P(F \cap G) \\ &= P(E) \underbrace{(P(F) + P(G) - P(F \cap G))}_{P(F \cup G)} \\ &= P(E)P(F \cup G). \end{aligned} \tag{1.4}$$

1.5 Formule de Bayes

1.5.1 Formule des probabilités totales

Propriétés 1.5.1 (Formule des probabilités totales) *Soient E et F deux événements quelconques. On a*

$$P(E) = P(E|F)P(F) + P(E|\bar{F})P(\bar{F}). \tag{1.5}$$

Remarque 1.5.1 *Cette formule est extrêmement utile puisqu'il nous permet de déterminer la probabilité d'un événement en commençant par le conditionner selon la réalisation ou non d'un autre événement. Autrement, il existe de nombreuses situations où il est difficile de calculer directement la probabilité d'un événement mais il est possible de la calculer connaissant ses probabilités conditionnelles si certains événements sont réalisés.*

1.5.2 Formule de Bayes

En fait, la formule précédente n'est rien d'autre que la formule de Bayes appliquée au cas de E, F_1 et F_2 :

$$\begin{aligned} P(F_1|E) &= \frac{P(E \cap F_1)}{P(E)} \\ &= \frac{P(F_1)P(E|F_1)}{P(F_1)P(E|F_1) + P(F_2)P(E|F_2)} \end{aligned}$$

la formule s'obtient à partir de la formule des probabilités conditionnelles $P(F_1|E) = \frac{P(E \cap F_1)}{P(E)}$ et de la formule de la probabilité totale $P(E) = P(F_1)P(E|F_1) + P(F_2)P(E|F_2)$.

On va utiliser le même principe pour généraliser cette formule, on a la proposition suivante :

Proposition 1.5.1 (Formule de Bayes) Supposons que les événements E_1, E_2, \dots, E_n forment une partition de l'ensemble fondamental Ω (c.à.d $\Omega = \cup_{i=1}^n E_i, E_i \cap E_j = \emptyset$); alors si on considère un événement quelconque E , on a :

$$P(E) = \sum_{i=1}^n P(E_i \cap E) = \sum_{i=1}^n P(E_i)P(E|E_i) \quad (1.6)$$

$$P(E_k|E) = \frac{P(E_k)P(E|E_k)}{\sum_{i=1}^n P(E_i)P(E|E_i)} \quad \text{où } k = 1, 2, \dots, n \quad (1.7)$$

Preuve : Comme E_1, E_2, \dots, E_n est une partition de Ω , les événements $E_1 \cap E, E_2 \cap E, \dots, E_n \cap E$ forment une partition de E c'est à dire :

$$E = (E_1 \cap E) \cup (E_2 \cap E) \cup \dots \cup (E_n \cap E)$$

et $(E_i \cap E) \cap (E_j \cap E) = \emptyset$ si $i \neq j$ (1.8)

D'après l'axiome 3 et la règle de multiplication on a :

$$\begin{aligned} P(E) &= P[(E_1 \cap E) \cup (E_2 \cap E) \cup \dots \cup (E_n \cap E)] \\ &= P(E_1 \cap E) + P(E_2 \cap E) + \dots + P(E_n \cap E) \\ &= P(E_1)P(E|E_1) + P(E_2)P(E|E_2) + \dots + P(E_n)P(E|E_n) \\ &= \sum_{i=1}^n P(E_i)P(E|E_i) \end{aligned} \quad (1.9)$$

Pour la deuxième formule :

$$P(E) = \sum_{i=1}^n P(E_i \cap E) = \sum_{i=1}^n P(E_i)P(E|E_i) \quad (1.10)$$

$$\begin{aligned} P(E_k|E) &= \frac{P(E_k \cap E)}{P(E)} \\ &= \frac{P(E_k \cap E)}{\sum_{i=1}^n P(E_i)P(E|E_i)} \end{aligned}$$

ou on a utilisé la règle de multiplication des probabilités et l'équation (1.6).

Chapitre 2

Théorie de la mesure

L'objet de ce chapitre est de rappeler les éléments de théorie de la mesure qui seront indispensables au développement du calcul des probabilités dans les chapitres suivants. Une mesure abstraite sur un ensemble Ω généralise la notion de longueur, d'aire ou de volume, sur la droite, le plan ou l'espace. Intuitivement, le lien avec les probabilités est qu'une probabilité mesure la vraisemblance d'un événement. Sur la droite (ou le plan, ou l'espace), la longueur (ou l'aire, ou le volume) est une fonction qui à un ensemble associe un nombre réel positif. Cette fonction est additive, au sens où appliquée à $A \cup B$, elle est la somme de la fonction appliquée en A et de la fonction appliquée en B , pourvu que A et B soient disjoints. On demandera à une mesure abstraite de vérifier cette additivité. Un fait peu intuitif est qu'il existe des sous-ensembles de la droite (ou du plan, ou de l'espace) pour lesquels on ne peut pas définir leur longueur (ou aire, ou volume). Il convient donc, dans un premier temps, de définir la classe d'ensembles que l'on veut (et peut) mesurer. Compte tenu de la propriété d'additivité décrite au paragraphe précédent, on imposera par exemple que cette classe soit stable par réunion finie.

Soit Ω un ensemble.

Exemple 2.0.1 (i) Ω pourra être \mathbb{R} ou \mathbb{R}^d , un espace métrique, ou plus généralement topologique.

(ii) On joue au dé en le lançant une fois. L'ensemble Ω peut être pris comme l'ensemble des faces du dé, $\Omega = \{1, 2, 3, 4, 5, 6\}$. Lorsque l'on lance le dé au hasard, cela revient à choisir au hasard un élément de Ω .

Remarque 2.0.1 Il convient de remarquer que l'on peut toujours ajouter des points à Ω . Dans l'exemple 2.0.1 (ii) nous pourrions tout aussi bien prendre $\Omega = \{1, 2, 3, 4, 5, 6, 7\}$. Mais intuitivement, 7 a une probabilité nulle d'être réalisé.

On considère $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω . Un sous-ensemble C de $\mathcal{P}(\Omega)$ est un ensemble de parties de Ω .

Le formalisme de la théorie des probabilités utilise les outils de la théorie de la mesure en adoptant un vocabulaire spécifique aux probabilités.

2.1 Tribu sur un ensemble

Définition 2.1.1 Une tribu ou σ -algèbre \mathcal{F} sur Ω est un sous-ensemble de $\mathcal{P}(\Omega)$ contenant Ω et vérifiant les propriétés suivantes :

$$- A \in \mathcal{F} \implies A^c.$$

— $A_1, \dots, A_n, \dots \in \mathcal{F} \implies (A_1 \cap \dots \cap A_n \cap \dots) \in \mathcal{F}$

Définition 2.1.2 Soit A une partie quelconque de $P(\Omega)$. On appelle tribu engendrée par A et on note $\sigma(A)$ l'intersection de toutes les tribus contenant A .

Définition 2.1.3 On appelle tribu des Boréliens de \mathbb{R} et on note $\mathcal{B}(\mathbb{R})$, la tribu engendrée par les intervalles ouverts $]a, b[$, $a > b \in \mathbb{R}$.

Définition 2.1.4 La tribu borélienne ou tribu de Borel de \mathbb{R}^d , notée $\mathcal{B}(\mathbb{R}^d)$, est la plus petite des tribus sur \mathbb{R}^d contenant tous les pavés de \mathbb{R}^d i.e. les parties de la forme $]a_1, b_1[\times]a_2, b_2[\times \dots \times]a_d, b_d[$.

Définition 2.1.5 Les éléments des tribus $\mathcal{B}(\mathbb{R})$, resp. $\mathcal{B}(\mathbb{R}^d)$, sont appelés boréliens de \mathbb{R} , resp. \mathbb{R}^d .

Définition 2.1.6 Le couple (Ω, \mathcal{A}) s'appelle un espace mesurable et les éléments de \mathcal{A} sont appelés les parties mesurables de Ω relativement à la tribu \mathcal{A} ou parties \mathcal{A} -mesurables de Ω

Définition 2.1.7 Soit (Ω, \mathcal{F}) et (E, \mathcal{E}) deux espaces mesurables. Une application $f : \Omega \rightarrow E$ est dite $(\mathcal{F}, \mathcal{E})$ -mesurable, si $f^{-1}(A) \in \mathcal{F}$ pour tout $A \in \mathcal{E}$, où

$$f^{-1}(A) = \{\omega \in \Omega \mid f(\omega) \in A\}.$$

Dans la suite du cours les ensembles \mathbb{R} et \mathbb{R}^d seront toujours supposés munis de leurs tribus boréliennes.

2.2 Mesures et probabilités

Définition 2.2.1 Une mesure positive sur (Ω, \mathcal{F}) est une application m de \mathcal{F} dans $\mathbb{R}^+ \times \{+\infty\}$ telle que

- $m[\emptyset] = 0$.
- Si A_1, \dots, A_n, \dots sont des parties de \mathcal{F} deux à deux disjointes, alors

$$m\left(\bigcup_{n=0}^{n=\infty} A_n\right) = \sum_{n=0}^{n=\infty} m(A_n).$$

Exemple 2.2.1 - Mesure de Dirac sur $(\Omega; \mathcal{P}(\Omega))$: soit $a \in \Omega$,

$$\delta_a(A) = 1_A(a) = \begin{cases} 1, & \text{si } a \in A; \\ 0, & \text{sinon} \end{cases}$$

- Mesure de Lebesgue sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$: c'est la mesure qui généralise la notion de longueur des intervalles

$$\lambda(]a_1, b_1[\times]a_2, b_2[\times \dots \times]a_d, b_d[) = (b_1 - a_1) \times (b_2 - a_2) \times \dots \times (b_d - a_d)$$

Définition 2.2.2 Une probabilité sur (Ω, \mathcal{F}) est une application P de \mathcal{F} dans $[0, 1]$ telle que

- $P[\Omega] = 1$

— Si $A_1 \dots A_n \dots$ sont des parties de \mathcal{F} deux à deux disjointes, alors

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

Exemple 2.2.2 1- La probabilité de Dirac au point $a \in \mathbb{R}$, δ_a .

2- La probabilité de Bernoulli de paramètre $0 < p < 1$:

$$\mathcal{B}(1, p) := p\delta_1 + (1 - p)\delta_0.$$

3- La probabilité binomiale de paramètre $0 < p < 1$ et n :

$$\mathcal{B}(n, p) := \sum_{k=0}^n C_n^k p^k (1 - p)^{n-k} \delta_k.$$

4- La probabilité de Poisson de paramètre λ :

$$\mathcal{P}(\lambda) := \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k.$$

5- La probabilité géométrique de paramètre $0 < p < 1$:

$$\mathcal{G}(p) := \sum_{k=1}^{\infty} p(1 - p)^{k-1} \delta_k.$$

6- La probabilité uniforme-discrète de paramètre n ou équiprobabilité sur $\{1, 2, \dots, n\}$

$$\mathcal{U}(n) := \frac{1}{n} \sum_{k=1}^n \delta_k.$$

7- La probabilité normale standard :

$$\mathcal{N}(0, 1)(A) := \frac{1}{\sqrt{2\pi}} \int_A e^{-\frac{x^2}{2}} dx$$

8- La probabilité uniforme-continue sur $[a, b]$

$$\mathcal{U}([a, b])(A) := \int_A \frac{1}{b - a} 1_{[a, b]}(x) dx$$

9- La probabilité exponentielle de paramètre $\theta > 0$,

$$\mathcal{E}(\theta)(A) := \int_A \theta e^{-\theta x} 1_{]0, +\infty[} dx$$

Exercice 2.2.1 Vérifier que les probabilités introduites dans l'exemple précédent sont bien des probabilités.

Exercice 2.2.2 1) Expliciter les expressions analytiques, pour tout $i \in \mathbb{N}$, de $\mathcal{B}(n, p)(\{i\})$ et $P(\lambda)(\{i\})$

2) Expliciter et calculer $P(\frac{1}{10})(\{1, 3, 5\})$, $\mathcal{B}(7, \frac{3}{10})(\{0, 3, 5\})$ et $\mathcal{N}(0, 1)(]0, \infty[)$.

Définition 2.2.3 Un espace mesurable (Ω, \mathcal{F}) muni d'une probabilité P sécrit (Ω, \mathcal{F}, P) et on parle alors d'espace de probabilité ou déspace probabilisé.

Propriétés 2.2.1 On a les propriétés suivantes.

(i) $P(\bar{A}) = 1 - P(A)$.

(ii) $P(\emptyset) = 0$.

(iii) Si A et B sont deux parties disjointes de Ω , on a

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

(iv) Si $A \subset B$ alors $P(A) \leq P(B)$.

(v) $0 \leq P(A) \leq 1$.

(vi) $P(A \cap \bar{B}) = P(A) - P(A \cap B)$.

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

Preuve. (i) on a $\Omega = A \cup \bar{A}$, comme $A \cap \bar{A} = \emptyset$ alors $1 = P(A \cup \bar{A}) = P(A) + P(\bar{A})$, ce qu'il fallait démontrait.

(ii) Prenons $A = \Omega$ et $B = \emptyset$. On a donc $A \cup B = \Omega$ et ainsi

$$P(\Omega) = 1, \quad P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = 1$$

ce qui implique $P(\emptyset) = 0$.

(iii) Il suffit de remarquer que $\{A, B\}$ est une sous famille de \mathcal{F} et d'appliquer la définition.

(iv) Si $A \subset B$ alors $B = A \cup C_B^A$ donc

$$P(B) = P(A) + P(C_B^A)$$

ce qui implique $P(A) \leq P(B)$ car $P(C_B^A) \geq 0$.

(v) On a $A \subset \Omega$ donc d'après (iv), il vient $0 \leq P(A) \leq P(\Omega) = 1$.

(vi) On a $A = (A \cap \bar{B}) \cup (A \cap B)$ donc d'après (iii) il vient

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

ainsi

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

(vii) On a

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

donc

$$P(A \cup B) = P(A \cap \bar{B}) + P(\bar{A} \cap B) + P(A \cap B),$$

donc d'après (vi)

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

Ainsi

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

Attention, cette formule ne se généralise pas immédiatement pour plus de deux événements, par exemple pour A, B, C , on a :

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

Plus généralement, on a le résultat suivant (exercice : il se montre par récurrence sur n) :

Propriétés 2.2.2 Soit (A_n) une suite monotone d'événements. Alors

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

où

$$\lim_{n \rightarrow \infty} A_n = \begin{cases} \bigcup_{n=1}^{\infty} A_n, & \text{si } (A_n) \text{ est croissante;} \\ \bigcap_{n=1}^{\infty} A_n, & \text{si } (A_n) \text{ est décroissante.} \end{cases}$$

Démonstration. Supposons, d'abord, la suite (A_n) croissante. Avec $A_0 = \emptyset$ on a :

$$P\left(\lim_{n \rightarrow \infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} (A_n \setminus A_{n-1})\right).$$

donc

$$P\left(\lim_{n \rightarrow \infty} A_n\right) = \sum_{k=1}^{\infty} P(A_k \setminus A_{k-1}) = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(A_k \setminus A_{k-1}).$$

Or

$$\sum_{k=1}^n P(A_k \setminus A_{k-1}) = P(A_n).$$

Donc

$$P\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Si (A_n) est décroissante on pose $B = A^c$ donc d'après ce qui précède on

$$P\left(\bigcup_{n=1}^{\infty} A_n^c\right) = \lim_{n \rightarrow \infty} P(A_n^c).$$

Ainsi

$$P\left(\left(\bigcap_{n=1}^{\infty} A_n\right)^c\right) = \lim_{n \rightarrow \infty} P(A_n^c).$$

donc

$$1 - P\left(\left(\bigcap_{n=1}^{\infty} A_n\right)\right) = 1 - \lim_{n \rightarrow \infty} P(A_n).$$

Finalement

$$P\left(\left(\bigcap_{n=1}^{\infty} A_n\right)\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Propriétés 2.2.3 (Formule de Poincaré) Pour tout entier $n \geq 2$, et tous événements A_1, A_2, \dots, A_n , on a :

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) + \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}).$$

Définition 2.2.4 Une probabilité P sur $\mathcal{B}(\mathbb{R}^d)$ est dite discrète et portée par l'ensemble Ω si elle peut s'écrire sous la forme

$$P = \sum_{n \in \mathbb{N}} p_n \delta_{a_n}.$$

où $(p_n)_{n \in \mathbb{N}}$ est une suite de réels positifs ou nuls, $(a_n)_{n \in \mathbb{N}}$ est une suite de vecteurs de $\mathcal{B}(\mathbb{R}^d)$ et Ω désigne l'ensemble des a_n pour lesquels $p_n > 0$.

Les probabilités de Dirac δ_a , binomiale $\mathcal{B}(n, p)$, Poisson $\mathcal{P}(\lambda)$, géométrique $\mathcal{G}(p)$, uniforme-discrète $\mathcal{U}(n)$ sont discrètes et portées respectivement par les ensembles $\{a\}$, $\{0, 1, 2, \dots, n\}$, \mathbb{N} , \mathbb{N}^* , $\{1, 2, \dots, n\}$.

