SCIENCES
SORBONNE
UNIVERSITÉ

Probabilités Approfondies

Thierry Lévy

Master de mathématiques et applications
UE UM4MA311 – Automne 2025

Version du 13 septembre 2025

Table des matières

Introduction

Ces notes sont celles du cours intitulé *Probabilités approfondies* et enseigné au premier semestre de l'année de Master 1 de mathématiques à Sorbonne Université. Cet enseignement donne lieu à 18 séances de 2 heures de cours, et 24 séances de 2 heures de travaux dirigés.

Ce cours se compose de trois parties de tailles comparables. La première est consacrée à la notion d'espérance conditionnelle, qui marque une sorte de frontière entre le contenu d'un premier cours de probabilités et une étude plus avancée de ce domaine.

Les deux autres parties sont largement indépendantes entre elles, mais directement subordonnées à la première. Il s'agit d'abord d'une étude des martingales (en temps discret), qui sont des suites de variables aléatoires réelles qui jouent, dans l'étude des systèmes dynamiques aléatoires, le rôle que jouent les quantités conservées dans l'étude des systèmes dynamiques déterministes.

La troisième et dernière partie est consacrée à l'étude des chaînes de Markov (en temps discret et à espace d'états discret). Il s'agit d'un modèle d'évolution aléatoire naturel, à la fois suffisamment souple pour pouvoir servir, au moins en première intention, de modèle à de nombreux phénomènes aléatoires, et suffisamment rigide pour donner lieu à une étude mathématique riche et détaillée.

Ce cours a été conçu en faisant l'hypothèse que vous ayez déjà étudié au moins une fois la théorie de la mesure et de l'intégration, et que vous ayez déjà suivi un cours de probabilités qui s'appuie explicitement sur la théorie de la mesure.

Concernant la théorie de la mesure, il est souhaitable que vous sachiez ce que sont : une tribu sur un ensemble et en particulier la tribu borélienne sur l'ensemble des nombres réels, une mesure sur un espace mesurable, une fonction mesurable sur un espace mesurable, l'intégrale d'une fonction mesurable positive par rapport à une mesure, une fonction intégrable sur un espace mesuré, l'intégrale d'une fonction intégrable ; et que vous ayez déjà étudié : pour les fonctions mesurables positives, le théorème de convergence monotone, le lemme de Fatou, l'inégalité de Hölder, et pour les fonctions intégrables le théorème de convergence dominée, et le théorème de Fubini.

Concernant les probabilités, il est souhaitable que vous sachiez ce que sont : un espace de probabilités, une variable aléatoire, sa loi, son espérance quand elle existe, que vous connaissiez l'inégalité de Markov, que vous connaissiez les notions usuelles de convergence de variables aléatoires (presque sûre, en moyenne, et en probabilités), et que vous ayez déjà étudié la notion d'indépendance.

Des notions un peu plus avancées comme le lemme de Borel–Cantelli, le théorème de la classe monotone, l'étude des produits d'espaces mesurés, seront étudiées ou revues pendant les premières séances de travaux dirigés. Certaines de ces notions sont aussi abordées dans les annexes de ce cours.

Des informations pratiques sur le déroulement et l'avancement du cours se trouvent sur la page qui lui est consacrée, à l'adresse

`https://www.lpsm.paris/users/levyt/probas_approfondies`

Première partie

Espérance conditionnelle

1 Définition dans le cas positif

1.1 Notations et conventions

Les termes **en couleur et en gras** dans ce paragraphe indiquent des notions dont il est souhaitable que vous les ayez déjà étudiées sérieusement avant d'aborder ce cours. Si ce n'est pas le cas pour certaines d'entre elles, ou si vous avez des doutes à leur sujet, prenez le temps, au début du semestre, de les étudier ou de les réviser.

Dans ce cours, nous écrirons souvent ∞ pour désigner $+\infty$. Par contre, $-\infty$ sera toujours noté $-\infty$. Les opérations arithmétiques (addition et multiplication) sur les réels s'étendent à l'ensemble $[0,\infty]$ des réels positifs étendu (pour lequel nous n'utiliserons pas de notation particulière). La seule convention sur laquelle on puisse avoir un doute concerne la valeur à donner à $\infty \times 0$, et dans le cadre de la théorie de la mesure, la convention appropriée est $\infty \times 0 = 0$.[1]

L'ensemble $\mathbb{R}$ des réels et l'intervalle $[0,\infty]$ seront toujours munis de leurs **tribus boréliennes** respectives, que nous noterons au besoin $\mathscr{B}_{\mathbb{R}}$ et $\mathscr{B}_{[0,\infty]}$. La tribu borélienne de $\mathbb{R}$, par exemple, est la **tribu engendrée** par la classe des parties ouvertes de $\mathbb{R}$. C'est aussi la tribu engendrée par la classe des intervalles.

Un espace mesurable est une paire $(\Omega, \mathscr{F})$ formée d'un ensemble Ω et d'une **tribu** $\mathscr{F}$ sur Ω. Un espace de probabilités est un triplet $(\Omega, \mathscr{F}, \mathbf{P})$ où $(\Omega, \mathscr{F})$ est un espace mesurable et $\mathbf{P}$ est une **mesure positive** de masse totale 1 sur $\mathscr{F}$.

