

Notes de cours
de
Probabilités Appliquées

Olivier François

Table des matières

1	Axiomes des probabilités	7
1.1	Introduction	7
1.2	Définitions et notions élémentaires	7
1.2.1	Espace fondamental	7
1.2.2	Opérations sur les événements	8
1.3	Axiomes des probabilités	9
1.3.1	Mesure de probabilité	10
1.3.2	Quelques conséquences des axiomes	11
1.3.3	Passage à la limite monotone	12
1.4	Variables aléatoires	13
1.5	Simulation et probabilités	16
2	Probabilités conditionnelles. Indépendance	19
2.1	Probabilités conditionnelles	19
2.1.1	Définition	19
2.1.2	Quelques propriétés	20
2.2	Indépendance	22
2.2.1	Définition	22
2.2.2	Epreuves indépendantes	24
2.2.3	Variables aléatoires de loi géométrique	25
2.2.4	Variables aléatoires de loi binomiale	26
2.2.5	Probabilité conditionnelle et rejet	27
3	Variables aléatoires réelles	31
3.1	Loi d'une variable aléatoire	31
3.2	Quelques exemples de loi	33
3.2.1	Lois discrètes	33
3.2.2	Loi à densité	38
3.2.3	Exemples de loi	39
3.3	Simulation	41
3.3.1	Méthode par inversion	42
3.3.2	Changement de variable	43
3.3.3	Méthode de conditionnement	44
3.3.4	Superposition de variables aléatoires	45

3.4	Moments et inégalités. Indépendance.	47
3.4.1	Moments	47
3.4.2	Variance	48
3.4.3	Inégalité de Bienaymé-Chebichev	49
3.4.4	Autres inégalités	51
3.5	Transformées	52
3.5.1	Fonctions génératrices	53
3.5.2	Fonctions caractéristiques	58
3.5.3	Transformée de Laplace	60
4	Couples et conditionnement	61
4.1	Couples de variables aléatoires	61
4.1.1	Lois conjointes discrètes	62
4.1.2	Couples de variables à densité	63
4.1.3	Indépendance	65
4.1.4	Variables aléatoires indépendantes	65
4.1.5	Sommes de variables aléatoires indépendantes	67
4.2	Changement de variables	69
4.3	Lois conditionnelles	71
4.3.1	Cas discret	71
4.3.2	Variables à densité	72
4.3.3	Autres cas de lois conditionnelles	73
4.3.4	Espérance conditionnelle	75
4.3.5	Calcul d'espérance par conditionnement	76
4.3.6	Calcul de probabilité par conditionnement	78
4.3.7	Espérance conditionnelle et prédiction	79
4.4	Simulation	81
4.4.1	Couples de variables	81
4.4.2	Méthodes de conditionnement	82
4.4.3	Méthode de rejet	83
5	Annexe : Théorie de la mesure	85
5.1	Introduction	85
5.2	Espaces mesurables et mesures positives	85
5.2.1	Définitions	85
5.2.2	Exemples élémentaires de mesures	86
5.2.3	Construction de mesures. Mesure de Lebesgue.	87
5.3	Applications mesurables	90
5.3.1	Définition	90
5.3.2	Mesure image	91
5.4	Construction d'une intégrale	91
5.4.1	Intégrale d'une fonction indicatrice d'événement	91
5.4.2	Intégrale d'une fonction mesurable positive	92
5.4.3	Fonctions intégrables	92

5.5	Théorème de convergence monotone	93
5.6	Calcul d'intégrales	95
5.6.1	L'intégrale de Lebesgue sur \mathbb{R}	95
5.6.2	Intégrale par rapport à une mesure discrète	95
5.6.3	Mesures définies par des densités	96
5.6.4	L'intégrale d'une fonction mesurable à valeurs dans \mathbb{N}	97
5.7	Théorème de convergence dominée	98
5.8	Mesures produit - Théorème de Fubini	100
5.9	Perspectives probabilistes	102

Chapitre 1

Axiomes des probabilités

1.1 Introduction

Ce chapitre présente une approche intuitive des concepts de probabilité et de variables aléatoires. Toutefois, les notions formelles d'espace fondamental, d'événement et de mesure de probabilité y sont introduites. Les conséquences immédiates des axiomes de définition sont aussi soulignées.

1.2 Définitions et notions élémentaires

1.2.1 Espace fondamental

Considérons une expérience dont l'issue n'est pas entièrement prévisible. Une telle expérience est aussi appelée *épreuve*. Bien que l'issue de l'épreuve en question ne soit pas connue d'avance, nous admettrons que l'ensemble des issues possibles est lui bien connu. Nous le désignons par la lettre grecque Ω . Cet ensemble est appelé *espace fondamental* de l'épreuve. Ses éléments constituent les *éventualités*. Voici des exemples concrets d'espaces fondamentaux. Il faut noter que ces exemples concrets sont rarement utilisés directement pour décrire des phénomènes aléatoires. Comme on le verra par la suite, on leur préférera souvent un espace abstrait contenant toutes les éventualités du hasard et sur lequel seront définies des variables aléatoires.

Exemple 1.2.1 *Exemples.*

Epreuve	Espace fondamental Ω
Lancer d'un dé	$\{1, 2, 3, 4, 5, 6\}$
Durée de vie d'un appareil	\mathbb{R}_+

Un *événement* A est un ensemble d'éventualités correspondant à différents résultats possibles de l'épreuve. Un événement est donc une partie de Ω . Lorsqu'un événement

A contient l'issue de l'épreuve, on dit qu'il est *réalisé*. Par exemple, lors du jet d'un dé à six faces, l'événement

$$A = \text{“ Le résultat est pair ”}$$

est réalisé pour les éventualités $\omega = 2, 4$ ou 6 . Dans le second exemple, l'intervalle $(2, 5)$ correspond à l'événement

$$C = \text{“ L'appareil tombe en panne après un temps d'utilisation compris entre 2 et 5 ”}.$$

Convention. Nous noterons systématiquement les événements par des lettres capitales A, B, C, D, E .

1.2.2 Opérations sur les événements

Diverses opérations de type ensembliste sont tolérées sur les événements.

Union. Pour toute paire d'événements A et B de Ω , nous imposerons au nouvel ensemble $A \cup B$ d'être un événement. En d'autres termes, l'événement $A \cup B$ est réalisé si A ou B l'est.

Complémentaire. Un événement A est susceptible de ne pas se réaliser. Ceci impose à l'ensemble \bar{A} d'être un événement.

En conséquence des deux postulats précédents, nous voyons que les ensembles Ω et \emptyset sont nécessairement des événements. De plus, si A et B sont deux événements, l'intersection $A \cap B$ est aussi un événement, réalisé si et seulement si A et B le sont simultanément.

Union dénombrable. Certaines épreuves sont constituées de répétitions d'expériences aléatoires. Imaginons qu'un joueur parie sur *pile* au jeu de *pile* ou *face* et se lance dans une série indéfiniment longue de parties. On souhaite décrire l'événement

$$E = \text{“ Le joueur perd une partie ”} .$$

Les ensembles

$$E_i = \text{“ Le joueur perd la } i\text{ème partie ”} , \quad i = 1, 2, \dots$$

sont par hypothèse des événements et nous avons

$$E = \bigcup_{i=1}^{\infty} E_i .$$

Nous devons donc imposer qu'une union dénombrable d'événements soit encore un événement.

En conclusion, l'ensemble des événements doit former un système cohérent du point de vue de l'algèbre des différentes opérations décrites ci-dessus. Ce système cohérent porte le nom de *tribu*. Nous en donnons maintenant une définition formelle.

Définition 1.2.1 - Tribu. Un ensemble non vide \mathcal{A} de parties de Ω est appelé tribu si les axiomes suivants sont vérifiés

- 1) $\forall A \in \mathcal{A}, \bar{A} \in \mathcal{A}$.
- 2) Pour toute famille dénombrable $\{A_i ; i \geq 1\}$ d'éléments de \mathcal{A} ,

$$\bigcup_{i \geq 1} A_i \in \mathcal{A}.$$

Exercice 1.2.1 Soient A, B, C trois événements. Trouver l'expression ensembliste des événements suivants

- a) A seul se réalise,
 - b) A et B se réalisent mais C ne se réalise pas,
 - c) un événement au moins se réalise,
 - d) un événement au plus se réalise,
 - e) au moins deux événements se réalisent,
 - f) les trois événements se réalisent,
 - g) aucun événement ne se réalise,
 - h) au plus deux événements se réalisent,
- à partir des événements A, B, C .

1.3 Axiomes des probabilités

Nous avons vu dans le paragraphe précédent que la structure de tribu est à même de modéliser un système cohérent d'événements. Le calcul des probabilités repose sur le principe de modèle. Il nous faut maintenant formuler des hypothèses concernant la probabilité qu'un événement donné se réalise. Un moyen intuitif de définir la probabilité d'un événement serait de le faire en terme de fréquence relative. Une telle définition se formulerait ainsi.

On suppose qu'une épreuve dont l'espace fondamental est Ω est répétée plusieurs fois dans les mêmes conditions. Pour chaque événement A , on définit $n(A)$ comme le nombre de fois où l'événement A se réalise lors des n premières répétitions de l'expérience. La probabilité $P(A)$ de l'événement A est alors définie par

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

Ceci signifie que la probabilité de l'événement A est définie comme la limite des fréquences des réalisations de l'événement A .

Commentaires. Cette définition comporte un inconvénient sérieux. Dans le cas du lancer d'une pièce par exemple, peut-on être sûr que la proportion de *pile* va tendre vers une limite et que cette limite sera indépendante de la suite de lancers effectuée ? Les partisans de cette définition que l'on qualifie de *fréquentiste* répondent en faisant remarquer que la convergence est un **axiome** des probabilités et ne peut être remise en question. Cependant cette hypothèse de modélisation est très complexe à manipuler

et il semble plus raisonnable d'admettre pour définition un système d'axiomes plus simples, intuitivement acceptable pour essayer de démontrer qu'une telle limite existe. Cette démarche fut adoptée par le mathématicien A. Kolmogorov et l'existence de la limite que nous démontrerons dans un chapitre ultérieur porte le nom de *loi des grands nombres*. Nous allons définir la probabilité d'un événement comme une mesure de cet événement.

1.3.1 Mesure de probabilité

Définition 1.3.1 Soit (Ω, \mathcal{A}) un ensemble fondamental muni d'une tribu d'événements. On appelle mesure de probabilité une application P de \mathcal{A} dans $[0, 1]$ telle que

- 1) $P(\Omega) = 1$,
- 2) pour toute suite d'événements mutuellement exclusifs A_1, A_2, \dots (c'est à dire d'événements pour lesquels $A_i \cap A_j = \emptyset$ si $i \neq j$),

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

La probabilité d'un événement est donc un nombre compris entre 0 et 1. Le premier axiome dit qu'on est certain qu'une éventualité de Ω se réalise lors de l'épreuve. Le second signifie que les probabilités d'événements exclusifs s'ajoutent pour obtenir la probabilité que l'un d'eux se réalise.

Convention. On dit souvent que le triplet (Ω, \mathcal{A}, P) est un espace muni d'une structure probabiliste.

Exemple 1.3.1 On lance deux dés équilibrés. On cherche à déterminer la probabilité de l'événement A : "obtenir au moins un six".

Solution. Nous sommes tous capables de déterminer cette probabilité sans faire appel à aucune notion formelle. Cet exercice est donc prétexte à discuter des différents modèles susceptibles d'être introduits pour faire ce calcul.

Modèle 1. Nous supposons que l'on peut identifier les deux dés : l'un d'eux est rouge et l'autre bleu (pour fixer les idées). On observe alors la réalisation d'un couple (i, j) correspondant aux numéros lus sur le premier et le second dé. Dans ce cas, nous pouvons choisir

$$\Omega_1 = \{(i, j), 1 \leq i, j \leq 6\}.$$

L'hypothèse d'uniformité (les dés sont équilibrés) se traduit par l'équiprobabilité des éventualités

$$\forall (i, j) \in \Omega_1, \quad P((i, j)) = \frac{1}{36}.$$

L'ensemble des événements est l'ensemble des parties de Ω_1 . Ainsi, nous avons

$$P(A) = \sum_{i=1}^5 P((i, 6)) + \sum_{j=1}^5 P((6, j)) + P((6, 6)) = \frac{11}{36}.$$

Modèle 2. Nous supposons que l'on ne peut pas identifier les deux dés. On observe alors la réalisation d'une paire $\{i, j\}$ correspondant aux numéros ordonnés. Dans ce cas, nous pouvons choisir

$$\Omega_2 = \{\{i, j\}, 1 \leq i \leq j \leq 6\}.$$

Il n'y a pas d'équiprobabilité des éventualités

$$\forall i < j, \quad P(\{i, j\}) = \frac{1}{18}$$

et

$$P(\{i, i\}) = \frac{1}{36}.$$

L'ensemble des événements est l'ensemble des parties de Ω_2 . Bien entendu, nous avons

$$P(A) = \sum_{i=1}^5 P(\{i, 6\}) + P(\{6, 6\}) = \frac{11}{36}.$$

Modèle 3. Nous supposons que le modèle probabiliste est un modèle abstrait (Ω, \mathcal{A}, P) . Comme dans le premier modèle, nous supposons que les dés sont identifiés et nous notons N_1 et N_2 les numéros obtenus. Nous supposons de plus que les sous-ensembles $(N_1 = i)$ et $(N_2 = j)$ sont des événements quels que soient $1 \leq i, j \leq 6$. Dans ce cas, nous avons

$$P(A) = \sum_{i=1}^5 P(N_1 = i \cap N_2 = 6) + \sum_{j=1}^5 P(N_1 = 6 \cap N_2 = j) + P(N_1 = 6 \cap N_2 = 6).$$

Pour tout $1 \leq i, j \leq 6$,

$$P(N_1 = i \cap N_2 = j) = \frac{1}{36}.$$

Les calculs, bien que formulés différemment, sont donc identiques dans les trois modèles.

1.3.2 Quelques conséquences des axiomes

Proposition 1.3.1 On a $P(\emptyset) = 0$.

Démonstration. Considérons la suite d'événements mutuellement exclusifs A_1, A_2, \dots où $A_1 = \Omega$ et $A_i = \emptyset$ pour tout $i \geq 2$. Par l'axiome 2,

$$P(\Omega) = \sum_{i \geq 1} P(A_i) = P(\Omega) + \sum_{i \geq 2} P(A_i).$$

Ceci implique $P(\emptyset) = 0$. ■

Il découle de cette proposition que l'axiome 2 peut être appliqué à une suite finie A_1, \dots, A_n d'événements mutuellement exclusifs. On a alors

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

Ce résultat est obtenu en prenant $A_i = \emptyset$, pour tout $i \geq n + 1$.

Proposition 1.3.2 *Pour tout $A \in \mathcal{A}$, $P(\bar{A}) = 1 - P(A)$.*

Démonstration. $1 = P(\Omega) = P(A) + P(\bar{A})$. ■

Notre troisième résultat exprime que si l'événement A est contenu dans l'événement B alors la probabilité qu'il se réalise doit être plus petite que celle de B . Notons dans ce cas que la réalisation de A implique celle de B .

Proposition 1.3.3 *Si $A \subset B \in \mathcal{A}$, $P(A) \leq P(B)$.*

Démonstration. On peut écrire

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

Ces deux événements étant mutuellement exclusifs, On tire de l'axiome 2 que

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

Ceci prouve le résultat puisque le second terme du membre de droite est positif. ■

Pour terminer ce paragraphe, nous présentons deux exercices faciles mais fort utiles.

Exercice 1.3.1 *Soit A et B deux événements. Démontrer que*

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

En déduire que,

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

Ce résultat constitue l'inégalité de Bonferroni.

Exercice 1.3.2 *Soit (A_n) une suite d'événements. Montrer par récurrence l'inégalité suivante (inégalité de Boole)*

$$P\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 1} P(A_n).$$

1.3.3 Passage à la limite monotone

Une suite d'événements A_1, A_2, \dots est dite croissante si

$$A_1 \subset A_2 \subset \dots \subset A_n \subset A_{n+1} \subset \dots$$

Elle est dite décroissante si

$$A_1 \supset A_2 \supset \dots \supset A_n \supset A_{n+1} \supset \dots$$

Pour de telles suites, nous obtenons le résultat suivant

Proposition 1.3.4 *Si A_1, A_2, \dots est une suite croissante, alors*

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

Si A_1, A_2, \dots est une suite décroissante, alors

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

Démonstration. Supposons tout d'abord que A_1, A_2, \dots est une suite croissante et définissons les événements B_n , $n \geq 1$, par

$$\begin{aligned} B_1 &= A_1 \\ B_n &= A_n \cap \overline{A_{n-1}} . \end{aligned}$$

Il est facile de voir que les événements B_n sont mutuellement exclusifs. De plus, nous avons

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i \quad \text{et} \quad \bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A_i \quad \forall n \geq 1 .$$

Ainsi,

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} B_i\right) \\ &= \sum_{i=1}^{\infty} P(B_i) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n B_i\right) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right) \\ &= \lim_{n \rightarrow \infty} P(A_n) \end{aligned}$$

puisque les A_i sont emboîtés. La deuxième assertion de la proposition est laissée en exercice au lecteur. ■

1.4 Variables aléatoires

Lors de l'expérience du *lancer d'un dé*, l'espace fondamental que nous avons choisi est celui des résultats du jet

$$\Omega_{\text{dé}} = \{1, 2, 3, 4, 5, 6\} .$$

Une seconde façon de voir les choses consiste à introduire un espace fondamental abstrait Ω . Cet espace abstrait modélise toutes les éventualités du hasard. On ne sait rien de cet ensemble hormis que des événements sont susceptibles de se produire. Ceci revient à supposer que Ω est muni d'une tribu \mathcal{A} . Lorsque l'on jette le dé, une certaine éventualité $\omega \in \Omega$ se réalise et le résultat du jet est une fonction de cette éventualité. Il s'agit d'un numero compris entre 1 et 6 que l'on note $N(\omega)$. A toute éventualité $\omega \in \Omega$ correspond un résultat $N(\omega)$ de l'expérience aléatoire. On définit ainsi une application N de Ω dans $\{1, 2, 3, 4, 5, 6\}$. Le modèle impose que les sous ensembles

$$(N = k) =_{def} \{\omega \in \Omega ; N(\omega) = k\} = N^{-1}(\{k\}) \quad k = 1, \dots, 6$$

soient des événements. Puisque les événements $(N = k)$ sont nécessairement disjoints, la plus petite tribu \mathcal{A}' qui contient tous ces événements est de cardinal fini car elle est constituée de l'ensemble de toutes les unions possibles de ces événements. La tribu \mathcal{A}' contient donc 2^6 événements et le modèle $(\Omega, \mathcal{A}', P)$ est parfaitement bien défini par la donnée des probabilités des événements $(N = k)$. En langage probabiliste, l'application N est une variable aléatoire à valeurs dans $\{1, 2, 3, 4, 5, 6\}$. En attendant une définition plus précise qui sera donnée au chapitre suivant, on appellera *variable aléatoire* toute application définie sur Ω (dont la valeur dépend du hasard). Cette définition heuristique cache des difficultés théoriques. L'exemple de la durée de vie donné au début du chapitre est moins évident à définir de manière précise. On peut noter X la durée de vie de l'appareil en question. On considère que X est une variable aléatoire à valeurs dans \mathbb{R}_+ . Nous pouvons formaliser l'événement C de l'exemple introductif

$$C = (X \in (a, b))$$

et lui affecter une probabilité (par exemple) égale à

$$P(C) = \int_a^b e^{-x} dx .$$

Malheureusement, il n'est pas évident dans ce cas d'explicitier tous les événements dont on pourra mesurer la probabilité. Une bonne définition du modèle nécessite un formalisme que l'on déduira de la théorie de la mesure (chapitre 3)

Convention. Nous noterons toujours les variables aléatoires par des lettres capitales : K, L, M, N pour des variables à valeurs entières (discrètes) et X, Y, Z, T pour des variables à valeurs réelles (continues).

Commentaires. Revenons au lancer du dé. Lorsque l'on choisit le point de vue abstrait et que l'on souhaite décrire complètement l'expérience, on définit la mesure de probabilité P sur (Ω, \mathcal{A}) en posant

$$P(N = k) = 1/6 .$$

Tandis que si l'on choisit $\Omega_{\text{dé}} = \{1, 2, 3, 4, 5, 6\}$, on définit une mesure de probabilité P_N sur $(\Omega_{\text{dé}}, \mathcal{P}(\Omega_{\text{dé}}))$ en posant

$$P_N(\{k\}) = 1/6 .$$

Les mesures de probabilités notées P et P_N ne sont pas définies sur le même ensemble. La mesure P vit sur l'espace abstrait, tandis que P_N vit sur $\{1, 2, 3, 4, 5, 6\}$. Les deux manières de décrire l'expérience conduisent bien entendu à des calculs et des conclusions identiques.

Exemple 1.4.1 *On lance deux dés équilibrés. Calculer la probabilité d'obtenir un numéro pair.*

Solution. Soit $1 \leq i, j \leq 6$. On note N_1 et N_2 les numéros de chacun des deux dés. On remarque que

$$P(N_1 = i \text{ } N_2 = j) = \frac{1}{36} = \frac{1}{6} \frac{1}{6} = P(N_1 = i) P(N_2 = j)$$

Cette propriété très importante traduit en fait l'**indépendance** des deux résultats. Elle permet de calculer

$$P(N_1 \text{ pair } \cap N_2 \text{ pair}) = P(N_1 \text{ pair})P(N_2 \text{ pair}) = \frac{1}{2} \frac{1}{2} = \frac{1}{4}.$$

■

Exemple 1.4.2 *Soit X un nombre positif mesuré à l'issue d'une épreuve aléatoire. On suppose que*

$$\forall 0 \leq a \leq b, \quad P(X \in [a, b]) = \int_a^b e^{-x} dx.$$

Calculer $P(X \geq t)$ pour tout $t \geq 0$. Calculer $P(\sin X \geq 0)$.

Solution. Soit $t \geq 0$. On a

$$P(X \geq t) = \int_t^\infty e^{-x} dx = e^{-t}$$

et l'on vérifie ainsi que $P(X \geq 0) = 1$. Pour calculer la probabilité de l'événement $(\sin X \geq 0)$, il suffit de décomposer ce dernier en union disjointe

$$(\sin X \geq 0) = \bigcup_{n \geq 0} P(X \in [2n\pi, (2n+1)\pi]).$$

La probabilité cherchée est donc égale à

$$\begin{aligned} P(\sin X \geq 0) &= \sum_{n \geq 0} P(X \in [2n\pi, (2n+1)\pi]) \\ &= \sum_{n \geq 0} e^{-2n\pi} - e^{-(2n+1)\pi} \\ &= \frac{1 - e^{-\pi}}{1 - e^{-2\pi}} \end{aligned}$$

■

Convention. Nous dirons qu'un réel U est un *nombre pris au hasard* dans l'intervalle $(0, 1)$ si

$$\forall 0 \leq a \leq b \leq 1, \quad P(U \in [a, b]) = b - a .$$

En d'autres termes, U est une variable aléatoire tirée uniformément dans l'intervalle $(0, 1)$.

1.5 Simulation et probabilités

Une place particulière dans ce cours est réservée à la simulation de variables aléatoires sur un ordinateur. Il s'agit de produire à l'aide d'un algorithme utilisant un générateur de nombres aléatoires une variable qui aura les mêmes caractéristiques probabilistes que la variable aléatoire souhaitée. Nous cherchons aussi à simuler des événements qui se produisent avec une probabilité donnée. Dans ce cours, nous décrivons les algorithmes en pseudo-code, utilisant par exemple les notations

$$X := a \quad \text{ou} \quad X \leftarrow a$$

pour l'instruction

X prend la valeur a .

Nous ne chercherons pas dans ce cours à formuler une théorie de la simulation. Nous procéderons au coup par coup et à l'aide d'exemples. En voici un qui est fondamental.

Exemple 1.5.1 Soit A un événement dont la probabilité est connue et égale à $P(A)$. Nous souhaitons simuler une variable aléatoire X qui prend la valeur 1 avec la probabilité $P(A)$ et la valeur 0 avec la probabilité $1-P(A)$.

Algorithmiquement, la situation se décrit de la manière suivante

```

X := 0
Simuler Epreuve
Si (A est réalisé)
    X := 1
FinSi.
```

Nous supposons tout au long de ce cours que le générateur de nombres aléatoires est une fonction que l'on conviendra de noter *ALEA*. Tout appel à cette fonction retourne un nombre réel U *au hasard* compris entre 0 et 1. Une telle hypothèse se formule de la manière suivante

$$\forall a, b \in [0, 1], \quad a \leq b, \quad P(U \in [a, b]) = \text{longueur}([a, b]) = b - a .$$

Ainsi les “instructions”

```
Simuler Epreuve  
Si (A est réalisé)  
     $X := 1$   
FinSi
```

peuvent être remplacées par

```
 $U := ALEA$   
Si ( $U < P(A)$ )  
     $X := 1$   
FinSi.
```

Lors de la simulation, l'épreuve est constituée d'un appel de la fonction *ALEA* ou si l'on préfère d'un tirage au hasard. La variable *U* est aléatoire. La condition

(*A* est réalisé)

est remplacée par

($U < P(A)$)

qui se réalise avec la même probabilité.

Chapitre 2

Probabilités conditionnelles. Indépendance

Nous allons présenter dans ce chapitre l'un des plus importants concepts de la théorie des probabilités : celui de *probabilité conditionnelle*. L'importance de ce concept est double. En premier lieu, on s'intéresse au calcul des probabilités lorsqu'une partie de l'information concernant le résultat de l'expérience est disponible. Dans une telle situation, les probabilités cherchées sont des probabilités conditionnelles. D'autre part, il est souvent utile d'utiliser un conditionnement pour effectuer le calcul de certaines probabilités. L'utilisation de conditionnement est alors plus un moyen technique pour résoudre un problème qu'une finalité. Ce moyen technique permet d'utiliser pleinement les informations dont on dispose pour un problème donné. Nous reviendrons largement sur cette technique lors du chapitre 5.

2.1 Probabilités conditionnelles

2.1.1 Définition

Voici un exemple introductif au problème du conditionnement.

Exemple 2.1.1 *Nous répétons le lancer d'un dé jusqu'à obtenir un numéro pair. Quelle est alors la probabilité d'obtenir un 2 ?*

Solution. Lors du lancer du dé, l'espace fondamental de l'épreuve est constitué des six éventualités

$$\{1, 2, 3, 4, 5, 6\} .$$

A la suite de l'épreuve décrite ci-dessus, on doit considérer l'espace fondamental restreint

$$\{2, 4, 6\}$$

obtenu en rejetant les éventualités 1, 3, 5. Il n'y a pas de raison de privilégier une éventualité plutôt qu'une autre après rejet. La probabilité de l'événement "le numéro obtenu est 2" est égale à $1/3$ pour cette épreuve. ■

Nous allons donner une définition des probabilités conditionnelles et préciser le lien entre ces probabilités et la procédure de rejet mise en évidence dans l'exemple précédent.