Définition 2.2.5 Soit (Ω, \mathcal{F}) et (E, \mathcal{E}) deux espaces mesurables. Une application $f : \Omega \rightarrow E$ est dite $(\mathcal{F}, \mathcal{E})$ -mesurable, si $f^{-1}(A) \in \mathcal{F}$ pour tout $A \in \mathcal{E}$, où

$$f^{-1}(A) = \{\omega \in \Omega \mid f(\omega) \in A\}.$$

Définition 2.2.6 Nous appellerons densité de probabilité sur \mathbb{R}^d toute application f positive de \mathbb{R} dans \mathbb{R}_+ , borélienne (i.e mesurable par rapport à la tribu borélienne), telle que

$$\int f(x) dx = 1.$$

Définition 2.2.7 Une probabilité P sur $\mathcal{B}(\mathbb{R}^d)$ est dite absolument continue si elle peut s'écrire sous la forme

$$P(A) = \int_A f(x) dx$$

où f est une densité de probabilité sur \mathbb{R}^d .

Les probabilités normale standard, uniforme-continue sur $[a, b]$, exponentielle de paramètre $\theta > 0$, sont des probabilités absolument continue.

Chapitre 3

Variables aléatoires

Les variables aléatoires sont l'objet de base de ce cours. Il s'agit de fonction du hasard dont nous présentons dans ce chapitre les outils clef pour leur étude : loi, fonction de répartition, les moments. Nous donnons la plupart des lois usuelles (discrètes et continues) et insistons sur leur interprétation en terme de modèle. Dans tout le chapitre, on considère un espace de probabilité (Ω, \mathcal{F}, P) .

3.1 Définitions

Définition 3.1.1 Soit (Ω, \mathcal{F}, P) un espace probabilisé. Un ensemble $A \in \mathcal{F}$ est dit négligeable pour P si $P[A] = 0$.

Définition 3.1.2 On dit qu'une propriété est vraie presque sûrement (p.s.) si elle est vraie en dehors d'un ensemble négligeable.

Définition 3.1.3 Soit (Ω, \mathcal{F}) et (E, \mathcal{E}) deux espaces mesurables. Une application $f : \Omega \rightarrow E$ est dite $(\mathcal{F}, \mathcal{E})$ -mesurable, si $f^{-1}(A) \in \mathcal{F}$ pour tout $A \in \mathcal{E}$, où

$$f^{-1}(A) = \{\omega \in \Omega | f(\omega) \in A\}.$$

3.2 Variable aléatoire

Définition 3.2.1 Soit (Ω, \mathcal{F}) un espaces mesurable. Une application $X : \Omega \rightarrow \mathbb{R}$ est dite variable aléatoire réelle (v.a.r) si X est mesurable de (Ω, \mathcal{F}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Exemple 3.2.1 (i) Soit $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$ la tribu borélienne de $[0, 1]$ et soit $P(A) = \lambda_{[0,1]}(A)$, $A \in \mathcal{B}(\mathbb{R})$, la mesure de Lebesgue sur $[0, 1]$. L'application identité de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable. C'est donc une variable aléatoire. On appelle aussi la probabilité uniforme sur $[0, 1]$, que l'on notera $U_{[0,1]}$.

Définition 3.2.2 La tribu engendrée par une variable aléatoire X définie sur (Ω, \mathcal{F}) est l'ensemble

$$\sigma(X) = \{X^{-1}(A), A \in \mathcal{B}(\mathbb{R})\}.$$

Définition 3.2.3 Soit X une v.a.r. définie sur (Ω, \mathcal{F}, P) . La loi de X est la probabilité P_X sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie comme mesure-image de P par X : pour tout $A \in \mathcal{B}(\mathbb{R})$,

$$P_X(A) = P[X \in A] = P[X^{-1}(A)] = P[\omega \in \Omega | X(\omega) \in A]$$

Pour les applications, en général, seule compte la mesure image, et l'on explicite rarement la variable aléatoire et l'espace (Ω, \mathcal{F}, P) . On écrira par exemple "soit X une variable de Bernoulli de paramètre p ", c'est-à-dire telle que $P\{X = 1\} = 1 - P\{X = 0\} = p$ au lieu de soit X une variable aléatoire de l'espace (Ω, \mathcal{F}, P) dans $\{0, 1\}$, de loi de Bernoulli, c'est-à-dire telle que $P_X(1) = 1 - P_X(0) = p$, ou plus exactement $P\{X = 1\} = 1 - P\{X = 0\} = p$."

Définition 3.2.4 Soient P et Q deux mesures sur (Ω, \mathcal{F})

— On dit que Q est absolument continue par rapport à P et on écrit $Q \ll P$ si pour tout $A \in \mathcal{F}$,

$$P[A] = 0 \implies Q[A] = 0$$

— On dit que P et Q sont équivalentes et on écrit $P \sim Q$ si pour tout $A \in \mathcal{F}$,

$$P[A] = 0 \iff Q[A] = 0$$

Théorème 3.2.1 Une Probabilité P est absolument continue par rapport à une mesure Q ssi il existe une fonction f mesurable positive telle que

$$P(A) = \int_A f dQ.$$

on dit que P a la densité f par rapport Q et on écrit $dP = f dQ$

Définition 3.2.5 On dit qu'une variable aléatoire est discrète si sa loi P_X est une combinaison linéaire finie ou dénombrable de masses de Dirac :

$$P_X = \sum_{n \in \mathbb{N}} p_n \delta_{a_n}.$$

où $(p_n)_{n \in \mathbb{N}}$ est une suite de réels positifs ou nuls, $(a_n)_{n \in \mathbb{N}}$ est une suite de réelles de $\mathcal{B}(\mathbb{R})$ et Ω désigne l'ensemble des a_n pour lesquels $p_n > 0$.

Définition 3.2.6 Si la loi de X est absolument continue par rapport à une mesure μ et que

$$dP = f d\mu,$$

on dit que X admet la densité f par rapport à μ . Si μ est la mesure de Lebesgue, on dit simplement que X est de densité f .

Exemple 3.2.2 - Variables aléatoires discrètes.

1- Variable aléatoire de Bernoulli de paramètre $0 < p < 1$ ($X \sim \mathcal{B}(1, p)$) :

$$P_X = p\delta_1 + (1 - p)\delta_0.$$

2- Variable aléatoire de loi binomiale de paramètre $0 < p < 1$ et n ($X \sim \mathcal{B}(n, p)$) :

$$P_X = \sum_{k=0}^n C_n^k p^k (1 - p)^{n-k} \delta_k.$$

3- Variable aléatoire de Poisson de paramètre λ ($X \sim \mathcal{P}(\lambda)$) :

$$P_X = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k.$$

4- Variable aléatoire géométrique de paramètre $0 < p < 1$ ($X \sim \mathcal{G}(p)$) :

$$P_X = \sum_{k=1}^{\infty} p(1-p)^{k-1} \delta_k.$$

5- Variable aléatoire uniforme-discrète de paramètre n ou équiprobabilité sur $\{1, 2, \dots, n\}$ ($X \sim \mathcal{U}(n)$)

$$P_X = \frac{1}{n} \sum_{k=1}^n \delta_k.$$

Exemple 3.2.3 - Variables aléatoires absolument continue.

1- Variable aléatoire normale standard $X \sim \mathcal{N}(0, 1)$ de densité

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

2- Variable aléatoire uniforme-continue sur $[a, b]$ $X \sim \mathcal{U}([a, b])$ de densité

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

3- Variable exponentielle de paramètre $\theta > 0$, $X \sim \mathcal{E}(\theta)$ de densité

$$f = \theta e^{-\theta x} 1_{]0, +\infty[}(x).$$

3.3 Fonctions de répartition

3.3.1 Fonctions de répartition d'une variable aléatoire réelle

Soit X une variable aléatoire réelle (i.e. X est à valeurs réelles), définie sur un espace probabilisé (Ω, \mathcal{F}, P) .

Définition 3.3.1 On appelle fonction de répartition de X , ou de sa loi P_X , et on note F_X , la fonction sur \mathbb{R} définie par

$$F_X(t) = P_X(]-\infty, t]) = P(\omega : X(\omega) \leq t) = P(X \leq t), t \in \mathbb{R}$$

La fonction de répartition de X ne dépend que de la loi de X puisque $F_X(t) = P_X(]-\infty, t])$. Un exercice classique d'intégration montre que la réciproque est également vraie : si on connaît la fonction de répartition de X , on peut calculer $P_X(]a; b])$ pour tout a et b , en vérifiant facilement que

$$F_X(b) - F_X(a) = P(]a, b]),$$

puis, puisque les intervalles du type $]a, b]$ engendrent les boréliens, on peut reconstituer P_X grâce à F_X .

Proposition 3.3.1 Une fonction de répartition F vérifie les propriétés suivantes :

(i) $0 \leq F \leq 1$,

(ii) F est croissante, continue à droite avec la limite à gauche suivante

$$F(t^-) = P_X(-\infty, t] = P(X < t).$$

De plus

$$P_X(\{t\}) = P(X = t) = F(t) - F(t^-),$$

donc F est continue en t ssi $P(X = t) = 0$.

(iii) $\lim_{t \rightarrow -\infty} F(t) = 0$ et $\lim_{t \rightarrow +\infty} F(t) = 1$.

Réciproquement, une fonction F vérifiant (i)-(iii) est la fonction de répartition d'une variable aléatoire réelle.

Démonstration. (i) vient de ce que P est à valeurs dans $[0, 1]$. La croissance dans (ii) découle de la croissance des mesures (i.e. $A \subset B \rightarrow P(A) \leq P(B)$). La continuité à droite peut être vue comme une conséquence de la Proposition 2.2.2 en remarquant que

$$\{X \leq t\} = \bigcap_{n \geq 1} \left\{ X \leq t + \frac{1}{n} \right\}.$$

Donc

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

et que la croissance de F implique

$$\lim_{h \rightarrow 0} F(t+h) = \lim_{n \rightarrow \infty} F\left(t + \frac{1}{n}\right) = F(t).$$

La limite à gauche est également une conséquence de la croissance de F . De plus

$$P(X = t) = P(X \leq t) - P(X < t) = F(t) - F(t^-),$$

donc F est continue en x ssi $P(X = t) = 0$.

La propriété (iii) vient encore de la Proposition 2.2.2, en remarquant que

$$\bigcap_{n \geq 1} \{X \leq -n\} = \emptyset.$$

Donc

$$0 = P(\emptyset) = \lim_{n \rightarrow +\infty} P(X \leq -n) = \lim_{n \rightarrow +\infty} F(-n) = \lim_{x \rightarrow -\infty} F(x).$$

Tandis que, d'après la Proposition 2.2.2,

$$1 = P(\Omega) = P\left(\bigcup_{n \geq 1} \{X \leq n\}\right) = \lim_{n \rightarrow +\infty} P(X \leq n) = \lim_{x \rightarrow +\infty} F(x)$$

De plus on a

$$\bigcap_{n \geq 1} \left\{ X \leq t - \frac{1}{n} \right\} = \{X < t\}.$$

Donc

$$F(x^-) = \lim_{n \rightarrow +\infty} F\left(x - \frac{1}{n}\right) = \lim_{n \rightarrow +\infty} P_X\left(\left]-\infty, x - \frac{1}{n}\right]\right) = P(X < x).$$

Propriétés 3.3.1 La fonction de répartition caractérise la loi, c'est-à-dire $F_X = F_Y$ si et seulement si $P_X = P_Y$.

Démonstration. En effet, si $F_X = F_Y$, alors P_X et P_Y coïncident sur les intervalles, donc la tribu engendrée par les intervalles. La tribu engendrée par les intervalles est la tribu borélienne et le résultat s'ensuit. La réciproque est triviale.

Propriétés 3.3.2 Une fonction de répartition admet au plus un nombre dénombrable de points de discontinuité.

Démonstration. Soit D_n l'ensemble des points de discontinuité avec un saut d'amplitude plus grande que $1/n$; en notant $F(t)$ la limite à gauche de F en t ,

$$D_n = \{t \in \mathbb{R}, F(t) - F(t^-) \geq 1/n\}.$$

L'ensemble des points de discontinuité est $D = \bigcup_{n \geq 1} D_n$. Puisque $0 \leq F \leq 1$, nécessairement

$$\text{card}(D_n) \leq n.$$

Sinon il existe au moins $t_1, t_2, \dots, t_{n+1} \in D_n$ ce qui implique la contradiction suivante

$$1 = P_X(\mathbb{R}) \geq (F(t_1) - F(t_1^-)) + (F(t_2) - F(t_2^-)) + \dots + (F(t_{n+1}) - F(t_{n+1}^-)) \geq \frac{n+1}{n} > 1.$$

Donc l'ensemble des points de discontinuité D est au plus dénombrable.

Propriétés 3.3.3 Soit X une variable de fonction de répartition F alors si

(i) X est discrète telle que $P_X = \sum_{n \in \mathbb{N}} p_n \delta_{a_n}$, alors

$$F(t) = \sum_{a_n \leq t} p_n,$$

(ii) X est continue de densité f (i.e absolument continue par rapport à la mesure de Lebesgue $dP_X = f d\lambda$) alors

$$F(t) = \int_{-\infty}^t f(x) dx, \text{ et } F(t_2) - F(t_1) = \int_{t_1}^{t_2} f(x) dx.$$

Remarque 3.3.1 On observe que la densité f doit vérifier $f(x) \geq 0$ et

$$\int_{\mathbb{R}} f(x) dx = 1.$$

Plus généralement on a la définition suivante.

Définition 3.3.2 Une fonction $f: \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable est appelée densité de probabilité si

- f est positive : en tout point où elle est définie, $f(t_1, \dots, t_d) \geq 0$,
- f est intégrable sur \mathbb{R}^d d'intégrale 1 :

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1$$

Propriétés 3.3.4 Soit X une variable de fonction de répartition F et de densité f , si

- (i) f est continue sur $[a, b]$, alors F est dérivable sur $[a, b]$ et on a $F' = f$ sur $[a, b]$.
- (ii) F est λ -prèsque partout dérivable, alors $f = F'$, λ -p.p.

Exemple 3.3.1 - Variables aléatoires discrètes.

1- $F = A_{[a, \infty[}$ est la fonction de répartition de la masse de Dirac δ_a .

2- Si $X \sim \mathcal{B}(1, p)$ (v.a. de bernoulli), alors

$$F_X(t) = \begin{cases} 0, & t < 0; \\ 1 - p, & 0 \leq x < 1; \\ 1, & 1 \leq t. \end{cases}$$

3- Si $X \sim \mathcal{B}(n, p)$ (v.a. de de loi binomiale), alors

$$F_X(t) = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} 1_{[k, \infty[}(t) = \sum_{k \leq \min\{n, [t]\}} C_n^k p^k (1-p)^{n-k}.$$

4- Si $X \sim \mathcal{P}(\lambda)$ (v.a. de poisson), alors

$$F_X(t) = \sum_{k \leq [t]} e^{-\lambda} \frac{\lambda^k}{k!}.$$

5- Si $X \sim \mathcal{G}(p)$ (v.a. géométrique), alors

$$F_X(t) = \sum_{k \leq [t]} p(1-p)^{k-1}.$$

6- Si $X \sim \mathcal{U}(n)$ (v.a. uniforme discrète), alors

$$F_X(t) = \frac{[t]}{n}.$$

Exemple 3.3.2 - Variables aléatoires continue.

1- Si $X \sim \mathcal{U}([a, b])$ (v.a. uniforme continue), alors

$$F_X(t) = \begin{cases} 0, & t < a; \\ \frac{x-a}{b-a}, & a \leq t < b; \\ 1, & b \leq t. \end{cases}$$

2- Si $X \sim \mathcal{E}(\theta)$ (v.a. exponentielle), alors de densité

$$F_X(t) = (1 - e^{-\theta t}) 1_{[0, +\infty[}(t).$$

3- Si $X \sim \mathcal{N}(0, 1)$ (v.a. normale standard) alors

$$F_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx.$$

Définition 3.3.3 Une variable aléatoire est dite continue si elle ne possède pas d'atome : $P(X = x) = 0 \forall x \in \mathbb{R}$ (i.e sa fonction de répartition est continue).

Propriétés 3.3.5 Une variable aléatoire réelle absolument continue est continue mais la réciproque est fausse.

Démonstration. Si X est absolument continue, on a alors

$$P_X(\{t\}) = \int_t^t f(x)dx,$$

pour tout x dans \mathbb{R} et la variable aléatoire X est bien continue.

Définition 3.3.4 Une variable aléatoire est dite *singulière* lorsqu'elle est continue mais pas absolument continue. C'est-à-dire qu'une loi singulière ne possède ni atome, ni densité.

On peut trouver des variables aléatoires continues sur \mathbb{R} mais qui ne sont pas absolument continues. Cependant, dans le cadre de ce cours, on se trouvera rarement dans cette situation. En revanche dans \mathbb{R}^2 , il est plus facile de trouver des variables aléatoires continues mais qui ne sont pas absolument continues.

Il existe des lois de probabilité qui ne sont ni discrètes, ni absolument continues, ni singulières, elles sont parfois appelées lois mixtes. Le théorème suivant, dû à Lebesgue, précise ceci.

Théorème 3.3.2 Soit F une fonction de répartition. Alors il existe trois fonctions de répartition F_1 discrète, F_2 absolument continue et F_3 singulière (i.e. continue mais non absolument continue) et trois nombres réels α_1, α_2 et α_3 positifs et de somme 1 tel que F puisse s'écrire sous la forme

$$F = \alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3.$$

Remarque 3.3.2 Il y a trois types de variables aléatoires :

- **Discrète**, est caractérisée par sa loi

$$P(X = x_i) = p_i.$$

- **Absolument continue**, est caractérisée par : sa fonction de répartition ou sa densité de probabilité :

$$F(t) = \int_{-\infty}^t f(x)dx.$$

- **Singulière**, est caractérisée par sa loi de probabilité et sa fonction de répartition $P(X \leq t) = F(t)$.