Sur un espace de probabilités $(\Omega, \mathscr{F}, \mathbf{P})$, une **variable aléatoire positive** est une **fonction mesurable** de $(\Omega, \mathscr{F})$ dans $([0,\infty], \mathscr{B}_{[0,\infty]})$. Une **variable aléatoire réelle** est une fonction mesurable de $(\Omega, \mathscr{F})$ dans $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$.

Une construction d'importance fondamentale associe à toute variable aléatoire positive X sur un espace de probabilités $(\Omega, \mathscr{F}, \mathbf{P})$ son espérance, qui est un élément de $[0,\infty]$ noté $\mathbf{E}[X]$, et qui est définie comme son **intégrale** (au sens de Lebesgue) par rapport à la mesure $\mathbf{P}$.

Une variable aléatoire réelle Y sur $(\Omega, \mathscr{F}, \mathbf{P})$ n'admet, en revanche, pas toujours d'espérance. Le mieux qu'on puisse faire est de considérer la partie positive et la partie négative de Y, qui sont les variables aléatoires $Y^+ = \max(Y, 0)$ et $Y^- = \max(-Y, 0)$. Ce sont des variables aléatoires réelles positives, qui satisfont la relation $Y = Y^+ - Y^-$, et ce sont les plus petites variables aléatoires positives qui satisfont cette relation. On peut alors considérer $\mathbf{E}[Y^+]$ et $\mathbf{E}[Y^-]$ et, à la condition que ces nombres ne soient pas tous les deux infinis, poser

$$\mathbf{E}[Y] = \mathbf{E}[Y^+] - \mathbf{E}[Y^-].$$

L'hypothèse que l'un au moins des deux nombres $\mathbf{E}[Y^+]$ et $\mathbf{E}[Y^-]$ est fini (on dit parfois dans ce cas que Y *admet une intégrale*) est trop faible pour beaucoup d'usages, et on lui préfère la

1. On peut réfléchir à des manières d'étendre partiellement la soustraction et la division à $[0,\infty]$, mais les expressions $0/0$, ∞/∞ et $\infty - \infty$ n'auront, dans ce cours, jamais de sens. Leur apparition inattendue doit toujours constituer un signal d'alerte.

condition plus forte que les deux nombres $\mathbf{E}[Y^+]$ et $\mathbf{E}[Y^-]$ sont finis. Cette hypothèse équivaut à ce que leur somme soit finie, c'est-à-dire à ce que

$$\mathbf{E}\big[|Y|\big] = \mathbf{E}[Y^+] + \mathbf{E}[Y^-] < \infty.$$

On dit que Y est **intégrable** lorsque cette condition est vérifiée : Y est intégrable si l'espérance de la variable aléatoire positive $|Y|$ est finie.

On note $\mathscr{L}^1(\Omega, \mathscr{F}, \mathbf{P})$ l'ensemble des variables aléatoires réelles intégrables sur $(\Omega, \mathscr{F}, \mathbf{P})$. C'est un sous-espace vectoriel de l'espace vectoriel de toutes les fonctions réelles sur Ω. On note $L^1(\Omega, \mathscr{F}, \mathbf{P})$ le quotient de cet espace vectoriel par le sous-espace formé des variables aléatoires nulles $\mathbf{P}$-presque sûrement. Il y a beaucoup à dire sur **cet espace**, mais ce n'est pas l'objet de ce cours.

1.2 Définition et unicité

Comme pour l'intégrale, on peut définir l'espérance conditionnelle pour des variables aléatoires positives ou pour des variables aléatoires intégrables, et aucune des deux définitions de ne contient l'autre. Nous allons commencer par étudier le cas des variables aléatoires positives.

Définition 1.1. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soit X une variable aléatoire positive sur cet espace. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$.*
On appelle espérance conditionnelle de X sachant $\mathscr{G}$ *une variable aléatoire positive Y sur $(\Omega, \mathscr{F}, \mathbf{P})$ qui vérifie les deux propriétés suivantes :*

1. *Y est $\mathscr{G}$-mesurable,*
2. *pour tout $B \in \mathscr{G}$, on a $\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B]$.*

Cette définition ne garantit ni l'existence ni l'unicité d'une espérance conditionnelle d'une variable aléatoire sachant une sous-tribu de la tribu ambiante.

Commençons par discuter de l'unicité, qui repose sur le très joli résultat suivant, que je vous conseille vivement d'essayer de démontrer par vous-même avant d'en lire la démonstration.