Définition 2.1.1 Soient A et B deux événements tels que $P(B) > 0$. On appelle probabilité conditionnelle de A sachant B la grandeur

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

On vérifie que l'application $P(\cdot | B) : \mathcal{A} \rightarrow [0, 1]$ est une mesure de probabilité sur (Ω, \mathcal{A}) . Tout événement disjoint de B est de probabilité nulle.

La définition (2.1.1) permet de justifier le calcul de l'exemple précédent. En effet, nous avons, dans l'espace fondamental lié à un seul lancer du dé,

$$P(\{2\} | \{2, 4, 6\}) = \frac{1/6}{1/6 + 1/6 + 1/6}$$

et cette dernière grandeur est égale à $1/3$.

Commentaires. Intuitivement, la probabilité conditionnelle $P(A|B)$ correspond à la proportion des éventualités pour lesquelles A est réalisé parmi celles pour lesquelles B est réalisé.

2.1.2 Quelques propriétés

Les résultats suivants ont de nombreuses applications pratiques.

Proposition 2.1.1 Formule des probabilités totales. Soit B_1, B_2, \dots une suite d'événements mutuellement exclusifs tels que

$$\forall n \geq 1, \quad P(B_n) \neq 0 \quad \bigcup_{n \geq 1} B_n = \Omega$$

alors

$$\forall A \in \mathcal{A}, \quad P(A) = \sum_{n \geq 1} P(A|B_n)P(B_n).$$

Démonstration. Elle résulte simplement de l'identité

$$P(A) = \sum_{n \geq 1} P(A \cap B_n)$$

puisque A est union disjointe des événements $A \cap B_n$ (Axiome 2 des probabilités). ■

Exemple 2.1.2 On note N le résultat du lancer d'un dé équilibré à six faces, puis X le résultat du lancer d'un dé équilibré à N faces. Nous souhaitons déterminer les probabilités $P(X = k)$, pour tout $k \in \{1, \dots, 6\}$.

Solution. Nous avons

$$\forall k \in \{1, \dots, 6\}, \quad P(X = k) = \sum_{n=1}^6 P(X = k | N = n)P(N = n) .$$

Or

$$P(X = k | N = n) = \begin{cases} 0 & \text{si } k > n \\ 1/n & \text{sinon.} \end{cases}$$

Finalement

$$P(X = k) = \sum_{n \geq k} \frac{1}{6n} .$$

■

Exemple 2.1.3 On lance deux dés équilibrés à n faces. On note N_i le résultat du lancer pour $i = 1, 2$. Déterminer la probabilité de l'événement $(N_1 > N_2)$.

Solution.

$$P(N_1 > N_2) = \sum_{k=2}^n P(N_1 > N_2 | N_1 = k) P(N_1 = k) .$$

Or, pour $k = 2, \dots, n$,

$$P(N_1 > N_2 | N_1 = k) = P(k > N_2) = \frac{k-1}{n}$$

et

$$P(N_1 > N_2) = \sum_{k=2}^n \frac{k-1}{n^2} = \frac{n-1}{2n} .$$

■

D'autres exemples d'applications de cette proposition seront étudiés en travaux dirigés.

Proposition 2.1.2 Formule de Bayes Soit B_1, B_2, \dots une suite d'événements mutuellement exclusifs tels que

$$\forall n \geq 1, \quad P(B_n) \neq 0 \quad \bigcup_{n \geq 1} B_n = \Omega$$

alors

$$P(B_i | A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(A | B_i)P(B_i)}{\sum_{n \geq 1} P(A | B_n)P(B_n)} .$$

Démonstration. Exercice. ■

Exemple 2.1.4 Un objet a disparu dans une zone déterminée que l'on peut séparer en trois régions de surfaces égales. Notons α_1 la probabilité de ne pas trouver l'objet dans la région 1 bien qu'il s'y trouve effectivement. Quelle est la probabilité que l'objet se trouve dans la région i sachant que les recherches dans la région 1 n'ont rien donné.

Solution. Soit A_i l'événement “l'objet se trouve dans la région i ” et B l'événement “les recherches dans la région 1 n'ont rien donné”. Nous avons *a priori*

$$\forall i = 1, \dots, 3 \quad P(A_i) = 1/3 .$$

D'après la formule de Bayes,

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^3 P(B|A_i)P(A_i)} \\ &= \frac{\alpha_1/3}{\alpha_1/3 + 1/3 + 1/3} \\ &= \frac{\alpha_1}{\alpha_1 + 2} \leq 1/3 \end{aligned}$$

Pour $j = 2, 3$

$$\begin{aligned} P(A_j|B) &= \frac{P(B|A_j)P(A_j)}{P(B)} \\ &= \frac{1}{\alpha_1 + 2} \geq 1/3 \end{aligned}$$

■

On remarquera que la probabilité *a posteriori* (c'est à dire conditionnelle) que l'objet se trouve dans la région j une fois que l'on sait que la fouille de la région 1 n'a rien donné est plus grande pour $j = 2, 3$ que la probabilité *a priori*. Pour la région 1, l'effet est inverse. On remarquera aussi que cette probabilité pour la région 1 est une fonction croissante de α_1 . Ceci est conforme à l'intuition.

2.2 Indépendance

2.2.1 Définition

Les exemples vus jusqu'à présent ont montré qu'une information de type “ B est réalisé” influe en général sur la mesure de probabilité qu'un événement A se réalise. La probabilité conditionnelle $P(A | B)$ de l'événement A sachant B n'est donc pas en général égale à $P(A)$. Dans le cas où

$$P(A | B) = P(A) ,$$

on dira que A est indépendant de B . Il est fréquent de rencontrer des événements indépendants. Considérons, par exemple, les événements A et B décrits par

$A =$ “L'équipe du Brésil gagne la prochaine coupe du monde”

et

$B =$ “Il y a de la vie sur Mars”.

Si l'on interrogeait des personnes au sujet de la relation entre A et B , la plupart d'entre elles accepteraient d'admettre qu'il s'agit d'événements indépendants. Puisque la définition de l'indépendance donnée précédemment est symétrique en A et B , on peut la reformuler de la manière suivante

Définition 2.2.1 Deux événements A et B (A et $B \in \mathcal{A}$) sont dits indépendants si

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

Ils sont dits dépendants lorsqu'ils ne sont pas indépendants.

Exemple 2.2.1 On jette deux pièces et on suppose les quatre résultats possibles équiprobables. On désigne par

$$A = \text{“La première pièce montre pile”}$$

et

$$B = \text{“La deuxième pièce montre face”}.$$

Les deux événements sont indépendants puisque

$$P(A \cap B) = \frac{1}{4} = \frac{1}{2} \frac{1}{2} = P(A)P(B).$$

■

Nous allons démontrer maintenant que si A est indépendant de B , il l'est aussi de \bar{B} .

Proposition 2.2.1 Si A et B sont indépendants, A et \bar{B} le sont aussi.

Démonstration. On peut partitionner A de la manière suivante

$$A = A \cap B \cup A \cap \bar{B}$$

Ainsi, il vient

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap \bar{B}) \\ &= P(A)P(B) + P(A \cap \bar{B}) \end{aligned}$$

d'où

$$\begin{aligned} P(A \cap \bar{B}) &= P(A)(1 - P(B)) \\ &= P(A)(P(\bar{B})) \end{aligned}$$

ce qui établit le résultat. ■

2.2.2 Epreuves indépendantes

Définition 2.2.2 Indépendance mutuelle de n événements. On dit de n événements A_1, \dots, A_n qu'ils sont mutuellement indépendants si pour tout sous-ensemble d'indices $1 \leq i_1 < \dots < i_r \leq n$

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

En d'autres termes, des épreuves sont indépendantes si tout ensemble d'événements issus d'épreuves distinctes est constitué d'événements mutuellement indépendants.

En conséquence, si A, B, C sont trois événements mutuellement indépendants, A sera indépendant de tout événement formé à partir de B et de C . Cette définition est à distinguer de l'indépendance de n événements deux à deux. Des événements deux à deux indépendants ne sont pas forcément indépendants mutuellement. Voici un contre-exemple. On jette deux dés équilibrés. En définissant les événements

$$\begin{aligned} A &= \text{“ La somme vaut 7”} \\ B &= \text{“ Le premier dé donne 4”} \\ C &= \text{“ Le second dé donne 3”} . \end{aligned}$$

On peut vérifier que A est indépendant de B et de C mais n'est pas indépendant de $B \cap C$ (à voir en cours).

Définition 2.2.3 On dit d'une suite qu'elle est constituée d'événements mutuellement indépendants si toute sous-suite extraite finie est constituée d'événements indépendants. On dit d'une suite $\mathcal{E}_1, \mathcal{E}_2, \dots$ qu'elle est constituée d'épreuves indépendantes si, pour toute suite finie d'indices $1 \leq i_1 < \dots < i_n$, les événements $A_{i_1}, A_{i_2}, \dots, A_{i_n}$, (A_{i_r} issu de l'épreuve \mathcal{E}_{i_r}) sont mutuellement indépendants.

Une répétition de jeux de *pile* ou *face* constitue une suite d'épreuves indépendantes. On peut citer nombre d'autres exemples. Ceux qui nous intéressent le plus sont liés à la simulation. Nous considérons lors d'une simulation que chaque appel au générateur aléatoire est une épreuve. La production d'un nombre au hasard par un générateur aléatoire est sensé ne pas influencer sur les nombres générés par la suite. Nous considérerons donc comme un **postulat** l'assertion suivante.

Les appels successifs à un générateur aléatoire constituent une suite d'épreuves indépendantes.

Ainsi les événements relatifs à n appels différents seront mutuellement indépendants, quels que soient les appels en question et ce pour tout entier n .

Dans les deux paragraphes suivants, nous nous intéressons à deux types de variables aléatoires liées à une répétition d'épreuves indépendantes. Tout d'abord, nous nous intéressons au temps d'attente de la première réalisation d'un événement de probabilité donnée. Ensuite, nous regardons le nombre de réalisations du même événement lors de n répétitions indépendantes.

Exemple 2.2.2 On lance 23 dés équilibrés. Calculer la probabilité pour obtenir au moins un 6.

Solution. On note A_{23} l'événement "obtenir au moins un 6". En passant au complémentaire, on obtient

$$P(A_{23}) = 1 - \prod_{i=1}^{23} P(N_i \neq 6) = 1 - \left(\frac{5}{6}\right)^{23}$$

■

2.2.3 Variables aléatoires de loi géométrique

Lors d'une suite de répétitions indépendantes d'une même expérience (suite d'épreuves indépendantes), on s'intéresse au moment où l'événement A se réalise pour la première fois. On suppose de plus

$$P(A) = p > 0.$$

Soit N le rang de l'épreuve où l'événement A se réalise pour la première fois. Il s'agit d'une variable aléatoire à valeurs dans

$$\mathbb{N}^* = \{1, 2, 3, \dots\}.$$

Nous avons

$$\begin{aligned} P(N = 1) &= P(A \text{ se réalise lors de la première épreuve}) = P(A) \\ P(N = k) &= P((A \text{ ne se réalise pas lors de la première épreuve}) \\ &\quad \cap \dots \cap (A \text{ ne se réalise pas lors de la } (k-1)^{\text{e}} \text{ épreuve}) \\ &\quad \cap (A \text{ se réalise lors de la } k^{\text{e}} \text{ épreuve})) \\ &= (1-p)^{k-1} p. \end{aligned}$$

Définition 2.2.4 Soit $0 < p < 1$. On dit qu'une variable N suit une loi géométrique de paramètre p , notée $\mathcal{G}(p)$ si

- a) N est à valeurs dans \mathbb{N}^* ;
- b) $\forall k \geq 1, P(N = k) = (1-p)^{k-1} p$.

Commentaires. Remarquons que

$$P\left(\bigcup_{k \geq 1} (N = k)\right) = \sum_{k \geq 1} P(N = k) = p \sum_{k \geq 1} (1-p)^{k-1} = \frac{p}{1-(1-p)} = 1.$$

les événements $(N = k)$ étant disjoints. On peut donc choisir comme espace fondamental $\Omega = \bigcup_{k \geq 1} (N = k)$ et le munir de la tribu formée des unions d'événements $(N = k)$. La formule précédente signifie aussi qu'un événement de probabilité positive finit toujours par se réaliser lors d'une suite infinie de répétitions d'épreuves indépendantes.

Exercice. Ecrire un algorithme de simulation d'une variable aléatoire de loi géométrique.

2.2.4 Variables aléatoires de loi binomiale

On s'intéresse maintenant au nombre de réalisations d'un événement A de probabilité p lors de la répétition de n épreuves indépendantes. On pose donc, pour $i = 1, \dots, n$,

$$X_i = \begin{cases} 1 & \text{si } A \text{ est réalisé à l'issue de l'épreuve } i \\ 0 & \text{sinon.} \end{cases}$$

On définit alors la variable aléatoire

$$X = \sum_{i=1}^n X_i .$$

Cette variable compte le nombre de fois où A est réalisé au cours des n premières épreuves. Le vecteur (X_1, \dots, X_n) est à valeurs dans l'ensemble $\{0, 1\}^n$. Il y a exactement

$$C_n^k = \frac{n!}{k!(n-k)!}$$

configurations possibles de ce vecteur comportant k fois un 1. Par indépendance, chacune de ces configurations possède une probabilité égale à

$$p^k(1-p)^{n-k} .$$

Définition 2.2.5 *On dit qu'une variable X suit une loi binomiale de paramètres n et p , notée $\mathcal{B}(n, p)$ si*

- a) X est à valeurs dans $\{0, 1, \dots, n\}$;
- b) $\forall 0 \leq k \leq n, P(X = k) = C_n^k p^k(1-p)^{n-k}$.

Remarquons finalement, d'après la formule du binôme, que

$$\sum_{k=0}^n P(N = k) = (p + (1-p))^n = 1 .$$

Exemple 2.2.3 *Au jeux olympiques, un athlète dispose de $n = 5$ essais pour se qualifier en finale d'une épreuve. Lors d'un essai, la probabilité pour qu'il réalise le minimum requis est $p = 0.2$, c'est à dire un sur cinq. Nous souhaitons déterminer la probabilité de l'événement*

$$Q = \text{"l'athlète se qualifie"} .$$

Solution. Nous proposons deux solutions à cet exercice. La première solution utilise une variable aléatoire de loi géométrique et la seconde utilise une variable aléatoire de loi binomiale.

Première solution. Supposons que l'athlète dispose en fait d'un nombre indéfini d'essais et que les n premiers seulement sont pris en compte pour la qualification.

Puisque la probabilité p est strictement positive, l'athlète finira par réaliser le minimum requis lors d'un essai. Notons X le numéro cet essai (qui n'est pas nécessairement qualificatif). La variable X est une variable de loi géométrique de paramètre p et

$$P(Q) = P(X \leq n) = 1 - P(X > n) = 1 - (1 - p)^n .$$

Numériquement, nous obtenons

$$P(Q) = 1 - (0.8)^5 = 0.67232 .$$

Cette réponse peut surprendre l'auteur du raisonnement grossier (et faux) suivant. Puisque l'athlète dispose de cinq essais et possède une chance sur cinq de se qualifier, il se qualifiera avec certitude (probabilité 1). Nous constatons que le résultat exact est assez éloigné de 1.

Seconde solution. Considérons cette fois que l'athlète effectue ses n essais sans se soucier de sa qualification. Il termine sa série d'essais quoiqu'il adviene. Notons X le nombre d'essais pour lesquels le minimum requis est réussi. La variable X est une variable aléatoire de loi binomiale de paramètre n et p . Nous obtenons donc

$$P(Q) = P(X \leq 1) = 1 - P(X = 0) = 1 - (1 - p)^n .$$

■

2.2.5 Probabilité conditionnelle et rejet

Nous traduisons dans ce paragraphe le lien entre procédures de rejet et probabilités conditionnelles. Admettons que l'on puisse répéter indéfiniment une épreuve (le lancer d'un dé, par exemple), de sorte que les épreuves successives soient indépendantes. Pour observer la réalisation d'un événement A (le résultat est 2, par exemple) de probabilité conditionnelle donnée, il faudra rejeter toutes les réalisations pour lesquelles la condition n'est pas réalisée. Il faut retenir, qu'en général, **en sortie** de l'algorithme

Répéter Epreuve
 $X :=$ résultat de l'épreuve
 Jusqu'à (*Condition* réalisée.)

on aura, pour tout ensemble B de résultats possibles de l'épreuve,

$$P(X \in B) = P(\text{"résultat de l'épreuve} \in B \mid \textit{Condition}) .$$

Nous donnons ci-dessous une justification formelle de ce point de vue.

Démonstration. Notons C la condition (il s'agit d'un événement observable lors d'une épreuve donnée) de probabilité non nulle. Nous cherchons à déterminer la mesure de probabilité P^C définie à l'issue de l'algorithme de rejet. Soit A un événement observable lors d'une épreuve donnée. Posons

$$A_n = A \text{ se réalise lors de l'épreuve } n$$

et

$C_n = C$ se réalise lors de l'épreuve n .

Nous avons alors,

$$\begin{aligned}
 P^C(A) &= \sum_{n=1}^{\infty} P(A_n \cap C_n \cap \bar{C}_{n-1} \cap \dots \cap \bar{C}_1) \\
 &= \sum_{n=1}^{\infty} P(A_n \cap C_n) P(\bar{C}_{n-1}) \dots P(\bar{C}_1) \\
 &= P(A \cap C) \sum_{n=1}^{\infty} P(\bar{C})^{n-1} \\
 &= P(A \cap C) \frac{1}{1 - P(\bar{C})} \\
 &= P(A | C).
 \end{aligned}$$

La seconde identité utilise l'indépendance des épreuves et la troisième le fait que les épreuves sont identiques. ■

Exemple 2.2.4 *Considérons la procédure $\text{dé}(n)$ qui simule pour tout entier n le lancer d'un dé à n faces et l'algorithme*

Répéter
 $U := \text{dé}(6)$
Jusqu'à ($U < 4$)
 $X := U$.

Solution. En sortie de cet algorithme, nous obtenons

$$\forall k = 1, 2, 3, \quad P(X = k) = P(U = k | U < 4) = \frac{1/6}{3/6} = 1/3.$$

Ainsi cet algorithme simule le lancer d'un dé à trois faces à partir d'un dé à six faces. ■

L'exemple précédent illustre l'idée de rejet. En conditionnant, on "restreint" l'espace fondamental de l'épreuve. Ce faisant, on change la mesure de probabilité en passant de P à $P(\cdot | \text{Condition})$. En résumé, lorsque l'on voudra en sortie d'une épreuve un événement avec une probabilité conditionnelle donnée, il faudra répéter l'épreuve autant de fois que c'est nécessaire pour voir la condition réalisée. Le passage d'une probabilité P à une probabilité $P(\cdot | \text{Condition})$ consistera du point de vue de la simulation à rejeter toutes les épreuves pour lesquelles la condition n'est pas réalisée.

Les probabilités conditionnelles apparaissent dans de nombreux problèmes pour lesquels il n'est pas toujours simple de mettre en évidence ce que l'on doit calculer. Nous laissons au lecteur méditer l'exemple détaillé ci-dessous.

Exemple 2.2.5 Dans un cachot médiéval, le geôlier informe trois prisonniers que l'un d'entre eux a été choisi au hasard pour être exécuté tandis que les deux autres seront libérés. Le premier des prisonniers demande au geôlier de lui communiquer lequel de ses compagnons sera libre, argumentant que cela ne changera rien pour lui puisqu'il sait déjà que l'un des deux ne sera pas exécuté. Le geôlier lui rétorque que s'il lui donne cette information sa probabilité d'être exécuté passera de $1/3$ à $1/2$ et qu'il n'a donc aucun intérêt à connaître cette information... Que doit penser le prisonnier quant à son propre sort ?

Solution. Voici le point de vue du geôlier. Si l'information n'est pas donnée, l'espace fondamental de l'épreuve (du point de vue du prisonnier) est constitué des trois triplets

$$\{(e, l, l), (l, e, l), (l, l, e)\} .$$

La lettre l (resp. e) en position i signifie “ i est libre (resp. est exécuté) ”. Puisqu'il y a trois éventualités et aucune raison de privilégier l'une par rapport aux autres, la probabilité de l'événement “le premier prisonnier est exécuté” est $1/3$. Le geôlier possède l'information “le deuxième prisonnier est libre” et considère donc que l'on peut rejeter l'éventualité (l, e, l) . Dans ce cas, l'espace fondamental se restreint à

$$\{(e, l, l), (l, l, e)\} .$$

Après rejet, la probabilité de l'événement “le premier prisonnier est exécuté” devient égale à $1/2$. Etant donné le lien entre les probabilités conditionnelles et la procédure de rejet mise en évidence précédemment, nous avons calculé

$$P(\text{“le premier prisonnier est exécuté”} \mid \text{“le deuxième est libre”})$$

et cette dernière grandeur est égale à

$$\frac{P((e, l, l))}{P((e, l, l)) + P((l, l, e))} = \frac{1}{2} .$$

Toutefois, en rejetant l'éventualité (l, e, l) , le geôlier privilégie systématiquement le second prisonnier. Il considère implicitement que le second prisonnier est protégé, que la sentence n'a pas d'effet sur lui et qu'il est libre quoiqu'il arrive. Ce n'est pas le cas en toute généralité.

En fait, le raisonnement ci-dessus ne résout pas réellement le problème du prisonnier. Voici le point de vue du prisonnier. La probabilité conditionnelle cherchée est la probabilité de l'événement “1 est exécuté” sachant “2 est désigné comme libre par le geôlier”. Cela fait une différence avec le calcul précédent... Examinons les différents cas possibles (on considère toujours que le prisonnier 1 demande l'information) et leurs probabilités.

Prisonnier exécuté	Choix du geôlier	Probabilité
1	2	$1/3 \times 1/2 = 1/6$
1	3	$1/6$
2	3	$1/3$
3	2	$1/3$

Dans les deux premier cas, c'est le prisonnier 1 qui est exécuté (avec probabilité $1/3$). Le geôlier a donc deux choix possibles : désigner le prisonnier 2 ou le prisonnier 3. Il n'a pas de raison de privilégier l'un plutôt que l'autre. La probabilité de chacune des 2 premières configurations est donc $1/6$. Pour les deux dernières configurations, le geôlier n'a pas le choix... La probabilité de chacune des deux dernières configurations est $1/3$. La probabilité conditionnelle cherchée est

$$P(“1 est exécuté” \mid “2 est désigné libre”) = \frac{1/6}{1/6 + 1/3} = 1/3.$$

Connaître l'information ne change donc rien pour le prisonnier dans ce cas.

Voici un point de vue “extérieur”. En général, considérons que le geôlier a une préférence dans le choix du prisonnier libre qu'il peut désigner. Quand il a le choix, il désigne 2 comme libre avec la probabilité p et 3 avec la probabilité $1 - p$. Dans ce cas, les probabilités des différents cas possibles sont données par le tableau suivant.

Prisonnier exécuté	Choix du geôlier	Probabilité
1	2	$1/3 \times p = p/3$
1	3	$(1 - p)/3$
2	3	$1/3$
3	2	$1/3$

et, donc

$$P(“1 est exécuté” \mid “2 est désigné libre”) = \frac{p/3}{p/3 + 1/3} = \frac{p}{1 + p}.$$

L'information apportée par le geôlier modifie en général la probabilité qu'a le prisonnier 1 d'être exécuté. Malheureusement, le paramètre p est inconnu du prisonnier. Il est donc impossible pour le prisonnier de quantifier l'information apportée. Notons enfin que la grandeur $p/(1 + p)$ varie entre 0 et $1/2$. Elle prend la valeur $1/3$ lorsque $p = 1/2$ (le geôlier est indifférent). Elle prend la valeur $1/2$ lorsque $p = 1$, c'est à dire lorsque le geôlier désigne systématiquement 2 comme libre lorsque 1 est exécuté. ■

Chapitre 3

Variables aléatoires réelles

L'objectif de toute modélisation est de fournir des prédictions sur l'expérience modélisée. Admettons qu'à toute éventualité ω de l'épreuve, on puisse associer un nombre réel $X(\omega)$. Que peut-on dire alors de la répartition des réalisations de ce nombre lors de répétitions de l'épreuve ? Quelle valeur prend-il en moyenne ? Comme quantifier l'écart à la moyenne ? Ce sont les réponses à ces questions qui font l'objet du présent chapitre.

3.1 Loi d'une variable aléatoire

Les notions nécessaires à la modélisation probabiliste ont été introduites dans le chapitre 3. Rappelons et complétons ces différentes notions. Soit (Ω, \mathcal{A}, P) une espace probabilisé. L'ensemble \mathbb{R} sera muni de sa tribu de Borel $\mathcal{B}(\mathbb{R})$.

Définition 3.1.1 Variable aléatoire. *On appelle variable aléatoire réelle toute application mesurable X de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On dira que la variable X est discrète si elle est à valeurs dans un sous-ensemble dénombrable de \mathbb{R} .*

Par définition, X est une variable aléatoire si

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad (X \in B) =_{def} X^{-1}(B) \in \mathcal{A}.$$

Ainsi, on peut calculer la probabilité de tout ensemble de la forme $(X \in B)$. On définit, par image, une mesure de probabilité notée P_X sur l'espace d'arrivée $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad P_X(B) = P(X \in B) = P(X^{-1}(B)).$$

Cette mesure de probabilité est appelée *loi de la variable X* . Il s'agit de la mesure image de P par l'application X . Comme toute mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, la loi P_X est caractérisée par ses valeurs sur les intervalles. Puisqu'il s'agit d'une mesure de probabilité, elle est aussi caractérisée par sa fonction de répartition que l'on note F_X (*cf. théorème de prolongement*)

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

Proposition 3.1.1 *La fonction de répartition F d'une variable aléatoire est croissante, continue à droite. On a de plus*

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

Démonstration. Nous ne démontrons que la continuité à droite (les autres points se vérifient facilement). Soit $h > 0$. La mesure de l'intervalle $]t, t + h]$ est égale à $F(t + h) - F(t)$. Lorsque h tend vers 0, la famille d'intervalles $]t, t + h]$ "tend" vers l'ensemble \emptyset . Donc, d'après la propriété de monotonie des mesures (cf. chapitre 1 et 3)

$$F(t + h) - F(t) \rightarrow 0 .$$

Ceci démontre la continuité à droite. ■

Définition 3.1.2 *Soit X une variable aléatoire réelle positive ou telle que*

$$\int_{\Omega} |X| dP < \infty \quad (X \text{ est intégrable}).$$

On définit l'espérance mathématique de la variable X comme l'intégrale de X sous la mesure P

$$E[X] = \int_{\Omega} X dP .$$

Commentaires. L'espérance de la variable X représente la valeur moyenne que peut prendre la variable X sous toutes les éventualités de l'épreuve. Bien entendu, on n'effectue aucun calcul d'intégrale sous la mesure P . Tous les calculs se feront à l'aide d'intégrales sur \mathbb{R} ou de sommations discrètes. Ceci se justifiera à l'aide du théorème suivant.