- **Mixte**, est caractérisée par : sa fonction de répartition : $F = \alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3$. Sa fonction de répartition est un mélange de trois fonctions de répartitions : F_1 discrète, F_2 absolument continue et F_3 singulière : $F = \alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3$.

Exemple 3.3.3 Soit la fonction de répartition F d'une loi variable aléatoire X , donnée par

$$F_X(t) = \begin{cases} 0, & t < 0; \\ \frac{t}{4}, & 0 \leq t < 1; \\ \frac{1}{2}, & 1 \leq t < 2; \\ \frac{2}{3} + \frac{1}{3}(1 - e^{-(t-2)}), & t \geq 2; \end{cases}$$

Le graphe de F comporte deux points de discontinuité en 1 et 2 d'amplitudes respectives $1/4$ et $1/6$. La partie continue est dérivable presque partout par rapport à la mesure de Lebesgue, de densité

$$f(x) = \frac{1}{4}1_{[0,1[}(x) + \frac{1}{3}e^{-(x-2)}1_{[2,+\infty[}(x).$$

La mesure de probabilité P se représente donc comme

$$P = \frac{1}{4}\delta_1 + \frac{1}{6}\delta_2 + \mu_c$$

avec μ_c la mesure de densité $f(x) = \frac{1}{4}1_{[0,1[}(x) + \frac{1}{3}e^{-(x-2)}1_{[2,+\infty[}(x)$. par rapport à la mesure de Lebesgue.

3.4 Espérance mathématique

3.4.1 Rappels d'intégration

On commence dans cette section par des rappels d'intégration en théorie de la mesure. On renvoie au cours d'intégration pour des détails ou des preuves.

Définition 3.4.1 Soit (Ω, \mathcal{F}, m) un espace mesurable muni d'une mesure m . L'intégrale d'une fonction f par rapport à la mesure m est définie comme suivant :

— f indicatrice : $f = 1_A$:

$$\int f dm = m(A)$$

— f étagée : $f = \sum_{k=1}^n \alpha_k 1_{A_k} \implies$

$$\int f dm = \sum_{k=1}^n \alpha_k m(A_k)$$

— f est mesurable positive : \implies

$$\int f dm = \sup_{0 \leq g \text{ étagée} \leq f} \int g dm$$

— Cas général : on dit que f est intégrable si elle est mesurable et que

$$f^+ = \max(f, 0) \text{ et } f^- = -\min(f, 0)$$

soient d'intégrale finie. On pose alors

$$\int f dm = \int f^+ dm - \int f^- dm.$$

3.4.2 Espérance d'une variable aléatoire

Définition 3.4.2 Soit X une v.a.r. définie sur (Ω, \mathcal{F}, P) telle que $\int_{\Omega} |X(\omega)| dP < \infty$. On définit alors l'espérance mathématique de X par

$$E(X) = \int_{\Omega} X dP = \int_{\mathbb{R}} x dP_X.$$

On dit que X est centrée si elle est intégrable et $E(X) = 0$.

L'espérance d'une variable aléatoire n'est donc rien d'autre que sa valeur moyenne. Une mesure de probabilité étant de masse totale égale à 1, l'espérance d'une variable aléatoire constante ou presque sûrement constante est égale à cette constante.

Si $\int |X| dP < +\infty$ et on note

$$L^1(\Omega, \mathcal{F}, P) = \left\{ X \text{ v.a.r. telle que } \int |X| dP < +\infty \right\}$$

Plus généralement, pour $p > 0$,

$$L^p(\Omega, \mathcal{F}, P) = \left\{ X \text{ v.a.r. telle que } \int |X|^p dP < +\infty \right\}.$$

Définition 3.4.3 Si $X \in L^p, p > 0$, on définit le moment absolu d'ordre p de X par

$$E(|X|^p) = \int |X|^p dP.$$

Si p est entier, on peut aussi définir le moment d'ordre p ,

$$E(X^p) = \int |X|^p dP.$$

Théorème 3.4.1 (de transport). Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}))$ et soit ϕ une fonction borélienne de \mathbb{R} dans \mathbb{R} . Si ϕ est à valeurs positives,

$$E(\phi(X)) = \int_{\Omega} \phi(X)(\omega) dP(d\omega) = \int_{\mathbb{R}} \phi(x) dP_X(x).$$

Si ϕ est à valeurs quelconques, $\phi(X) \in (\Omega, \mathcal{A}, P)$ si et seulement si $\phi \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$. Dans ce cas, l'égalité précédente a lieu.

Remarque 3.4.1 i) Si X est une variable aléatoire à valeurs dans \mathbb{R} , admettant une densité f , alors

$$E(\phi(X)) = \int_{\mathbb{R}} \phi(x) f(x) dx.$$

ii) Si X est une variable aléatoire à valeurs dans \mathbb{R}

$$P(X \in A) = E(1_A(X)) = \int_{\Omega} 1_A(X(\omega)) dP(\omega) = \int_{\mathbb{R}^d} 1_A(x) dP_X(x) = P_X(A)$$

3.5 Changement de variables

Soit X une v.a.r dans \mathbb{R} et g une application mesurable de \mathbb{R} vers \mathbb{R} . On veut déterminer la loi du vecteur aléatoire $g(X)$.

Théorème 3.5.1 Soit X une variable aléatoire à valeurs dans \mathbb{R} de densité f_X et soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une bijection de classe C^1 (ainsi que son inverse). Alors $Y = g(X)$ est une variable aléatoire à valeurs dans \mathbb{R} qui admet pour densité la fonction

$$f_Y(y) = \left| \frac{1}{g'(g^{-1}(y))} \right| \cdot f_X(g^{-1}(y)).$$

Démonstration. Il suffit d'écrire

$$\begin{aligned} F_Y(y) &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ &= \int_{-\infty}^{g^{-1}(y)} f_X(x) dx \end{aligned}$$

Puisque $g : \mathbb{R} \rightarrow \mathbb{R}$ une bijection de classe C^1 (ainsi que son inverse), le changement de variable $y = g(x)$ implique

$$F_Y(y) = \int_{-\infty}^y f_X(g^{-1}(y)) |(g^{-1})'(y)| dy,$$

avec

$$(g^{-1})'(y) = \frac{1}{g'(g^{-1}(y))}.$$

Nous avons également une méthode assez générale, valable même dans le cas où g n'est pas bijective, pour déterminer la loi d'une transformée d'une variable aléatoire, basée sur la caractérisation suivante.

Théorème 3.5.2 Soit Y une variable aléatoire à valeurs dans \mathbb{R} . Pour que sa loi P_Y , absolument continue, soit de densité f_Y il faut et il suffit que pour toute fonction borélienne $\phi : \mathbb{R} \rightarrow \mathbb{R}$ telle que $\phi(Y)$ soit intégrable ou positive, on ait :

$$E(\phi(Y)) = \int_{\mathbb{R}} \phi(y) f(y) dy.$$

Définition 3.5.1 Soit X une variable aléatoire réelle dont le carré est intégrable. On appelle variance de X , ou de sa loi P_X , et on note $V(X)$, la quantité

$$V(X) = E((X - E(X))^2)$$

La racine $\sqrt{V(X)}$ est appelée l'écart type, parfois noté $\sigma(X)$. Une variable aléatoire d'écart type 1 est dite réduite.

Une expression équivalente de la variance est

$$V(X) = E(X^2) - (E(X))^2.$$

La définition 3.5.1 montre que plus la variance est grande, plus la variable aléatoire est "dispersée", c'est-à-dire prend avec forte probabilité des valeurs éloignées de sa moyenne.

Inégalité 3.5.1 (Markov) Si X est intégrable et $t > 0$, alors

$$P(X \geq t) \leq \frac{E(|X|)}{t}.$$

Exemple 3.5.1 (i) Si $X \in L^p$, $p > 0$, alors

$$P(X \geq t) \leq \frac{E(|X|^p)}{t^p}.$$

(ii) Si $X \in L^2$, l'inégalité de Markov implique l'inégalité de Tchebitchev

$$P(|X - E(X)| \geq t) \leq \frac{V(X)}{t^2}.$$

3.6 Fonctions caractéristiques

Nous savons que la fonction de répartition d'une variable aléatoire réelle ou vectorielle X caractérise sa loi. Autrement dit, sur \mathbb{R} par exemple, la donnée de

$$F_X(t) = E\left(1_{]-\infty, t]}(X)\right), \quad t \in \mathbb{R}$$

détermine la loi de X . Puisque les indicatrices sont des fonctions boréliennes bornées. La fonction indicatrice $1_{]-\infty, t]}$ peut être approchée par une suite de fonctions continues bornées

$$\phi_n(x) = \begin{cases} 1, & \text{si } x \leq t, \\ 1 + n(t - x), & \text{si } t \leq x \leq t + \frac{1}{n}, \\ 0, & \text{si } x > t + \frac{1}{n}. \end{cases}$$

Il s'ensuit, d'après le théorème de convergence dominée, que la donnée de $E(\phi(X))$ pour toute fonction continue bornée sur \mathbb{R} caractérise P_X . Plus généralement, les fonctions indicatrices peuvent être approchées simplement par des fonctions C^∞ bornées; et donc la donnée de $E(\phi(X))$ pour toute fonction ϕ infiniment dérivable caractérise également P_X . On pourrait même se restreindre aux fonctions C^∞ à support compact. Ces raisonnements et conclusions s'appliquent de la même façon aux vecteurs aléatoires que nous abordons dans la suite.

Une autre caractérisation intéressante en pratique est celle des fonctions caractéristiques, ou transformées de Fourier, qui remplace la classe des fonctions C^∞ bornées par la famille des fonctions sinus et cosinus.

Définition 3.6.1 Soit X une variable aléatoire à valeurs dans \mathbb{R} . On appelle fonction caractéristique de X ou de la loi de X , ou transformée de Fourier, et on note ϕ_X , la fonction à valeurs complexes

$$t \in \mathbb{R}^d \longrightarrow \phi_X(t) = E(e^{itX}) = \int_{\mathbb{R}} e^{itx} dP_X(x).$$

La fonction caractéristique est à valeurs complexes, de module majoré par 1 (d'après l'inégalité de Jensen), et $\phi_X(0) = 1$. Si la loi de X a une densité f par rapport à la mesure de Lebesgue sur \mathbb{R}^d , alors

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx.$$

est aussi appelée la transformée de Fourier de la fonction f . Comme son nom l'indique, la fonction caractéristique caractérise la loi.

Théorème 3.6.1 Si X et Y sont deux variables aléatoires de lois P_X et P_Y telles que $\phi_X = \phi_Y$, alors $P_X = P_Y$.

Exemple 3.6.1 (i) Si $X = a$ p.s., i.e. $P_X = \delta_a$, $a \in \mathbb{R}$, alors $\phi_X(t) = e^{ita}$.
 ii) Si X suit une loi normale $\mathcal{N}(0, 1)$, alors

$$\phi_X(t) = E(e^{itX}) = \int_{\mathbb{R}} e^{itx-x^2/2} dx = e^{-t^2/2}.$$

Si Y est de loi $\mathcal{N}(m, \sigma^2)$, Y a la même loi que $\sigma X + m$, et donc

$$\phi_Y(t) = E(e^{it(\sigma X + m)}) = e^{itm - \sigma^2 t^2/2}.$$

iii) Si X est de loi exponentielle de densité e^{-x} sur \mathbb{R}_+ , alors

$$\phi_Y(t) = E(e^{itX}) = \int_0^{+\infty} e^{itx-x} dx = \frac{1}{1-it}.$$

v) Si X est de loi de Poisson de paramètre λ ,

$$\phi_Y(t) = E(e^{itX}) = \int e^{itx} dP_{\mathcal{P}(\lambda)}(x) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} e^{itk} = e^{\lambda(e^{it}-1)}.$$

vi) Si X est de loi binomiale de paramètres n et p , alors

$$\phi_Y(t) = E(e^{itX}) = \int e^{itx} dP_{\mathcal{B}(n,p)}(x) = \sum_{k=0}^n C_n^k e^{itk} p^k (1-p)^{n-k} = (1-p + pe^{it})^n.$$

Puisque la transformée de Fourier caractérise la loi, il est souhaitable d'avoir une formule d'inversion permettant d'obtenir effectivement la loi à partir de la fonction caractéristique. Il existe plusieurs formules de ce type permettant de calculer la densité si elle existe, ou la fonction de répartition. En voici une possible.

Théorème 3.6.2 (Formule d'inversion de Fourier) Soit ϕ_X la fonction caractéristique d'une v.a.r X , supposée intégrable par rapport à la mesure de Lebesgue sur \mathbb{R} . Alors, la loi de X admet une densité continue bornée f_X par rapport à la mesure de Lebesgue sur \mathbb{R} , donnée, pour tout $x \in \mathbb{R}$, par

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

La formule de Taylor montre alors que les moments de la variable sont proportionnels aux dérivées de la transformée de Fourier. Le résultat rigoureux est le suivant.

Proposition 3.6.3 Soit X une variable aléatoire réelle, de fonction caractéristique $\phi = \phi_X$ et de loi P_X . Si $E(|X|^n) < \infty$, alors ϕ est n -fois dérivable, de dérivée k -ième ($k \leq n$)

$$\phi^{(k)}(t) = i^k \int x^k e^{itx} dP_X(x) = i^k E(X^k e^{itX}).$$

En particulier,

$$\phi^{(k)}(0) = i^k E(X^k).$$

Réciproquement, si n est pair et si ϕ est n -fois dérivable en 0, alors X admet tout moment d'ordre plus petit ou égal à n .

Preuve. Montrons que ϕ est dérivable en tout point $t \in \mathbb{R}$ lorsque

$$E(|X|) < \infty.$$

Pour tout $h \neq 0$,

$$\frac{\phi(t+h) - \phi(t)}{h} = \int e^{itx} \frac{e^{ihx} - 1}{h} dP_X(x).$$

On a

$$\left| e^{itx} \frac{e^{ihx} - 1}{h} \right| \leq |x|$$

qui est intégrable pour P_X indépendamment de h . D'après le théorème de convergence dominée,

$$\phi'(t) = \lim_{h \rightarrow 0} \int e^{itx} \frac{e^{ihx} - 1}{h} dP_X(x) = \int ixe^{itx} dP_X(x) = iE(Xe^{itX}).$$

Les dérivées d'ordre supérieur se calculent de la même façon.

Chapitre 4

Vecteurs aléatoires

Un vecteur aléatoire est une généralisation à n dimensions d'une variable aléatoire réelle. Alors qu'une variable aléatoire réelle est une fonction qui à chaque éventualité fait correspondre un nombre réel, le vecteur aléatoire est une fonction X qui à chaque éventualité fait correspondre un vecteur de \mathbb{R}^n

Définition 4.0.1 Soit (Ω, \mathcal{F}) un espaces mesurable. Une application $X : \Omega \rightarrow \mathbb{R}^d$ est dite variable aléatoire vectorielle si X est mesurable de (Ω, \mathcal{F}) dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Remarque 4.0.1 Soient X_1, X_2, \dots, X_d des variables aléatoires : $X_i : \Omega \rightarrow \mathbb{R}$ alors $X = (X_1, X_2, \dots, X_d)$ est une variable aléatoire vectorielle de (Ω, \mathcal{F}) dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ appelé vecteur aléatoire.

4.1 Loi d'un vecteur aléatoire

Définition 4.1.1 Soit une variable aléatoire X sur l'espace probabilisé (Ω, \mathcal{F}, P) , à valeurs dans \mathbb{R}^d muni de la tribu borélienne réelle $\mathcal{B}(\mathbb{R}^d)$. La loi de la variable aléatoire X est la mesure de probabilité, P_X , définie par pour tout $B \in \mathcal{B}(\mathbb{R}^d)$:

$$P_X(B) = P(X^{-1}(B)) = P(X \in B).$$

Remarque 4.1.1 Si X est un vecteur aléatoire à valeurs dans \mathbb{R}^d

$$P(X \in A) = E(1_A(X)) = \int_{\Omega} 1_A(X(\omega)) dP(\omega) = \int_{\mathbb{R}^d} 1_A(x) dP_X(x) = P_X(A)$$

Remarque 4.1.2 i) Puisque les ensembles $A_1 \times \dots \times A_d$, où les A_i sont des boréliens, engendrent la tribu borélienne $\mathcal{B}(\mathbb{R}^d)$, alors P_X est caractérisée par pour tout $A_1, \dots, A_d \in \mathcal{B}(\mathbb{R})$:

$$\begin{aligned} P_X(A_1 \times \dots \times A_d) &= P(X^{-1}(A_1 \times \dots \times A_d)) \\ &= P(X \in A_1 \times \dots \times A_d) \\ &= E(1_{A_1 \times \dots \times A_d}(X)). \end{aligned}$$

ii) Si $X = (X_1, X_2, \dots, X_d)$ alors, puisque les pavés $]-\infty, x_1] \times \dots \times]-\infty, x_d]$ engendrent la tribu borélienne $\mathcal{B}(\mathbb{R}^d)$, la loi du vecteur aléatoire X P_X est caractérisée par pour tout $x_1, \dots, x_d \in \mathbb{R}$:

$$P_X(]-\infty, x_1] \times \dots \times]-\infty, x_d]) = P(X_1 \leq x_1, \dots, X_d \leq x_d).$$

La formule de transfert se généralise à d variables.