Proposition 1.2. *Soient $X, Y : (\Omega, \mathscr{F}, \mathbf{P}) \to [0, \infty]$ deux variables aléatoires positives.*

1. *Si pour tout événement $A \in \mathscr{F}$ on a $\mathbf{E}[X\mathbf{1}_A] \leqslant \mathbf{E}[Y\mathbf{1}_A]$, alors $X \leqslant Y$ p.s.*
2. *Si pour tout événement $A \in \mathscr{F}$ on a $\mathbf{E}[X\mathbf{1}_A] = \mathbf{E}[Y\mathbf{1}_A]$, alors $X = Y$ p.s.*

Démonstration. 1. Nous voulons montrer que $\mathbf{P}(Y < X) = 0$. Commençons par écrire

$$\mathbf{P}(Y < X) = \mathbf{P}(\{Y < X\} \cap \{X < \infty\}) + \mathbf{P}(\{Y < X\} \cap \{X = \infty\})$$

et montrons que le deuxième terme du membre de droite est nul. Pour cela, écrivons

$$\{Y < X\} \cap \{X = \infty\} = \{Y < \infty\} \cap \{X = \infty\} = \bigcup_{n \geqslant 1} \big(\{Y \leqslant n\} \cap \{X = \infty\}\big)$$

et appliquons, pour tout entier $n \geqslant 1$, l'hypothèse à l'événement $A_n = \{Y \leqslant n\} \cap \{X = \infty\}$, qui appartient à $\mathscr{F}$. Nous trouvons

$$\infty \times \mathbf{P}(A_n) = \mathbf{E}[X\mathbf{1}_{A_n}] \leqslant \mathbf{E}[Y\mathbf{1}_{A_n}] \leqslant n\mathbf{P}(A_n),$$

et l'inégalité entre le premier et le dernier terme impose que $\mathbf{P}(A_n) = 0$.

Puisqu'une union dénombable d'événements négligeables est négligeable, nous avons montré que $\mathbf{P}(\{Y < X\} \cap \{X = \infty\}) = 0$, si bien que

$$\mathbf{P}(Y < X) = \mathbf{P}(\{Y < X\} \cap \{X < \infty\}).$$

Écrivons maintenant

$$\{Y < X\} \cap \{X < \infty\} = \bigcup_{n,m \geqslant 1} \{Y + \tfrac{1}{n} \leqslant X \leqslant m\}$$

et appliquons, comme nous l'avons fait plus haut, pour tous entiers $n, m \geqslant 1$, l'hypothèse à l'événement $B_{n,m} = \{Y + \frac{1}{n} \leqslant X \leqslant m\}$, qui appartient à $\mathscr{F}$. Nous obtenons

$$\mathbf{E}[Y\mathbf{1}_{B_{n,m}}] \geqslant \mathbf{E}[X\mathbf{1}_{B_{n,m}}] \geqslant \mathbf{E}[(Y + \tfrac{1}{n})\mathbf{1}_{B_{n,m}}] = \mathbf{E}[Y\mathbf{1}_{B_{n,m}}] + \tfrac{1}{n}\mathbf{P}(B_{n,m}).$$

Sur l'événement $B_{n,m}$, on a $Y \leqslant m$, donc $\mathbf{E}[Y\mathbf{1}_{B_{n,m}}] < \infty$. Ainsi, dans l'inégalité

$$\mathbf{E}[Y\mathbf{1}_{B_{n,m}}] \geqslant \mathbf{E}[Y\mathbf{1}_{B_{n,m}}] + \tfrac{1}{n}\mathbf{P}(B_{n,m})$$

que nous venons de démontrer, nous pouvons soustraire $\mathbf{E}[Y\mathbf{1}_{B_{n,m}}]$ à chaque membre pour trouver que $\frac{1}{n}\mathbf{P}(B_{n,m}) \leqslant 0$, donc $\mathbf{P}(B_{n,m}) = 0$.

Nous en concluons, en prenant l'union sur $n, m \geqslant 1$ des événements négligeables $B_{n,m}$, que

$$\mathbf{P}(Y < X) = 0,$$

ce qui conclut la démonstration.

2. L'hypothèse entraîne d'une part que $\mathbf{E}[X\mathbf{1}_A] \leqslant \mathbf{E}[Y\mathbf{1}_A]$ pour tout $A \in \mathscr{F}$, donc que $X \leqslant Y$ presque sûrement, et d'autre part que $\mathbf{E}[Y\mathbf{1}_A] \leqslant \mathbf{E}[X\mathbf{1}_A]$ pour tout $A \in \mathscr{F}$, donc que $Y \leqslant X$ presque sûrement. □

Il découle immédiatement de la deuxième assertion de cette proposition que deux espérances conditionnelles d'une même variable aléatoire positive sachant une même sous-tribu sont égales presque sûrement. Nous énoncerons ce fait en même temps que l'existence de l'espérance conditionnelle, dont nous nous occuperons plus tard (voir le théorème 2.1).

1.3 Premiers exemples

1.3.1 Le cas mesurable

Le cas le plus simple est celui où la variable aléatoire X est $\mathscr{G}$-mesurable. Dans ce cas, la variable X elle-même est une espérance conditionnelle de X sachant $\mathscr{G}$. En effet,

1. X est $\mathscr{G}$-mesurable, par hypothèse, et
2. pour tout $B \in \mathscr{G}$, il est bien vrai que $\mathbf{E}[X\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B]$.

Un sous-cas particulier de ce cas est celui où $\mathscr{G} = \mathscr{F}$, car alors toute variable aléatoire sur $(\Omega, \mathscr{F}, \mathbf{P})$, étant $\mathscr{F}$-mesurable par définition, est $\mathscr{G}$-mesurable.