Théorème 3.1.1 Théorème de la mesure image. *Soit X une variable aléatoire réelle et φ une application mesurable de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans lui-même telle que la variable $\varphi(X)$ est intégrable. Alors*

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

Démonstration.

1) Si φ est la fonction indicatrice d'un borélien $B \in \mathcal{B}(\mathbb{R})$

$$\varphi = \mathbb{1}_B \quad B \subset \mathbb{R}$$

la relation s'écrit

$$\int_{\Omega} \mathbb{1}_B(X) dP = \int_{\mathbb{R}} \mathbb{1}_B(x) dP_X(x) .$$

Ceci est équivalent à

$$P(X \in B) = \int_B dP_X(x) .$$

2) Si φ est une fonction étagée

$$\varphi = \sum_i \alpha_i \mathbb{1}_{B_i} \quad B_i \in (\mathbb{R}) \quad \alpha_i \geq 0$$

la relation s'écrit

$$\int_{\Omega} \varphi(X) dP = \sum_i \alpha_i \int_{B_i} dP_X(x) = \int_{\mathbb{R}} \varphi(x) dP_X(x).$$

3) Si φ est une fonction mesurable positive de \mathbb{R} dans \mathbb{R} , nous considérons une suite (u_n) de fonctions étagées positives qui convergent en croissant vers φ . Nous appliquons ensuite le théorème de convergence monotone pour terminer la démonstration. Pour une fonction intégrable, le résultat suit naturellement. ■

Corollaire 3.1.1 *L'espérance mathématique d'une variable réelle X intégrable se calcule de la manière suivante*

$$E[X] = \int_{\mathbb{R}} x dP_X(x).$$

Commentaires. Attention! Pour calculer l'espérance d'une variable aléatoire, la formule du corollaire précédent n'est pas forcément la meilleure solution. En particulier, il n'est pas toujours nécessaire de connaître la loi de la variable pour calculer son espérance. On peut constater ce fait par la formule du théorème de la mesure image.

Nous définissons maintenant la notion d'indépendance de n variables aléatoires. Cette notion sera utile au cours de ce chapitre.

Définition 3.1.3 *On dit que les variables aléatoires X_1, \dots, X_n sont indépendantes si pour tous ensembles $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$, les événements $(X_i \in B_i)$, $i = 1, \dots, n$ sont mutuellement indépendants.*

3.2 Quelques exemples de loi

3.2.1 Lois discrètes

Une variable aléatoire réelle X de loi discrète est par définition une variable aléatoire à valeurs dans une partie finie ou dénombrable de \mathbb{R} . Cet ensemble de valeurs prises par la variable X est souvent égal à $\{0, 1, 2, \dots, n\}$ pour un n entier ou alors à \mathbb{N} tout entier. La loi d'une variable aléatoire à valeurs dans \mathbb{N} se caractérise par la donnée des probabilités $\{p_n ; n \in \mathbb{N}\}$ définies par

$$\forall n \in \mathbb{N}, \quad p_n = P(X = n).$$

Pour tout ensemble borélien B de \mathbb{R} , nous avons

$$P(X \in B) = \sum_{n \in B \cap \mathbb{N}} p_n .$$

Ainsi, la mesure image P_X de P par X n'est autre que la mesure de probabilité discrète

$$P_X = \sum_{n \geq 0} p_n \delta_n .$$

En effet,

$$P_X(B) = \sum_{n \geq 0} p_n \delta_n(B) = \sum_{n \geq 0} p_n \mathbb{1}_B(n) = \sum_{n \in B} p_n .$$

L'espérance de la variable X se calcule donc de la manière suivante

Proposition 3.2.1 *Soit X une variable aléatoire à valeurs dans \mathbb{N} , alors*

$$E[X] = \sum_{n \geq 1} n p_n .$$

Pour toute fonction φ définie sur \mathbb{N} , nous avons

$$E[\varphi(X)] = \sum_{n \geq 0} \varphi(n) p_n .$$

Démonstration. (Nous avons déjà démontré ce résultat lors du chapitre précédent.)

$$\begin{aligned} E[X] &= \int_{\Omega} X dP \\ &= \int_{\mathbb{R}} x dP_X(x) \quad \text{d'après le théorème de la mesure image,} \\ &= \int_{\mathbb{R}} x \sum_{n \geq 0} p_n d\delta_n(x) \\ &= \sum_{n \geq 0} p_n \int_{\mathbb{R}} x d\delta_n(x) \\ &= \sum_{n \geq 1} p_n n . \end{aligned}$$

Démontrer la deuxième assertion est un exercice. ■

Commentaires. On retrouve à travers cette proposition l'intuition de valeur moyenne que contient la notion d'espérance. Ce résultat a déjà été démontré au chapitre 3 sans utiliser explicitement le théorème de la mesure image.

L'exemple suivant est très intéressant car il fournit une alternative au calcul précédent de l'espérance d'une variable aléatoire discrète.

Exemple 3.2.1 Soit (Ω, \mathcal{A}, P) un espace muni d'une structure probabiliste. Soit X une variable aléatoire à valeurs dans \mathbb{N} . Montrer que

$$\int_{\Omega} X \, dP = \sum_{k=0}^{\infty} P(X > k) \quad .$$

Solution. Appliquons le théorème de convergence monotone à la suite (X_n) définie par

$$\forall n \geq 0 \quad X_n = \min(X, n) \quad .$$

Cette suite est positive, croissante et converge vers X . De plus

$$\forall n \geq 0 \quad X_n = \sum_{k=0}^n \mathbb{1}_{X > k}$$

Ainsi, nous obtenons à la limite

$$\lim_{n \rightarrow \infty} \int_{\Omega} X_n \, dP = \int_{\Omega} X \, dP \quad .$$

et, d'autre part

$$\lim_{n \rightarrow \infty} \int_{\Omega} X_n \, dP = \lim_{n \rightarrow \infty} \sum_{k=0}^n P(X > k) = \sum_{k=0}^{\infty} P(X > k) \quad .$$

Nous venons de montrer le résultat. Il est intéressant de noter que la somme diverge si et seulement si X n'est pas intégrable. ■

La fonction de répartition d'une variable à valeurs dans \mathbb{N} se calcule de la manière suivante

$$\forall t \in \mathbb{R}, \quad P(X \leq t) = \sum_{n \leq t} P(X = n) \quad .$$

Bien entendu, cette fonction n'est pas continue. Elle présente un saut à chaque entier. Nous verrons en général qu'un saut de la fonction de répartition correspond à la présence d'une masse de Dirac à l'endroit du saut. Nous donnons quelques exemples de variables aléatoires discrètes. Il s'agit de celles les plus fréquemment rencontrées.

Exemple 3.2.2 Variable certaine.

Supposons qu'une variable X prenne la valeur 0 avec la probabilité 1

$$P(X = 0) = 1 \quad .$$

Ceci traduit l'absence de hasard dans la réalisation de l'épreuve associée à la variable X . La loi de X est la masse de Dirac δ_0 , son espérance est

$$E[X] = E[0] = 0$$

et sa fonction de répartition est la fonction de *Heavyside* (ou échelon)

$$\forall t \in \mathbb{R}, \quad F(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 & \text{si } t \geq 0 \end{cases} \quad .$$

■

Exemple 3.2.3 Loi de Bernoulli

Nous pouvons modéliser le résultat d'un jeu de *pile* ou *face* à l'aide d'une variable aléatoire X à valeurs dans $\{0, 1\}$ (cf. chapitre 1). Nous avons alors

$$P(X = 1) = p \quad \text{où } p \in [0, 1]$$

et

$$P(X = 0) = 1 - p$$

Lorsque $p = 1/2$, on dit que le jeu est équilibré. La loi de la variable X est la mesure de probabilité discrète

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

Elle est appelée *loi de Bernoulli* de paramètre p . On la note $\mathcal{B}(p)$. Son espérance est égale à

$$E[X] = \int_{\Omega} X dP = \int_{(X=1)} 1 dP = P(X = 1) = p .$$

Sa fonction de répartition est la suivante

$$\forall t \in \mathbb{R}, \quad F(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - p & \text{si } 0 \leq t < 1 \\ 1 & \text{si } t \geq 1 . \end{cases}$$

La loi de Bernoulli est aussi la loi de la variable $\mathbb{1}_A$ o A est un événement quelconque de \mathcal{A} . ■

Exemple 3.2.4 Loi géométrique et loi binomiale.

Nous avons rencontré des variables aléatoires de lois géométrique et binomiale au cours du chapitre 2. Rappelons qu'une variable aléatoire X de loi géométrique $\mathcal{G}(p)$ est à valeurs dans \mathbb{N}^* . Elle modélise le temps de sortie d'un événement de probabilité $p > 0$ lors d'une répétition indéfinie d'épreuves indépendantes. Elle se caractérise par

$$\forall n \geq 1 \quad P(X = n) = p(1 - p)^{n-1} .$$

Son espérance se calcule de la manière suivante

$$\begin{aligned} E[X] &= p \sum_{n \geq 1} n(1 - p)^{n-1} \\ &= p \sum_{n \geq 1} -((1 - p)^n)' \\ &= p \left(\sum_{n \geq 1} -(1 - p)^n \right)' \\ &= p \left(\frac{-1}{p} \right)' \\ &= \frac{p}{p^2} = \frac{1}{p} \end{aligned}$$

Nous aurons par la suite un moyen pratique de calculer l'espérance d'une variable discrète X à l'aide de sa fonction génératrice (paragraphe (3.5.1)). La méthode de dérivation utilisée ci-dessus sera systématisée. Il faut noter que lors du calcul précédent nous avons dérivé sous le signe *somme*. Il est bon à titre d'application des théorèmes du chapitre 3 de s'exercer à justifier les différentes opérations de dérivation abusivement effectuées lors du calcul précédent.

Rappelons qu'une variable aléatoire X de loi binomiale $\mathcal{B}(n, p)$ est à valeurs dans $\{0, 1, 2, \dots, n\}$. Elle modélise le nombre de réalisations d'un événement de probabilité $p > 0$ donnée lors d'une répétition de n épreuves indépendantes

$$\forall 0 \leq k \leq n, \quad \mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}.$$

Son espérance vaut

$$E[X] = np.$$

Ce résultat ne nécessite pas de calculs compliqués. En effet, X peut s'écrire comme une somme de n variables de Bernoulli de paramètre p (*cf.* chapitre 2)

$$X = \sum_{i=1}^n X_i.$$

D'où

$$E[X] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = np.$$

par linéarité de l'intégrale. ■

Exemple 3.2.5 Loi de Poisson.

Une variable aléatoire X à valeurs dans \mathbb{N} suit la *loi de Poisson* de paramètre $\lambda > 0$ si

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

La notation pour la loi de Poisson est $\mathcal{P}(\lambda)$. L'espérance d'une variable aléatoire de loi de Poisson est

$$E[X] = \lambda.$$

Ce calcul de cette espérance sera détaillé au paragraphe (3.5.1). Nous verrons dans la suite du cours que la loi de Poisson modélise le nombre de clients arrivés à un serveur durant une unité de temps, sous l'hypothèse selon laquelle les arrivées sont *au hasard* et indépendantes les unes des autres. Nous reviendrons sur cette modélisation lors du chapitre 8 (début avril en principe). ■

3.2.2 Loi à densité

Après avoir étudié des variables à valeurs entières, nous nous intéressons à des variables dites *continues* par rapport à la mesure de Lebesgue.

Définition 3.2.1 *On appelle densité de probabilité une fonction mesurable positive f définie sur \mathbb{R} dont l'intégrale par rapport à la mesure de Lebesgue vaut 1*

$$\int_{\mathbb{R}} f(x) d\lambda(x) = \int_{-\infty}^{+\infty} f(x) dx = 1 .$$

On dit qu'une variable aléatoire réelle X admet pour densité f si

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad P_X(B) = \int_B f(x) dx .$$

Commentaires. Cette terminologie est abusive. Si l'on se réfère au chapitre 3, c'est en fait la mesure P_X qui admet une densité. Dans le langage probabiliste, les variables sont souvent assimilées à leur loi. Il faut prendre garde de ne pas faire de confusion.

L'espérance d'une variable aléatoire de densité f se calcule de la manière suivante.

Proposition 3.2.2 *Supposons X positive ou intégrable, alors*

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

Soit g une application mesurable de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans lui-même telle que la variable $g(X)$ est intégrable. Alors

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

Démonstration. On ne démontre que la deuxième assertion de cette proposition. La première est un cas particulier de la seconde. Nous avons, d'après le théorème de la mesure image

$$E[g(X)] = \int_{\mathbb{R}} g(x) dP_X(x) .$$

Lorsque g est l'indicatrice d'un borélien B de \mathbb{R} , $g = \mathbb{1}_B$, nous avons

$$\int_{\mathbb{R}} \mathbb{1}_B(x) dP_X(x) = P(X \in B) = \int_{\mathbb{R}} \mathbb{1}_B(x) f(x) dx .$$

Lorsque g est une fonction étagée positive de \mathbb{R} , $g = \sum_i \alpha_i \mathbb{1}_{B_i}$, nous avons

$$\begin{aligned} \int_{\mathbb{R}} g(x) dP_X(x) &= \sum_i \alpha_i P(X \in B_i) \\ &= \sum_i \alpha_i \int_{\mathbb{R}} \mathbb{1}_{B_i}(x) f(x) dx \\ &= \int_{\mathbb{R}} \sum_i \alpha_i \mathbb{1}_{B_i}(x) f(x) dx \end{aligned}$$

La proposition est donc démontrée pour les fonctions étagées positives. Par le procédé de limite croissante, on montre qu'elle l'est sous les hypothèses de la proposition. ■

Définition 3.2.2 La fonction de répartition d'une variable aléatoire réelle X de densité f est la fonction F_X définie par

$$\forall t \in \mathbb{R}, \quad F_X(t) = P(X \leq t) = \int_{-\infty}^t f(x) dx .$$

Commentaires. La fonction de répartition est la primitive de la densité qui s'annule en $-\infty$. Les propriétés suivantes sont immédiates.

- 1) $\forall a \leq b, \quad P(X \in [a, b]) = F(b) - F(a)$.
- 2) F est continue.
- 3) En tout point t où la densité f est continue, F est dérivable et nous avons

$$F'(t) = f(t) .$$

3.2.3 Exemples de loi

Quelques exemples élémentaires de loi à densité sont à connaître.

Exemple 3.2.6 Loi uniforme sur l'intervalle $[a, b]$, $a < b$.

Une variable aléatoire X suit la *loi uniforme* sur l'intervalle $[a, b]$, $a < b$, si sa densité est

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

La loi uniforme sur l'intervalle $[a, b]$ est notée $\mathcal{U}(a, b)$. L'espérance d'une variable aléatoire de loi $\mathcal{U}(a, b)$ est

$$E[X] = \frac{1}{(b-a)} \int_a^b x dx = \frac{a+b}{2} .$$

La fonction de répartition d'une variable aléatoire de loi $\mathcal{U}(a, b)$ est

$$\forall t \in \mathbb{R}, \quad F(t) = \begin{cases} 0 & \text{si } t \leq a ; \\ (t-a)/(b-a) & \text{si } a \leq t \leq b ; \\ 1 & \text{si } t \geq b . \end{cases}$$

■

Exemple 3.2.7 Loi exponentielle de paramètre $\lambda > 0$.

Une variable aléatoire X suit la *loi exponentielle* de paramètre $\lambda > 0$ si sa densité est donnée par

$$\forall x \in \mathbb{R}, \quad f(x) = \lambda \exp(-\lambda x) \mathbb{1}_{\mathbb{R}_+}(x) .$$

La loi exponentielle est notée $\mathcal{E}(\lambda)$. Il s'agit du modèle le plus simple de la durée de vie d'un matériel. Nous reviendrons en détail sur l'intérêt de cette loi dans la modélisation

des processus à mémoire courte (chapitre 8). L'espérance d'une variable aléatoire de loi $\mathcal{E}(\lambda)$ est

$$E[X] = \lambda \int_{\mathbb{R}_+} x e^{-\lambda x} dx = \frac{1}{\lambda}.$$

La fonction de répartition d'une variable aléatoire de loi $\mathcal{E}(\lambda)$ est

$$\forall t \in \mathbb{R}, \quad F(t) = \begin{cases} 0 & \text{si } t \leq 0; \\ 1 - \exp(-\lambda t) & \text{si } t \geq 0. \end{cases}$$

■

Exemple 3.2.8 Loi normale.

Une variable aléatoire X suit la *loi normale* $\mathcal{N}(0, 1)$ si sa densité est donnée par

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2).$$

La loi normale est une loi utile au statisticien. Mise en évidence par le mathématicien F. Gauss, elle apparaît dans un grand nombre de phénomènes aléatoires. Nous démontrerons dans la suite de ce cours (chapitre 7) un théorème de convergence vers cette loi qui justifiera son importance dans les applications en statistique. L'espérance d'une variable aléatoire $\mathcal{N}(0, 1)$ est

$$E[X] = \int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx = 0.$$

par parité. La fonction de répartition d'une variable aléatoire de loi normale ne connaît pas d'expression analytique explicite. Toutefois des tables très précises des valeurs de cette intégrale sont disponibles. ■

Exemple 3.2.9 Loi exotique.

La plupart des variables aléatoires que nous rencontrerons seront à densité continue par morceaux sur \mathbb{R} et leurs fonctions de répartition seront continues. Toutefois, il est possible de rencontrer des variables aléatoires dont les lois peuvent être nettement plus *exotiques*. Considérons par exemple la mesure de probabilité μ définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ par

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad \mu(B) = \frac{1}{2} \delta_0(B) + \frac{1}{2} \int_{[0,1] \cap B} d\lambda(x).$$

Une variable aléatoire X de loi μ satisfera

$$P(X = 0) = \frac{1}{2}$$

et

$$P(0 < X \leq 1) = \frac{1}{2}.$$

La loi d'une telle variable aléatoire n'admet pas de densité par rapport à la mesure de Lebesgue. La fonction de répartition est donnée par

$$\forall t \in \mathbb{R}, \quad F(t) = \begin{cases} 0 & \text{si } t < 0; \\ (1+t)/2 & \text{si } 0 \leq t \leq 1; \\ 1 & \text{si } t \geq 1. \end{cases}$$

La discontinuité en 0 de la fonction F correspond à la présence de la masse de Dirac en 0.

Exercice 3.2.1 Soit U une variable aléatoire de loi uniforme sur $(0, 1)$. Déterminer la loi des variables $X = -\ln(U)/\lambda$, $\lambda > 0$ et $Y = U^2$. Calculer les espérances des variables X et Y .

Nous terminons ce paragraphe par un résultat utile dans bien des exercices.

Proposition 3.2.3 Soit X une variable aléatoire réelle positive de fonction de répartition F . Alors

$$E[X] = \int_0^\infty 1 - F(t) dt.$$

Démonstration. Notons que l'on peut écrire

$$X = \int_0^X dt = \int_{(X>t)} dt = \int_0^\infty \mathbb{1}_{(X>t)} dt$$

Ainsi, nous avons

$$E[X] = \int_\Omega \int_0^\infty \mathbb{1}_{(X>t)} dt dP$$

et, d'après le théorème de Fubini,

$$E[X] = \int_0^\infty \int_\Omega \mathbb{1}_{(X>t)} dP dt = \int_0^\infty P(X > t) dt.$$

■

3.3 Simulation

Simuler une variable aléatoire X de loi donnée consiste à écrire un algorithme tel que l'une des variables en sortie de cet algorithme possède la même loi que X .

Convention. Nous postulons qu'une succession d'appels du générateur *ALEA* peut être modélisée par une suite de variables aléatoires U_1, U_2, \dots indépendantes et de loi $\mathcal{U}(0, 1)$.

3.3.1 Méthode par inversion

La méthode par inversion est une méthode de simulation conceptuellement très simple. Elle nécessite d'inverser la fonction de répartition F de la variable à simuler. Puisque cette fonction de répartition est croissante, son inverse au sens large se définit de la manière suivante

$$F^{-1}(x) = \inf\{t \in \mathbb{R} ; F(t) \geq x\} .$$

Théorème 3.3.1 Soit U une variable de loi $\mathcal{U}(0, 1)$ et F une fonction de répartition quelconque. La variable aléatoire X définie par

$$X = F^{-1}(U)$$

a pour fonction de répartition F .

Démonstration. Soit $t \in \mathbb{R}$,

$$\begin{aligned} \mathbb{P}(X \leq t) &= \mathbb{P}(F^{-1}(U) \leq t) \\ &= \mathbb{P}(U \leq F(t)) \\ &= F(t) . \end{aligned}$$

■

Lorsque F^{-1} à une expression analytique simple, l'algorithme de simulation par inversion prend la forme suivante

$$U := ALEA ; X := F^{-1}(U) .$$

Pour une variable aléatoire discrète à valeurs dans $\{1, 2, \dots, n\}$, la fonction de répartition s'exprime de la manière suivante.

$$\forall t \in \mathbb{R}, \quad F(t) = \begin{cases} 0 & \text{si } t < 1 \\ F(k) = p_1 + \dots + p_k & \text{si } k \leq t < (k + 1) \\ 1 & \text{si } t \geq n . \end{cases}$$

L'algorithme de simulation par inversion est donc l'algorithme suivant

```

X := 0
U := ALEA
Répéter
    X := X + 1
Jusqu'à (F(X) > U)

```

En sortie de cet algorithme, nous avons bien

$$\forall k = 1, \dots, n, \quad \mathbb{P}(X = k) = \mathbb{P}(F(k - 1) \leq U < F(k)) = F(k) - F(k - 1) = p_k .$$

Il faut noter toutefois que le calcul de l'inverse de F nécessite parfois des opérations coûteuses du point de vue de la machine. En pratique, la simulation aléatoire est rarement utilisée de manière ponctuelle. Il s'agit presque toujours d'utilisation intense. Un code de calcul fera certainement un grand nombre d'appels à de tels algorithmes. Ainsi, il est bon d'avoir à disposition une panoplies de méthodes de simulation utilisables dans des circonstances différentes avec pour objectif de réduire le temps de calcul. Nous développons dans les paragraphes suivants d'autres méthodes de simulation.

3.3.2 Changement de variable

On peut utiliser des transformations de variables aléatoires connues pour générer de nouvelles variables. Toutefois, il faut faire attention de choisir des transformations inversibles.

Proposition 3.3.1 *Soit X une variable aléatoire réelle admettant une densité f_X et φ une application dérivable strictement croissante de \mathbb{R} dans \mathbb{R} . Alors, la variable aléatoire*

$$Y = \varphi(X)$$

admet pour densité la fonction f_Y définie par

$$\forall y \in \mathbb{R}, \quad f_Y(y) = \frac{1}{\varphi'(\varphi^{-1}(y))} f_X(\varphi^{-1}(y)).$$

Commentaires. On reconnaît le terme qui intervient dans la formule de changement de variable pour l'intégrale de Riemann.

Démonstration. Soit $y \in \mathbb{R}$,

$$P(Y \leq y) = P(\varphi(X) \leq y) = P(X \leq \varphi^{-1}(y)) = F_X(\varphi^{-1}(y)).$$

En dérivant, on obtient la formule souhaitée pour la densité de la variable Y . ■

Exemple 3.3.1 *Soit U une variable aléatoire de loi $\mathcal{U}(0, 1)$. La variable*

$$Y = a + (b - a)U, \quad a < b$$

suit la loi $\mathcal{U}(a, b)$.

Solution. Nous avons

$$\forall y \in \mathbb{R}; \quad f_Y(y) = \frac{1}{b - a} f_U\left(\frac{y - a}{b - a}\right) = \frac{1}{b - a} \mathbb{1}_{[a, b]}(y).$$

■

Exemple 3.3.2 Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$. La variable aléatoire $Y = \sigma X + m$, $m \in \mathbb{R}$, $\sigma > 0$ admet pour densité

$$\forall y \in \mathbb{R}, \quad f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-(y - m)^2/2\sigma^2).$$

Il s'agit de la loi $\mathcal{N}(m, \sigma^2)$.

Solution. dans cet exemple, nous avons $\varphi(x) = \sigma x + m$. Ainsi

$$\varphi^{-1}(y) = \frac{y - m}{\sigma}$$

et

$$[\varphi'(\varphi^{-1}(y))]^{-1} = 1/\sigma.$$

■

3.3.3 Méthode de conditionnement

La méthode de conditionnement est essentiellement fondée sur l'idée de rejet. Avant de décrire cette méthode, revenons sur les notions vues au chapitre 2. Soit (Ω, \mathcal{A}, P) un espace muni d'une mesure de probabilité que l'on considère comme modèle probabiliste pour une certaine épreuve. Soit A un événement de probabilité positive. Supposer l'événement A réalisé lors de l'épreuve revient à remplacer l'espace (Ω, \mathcal{A}, P) par $(\Omega, \mathcal{A}, P(\cdot | A))$ où $P(\cdot | A)$ désigne la mesure de probabilité conditionnelle sachant A .

De nombreux algorithmes de simulation peuvent être obtenus en imposant à une variable aléatoire X de loi donnée P_X de satisfaire une condition A . L'algorithme en question prendra la forme suivante (cf. chapitre 2)

Répéter
 Simuler X selon la loi P_X
 Jusqu'à (A réalisé)
 $Y := X$

La variable obtenue en sortie de cet algorithme admettra pour loi, la mesure de probabilité P_X^A définie de la manière suivante

$$\forall t \in \mathbb{R}, \quad P(Y \leq t) = P_X^A([-\infty, t]) = P(X \leq t | A).$$

Exemple 3.3.3 Simuler la loi uniforme sur l'intervalle $[0, 1/2]$ en tirant un nombre au hasard sur $[0, 1]$ et en rejetant ce nombre lorsque sa valeur est supérieure à $1/2$ (méthode de rejet).

Solution.

Répéter
 $U := ALEA$
 Jusqu'à ($U < 0.5$)
 $X := U$

En sortie d'un tel algorithme, la fonction de répartition de X est

$$\forall t \in \mathbb{R}, \quad P(X \leq t) = P(U \leq t \mid U < 0.5) = 2P(U \leq t \cap U < 0.5).$$

Pour les différentes valeurs de $t \in \mathbb{R}$, nous avons

$$F(t) = \begin{cases} 0 & \text{si } t \leq 0; \\ 2t & \text{si } 0 \leq t \leq 1/2; \\ 1 & \text{sinon.} \end{cases}$$

Il s'agit bien de la fonction de répartition de la loi $\mathcal{U}(0, 1/2)$. ■

3.3.4 Superposition de variables aléatoires

Nous présentons une dernière technique de simulation. L'idée consiste à superposer des variables aléatoires de densités différentes. Soient X_1, \dots, X_n des variables aléatoires de densités respectives f_1, \dots, f_n et N une variable aléatoire discrète à valeurs dans $\{1, \dots, n\}$ indépendante des X_i telle que

$$\forall k = 1, \dots, n, \quad P(N = k) = p_k > 0.$$

En sortie de l'algorithme

Simuler N
 Si ($N = k$)
 Simuler X_k de densité f_k
 $X := X_k$
 FinSi.