Théorème 4.1.1 (de transfert) Soit (X_1, \dots, X_d) un vecteur aléatoire sur (Ω, \mathcal{A}, P) à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et soit ϕ une fonction borélienne de \mathbb{R}^d dans \mathbb{R} . Si ϕ est à valeurs positives,

$$E(\phi(X_1, \dots, X_d)) = \int_{\Omega} \phi(X_1(\omega), \dots, X_d(\omega)) dP(d\omega) = \int_{\mathbb{R}^d} \phi(x_1, \dots, x_d) dP_{(X_1, \dots, X_d)}(x_1, \dots, x_d).$$

Si ϕ est à valeurs quelconques, $\phi(X_1, \dots, X_d) \in L^1(\Omega, \mathcal{A}, P)$ si et seulement si $\phi \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P_{(X_1, \dots, X_d)})$. Dans ce cas, l'égalité précédente a lieu.

La preuve est similaire à celle faite précédemment dans le cas d'une variable, on procède en approchant g par une fonction étagée.

4.1.1 Loi marginale

Définition 4.1.2 Soit (X_1, X_2, \dots, X_d) un vecteur aléatoire sur l'espace probabilisé (Ω, \mathcal{F}, P) , à valeurs dans \mathbb{R}^d , de loi $P_{(X_1, X_2, \dots, X_d)}$. La loi de probabilité de la i -ème coordonnée d'un vecteur aléatoire, notée P_{X_i} , est appelée la i -ème loi marginale.

Proposition 4.1.2 La loi marginale P_{X_i} s'obtient par la formule : pour tout $A \in \mathcal{B}(\mathbb{R})$,

$$P_{X_i}(A) = \int 1_A(x_i) dP_{(X_1, X_2, \dots, X_d)}(x_1, \dots, x_d),$$

où pour toute fonction ϕ borélienne positive de \mathbb{R} dans \mathbb{R}_+

$$E(\phi(X_i)) = \int_{\mathbb{R}^d} \phi(x_i) dP_{(X_1, \dots, X_d)}(x_1, \dots, x_d).$$

Démonstration. Il suffit de remarquer que pour tout $A \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned} P_{X_i}(A) &= P_{(X_1, X_2, \dots, X_d)}(\mathbb{R} \times \dots \times A \times \dots \times \mathbb{R}) \\ &= \int 1_{\mathbb{R} \times \dots \times A \times \dots \times \mathbb{R}}(x_1, \dots, x_d) dP_{(X_1, X_2, \dots, X_d)}(x_1, \dots, x_d) \\ &= \int 1_A(x_i) dP_{X_1, X_2, \dots, X_d}(x_1, \dots, x_d). \end{aligned}$$

En appliquant le théorème de transfert pour $\phi'((x_1, x_2, \dots, x_d)) = \phi(x_i)$ il vient

$$E(\phi(X_i)) = \int_{\mathbb{R}^d} \phi(x_i) dP_{(X_1, \dots, X_d)}(x_1, \dots, x_d).$$

Plus généralement on a la proposition suivante.

Proposition 4.1.3 Soient $k \leq d$ et $i_1 < i_2 < \dots < i_k$. La loi (ou loi marginale) du vecteur $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$, notée $P_{(X_{i_1}, X_{i_2}, \dots, X_{i_k})}$ s'obtient par la formule : pour tout $A_{i_1}, \dots, A_{i_k} \in \mathcal{B}(\mathbb{R})$,

$$P_{(X_{i_1}, X_{i_2}, \dots, X_{i_k})}(A_{i_1} \times \dots \times A_{i_k}) = \int 1_{A_{i_1} \times \dots \times A_{i_k}}(x_{i_1}, \dots, x_{i_k}) dP_{X_1, X_2, \dots, X_d}(x_1, \dots, x_d),$$

où pour toute fonction ϕ borélienne positive de \mathbb{R} dans \mathbb{R}_+

$$E(\phi(X_{i_1}, X_{i_2}, \dots, X_{i_k})) = \int_{\mathbb{R}^d} \phi(x_{i_1}, x_{i_2}, \dots, x_{i_k}) dP_{(X_1, \dots, X_d)}(x_1, \dots, x_d).$$

Définition 4.1.3 Un vecteur aléatoire (X_1, X_2, \dots, X_d) sur l'espace probabilisé (Ω, \mathcal{F}, P) , à valeurs dans \mathbb{R}^d , est dit discret si elle existe des parties $\mathcal{D}_i \subset \mathbb{R}$, $1 \leq i \leq d$ au plus dénombrable telle que

$$P_{(X_1, X_2, \dots, X_d)} = \sum_{(x_1, \dots, x_d) \in \mathcal{D}_1 \times \dots \times \mathcal{D}_d} p_{(x_1, \dots, x_d)} \delta_{(x_1, \dots, x_d)},$$

où

$$\sum_{(x_1, \dots, x_d) \in \mathcal{D}_1 \times \dots \times \mathcal{D}_d} p_{(x_1, \dots, x_d)} = 1 \quad \text{et} \quad p_{(x_1, \dots, x_d)} > 0.$$

Remarque 4.1.3 Si (X_1, X_2, \dots, X_d) est un vecteur aléatoire discret alors la loi, les lois marginales et les espérances sont caractérisées par :

1) Loi (X_1, X_2, \dots, X_d)

$$P(X_1 = x_1 = \dots = X_d = x_d) = p_{(x_1, \dots, x_d)} \quad \forall (x_1, x_2, \dots, x_d) \in \mathcal{D}_1 \times \dots \times \mathcal{D}_d.$$

2) Lois marginales : pour $k \leq d$ et $i_1 < i_2 < \dots < i_k$

$$P_{(X_{i_1}, X_{i_2}, \dots, X_{i_k})} = \sum_{(x_1, \dots, x_d) \in \mathcal{D}_1 \times \dots \times \mathcal{D}_d} p_{(x_1, \dots, x_d)} \delta_{(x_{i_1}, x_{i_2}, \dots, x_{i_k})},$$

en particulier

$$P_{X_i} = \sum_{(x_1, \dots, x_d) \in \mathcal{D}_1 \times \dots \times \mathcal{D}_d} p_{(x_1, \dots, x_d)} \delta_{x_i},$$

2) Si $\phi(X_i)$ est positive ou intégrable alors

$$E(\phi(X_i)) = \sum_{(x_1, \dots, x_d) \in \mathcal{D}_1 \times \dots \times \mathcal{D}_d} p_{(x_1, \dots, x_d)} \phi(x_i),$$

Comme il ressort de cette définition, la loi d'un vecteur $X = (X_1, X_2, \dots, X_d)$ détermine chacune des lois marginales (loi de X_i , $1 \leq i \leq d$). L'exemple suivant montre que la réciproque est fautive en général.

Exemple 4.1.1 Supposons que $X = (X_1, X_2)$ soit de loi discrète dans \mathbb{R}^2 concentrée en les points $(-1, 0), (0, 1), (0, -1), (1, 0)$ tous de probabilité $1/4$. Autrement dit,

$$P_X = \frac{1}{4} \delta_{(-1,0)} + \frac{1}{4} \delta_{(0,1)} + \frac{1}{4} \delta_{(0,-1)} + \frac{1}{4} \delta_{(1,0)},$$

ce qui se résume dans le tableau ci-contre. Les lois marginales P_{X_1} et P_{X_2} de P_X sont égales, et données par

$$P_{X_1} = P_{X_2} = \frac{1}{4} \delta_{-1} + \frac{1}{2} \delta_0 + \frac{1}{4} \delta_1.$$

	-1	0	1	$P_{X_1}(\cdot)$
-1	0	1/4	0	1/4
0	1/4	0	1/4	1/2
1	0	1/4	0	1/4
$P_{X_2}(\cdot)$	1/4	1/2	1/4	1

On peut produire un autre vecteur, (Y_1, Y_2) , ayant les mêmes lois marginales, dont les probabilités sont données par le tableau ci-contre.

	-1	0	1	$P_{Y_1}(\cdot)$
-1	1/16	1/8	1/16	1/4
0	1/8	1/4	1/8	1/2
1	1/16	1/8	1/16	1/4
$P_{Y_2}(\cdot)$	1/4	1/2	1/4	1

Remarque 4.1.4 On pourra noter que l'on obtient les lois marginales en sommant les probabilités respectivement sur les lignes et les colonnes de la table.

Définition 4.1.4 Soit une variable aléatoire X sur l'espace probabilisé (Ω, \mathcal{F}, P) , à valeurs dans \mathbb{R}^d muni de la tribu borélienne réelle $\mathcal{B}(\mathbb{R}^d)$. La loi de la variable aléatoire X , P_X , est dite absolument continue si elle est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^d , $d\lambda_{\mathbb{R}^d} = dx_1 \dots dx_d$, i.e

$$P_X \ll \lambda_{\mathbb{R}^d}.$$

Théorème 4.1.4 (Radon-Nikodim) Si la loi de la variable aléatoire X , P_X , est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^d , alors il existe une fonction intégrable $f: \mathbb{R}^d \rightarrow \mathbb{R}_+$, appelée densité, telle que

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_d) dx_1 \dots dx_d = 1,$$

et

$$\begin{aligned} P(X_1 \leq x_1, \dots, X_d \leq x_d) &= P_X([-\infty, x_1] \times \dots \times [-\infty, x_d]) \\ &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f(x_1, \dots, x_d) dx_1 \dots dx_d. \end{aligned}$$

La proposition suivante, qui est une conséquence du théorème de Radon-Nikodim et la Proposition 4.1.3, montre que si X est un vecteur aléatoire absolument continu, tout vecteur aléatoire marginal est également absolument continu et sa densité est obtenue en intégrant la densité conjointe de X par rapport aux coordonnées restantes.

Proposition 4.1.5 Soient $k \leq d$ et $i_1 < i_2 < \dots < i_k$. La densité marginale du vecteur $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$, notée $f_{(X_{i_1}, X_{i_2}, \dots, X_{i_k})}$ s'obtient par la formule : $\forall (x_{i_1}, \dots, x_{i_k}) \in \mathbb{R}^k$

$$f_{(X_{i_1}, X_{i_2}, \dots, X_{i_k})}(x_{i_1}, \dots, x_{i_k}) = \int_{\mathbb{R}^{d-k}} f_{(X_1, X_2, \dots, X_d)}(x_1, \dots, x_d) dx_1 \dots dx_{i_1-1} dx_{i_1+1} \dots dx_{i_k-1} dx_{i_k+1} \dots dx_d,$$

où pour toute fonction ϕ borélienne positive de \mathbb{R} dans \mathbb{R}_+

$$E(\phi(X_{i_1}, X_{i_2}, \dots, X_{i_k})) = \int_{\mathbb{R}^d} \phi(x_{i_1}, x_{i_2}, \dots, x_{i_k}) f_{(X_{i_1}, X_{i_2}, \dots, X_{i_k})}(x_{i_1}, \dots, x_{i_k}) dx_1 \dots dx_d.$$

Exemple 4.1.2 Considérons le vecteur aléatoire gaussien

$$X = (X_1, X_2, \dots, X_d) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_{\mathbb{R}^d})$$

dont la densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d est donnée par

$$f_X(x) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|x\|^2}{2}},$$

avec $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ et la norme euclidienne $\|x\|^2 = x_1^2 + \dots + x_d^2$. D'après la proposition précédente, les densités marginales sont des lois gaussiennes. en effet pour $k \leq d$ et $i_1 < i_2 < \dots < i_k$, La densité marginale du vecteur $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$, est

$$\begin{aligned}
f_{(X_{i_1}, X_{i_2}, \dots, X_{i_k})}(x_{i_1}, \dots, x_{i_k}) &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^{d-k}} e^{-\frac{x_1^2 + \dots + x_d^2}{2}} dx_1 \dots dx_{i_1-1} dx_{i_1+1} \dots dx_{i_k-1} dx_{i_k+1} \dots dx_d \\
&= \left(\frac{1}{(2\pi)^{\frac{k}{2}}} e^{-\frac{x_{i_1}^2 + x_{i_2}^2 + \dots + x_{i_k}^2}{2}} \right) \frac{1}{(2\pi)^{\frac{d-k}{2}}} \int_{\mathbb{R}^{d-k}} e^{-\frac{x_1^2 + x_2^2 + \dots + x_{d-k}^2}{2}} \prod_{i \notin \{i_1, i_2, \dots, i_k\}} dx_i \\
&= \left(\frac{1}{(2\pi)^{\frac{k}{2}}} e^{-\frac{x_{i_1}^2 + x_{i_2}^2 + \dots + x_{i_k}^2}{2}} \right) \left(\frac{1}{(2\pi)^{\frac{1}{2}}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx \right)^{d-k} \\
&= \frac{1}{(2\pi)^{\frac{k}{2}}} e^{-\frac{x_{i_1}^2 + x_{i_2}^2 + \dots + x_{i_k}^2}{2}}.
\end{aligned}$$

4.2 Fonctions de répartition d'un vecteur aléatoire

Définition 4.2.1 Soient X_1, X_2, \dots, X_d des variables aléatoires : $X_i : \Omega \rightarrow \mathbb{R}$. On appelle fonction de répartition de $X = (X_1, X_2, \dots, X_d)$, ou de la loi de X , la fonction

$$t = (t_1, \dots, t_d) \in \mathbb{R}^d \longrightarrow F_X(t) = P((X_1 \leq t_1) \cap \dots \cap (X_d \leq t_d)) = P(X_1 \leq t_1, \dots, X_d \leq t_d).$$

La fonctions de répartition marginale de la variable aléatoire X_i est donnée par

$$F_{X_i}(t_i) = \lim_{t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_d \rightarrow +\infty} F_X(t).$$

Proposition 4.2.1 i) Si $X = (X_1, X_2, \dots, X_d)$ est un vecteur discret alors

$$F_{(X_1, X_2, \dots, X_d)}(t_1, \dots, t_d) = \sum_{x_1 \in]\infty, t_1]} \dots \sum_{x_d \in]\infty, t_d]} p_{(x_1, \dots, x_d)}.$$

ii) Si $X = (X_1, X_2, \dots, X_d)$ est absolument continue par rapport à la mesure de Lebesgue de densité f alors

$$F_{(X_1, X_2, \dots, X_d)}(t_1, \dots, t_d) = \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_d} f_{(X_1, X_2, \dots, X_d)}(x_1, \dots, x_d) dx_1 \dots dx_d.$$

Proposition 4.2.2 Soient $k \leq d$ et $i_1 < i_2 < \dots < i_k$. La fonction de répartition marginale du vecteur $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$, notée $F_{(X_{i_1}, X_{i_2}, \dots, X_{i_k})}$ s'obtient par la formule : $\forall (t_{i_1}, \dots, t_{i_k}) \in \mathbb{R}^k$

$$F_{(X_{i_1}, X_{i_2}, \dots, X_{i_k})}(t_{i_1}, \dots, t_{i_k}) = \int_{-\infty}^{t_{i_1}} \dots \int_{-\infty}^{t_{i_k}} \int_{\mathbb{R}^{d-k}} f_{(X_1, X_2, \dots, X_d)}(x_1, \dots, x_d) dx_1 \dots dx_d.$$

Exemple 4.2.1 Soit (X, Y) un couple aléatoire de densité $f(x, y)$ par rapport à la mesure de Lebesgue sur \mathbb{R}^2 , i.e.

$$F(t_1, t_2) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} f(x, y) dx dy, \quad t_1, t_2 \in \mathbb{R}.$$

alors les fonctions de répartition marginales et les densités marginales sont données par

$$F_X(t_1) = \int_{-\infty}^{t_1} \int_{-\infty}^{+\infty} f(x, y) dy dx, \quad f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy,$$

et

$$F_Y(t_2) = \int_{-\infty}^{t_2} \int_{-\infty}^{+\infty} f(x, y) dx dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

4.3 Changement de variables

Soit X un vecteur aléatoire dans \mathbb{R}^d et g une application mesurable de \mathbb{R}^d vers \mathbb{R}^d . On veut déterminer la loi du vecteur aléatoire $g(X)$.

Théorème 4.3.1 Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^d de densité f_X et soit $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ un difféomorphisme de \mathbb{R}^d (bijection continûment différentiable ainsi que son inverse). Alors $Y = g(X)$ est une variable aléatoire à valeurs dans \mathbb{R}^d qui admet pour densité la fonction

$$f_Y(y) = f_X(g^{-1}(y)) |Jac g^{-1}(y)|; \quad y \in \mathbb{R}^d$$

où

$$Jac g^{-1}(y) = \det \left(\frac{(g^{-1}(y))_i}{dy_j}, \quad 1 \leq i, j \leq d \right).$$

est le jacobien de g^{-1} .

Démonstration. Il suffit d'écrire

$$\begin{aligned} F_Y(y_1, \dots, y_d) &= P((g(X))_1 \leq y_1, \dots, (g(X))_d \leq y_d) \\ &= \int_{g^{-1}([-\infty, y_1] \times \dots \times [-\infty, y_d])} f_X(x_1, \dots, x_d) dx_1 \dots dx_d. \end{aligned}$$

Puisque $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ est un difféomorphisme, le changement de variable $y = g(x)$ implique

$$\begin{aligned} F_Y(y_1, \dots, y_d) &= \int_{-\infty}^{y_1} \dots \int_{-\infty}^{y_d} f_X((g^{-1}(y))_1, \dots, (g^{-1}(y))_d) |Jac g^{-1}(y)| dy_1 \dots dy_d \\ &= \int_{-\infty}^{y_1} \dots \int_{-\infty}^{y_d} f_X(g^{-1}(y)) |Jac g^{-1}(y)| dy. \end{aligned}$$

Nous avons également une méthode assez générale, valable même dans le cas où g n'est pas difféomorphisme, pour déterminer la loi d'une transformée d'un vecteur aléatoire, basée sur la caractérisation suivante.