1.3.2 Le cas indépendant

Un autre cas simple (mais un tout petit peu moins simple que le précédent) est celui où la variable aléatoire X est indépendante de la sous-tribu $\mathscr{G}$. Par définition, cela signifie que pour tout borélien $C \subseteq [0,\infty]$ et tout événement $B \in \mathscr{G}$, on a l'égalité

$$\mathbf{P}(\{X \in C\} \cap B) = \mathbf{P}(X \in C)\,\mathbf{P}(B).$$

La conséquence de cette indépendance qui nous importe est que pour tout événement $B \in \mathscr{G}$, on a l'égalité

$$\mathbf{E}[X\mathbf{1}_B] = \mathbf{E}[X]\mathbf{P}(B).$$

Ceci entraîne que la variable aléatoire constante $Y = \mathbf{E}[X]$ est une espérance conditionnelle de X sachant $\mathscr{G}$. En effet,

1. Y est $\mathscr{G}$-mesurable, parce qu'une variable aléatoire constante est mesurable par rapport à toute tribu sur Ω, et
2. pour tout $B \in \mathscr{G}$, on a $\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[X]\mathbf{P}(B) = \mathbf{E}[X\mathbf{1}_B]$.

Un sous-cas particulier de ce cas est celui où $\mathscr{G} = \{\varnothing, \Omega\}$. En effet, dans ce cas, toute variable aléatoire sur $(\Omega, \mathscr{F}, \mathbf{P})$ est indépendante de $\mathscr{G}$.

1.4 L'exemple fondamental

Beaucoup de notions mathématiques, même aussi raffinées que celle d'espérance conditionnelle (et raffinée, elle l'est, comme nous aurons amplement l'occasion de le voir), sont construites par élaborations successives à partir d'un ou quelques exemples d'importance fondamentale, en général assez simples. Le plus souvent, de par la nature même du discours mathématique qui tend vers la généralité et l'abstraction, ces exemples disparaissent complètement de la forme finale et très lisse de la définition de la notion. Néanmoins, leur connaissance et leur étude est la première étape indispensable à sa compréhension.

Pour l'espérance conditionnelle, l'exemple fondamental est celui où la sous-tribu $\mathscr{G}$ est engendrée par une partition finie ou dénombrable de Ω. Parmi les nombreuses choses que nous allons dire dans ce chapitre sur la notion d'espérance conditionnelle, le contenu de cette section est donc une des plus importantes, et il est indispensable de très bien le comprendre, et d'y revenir aussi souvent que nécessaire.

Cet exemple fondamental n'est heureusement pas le plus compliqué de la notion, il est même dans ce cas assez élémentaire, dans la mesure ou tout y est entièrement explicite, comme le montre un coup d'œil à la proposition 1.4.

Un excellente habitude à prendre consiste à éprouver, à tester, tout énoncé concernant l'espérance conditionnelle dans ce cas particulier très simple où la sous-tribu est engendrée par une partition finie.

1.4.1 Tribu engendée par une partition

Une *partition* d'un ensemble Ω est un ensemble de parties non vides deux à deux disjointes de cet ensemble, dont l'union est égale à Ω. Nous nous intéresserons à des partitions finies, de la forme $\{A_1, \ldots, A_n\}$, et à des partitions infinies dénombrables $\{A_1, A_2, \ldots\}$. Pour traiter ces deux cas d'un coup, et éviter d'écrire deux fois des formules quasiment identiques, nous noterons $(A_i)_{i\in I}$ une partition de Ω, en supposant l'ensemble d'indices I fini ou infini dénombrable. Nous appellerons les parties A_i de Ω les *blocs* de la partition.

Étant donné une partition $(A_i)_{i\in I}$ de Ω, on peut considérer la tribu qu'elle engendre :

$$\mathscr{G} = \sigma(A_i : i \in I).$$

Par définition, c'est la plus petite tribu sur Ω (plus petite au sens de l'inclusion) qui contient A_i pour tout $i \in I$.

L'exercice suivant est fondamental. Il dit, de manière précise, que la tribu sur Ω engendrée par une partition finie ou infinie dénombrable est l'ensemble des parties de Ω qui sont réunion d'un ensemble arbitraire de blocs de la partition. (Attention, cet énoncé est faux si la partition est infinie non dénombrable, c'est l'objet de la question 2 de l'exercice.)

Exercice 1.1. *1. Soit $(A_i)_{i\in I}$ une partition finie ou infinie dénombrable de Ω. Montrer que la tribu sur Ω engendrée par cette partition est*

$$\sigma(A_i : i \in I) = \Big\{ \bigcup_{j\in J} A_j : J \subseteq I \Big\}.$$

En particulier, si I est fini de cardinal n, alors cette tribu a 2^n éléments.