La variable X admet pour densité la fonction f définie par combinaison linéaire convexe des densités f_k

$$\forall x \in \mathbb{R}, \quad f(x) = \sum_{k=1}^n p_k f_k(x)$$

En effet, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} P(X \leq t) &= \sum_{k=1}^n P(X \leq t \mid N = k) p_k && \text{d'après la formule des probabilités totales,} \\ &= \sum_{k=1}^n P(X_k \leq t) p_k && \text{par indépendance de } N. \end{aligned}$$

Le résultat précédemment annoncé s'obtient alors par dérivation.

Exemple 3.3.4 On cherche à simuler une variable aléatoire de loi de densité

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{4}(4x + 1)\mathbb{1}_{[0,1]}(x) + \frac{1}{4}\mathbb{1}_{[1,2]}(x).$$

Solution. Nous pouvons écrire

$$\begin{aligned} f(x) &= \frac{1}{2} \frac{1}{2} \mathbb{1}_{[0,2]}(x) + \frac{1}{2} 2x \mathbb{1}_{[0,1]}(x) \\ &= \frac{1}{2} f_1(x) + \frac{1}{2} f_2(x) \end{aligned}$$

Compte-tenu du fait que f_1 est la densité d'une variable de loi $\mathcal{U}(0, 2)$ et que f_2 est la densité du maximum de deux variables indépendantes de loi $\mathcal{U}(0, 1)$, nous pouvons écrire l'algorithme de simulation suivant

```

Si (ALEA < 0.5)
    X := 2 * ALEA
Sinon
    X := max(ALEA, ALEA)
FinSi.
```

■

Bien entendu, la méthode se généralise à des lois définies comme combinaison convexe infinie.

Exemple 3.3.5 Soit $0 < p < 1$. On cherche à simuler une variable aléatoire de loi de densité

$$\forall x \in \mathbb{R}_+, \quad f(x) = \frac{pe^{-x}}{(1 - (1 - p)e^{-x})^2}.$$

Solution. Ecrivons l'algorithme suivant, en admettant que la fonction $\text{Geom}(p)$ retourne une variable discrète de loi $\mathcal{G}(p)$ (chapitre 2).

```

N := Geom(p)
X := -ln(ALEA)/N.
```

Soit $x > 0$ et f_X la densité de la variable en sortie de l'algorithme

$$\begin{aligned} f_X(x) &= \sum_{n=1}^{\infty} \mathbb{P}(N = n) f_n(x) \\ &= \sum_{n=1}^{\infty} p(1 - p)^{n-1} n e^{-nx} \\ &= p e^{-x} \sum_{n=1}^{\infty} n(1 - p)^{n-1} e^{-(n-1)x} \\ &= \frac{p e^{-x}}{(1 - (1 - p)e^{-x})^2} \end{aligned}$$

■

3.4 Moments et inégalités. Indépendance.

3.4.1 Moments

Définition 3.4.1 Soit X une variable aléatoire réelle et k un entier tel que

$$E[|X|^k] < \infty .$$

On appelle moment d'ordre k de X l'espérance $E[X^k]$.

Bien entendu, l'espérance d'une variable aléatoire est égale à son moment d'ordre 1. La terminologie de "moment" provient d'une analogie avec la mécanique. La distribution de probabilité peut être vue comme la distribution de masse d'un solide après normalisation à 1. Le moment d'ordre 1 est alors l'équivalent du barycentre des masses. Voici deux propriétés de l'espérance

Proposition 3.4.1 1) Soit X et Y deux variables aléatoires intégrables et a, b deux réels quelconques. Alors, la variable

$$Z = aX + bY$$

est intégrable et

$$E[Z] = aE[X] + bE[Y] .$$

2) Si de plus X et Y sont indépendantes, alors

$$E[XY] = E[X]E[Y] .$$

Démonstration. L'assertion 1) est évidente. Il s'agit de la traduction de la linéarité de l'intégrale. Le fait que Z soit intégrable provient de l'inégalité triangulaire. L'assertion 2) résulte de la définition de l'indépendance lorsque X et Y sont les fonctions indicatrices des événements A et B . Montrons que cette identité est encore vraie lorsque X et Y sont étagées

$$X = \sum_{i=1}^m \alpha_i \mathbb{1}_{A_i} \quad ; \quad Y = \sum_{j=1}^n \beta_j \mathbb{1}_{B_j} .$$

On a dans ce cas

$$E[XY] = \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j E[\mathbb{1}_{A_i} \mathbb{1}_{B_j}] .$$

Or,

$$A_i = (X = \alpha_i) \quad \text{et} \quad B_j = (Y = \beta_j)$$

sont, par définition, des événements indépendants. Ainsi, on a

$$E[\mathbb{1}_{A_i} \mathbb{1}_{B_j}] = P(A_i \cap B_j) = P(A_i)P(B_j) = E[\mathbb{1}_{A_i}]E[\mathbb{1}_{B_j}] .$$

En réordonnant les sommations, nous obtenons

$$E[XY] = E[X]E[Y] .$$

Pour démontrer le résultat en général, il suffit de considérer des suites de fonctions étagées qui convergent en croissant vers les parties positives et négatives de X et Y . ■

Il est important de noter que l'indépendance n'est qu'une condition suffisante à l'assertion 2) du précédent théorème. Il est facile de donner des contre-exemples à la réciproque de ce résultat (cf. TDs). En revanche, nous avons

Théorème 3.4.1 *Soient X et Y deux variables aléatoires réelles. Elles sont indépendantes si et seulement si l'identité*

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$

est satisfaite pour toutes fonctions f et g de \mathbb{R} continues bornées.

Démonstration. Admis. ■

Exemple 3.4.1 *Soient U et V deux variables de loi uniforme indépendantes. Alors*

$$E[U(1 + V)] = 1/2 \cdot 3/2 = 3/4 .$$

3.4.2 Variance

Définition 3.4.2 *Soit X une variable aléatoire dont le carré est intégrable*

$$E[X^2] < \infty .$$

On appelle variance de X et on note $Var(X)$ la grandeur

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

Commentaires. La variance est bien définie. Nous anticipons en affirmant qu'une variable aléatoire de carré intégrable est intégrable. Ceci découlera de l'inégalité de Cauchy-Schwarz que nous établirons un peu plus loin. La variance représente l'écart quadratique moyen à la moyenne d'une variable aléatoire. En mécanique, on parlerait de moment d'inertie.

Proposition 3.4.2 *Soit X une variable aléatoire de carré intégrable. Nous avons*

- a) $Var(X) = E[X^2] - E[X]^2$;
- b) pour tout réel a , $Var(X + a) = Var(X)$ et $Var(aX) = a^2Var(X)$,
- c) si Y est une variable aléatoire de carré intégrable et indépendante de X

$$Var(X + Y) = Var(X) + Var(Y) .$$

Démonstration. Il s'agit de résultats immédiats. Pour c), nous avons

$$Var(X + Y) = E[(X + Y)^2] - (E[X] + E[Y])^2 .$$

Après développement et regroupement des termes de cette expression, nous obtenons

$$Var(X + Y) = Var(X) + Var(Y) + 2(E[XY] - E[X]E[Y]) .$$

Par indépendance, le terme $E[XY] - E[X]E[Y]$ (qui sera noté $Cov(X, Y)$ par la suite) est nul. ■

Exemple 3.4.2 *Montrer que la variance d'un variable X de loi de Bernoulli de paramètre p est*

$$\text{Var}(X) = p(1 - p) .$$

Solution. Nous avons

$$E[X^2] = E[X] = P(X = 1) = p .$$

Le résultat est immédiat. ■

Exemple 3.4.3 *Montrer que la variance d'un variable X de loi exponentielle de paramètre $\lambda > 0$ est*

$$\text{Var}(X) = \frac{1}{\lambda^2} .$$

Solution. Nous avons $E[X] = 1/\lambda$ et, en utilisant une intégration par parties,

$$E[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda} E[X] = \frac{2}{\lambda^2} .$$

■

Exemple 3.4.4 *Appliquons c) pour calculer la variance d'une variable de loi binomiale $\mathcal{B}(n, p)$.*

Solution. Une variable aléatoire X de loi binomiale $\mathcal{B}(n, p)$ peut s'écrire comme la somme de n variables de Bernoulli $\mathcal{B}(p)$ indépendantes.

$$X = \sum_{i=1}^n X_i$$

Ainsi, sa variance est

$$\text{Var}(X) = n\text{Var}(X_1) = np(1 - p) .$$

■

3.4.3 Inégalité de Bienaymé-Chebichev

L'inégalité suivante permet de quantifier quelque peu le fait que la variance contrôle l'écart d'une variable à son espérance.

Proposition 3.4.3 *Soit X une variable aléatoire d'espérance m et de variance σ^2 finie. Alors*

$$\forall \epsilon > 0 , \quad P(|X - m| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2} .$$

Commentaires. Cette inégalité peut s'interpréter de la manière suivante. Plus la variance est petite et plus la variable se concentre autour de sa moyenne.

Démonstration. Il est utile de connaître la démonstration de ce résultat. On a

$$\begin{aligned} \mathbb{P}(|X - m| > \epsilon) &= \mathbb{P}(|X - m|^2 > \epsilon^2) = \int_{(|X-m|^2 > \epsilon^2)} d\mathbb{P} \\ &= \int_{(\frac{|X-m|^2}{\epsilon^2} > 1)} 1 d\mathbb{P} \\ &\leq \int_{(\frac{|X-m|^2}{\epsilon^2} > 1)} \frac{|X - m|^2}{\epsilon^2} d\mathbb{P} \\ &\leq \int_{\Omega} \frac{|X - m|^2}{\epsilon^2} d\mathbb{P} = \frac{\text{Var}(X)}{\epsilon^2}. \end{aligned}$$

■

Revenons maintenant sur l'intuition fréquentiste en probabilités. L'inégalité de Bienaymé-Chebichev permet de démontrer en un *certain sens* que la moyenne de variables aléatoires indépendantes et de même loi converge vers l'espérance commune de ces variables.

Théorème 3.4.2 Loi faible des grands nombres. *Soit X une variable aléatoire d'espérance m et de variance σ^2 finie. Soit $(X_n)_{n \geq 1}$ une suite de variables indépendantes de même loi que X . Alors*

$$\forall \epsilon > 0, \quad \mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - m\right| > \epsilon\right) \rightarrow 0 \quad \text{lorsque } n \rightarrow \infty.$$

Démonstration. Appliquons l'inégalité de Bienaymé-Chebichev à la variable

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}.$$

On a

$$E\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{E[X_1] + \dots + E[X_n]}{n} = m$$

et

$$\text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\text{Var}(X_1) + \dots + \text{Var}(X_n)}{n^2} = \frac{\sigma^2}{n}.$$

Donc

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - m\right| > \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2 n}$$

et,

$$\forall \epsilon > 0, \quad \frac{\sigma^2}{\epsilon^2 n} \rightarrow 0 \quad \text{lorsque } n \rightarrow \infty.$$

■

Soit A un événement dont la probabilité est égale à $p > 0$. Appliquons le théorème précédent à la variable $X = \mathbb{1}_A$ qui admet pour loi la loi $\mathcal{B}(p)$. Dans une répétition d'épreuves indépendantes où A est susceptible de se réaliser avec probabilité p , le théorème précédent indique que la fréquence de réalisation de A approche la valeur p .

Application à la simulation. Il n'est pas toujours possible de calculer la probabilité d'un événement A donné de manière explicite. On peut penser alors que des méthodes numériques sont susceptibles de fournir un calcul approché. La loi des grands nombres (ou plutôt sa version forte que nous verrons plus tard, chapitre 7) justifie le calcul approché de cette probabilité par la fréquence de réalisation de A . Si l'on admet que la variable $\mathbb{1}_A$ est simulable, nous obtenons l'algorithme suivant

```

 $S := 0$ 
Répéter  $n$  fois
   $X := \mathbb{1}_A$ 
   $S := S + X$ 
FinRépéter
 $\bar{X}_n := S/n$ 

```

En sortie de cet algorithme, la variable aléatoire \bar{X}_n fournit une valeur approchée de la valeur $P(A) = p$.

3.4.4 Autres inégalités

L'inégalité de Bienaymé-Chebichev porte sur l'estimation de probabilités. Les inégalités présentées dans ce paragraphe portent sur des espérances. Nous aurons besoin de la notion de convexité d'une fonction réelle. Une fonction réelle f deux fois différentiable est dite convexe si

$$\forall x \in \mathbb{R}, \quad f''(x) \geq 0.$$

Par exemple, les fonctions $x \rightarrow x^2$ et $x \rightarrow e^x$ sont convexes.

Théorème 3.4.3 Inégalité de Jensen. *Soit X une variable aléatoire réelle et f une fonction réelle convexe. Alors,*

$$E[f(X)] \geq f(E[X])$$

pour autant que les espérances existent et sont finies.

Démonstration. Ecrivons le développement de Taylor de f autour de $m = E(X)$

$$f(x) = f(m) + f'(m)(x - m) + \frac{f''(\theta)}{2}(x - m)^2$$

où θ est un réel compris entre x et m . Comme $f''(\theta) \geq 0$, on aura

$$f(x) \geq f(m) + f'(m)(x - m)$$

et donc

$$f(X) \geq f(m) + f'(m)(X - m).$$

Il suffit alors de prendre l'espérance des deux membres de cette dernière inégalité pour conclure. ■

Théorème 3.4.4 Inégalité de Cauchy-Schwarz. *Soit X et Y deux variables aléatoires réelles de carré intégrable. Alors*

$$\|E[XY]\| \leq \sqrt{E[X^2]} \sqrt{E[Y^2]}.$$

Commentaires. Cette inégalité est l'analogie de l'inégalité de Cauchy-Schwarz pour les espaces vectoriels euclidiens. Remarquons que l'ensemble des variables aléatoires de carré intégrable forme un espace vectoriel de dimension infini pour lequel l'application

$$(X, Y) \rightarrow E[XY]$$

joue le rôle d'un produit scalaire (espace de Hilbert).

Démonstration. On distingue les cas $Y = -tX$ pour un t réel et $Y \neq -tX$. Dans le premier cas, on vérifie que l'inégalité annoncée est une égalité. Dans le second cas,

$$0 < E[(tX + Y)^2] = E[X^2]t^2 + 2E[XY]t + E[Y^2].$$

Le discriminant du polynôme de degré 2

$$t \rightarrow E[X^2]t^2 + 2E[XY]t + E[Y^2]$$

est négatif car le polynôme ne s'annule pas. On vérifie facilement que cette condition est équivalente à la conclusion du théorème. ■

L'inégalité de Cauchy-Schwarz se généralise de la manière suivante.

Théorème 3.4.5 Inégalité de Hölder. *Soient X et Y deux variables aléatoires positives et $p \geq 1, q \geq 1$ deux réels tels que $1/p + 1/q = 1$. Alors,*

$$E[XY] \leq E[X^p]^{1/p} E[Y^q]^{1/q}.$$

Démonstration. Exercice. (Utiliser la convexité de la fonction $-\ln$.) ■

3.5 Transformées

Nous avons vu dans ce chapitre que la loi d'une variable aléatoire réelle se caractérise

- 1) par la probabilité de chaque valeur prise par la variable dans le cas discret,
- 2) par la fonction de répartition de la variable dans tous les cas et en particulier dans le cas de variables à densité.

Il existe d'autres fonctions susceptibles de caractériser la loi de la variable X . Dans le cas d'une variable à valeurs dans \mathbb{N} , on parlera de *fonction génératrice*. En général, on parlera de *fonction caractéristique* et de *transformée de Laplace* (si la variable est positive).

3.5.1 Fonctions génératrices

Définition 3.5.1 La fonction génératrice d'une variable aléatoire N à valeurs dans \mathbb{N} est définie par la série entière

$$G(z) = \sum_{n=0}^{\infty} P(N = n)z^n .$$

Le rayon de convergence ρ de cette série vérifie

$$\rho \geq 1 .$$

De manière plus synthétique, la fonction génératrice d'une variable discrète N s'exprime

$$\forall |z| \leq 1, \quad G(z) = E[z^N] .$$

Proposition 3.5.1 Soit N une variable aléatoire à valeurs dans \mathbb{N} et G sa fonction génératrice.

- 1) Pour tout $n \in \mathbb{N}$, $P(N = n) = \frac{1}{n!}G^{(n)}(0)$.
- 2) Si G est dérivable au point $z = 1$, N admet une espérance et

$$E[N] = G'(1) .$$

- 3) Si G est deux fois dérivable au point $z = 1$, N est de carré intégrable

$$E[N(N - 1)] = G''(1) \quad \text{et} \quad \text{Var}(N) = G''(1) + G'(1) - (G'(1))^2 .$$

Commentaires. La fonction génératrice G caractérise la loi de la variable N d'après 1). Elle peut s'avérer utile pour identifier une loi donnée. Elle permet aussi de calculer les moments de la variable N par dérivations successives.

Démonstration. La première assertion est une conséquence de la formule de Taylor. Pour les deux autres, notons que

$$G'(z) = \sum_{n=1}^{\infty} n z^{n-1} P(N = n)$$

et

$$G''(z) = \sum_{n=2}^{\infty} n(n-1) z^{n-2} P(N = n) .$$

On conclut en prenant $z = 1$ dans ces deux expressions. ■

Exemple 3.5.1 Loi de Bernoulli $\mathcal{B}(p)$, $0 < p < 1$.

On a

$$G(z) = (1 - p) + pz .$$

On retrouve $E[N] = p$ et $\text{Var}(N) = p - p^2 = p(1 - p)$. ■

Exemple 3.5.2 Loi géométrique $\mathcal{G}(p)$.

On a

$$G(z) = pz \sum_{n=1}^{\infty} (1-p)^{n-1} z^{n-1} = \frac{pz}{1 - (1-p)z}.$$

On vérifie que

$$E[N] = 1/p \quad \text{et} \quad \text{Var}(N) = \frac{(1-p)}{p^2}.$$

■

Exemple 3.5.3 Loi de Poisson $\mathcal{P}(\lambda)$, $\lambda > 0$.

On a

$$G(z) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} z^n = \exp(\lambda(z-1)).$$

On vérifie que

$$E[N] = \lambda \quad \text{et} \quad \text{Var}(N) = \lambda.$$

■

Proposition 3.5.2 Soient M et N deux variables aléatoires indépendantes à valeurs dans \mathbb{N} de fonctions génératrices respectives G_M et G_N . On note G_{M+N} la fonction génératrice de la somme $M + N$. Alors,

$$G_{M+N}(z) = G_M(z)G_N(z).$$

Démonstration. D'après la proposition (3.4.1),

$$G_{M+N}(z) = E[z^{M+N}] = E[z^M]E[z^N] = G_M(z)G_N(z).$$

■

Exemple 3.5.4 Soit X et Y deux variables aléatoires indépendantes de loi de Poisson de paramètres respectifs λ et μ . La variable $Z = X + Y$ est une variable aléatoire de Poisson de loi $\lambda + \mu$.

Solution. En effet,

$$G_Z(z) = \exp(\lambda(z-1)) \exp(\mu(z-1)) = \exp((\lambda + \mu)(z-1)).$$

■

Exemple 3.5.5 Une variable aléatoire de loi binomiale $\mathcal{B}(n, p)$ peut s'écrire comme la somme de n variables de Bernoulli $\mathcal{B}(p)$ indépendantes. Ainsi, la fonction génératrice de la loi $\mathcal{B}(n, p)$ est

$$G(z) = ((1-p) + pz)^n.$$

Exemple 3.5.6 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes à valeurs dans \mathbb{N} , de même loi. Soit N une variable aléatoire à valeurs dans \mathbb{N} indépendante des X_n .

On note G la fonction génératrice commune des X_n , μ et σ^2 l'espérance et la variance des X_n et F la fonction génératrice de N .

On définit

$$Z = \sum_{i=1}^N X_i .$$

- Vérifier que Z est une variable aléatoire.
- Déterminer la fonction génératrice de Z en fonction de G et F .
- En déduire $E[Z]$ et $\text{Var}(Z)$.

Solution.

- Z est une application de Ω à valeurs dans \mathbb{N} . Pour vérifier que Z est une variable aléatoire, il suffit de montrer que les sous-ensembles $(Z = k)$ de Ω sont des événements. Or, pour tout $k \in \mathbb{N}$,

$$(Z = k) = \bigcup_{n \in \mathbb{N}} (N = n) \cap (X_1 + \cdots + X_n = k).$$

Par définition, $(N = n)$ et $(X_1 + \cdots + X_n = k)$ sont des événements puisque N et X_1, X_2, \dots sont des variables aléatoires. Par les axiomes de définition de la tribu des événements, une union dénombrable d'intersections d'événements est encore un événement donc $(Z = k)$ est un événement.

- On calcule, pour tout $|z| \leq 1$, $H(z) = E[z^Z]$.

$$\begin{aligned} H(z) &= \sum_{k \in \mathbb{N}} P(Z = k) z^k \\ &= \sum_{k \in \mathbb{N}} P\left(\bigcup_{n \in \mathbb{N}} (N = n) \cap (X_1 + \cdots + X_n = k)\right) z^k \\ &= \sum_{k \in \mathbb{N}} \sum_{n \in \mathbb{N}} P(N = n) P(X_1 + \cdots + X_n = k) z^k \\ &= \sum_{n \in \mathbb{N}} P(N = n) \sum_{k \in \mathbb{N}} P(X_1 + \cdots + X_n = k) z^k \\ &= \sum_{n \in \mathbb{N}} P(N = n) E(z^{(X_1 + \cdots + X_n)}) \\ &= \sum_{n \in \mathbb{N}} P(N = n) \prod_{i=1}^n E(z^{X_i}) \\ &= \sum_{n \in \mathbb{N}} P(N = n) G(z)^n \\ &= F(G(z)) \end{aligned}$$

Pour passer de la deuxième ligne à la troisième, on utilise le fait que les événements $(N = n)$ sont disjoints et que la variable N est indépendante des X_i . A la quatrième ligne, l'inversion des deux signes \sum est possible puisque la série $H(z)$ est normalement convergente lorsque $|z| \leq 1$. L'indépendance des X_i est utilisée à la sixième ligne. On conclut grâce au fait que les X_i ont même loi et donc même fonction génératrice.

- Afin de simplifier la discussion, on suppose que F et G sont deux fois dérivables au point $|z| = 1$. Les relations du cours donnent

$$\begin{aligned} E[Z] &= H'(1) = E[N] \mu, \\ \text{Var}(Z) &= H''(1) - H'(1)^2 + H'(1) = E[N] \sigma^2 + \text{Var}(N) \mu^2. \end{aligned}$$

Le premier résultat n'est pas surprenant si l'on se rappelle que $E[X_1 + \dots + X_n] = n\mu$. En revanche, sommer un nombre aléatoire de variables indépendantes introduit un "aléa" supplémentaire que l'on peut voir sur la variance de Z . En effet, $\text{Var}(X_1 + \dots + X_n) = n\sigma^2$ et nous avons démontré qu'un terme s'ajoute à "l'effet moyen" $E[N] \sigma^2$. ■

Nous terminons ce paragraphe en développant un exemple qui montre que la notion de fonction génératrice permet d'obtenir certaines espérances dont le calcul serait extrêmement complexe si l'on devait s'en passer. Nous partons d'une situation concrète liée à la finance. La préoccupation centrale d'un investisseur dans un marché financier aléatoire est d'estimer s'il peut réaliser un bénéfice par rapport à son investissement initial. Soit S_0 le capital investi initialement et S_t la valeur de ce capital au temps t . Typiquement, l'individu s'intéresse au premier instant T_1 où $S_t \geq S_0 + \alpha$, $\alpha > 0$. Bien entendu, il cherchera à calculer la probabilité $P(T_1 < \infty)$ pour que le bénéfice se réalise un jour et le temps moyen $E[T_1]$ qu'il lui faudra attendre. Nous ne disposons pas, à ce point d'avancée du cours, des outils mathématiques et de modélisation nécessaires pour aborder ce problème sous une forme raisonnable. Toutefois, nous pouvons caricaturer cette situation pour l'étudier à l'aide des outils dont nous disposons.

Exemple 3.5.7 *Supposons qu'un joueur soit impliqué dans un jeu de pile ou face et soit S_n le gain cumulé réalisé par le joueur à l'issue de la n^e partie. Nous avons*

$$\begin{aligned} S_0 &= 0 \\ S_n &= X_1 + \dots + X_n \end{aligned}$$

où les (X_i) sont des variables aléatoires indépendantes représentant le gain (ou la perte) du joueur à l'issue de chaque partie. Nous imposons $X_i \in \{-1, +1\}$ et

$$P(X_i = +1) = 1 - P(X_i = -1) = p \in]0, 1[.$$

On s'intéresse au temps pour lequel le joueur est bénéficiaire pour la première fois

$$T_1 = \min\{n \geq 1 \text{ t.q. } S_n = +1\}.$$

Solution. On définit de plus

$$T_0 = \min\{n \geq 1 \text{ t.q. } S_n = 0\} .$$

Il est évident que T_1 est impair et que T_0 est pair. Soit

$$\forall n \geq 1, \quad p_n = P(T_1 = n)$$

Nous cherchons à déterminer la loi de la variable T_1 . Pour les valeurs paires, nous avons $p_{2k} = 0$. Pour $n = 1$, nous avons

$$P(T_1 = 1) = p .$$

Soit $n \geq 3$ (n impair), nous avons

$$\begin{aligned} P(T_1 = n) &= P(T_1 = n \cap S_1 = -1) + P(T_1 = n \cap S_1 = +1) \\ &= P(T_1 = n \mid S_1 = -1)(1 - p) \end{aligned}$$

or

$$\begin{aligned} P(T_1 = n \mid S_1 = -1) &= \sum_{k=2}^{n-1} P(T_1 = n \cap T_0 = k \mid S_1 = -1) \\ &= \sum_{k=2}^{n-1} P(T_1 = n \mid T_0 = k ; S_1 = -1)P(T_0 = k \mid S_1 = -1) . \end{aligned}$$

De plus

$$\begin{aligned} P(T_0 = k \mid S_1 = -1) &= P(T_1 = k - 1) \\ P(T_1 = n \mid T_0 = k ; S_1 = -1) &= P(T_1 = n - k) . \end{aligned}$$

Ainsi

$$p_n = (1 - p) \sum_{k=2}^{n-1} p_{k-1} p_{n-k} .$$

Cette formule permet de calculer p_n par récurrence mais n'est pas explicite. En revanche, on reconnaît dans p_n le terme général d'une série produit. Ainsi

$$G_{T_1}(z) = \sum_{n=1}^{\infty} p_n z^n = pz + z(1 - p)G_{T_1}^2(z) .$$

Après résolution de cette équation du second degré, nous obtenons

$$G'_{T_1}(1) = E[T_1] = \begin{cases} \frac{1}{2p-1} & \text{si } p > 1/2 \quad (\text{jeu favorable}) \\ \infty & \text{si } p \leq 1/2 \quad (\text{jeu équilibré ou défavorable.}) \end{cases}$$

La probabilité pour que le joueur réalise un jour un bénéfice est

$$P(T_1 < \infty) = \sum_{n=1}^{\infty} p_n = G_{T_1}(1) = \begin{cases} 1 & \text{si } p \geq 1/2 \\ \frac{p}{1-p} & \text{si } p < 1/2 . \end{cases}$$

■

3.5.2 Fonctions caractéristiques

Pour une variable aléatoire quelconque, la notion de *fonction caractéristique* remplace celle de fonction génératrice introduite pour les variables discrètes.