Théorème 4.3.2 Soit Y un vecteur aléatoire à valeurs dans \mathbb{R}^d . Pour que sa loi P_Y , absolument continue, soit de densité f_Y il faut et il suffit que pour toute fonction borélienne bornée $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ telle que $\phi(Y)$ soit intégrable ou positive, on ait :

$$E(\phi(Y)) = \int_{\mathbb{R}^d} \phi(y) f(y) dy.$$

Définition 4.3.1 Soient X_1, X_2, \dots, X_d des variables aléatoires sur un espace probabilisé : $X_i : \Omega \rightarrow \mathbb{R}$ positives ou de carré intégrable. L'espérance du vecteur $X = (X_1, X_2, \dots, X_d)$ est le vecteur

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

Sa matrice de covariance est définie par

$$\text{Cov}(X) = (E(X_i - E(X_i))(X_j - E(X_j)))_{1 \leq i, j \leq d} = E((X - E(X))(X - E(X))^t),$$

où A^t est la transposée de la matrice A .

Remarque 4.3.1 (i) Si X_i et X_j sont deux variables aléatoires réelles alors

$$\begin{aligned} \text{Cov}(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) \\ &= \int_{\mathbb{R}^d} x_i x_j dP_{(X_1, \dots, X_d)}(x_1, \dots, x_d) - \int_{\mathbb{R}^d} x_i dP_{(X_1, \dots, X_d)}(x_1, \dots, x_d) \times \int_{\mathbb{R}^d} x_j dP_{(X_1, \dots, X_d)}(x_1, \dots, x_d). \end{aligned}$$

(ii) Si X_1 est une variable aléatoire réelle alors

$$\text{Cov}(X_i, X_i) = V(X_i) = \int_{\mathbb{R}^d} x_i^2 dP_{(X_1, \dots, X_d)}(x_1, \dots, x_d) - \left(\int_{\mathbb{R}^d} x_i dP_{(X_1, \dots, X_d)}(x_1, \dots, x_d) \right)^2.$$

(iii) Si (X_1, X_2, \dots, X_d) est dit discret de loi

$$P_{(X_1, X_2, \dots, X_d)} = \sum_{(x_1, \dots, x_d) \in \mathcal{D}_1 \times \dots \times \mathcal{D}_d} p(x_1, \dots, x_d) \delta_{(x_1, \dots, x_d)},$$

alors les espérances marginales sont données par

$$E(X_i) = \sum_{(x_1, \dots, x_d) \in \mathcal{D}_1 \times \dots \times \mathcal{D}_d} x_i p(x_1, \dots, x_d),$$

$$E(X_i^2) = \sum_{(x_1, \dots, x_d) \in \mathcal{D}_1 \times \dots \times \mathcal{D}_d} x_i^2 p(x_1, \dots, x_d),$$

et

$$E(X_i X_j) = \sum_{(x_1, \dots, x_d) \in \mathcal{D}_1 \times \dots \times \mathcal{D}_d} x_i x_j p(x_1, \dots, x_d),$$

(iii) Si (X_1, X_2, \dots, X_d) est absolument continu de densité $f_{(X_1, X_2, \dots, X_d)}$ alors les espérances marginales sont données par

$$E(X_i) = \int_{\mathbb{R}^d} x_i f_{(X_1, X_2, \dots, X_d)}(x_1, \dots, x_d) dx_1 \dots dx_d,$$

$$E(X_i^2) = \int_{\mathbb{R}^d} x_i^2 f_{(X_1, X_2, \dots, X_d)}(x_1, \dots, x_d) dx_1 \dots dx_d,$$

et

$$E(X_i X_j) = \int_{\mathbb{R}^d} x_i x_j f_{(X_1, X_2, \dots, X_d)}(x_1, \dots, x_d) dx_1 \dots dx_d,$$

Exemple 4.3.1 Si $X = (X_1, X_2)$ est de loi discrète dans \mathbb{R}^2 telle que

$$P_X = \frac{1}{4}\delta_{(-1,0)} + \frac{1}{4}\delta_{(0,1)} + \frac{1}{4}\delta_{(0,-1)} + \frac{1}{4}\delta_{(1,0)},$$

alors

$$E(X_1) = -1 \times p_{(-1,0)} + 0 \times p_{(0,1)} + 0 \times p_{(0,-1)} + 1 \times p_{(1,0)} = \frac{1}{4} - \frac{1}{4} = 0,$$

$$E(X_2) = \frac{1}{4} \times 0 \times p_{(-1,0)} + 1 \times p_{(0,1)} + (-1) \times p_{(0,-1)} + 0 \times p_{(1,0)} = \frac{1}{4} - \frac{1}{4} = 0,$$

$$V(X_1) = E(X_1^2) = (-1)^2 \times p_{(-1,0)} + 0^2 \times p_{(0,1)} + 0^2 \times p_{(0,-1)} + 1^2 \times p_{(1,0)} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

$$V(X_2) = E(X_2^2) = (0)^2 \times p_{(-1,0)} + 1^2 \times p_{(0,1)} + (-1)^2 \times p_{(0,-1)} + 0^2 \times p_{(1,0)} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

$$\text{Cov}(X_1, X_2) = E(X_1 X_2) = -1 \times 0 \times p_{(-1,0)} + 0 \times 1 \times p_{(0,1)} + 0 \times (-1) \times p_{(0,-1)} + 1 \times 0 \times p_{(1,0)} = 0.$$

Exemple 4.3.2 Considérons le vecteur aléatoire gaussien $X = (X_1, X_2, \dots, X_d)$ dont la densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d est donnée par

$$f_X(x_1, \dots, x_d) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{x_1^2 + \dots + x_d^2}{2}},$$

alors

$$\begin{aligned} E(X_i) &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} x_i e^{-\frac{x_1^2 + \dots + x_d^2}{2}} dx_1 \dots dx_d \\ &= \left(\frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx \right) \left(\frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \right)^{d-1} \\ &= 0, \end{aligned}$$

$$\begin{aligned} V(X_i) &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} x_i^2 e^{-\frac{x_1^2 + \dots + x_d^2}{2}} dx_1 \dots dx_d \\ &= \left(\frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx \right) \left(\frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \right)^{d-1} \\ &= 1, \end{aligned}$$

et pour $i \neq j$

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} x_i x_j e^{-\frac{x_1^2 + \dots + x_d^2}{2}} dx_1 \dots dx_d \\ &= \left(\frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx \right)^2 \left(\frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \right)^{d-2} \\ &= 0. \end{aligned}$$

Définition 4.3.2 On dit qu'un vecteur aléatoire $X = (X_1, \dots, X_n)$ suit la loi multinomiale de paramètres $n > 0, p_1 > 0, \dots, p_m > 0$ tels que $p_1 + \dots + p_m = 1$, on écrit $X \sim \mathcal{B}(n, p_1, \dots, p_m)$ si

$$\mathbb{P}(X_1 = n_1, \dots, X_n = n_m) = \frac{n!}{n_1! \dots n_m!} p_1^{n_1} \dots p_m^{n_m}$$

pour tout (n_1, \dots, n_m) tel que

$$\sum_{i=1}^m n_i = n$$

Exercice 4.3.1 Soit $X \sim \mathcal{B}(n, p_1, \dots, p_m)$ alors

1. Montrer que chacune des variables reste une variable binomiale.
2. Montrer que

$$E(X_i) = np_i \quad \text{var}(X_i) = np_i(1 - p_i)$$

tandis que les covariances s'écrivent

$$\text{cov}(X_i, X_j) = -np_i p_j$$

4.4 Fonctions caractéristiques

Définition 4.4.1 Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^d . On appelle fonction caractéristique de X ou de la loi de X , ou transformée de Fourier, et on note ϕ_X , la fonction à valeurs complexes

$$t \in \mathbb{R}^d \longrightarrow \phi_X(t) = E(e^{i\langle t, X \rangle}) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} dP_X(x).$$

Théorème 4.4.1 Si X et Y sont deux vecteurs aléatoires de lois P_X et P_Y telles que $\phi_X = \phi_Y$, alors $P_X = P_Y$.

Exemple 4.4.1 (i) Si $X = a$ p.s., i.e. $P_X = \delta_a$, $a \in \mathbb{R}^d$, alors $\phi_X(t) = e^{i\langle t, a \rangle}$.

(ii) Si X est un vecteur aléatoire à valeurs dans \mathbb{R}^d , A une matrice opérant sur \mathbb{R}^d et $m \in \mathbb{R}^d$, alors $Y = AX + m$ est un vecteur aléatoire de fonction caractéristique

$$\phi_Y(t) = e^{i\langle t, m \rangle} \phi_X(A^t t)$$

puisque

$$\langle it, AX + m \rangle = \langle iA^t t, X \rangle + \langle it, m \rangle.$$

Donnons la version vectorielle de la Formule d'inversion de Fourier.

Théorème 4.4.2 (Formule d'inversion de Fourier). Soit ϕ_X la fonction caractéristique d'un vecteur aléatoire X , supposée intégrable par rapport à la mesure de Lebesgue sur \mathbb{R}^d . Alors, la loi de X admet une densité continue bornée f_X par rapport à la mesure de Lebesgue sur \mathbb{R}^d , donnée, pour tout $x \in \mathbb{R}^d$, par

$$f_X(x) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-i\langle t, x \rangle} \phi_X(t) dt.$$

Proposition 4.4.3 Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire, de fonction caractéristique $\phi = \phi_X$ et de loi P_X . Si $E(\|X\|^2) < \infty$, alors

$$E(X_k) = -i \frac{\partial \phi}{\partial t_k}(0), \quad E(X_k X_j) = -i \frac{\partial^2 \phi}{\partial t_k \partial t_j}(0).$$

Définition 4.4.2 Soit $X = (X_1, X_2, \dots, X_d)$ un vecteur aléatoire de loi P_X . X est dit de loi normale multidimensionnelle de paramètres $\boldsymbol{\mu} \in \mathbb{R}^d$ et $\Sigma \in \mathcal{M}_d(\mathbb{R})$, ou Normale multivariée ou multinormale ou loi de Gauss à plusieurs variables si P_X est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^d de densité définie par pour tout $x \in \mathbb{R}^d$

$$f_{\boldsymbol{\mu}, \Sigma}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}.$$

Cette loi est habituellement notée $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ par analogie avec la loi normale unidimensionnelle.

Remarque 4.4.1 Si $\boldsymbol{\mu} = (0, \dots, 0)$, $\Sigma = I_d$, alors $X = (X_1, X_2, \dots, X_d)$ est dit de loi normale multidimensionnelle centrée réduite dont la densité est définie par pour tout $x \in \mathbb{R}^d$

$$f_{\mathcal{N}(0, I_d)}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{x_1^2 + \dots + x_d^2}{2}}.$$

Exercice 4.4.1 Soit $X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ alors

1. Montrer que

$$E(X) = \boldsymbol{\mu}, \quad \text{Cov}(X) = \Sigma.$$

2. Montrer que la fonction caractéristique s'écrit

$$\phi_{\boldsymbol{\mu}, \Sigma}(\mathbf{t}) = \exp\left(i\mathbf{t}^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top \Sigma \mathbf{t}\right).$$

Chapitre 5

Indépendance

5.1 Définitions et critères d'indépendance

Définition 5.1.1 Une famille quelconque d'évènements $(A_i)_{i \in I} \in \mathcal{F}$ est mutuellement indépendante si :

$$\forall J \subset I \text{ fini} \quad P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

Exercice 5.1.1 (Jet de deux dés (Bleu, Rouge)) Soient A, B, C trois évènements représentant respectivement le dé rouge est impaire, le dé bleu est impair, la somme des deux dés est impair. $P(A) = P(B) = P(C) = 1/2$ et $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4$, mais

$$P(A \cap B \cap C) = 0 \neq 1/8.$$

Ainsi, A, B, C sont deux à deux indépendants mais pas mutuellement indépendants.

Définition 5.1.2 Une famille de sous-tribus $\{\mathcal{F}_1, \dots, \mathcal{F}_n\}$ de \mathcal{F} est (mutuellement) indépendante si pour tout $A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$ la famille $\{A_1, \dots, A_n\}$ est indépendante. Autrement dit

$$P[A_1 \cap \dots \cap A_n] = P[A_1] \times \dots \times P[A_n] \quad \forall A_1 \in \mathcal{F}_1 \dots \forall A_n \in \mathcal{F}_n.$$

Définition 5.1.3 Une famille X_1, \dots, X_n de v.a. (\mathcal{F} -mesurables) est (mutuellement) indépendante si $(\sigma(X_1), \dots, \sigma(X_n))$ est indépendante. Autrement dit, $\forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$

$$P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n P(X_i \in B_i).$$

Définition 5.1.4 Une v. a. X est indépendante d'une sous-tribu \mathcal{F} si $\sigma(X) \perp \mathcal{F}$. En particulier, deux variables X et Y sont indépendantes si $\sigma(X) \perp \sigma(Y)$ et on notera alors $X \perp Y$.

Exercice 5.1.2 jet de deux dés.

Soient X_a, X_b deux variables aléatoires donnant respectivement le résultat du dé rouge et celui du dé bleu. Ces deux variables sont indépendantes.

Remarque : L'indépendance de tribus est difficile à prouver en général, pour la prouver en pratique on la résultat suivant :

Proposition 5.1.1 Si C_1 et C_2 sont deux famille indépendantes alors $\sigma(C_1)$ et $\sigma(C_2)$ sont indépendantes.

5.1.1 Indépendance de variables aléatoires et corrélation.

Proposition 5.1.2 Soit (X_1, \dots, X_d) une famille de variables aléatoires indépendantes réelles, alors la loi du vecteur aléatoire sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est égale au produit des lois marginales

$$P_{(X_1, \dots, X_d)} = P_{X_1} \otimes \dots \otimes P_{X_d}$$

Réciproquement si $P_{(X_1, \dots, X_d)} = P_{X_1} \otimes \dots \otimes P_{X_d}$ alors (X_1, \dots, X_d) est indépendante.

Exercice 5.1.3 Dans \mathbb{R}^2

		X_1			
		0	-1	0	1
X_2	-1	0	1/4	0	
	0	1/4	0	1/4	
	1	0	1/4	0	

On pose $X = (X_1, X_2)$, on a :

$$P_X = \frac{1}{4}\delta_{(-1,0)} + \frac{1}{4}\delta_{(0,1)} + \frac{1}{4}\delta_{(1,0)} + \frac{1}{4}\delta_{(0,-1)},$$

alors

$$P(X = (0, 0)) = 0 \neq \frac{1}{4} = P(X_1 = 0)P(X_2 = 0),$$

ainsi X_1 et X_2 ne sont pas indépendantes.

Corollaire 5.1.3 X_1, \dots, X_d indépendantes si et seulement si $\forall \phi_i, i \in \{1, \dots, n\}$ fonctions boréliennes telles que les $\phi_i(X_i)$ sont intégrables, on ait :

$$E\left(\prod_{i=1}^n \phi_i(X_i)\right) = \prod_{i=1}^n (E(\phi_i(X_i))).$$

5.1.2 Indépendance, fonctions caractéristiques et de répartitions.

Proposition 5.1.4 : Une famille (X_1, \dots, X_n) de variables aléatoires est une famille indépendante si et seulement si : $\forall (t_1, \dots, t_n) \in \mathbb{R}^n$,

$$\varphi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \prod_{i=1}^n \varphi_{X_i}(t_i),$$

où $\varphi_X : \mathbb{R}^d \rightarrow \mathbb{R}$ est la fonction caractéristique de X : $\varphi_X(t) = E(e^{i\langle t, X \rangle})$

Proposition 5.1.5 : Soient X_1 et X_2 deux variables aléatoires, elles sont indépendantes si et seulement si

$$F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2).$$

Proposition 5.1.6 Soient (X, Y) un couple aléatoires de densité conjointe $f(x, y)$, alors X et Y sont indépendantes si et seulement si :

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

Proposition 5.1.7 Soit X et Y deux variables aléatoires indépendantes telle que

$$X(\Omega) = \{x_1, \dots, x_n\}, \quad Y(\Omega) = \{y_1, \dots, y_m\}.$$

Alors X et Y sont indépendantes si et seulement si

$$p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j), \quad \forall (i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}.$$

5.1.3 Corrélacion et vecteur gaussien.

Définition 5.1.5 : Deux variables aléatoires X et Y dans $L^2(\Omega, \mathcal{A}, P)$ sont non corrélées si :

$$E(XY) = E(X)E(Y) \Leftrightarrow E((X - E(X))(Y - E(Y))) = Cov(X, Y) = 0$$

Remarque : D'après ce qui précède on a X, Y indépendantes implique X, Y non corrélées.

Contre-exemple : $Y = X^2$, $X \sim \mathcal{N}(0, 1)$, $E(X) = 0$ et $E(XY) = 0$, ainsi $E(X)E(Y) = E(XY)$ mais X et Y ne sont pas indépendantes :

$$P(|X| \leq 1, Y \geq 1) = 0 \neq P(|X| \leq 1)P(Y \geq 1).$$

Théorème 5.1.8 Un vecteur $X = (X_1, \dots, X_d)$ est gaussien si $\forall \alpha \in \mathbb{R}^d$, $\sum_{i=1}^d \alpha_i X_i$ est une variable aléatoire gaussienne.