♡ *2. Soit $(A_i)_{i\in I}$ une partition quelconque de Ω. Montrer que*

$$\sigma(A_i : i \in I) = \Big\{ \bigcup_{j\in J} A_j : J \subseteq I, J \text{ dénombrable ou } I \setminus J \text{ dénombrable} \Big\}.$$

Il découle de cette description de la tribu engendrée par une partition que si un élément de cette tribu rencontre un des blocs de la partition (c'est-à-dire qu'il a avec ce bloc une intersection non vide), alors il inclut ce bloc. En symboles,

$$\forall B \in \sigma(A_i : i \in I), \forall i \in I, \;\; B \cap A_i \neq \varnothing \Rightarrow A_i \subseteq B.$$

Cette remarque va nous permettre de démontrer le résultat suivant.

Lemme 1.3. *Soit Ω un ensemble, $(A_i)_{i\in I}$ une partition finie ou dénombrable de Ω, et $\mathscr{G}$ la tribu sur Ω engendrée par cette partition. Soit $(E, \mathscr{E})$ un espace mesurable tel que pour tout $x \in E$, le singleton $\{x\}$ appartienne à la tribu $\mathscr{E}$.*
Une fonction $(\Omega, \mathscr{G}) \to (E, \mathscr{E})$ est mesurable si et seulement si elle est constante sur chacun des blocs de la partition.

Démonstration. Soit $Y : \Omega \to E$ une fonction qui est constante sur A_i pour chaque $i \in I$. Montrons que Y est mesurable relativement aux tribus $\mathscr{G}$ et $\mathscr{E}$.

Pour chaque $i \in I$, notons $y_i \in E$ la valeur de Y sur le bloc A_i. Soit B un élément de $\mathscr{E}$. Notons $J = \{i \in I : y_i \in B\}$. Alors

$$Y^{-1}(B) = \bigcup_{i\in J} A_i$$

appartient à $\mathscr{G}$, ce qu'il fallait démontrer.

Montrons maintenant la réciproque : considérons $Y : (\Omega, \mathscr{G}) \to (E, \mathscr{E})$ une fonction mesurable. Soit $i \in I$. Choisissons un élément $a_i \in A_i$ et posons $y_i = Y(a_i)$. Alors par hypothèse, le singleton $\{y_i\}$ appartient à $\mathscr{E}$. Puisque Y est mesurable, ceci entraîne que $Y^{-1}(\{y_i\})$ est un élément de $\mathscr{G}$. Cet élément de $\mathscr{G}$ a une intersection non vide avec A_i, donc il inclut A_i. Ainsi, Y est constante sur A_i, égale à y_i. Ceci est vrai pour tout $i \in I$, et la démonstration est terminée. □

1.4.2 Espérance conditionnelle sachant la tribu engendrée par une partition

Considérons un espace de probabilités $(\Omega, \mathscr{F}, \mathbf{P})$ et donnons-nous une partition $(A_i)_{i\in I}$ finie ou dénombrable de Ω. Faisons les hypothèses supplémentaires suivantes :

— pour tout $i \in I$, l'ensemble A_i appartient à $\mathscr{F}$,

— pour tout $i \in I$, on a $\mathbf{P}(A_i) > 0$.

Exercice 1.2. *Montrer que même si l'on n'avait rien supposé sur l'ensemble I, ces deux dernières hypothèses entraînent que I est fini ou dénombrable. (Indication : montrer que pour tout $n \geqslant 1$, il ne peut pas y avoir plus de n indices $i \in I$ pour lesquels $\mathbf{P}(A_i) \geqslant \frac{1}{n}$.)*

Considérons sur Ω la tribu

$$\mathscr{G} = \sigma(A_i : i \in I)$$

engendrée par la partition. C'est une sous-tribu de $\mathscr{F}$.

Donnons-nous une variable aléatoire positive X sur $(\Omega, \mathscr{F}, \mathbf{P})$ et cherchons-en une espérance conditionnelle sachant $\mathscr{G}$.

Tout d'abord, si une telle espérance conditionnelle existe, la première chose que nous dit la définition 1.1 est qu'elle est $\mathscr{G}$-mesurable. D'après le lemme 1.3, elle est donc constante sur chaque bloc de la partition. Elle doit donc être de la forme

$$Y = \sum_{i \in I} y_i \mathbf{1}_{A_i},$$

pour une certaine famille $(y_i)_{i \in I}$ d'éléments de $[0, \infty]$.

La deuxième partie de la définition 1.1 va nous permettre de déterminer ces éléments. En effet, considérons un certain $i \in I$. En appliquant la définition des l'espérance conditionnelle à $B = A_i$, nous trouvons

$$y_i \mathbf{P}(A_i) = \mathbf{E}[Y \mathbf{1}_{A_i}] = \mathbf{E}[X \mathbf{1}_{A_i}],$$

ce qui, puisque nous avons supposé que $\mathbf{P}(A_i)$ n'est pas nulle, nous donne

$$y_i = \frac{\mathbf{E}[X \mathbf{1}_{A_i}]}{\mathbf{P}(A_i)}.$$

Ainsi, toujours en supposant que l'espérance conditionnelle que nous cherchons existe, y_i doit être égal à la valeur moyenne de X sur A_i.