Définition 3.5.2 Soit X une variable aléatoire réelle. La fonction caractéristique de X est la fonction à valeurs complexes définie par

$$\forall t \in \mathbb{R}, \quad \phi_X(t) = E[e^{itX}].$$

Commentaires. On dira qu'une application à valeurs complexes est intégrable si ses parties réelles et imaginaires le sont. Or, nous avons

$$e^{itX} = \cos(tX) + i \sin(tX)$$

et

$$|\cos(tX)| \leq 1 \quad \text{et} \quad |\sin(tX)| \leq 1.$$

Ceci implique que la fonction caractéristique est bien définie.

En termes de mesures, la fonction caractéristique de X est la *transformée de Fourier* de la mesure de probabilité P_X . Soit X est une variable aléatoire admettant pour densité f_X . Nous avons

$$\forall t \in \mathbb{R}, \quad \phi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x) dx.$$

On identifie la densité de X à l'aide de sa fonction caractéristique en utilisant la formule d'inversion de Fourier

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

La fonction caractéristique caractérise donc la loi de la variable X . Cette fonction possède des propriétés analogues à celles des fonctions génératrices. En ce qui concerne par exemple les différents moments de la variable X , nous avons le résultat suivant.

Proposition 3.5.3 La variable aléatoire X admet un moment d'ordre r si et seulement si ϕ est dérivable r fois au point $t = 0$. On a alors

$$E[X^r] = (-i)^r \phi_X^{(r)}(0).$$

Démonstration. Il suffit de permuter les symboles de dérivation et d'espérance (on justifie cette opération grâce aux théorèmes du chapitre 3). ■

Exemple 3.5.8 Montrer que la fonction caractéristique d'une variable aléatoire X de loi $\mathcal{N}(0, 1)$ est donnée par

$$\forall t \in \mathbb{R}, \quad \phi_X(t) = e^{-t^2/2}.$$

Commentaires. Ce calcul prendra une importance toute particulière lors des chapitres suivants (vecteurs gaussiens et convergences).

Solution. Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$, nous avons

$$\begin{aligned}\phi_X(t) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{itx} e^{-x^2/2} dx \\ &= e^{-t^2/2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{t^2/2+itx-x^2/2} dx \\ &= e^{-t^2/2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-(x-it)^2/2} dx \\ &= e^{-t^2/2} .\end{aligned}$$

■

Exercice 3.5.1 Soit X une variable aléatoire de loi uniforme sur l'intervalle (a, b) , $a < b$. Montrer que

$$\forall t \in \mathbb{R}, \quad \phi_X(t) = \frac{e^{ibt} - e^{iat}}{i(b-a)t} .$$

Soit X une variable aléatoire de loi exponentielle de paramètre $\lambda > 0$. Montrer que

$$\forall t \in \mathbb{R}, \quad \phi_X(t) = \frac{1}{1 - it/\lambda} .$$

Exemple 3.5.9 Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$, m et σ des réels quelconques. On pose

$$Y = \sigma X + m .$$

Montrer que la fonction caractéristique de la variable Y est donnée par

$$\forall t \in \mathbb{R}, \quad \phi_Y(t) = e^{itm - \sigma^2 t^2/2} .$$

Commentaires. La variable Y est une variable de loi normale $\mathcal{N}(m, \sigma^2)$.

Solution. Nous avons

$$\forall t \in \mathbb{R}, \quad \phi_Y(t) = E[e^{it(\sigma X + m)}] = e^{itm} E[e^{it\sigma X}] = e^{itm} \phi_X(t\sigma) .$$

Puisque X admet pour loi $\mathcal{N}(0, 1)$, nous obtenons

$$\phi_Y(t) = e^{itm} e^{-\sigma^2 t^2/2} .$$

■

Pour terminer ce paragraphe, nous pouvons énoncer le théorème suivant.

Proposition 3.5.4 Soient X et Y deux variables aléatoires réelles indépendantes de fonctions caractéristiques respectives ϕ_X et ϕ_Y . On note ϕ_{X+Y} la fonction caractéristique de la somme $X + Y$. Alors,

$$\phi_{X+Y} = \phi_X \phi_Y .$$

Démonstration. Exercice. ■

Exemple 3.5.10 Soit X_1 une variable aléatoire de loi $\mathcal{N}(m_1, \sigma_1^2)$ et X_2 une variable aléatoire de loi $\mathcal{N}(m_2, \sigma_2^2)$ indépendante de X_1 . On pose

$$Y = X_1 + X_2 .$$

Montrer que Y admet pour loi $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Solution. Nous avons

$$\forall t \in \mathbb{R}, \quad \phi_Y(t) = \phi_{X_1}(t) \phi_{X_2}(t) .$$

D'après l'exemple précédent et après factorisation, nous obtenons

$$\phi_Y(t) = e^{it(m_1+m_2) - (\sigma_1^2 + \sigma_2^2)t^2/2} .$$

Il s'agit bien de la fonction caractéristique de la loi $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$. ■

3.5.3 Transformée de Laplace

Pour une variable réelle positive X , une notion voisine de la fonction caractéristique peut-être définie. Il s'agit de la *transformée de Laplace*

$$\forall s \in \mathbb{R}_+, \quad L_X(s) = E[e^{-sX}] .$$

Cette fonction est à valeurs réelles donc plus facile à manipuler sur le plan pratique que la fonction caractéristique. Notons que les calculs à effectuer sont identiques dans les deux cas. Nous admettrons que cette transformée caractérise la loi de la variable aléatoire X et nous laissons au lecteur le soin d'en établir les principales propriétés à titre d'exercice. Cette transformée est souvent préférée à la fonction caractéristique lorsque l'on travaille avec des variables à valeurs dans \mathbb{R}_+ .

Chapitre 4

Couples et conditionnement

4.1 Couples de variables aléatoires

Nous étudions dans ce chapitre la loi conjointe d'un couple (X, Y) de variables aléatoires à valeurs réelles. Il s'agit d'une mesure de probabilité sur \mathbb{R}^2 muni de sa tribu de Borel $\mathcal{B}(\mathbb{R}^2)$. Nous notons cette mesure $P_{(X,Y)}$ et nous la définissons de la manière suivante

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

D'après le théorème de prolongement, cette mesure de probabilité se caractérise par sa fonction de répartition que nous appellerons *fonction de répartition conjointe* des variables X et Y

$$\forall s, t \in \mathbb{R}, \quad F(s, t) = P(X \leq s; Y \leq t).$$

La loi de chacune des variables du couple peut se déduire de la fonction de répartition conjointe. Par exemple, la fonction de répartition de la variable X se calcule de la manière suivante

$$\begin{aligned} F_X(s) &= P(X \leq s) \\ &= P(\cup_{n \in \mathbb{N}} (X \leq s; Y \leq n)) \\ &= \lim_{n \rightarrow \infty} P(X \leq s; Y \leq n) \\ &= \lim_{n \rightarrow \infty} F(s, n). \end{aligned}$$

De la même manière, nous avons

$$\forall s \in \mathbb{R}, \quad F_X(s) = \lim_{t \rightarrow \infty} F(s, t)$$

où la limite est prise pour t réel. Les lois de chacune des variables X et Y s'appellent les *lois marginales* du couple (X, Y) .

Définition 4.1.1 Soit (X, Y) un couple de variables aléatoires réelles positives ou intégrables. On définit l'**espérance mathématique** du couple (X, Y) comme le couple

$$(m_1, m_2) = (E[X], E[Y]).$$

Considérons maintenant une fonction mesurable φ définie de \mathbb{R}^2 dans \mathbb{R} et la variable $Z = \varphi(X, Y)$ supposée intégrable ($E[|Z|] < \infty$). Alors, le théorème de la mesure image conduit au résultat suivant

$$E[Z] = E[\varphi(X, Y)] = \int_{\mathbb{R}^2} \varphi(x, y) dP_{(X, Y)}(x, y).$$

4.1.1 Lois conjointes discrètes

Dans le cas où X et Y sont des variables aléatoires discrètes, à valeurs dans \mathbb{N} par exemple, il est commode de définir la fonction p suivante

$$\forall k, l \in \mathbb{N}, \quad p(k, l) = P(X = k ; Y = l).$$

Dans ce cas, la fonction p est assimilée à la loi conjointe des variables X et Y et caractérise cette loi. On retrouve la loi marginale de la variable X en sommant sur toutes les valeurs prises par Y

$$\forall k \geq 0, \quad P(X = k) = \sum_{l \geq 0} p(k, l).$$

Les variables X et Y sont indépendantes si et seulement si

$$\forall k, l \geq 0, \quad P(X = k ; Y = l) = P(X = k)P(Y = l).$$

En d'autres termes, la loi du couple se *factorise*.

Exemple 4.1.1 Une urne contient r boules dont r_1 sont blanches et r_2 sont noires ($r_1 + r_2 = r$). On effectue n tirages successifs en remplaçant à chaque fois la boule tirée dans l'urne. A l'issue des n tirages, on note X le nombre de boules blanches et Y le nombre de boules noires. Déterminer la loi conjointe du couple (X, Y) et les lois de X et Y .

Solution. Pour tout k, l positifs tels que $k + l = n$,

$$P(X = k, Y = l) = \frac{(k+l)!}{k!l!} \left(\frac{r_1}{r}\right)^k \left(\frac{r_2}{r}\right)^l.$$

On vérifie facilement que la loi (marginale) de X est donnée par

$$\forall k = 0, \dots, n, \quad P(X = k) = P(X = k, Y = n - k) = C_n^k p_1^k (1 - p_1)^{n-k}$$

où $p_1 = \frac{r_1}{r}$. La variable X suit la loi $\mathcal{B}(n, p_1)$. Par symétrie, la variable Y suit la loi $\mathcal{B}(n, p_2)$ où $p_2 = \frac{r_2}{r}$. ■

4.1.2 Couples de variables à densité

Définition 4.1.2 Soit f une fonction mesurable de \mathbb{R}^2 dans \mathbb{R}_+ telle que

$$\iint_{\mathbb{R}^2} f(x, y) dx dy = 1 .$$

Le couple de variables aléatoires réelles (X, Y) admet f pour densité si

$$\forall C \in \mathcal{B}(\mathbb{R}^2) , \quad P_{(X, Y)}(C) = \iint_C f(x, y) dx dy .$$

Par la suite, nous considérons des fonctions densité continues (sauf peut-être sur un ensemble de mesure de Lebesgue nulle). Supposons que C soit le produit cartésien de deux sous-ensembles A et B boréliens de \mathbb{R} . On obtient par le théorème de Fubini

$$P(X \in A ; Y \in B) = \int_B \left(\int_A f(x, y) dx \right) dy .$$

La fonction de répartition d'un couple de densité f est donc

$$F(s, t) = P(X \leq s ; Y \leq t) = \int_{-\infty}^t \int_{-\infty}^s f(x, y) dx dy .$$

Réciproquement, si F est donnée, il suffit de dériver (lorsque c'est possible) pour obtenir

$$f(x, y) = \frac{\partial^2 F}{\partial s \partial t}(x, y) .$$

Commentaires. Il est possible de donner une idée intuitive (parfois utile pour certains calculs) de la notion de densité conjointe. On peut vérifier en effet que

$$P(X \in (x, x + dx) ; Y \in (y, y + dy)) \simeq f(x, y) dx dy$$

à condition que dx et dy soient petits et que f soit continue au point (x, y) .

Lorsque (X, Y) est un couple de variables aléatoires de densité f , chacune des variables X et Y admettent des densités que l'on appelle *densités marginales*. On obtient les densités marginales de X et de Y de la manière suivante

$$\begin{aligned} \forall A \in \mathcal{B}(\mathbb{R}) , \quad P(X \in A) &= P(X \in A ; Y \in \mathbb{R}) \\ &= \int_A \left(\int_{\mathbb{R}} f(x, y) dy \right) dx \\ &= \int_A f_X(x) dx . \end{aligned}$$

Par conséquent, nous avons

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy .$$

Par symétrie des rôles de x et de y , on obtient

$$f_Y(y) = \int_{\mathbb{R}} f(x, y) dx .$$

Nous terminons ce paragraphe par quelques exemples.

Exemple 4.1.2 Soit (X, Y) un couple de variables aléatoires dont la densité conjointe est donnée par

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y} & x \in \mathbb{R}_+, y \in \mathbb{R}_+, \\ 0 & \text{sinon.} \end{cases}$$

Quelle est la loi marginale de la variable X ? Calculer la probabilité de l'événement $(X < Y)$.

Solution. On vérifie que

$$\forall s \geq 0, \quad P(X \leq s) = \int_0^s \int_0^\infty 2e^{-x}e^{-2y} dy dx = 1 - e^{-s}.$$

Ainsi, la loi marginale de la variable X est la loi exponentielle $\mathcal{E}(1)$. Il est possible de calculer la probabilité de différents événements faisant intervenir les variables X et Y . Par exemple

$$P(X < Y) = \iint_{(x < y)} f(x, y) dx dy = \int_0^\infty \left(\int_0^y 2e^{-x}e^{-2y} dx \right) dy = 1/3.$$

■

Exemple 4.1.3 Loi uniforme sur le disque unité On considère un disque de rayon 1 et on choisit un point au hasard dans ce disque. La distribution de la position de ce point sera dite uniforme. Les variables X et Y représentent les coordonnées cartésiennes du point choisi. Nous cherchons l'expression de la densité du couple (X, Y) et celles des densités des lois marginales.

Solution. Puisque la distribution des points dans le disque est uniforme, la densité du couple (X, Y) doit être constante sur le disque

$$f(x, y) = \begin{cases} c & \text{si } x^2 + y^2 \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

Or,

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

Donc

$$c \iint_{(x^2+y^2 \leq 1)} dx dy = 1.$$

Cette dernière intégrale est égale à l'aire du disque. Ainsi, nous avons

$$c = \frac{1}{\pi}.$$

Connaissant la densité marginale de X , celle de Y s'obtient par symétrie des rôles de x et y . Détaillons ce calcul pour la marginale en X . Soit $|x| \leq 1$,

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy$$

$$\begin{aligned}
&= \frac{1}{\pi} \int_{(x^2+y^2 \leq 1)} dy \\
&= \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} dy \\
&= \frac{2}{\pi} \sqrt{1-x^2}
\end{aligned}$$

Lorsque $|x| \geq 1$, la densité de la variable X est nulle. Nous avons donc

$$\forall x \in \mathbb{R}, \quad f_X(x) = \frac{2}{\pi} \sqrt{1-x^2} \mathbb{1}_{[-1,+1]}(x).$$

■

Exemple 4.1.4 Supposons la densité conjointe de (X, Y) donnée par

$$f(x, y) = \begin{cases} e^{-(x+y)} & x \in \mathbb{R}_+, \quad y \in \mathbb{R}_+, \\ 0 & \text{sinon.} \end{cases}$$

On souhaite déterminer la loi de la variable X/Y .

Solution. Cherchons tout d'abord la fonction de répartition de X/Y . Pour $t \geq 0$,

$$\begin{aligned}
F_{X/Y}(t) &= P(X/Y \leq t) \\
&= P((X, Y) \in \{(x, y) ; x/y \leq t\}) \\
&= \iint_{(x/y \leq t)} e^{-(x+y)} dx dy \\
&= \int_0^\infty \left(\int_0^{ty} e^{-(x+y)} dx \right) dy \\
&= \int_0^\infty (1 - e^{-ty}) e^{-y} dy \\
&= 1 - 1/(t+1).
\end{aligned}$$

On obtient la densité de X/Y par dérivation

$$f_{X/Y}(x) = \frac{1}{(x+1)^2} \mathbb{1}_{\mathbb{R}_+}(x).$$

■

4.1.3 Indépendance

4.1.4 Variables aléatoires indépendantes

Lors du chapitre précédent, nous avons défini la notion de variables indépendantes. Regardons maintenant quelle est la conséquence de cette définition sur la densité conjointe d'un couple de variables indépendantes. Soit (X, Y) un couple de variables

aléatoires réelles de densité conjointe f . Elles sont indépendantes si pour tous boréliens A et B de $\mathcal{B}(\mathbb{R})$, on a

$$P(X \in A ; Y \in B) = P(X \in A)P(Y \in B) .$$

En s'appuyant sur le théorème de prolongement, cette propriété est équivalente à

$$\forall s, t \in \mathbb{R}, \quad F(s, t) = P(X \leq s ; Y \leq t) = P(X \leq s) P(Y \leq t) = F_X(s)F_Y(t) .$$

En dérivant par rapport à chacune des variables, on obtient

$$f(x, y) = f_X(x)f_Y(y)$$

où f_X et f_Y sont les densités marginales des variables X et Y . Autrement dit, la densité du couple (X, Y) se *factorise*. Nous venons d'établir la proposition suivante.

Proposition 4.1.1 *Soit (X, Y) un couple de densité f . Les variables X et Y sont indépendantes si et seulement si*

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

Voici des exemples de calculs effectués avec des variables indépendantes.

Exemple 4.1.5 *Deux personnes se donnent un rendez-vous. L'heure d'arrivée de ces deux personnes est une variable aléatoire uniforme répartie sur l'intervalle $(20h, 21h)$. Ces deux variables sont indépendantes. Quelle est la probabilité pour que la première personne arrivée attende plus de 10 minutes ?*

Solution. Désignons par X et Y les deux variables correspondant aux dates d'arrivée de chacun des individus, exprimées en minutes. Il s'agit de deux variables indépendantes de loi $\mathcal{U}(0, 60)$. La probabilité cherchée est

$$P(X + 10 < Y) + P(Y + 10 < X) = 2P(X + 10 < Y) .$$

On obtient alors

$$\begin{aligned} 2 P(X + 10 < Y) &= 2 \iint_{x+10 < y} f(x, y) dx dy \\ &= 2 \iint_{(x+10 < y)} f_X(x) f_Y(y) dx dy \\ &= 2 \int_{10}^{60} \int_0^{y-10} \left(\frac{1}{60}\right)^2 dx dy \\ &= \frac{25}{36} . \end{aligned}$$

■

L'exemple qui suit est l'un des plus anciens problèmes de probabilité en relation avec la géométrie.

Exemple 4.1.6 Problème de Buffon. *Sur une table, on trace des lignes parallèles espacées d'un écart D les unes des autres. On y jette une aiguille de longueur L , avec $L \leq D$. Quelle est la probabilité que l'aiguille coupe une ligne ? (L'alternative est que l'aiguille soit complètement située dans l'une des bandes délimitée par les lignes.)*

Solution. On repère la position de l'aiguille grâce aux coordonnées X et Θ suivantes. La variable X représente la distance entre le milieu de l'aiguille et la parallèle la plus proche. L'angle Θ est l'angle formé entre l'aiguille et une perpendiculaire aux lignes. L'aiguille chevauchera une parallèle si

$$\frac{X}{\cos \Theta} < \frac{L}{2}$$

ou

$$X < \frac{L}{2} \cos \Theta .$$

La variable X varie entre 0 et $D/2$ tandis que Θ varie entre 0 et $\pi/2$. Les variables X et Θ sont des variables indépendantes de loi uniforme. Aussi aura-t-on

$$\begin{aligned} P(X < \frac{L}{2} \cos \Theta) &= \iint_{(x < \frac{L}{2} \cos \theta)} f_X(x) f_\Theta(\theta) dx d\theta \\ &= \frac{4}{\pi D} \int_0^{\pi/2} \int_0^{L/2 \cos \theta} dx d\theta \\ &= \frac{4}{\pi D} \int_0^{\pi/2} \frac{L}{2} \cos \theta d\theta \\ &= \frac{2L}{\pi D} \end{aligned}$$

■

Application : Calcul de π . L'exemple précédent permet d'imaginer une méthode de calcul de π faisant intervenir la loi des grands nombres. On répète n fois le jet de l'aiguille sur la table de manière à ce que l'indépendance soit garantie. Après chaque jet, on compte le nombre N de fois où l'aiguille coupe une ligne parallèle. D'après la loi des grands nombres, on peut considérer que la fréquence empirique N/n converge vers $\frac{2L}{\pi D}$. Ce résultat sera précisé par la loi forte des grands nombres dans le chapitre sur la convergence.

4.1.5 Sommes de variables aléatoires indépendantes

Soit (X, Y) un couple de variables aléatoires indépendantes de densité conjointe f . Nous avons donc

$$\forall (x, y) \in \mathbb{R}^2, \quad f(x, y) = f_X(x) f_Y(y) .$$

Nous souhaitons déterminer la densité de la variable $X + Y$. Cherchons à calculer la fonction de répartition de $X + Y$

$$\forall t \in \mathbb{R}, \quad F_{X+Y}(t) = P(X + Y \leq t)$$

$$\begin{aligned}
&= \iint_{(x+y \leq t)} f_X(x) f_Y(y) dx dy \\
&= \int_{\mathbb{R}} \left(\int_{-\infty}^{t-y} f_X(x) f_Y(y) dx \right) dy \\
&= \int_{\mathbb{R}} F_X(t-y) f_Y(y) dy
\end{aligned} \tag{4.1}$$

La fonction F_{X+Y} obtenue dans l'équation (4.1) est appelée *produit de convolution* des fonctions F_X et f_Y . La densité de la variable $X + Y$ est obtenue comme produit de convolution des densités marginales f_X et f_Y . Ceci s'obtient par dérivation

$$\begin{aligned}
\forall x \in \mathbb{R}, \quad f_{X+Y}(x) &= \frac{d}{dt} \int_{\mathbb{R}} F_X(t-y) f_Y(y) dy \Big|_{t=x} \\
&= \int_{\mathbb{R}} \frac{d}{dt} F_X(t-y) \Big|_{t=x} f_Y(y) dy \\
&= \int_{\mathbb{R}} f_X(x-y) f_Y(y) dy
\end{aligned}$$

■

Application. La densité d'une variable de loi Gamma de paramètre $a > 0$ et $\lambda > 0$ notée $G(a, \lambda)$ est de la forme

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{\lambda}{\Gamma(a)} e^{-\lambda x} (\lambda x)^{a-1} \mathbb{1}_{\mathbb{R}_+}(x)$$

ou $\Gamma(a)$ est une constante qui s'obtient par intégration

$$\Gamma(a) = \int_0^{\infty} \lambda^a e^{-\lambda x} x^{a-1} dx .$$

Pour $n \geq 1$, nous avons $\Gamma(n) = (n-1)!$. De plus $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ et pour tout $a > 0$, $\Gamma(a+1) = a\Gamma(a)$.

Proposition 4.1.2 *Si X et Y sont deux variables indépendantes de loi Gamma de paramètres respectifs (a, λ) et (b, λ) , $a, b > 0$, alors $X + Y$ sera également une variable de loi Gamma avec pour paramètres $(a + b, \lambda)$.*

Commentaires. La famille des lois Gamma possède une propriété d'additivité lorsque λ est constant. On dit, pour λ constant, que la famille est stable par le produit de convolution.

Démonstration. On utilise le produit de convolution obtenu ci-dessus. Soit $x > 0$, nous avons

$$\begin{aligned}
f_{X+Y}(x) &= \frac{1}{\Gamma(a)\Gamma(b)} \int_0^x \lambda e^{-\lambda(x-y)} (\lambda(x-y))^{b-1} \lambda e^{-\lambda y} (\lambda y)^{a-1} dy \\
&= C(\lambda, a, b) e^{-\lambda x} \int_0^x (x-y)^{b-1} y^{a-1} dy .
\end{aligned}$$

En posant $z = y/x$, on obtient

$$\forall x \in \mathbb{R}, \quad f_{X+Y}(x) = K(\lambda, a, b)e^{-\lambda x}x^{a+b-1}.$$

La valeur de la constante K ne dépend pas de x . De plus, cette dernière expression est celle d'une densité. Son intégration devra donner 1. Nous avons donc nécessairement

$$K(\lambda, a, b) = \frac{\lambda^{a+b}}{\Gamma(a+b)} \quad \text{et} \quad f_{X+Y}(x) = \frac{\lambda}{\Gamma(a+b)} e^{-\lambda x} (\lambda x)^{a+b-1} \mathbb{1}_{\mathbb{R}_+}(x).$$

Exercice. Déterminer la loi de la somme de n variables indépendantes de loi exponentielle de paramètre 1. ■

Commentaires. Bien entendu, il peut être intéressant du point de vue calculatoire d'utiliser les fonctions caractéristiques pour déterminer la loi de la somme de deux variables indépendantes (cf. proposition 4.5.4).

4.2 Changement de variables

Considérons un couple de variables aléatoires réelles (X, Y) de densité conjointe $f_{(X,Y)}$. On appellera *support* de $f_{(X,Y)}$ le domaine D , que l'on suppose ouvert, défini par

$$D = \{(x, y) \in \mathbb{R}^2 ; f_{(X,Y)}(x, y) > 0\}.$$

On s'intéresse à la densité conjointe du couple (Z, T) obtenu par une transformation notée $\varphi(X, Y)$ où

$$\forall (x, y) \in D, \quad \varphi(x, y) = \begin{cases} z = \varphi_1(x, y) \\ t = \varphi_2(x, y) \end{cases}.$$

Nous supposons que φ est bijective de D sur son image $\varphi(D)$ et nous noterons son inverse de la manière suivante

$$\forall (z, t) \in \varphi(D), \quad \varphi^{-1}(z, t) = \begin{cases} x = \varphi_1^{-1}(z, t) \\ y = \varphi_2^{-1}(z, t) \end{cases}.$$

Nous supposons de plus φ et φ^{-1} de classe C^1 , et nous notons

$$Jac(\varphi)(x, y) = \begin{vmatrix} \frac{\partial \varphi_1}{\partial x} & \frac{\partial \varphi_1}{\partial y} \\ \frac{\partial \varphi_2}{\partial x} & \frac{\partial \varphi_2}{\partial y} \end{vmatrix}$$

le déterminant de la matrice des dérivées partielles (il ne s'annule pas sur le support D).

Proposition 4.2.1 *La densité du couple (Z, T) est donnée par la formule suivante*

$$\forall (z, t) \in \varphi(D), \quad f_{(Z,T)}(z, t) = |Jac(\varphi^{-1})(z, t)| f_{(X,Y)}(\varphi^{-1}(z, t)). \quad (4.2)$$

Commentaires. Le point délicat dans l'application de ce résultat réside souvent dans la détermination du domaine $\varphi(D)$. En général, on pourra déterminer $\Delta = \varphi(D)$ en fixant un Δ *a priori* puis en montrant les inclusions $\varphi(D) \subset \Delta$ et $\varphi^{-1}(\Delta) \subset D$.

Exemple 4.2.1 *On cherche à exprimer la densité du couple*

$$\begin{cases} Z = X + Y \\ T = X - Y \end{cases}$$

en fonction de la densité $f_{(X,Y)}$ du couple (X, Y) .