Dans ce cas,

$$\sum_{i=1}^d \alpha_i X_i \sim \mathcal{N} \left(\sum_{i=1}^d \alpha_i E(X_i), \sum_{i=1}^d \alpha_i \alpha_j Cov(X_i, X_j) \right).$$

Proposition 5.1.9 Un vecteur gaussien est entièrement déterminé par

$$m = E(X_i), \quad \Gamma = Cov(X_i, X_j).$$

Théorème 5.1.10 Soit X un vecteur gaussien de \mathbb{R}^d de matrice de covariance Γ . Alors, Γ est diagonale si et seulement si (X_1, \dots, X_d) est indépendante.

Remarque : La réciproque fautive dans le cas général.

Contre-exemple : Il ne suffit pas, pour que X soit gaussien que ses marginales soient gaussiennes :

$Z \sim \mathcal{N}(0, 1)$, ϵ de Rademacher :

$$\epsilon = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1.$$

Z et ϵ sont indépendants, alors $(Z, \epsilon Z)$ n'est pas gaussien mais de marginales gaussiennes.

5.2 Loi conditionnelle

Définition 5.2.1 Soit X et Y deux variables aléatoires indépendantes telle que

$$X(\Omega) = \{x_1, \dots, x_n\}, \quad Y(\Omega) = \{y_1, \dots, y_m\}.$$

Soit $x_i \in X(\Omega)$. La loi conditionnelle de Y sachant $X = x_i$ est la loi discrète prenant les valeurs y_j avec les probabilités

$$p_{j/i} = P(Y = y_j | X = x_i).$$

Exemple 5.2.1 Soient X, Y des v.a. entières indépendantes de loi de Poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$. Soit $T = X + Y$. Alors

$$X_{/T=t} \sim B \left(t, \frac{\lambda}{\lambda + \mu} \right).$$

On veut maintenant définir l'analogue des probabilités conditionnelles vues dans le cas discret.

Définition 5.2.2 Soient X et Y deux variables aléatoires de densité conjointe $f(x, y)$. La densité conditionnelle de X sachant $Y = y$ est définie par

$$f_{X/Y=y}(x) = \begin{cases} \frac{f(x,y)}{\int f(x,y)dx}, & \text{si } f(y) \neq 0, \\ 0, & \text{si } f(y) = 0. \end{cases}$$

Remarque 5.2.1 i) on a

$$f_Y(y) = \int f(x,y)dx, \quad f_X(x) = \int f(x,y)dy.$$

ii) Si $f(x) \neq 0$, et $f(y) \neq 0$, alors

$$f(x, y) = f(y)f_{X/Y=y}(x) = f(x)f_{Y/X=x}(y).$$

5.3 Noyaux et lois conditionnelles

Définition 5.3.1 $\nu : E \times \mathcal{F} \rightarrow [0, 1]$ est un noyau (ou probabilité) de transition si il satisfait :

1. $\forall x \in E, \nu(x, \cdot)$ est une probabilité sur (F, \mathcal{F}) .
2. $\forall B \in \mathcal{F}, \nu(\cdot, B)$ est \mathcal{E} -mesurable.

Définition 5.3.2 Si $X \in \mathbb{R}^d$ et $Y \in \mathbb{R}^k$ sont deux variables aléatoires, on appelle loi conditionnelle de Y sachant X un noyau ν sur $(\mathbb{R}^d \times \mathcal{B}(\mathbb{R}^k))$ tel que :

$$P_{(X,Y)} = P_X \cdot \nu$$

c'est-à-dire

$$P(X \in A; Y \in B) = \int_{x \in A} \nu(x, B) dP_X(x)$$

On notera, **et ce n'est qu'une notation**, un tel noyau ν de la manière suivante :

$$\nu(x, B) = P(Y \in B | X = x) \text{ ou } \nu(x, B) = P_Y(B | X = x)$$

Remarquer par exemple que si X est à densité, $\{X = x\}$ est un ensemble de probabilité nulle. Il est donc hors de question de la définir par :

$$P(Y \in B | X = x) = \frac{P(Y \in B; X = x)}{P(X = x)}$$

Cependant cette formule est vraie dans le cas où $P(X = x) \neq 0$.

Il reste à savoir quand on a l'existence de lois conditionnelles. La réponse à cette question est apportée par le théorème non trivial suivant.

Théorème 5.3.1 : théorème de Jirina

Si X et Y sont deux variables aléatoires à valeurs dans des $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ alors il existe toujours une loi conditionnelle de Y sachant X .

Définition 5.3.3 Soit (X, Y) un couple aléatoires de Loi image $P_{X,Y}$. La loi conditionnelle de X sachant $Y = y$, notée $P_{X/Y=y}$, est la mesure de probabilité qui vérifie pour ϕ mesurable positive

$$E(\phi(X, Y)) = \int \phi(x, y) dP_{X,Y}(x, y) = \int \left(\int \phi(x, y) dP_{X/Y=y}(x) \right) dP_Y(y).$$

Remarque 5.3.1 i) En pratique on travail par les $\phi(x, y) = g(x) \times h(y)$ et donc l'égalité précédente devient

$$E(\phi(X, Y)) = \int \left(\int g(x) dP_{X/Y=y}(x) \right) h(y) dP_Y(y).$$

(i) Il est facile de remarquer qu'au cas où (X, Y) admet une densité $f(x, y)$ alors, le bon choix de la loi conditionnelle de X sachant $Y = y$ dans la définition précédente est bien

$$dP_{X/Y=y}(x) = f_{X/Y=y}(x) d\lambda(x)$$

où

$$f_{X/Y=y}(x) = \begin{cases} \frac{f(x,y)}{\int f(x,y) dx}, & \text{si } f(y) \neq 0, \\ 0, & \text{si } f(y) = 0. \end{cases}$$

Exemple 5.3.1 Soient X et Y des v.a. positives, indépendantes, de même loi de densité e^{-x} par rapport à λ_+ mesure de lebesgue sur \mathbb{R}^+ . Soient

$$T = X + Y \quad \text{et} \quad U = \max(X, Y).$$

On veut calculer les lois conditionnelles

i) X sachant que $T = t$

ii) X sachant que $U = u$.

(i) Il est facile de calculer la densité de (T, X) par rapport à $\lambda_+ \otimes \lambda_+$. On a, pour $g \geq 0$ arbitraire,

$$\begin{aligned} E(g(T, X)) &= \int_0^\infty \int_0^\infty g(x+y, x) e^{-(x+y)} dx dy \\ &= \int_0^\infty \int_0^t g(t, x) e^{-t} dt dx \end{aligned}$$

et (T, X) a pour densité

$$h(t, x) = e^{-t} I_{\{0 < x < t\}}.$$

Alors la densité de T est

$$\phi(t) = \int h(t, x) dx = te^{-t}.$$

On a donc

$$h(x/t) = \frac{h(t, x)}{\phi(t)} = \frac{1}{t} I_{\{]0, t[\}}(x).$$

(ii) Si on conditionne par U , la situation est plus délicate. En effet $P(X = U) = \frac{1}{2}$ et le couple (U, X) n'a pas de densité par rapport à une mesure produit. Il faut utiliser la formule générale. Calculons d'abord la loi de U . On a

$$\begin{aligned} P(U \leq u) &= P(X \leq u, Y \leq u) \\ &= P(X \leq u)P(Y \leq u) \\ &= (1 - e^{-u})^2. \end{aligned}$$

U a pour densité $\phi(u) = 2e^{-u}(1 - e^{-u})$. Pour f, g positives, on a

$$\begin{aligned}
 E(g(U)f(X)) &= E(g(U)f(X)I_{\{X \leq Y\}}) + E(g(U)f(X)I_{\{X > Y\}}) \\
 &= E(g(Y)f(X)I_{\{X \leq Y\}}) + E(g(X)f(X)I_{\{X > Y\}}) \\
 &= \int_0^\infty \int_0^\infty g(y)f(x)I_{\{x \leq y\}}e^{-(x+y)} dx dy + \int_0^\infty \int_0^\infty g(x)f(x)I_{\{x > y\}}e^{-(x+y)} dx dy \\
 &= \int_0^\infty g(y)e^{-y} \int_0^y f(x)e^{-x} dx dy + \int_0^\infty g(x)f(x)e^{-x}(1 - e^{-x}) dx \\
 &= \int_0^\infty g(u) \frac{1}{2(1 - e^{-u})} \left[\int_0^u f(x)e^{-x} dx \right] \phi(u) du + \int_0^\infty g(u) \frac{1}{2} f(u) \phi(u) du \\
 &= \int_0^\infty g(u) \left[\frac{1}{2(1 - e^{-u})} \int_0^u f(x)e^{-x} dx + \frac{1}{2} f(u) \right] \phi(u) du \\
 &= \int_0^\infty g(u) \left[\frac{1}{2(1 - e^{-u})} \int_0^u f(x)e^{-x} dx + \frac{1}{2} f(u) \right] \phi(u) du.
 \end{aligned}$$

donc la loi conditionnelle de X sachant que $U = u$ est

$$P_{X/U=u} = \frac{1}{2(1 - e^{-u})} e^{-x} I_{\{]0, u[\}}(x) dx + \frac{1}{2} \delta_u(dx).$$

où δ_u est la mesure de Dirac du point u .

Remarque 5.3.2 1) $X \perp Y$ ssi

$$F(x, y) = F_X(x)F_Y(y).$$

2) $X \perp Y$ ssi

$$E(f(X)g(Y)) = E(f(X))E(g(Y)).$$

3) cas continue : $X \perp Y$ ssi

$$f(x, y) = f(x)f(y).$$

4) cas continue : $X \perp Y$ ssi

$$f(y) = f_{Y/X=x}(y).$$

5) cas continue : $X \perp Y$ ssi

$$f(x) = f_{X/Y=y}(x).$$

6) cas discret : $X \perp Y$ ssi

$$p_{ij} = p_i \cdot p_j.$$

7) cas discret : $X \perp Y$ ssi

$$p_{i/j} = p_i.$$

8) cas discret : $X \perp Y$ ssi

$$p_{j/i} = p_j.$$

En probabilités, la corrélation entre deux ou plusieurs variables aléatoires, représente l'intensité de la liaison qui peut exister entre ces variables.

La mesure de la corrélation linéaire entre les deux variables se fait alors par le calcul du coefficient de corrélation linéaire suivant.

Définition 5.3.4 Si X et Y admettent chacune une variance non nulle, on appelle coefficient de corrélation linéaire de X et Y le réel noté $\rho(X, Y)$ et définie par :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Propriétés 5.3.1 i) $\rho(X, Y)$ est compris dans $[-1, 1]$.

ii) Les deux variables ne sont pas corrélées si $\rho(X, Y)$ est nul.

iii) Les deux variables sont d'autant mieux corrélées que $\rho(X, Y)$ est proche de 1 ou de -1.

5.4 Somme de variables aléatoires indépendantes.

Démontrons d'abord un important résultat sur la variance de la somme.

Proposition 5.4.1 Si X_1, X_2, \dots, X_n est une suite indépendante de variables aléatoires réelles de carré intégrable, alors

$$E((X_1 + \dots + X_n)^2) < \infty.$$

et

$$V(X_1 + \dots + X_n) = \sum_{k=1}^n V(X_k).$$

La réciproque est fautive.

Preuve : Laisée en exercice.

Proposition 5.4.2 Si (X_1, \dots, X_n) sont des variables aléatoires indépendantes alors : $\forall t \in \mathbb{R}$,

$$\varphi_{(X_1 + \dots + X_n)}(t) = \prod_{i=1}^n \varphi_{X_i}(t),$$

où $\varphi_X : \mathbb{R} \rightarrow \mathbb{R}$ est la fonction caractéristique de X : $\varphi_X(t) = E(e^{itX})$.

En particulier on a le résultat suivant

Proposition 5.4.3 Si X_1, \dots, X_n, X sont des variables aléatoires indépendantes de même loi (i.i.d) alors : $\forall t \in \mathbb{R}$,

$$\varphi_{(X_1 + \dots + X_n)}(t) = (\varphi_X(t))^n.$$

Théorème 5.4.4 Soient X_1, \dots, X_m des variables aléatoires indépendantes suivantes chacune une loi Binomiale de paramètres n_i et p , alors leur somme X suit la loi binomiale :

$$X = \sum_{k=1}^m X_k \sim \mathcal{B}(n_1 + \dots + n_m, p).$$

Preuve. On a

$$\begin{aligned} \varphi_{(X_1 + \dots + X_m)}(t) &= \prod_{i=1}^m \varphi_{X_i}(t) \\ &= \prod_{i=1}^m (pe^{it} + 1 - p)^{n_i} \\ &= (pe^{it} + 1 - p)^{n_1 + \dots + n_m}. \end{aligned}$$

En particulier on a le résultat suivant

Théorème 5.4.5 Si X_1, \dots, X_m sont des variables aléatoires de Bernoulli avec paramètre p , indépendantes et identiquement distribuées, alors leur somme X suit la loi binomiale :

$$X = \sum_{k=1}^m X_k \sim \mathcal{B}(m, p).$$

Preuve. il suffit de remarquer que $\mathcal{B}(p) \sim \mathcal{B}(1, p)$.

Théorème 5.4.6 Soient X_1, \dots, X_m des variables aléatoires indépendantes suivant chacune une loi Poisson $\mathcal{P}(\lambda_i)$ de paramètre $\lambda_i > 0$. Alors leur somme X suit la loi Poisson :

$$X = \sum_{k=1}^m X_k \sim \mathcal{P}(\lambda_1 + \dots + \lambda_m).$$

Preuve. La fonction caractéristique d'une loi Poisson de paramètre $\lambda > 0$ est

$$\varphi_X(t) = \exp(\lambda(e^{it} - 1)).$$

Ainsi, on a

$$\begin{aligned} \varphi_{(X_1 + \dots + X_m)}(t) &= \prod_{i=1}^m \varphi_{X_i}(t) \\ &= \prod_{i=1}^m \exp(\lambda_i(e^{it} - 1)) \\ &= \exp((\lambda_1 + \dots + \lambda_m)(e^{it} - 1)). \end{aligned}$$

Proposition 5.4.7 Soient X et Y deux variables aléatoires réelles, indépendantes. La loi de la somme $X + Y$ est donnée par le produit de convolution $P_X * P_Y$ des lois P_X et P_Y , défini, pour toute fonction borélienne bornée ϕ de \mathbb{R} dans \mathbb{R} , par

$$\int_{\mathbb{R}} \phi dP_X * P_Y = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \phi(x+y) dP_X(x) \right) P_Y(y).$$

Preuve. L'indépendance implique $P_{(X,Y)} = P_X \otimes P_Y$ donc

$$\begin{aligned} E(\phi(X+Y)) &= \int \phi(X+Y) dP \\ &= \int \phi(x+y) dP_{(X,Y)}(x,y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \phi(x+y) dP_X(x) \otimes P_Y(y) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \phi(x+y) dP_X(x) \right) P_Y(y) \\ &= \int_{\mathbb{R}} \phi dP_X * P_Y. \end{aligned}$$

Proposition 5.4.8 Soient X et Y deux variables aléatoires réelles indépendantes de densités f_X et g_Y respectivement. La densité de la somme $X + Y$ est donnée par le produit de convolution $h_{X+Y} = f_X * g_Y$ défini par

$$h_{X+Y}(t) = \int_{\mathbb{R}} f_X(x)g_Y(t-x)dx.$$

Preuve. On a

$$\begin{aligned} E(\phi(X+Y)) &= \int_{\mathbb{R}} \phi dP_X * P_Y \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \phi(x+y)dP_X(x) \right) P_Y(y) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \phi(x+y)f(x)dx \right) g(y)dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \phi(t)f(x)g(t-x)dxdt \\ &= \int_{\mathbb{R}} \phi(t)h(t)dt. \end{aligned}$$

Théorème 5.4.9 Soient X_1, \dots, X_n des variables aléatoires indépendantes suivants chacune une loi normale d'espérance μ_i et de variance σ_i^2 . Alors la somme $X_1 + \dots + X_n$ suit une loi normale d'espérance $\mu_1 + \dots + \mu_n$ et de variance $\sigma_1^2 + \dots + \sigma_n^2$

Preuve. Nous allons considérer le cas $n = 2$, le cas général s'en déduit par récurrence. Il est plus commode de travailler avec

$$Y_1 = X_1 - \mu_1 \quad \text{et} \quad Y_2 = X_2 - \mu_2,$$

qui suivent des lois normales centrées (c'est-à-dire de moyenne nulle). Soit

$$u = u(y_2) = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1 \sigma_2} y_2 - \frac{\sigma_2}{\sigma_1 \sqrt{\sigma_1^2 + \sigma_2^2}} z.$$

Nous allons utiliser le fait que

$$u^2 + \frac{z^2}{\sigma_1^2 + \sigma_2^2} = \frac{(z - y_2)^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2},$$

qui se vérifie par un calcul élémentaire. La densité g de $Y_1 + Y_2$ est alors donnée par la convolution

$$\begin{aligned} g(z) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma_1} e^{-(z-y_2)^2/2\sigma_1^2} \frac{1}{\sqrt{2\pi} \sigma_2} e^{-y_2^2/2\sigma_2^2} dy_2 \\ &= \frac{1}{2\pi \sigma_1 \sigma_2} \int_{-\infty}^{\infty} \exp^{-\frac{1}{2} \left(\frac{(z-y_2)^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2} \right)} dy_2 \\ &= \frac{1}{2\pi \sigma_1 \sigma_2} \int_{-\infty}^{\infty} e^{-u^2/2} e^{-z^2/2(\sigma_1^2 + \sigma_2^2)} \frac{\sigma_1 \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} du \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-z^2/2(\sigma_1^2 + \sigma_2^2)}, \end{aligned}$$

où nous avons utilisé le changement de variable $y_2 \mapsto u(y_2)$. Ceci montre que $Y_1 + Y_2$ suit la loi $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$, et par conséquent $X_1 + X_2$ suit la loi $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Autre Preuve. Puisque X et Y sont indépendantes alors le vecteur (X, Y) est gaussien donc toute combinaison est gaussienne. Ainsi $X + Y$ est gaussien d'espérance $E(X_1 + X_2) = \mu_1 + \mu_2$ et de variance $V(X_1 + X_2) = \sigma_1^2 + \sigma_2^2$ et par conséquent $X_1 + X_2$ suit la loi $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Autre Preuve. La fonction caractéristique de $X_1 + X_2$ est donnée par

$$\begin{aligned}\varphi_{X_1+X_2}(t) &= E(e^{it(X_1+X_2)}) \\ &= E(e^{itX_1})E(e^{itX_2}) \\ &= \varphi_{X_1}(t)\varphi_{X_2}(t) \\ &= e^{i\mu_1 t - \frac{\sigma_1^2 t^2}{2}} e^{i\mu_2 t - \frac{\sigma_2^2 t^2}{2}} \\ &= e^{i(\mu_1 + \mu_2)t - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}},\end{aligned}$$

par conséquent $X_1 + X_2$ suit la loi $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. \square

Chapitre 6

Espérance conditionnelle

Dans ce chapitre, (Ω, \mathcal{F}, P) désignera un espace probabilisé et \mathcal{G} sera une sous-tribu de \mathcal{F} . Nous donnerons pour simplifier la notion d'espérance conditionnelle en dimension 1 (variables aléatoires réelles). Bien entendu, tout ceci se généralise sans peine au cas multidimensionnel (vecteurs aléatoires).