Une espérance conditionnelle de X sachant $\mathscr{G}$ ne peut donc être que la variable aléatoire

$$\sum_{i \in I} \frac{\mathbf{E}[X \mathbf{1}_{A_i}]}{\mathbf{P}(A_i)} \mathbf{1}_{A_i}.$$

Il ne reste plus qu'à vérifier que cette variable aléatoire est bien, comme nous nous en doutions, une espérance conditionnelle de X sachant $\mathscr{G}$.

Proposition 1.4. *Soit $(A_i)_{i \in I}$ une partition finie ou infinie dénombrable d'un espace de probabilités $(\Omega, \mathscr{F}, \mathbf{P})$ par des éléments de $\mathscr{F}$ de probabilité strictement positive. Soit $\mathscr{G}$ la sous-tribu de $\mathscr{F}$ engendrée par cette partition. Soit $X : (\Omega, \mathscr{F}) \to [0, \infty]$ une variable aléatoire positive.*
Alors la variable aléatoire

$$Y = \sum_{i \in I} \frac{\mathbf{E}[X \mathbf{1}_{A_i}]}{\mathbf{P}(A_i)} \mathbf{1}_{A_i}$$

est l'unique espérance conditionnelle de X sachant $\mathscr{G}$.

Démonstration. Nous savons que Y est le seul candidat possible à être une espérance conditionnelle de X sachant $\mathscr{G}$. Montrons que c'en est une.

Par construction, la variable aléatoire Y est constante sur chaque bloc de la partition que engendre $\mathscr{G}$, donc d'après le lemme 1.3, elle est $\mathscr{G}$-mesurable.

Donnons-nous maintenant un élément B de la tribu $\mathscr{G}$ et montrons que $\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B]$.

Soit $J \subseteq I$ la partie de I telle que

$$B = \bigcup_{i\in J} A_i.$$

Alors

$$\begin{aligned}
\mathbf{E}[Y\mathbf{1}_B] &= \sum_{j\in J} \mathbf{E}[Y\mathbf{1}_{A_j}] && \text{(additivité et convergence monotone [2])}\\
&= \sum_{j\in J} \mathbf{E}\left[\frac{\mathbf{E}[X\mathbf{1}_{A_j}]}{\mathbf{P}(A_j)}\mathbf{1}_{A_j}\right] && \text{(définition de } Y\text{)}\\
&= \sum_{j\in J} \mathbf{E}[X\mathbf{1}_{A_j}] = \mathbf{E}[X\mathbf{1}_B],
\end{aligned}$$

ce qui conclut la démonstration. □

1.4.3 Commentaires

La proposition 1.4 peut se paraphraser de la manière suivante : lorsque la sous-tribu $\mathscr{G}$ est engendrée par une partition (finie ou infinie dénombrable, par des éléments de $\mathscr{F}$ de probabilité strictement positive) de Ω, alors l'espérance conditionnelle d'une variable aléatoire se calcule en remplaçant, sur chaque bloc de la partition, la variable aléatoire par sa valeur moyenne sur ce bloc.

Prendre l'espérance conditionnelle s'apparente donc ici à une opération de moyennisation partielle, comme on calcule par exemple une moyenne semestre par semestre avant de calculer la moyenne annuelle correspondante.

Bien entendu, toutes les tribus ne sont pas engendrées par des partitions, si bien que la proposition 1.4 est loin d'épuiser le sujet. Il n'en reste pas moins que l'idée de moyenne partielle, ou de moyenne par bloc, est l'une des idées fondamentales (pas la seule, comme nous le verrons) de la notion d'espérance conditionnelle.

Exercice 1.3. *Dans ce qui précède, on a toujours supposé que les blocs des partitions étaient de probabilité strictement positive. Que devient la proposition 1.4 si l'on autorise certains des blocs de la suite $(A_n)_{n\geqslant 0}$ à être de probabilité nulle ?*

Exercice 1.4. *Est-il possible qu'une espérance conditionnelle d'une variable aléatoire positive qui est finie presque sûrement prenne la valeur ∞ avec probabilité strictement positive ? Autrement dit, avec les notations de la proposition 1.4, est-il possible que d'avoir $\mathbf{P}(X < \infty) = 1$ et $\mathbf{P}(Y = \infty) > 0$?*

1.5 Le cas d'une tribu engendrée par une variable aléatoire

Nous allons maintenant étudier le cas où la tribu $\mathscr{G}$ est engendrée par une variable aléatoire. C'est en fait un déguisement du cas général, au sens où toute tribu est engendrée par une variable aléatoire. Mais c'est une manière d'envisager l'espérance conditionnelle qui est à la fois fréquente et utile. Par contre, ce cas est trop général pour qu'on puisse y énoncer une proposition aussi parfaite que la proposition 1.4 : il n'y aura pas ici, il n'existe pas, de formule générale pour l'epérance conditionnelle.

2. On n'a ici besoin du théorème de convergence monotone seulement si la partie J est infinie.

Considérons un espace de probabilités $(\Omega, \mathscr{F}, \mathbf{P})$ et une variable aléatoire positive X définie sur cet espace. Donnons-nous une variable aléatoire $Z : (\Omega, \mathscr{F}) \to (E, \mathscr{E})$ à valeurs dans un espace mesurable $(E, \mathscr{E})$ arbitraire. Posons $\mathscr{G} = \sigma(Z)$, la plus petite tribu sur Ω qui rende la fonction Z mesurable. Nous allons chercher une espérance conditionnelle de X sachant $\mathscr{G}$.