Solution. Ici, nous avons $\varphi(\mathbb{R}^2) = \mathbb{R}^2$. On vérifie que le jacobien du changement de variable (linéaire) est non nul

$$\text{Jac}(\varphi) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2.$$

La formule (4.2) s'applique donc et nous obtenons

$$\forall z, t \in \mathbb{R}, \quad f_{(Z,T)}(z, t) = \frac{1}{2} f_{(X,Y)}\left(\frac{z+t}{2}, \frac{z-t}{2}\right).$$

■

Dans le cas où X et Y sont de loi normale centrée réduite $\mathcal{N}(0, 1)$ et indépendantes, on a

$$\begin{aligned} \forall z, t \in \mathbb{R}, \quad f_{(Z,T)}(z, t) &= \frac{1}{4\pi} e^{-((z+t)^2/8 + (z-t)^2/8)} \\ &= \frac{1}{4\pi} e^{-(z^2+t^2)/4} \\ &= \frac{1}{\sqrt{4\pi}} e^{-t^2/4} \frac{1}{\sqrt{4\pi}} e^{-z^2/4}. \end{aligned}$$

On découvre ainsi, un résultat secondaire : les variables $X + Y$ et $X - Y$ sont indépendantes, de loi $\mathcal{N}(0, 2)$. Dans le chapitre 6, nous nous consacrerons plus largement aux variables gaussiennes et nous retrouverons ce résultat facilement.

Exemple 4.2.2 Coordonnées polaires. *Soit (X, Y) un couple de variables de loi uniforme sur le disque unité*

$$D = \{(x, y) \in \mathbb{R}^2 ; x^2 + y^2 < 1\}.$$

La densité de ce couple s'exprime de la manière suivante

$$f(x, y) = \begin{cases} \frac{1}{\pi} & \text{si } (x, y) \in D \\ 0 & \text{sinon.} \end{cases}$$

Soit (R, Θ) le couple de coordonnées polaires définies par

$$\begin{cases} X = R \cos \Theta \\ Y = R \sin \Theta. \end{cases}$$

On cherche à exprimer la densité du couple (R, Θ) .

Solution. Remarquons tout d'abord que l'on ne change pas la loi du couple (X, Y) en enlevant au domaine D le segment $[0, 1[\times\{0\}$ qui est de probabilité nulle. On cherche à vérifier les hypothèses du théorème (4.2.1) par inversion locale. Les ensembles D et Δ doivent être ouverts. On considère le changement de variable φ défini de la manière suivante

$$\varphi : \begin{array}{ccc} D \setminus [0, 1[\times\{0\} & \longrightarrow &]0, 1[\times]0, 2\pi[\\ (x, y) & \longrightarrow & (r, \theta) \end{array}$$

où

$$\varphi^{-1} : \begin{array}{ccc}]0, 1[\times]0, 2\pi[& \longrightarrow & D \setminus [0, 1[\times\{0\} \\ (r, \theta) & \longrightarrow & (x = r \cos \theta, y = r \sin \theta) . \end{array}$$

Nous avons

$$Jac(\varphi^{-1})(r, \theta) = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix}$$

et donc

$$Jac(\varphi^{-1})(r, \theta) = r .$$

D'après le théorème (4.2.1), nous avons

$$f_{(R, \Theta)}(r, \theta) = \frac{r}{\pi} \mathbb{1}_{]0, 1[\times]0, 2\pi[}(r, \theta) = 2r \mathbb{1}_{]0, 1[}(r) \frac{1}{2\pi} \mathbb{1}_{]0, 2\pi[}(\theta) .$$

Lorsque l'on tire un point au hasard dans le disque unité, le rayon et l'angle sont indépendants. De plus, la variable Θ suit la loi $\mathcal{U}(0, 2\pi)$ et R admet pour densité

$$\forall r \in \mathbb{R}, \quad f_R(r) = 2r \mathbb{1}_{]0, 1[}(r) .$$

■

4.3 Lois conditionnelles

Nous avons souvent besoin de calculer des probabilités ou des espérances lorsque des informations partielles sont disponibles. Dans ce cas, les grandeurs recherchées sont des probabilités ou des espérances conditionnelles. De plus, il peut s'avérer extrêmement utile pour faire des calculs en probabilité de *conditionner* selon une variable aléatoire appropriée.

4.3.1 Cas discret

Soient X et Y deux variables à valeurs dans \mathcal{N} . On se souvient que, pour toute paire d'événements A et B , la probabilité de A sachant B réalisé est

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

pourvu que $P(B) > 0$. Il est naturel de définir la loi de probabilité conditionnelle de la variable X sachant $Y = l$ par

$$\begin{aligned} \forall k \in \mathbb{N}, \quad p_X^{Y=l}(k) &= P(X = k | Y = l) \\ &= \frac{p(k, l)}{p_Y(l)} \quad \text{si } p_Y(l) > 0. \end{aligned}$$

Exemple 4.3.1 Soient X et Y deux variables indépendantes de loi de Poisson de paramètres respectifs λ et μ . Alors, la loi conditionnelle de la variable X sachant $X + Y = n$ est la loi binomiale de paramètres n et $\lambda/(\lambda + \mu)$.

Solution. Cette loi s'obtient ainsi

$$\begin{aligned} \forall k = 0, \dots, n, \quad P(X = k | X + Y = n) &= \frac{P(X = k, Y = n - k)}{P(X + Y = n)} \\ &= \frac{P(X = k) P(Y = n - k)}{P(X + Y = n)} \end{aligned}$$

La somme de deux variables de Poisson indépendantes est encore une variable de Poisson dont le paramètre est égal à la somme des paramètres. L'expression précédente devient donc

$$\begin{aligned} P(X = k | X + Y = n) &= \frac{e^{-\lambda} \lambda^k e^{-\mu} \mu^{n-k}}{k! (n-k)!} \frac{n!}{e^{-(\lambda+\mu)} (\lambda + \mu)^n} \\ &= C_n^k \left(\frac{\lambda}{\mu + \lambda} \right)^k \left(\frac{\mu}{\mu + \lambda} \right)^{n-k}. \end{aligned}$$

■

4.3.2 Variables à densité

Définition 4.3.1 Soit (X, Y) un couple de variables aléatoires de densité conjointe f . On définit la densité conditionnelle de X sachant $Y = y$ lorsque $f_Y(y) > 0$ par la relation

$$\forall x \in \mathbb{R}, \quad f_X^{Y=y}(x) = \frac{f(x, y)}{f_Y(y)}.$$

Cette définition se calcule sur le cas discret. On peut toutefois lui trouver une justification intuitive par le raisonnement heuristique suivant. Multiplions le membre de gauche par dx et le membre de droite par $dxdy/dy$ pour obtenir

$$\begin{aligned} f_X^{Y=y}(x) dx &= \frac{f(x, y) dxdy}{f_Y(y) dy} \\ &\simeq \frac{P(x \leq X \leq x + dx; y \leq Y \leq y + dy)}{P(y \leq Y \leq y + dy)} \\ &= P(x \leq X \leq x + dx | y \leq Y \leq y + dy). \end{aligned}$$

En d'autres termes, lorsque dx et dy sont suffisamment petits, la quantité $f_X^{Y=y}(x)dx$ représente la probabilité conditionnelle que X se trouve entre x et $x + dx$ sachant que Y est entre y et $y + dy$.

Pour tout $B \in \mathcal{B}(\mathbb{R})$ ("relatif" à X), on aura par définition

$$P(X \in B | Y = y) = \int_B f_X^{Y=y}(x) dx .$$

Il faut noter que les notions qui viennent d'être présentées permettent le calcul de probabilités conditionnelles même dans le cas où la probabilité de la condition $Y = y$ est nulle...

Exemple 4.3.2 Soit D le domaine défini par

$$D = \{0 < x < 1, 0 < y < 1\}$$

et (X, Y) un couple de densité

$$f(x, y) = \begin{cases} \frac{12}{5}x(2 - x - y) & \text{si } (x, y) \in D \\ 0 & \text{sinon.} \end{cases}$$

On cherche la loi conditionnelle de la variable X , sachant $Y = y$, $0 < y < 1$.

Solution. On a, pour tout $0 < x < 1$

$$\begin{aligned} f_X^{Y=y}(x) &= \frac{x(2 - x - y)}{\int_0^1 x(2 - x - y) dx} \\ &= \frac{6x(2 - x - y)}{4 - 3y} . \end{aligned}$$

■

4.3.3 Autres cas de lois conditionnelles

Définition 4.3.2 On définit la loi conditionnelle de la variable X sachant l'événement A réalisé par la relation

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad P(X \in B | A) = \frac{P((X \in B) \cap A)}{P(A)} .$$

Lorsque cette loi admet une densité, on la note f^A et on a

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

Commentaires. Nous venons en fait de définir la conditionnelle de X sachant ($\mathbb{1}_A = 1$). On peut donc parler de distribution conditionnelle même lorsque les deux variables mises en jeu ne possèdent pas de densité conjointe ou ne sont pas toutes deux discrètes.

Imaginons le cas de deux variables l'une notée X de densité f et l'autre discrète notée N . On s'intéresse à la densité conditionnelle de X sous la condition $N = n$. Tout d'abord pour $h > 0$,

$$\frac{P(x < X < x + h \mid N = n)}{h} = \frac{P(N = n \mid x < X < x + h) P(x < X < x + h)}{P(N = n) h}.$$

Lorsque $h \rightarrow 0$, on peut écrire

$$\lim_{h \rightarrow 0} \frac{P(x < X < x + h \mid N = n)}{h} = \frac{P(N = n \mid X = x)}{P(N = n)} f(x).$$

où nous avons défini

$$P(N = n \mid X = x) = \lim_{h \rightarrow 0} P(N = n \mid x \leq X < x + h).$$

Ainsi, nous avons

Définition 4.3.3 Soit X une variable aléatoire réelle admettant pour densité f et N une variable aléatoire entière. La densité conditionnelle de X sachant ($N = n$) est définie par

$$\forall x \in \mathbb{R}, \quad f_X^{N=n}(x) = \frac{P(N = n \mid X = x)}{P(N = n)} f(x).$$

Commentaires. Bien entendu, lorsque N est égale à la variable de Bernoulli $\mathbb{1}_A$ pour un événement $A \in \mathcal{A}$, nous retrouvons la définition précédente.

Exemple 4.3.3 On considère une suite de $m+n$ épreuves de Bernoulli indépendantes identiques. Leur probabilité de succès est aléatoire et distribuée de manière uniforme sur l'intervalle $(0, 1)$. Que devient cette distribution si l'on sait que l'on a exactement n succès et m échecs ?

Solution. Désignons par U la probabilité de succès associée à une épreuve donnée. Sachant $U = u$, le nombre N de succès suit la loi binomiale de $\mathcal{B}(n+m, u)$. La densité conditionnelle de U , sachant ($N = n$) est donc

$$\begin{aligned} f_U^{N=n}(u) &= \frac{P(N = n \mid U = u)}{P(N = n)} f_U(u) \\ &= K(n, m) u^n (1-u)^m \quad 0 < u < 1 \end{aligned}$$

où $K(n, m)$ est une constante que l'on obtient par intégration

$$1/K(n, m) = \int_0^1 u^n (1-u)^m du$$

La loi conditionnelle est donc la loi béta $\beta(1+n, 1+m)$. ■

4.3.4 Espérance conditionnelle

Définition 4.3.4 Soient M et N deux variables aléatoires à valeurs dans \mathbb{N} et n un entier tel que $P(N = n) > 0$. On appelle espérance conditionnelle de M sachant $N = n$ la grandeur

$$E[M|N = n] = \sum_{m \geq 0} mP(M = m | N = n) .$$

On constate que l'espérance conditionnelle de M sachant $N = n$ n'est autre que l'espérance de M prise par rapport à la mesure de probabilité conditionnelle $P(\cdot | N = n)$. Dans le cas de variables à densité, la définition est la suivante.

Définition 4.3.5 Soit (X, Y) un couple de variables aléatoires réelles admettant une densité et y un réel tel que $f_Y(y) > 0$. On appelle espérance conditionnelle de X sachant $Y = y$ la grandeur

$$E[X|Y = y] = \int_{\mathbb{R}} x f_X^{Y=y}(x) dx .$$

Exemple 4.3.4 Supposons que la densité du couple (X, Y) soit donnée par

$$\forall (x, y) \in \mathbb{R}^2, \quad f(x, y) = \frac{1}{y} \exp(-y - x/y) \mathbf{1}_{\mathbb{R}_+ \times \mathbb{R}_+}(x, y) .$$

On souhaite calculer $E[X|Y = y]$.

Solution. Calculons tout d'abord la densité conditionnelle

$$\begin{aligned} \forall x \geq 0, \quad f_X^{Y=y}(x) &= \frac{f(x, y)}{\int_{\mathbb{R}_+} f(x, y) dx} \\ &= \frac{1}{y} e^{-x/y} . \end{aligned}$$

Cette loi conditionnelle est donc la loi exponentielle de paramètre $1/y$. Nous avons donc

$$E[X|Y = y] = \int_0^\infty x f_X^{Y=y}(x) dx = y .$$

■

Définition 4.3.6 Soit (X, Y) un couple de variables aléatoires à valeurs dans \mathbb{R}^2 . On appelle espérance conditionnelle de X sachant Y et on note $E[X|Y]$ la composition de la fonction

$$y \longrightarrow E[X|Y = y]$$

et de la variable Y .

Attention ! L'espérance conditionnelle $E[X|Y]$ est une variable aléatoire à la différence de l'espérance conditionnelle sachant $Y = y$ qui est déterministe. Ainsi, dans l'exemple précédent, nous avons

$$E[X|Y] = Y .$$

4.3.5 Calcul d'espérance par conditionnement

Le résultat qui suit énonce une propriété fondamentale de l'espérance conditionnelle.

Théorème 4.3.1

$$E[X] = E[E[X|Y]]$$

Commentaires. Lorsque Y est une variable discrète, ce théorème signifie que

$$E[X] = \sum_{n \in \mathbb{N}} E[X|Y = n]P(Y = n).$$

Tandis que lorsque Y est continue, le théorème signifie

$$E[X] = \int_{\mathbb{R}} E[X|Y = y]f_Y(y)dy.$$

Démonstration. On suppose pour cette démonstration que X et Y sont des variables discrètes. Nous avons donc,

$$\begin{aligned} \sum_{n \in \mathbb{N}} E[X|Y = n]P(Y = n) &= \sum_{n \in \mathbb{N}} \sum_{k \in \mathbb{N}} kP(X = k | Y = n)P(Y = n) \\ &= \sum_{n \in \mathbb{N}} \sum_{k \in \mathbb{N}} kP(X = k ; Y = n) \\ &= \sum_{k \in \mathbb{N}} k \sum_{n \in \mathbb{N}} P(X = k ; Y = n) \\ &= \sum_{k \in \mathbb{N}} kP(X = k) = E[X]. \end{aligned}$$

■

Exemple 4.3.5 Reprenons l'exemple (4.3.4). Rappelons que nous avons

$$\forall y \in \mathbb{R}_+ \quad E[X|Y = y] = y.$$

Nous cherchons à calculer $E[X]$.

Solution. D'après le théorème (4.3.1), nous avons

$$\begin{aligned} E[X] &= \int_{\mathbb{R}} E[X|Y = y]f_Y(y)dy \\ &= \int_{\mathbb{R}_+} yf_Y(y)dy = E[Y] \\ &= \int_{\mathbb{R}_+} e^{-y} \int_{\mathbb{R}_+} e^{-x/y} dx dy = 1. \end{aligned}$$

■

Commentaires. Bien entendu, on peut établir ce résultat directement à partir de la loi marginale de X . Toutefois, il est important d'avoir à l'esprit la méthode de conditionnement car souvent les lois marginales s'expriment difficilement. Voici un autre exemple significatif.

Exemple 4.3.6 *Un rat se trouve dans un labyrinthe face à deux portes. Il choisit la première de ces deux portes avec probabilité $\frac{1}{3}$ et la deuxième avec probabilité $\frac{2}{3}$. Quand il choisit la première porte, il revient à son point de départ en une minute. Quand il choisit la deuxième porte, il effectue un trajet d'une minute (jusqu'à un point intermédiaire) puis rebrousse chemin avec probabilité $\frac{1}{2}$ (le retour lui prend alors une minute) ou sort du labyrinthe en une minute. Tous les choix du rat se font indépendamment les uns des autres. Soit T le temps passé par le rat dans le labyrinthe. Quelle est la valeur de $E[T]$.*

Solution. la variable N représente le numéro de la porte choisie. Nous avons

$$E[T] = E[T|N = 1]1/3 + E[T|N = 2]2/3 .$$

et

$$\begin{aligned} E[T|N = 1] &= 1 + E[T] \\ E[T|N = 2] &= 1 + (E[T] + 1)1/2 + 1/2 \end{aligned}$$

Ainsi

$$E[T] = (1 + E[T])/3 + (2 + E[T]/2)2/3$$

soit

$$E[T] = 5 .$$

Pour formuler les équations (4.3), nous avons raisonné de la manière suivante. Si le rat emprunte la porte 1, il fait un trajet d'une minute et revient au point de départ. Le temps qu'il doit passer dans le labyrinthe est alors le même qu'à l'origine. Pour la porte 2, c'est à peine plus compliqué ... ■

Exemple 4.3.7 **Espérance de la somme d'un nombre aléatoire de variables aléatoires.** *Soit (X_n) une suite de variables réelles indépendantes de même loi et N une variable à valeurs dans \mathbb{N} indépendante de la suite (X_n) . On souhaite calculer l'espérance*

$$E\left[\sum_{i=1}^N X_i\right] .$$

Solution. On utilise le théorème (4.3.1), en conditionnant selon la variable N

$$\begin{aligned} E\left[\sum_{i=1}^N X_i\right] &= E\left[E\left[\sum_{i=1}^N X_i \mid N\right]\right] \\ &= \sum_{n \in \mathbb{N}} E\left[\sum_{i=1}^N X_i \mid N = n\right] P(N = n) \\ &= \sum_{n \in \mathbb{N}} E\left[\sum_{i=1}^n X_i\right] P(N = n) \quad \text{par indépendance} \\ &= \sum_{n \in \mathbb{N}} n E[X_1] P(N = n) \\ &= E[X_1] E[N] \end{aligned}$$

■

4.3.6 Calcul de probabilité par conditionnement

La technique utilisée pour calculer des espérances par conditionnement peut aussi s'appliquer pour calculer des probabilités. Soit A un événement quelconque et Y une variable aléatoire réelle. Nous posons

$$\forall \omega \in \Omega, \quad X(\omega) = \mathbb{1}_A(\omega)$$

où $\mathbb{1}_A$ est la fonction indicatrice de A . Du théorème 4.3.1 (valable en toute généralité), il vient

$$P(A) = E[\mathbb{1}_A] = E[E[\mathbb{1}_A | Y]] =_{conv} E[P(A|Y)].$$

Ainsi, lorsque Y est discrète, nous avons

$$P(A) = \sum_{n \in \mathbb{N}} E[\mathbb{1}_A | Y = n] P(Y = n) = \sum_{n \in \mathbb{N}} P(A | Y = n) P(Y = n).$$

Cette relation peut, par ailleurs, facilement se déduire de la formule des probabilités totales (chapitre 2). Lorsque Y admet une densité, nous avons

$$P(A) = \int_{\mathbb{R}} P(A | Y = y) f_Y(y) dy$$

où nous avons défini

$$P(A | Y = y) = \lim_{h \rightarrow 0} P(A | y \leq Y < y + h).$$

Exemple 4.3.8 Soient X et Y deux variables indépendantes et de densités f_X et f_Y . On peut expliciter par conditionnement la loi de $X + Y$.

Solution. En conditionnant selon Y , on obtient, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} P(X + Y < t) &= \int_{\mathbb{R}} P(X + Y < t | Y = y) f_Y(y) dy \\ &= \int_{\mathbb{R}} P(X + y < t | Y = y) f_Y(y) dy \\ &= \int_{\mathbb{R}} P(X < t - y) f_Y(y) dy \quad \text{par indépendance,} \\ &= \int_{\mathbb{R}} F_X(t - y) f_Y(y) dy. \end{aligned}$$

■

Exemple 4.3.9 Processus de branchement. Un processus de branchement décrit l'évolution de la taille d'une population par générations successives. A la génération 0, on suppose que la population est composée d'un unique individu. Cet individu est susceptible de donner naissance à X_0 descendants ($X_0 \in \{0, 1, 2\}$), selon une loi de descendance de fonction génératrice G_{X_0} . A la génération 1, les X_0 individus sont susceptibles de donner naissance à un nombre aléatoire d'individus, de manière indépendante les uns des autres et avec la même loi de descendance qu'à la génération initiale. Le processus se poursuit à l'identique de génération en génération. On s'intéresse à la probabilité d'extinction de la population, c'est-à-dire, à la probabilité qu'il existe une génération sans descendance.

Solution. Bien entendu, nous conditionnons selon la variable X_0

$$p_e = \text{P}(\text{extinction}) = \sum_{i=0}^2 \text{P}(\text{extinction} \mid X_0 = i) \text{P}(X_0 = i) .$$

Nous avons alors

$$\begin{aligned} \text{P}(\text{extinction} \mid X_0 = 0) &= 1 \\ \text{P}(\text{extinction} \mid X_0 = 1) &= p_e \\ \text{P}(\text{extinction} \mid X_0 = 2) &= p_e^2 . \end{aligned}$$

En reportant dans l'équation précédente, ceci conduit à

$$p_e = \text{P}(X_0 = 0) + \text{P}(X_0 = 1)p_e + \text{P}(X_0 = 2)p_e^2 = G_{X_0}(p_e) .$$

En conséquent, la probabilité p_e est un point fixe de la fonction génératrice de la variable X_0 . Ceci constitue un résultat tout à fait général des processus de branchement. En particulier, si

$$\text{P}(X_0 = 0) = \text{P}(X_0 = 1) = 1/4$$

alors, la résolution de l'équation du second degré donne

$$p_e = 1/2$$

■

4.3.7 Espérance conditionnelle et prédiction

Il arrive que l'on puisse observer la valeur d'une variable aléatoire Y et qu'ensuite, on souhaite prédire la valeur d'une autre variable X en se fondant sur ce que l'on sait sur Y . Admettons par exemple que Y soit une variable financière observée au temps t et que l'on cherche à prédire l'évolution de cette variable au temps $t + 1$. On note donc X la valeur au temps $t + 1$. On cherche un *prédicteur* de X , c'est-à-dire une variable $g(Y)$, la plus *proche* possible de X . Ce prédicteur doit être le plus précis possible. Nous imposons comme critère de précision que *l'écart quadratique moyen*

$$E[(X - g(Y))^2]$$

soit minimal. Nous allons montrer que selon ce critère la meilleure prédiction de X au vu de l'observation de Y est $E[X|Y]$.

Proposition 4.3.1 *Soit X et Y deux variables aléatoires de carré intégrable. Pour toute fonction mesurable g telle que $E[g^2(Y)] < \infty$, nous avons*

$$E[(X - g(Y))^2] \geq E[(X - E[X|Y])^2] .$$

Démonstration. Nous allons démontrer que

$$E[(X - g(Y))^2 | Y] \geq E[(X - E[X|Y])^2 | Y].$$

Nous pourrions alors conclure en prenant l'espérance des deux termes grâce au théorème (4.3.1). Développons

$$\begin{aligned} E[(X - g(Y))^2 | Y = y] &= E[(X - E[X|Y = y] + E[X|Y = y] - g(Y))^2 | Y = y] \\ &= E[(X - E[X|Y = y])^2 | Y = y] \\ &\quad + (E[X | Y = y] - g(y))^2 \\ &\quad + 2(E[X|Y = y] - g(y)) E[X - E[X|Y = y] | Y = y]. \end{aligned}$$

Le dernier terme est nul par linéarité de l'espérance. Donc,

$$E[(X - g(Y))^2 | Y = y] - E[(X - E[X|Y = y])^2 | Y = y] \geq (E[X | Y = y] - g(y))^2 \geq 0.$$

■

Commentaires. Considérons un nombre U pris au hasard dans l'intervalle $(0, 1)$ (U admet pour loi $\mathcal{U}(0, 1)$). Admettons que l'on sache que l'événement $A = (U > 1/2)$ est réalisé ou de manière équivalente que $\mathbf{1}_A = 1$. Calculons alors

$$E[U | \mathbf{1}_A = 1] = \int_{1/2}^1 u \frac{du}{P(U > 1/2)} = 3/4.$$

La variable aléatoire $E[U | \mathbf{1}_A]$ prend les valeurs $1/4$ et $3/4$ avec les probabilités $1/2$ et $1/2$. Ainsi, lorsque l'on pose la question “ U est-il plus grand que $1/2$?” et que l'on obtient une réponse affirmative, la meilleure prédiction de U est $3/4$. Cela justifie le choix de la méthode de dichotomie comme stratégie optimale de recherche de U dans l'intervalle $(0, 1)$ (voir examen janvier 1997). Bien entendu, la méthode de dichotomie n'est pas forcément optimale lorsque la loi de U n'est pas uniforme.

Nous terminons ce paragraphe avec le constat suivant (facultatif). On trouve dans certains ouvrages de référence une autre définition de l'espérance conditionnelle.

Définition 4.3.7 Soit (X, Y) un couple de variables aléatoires réelles. On appelle espérance conditionnelle et l'on note $E[X|Y]$ l'unique variable aléatoire réelle Z telle que

$$E[\phi(Y)X] = E[\phi(Y)Z]$$

pour toute fonction réelle ϕ mesurable pour la tribu de Borel sur \mathbb{R} .

L'équivalence des deux définitions est donnée par la proposition suivante qui généralise le théorème (4.3.1).

Proposition 4.3.2 Soit (X, Y) un couple de variables aléatoires réelles. Pour toute fonction réelle ϕ mesurable pour la tribu de Borel sur \mathbb{R} , on a

$$E[\phi(Y)X] = E[\phi(Y)E[X|Y]].$$

Démonstration. Nous démontrons la proposition lorsque le couple (X, Y) admet une densité. On a

$$\begin{aligned}
 E[\phi(Y)X] &= E[E[\phi(Y)X|Y]] \\
 &= \int_{\mathbb{R}} E[\phi(y)X|Y=y]f_Y(y)dy \\
 &= \int_{\mathbb{R}} \phi(x)E[Y|X=x]f_X(x)dx \\
 &= E[\phi(Y)E[X|Y]]
 \end{aligned}$$

■

4.4 Simulation

Les techniques de conditionnement sont très utiles pour la simulation de couples de variables aléatoires ou de variables elles-mêmes (méthode de rejet).

4.4.1 Couples de variables

Soit à simuler un couple (X, Y) de densité f . On utilisera la méthode suivante. On cherchera à simuler tout d'abord X selon la loi marginale de densité f_X . Ensuite, sachant $X = x$, on cherchera à simuler la loi conditionnelle de densité $f_Y^{X=x}$.