6.1 Introduction : variables aléatoires discrètes

On se donne deux variables aléatoires discrètes X et Z à valeurs dans \mathbb{N} à support fini. L'idée intuitive de probabilité conditionnelle nous amène à prendre pour valeur de la "probabilité de $\{X = i\}$ sachant " $\{Z = j\}$ " la quantité suivante

$$P(X = i|Z = j) = \frac{P(X = i, Z = j)}{P(Z = j)}$$

pour j appartenant au support de Z . Si j n'appartient pas au support de Z , cette définition n'a pas de sens. Par suite, l'espérance de X sachant $\{Z = j\}$ sera définie comme

$$\mathbb{E}(X|Z = j) = \sum_{i \in \mathbb{N}} iP(X = i|Z = j)$$

Définissons maintenant l'espérance conditionnelle de X sachant Z comme la variable aléatoire $Y(\omega) = \mathbb{E}(X|Z)(\omega)$ donnée par

$$Y(\omega) = \sum_j \mathbb{I}_{Z^{-1}(\{j\})}(\omega) \mathbb{E}(X|Z = j)$$

En particulier, si ω n'appartient pas au support de Z , $Y(\omega) = 0$.

1) Remarquons que pour tout borélien $B \in \sigma(Z)$, c'est-à-dire $B \subset \mathcal{P}(Z^{-1}(\{z\}); z \in \text{Support}(Z))$, on a

$$\mathbb{E}(Y \mathbb{I}_B) = \mathbb{E}(X \mathbb{I}_B)$$

2) Y est $\sigma(Z)$ -mesurable.

La question qui se pose est de savoir comment généraliser cette notion d'espérance conditionnelle lorsque Z n'est pas à valeurs discrètes. En particulier, on remarquera que si Z est à densité, l'événement $\{Z = z\}$ est pour tout réel z de probabilité nulle si bien que la définition même de probabilité conditionnelle devient délicate. La première idée consisterait à approximer l'événement $\{Z = z\}$ par $\lim_{\varepsilon \rightarrow 0} \{Z \in (z - \varepsilon; z + \varepsilon)\}$ et l'on peut effectivement procéder ainsi mais cela devient assez vite compliqué. On va procéder tout autrement.

6.2 Théorème d'existence et d'unicité

Théorème 6.2.1 Soit X une variable aléatoire sur (Ω, \mathcal{F}, P) intégrable. Il existe une variable aléatoire Y telle que :

- 1) Y est \mathcal{G} -mesurable.
- 2) $\mathbb{E}(|Y|) < +\infty$.
- 3) $\forall G \in \mathcal{G}, \int_G Y dP = \int_G X dP$.

De plus, si \tilde{Y} est une autre variable aléatoire vérifiant ces trois propriétés alors $Y = \tilde{Y}$ P p.s. On parlera alors de Y comme d'une version de l'espérance conditionnelle $\mathbb{E}(X|\mathcal{G})$ qui est un élément de $L^1(\Omega, \mathcal{G}, P)$. Par abus de langage, on parlera de Y (ou toute autre version) comme de l'espérance conditionnelle de X sachant \mathcal{G} .

Avant d'entamer la preuve de l'existence de l'espérance conditionnelle, faisons quelques remarques.

Remarque 6.2.1 Si l'on veut prouver 3), on peut se restreindre à montrer l'égalité pour des G qui appartiennent à une famille contenant Ω et générant \mathcal{G} (Théorème de la classe monotone). D'autre part, si on a 3), par le théorème de la classe monotone, on a aussi que

$$\mathbb{E}(YU) = \mathbb{E}(XU)$$

pour toute variable aléatoire U qui est \mathcal{G} -mesurable bornée.

Dans la démonstration, nous aurons besoin de considérer les espaces quotients $L^2(\Omega, \mathcal{G}, P)$ et $L^2(\Omega, \mathcal{F}, P)$. On a bien sûr l'inclusion triviale $L^2(\Omega, \mathcal{G}, P) \subset L^2(\Omega, \mathcal{F}, P)$. Cette inclusion passe au quotient puisque si $X, Y \in L^2(\Omega, \mathcal{G}, P)$ définissent le même élément dans $\mathbb{L}^2(\Omega, \mathcal{G}, P)$, c'est que $\{Y \neq X\} \in \mathcal{G}$ est de probabilité nulle donc que X et Y définissent aussi le même élément dans $L^2(\Omega, \mathcal{F}, P)$. D'autre part, $L^2(\Omega, \mathcal{G}, P)$ est fermé dans $L^2(\Omega, \mathcal{F}, P)$.

Preuve 6.2.1 (du théorème 6.2.1).

Existence : On commence par traiter le cas où X admet un moment d'ordre 2. $\mathbb{L}^2(\Omega, \mathcal{G}, P)$ est un sev complet de $\mathbb{L}^2(\Omega, \mathcal{F}, P)$ pour la norme $\|\cdot\|_2$ dérivant du produit-scalaire noté $\langle \cdot, \cdot \rangle$. On désigne alors par Y la projection orthogonale de X sur $\mathbb{L}^2(\Omega, \mathcal{G}, P)$. Par définition de la projection orthogonale,

$$\forall Z \in \mathbb{L}^2(\Omega, \mathcal{G}, P), \langle X - Y, Z \rangle = 0$$

donc $\mathbb{E}(X\mathbb{I}_G) = \mathbb{E}(Y\mathbb{I}_G)$ pour tout $G \in \mathcal{G}$. Ceci prouve l'existence de l'espérance conditionnelle pour les variables aléatoires ayant des moments d'ordre 2. Passons maintenant au cas général.

Soit $X \in \mathbb{L}^1(\Omega, \mathcal{F}, P)$. En l'écrivant comme différence de sa partie positive et de sa partie négative : $X = X^+ - X^-$, il est clair que l'on peut se ramener au cas où X est positive. On suppose donc maintenant que X est positive. Soit pour tout entier n , $X_n = X \wedge n \in L^2$. On peut donc par ce qui précède prendre son espérance conditionnelle par rapport à \mathcal{G} , $Y_n = \mathbb{E}(X_n|\mathcal{G})$. D'autre part, X_n tend simplement en croissant vers X et on va voir que de même, $(Y_n)_n$ tend en croissant vers une certaine limite Y .

Lemme 6.2.2 $0 \leq Y_n \leq Y_{n+1}$, P p.s.

Preuve 6.2.2 On prouve que si $U \geq 0$ admet un moment d'ordre 2 alors $W = \mathbb{E}(U|\mathcal{G}) \geq 0$. Si $P(W < 0) > 0$ alors il existe n tel que $P(W < -1/n) > 0$ et $\{W < -1/n\} \in \mathcal{G}$ donc

$$0 \leq \mathbb{E}(U\mathbb{I}_{W < -1/n}) = \mathbb{E}(W\mathbb{I}_{W < -1/n}) \leq -1/n$$

ce qui est exclu. On en déduit immédiatement le lemme (par linéarité de l'espérance conditionnelle sur $L^2(\Omega, \mathcal{F}, P)$).

On pose $Y = \limsup Y_n$ qui est \mathcal{G} -mesurable. Pour tout $G \in \mathcal{G}$,

$$\begin{aligned} \mathbb{E}(Y\mathbb{I}_G) &= \lim_{n \rightarrow \infty} \mathbb{E}(Y_n\mathbb{I}_G) \text{ par convergence monotone} \\ &= \lim_{n \rightarrow \infty} \mathbb{E}(X_n\mathbb{I}_G) \\ &= \mathbb{E}(X\mathbb{I}_G) \text{ par convergence monotone} \end{aligned}$$

En particulier, avec $G = \Omega$, $\mathbb{E}(Y) = \mathbb{E}(X) < +\infty$.

Unicité : Soit Y et \tilde{Y} deux variables aléatoires vérifiant 1), 2) et 3). On veut montrer que $Y = \tilde{Y}$, P presque sûrement. On a donc :

$$\forall G \in \mathcal{G}, \mathbb{E}(\mathbb{I}_G(Y - \tilde{Y})) = 0$$

Si $P(Y = \tilde{Y}) < 1$ alors quitte à permuter Y et \tilde{Y} , on a

$$P(Y > \tilde{Y}) > 0$$

Or $\lim_{n \rightarrow \infty} P(Y - \tilde{Y} > 1/n) = P(Y > \tilde{Y})$ et il existe donc n tel que $P(Y - \tilde{Y} > 1/n) > 0$. Mais $G_n = \{Y - \tilde{Y} > 1/n\} \in \mathcal{G}$ par 1) donc

$$\mathbb{E}(\mathbb{I}_{G_n}(Y - \tilde{Y})) = 0 \Rightarrow \frac{1}{n}P(Y - \tilde{Y} > 1/n) \leq 0 \quad : \text{exclu}$$

Ceci conclut la preuve de l'unicité et du théorème.

Remarque 6.2.2 1) Quand on considère au lieu de \mathcal{G} la tribu engendrée par une variable aléatoire Z (resp. une famille de variables aléatoires $(Z_i)_{i \in I}$), il est d'usage de noter l'espérance conditionnelle de X sachant $\sigma(Z)$ par $\mathbb{E}(X|Z)$ (resp. $\mathbb{E}(X|Z_i, i \in I)$).

2) On remarquera que $\mathbb{E}(X|Z)$ est $\sigma(Z)$ -mesurable donc elle s'écrit sous la forme $f(Z)$ où f est une certaine fonction borélienne.

Propriétés 6.2.1 On a les propriétés suivantes

1. $E[E[X|\mathcal{G}]] = E[X]$
2. Si X est \mathcal{G} -mesurable, $E[X|\mathcal{G}] = X$.
3. Si X est indépendante de \mathcal{G} , $E[X|\mathcal{G}] = E[X]$.
4. Si Y est \mathcal{H} -mesurable et si XY est intégrable, $E[XY|\mathcal{H}] = YE[X|\mathcal{H}]$
5. Si \mathcal{G} et \mathcal{H} sont deux tribus telles que $\mathcal{H} \subset \mathcal{G}$, alors

$$E[X|\mathcal{H}] = E[E[X|\mathcal{G}]|\mathcal{H}] = E[E[X|\mathcal{H}]|\mathcal{G}].$$

6.3 Variables à densité

Soit X et Z deux variables aléatoires réelles (pour simplifier les écritures) telles qu'elles aient une densité jointe $f_{X,Z}(x,z)$. On rappelle que X (resp. Z) a alors une densité notée $f_Z(z) = \int f_{X,Z}(x,z)dx$ (resp. $f_X(x) = \int f_{X,Z}(x,z)dz$). On pose alors

$$f_{X|Z}(x|z) = \mathbb{I}_{f_Z(z) \neq 0} \frac{f_{X,Z}(x,z)}{f_Z(z)}.$$

Soit h une fonction borélienne (bornée pour simplifier); quelle est l'espérance conditionnelle de $h(X)$ sachant Z ?

Proposition 6.3.1 Soit X et Z deux v.a.r. ayant une densité jointe $f_{X,Z}(x,z)$. Alors pour h une fonction borélienne bornée

$$\mathbb{E}(h(X)|Z)(\omega) = \int_{\mathbb{R}} h(x)f_{X|Z}(x)dx$$

Preuve. Si g est une fonction borélienne bornée, par le théorème de Fubini, on a la suite d'égalités suivantes :

$$\begin{aligned} \mathbb{E}(h(X)g(Z)) &= \int h(x)g(z)f_{X,Z}(x,z)dx dz \\ &= \int_{\mathbb{R}^2} g(z)f_Z(z)f_{X|Z}(x,z)h(x)dx dz \\ &= \int_{\mathbb{R}} g(z)f_Z(z) \left(\int_{\mathbb{R}} f_{X|Z}(x,z)h(x)dx \right) \\ &= \mathbb{E}(g(Z)\theta(Z)) \end{aligned}$$

où $\theta(z) = \int_{\mathbb{R}} f_{X|Z}(x,z)h(x)dx$. On a donc prouvé que

$$\mathbb{E}(h(X)|Z)(\omega) = \theta(Z(\omega)) = \int_{\mathbb{R}} f_{X|Z}(x, Z(\omega))h(x)dx.$$

6.4 Variables discrètes

Proposition 6.4.1 (et définition) Si Z prend un nombre fini ou dénombrable de valeurs $i \in I$, alors, l'espérance conditionnelle de la variable aléatoire X sachant $Z = i$ est définie par

$$\begin{aligned} E(X|Z = i) &= \frac{1}{P(Z = i)} \int_{Z=i} X dP \\ &= \int X dP(.|Z = i). \end{aligned}$$

où pour A mesurable $P(.|A)$ est la probabilité conditionnelle \tilde{P} à A définie par

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

où d'une façon équivalente par pour f mesurable positive ou intégrable :

$$\int f dP(.|A) = \frac{1}{P(A)} \int_A f dP.$$

On en déduit que si X est également discrètes alors, on a :

$$\begin{aligned} \mathbb{E}(X|Z) &= \sum_i \mathbb{I}_{Z=i} \mathbb{E}(X|Z = i) \\ &= \sum_j j \left(\sum_i \mathbb{I}_{Z=i} P(X = j|Z = i) \right), \end{aligned}$$

car

$$\mathbb{E}(X|Z = i) = \sum_j jP(X = j|Z = i).$$

6.5 Variables quelconques

Théorème 6.5.1 *L'espérance conditionnelle de la variable aléatoire X relativement à une autre variable aléatoire Y est*

$$\mathbb{E}(X|Y)(\omega) = \int_{\mathbb{R}} x dP_X(dx|Y = Y(\omega)).$$

Preuve. Il suffit en effet de remarquer que la fonction définie par le membre de droite vérifie les propriétés caractéristiques de l'espérance conditionnelle. L'espérance conditionnelle apparaît donc comme "l'espérance relativement à la loi conditionnelle".

Chapitre 7

Convergence de suites de variables aléatoires

Dans la théorie des probabilités, il existe différentes notions de convergence de variables aléatoires. La convergence (dans un des sens décrits ci-dessous) de suites de variables aléatoires est un concept important de la théorie des probabilités utilisé notamment en statistique et dans l'étude des processus stochastiques. Par exemple, la moyenne de n variables aléatoires indépendantes et identiquement distribuées converge presque sûrement vers l'espérance commune de ces variables aléatoires (si celle-ci existe). Ce résultat est connu sous le nom de loi forte des grands nombres.

Dans tout ce chapitre, les suites de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ sont supposées construites sur un espace de probabilité (Ω, \mathcal{A}, P) . Pour plus de simplicité, nous ne considérons que des variables aléatoires à valeurs réelles. Les énoncés et les résultats subsistent sans modifications pour des vecteurs aléatoires à valeurs dans \mathbb{R}^d (pour l'essentiel, remplacer les valeurs absolues par une norme sur \mathbb{R}^d). Rappelons les inégalités de concentration suivantes :

1. Soit X une variable aléatoire réelle définie sur un espace probabilisé (Ω, \mathcal{F}, P) .
— **Inégalité de Markov** Si $X(\Omega) \subset \mathbb{R}^+$ et l'espérance $m = E(X)$ existe, alors on a

$$\forall \lambda > 0 \quad P[X \geq \lambda] \leq \frac{E(X)}{\lambda}.$$

- **Inégalité de Bienaymé-Tchebychev** Si l'espérance $m = E(X)$ et la variance $\sigma^2 = \text{Var}(X)$ existent, alors on a

$$\forall \epsilon > 0 \quad P[|X - m| \geq \epsilon] \leq \frac{\sigma^2}{\epsilon^2}.$$

- Si $\text{Var}(X) = 0$ alors $P(X = m) = 1$.