Avant cela, observons que le cas d'une tribu engendrée par une partition est un cas particulier du cas que nous étudions maintenant.

Exercice 1.5. *Soit $(A_i)_{i\in I}$ une partition finie ou dénombrable de Ω par des éléments de $\mathscr{F}$. Soit $Z : (\Omega, \mathscr{F}) \to (I, \mathscr{P}(I))$ la fonction qui, pour tout $i \in I$, est constante égale à i sur le bloc A_i. Montrer que $\sigma(Z) = \sigma(A_i : i \in I)$.*

En fait, le cas d'une tribu engendrée par une partition est complètement général.

Exercice 1.6. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. On note $Z : (\Omega, \mathscr{F}) \to (\Omega, \mathscr{G})$ l'application identité de Ω. Montrer que $\sigma(Z) = \mathscr{G}$. Comparer cette construction à la précédente dans le cas où $\mathscr{G}$ est engendrée par une partition.*

Revenons à notre problème : chercher une espérance conditionnelle de X sachant $\sigma(Z)$. La première propriété d'une telle espérance conditionnelle est d'être mesurable par rapport à $\sigma(Z)$. La proposition suivante nous dit que ceci impose à cette espérance conditionnelle d'être une fonction de Z.

Je vous invite à réfléchir au fait, formulé ici de manière un peu vague, que pour une variable aléatoire, être une fonction de Z est l'analogue, dans la situation présente où $\mathscr{G} = \sigma(Z)$, de la propriété d'être constante sur chaque bloc dans le cas où $\mathscr{G}$ est engendrée par une partition.

Proposition 1.5. *Soit $Z : (\Omega, \mathscr{F}) \to (E, \mathscr{E})$ une application mesurable. Soit $Y : (\Omega, \sigma(Z)) \to ([0,\infty], \mathscr{B}_{[0,\infty]})$ une fonction mesurable. Alors il existe une fonction mesurable $h : (E, \mathscr{E}) \to ([0,\infty], \mathscr{B}_{[0,\infty]})$ telle que $Y = h(Z)$.*
De plus, si Y ne prend pas la valeur ∞, alors il est possible de choisir h qui ne prenne pas non plus la valeur ∞.

Démonstration. Pour démontrer cette proposition, nous allons utiliser l'écriture en base 2 d'un réel positif. Numérotons les "décimales" en base 2 par la puissance de 2 à laquelle elles correspondent :

$$\dots \underset{3}{1}\ \underset{2}{0}\ \underset{1}{1}\ \underset{0}{1},\ \underset{-1}{0}\ \underset{-2}{1}\ \underset{-3}{0} \dots$$

L'ensemble des réels positifs dont la 0-ième décimale, c'est-à-dire le chiffre juste à gauche de la virgule, dans l'écriture en base 2 est 1 est

$$[1,2[\cup [3,4[\cup \dots = \bigcup_{k\geqslant 0} [2k+1, 2k+2[.$$

Avec l'idée que l'écriture en base 2 de ∞ aurait tous ses chiffres égaux à 1, ajoutons ∞ à cet ensemble pour former

$$A = \{\infty\} \cup \bigcup_{k\geqslant 0} [2k+1, 2k+2[.$$

L'ensemble $2^n A$ est maintenant l'ensemble des réels positifs dont la n-ième décimale vaut 1, auquel on a ajouté ∞. On a donc, pour tout $x \in [0,\infty]$, l'égalité

$$x = \sum_{n\in\mathbb{Z}} 2^n \mathbf{1}_{2^n A}(x)$$

qui n'est autre, pour x réel positif, que l'écriture en base 2 de x.

Appliquons maintenant cette identité à la variable aléatoire Y. Nous trouvons

$$Y = \sum_{n\in\mathbb{Z}} 2^n \mathbf{1}_{2^n A}(Y).$$

Fixons un entier $n \in \mathbb{Z}$. La variable aléatoire $\mathbf{1}_{2^n A}(Y)$ est la fonction sur Ω qui vaut 1 en un point si l'image par Y de ce point appartient à $2^n A$, et 0 sinon. Elle est donc égale à l'indicatrice de $Y^{-1}(2^n A)$:

$$\mathbf{1}_{2^n A}(Y) = \mathbf{1}_{Y^{-1}(2^n A)}.$$

Or l'ensemble $2^n A$ est un borélien de $[0,\infty]$ et puisque Y est mesurable par rapport à la tribu $\sigma(Z)$, l'ensemble $Y^{-1}(2^n A)$ est un élément de $\sigma(Z)$. Il existe donc $C_n \in \mathscr{E}$ tel que

$$Y^{-1}(2^n A) = Z^{-1}(C_n).$$

Choisissons un tel $C_n \in \mathscr{E}$ pour tout $n \in \mathbb{Z}$. Nous obtenons l'écriture suivante de Y :

$$Y = \sum_{n\in\mathbb{Z}} 2^n \mathbf{1}_{Z^{-1}(C_n)} = \sum_{n\in\mathbb{Z}} 2^n \mathbf{1}_{C_n}(Z).$$

En définissant la fonction $h : (E,\mathscr{E}) \to ([0,\infty],\mathscr{B}_{[0,\infty]})$ par

$$h = \sum_{n\in\mathbb{Z}} 2^n \mathbf{1}_{C_n},$$

on a finalement écrit Y sous la forme $Y = h(Z)$, ce qui est ce qu'on souhaitait.