Exemple 4.4.1 Soit Δ le domaine de \mathbb{R}^2 défini par

$$\Delta = \{(x, y) \in \mathbb{R}^2, 0 < y < x\}.$$

Nous cherchons à simuler un couple de variables de densité

$$f(x, y) = \begin{cases} \frac{1}{x}e^{-x} & \text{si } (x, y) \in \Delta \\ 0 & \text{sinon.} \end{cases}$$

Solution. Nous pouvons vérifier que

$$\forall x \in \mathbb{R}, \quad f_X(x) = e^{-x} \mathbb{1}_{\mathbb{R}_+^*}(x)$$

et

$$\forall y \in \mathbb{R}, \quad f_Y^{X=x}(y) = \frac{1}{x} \mathbb{1}_{]0, x]}(y).$$

Nous pouvons utiliser la méthode d'inversion pour simuler la variable X . Ainsi, l'algorithme suggéré s'écrit

$$\begin{aligned}
 X &:= -\ln(ALEA) \\
 Y &:= X * ALEA
 \end{aligned}$$

On peut vérifier que cet algorithme est correct en utilisant la formule de changement de variables. Soit (U, V) un couple de variables de loi uniforme sur $(0, 1)$ et indépendantes. Soit φ l'application définie sur le domaine $D = [0, 1]^2$ par

$$\forall (u, v) \in D, \quad \varphi(u, v) = \begin{cases} x & = -\ln u \\ y & = -v \ln u \end{cases}$$

L'inverse de φ est défini par

$$\forall (x, y) \in \varphi(D), \quad \varphi^{-1}(x, y) = \begin{cases} u & = e^{-x} \\ v & = y/x \end{cases}$$

Les deux inclusions $\varphi(D) \subset \Delta$ et $\varphi^{-1}(\Delta) \subset D$ sont immédiates. De plus, nous avons

$$\text{Jac}(\varphi) = \begin{vmatrix} -1/u & 0 \\ -v/u & -\ln u \end{vmatrix} = \ln u / u = -xe^x$$

et, d'après la formule de changement de variables

$$f_{(X,Y)}(x, y) = |\text{Jac}(\varphi^{-1})| \mathbb{1}_{\Delta}(x, y) = \frac{1}{x} e^{-x} \mathbb{1}_{\Delta}(x, y) = f(x, y).$$

■

4.4.2 Méthodes de conditionnement

Les méthodes de conditionnement permettent de simuler des variables aléatoires réelles (chapitre 4) en utilisant le principe de rejet. Nous allons voir que, sur le même principe, nous pouvons aussi simuler des couples de variables aléatoires.

Nous avons vu lors du chapitre 2 qu'imposer lors d'une épreuve une condition de type *l'événement A doit être réalisé* revenait à changer de mesure de probabilité. On passe alors de la mesure P à la mesure $P(\cdot | A)$ et tous les calculs se font sous cette nouvelle mesure. Admettons que l'épreuve consiste à simuler une variable U de loi de densité f_U de manière répétitive jusqu'à ce que la condition $(U \in C)$ soit réalisée. La variable X obtenue en sortie aura pour loi

$$\forall B \in \mathcal{B}(\mathbb{R}^2), \quad P(X \in B) = \int_{B \cap C} \frac{f_U(x)}{P(U \in C)} dx.$$

Exemple 4.4.2 Loi uniforme sur $(0, \alpha)$, $0 < \alpha < 1$. On considère l'algorithme suivant

Répéter

$U := \text{ALEA}$

Jusqu'à $(U < \alpha)$

$X := U$.

Solution. Dans cet exemple, l'épreuve consiste en un appel du générateur aléatoire. Le résultat de l'épreuve est donc une variable aléatoire U de loi $\mathcal{U}(0, 1)$. Nous avons en sortie

$$\forall 0 \leq t \leq \alpha, \quad \mathbb{P}(X \leq t) = \mathbb{P}(U \leq t \mid U < \alpha) = t/\alpha$$

et

$$\forall t \geq \alpha, \quad \mathbb{P}(X \leq t) = \mathbb{P}(U \leq t \mid U < \alpha) = \alpha/\alpha = 1.$$

Ainsi, X admet pour loi la loi $\mathcal{U}(0, \alpha)$. ■

Admettons que l'épreuve consiste à simuler un couple de variables (U, V) de loi de densité $f_{(U, V)}$ de manière répétitive jusqu'à ce que la condition $((U, V) \in D)$ soit réalisée. Le couple (X, Y) obtenu en sortie aura pour loi

$$\forall B \in \mathcal{B}(\mathbb{R}^2), \quad \mathbb{P}((X, Y) \in B) = \iint_{B \cap D} \frac{f_{(U, V)}(x, y)}{\mathbb{P}((U, V) \in D)} dx dy.$$

Exemple 4.4.3 Loi uniforme sur le quart de disque unité. *On considère le domaine D défini par*

$$D = \{x > 0, y > 0, x^2 + y^2 < 1\},$$

et l'algorithme suivant

Répéter

$U := ALEA$

$V := ALEA$

*Jusqu'à $(U * U + V * V < 1)$*

$X := U ; Y := V.$

Solution. L'épreuve consiste à générer un couple de variables aléatoires indépendantes (U, V) de loi $\mathcal{U}(0, 1)$. Nous avons en sortie de l'algorithme

$$\begin{aligned} \forall B \in \mathcal{B}(\mathbb{R}^2), \quad \mathbb{P}((X, Y) \in B) &= \mathbb{P}((U, V) \in B \mid (U, V) \in D) \\ &= \frac{\mathbb{P}((U, V) \in B \cap D)}{\text{Aire}(D)} \\ &= \int_B \frac{4}{\pi} \mathbb{1}_D(x, y) dx dy. \end{aligned}$$

La loi du couple (X, Y) est donc la loi uniforme sur le domaine D . ■

4.4.3 Méthode de rejet

La méthode de conditionnement appliquée aux variables réelles se généralise. La méthode obtenue pour les variables réelles est parmi les plus populaires. Il s'agit de la *méthode de rejet*. Supposons que nous disposons d'une méthode pour simuler une

variable aléatoire réelle de densité g . On peut se servir de cette simulation comme point de départ pour simuler une variable qui aura pour densité f . L'idée consiste à simuler selon g puis d'accepter ou de rejeter le résultat y de cette simulation selon une probabilité proportionnelle à $f(y)/g(y)$. Nous supposons donc qu'il existe une constante $c > 0$ telle que

$$\forall y \in \mathbb{R}, \quad \frac{f(y)}{g(y)} \leq c.$$

L'algorithme de rejet s'écrit alors

```

Répéter
  simuler  $Y$  de densité  $g$ 
   $U := ALEA$ 
  Jusqu' à  $(U \leq f(Y)/cg(Y))$ 
 $X := Y$ 

```

Proposition 4.4.1 *La variable aléatoire X en sortie de l'algorithme de rejet admet pour densité f . Le nombre d'appels au générateur aléatoire effectués par l'algorithme de rejet suit la loi géométrique $\mathcal{G}(1/c)$.*

Démonstration. Soit Y une variable aléatoire de densité g . Nous avons

$$\begin{aligned}
 P(X \leq t) &= P(Y \leq t \mid U \leq f(Y)/cg(Y)) \\
 &= \frac{P(Y \leq t ; U \leq f(Y)/cg(Y))}{K} \\
 &= \frac{1}{K} \int_{\mathbb{R}} P(Y \leq t ; U \leq f(Y)/cg(Y) \mid Y = y) g(y) dy \\
 &= \frac{1}{K} \int_{-\infty}^t \frac{f(y)}{cg(y)} g(y) dy \\
 &= \frac{1}{Kc} \int_{-\infty}^t f(y) dy
 \end{aligned}$$

où l'on a noté $K = P(U \leq f(Y)/cg(Y))$. En faisant tendre t vers l'infini, on obtient $K = 1/c$. ■

L'espérance du nombre d'appels au générateur aléatoire est c . On a donc intérêt à choisir cette constante la plus petite possible dans l'optique de minimiser le coût de l'algorithme.

Chapitre 5

Annexe : Théorie de la mesure

5.1 Introduction

Ce chapitre constitue un aperçu de la *théorie de la mesure*. Nous admettrons la plupart des résultats issus de cette théorie. Toutefois, il paraît obligatoire d'en comprendre l'articulation et de maîtriser les principales applications qu'elle propose. En premier lieu, nous aurons besoin de définir rigoureusement quels sont les sous-ensembles de \mathbb{R}^n (ils constitueront une *tribu*) que l'on pourra raisonnablement mesurer. Puis nous verrons comment définir une mesure sur la tribu en question en spécifiant simplement la valeur des intervalles de \mathbb{R} ou des pavés de \mathbb{R}^n . En second lieu, la théorie des probabilités repose sur une notion abstraite d'*intégration*. L'espace d'intégration est constitué des éventualités d'une expérience. En général, on ne peut faire d'autre hypothèse de structure sur cet ensemble que celle qui consiste à le munir d'une tribu d'événements et d'une mesure de probabilité. C'est à partir de cette structure que nous allons construire une notion d'*intégrale* qui nous sera utile en probabilités. Enfin, le formalisme puissant de la théorie de la mesure nous permettra d'aboutir à des théorèmes dont l'intérêt pratique dépasse le cadre strict des probabilités. En particulier, on disposera d'un formalisme unifié pour traiter de sommes discrètes, d'intégrales sur \mathbb{R} , \mathbb{R}^n et d'intégrales abstraites. On pourra ensuite, sous des hypothèses relativement souples, intervertir signes *somme* et *limite*, dériver sous le signe *somme*, intervertir différentes intégrations. Ces opérations nécessitent souvent de lourdes justifications lorsque l'on ne dispose pas de la théorie de la mesure.

5.2 Espaces mesurables et mesures positives

5.2.1 Définitions

Nous avons vu dans le premier chapitre que le formalisme des probabilités nécessite

- 1) un espace fondamental Ω constitué des éventualités de l'expérience aléatoire,
- 2) un ensemble appelé *tribu* constitué d'événements cohérents entre eux,
- 3) une mesure de probabilité définie sur la tribu des événements.

Nous reproduirons une structure similaire pour élaborer la théorie de la mesure.

Définition 5.2.1 *Un ensemble Ω muni d'une tribu de parties \mathcal{A} sera dit espace mesurable. On notera cet espace sous forme d'une paire (Ω, \mathcal{A}) . Les éléments d'une tribu seront appelés événements ou encore ensembles mesurables relativement à cette tribu.*

Dans ce chapitre, les notions de *mesure* et d'*intégrale* seront introduites pour un espace mesurable (Ω, \mathcal{A}) quelconque. Toutefois, il faut garder à l'esprit les cas $\Omega = \mathbb{N}$ (cas *discret*) et $\Omega = \mathbb{R}, \mathbb{R}^n$ (cas *continus*). Il s'agit des situations dans lesquelles nous effectuerons des calculs concrets.

Définition 5.2.2 *On appelle mesure positive sur l'espace (Ω, \mathcal{A}) toute application μ de \mathcal{A} dans $\bar{\mathbb{R}}_+$ satisfaisant les axiomes suivants :*

- 1) $\mu(\emptyset) = 0$,
- 2) $\mu(\bigcup_{n \geq 1} A_n) = \sum_{i=1}^{\infty} \mu(A_n)$ pour toute suite $(A_n)_{n \geq 1}$ d'événements mutuellement exclusifs.

Le triplet $(\Omega, \mathcal{A}, \mu)$ est dit espace mesuré.

Une mesure de probabilité P est une mesure particulière, dite bornée. Elle satisfait à la condition

$$P(\Omega) = 1$$

en plus des axiomes 1) et 2) de la définition précédente. Le triplet (Ω, \mathcal{A}, P) est dit *espace probabilisé*.

Voici quelques propriétés élémentaires d'une mesure. Ces propriétés ont été démontrées pour une probabilité (cf. chapitre 1). Nous ne reviendrons pas sur leurs démonstrations.

Proposition 5.2.1 *Soit μ une mesure positive sur un espace mesurable (Ω, \mathcal{A}) . On a*

- 1) $\forall A, B \in \mathcal{A}, \quad A \subset B \Rightarrow \mu(A) \leq \mu(B)$,
- 2) pour toute suite croissante d'ensembles mesurables $(A_n)_{n \geq 1}$

$$\mu\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n).$$

- 3) pour toute suite décroissante $(A_n)_{n \geq 1}$ d'ensembles de mesure finie

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

5.2.2 Exemples élémentaires de mesures

Exemple 5.2.1 Mesure de Dirac. Soit ω_0 un point de Ω fixé. La mesure de Dirac (on dit aussi "masse de Dirac") au point ω_0 est la mesure définie de la manière suivante :

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

La masse de Dirac au point ω_0 est une mesure de probabilité. Pour une épreuve associée à cette mesure de probabilité, l'éventualité ω_0 peut être considérée comme certaine. En fait, il n'y a pas de hasard dans une telle épreuve. Le résultat est toujours ω_0 .

Exemple 5.2.2 Mesure de dénombrement. Soit $(\omega_n)_{n \geq 1}$ une suite de points de Ω . La mesure de dénombrement associée à cette suite est la mesure

$$\mu = \sum_{n \geq 1} \delta_{\omega_n} .$$

Par définition, nous avons

$$\forall A \in \mathcal{A}, \quad \mu(A) = \text{card} \{n \text{ t.q. } \omega_n \in A\} .$$

Les deux mesures que nous venons de définir sont des *mesures discrètes*. Les mesures discrètes peuvent être définies en général comme des combinaisons linéaires à coefficients positifs de masses de Dirac. Elles interviennent souvent dans le calcul de probabilités et nous les avons utilisées à plusieurs reprises dans les chapitres précédents. Lors du jeu de *pile* ou *face*, par exemple, nous avons travaillé avec l'espace probabilisé

$$\left(\{0, 1\}, \mathcal{P}(\{0, 1\}), \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 \right) ,$$

où la valeur 0 code *pile* et 1 code *face*. Pour décrire l'épreuve du lancer de dé équilibré, l'espace probabilisé utilisé est

$$\left(\{1, 2, \dots, 6\}, \mathcal{P}(\{1, 2, \dots, 6\}), \frac{1}{6} \sum_{i=1}^6 \delta_i \right) .$$

En revanche, nous ne sommes pas encore capables de décrire la mesure qui est liée au choix d'un nombre au hasard dans $(0, 1)$, pas plus que celle qui correspond à l'aire d'une partie de \mathbb{R}^2 .

5.2.3 Construction de mesures. Mesure de Lebesgue.

Dans ce paragraphe, nous souhaitons construire une mesure μ sur \mathbb{R}^n , $n \geq 1$. Afin de définir une mesure, il suffit de connaître la valeur $\mu(A)$ de tout ensemble mesurable A . C'est bien là que réside la difficulté car une telle application ne possède pas en général d'expression analytique. Il est en toute généralité impossible de spécifier une valeur pour chaque ensemble mesurable. Nous chercherons à définir la mesure de volume dans \mathbb{R}^n , $n \geq 1$. Pour cela, il est naturel de définir la mesure d'un *pavé*¹ (segment pour \mathbb{R} , rectangle pour \mathbb{R}^2) par le volume de ce pavé (longueur pour un segment, aire pour un rectangle). Pour tout $a_i < b_i$, $i = 1, \dots, n$

$$\text{Vol}([a_1, b_1] \times \dots \times [a_n, b_n]) = \prod_{i=1}^n (b_i - a_i) .$$

1. Un pavé est un produit cartésien d'intervalles de \mathbb{R}

Une axiomatique puissante devrait permettre à partir de cette notion élémentaire de mesure d'un pavé de fournir une mesure de volume de quasiment tous les sous-ensembles de \mathbb{R}^n ... Malheureusement, l'ensemble des pavés de \mathbb{R}^n ne constitue pas une tribu. Il s'agit donc de considérer la plus petite tribu qui contienne tous les pavés et de construire la mesure de volume sur cette tribu de sorte que ses valeurs coïncident pour tout pavé. Une telle construction nécessitera un résultat d'extension que nous admettrons. La mesure résultant de cette construction s'appellera la *mesure de Lebesgue* sur \mathbb{R}^n .

Définition 5.2.3 Soit \mathcal{F} une famille de parties de Ω . On appelle tribu engendrée par \mathcal{F} la plus petite tribu $\sigma(\mathcal{F})$ contenant \mathcal{F} .

Définition 5.2.4 On appelle tribu borélienne de \mathbb{R}^n ($n \geq 1$) et on note $\mathcal{B}(\mathbb{R}^n)$ la tribu engendrée par la famille des ouverts de \mathbb{R}^n .

Remarque. Il est possible de démontrer que

$$\mathcal{B}(\mathbb{R}) = \sigma(\{]a, b[; a < b \in \mathbb{R}\}) ,$$

et

$$\mathcal{B}(\mathbb{R}^n) = \sigma(\{\otimes_{i=1}^n]a_i, b_i[; a_i < b_i \in \mathbb{R}\}) .$$

Nous chercherons à mesurer n'importe quel élément de $\mathcal{B}(\mathbb{R}^n)$. La remarque précédente suggère qu'il suffit de définir la mesure des pavés (éléments générateurs) puis d'étendre cette définition aux éléments de $\mathcal{B}(\mathbb{R}^n)$. Nous avons besoin du formalisme des semi-anneaux afin d'énoncer ce résultat rigoureusement.

Définition 5.2.5 Une famille de parties \mathcal{S} de Ω est un semi-anneau sur Ω si

- 1) $A \in \mathcal{S}, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S}$.
- 2) Soit $A, B \in \mathcal{S}$ t.q. $B \subset A$, alors $A \setminus B = \bigcup_{i=1}^n A_i$, où les $A_i \in \mathcal{S}$ sont deux à deux disjoints.

Voici deux exemples importants de semi-anneaux qui ne sont pas des tribus

$$\Omega = \mathbb{R} \quad \text{et} \quad \mathcal{S} = \{]a, b[, a < b \in \mathbb{R}\}$$

et

$$\Omega = \mathbb{R}^n \quad \text{et} \quad \mathcal{S} = \{\otimes_{i=1}^n]a_i, b_i[, a_i < b_i \in \mathbb{R}\} .$$

Théorème 5.2.1 Théorème de prolongement. Soit \mathcal{S} un semi-anneau de parties de Ω et \mathcal{A} la tribu engendrée par \mathcal{S} . On suppose de plus que Ω peut être recouvert par une suite (peut-être finie) d'éléments de \mathcal{S} . Soit μ une application de \mathcal{S} dans $\overline{\mathbb{R}}_+$ vérifiant la propriété suivante

- (*) Si $(A_n)_{n \geq 1}$ est une suite (peut-être finie) d'ensembles deux à deux disjoints dans \mathcal{S} telle que $\bigcup_{n \geq 1} A_n \in \mathcal{S}$ alors

$$\mu\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mu(A_n) .$$

Alors, il existe une mesure sur \mathcal{A} unique dont la restriction à \mathcal{S} coïncide avec μ .

Le théorème (5.2.1) permet de définir une mesure par ses valeurs sur un semi-anneau. Considérons un exemple qui nous intéresse : celui d'une mesure de probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La tribu $\mathcal{B}(\mathbb{R})$ est la tribu engendrée par $\mathcal{S} = \{]a, b[, a < b \in \mathbb{R}\}$. La mesure μ est donc caractérisée par ses valeurs sur les intervalles de \mathbb{R} . Allons un peu plus loin. La fonction

$$\forall t \in \mathbb{R}, \quad F(t) = \mu(]-\infty, t]),$$

est bien définie par la donnée de la mesure des intervalles de longueur finie. En effet, d'après la proposition (5.2.1), nous avons

$$F(t) = \mu(\cup_{n \geq 1}]-n, t]) = \lim_{n \rightarrow \infty} \mu(]-n, t])$$

et cette dernière suite est croissante, bornée donc convergente. Compte-tenu du théorème ci-dessus, F caractérise entièrement la probabilité μ puisque $\mu(]a, b]) = F(b) - F(a)$. La fonction F s'appelle *fonction de répartition* de μ . Nous reviendrons largement sur cette fonction lors des chapitres suivants.

Définition 5.2.6 On appelle *mesure de Lebesgue sur \mathbb{R}* et on note λ l'unique mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ prolongeant l'application définie sur le semi-anneau des intervalles $]a, b[$ par

$$\lambda(]a, b]) = (b - a) \quad (\text{longueur de l'intervalle }]a, b]).$$

On appelle *mesure de Lebesgue sur \mathbb{R}^n* et on note λ l'unique mesure sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ prolongeant l'application définie sur le semi-anneau des pavés $\otimes_{i=1}^n]a_i, b_i[$ par

$$\lambda(\otimes_{i=1}^n]a_i, b_i]) = \prod_{i=1}^n (b_i - a_i) \quad (\text{volume du pavé } \otimes_{i=1}^n]a_i, b_i]).$$

Nous sommes désormais à même de donner une définition rigoureuse de l'espace probabilisé associé aux réalisations d'une variable U associée à un générateur aléatoire uniforme. Il s'agit de l'espace

$$([0, 1], \mathcal{B}([0, 1]), \lambda)$$

où $\mathcal{B}([0, 1]) = \mathcal{B}(\mathbb{R}) \cap [0, 1]$ et λ est la mesure de Lebesgue restreinte à $[0, 1]$. Dans ce cas, la mesure de Lebesgue est aussi appelée *mesure uniforme* sur l'intervalle $[0, 1]$. On parlera de *loi uniforme* en donnant au terme *loi* la même définition qu'au terme de mesure de probabilité (sur $([0, 1], \mathcal{B}([0, 1]))$ dans ce cas). Nous avons, en conclusion,

$$\forall B \in \mathcal{B}([0, 1]), \quad P(U \in B) = \lambda(B).$$

5.3 Applications mesurables

5.3.1 Définition

Définition 5.3.1 Une application f d'un espace mesurable (Ω, \mathcal{A}) dans un espace mesurable (E, \mathcal{B}) est dite mesurable si

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

En probabilité, on choisira d'appeler les fonctions mesurables *variables aléatoires*. On notera X plutôt que f . On dira que X est une variable aléatoire à valeurs dans (E, \mathcal{B}) si

$$\forall B \in \mathcal{B}, \quad X^{-1}(B) = \{X \in B\} \text{ est un événement de } \mathcal{A}.^2$$

Nous rassemblons dans la proposition suivante des propriétés élémentaires des fonctions mesurables à valeurs dans \mathbb{R} car ce sont elles qui seront les plus utilisées par la suite. L'ensemble \mathbb{R} est muni de la tribu $\mathcal{B}(\mathbb{R})$. Rappelons que les applications à valeurs dans \mathbb{R} sont appelées *fonctions*.

Proposition 5.3.1 Soit (Ω, \mathcal{A}) un espace mesuré, f, g deux fonctions mesurables et α un réel. Les applications αf , $f + g$, $\sup(f, g)$, $\inf(f, g)$, $f \cdot g$, $f^+ = \sup(f, 0)$, $f^- = \inf(f, 0)$ sont mesurables. Si $(f_n)_{n \geq 1}$ est une suite de fonctions mesurables, les fonctions $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, $\liminf_n f_n$ sont mesurables.

Démonstration. Admis. ■

Pour un événement $A \in \mathcal{A}$, la *fonction indicatrice* de A est définie par

$$\forall \omega \in \Omega, \quad \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon.} \end{cases}$$

Il s'agit d'une fonction mesurable (à voir en cours). D'après la proposition précédente, l'application

$$u_n = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$$

où $\alpha_1, \dots, \alpha_n$ sont des réels et A_1, \dots, A_n des événements de \mathcal{A} , est aussi une fonction mesurable. Une telle fonction est appelée *fonction étagée*. Les fonctions étagées permettent d'*approcher* n'importe quelle fonction mesurable.

Proposition 5.3.2 Soit (Ω, \mathcal{A}) un espace mesuré et f une fonction mesurable positive. Il existe une suite croissante $(u_n)_{n \geq 1}$ de fonctions étagées convergeant simplement vers f .

Démonstration. Pour tout $n \geq 1$, posons

$$u_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{1}_{\left(\frac{k}{2^n} \leq f < \frac{k+1}{2^n}\right)} + n \mathbb{1}_{\{f \geq n\}}.$$

On vérifie que cette suite convient. ■

2. L'ensemble $\{\omega \in \Omega : X(\omega) \in B\}$ se note $\{X \in B\}$

5.3.2 Mesure image

Les fonctions mesurables (ou de manière équivalente les variables aléatoires) à valeurs dans (E, \mathcal{B}) définies sur un espace mesuré $(\Omega, \mathcal{A}, \mu)$ permettent le *transport* de la structure mesurée sur l'espace d'arrivée (E, \mathcal{B}) . En d'autres termes, l'espace (E, \mathcal{B}) est muni naturellement d'une mesure. Cette mesure notée μ_f est définie par la relation suivante

$$\forall B \in \mathcal{B}, \mu_f(B) = \mu(f^{-1}(B)).$$

La mesure μ_f ainsi définie sur l'espace mesurable (E, \mathcal{B}) porte le nom de *mesure image* de μ par f .

Lorsque $\mu = P$ est une mesure de probabilité et $f = X$ une variable aléatoire à valeurs dans $(E, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, la mesure image notée P_X est entièrement caractérisée par sa fonction de répartition

$$\forall t \in \mathbb{R}, F(t) = P_X(]-\infty, t]) = P(X \in]-\infty, t]) = P(X \leq t).$$

Il s'agit d'une conséquence du théorème (5.2.1) comme nous l'avons vu précédemment. Cette mesure image qui est donc une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est appelée *loi de probabilité de la variable X*. Nous reviendrons largement sur ces lois de probabilité au chapitre 4.

5.4 Construction d'une intégrale

Soit Ω un espace muni d'une tribu \mathcal{A} et d'une mesure μ . Nous souhaitons définir une notion d'intégrale pour une application mesurable à valeurs dans un espace (E, \mathcal{B}) . Cette notion prendra un sens précis en calcul des probabilités. Lorsque $\mu = P$ est une mesure de probabilité et X une variable aléatoire, nous souhaitons définir la valeur moyenne de la variable X à l'issue de l'épreuve. Cette valeur moyenne doit se calculer intuitivement comme une intégrale de la variable X sous toutes les éventualités de l'épreuve, pondérée par la probabilité de réalisation de chaque éventualité

$$\int_{\Omega} X(\omega) dP(\omega).$$

Tout d'abord nous allons voir comment intégrer des fonctions élémentaires : les indicatrices d'ensembles mesurables. Ensuite, nous intégrerons les combinaisons linéaires de telles indicatrices et enfin n'importe quelle fonction positive par un procédé de limite croissante.