7.1 Convergence presque sûre

Un des outils classiques pour montrer la convergence presque sûre est le lemme de Borel-Cantelli. Commençons d'abord par le lemme de Borel-Cantelli pour une suite d'événements $A_n, n \in \mathbb{N}$. Remarquons

$$A = \bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} A_n = \{A_n \text{ a lieu une infinité de fois}\}$$

On abrège souvent l'expression " A_n a lieu une infinité de fois" par " A_n infiniment souvent" ou " A_n i.s.". Remarquer que $P(A_n \text{ i.s.}) = 0$ signifie que presque sûrement seulement un nombre fini d'événements A_n surviennent. C'est-à-dire que pour presque tout $\omega \in \Omega$, il existe un $n(\omega)$ fini tel que si $n \geq n(\omega)$ alors $\omega \notin A_n$, i.e. A_n n'a pas lieu. On fera très attention au fait que l'entier $n(\omega)$ dépend de ω .

Théorème 7.1.1 (Lemme de Borel-Cantelli) Soit $A_n, n \in \mathbb{N}$, une suite d'événements sur un espace probabilisé (Ω, \mathcal{A}, P) .

(i) Si $\sum_{n \in \mathbb{N}} P(A_n) < \infty$, alors $P(A_n \text{ i.s.}) = 0$

(ii) Si les $A_n, n \in \mathbb{N}$ sont indépendantes, et $\sum_{n \in \mathbb{N}} P(A_n) = \infty$, alors $P(A_n \text{ i.s.}) = 1$

Preuve (i) est évidente : pour tout n , on a

$$\{A_n \text{ i.s.}\} = \bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} A_n \subset \bigcup_{n \geq m} A_n,$$

donc

$$P(A_n \text{ i.s.}) \leq \sum_{n \geq m} P(A_n)$$

qui tend vers 0 si la série converge.

ii) En TD.

Définition 7.1.1 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles et X une variable aléatoire réelles définies sur un espace probabilisé (Ω, \mathcal{F}, P) , on dit que (X_n) converge presque sûre vers X , si pour P presque tout ω ,

$$X_n(\omega) \longrightarrow X(\omega), \quad n \rightarrow \infty.$$

Théorème 7.1.2 Loi forte des grands nombres Soit $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur un espace probabilisé (Ω, \mathcal{F}, P) indépendantes, intégrables de même loi d'espérance $m = E(X)$, alors $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ converge presque sûre (p.s.) vers m :

$$\frac{\sum_{i=1}^n X_i}{n} \rightarrow m \text{ p.s.}$$

Observons que l'événement

$$\{\lim_{n \rightarrow \infty} X_n = X\}$$

est bien mesurable puisque égal à

$$\bigcap_{p \geq 1} \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \left\{ |X_n - X| < \frac{1}{p} \right\},$$

ce qui implique que pour tout $p \geq 1$, on

$$P \left(\bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \left\{ |X_n - X| < \frac{1}{p} \right\} \right) = 1,$$

il s'ensuit que X_n converge vers X p.s. si et seulement si pour tout $\varepsilon > 0$, on a

$$P \left(\bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \{ |X_n - X| < \varepsilon \} \right) = 1.$$

Par passage au complémentaire, il vient pour tout $\varepsilon > 0$, on a

$$P\left(\bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} \{|X_n - X| \geq \varepsilon\}\right) = 0. \quad (7.1)$$

ce qui est équivalent à

$$P(|X_n - X| \geq \varepsilon \text{ i.s.}) = 0,$$

Par convergence monotone, X_n converge vers X p.s. est encore équivalent à

$$\lim_{m \rightarrow \infty} P\left(\sup_{n \geq m} \{|X_n - X| \geq \varepsilon\}\right) = 0.$$

La convergence p.s. peut aussi être décrite à l'aide du critère de Cauchy. Par exemple $X_n \rightarrow X$ p.s. si et seulement si $\forall \varepsilon > 0$,

$$P\left(\bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \{|X_n - X_m| < \varepsilon\}\right) = 1.$$

Proposition 7.1.3 (Lemme de Borel-Cantelli) Soient $X_n, n \in \mathbb{N}$, et X , des variables aléatoires réelles définies sur (Ω, \mathcal{A}, P) .

(i) Si pour tout $\varepsilon > 0$,

$$\sum_{n \in \mathbb{N}} P(|X_n - X| \geq \varepsilon) < \infty,$$

alors $X_n \rightarrow X$ p.s.

(ii) Si les $X_n, n \in \mathbb{N}$ sont mutuellement indépendantes, alors $X_n \rightarrow 0$ p.s. si et seulement si pour tout $\varepsilon > 0$,

$$\sum_{n \in \mathbb{N}} P(|X_n| \geq \varepsilon) < \infty.$$

Preuve (i) soit $\varepsilon > 0$ et les événements

$$A_n = \{|X_n - X| \geq \varepsilon\}, n \in \mathbb{N}.$$

Appliquons le Lemme de Borel-Cantelli, il vient

$$P\left(\bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} A_n\right) = P(\{A_n \text{ i.s.}\}) = 0,$$

ce qui fournit le résultat d'après (7.1)

ii) se démontre de façon analogue à partir de la partie indépendante du lemme de Borel-Cantelli. (Noter qu'il convient de supposer X nulle, ou constante, sans quoi les événements A_n ne sont pas nécessairement indépendants.) \square

7.2 Convergence en probabilité

Définition 7.2.1 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles et X une variable aléatoire réelles définies sur un espace probabilisé (Ω, \mathcal{F}, P) , on dit que (X_n) converge en probabilité vers X si

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

Remarque 7.2.1 La convergence presque sûre implique la convergence en probabilité, en effet X_n converge vers X p.s. implique

$$\lim_{m \rightarrow \infty} P(\{|X_m - X| \geq \epsilon\}) \leq \lim_{m \rightarrow \infty} P\left(\bigcup_{n \geq m} \{|X_n - X| \geq \epsilon\}\right) = P\left(\bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} \{|X_n - X| \geq \epsilon\}\right) = 0.$$

Exemple 7.2.1 Montrer que si, pour tout entier $n \geq 1$, X_n est une variable aléatoire réelle de densité $n1_{[0,1/n]}$. alors la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers la variable aléatoire réelle constante 0.

Rép : Soit ϵ et n_0 tels que $1/n_0 < \epsilon$, donc pour $n \geq n_0$ on a $1/n < \epsilon$ ce qui implique

$$P(|X_n - 0| > \epsilon) = P(X_n > \epsilon) = n \int_{\epsilon}^{+\infty} n1_{[0,1/n]}(x) dx = 0 \rightarrow 0.$$

Théorème 7.2.1 Loi faible des grands nombres Soit $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur un espace probabilisé (Ω, \mathcal{F}, P) indépendantes, de même loi, possédant une espérance $m = E(X)$ et une variance $\sigma^2 = Var(X)$, alors $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ converge en probabilité vers m .

Preuve D'après l'inégalité de Bienaymé-Tchebychev, on a

$$\begin{aligned} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - m\right| > \epsilon\right) &\leq \frac{1}{\epsilon^2} V\left(\frac{\sum_{i=1}^n X_i}{n}\right) \\ &= \frac{\sigma^2}{n\epsilon^2} \rightarrow 0. \end{aligned}$$

Théorème 7.2.2 Soient $X_n, n \in \mathbb{N}, X$, des variables aléatoires réelles définies sur (Ω, \mathcal{A}, P) . Alors X_n converge en probabilité vers X :

$$X_n \rightarrow X \text{ en probabilité.}$$

Si et seulement si de toute suite déterministe croissante d'entiers (n') , on peut extraire une sous-suite n'_k telle que

$$X_{n'_k} \rightarrow X \text{ p.s.}$$

Preuve. Nécessité : Il suffit de considérer $(n') = (n)$. Soit $\epsilon > 0$. Pour tout $k \geq 1$, soit $n_k = n_k(\epsilon)$ le plus petit entier tel

$$P(|X_{n_k} - X| \geq \epsilon) \leq 2^{-k}.$$

Alors, (X_{n_k}) est telle que pour tout $\epsilon > 0$,

$$\sum_{k \geq 1} P(|X_{n_k} - X| \geq \epsilon) < \infty.$$

Par le lemme de Borel-Cantelli $X_{n_k} \rightarrow X$ p.s.

Suffisance : Si X_n ne converge pas en proba vers X , alors il existe $\varepsilon > 0$ et $\eta > 0$ pour lequel $P(|X_{n_k} - X| > \varepsilon) > \eta$ le long d'une sous suite n_k . Aucune sous suite extraite de $(X_{n_k})_k$ ne peut converger en proba vers X et donc aucune ne peut converger vers X avec proba 1. \square

La convergence en probabilité est stable par les opérations algébriques usuelles.

Proposition 7.2.3 Soient $(X_n)_{n \in \mathbb{N}}, (Y_n)_{n \in \mathbb{N}}$, deux suites de variables aléatoires réelles définies sur un espace (Ω, \mathcal{A}, P) . Supposons que X_n (resp. Y_n) converge en probabilité vers une variable aléatoire X (resp. Y) définie sur (Ω, \mathcal{A}, P) .

(i) Si φ est une application continue de \mathbb{R} à valeurs dans R , alors

$$\varphi(X_n) \rightarrow \varphi(X), \text{ en probabilité.}$$

(ii) Pour tout $\alpha, \beta \in \mathbb{R}$,

$$\alpha X_n + \beta Y_n \rightarrow \alpha X + \beta Y, \text{ en probabilité.}$$

(iii) De plus,

$$\langle X_n, Y_n \rangle \rightarrow \langle X, Y \rangle, \text{ en probabilité.}$$

Preuve Vérifions par exemple (ii). Soit (n') une suite partielle. On peut extraire une sous-suite (n'') telle que

$$X_{n''} \rightarrow X \text{ p.s.}$$

De (n'') , on peut extraire une sous-suite (n''') tel que

$$Y_{n'''} \rightarrow Y \text{ p.s.}$$

Alors

$$\alpha X_{n'''} + \beta Y_{n'''} \rightarrow \alpha X + \beta Y \text{ p.s.}$$

On conclut à l'aide du Théorème 7.2.2 (Il peut être instructif de démontrer cette proposition sans l'aide du Théorème 7.2.2).

Théorème 7.2.4 (Critère de Cauchy) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur (Ω, \mathcal{A}, P) . Supposons qu'elle vérifie le critère de Cauchy en probabilité, c'est-à-dire que

$$\forall \varepsilon > 0, \exists n_0, \forall n \geq n_0 \quad P(|X_n - X_{n_0}| \geq \varepsilon) \leq \varepsilon,$$

Alors X_n converge en probabilité.

Preuve En TD.

Comme pour les suites usuelles (non aléatoires), l'intérêt du critère de Cauchy et du Théorème 7.2.4 est qu'il assure l'existence d'une limite sans que nous ayons besoin de la calculer explicitement.

7.3 Convergence dans L^p

Nous avons introduit les espaces L^p au chapitre III. Rappelons qu'une variable aléatoire réelle X , définie sur (Ω, \mathcal{A}, P) , est dans $L^p(\Omega, \mathcal{A}, P)$, $p > 0$, si $E(|X|^p)$ est fini. L'espace $L^p(\Omega, \mathcal{A}, P)$ est muni de la norme,

$$\|X\|_p = E(|X|^p)^{\frac{1}{p}},$$

qui en fait un espace complet. En particulier, on peut définir une notion de convergence.

Définition 7.3.1 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles et X une variable aléatoire réelles définies sur un espace probabilisé (Ω, \mathcal{F}, P) . On dit que la suite X_n converge vers la v.a. X dans L^p , $0 < p < +\infty$, si $X_n, X \in L^p$ et si

$$\mathbb{E}|X_n - X|^p \rightarrow_n 0,$$

ou de façon équivalente

$$\|X_n - X\|_p \rightarrow_n 0,$$

Remarque 7.3.1 L'inégalité de Markov montre que pour tout $p > 0$, la convergence dans L^p implique la convergence en probabilité.

7.4 Convergence en loi

Définition 7.4.1 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles et X une variable aléatoire réelles définies sur un espace probabilisé (Ω, \mathcal{F}, P) , on note F_n la fonction de répartition de X_n et F celle de X . on dit que (X_n) converge en loi vers X (ou que les lois P_{X_n} convergent étroitement vers la loi P_X .) si en tout point x où F est continue

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

Dans le cas d'une variable discrète cette définition devient : Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires discrètes et X une variable aléatoire discrètes définies sur un espace probabilisé (Ω, \mathcal{F}, P) , on suppose que $\forall n \in \mathbb{N} X_n(\Omega) \subset X(\Omega) = \{x_i\}_{i \in I}$. On dit que (X_n) converge en loi vers X si

$$\forall i \in N \lim_{n \rightarrow \infty} P(X_n = x_i) = P(X = x_i).$$

Remarque 7.4.1 La convergence en loi ne vérifie pas de bonnes propriétés arithmétiques : si par exemple X_n tend en loi vers X , il n'est pas vrai que $X_n - X$ tend vers 0.

Théorème 7.4.1 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles et X une variable aléatoire réelles définies sur un espace probabilisé (Ω, \mathcal{F}, P) . La suite (X_n) converge en loi vers X si et seulement si

$$E[\phi(X_n)] \rightarrow E[\phi(X)], \quad \forall \phi \text{ continue bornée.}$$

Preuve Il suffit de vérifier le Théorème pour les fonctions indicatrices $1_{]-\infty, x]}$. En effet, la définition de la convergence en loi implique que $E[\phi(X_n)] \rightarrow E[\phi(X)]$ est vérifiée pour les indicatrice $1_{]-\infty, x]}$. En particulier on a le théorème de Lévy qui se restreint sur les fonctions caractéristiques

Théorème 7.4.2 (Théorème de Lévy) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles et X une variable aléatoire réelles définies sur un espace probabilisé (Ω, \mathcal{F}, P) . La suite (X_n) converge en loi vers X si et seulement si

$$\phi_{X_n}(t) \longrightarrow \phi_X(t), \quad \forall t \geq 0.$$

La convergence faible est bien le mode le plus faible de convergence que nous avons introduit jusqu'à présent.

Remarque 7.4.2 (i) Si X_n converge p.s. vers X , alors X_n converge en loi vers X . Cela se déduit par exemple du théorème de convergence dominée et du théorème précédent.

ii) Si X_n converge en probabilité vers X , alors X_n converge en loi vers X . En effet, pour tout $\varepsilon > 0$, et tout t ,

$$F_X(t - \varepsilon) - P(|X_n - X| \geq \varepsilon) \leq F_{X_n}(t) \leq F_X(t + \varepsilon) + P(|X_n - X| \geq \varepsilon),$$

donc si X_n converge en probabilité vers X

$$F_X(t - \varepsilon) + o(1) \leq F_{X_n}(t) \leq F_X(t + \varepsilon) + o(1)$$

et donc

$$F_{X_n}(t) \rightarrow F_X(t).$$

iii) Rappelons que la convergence dans L^p , $p > 0$, entraîne la convergence en probabilité, et donc la convergence en loi.

• **Approximations**

— On approche une loi binomiale $\mathcal{B}(n, p_n)$ avec $\lim_{n \rightarrow \infty} np_n = \lambda > 0$, par une loi de poisson $\mathcal{P}(\lambda)$ dès que $n \geq 30$, $np < 15$ et $p < 0,1$.

Théorème 7.4.3 Théorème de la limite centrée Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur un espace probabilisé (Ω, \mathcal{F}, P) indépendantes, de même loi, possédant une espérance $m = E(X)$ et une variance $\sigma^2 = \text{Var}(X)$, alors $Z_n = \frac{\bar{X}_n - m}{\sigma/\sqrt{n}}$ converge en loi vers une variable aléatoire réelle de loi $\mathcal{N}(0, 1)$.

Preuve : Soit φ_{Z_n} la fonction caractéristique de φ_{Z_n} . Remarquons que pour une variable aléatoire Y d'espérance 0 et de variance 1, la fonction caractéristique de Y admet le développement limité :

$$\varphi_Y(t) = 1 - \frac{t^2}{2} + o(t^2), \quad t \rightarrow 0.$$

Si Y_i vaut $\frac{X_i - \mu}{\sigma}$, alors

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \sum_{i=1}^n \frac{Y_i}{\sqrt{n}}.$$

D'après les propriétés élémentaires des fonctions caractéristiques, la fonction caractéristique de Z_n est

$$\varphi_{Z_n}(t) = \left[\varphi_Y\left(\frac{t}{\sqrt{n}}\right) \right]^n = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right]^n \longrightarrow e^{-t^2/2}.$$

Mais cette limite est la fonction caractéristique de la loi normale centrée réduite $\mathcal{N}(0, 1)$, d'où l'on déduit le théorème central limite grâce au théorème de convergence de Lévy, qui affirme que la convergence simple des fonctions caractéristiques implique la convergence en loi.

• **Approximations**

- On approche une loi binomiale $\mathcal{B}(n, p)$ par une loi normale $\mathcal{N}(np, npq)$ dès que $n \geq 30$, $np > 15$ et $npq > 5$.
- On approche une loi de Poisson $\mathcal{P}(\lambda)$ par une loi normale $\mathcal{N}(\lambda, \lambda)$ dès que $\lambda \geq 15$.