5.4.1 Intégrale d'une fonction indicatrice d'événement

Soit $A \in \mathcal{A}$, on définit l'intégrale de la fonction $\mathbb{1}_A$ de la manière suivante

$$\int_{\Omega} \mathbb{1}_A(\omega) d\mu(\omega) =_{conv} \int_{\Omega} \mathbb{1}_A d\mu =_{conv} \int_A d\mu =_{def} \mu(A).$$

Ainsi lorsque $A, B \in \mathcal{A}$ sont deux événements disjoints, on a

$$\int_{\Omega} \mathbb{1}_{A \cup B} d\mu = \mu(A \cup B) = \mu(A) + \mu(B) = \int_{\Omega} \mathbb{1}_A d\mu + \int_{\Omega} \mathbb{1}_B d\mu .$$

Or nous avons

$$\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B .$$

On obtient ainsi

$$\int_{\Omega} \mathbb{1}_A + \mathbb{1}_B d\mu = \int_{\Omega} \mathbb{1}_A d\mu + \int_{\Omega} \mathbb{1}_B d\mu .$$

L'objet ainsi défini sur les fonctions indicatrices possède donc la propriété de linéarité habituelle de l'intégrale

$$\int_{\Omega} f + g d\mu = \int_{\Omega} f d\mu + \int_{\Omega} g d\mu .$$

5.4.2 Intégrale d'une fonction mesurable positive

Soit $\alpha_1, \dots, \alpha_n$ des réels **positifs** et $A_1, \dots, A_n \in \mathcal{A}$, n événements. Posons

$$\forall \omega \in \Omega, \quad u_n(\omega) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(\omega) .$$

D'après le paragraphe précédent, il est naturel de poser

$$\int_{\Omega} u_n d\mu = \sum_{i=1}^n \alpha_i \mu(A_i) .$$

car nous souhaitons obtenir la propriété de linéarité de l'intégrale. D'après la proposition (5.3.2), toute fonction mesurable positive f est limite croissante de fonctions étagées positives. Notons $(u_n)_{n \geq 1}$ cette suite, nous avons

$$f(\omega) = \lim_{n \rightarrow \infty} u_n(\omega) .$$

Définition 5.4.1 *On appelle intégrale d'une fonction positive f par rapport à la mesure μ la limite*

$$\int_{\Omega} f d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} u_n d\mu$$

où (u_n) est la suite de fonctions étagées définies précédemment.

5.4.3 Fonctions intégrables

On sait désormais définir l'intégrale d'une fonction mesurable positive. Nous souhaitons définir la notion d'intégrabilité d'une fonction mesurable quelconque.

Définition 5.4.2 Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f une fonction mesurable. On dit que f est intégrable si

$$\int_{\Omega} |f| d\mu < \infty .$$

Dans ce cas, on a

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

où $f^+ = \max(f, 0)$ et $f^- = \min(f, 0)$.

Lorsque f est mesurable, sa partie positive et sa partie négative le sont aussi. Par conséquent, il en va de même de la valeur absolue de f . L'intégrale $\int_{\Omega} |f| d\mu$ est bien définie et vaut éventuellement $+\infty$. On dit que f est intégrable si l'intégrale précédente est finie.

5.5 Théorème de convergence monotone

L'intérêt de l'intégrale définie dans le paragraphe précédent est qu'elle tolère un certain nombre d'opérations de limites qui sont très utiles en pratique.

Théorème 5.5.1 Théorème de convergence monotone Soit $(f_n)_{n \geq 1}$ une suite croissante de fonctions mesurables positives. Soit f la limite simple de cette suite (en supposant qu'elle existe). Alors

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

Démonstration. Pour tout $n \geq 1$, on a $f_n \leq f$ et donc

$$\int f_n d\mu \leq \int f d\mu .$$

Par suite, nous avons

$$\sup_n \int f_n d\mu = \lim_n \int f_n d\mu \leq \int f d\mu .$$

Pour démontrer l'inégalité inverse, on choisit une fonction étagée positive que l'on note u telle que

$$u \leq f .$$

Soit $0 < \epsilon < 1$ et

$$\forall n \geq 1, \quad E_n = \{\omega \in \Omega : \epsilon u(\omega) \leq f_n(\omega)\} .$$

Les sous ensembles E_n sont mesurables et

$$E_1 \subset E_2 \subset \dots \subset E_n \subset \dots$$

De plus, il est facile de vérifier que

$$\bigcup_{n \geq 1} E_n = \Omega .$$

Alors,

$$\int f_n d\mu \geq \int_{E_n} f_n d\mu \geq \epsilon \int_{E_n} u d\mu .$$

Donc

$$\sup_n \int f_n d\mu \geq \epsilon \int_{E_n} u d\mu \quad \forall n \geq 1 ; .$$

Or, la fonction u s'écrit

$$\forall \omega \in \Omega , \quad u(\omega) = \sum_{i=1}^m \alpha_i \mathbb{1}_{A_i}$$

où m est un entier et les A_i sont des ensembles mesurables. Nous avons donc

$$\int_{E_n} u d\mu = \sum_{i=1}^m \alpha_i \int_{E_n} \mathbb{1}_{A_i} d\mu = \sum_{i=1}^m \alpha_i \mu(A_i \cap E_n) .$$

Après application de la proposition (5.3.2) de passage à la limite croissante pour la mesure μ , nous obtenons

$$\sup_n \int_{E_n} u d\mu = \int u d\mu .$$

Finalement, puisque ϵ est choisi arbitrairement,

$$\sup_n \int f_n d\mu \geq \int u d\mu \quad \Rightarrow \quad \sup_n \int f_n d\mu \geq \sup_{u \leq f} \int u d\mu = \int f d\mu .$$

■

Commentaires. Le théorème (5.5.1) joue un grand rôle dans la théorie de l'intégration. En fait, le théorème de convergence monotone va permettre de généraliser à toute fonction mesurable positive, puis à toute fonction intégrable des propriétés qui sont vérifiées par les fonctions indicatrices

Proposition 5.5.1 Soient f et g deux fonctions mesurables positives et $\alpha \in \mathbb{R}$. Alors

$$a) \int (f + g) d\mu = \int f d\mu + \int g d\mu$$

$$b) \int \alpha f d\mu = \alpha \int f d\mu$$

Démonstration. Nous avons vu dans le paragraphe précédent que la linéarité était vérifiée pour les fonctions indicatrices. Il en va de même pour les fonctions étagées. Prenons deux suites de fonctions étagées (u_n) et (v_n) croissantes, positives qui convergent vers f et g respectivement. D'après le théorème 5.5.1, nous avons

$$\int f + g d\mu = \lim_n \int u_n + v_n d\mu = \lim_n \int u_n d\mu + \lim_n \int v_n d\mu = \int f d\mu + \int g d\mu .$$

■

Proposition 5.5.2 Soit $(f_n)_{n \geq 1}$ une suite de fonctions mesurables positives. Alors,

$$\int \sum_{n \geq 1} f_n d\mu = \sum_{n \geq 1} \int f_n d\mu .$$

Démonstration. Exercice. Utiliser le théorème (5.5.1). ■

5.6 Calcul d'intégrales

5.6.1 L'intégrale de Lebesgue sur \mathbb{R}

Dans la plupart des cas, il n'est pas possible de donner une expression simple de l'intégrale d'une fonction mesurable positive pour la tribu des boréliens de \mathbb{R} , munie de la mesure de Lebesgue λ . Nous allons cependant remarquer que dans le cas d'une fonction positive continue sur un intervalle $[a, b]$, l'intégrale de Lebesgue sur l'intervalle $[a, b]$ coïncide avec l'intégrale au sens classique (*i.e.*, *au sens de Riemann*)

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

En effet, l'intégrale classique d'une fonction continue sur un intervalle $[a, b]$ est définie comme

$$\int_a^b f(x) dx = \lim_n s_n \quad \text{où} \quad s_n = \sum_{k=0}^{2^n-1} \frac{(b-a)}{2^n} \inf_{x \in I_k} f(x)$$

avec, pour tout $k = 0, \dots, 2^n - 1$,

$$I_k = \left[a + k \frac{b-a}{2^n}, a + (k+1) \frac{b-a}{2^n} \right[.$$

Définissons, pour tout $n \geq 1$,

$$f_n = \sum_{k=0}^{2^n-1} \mathbb{1}_{I_k} \inf_{x \in I_k} f(x) .$$

Les fonctions positives mesurables f_n convergent en croissant vers f qui est mesurable pour la tribu de Borel. D'après le théorème de convergence monotone

$$\int_a^b f(x) dx = \lim_n s_n = \lim_n \int f_n d\lambda = \int_{[a,b]} f d\lambda .$$

■

Bien entendu, il est aussi possible de calculer l'intégrale de fonctions non continues ayant un cardinal fini ou infini dénombrable de discontinuités. L'exemple suivant est célèbre.

Exemple 5.6.1 (*Fonction de Dirac*) Soit \mathcal{Q} l'ensemble des nombres rationnels dans \mathbb{R} . Montrer que la fonction indicatrice $\mathbb{1}_{\mathcal{Q}}$ est mesurable pour la tribu de Borel et calculer son intégrale de Lebesgue. Cette fonction est-elle intégrable au sens classique ?

5.6.2 Intégrale par rapport à une mesure discrète

Proposition 5.6.1 *Masse de Dirac.* Soit (Ω, \mathcal{A}) un espace mesurable et f une fonction mesurable positive. Soit ω_0 un point fixé de Ω et δ_{ω_0} la masse de Dirac au point ω_0 . Alors

$$\int f d\delta_{\omega_0} = f(\omega_0) .$$

Démonstration. Soit $A \in \mathcal{A}$. On a

$$\delta_{\omega_0}(A) = \int \mathbb{1}_A d\delta_{\omega_0} = \mathbb{1}_A(\omega_0) .$$

Par conséquent, la proposition peut être établie pour toute fonction indicatrice et par suite pour toute fonction étagée. Considérons ensuite une suite (u_n) croissante de fonctions étagées positives qui converge vers f . D'après le théorème de convergence monotone,

$$\int f d\delta_{\omega_0} = \lim_{n \rightarrow \infty} \int u_n d\delta_{\omega_0} = \lim_{n \rightarrow \infty} u_n(\omega_0) = f(\omega_0) .$$

■

Soit maintenant une suite (ω_n) de points de Ω . Considérons la mesure discrète définie par

$$\mu = \sum_{n \geq 1} p_n \delta_{\omega_n}$$

où les p_n sont des réels positifs.

Proposition 5.6.2 Mesure discrète. *Soit f une fonction mesurable positive et μ la mesure discrète ci-dessus. Alors*

$$\int_{\Omega} f d\mu = \sum_{n \geq 1} p_n f(\omega_n) .$$

Commentaires. En d'autres termes, intégrer une fonction mesurable par rapport à une mesure discrète consiste simplement à effectuer la somme pondérée des valeurs que prend la fonction f aux points chargés par la mesure.

A travers les exemples précédents se dégage l'intérêt pratique de la théorie de la mesure. Cette théorie propose un formalisme unique pour traiter de sommes et d'intégrales. Ainsi les théorèmes que nous énonçons dans ce chapitre sont valables dans les deux contextes.

5.6.3 Mesures définies par des densités

Proposition 5.6.3 *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f une fonction mesurable positive. L'application ν de \mathcal{A} dans $\overline{\mathbb{R}}_+$ qui associe à tout ensemble mesurable A la valeur*

$$\nu(A) = \int_A f d\mu$$

est une mesure positive dite mesure de densité f par rapport à μ .

Démonstration. Exercice. ■

Pour toute fonction mesurable positive g , l'intégrale par rapport à la mesure ν se calcule de la manière suivante

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

Ainsi, lorsque μ est une mesure discrète attribuant un poids $p_n = f(\omega_n)$ à tout point de la suite $(\omega_n)_{n \geq 1}$, nous avons

$$\int g d\nu = \sum_{n \geq 1} g(\omega_n) p_n .$$

En fait, toute mesure discrète est une mesure à densité pour une mesure de dénombrement

$$\mu = \sum_n \delta_{\omega_n} .$$

Si f est une fonction réelle positive continue qui satisfait

$$\int_{\mathbb{R}} f(x) d\lambda(x) = \int_{-\infty}^{+\infty} f(x) dx = 1 ,$$

la mesure ν de densité f par rapport à la mesure de Lebesgue sur \mathbb{R} est une mesure de probabilité. La fonction f est dite alors *densité de probabilité*. En conséquence du théorème de prolongement, la mesure ν est caractérisée par sa fonction de répartition

$$\forall t \in \mathbb{R} , \quad F(t) = \nu(] - \infty, t]) = \int_{-\infty}^t f(x) dx .$$

Dans ce cas, la fonction de répartition F est la primitive de la densité f qui s'annule en $-\infty$. Nous donnerons un grand nombre d'exemples de mesures à densité dans le chapitre 4.

5.6.4 L'intégrale d'une fonction mesurable à valeurs dans \mathbb{N}

Nous allons montrer dans ce paragraphe que la théorie établie permet de calculer des intégrales de manière explicite lorsque l'espace sur lequel on intègre est un espace probabilisé abstrait (Ω, \mathcal{A}, P) . Soit X une variable aléatoire (rappelons qu'une variable aléatoire n'est rien d'autre qu'une application mesurable définie sur Ω) à valeurs dans \mathbb{N} . Supposons connues les probabilités des événements $(X = n)$

$$\forall n \in \mathbb{N} \quad P(X = n) = p_n .$$

Nous souhaitons calculer l'intégrale

$$\int_{\Omega} X dP .$$

Remarquons que

$$X = \sum_{n \in \mathbb{N}} n \mathbb{1}_{(X=n)}$$

et appliquons la proposition 5.5.2

$$\int_{\Omega} X dP = \int_{\Omega} \sum_{n \in \mathbb{N}} n \mathbb{1}_{(X=n)} dP = \sum_{n \in \mathbb{N}} n \int_{\Omega} \mathbb{1}_{(X=n)} dP .$$

Par définition, nous avons

$$\int_{\Omega} \mathbb{1}_{(X=n)} dP = P(X = n) = p_n .$$

Ainsi,

$$\int_{\Omega} X dP = \sum_{n \in \mathbb{N}} np_n .$$

Commentaires. Nous justifions par ce calcul l'intuition qui correspond à l'intégrale d'une variable aléatoire. Il s'agit de la moyenne de toutes les éventualités prises par X pondérées par les probabilités de ces éventualités. Pour les variables aléatoires à valeurs dans \mathbb{R} , nous énoncerons un résultat similaire dans le chapitre 4.

5.7 Théorème de convergence dominée

La puissance donnée par la construction de l'intégrale apparaît à travers les théorèmes suivants. Ces théorèmes permettent sous des hypothèses raisonnables d'invertir signes *somme* et *limite* et de dériver sous le signe *somme*. De telles opérations nécessitaient avec l'intégrale classique de lourdes justifications.

Théorème 5.7.1 Théorème de convergence dominée. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $(f_n)_{n \geq 1}$ une suite de fonctions mesurables. On suppose que

a) la suite (f_n) converge simplement vers une limite f , i.e.,

$$\forall \omega \in \Omega, \quad f_n(\omega) \rightarrow f(\omega) ;$$

b) il existe une fonction g intégrable telle que

$$\forall \omega \in \Omega, \quad |f_n(\omega)| \leq g(\omega) .$$

Alors f est intégrable et

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

Donnons un exemple d'application de ce théorème.

Exemple 5.7.1 Calculer la limite quand $n \rightarrow \infty$ de l'intégrale

$$\int_0^n \left(1 - \frac{x}{n}\right)^n dx .$$

Solution. Dans cet exemple $\Omega = \mathbb{R}$. Nous avons, pour tout $n \geq 1$,

$$\int_0^n \left(1 - \frac{x}{n}\right)^n dx = \int_{\mathbb{R}} \mathbb{1}_{[0,n]}(x) \left(1 - \frac{x}{n}\right)^n dx .$$

Nous appliquons le théorème de convergence dominée à la suite de fonctions définies par

$$\forall n \geq 1, \forall x \in \mathbb{R}, \quad f_n(x) = \mathbb{1}_{[0,n]}(x) \left(1 - \frac{x}{n}\right)^n.$$

La suite f_n converge simplement vers la fonction f définie par

$$\forall x \in \mathbb{R}, \quad f(x) = e^{-x} \mathbb{1}_{\mathbb{R}_+}(x).$$

La condition b) est satisfaite par la fonction $g = f$. En effet, pour tout $0 \leq x \leq n$,

$$n \ln\left(1 - \frac{x}{n}\right) + x = -\left(\frac{x^2}{2n} + \frac{x^3}{3n^2} + \dots\right) \leq 0$$

et donc

$$\left(1 - \frac{x}{n}\right)^n = e^{n \ln\left(1 - \frac{x}{n}\right)} \leq e^{-x}.$$

Ainsi, nous avons

$$f_n \leq f.$$

La fonction f est intégrable et finalement

$$\lim_n \int_0^n \left(1 - \frac{x}{n}\right)^n dx = \int_{\mathbb{R}_+} e^{-x} dx = 1.$$

■

Les deux résultats suivants sont des conséquences du théorème de convergence dominée.

Théorème 5.7.2 *Intégrales dépendant d'un paramètre.* Soit I un intervalle de \mathbb{R} et une famille de fonctions mesurables $\{f(\cdot, t), t \in I\}$ dépendant d'un paramètre t pris dans I . On suppose que

a) $\forall \omega \in \Omega$, l'application

$$\begin{aligned} I &\rightarrow \mathbb{R} \\ t &\rightarrow f(\omega, t) \end{aligned}$$

est continue en tout point de l'intervalle I .

b) Il existe une fonction intégrable g telle que

$$\forall t \in I, \forall \omega \in \Omega, \quad |f(\omega, t)| \leq g(\omega).$$

Alors l'application

$$t \rightarrow \int_{\Omega} f(\omega, t) d\mu(\omega)$$

est continue en tout point de I .

Théorème 5.7.3 *Dérivation sous le signe somme.* Soit I un intervalle de \mathbb{R} et $\{f(\cdot, t), t \in I\}$ une famille de fonctions mesurables dépendant d'un paramètre t pris dans I . On suppose que

a) $\forall \omega \in \Omega$, l'application

$$\begin{aligned} I &\rightarrow \mathbb{R} \\ t &\rightarrow f(\omega, t) \end{aligned}$$

est dérivable en tout point de l'intervalle I .

b) Il existe une fonction intégrable g telle que

$$\forall t \in I, \forall \omega \in \Omega, \quad \left| \frac{\partial}{\partial t} f(\omega, t) \right| \leq g(\omega).$$

Alors l'application

$$t \rightarrow \int_{\Omega} f(\omega, t) d\mu(\omega)$$

est dérivable en tout point de I et

$$\frac{\partial}{\partial t} \int_{\Omega} f(\omega, t) d\mu(\omega) = \int_{\Omega} \frac{\partial}{\partial t} f(\omega, t) d\mu(\omega).$$

5.8 Mesures produit - Théorème de Fubini

Nous aurons besoin, dans la suite du cours d'intégrer sur des espaces constitués de produits d'espaces élémentaires ($\mathbb{R} \times \mathbb{R}$ par exemple) et de définir des intégrales multiples.

Soit $(\Omega_1, \mathcal{A}_1)$ et $(\Omega_2, \mathcal{A}_2)$ deux espaces mesurables. On considère le produit cartésien $\Omega_1 \times \Omega_2$. Dans ce produit, l'ensemble des parties

$$\{A_1 \times A_2 \mid A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$$

forme un semi-anneau. On appellera *tribu produit* de \mathcal{A}_1 et \mathcal{A}_2 la tribu notée $\mathcal{A}_1 \otimes \mathcal{A}_2$ engendrée par ce semi-anneau.

Définition 5.8.1 Soit $(\Omega_1, \mathcal{A}_1, \mu_1)$ et $(\Omega_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés. On appelle *mesure produit* de μ_1 et μ_2 la mesure notée $\mu_1 \otimes \mu_2$ définie sur la tribu $\mathcal{A}_1 \otimes \mathcal{A}_2$ par

$$\forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2, \quad \mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2).$$

Cette définition fait appel au théorème de prolongement (5.2.1). Bien entendu, l'exemple fondamental qui nous concerne est celui de la mesure de Lebesgue dans \mathbb{R}^n . La tribu produit $\mathcal{B}(\mathbb{R})^{\otimes n}$ est égale à la tribu de Borel $\mathcal{B}(\mathbb{R}^n)$. La mesure de Lebesgue sur \mathbb{R}^n (volume) peut être vue comme le produit de n mesures de Lebesgue sur \mathbb{R} . L'intégrale par rapport à une mesure produit se définit grâce au théorème suivant qui est aussi très utile pour faire des calculs.

Théorème 5.8.1 Théorème de Fubini. Soit $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$ un espace produit mesuré et f une fonction mesurable. On suppose

$$g(\omega_1) = \int_{\Omega_2} |f(\omega_1, \omega_2)| d\mu_2(\omega_2) \quad \mu_1 - \text{intégrable}$$

ou bien

$$g(\omega_2) = \int_{\Omega_1} |f(\omega_1, \omega_2)| d\mu_1(\omega_1) \quad \mu_2 - \text{intégrable.}$$

Alors f est $\mu_1 \otimes \mu_2$ -intégrable. De plus,

$$\int_{\Omega_1 \times \Omega_2} f d\mu_1 \otimes \mu_2 = \int_{\Omega_1} \left(\int_{\Omega_2} f d\mu_1 \right) d\mu_2 = \int_{\Omega_2} \left(\int_{\Omega_1} f d\mu_2 \right) d\mu_1 .$$

Commentaires. En clair, pour que la fonction f soit intégrable il suffit que l'une des intégrales doubles

$$\int_{\Omega_1} \left(\int_{\Omega_2} |f| d\mu_1 \right) d\mu_2 \quad \text{ou} \quad \int_{\Omega_2} \left(\int_{\Omega_1} |f| d\mu_2 \right) d\mu_1$$

soit finie. Dans ce cas, on pourra intervertir l'ordre des intégrations. Le théorème de Fubini permet donc de permuter les signes d'intégration sous hypothèse que la fonction soit intégrable pour la mesure produit.

Toutefois, l'exemple suivant montre qu'il est dangereux de permuter les intégrales sans prendre la précaution de vérifier les hypothèses du théorème. Munissons le carré $[0, 1]^2$ de la mesure de Lebesgue et définissons la fonction f sur $[0, 1]^2$ de la manière suivante

$$\forall (x, y) \in [0, 1]^2, \quad f(x, y) = \frac{x^2 - y^2}{(x^2 + y^2)^2} .$$

Après calculs, nous obtenons, pour tout $x \neq 0$,

$$\int_0^1 f(x, y) dy = \frac{1}{x^2 + 1}$$

et, puisque le segment $x = 0$ est de mesure nulle

$$\int_0^1 \int_0^1 f(x, y) dy dx = \int_0^1 \frac{1}{x^2 + 1} dx = [\arctg(x)]_0^1 = \pi/4 .$$

De plus, nous avons

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

et donc

$$\int_0^1 \int_0^1 f(x, y) dx dy = -\pi/4 .$$

Ceci est contraire aux conclusions du théorème de Fubini. La fonction f n'est donc pas intégrable sur $[0, 1]^2$!

■

5.9 Perspectives probabilistes

Avant de poursuivre ce texte et débiter avec les notions probabilistes, il est sans doute utile de donner en quelques grandes lignes une perspective de l'apport du présent chapitre pour les probabilités. Nous avons vu, lors du chapitre 1, que tout raisonnement en calcul des probabilités se construit sur la base d'un modèle formel. Il s'agit de la donnée d'un triplet abstrait (Ω, \mathcal{A}, P) . Cette structure ne traduit rien d'autre que des événements sont susceptibles de se réaliser et que l'on est capable d'en mesurer les probabilités. Toutefois, en pratique, on est capable d'observer des variables dont la valeur dépend du hasard. Formellement, ces variables sont des applications qui, à toute réalisation $\omega \in \Omega$ associent un résultat à valeurs dans un ensemble E . Puisque le hasard intervient, il est naturel de munir l'ensemble des résultats E d'une tribu \mathcal{B} . La tribu \mathcal{B} décrit l'ensemble des événements susceptibles d'être observés lors de la réalisation d'une variable aléatoire à valeurs dans E . Soit X une variable aléatoire et B un événement "résultat". Il est raisonnable de supposer que $(X \in B)$ est un événement de Ω et donc que $X^{-1}(B) \in \mathcal{A}$. Nous venons en fait de supposer que X était une application mesurable de (Ω, \mathcal{A}) à valeurs dans (E, \mathcal{B}) (voir définition 3.3.1). Montrons maintenant comment apparait naturellement la notion de mesure image. Nous examinons tout d'abord le cas simple d'une variable qui ne prend qu'un nombre fini de valeurs puis le cas d'une variable "continue" pour laquelle il sera nécessaire d'invoquer le théorème de prolongement (3.2.1). Prenons d'abord l'exemple du lancer d'un dé équilibré. Dans ce cas, nous avons

$$E = \{1, \dots, 6\}, \quad \mathcal{B} = \mathcal{P}(E)$$

et X désigne le résultat du lancer. Nous prenons pour hypothèse que les sous-ensembles de Ω égaux à $(X = k)$, ($k = 1, 6$) sont des éléments de \mathcal{A} . Nous posons alors

$$\forall k, \quad \mu(k) = P(X = k)$$

et

$$\forall B \in \mathcal{B}, \quad \mu(B) = \sum_{k \in B} \mu(k).$$

Nous venons ainsi de définir une mesure μ sur (E, \mathcal{B}) qui n'est autre que la mesure image de P par X . On dit que la variable X effectue un transfert de structure. Bien entendu, tous les calculs de probabilité peuvent se faire indifféremment dans l'un ou l'autre des modèles (Ω, \mathcal{A}, P) ou (E, \mathcal{B}, μ) . Réciproquement, introduire un modèle qui n'est pas abstrait revient à considérer qu'une variable aléatoire est observée. Ceci explique que l'on ne précise rarement le modèle utilisé en probabilité : l'espace abstrait est sous-entendu et on utilise des variables aléatoires. Considérons maintenant un nombre U pris au hasard dans $E = (0, 1)$. Bien entendu, nous allons chercher à ce que U définisse une application mesurable. Ainsi les calculs de probabilité concernant ce nombre pourront être fait (par transfert de structure) sur E . Dans ce cas, le problème se pose de définir \mathcal{B} et μ . Nous souhaitons avoir

$$\forall [a, b] \subset E, \quad P(U \in [a, b]) = \mu([a, b]) = b - a.$$

Ceci suppose que $[a, b] \in \mathcal{B}$. La tribu \mathcal{B} doit contenir tout les intervalles. La plus petite qui convient est la tribu engendrée par

$$\mathcal{S} = \{]a, b] ; a < b\}$$

(qui n'est pas lui-même une tribu!). Afin de définir la mesure μ , il est utile d'invoquer le théorème de prolongement. Ce dernier garantit l'existence et l'unicité d'une mesure sur (E, \mathcal{B}) lorsque l'on ne connaît sa valeur que sur \mathcal{S} . Dans le cas de la variable U , la mesure image ainsi définie est la mesure de Lebesgue sur $(0, 1)$

$$\forall B \in \mathcal{B}, \quad \mu(B) = P(U \in B) = \lambda(B) .$$

Nous terminons ce survol en attirant l'attention sur l'importance de l'intégrale construite lors du chapitre écoulé. Nous allons être à même de définir la valeur moyenne d'une variable X sur toutes les éventualités d'une épreuve. Symboliquement, cette intégrale se note

$$\int_{\Omega} X(\omega) dP(\omega)$$

ou

$$\int_E x d\mu_X(x) .$$

Nous établirons dans le chapitre à venir l'égalité de ces deux expressions (modulo existence).