Montrons maintenant la deuxième assertion. Si Y ne prend pas la valeur ∞, il est malgré tout possible que notre construction nous fournisse une fonction h qui prend cette valeur. En effet, nous savons très peu de choses sur les ensembles C_n. Il n'est pas exclu, par exemple, qu'un point de E qui n'appartient pas à l'image de Z appartienne à tous les C_n. Dans ce cas, h est infinie en ce point, sans que cela n'empêche la relation $Y = h(Z)$ d'être vraie.

Supposons donc Y finie en tout point, et considérons une fonction h telle que $Y = h(Z)$. L'ensemble $\{h < \infty\}$ appartient à $\mathscr{E}$ et il contient l'image de Z, sans quoi Y prendrait une valeur infinie. Posons

$$\tilde{h} = h\mathbf{1}_{\{h<\infty\}}.$$

C'est une fonction mesurable de $(E,\mathscr{E})$ dans $[0,\infty]$ qui par construction est en fait à valeurs dans $[0,\infty[$. De plus, $\tilde{h}(Z) = h(Z) = Y$. La fonction $\tilde{h}$ convient donc. □

Exercice 1.7. *Dans la situation de l'exercice 1.5, comparer les descriptions données par le lemme 1.3 et le proposition 1.5 de ce qu'est une variable aléatoire $\sigma(Z)$-mesurable.*

Exercice 1.8. *Voici un énoncé purement ensembliste qui vous aidera peut-être à réfléchir à la remarque faite juste avant l'énoncé de la proposition 1.5. Il n'y a ici ni tribus ni mesure, que des ensembles et des fonctions, dont les noms sont choisis pour aider la comparaison.*

Soient Ω, E, R des ensembles. Soit $Z : \Omega \to E$ une fonction. Pour tout $e \in Z(\Omega)$, c'est-à-dire pour tout e dans l'image de Z, on pose

$$A_e = Z^{-1}(\{e\}).$$

On définit ainsi une partition $(A_e : e \in Z(\Omega))$ de Ω.

Soit maintenant $Y : \Omega \to R$ une fonction. Montrer que les deux assertions suivantes sont équivalentes.

1. *Il existe une fonction $h : E \to R$ telle que $Y = h \circ Z$.*

2. *Pour tout $e \in Z(\Omega)$, la fonction Y est constante sur A_e.*

Nous savons maintenant qu'il nous faut chercher une espérance conditionnelle de X sachant la tribu $\sigma(Z)$ sous la forme d'une fonction de Z, c'est-à-dire sous la forme $h(Z)$. Il reste donc à déterminer h.

S'il existait une formule qui nous donnait cette fonction h dans toutes les situations, il existerait une formule qui donne l'espérance conditionnelle d'une variable aléatoire sachant n'importe quelle sous-tribu de la tribu ambiante. Or il n'existe pas de telle formule.

Il y a en revanche, comme nous avons commencé et continuerons à le voir, un certain nombre de cas particuliers dans lesquels on peut écrire des formules, il y a des méthodes, plus ou moins générales, pour calculer des espérances conditionnelles, et il y a des règles formelles de calcul avec l'espérance conditionnelle.

Avant d'étudier tout cela, nous allons apporter une réponse générale à la question de l'existence de l'espérance conditionnelle.

2
Existence dans le cas positif

Le théorème principal de cette section est le suivant, et nous en donnerons deux démonstrations.

Théorème 2.1. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Pour toute variable aléatoire positive $X : \Omega \to [0, \infty]$ et toute sous-tribu $\mathscr{G}$ de $\mathscr{F}$, il existe une espérance conditionnelle de X sachant $\mathscr{G}$. De plus, deux espérances conditionnelles de X sachant $\mathscr{G}$ sont égales $\mathbf{P}$-presque sûrement.*

L'unicité est une conséquence de la définition de l'espérance conditionnelle et de la proposition 1.2. C'est l'existence qui va nous occuper dans ce qui suit.

2.1 L'espérance conditionnelle comme dérivée de Radon–Nikodym

La première démonstration repose sur le théorème de Radon–Nikodym, dont nous rappelons ci-dessous la partie de l'énoncé qui nous importe. L'énoncé qui suit n'est ni le plus complet ni le plus général de ce théorème.

Théorème 2.2 (Théorème de Radon–Nikodym)**.** *Soient μ et ν deux mesures finies sur un espace mesurable $(\Omega, \mathscr{G})$. Supposons que ν soit absolument continue par rapport à μ, c'est-à-dire que*

$$\forall B \in \mathscr{G},\ \mu(B) = 0 \implies \nu(B) = 0.$$

Alors il existe une fonction mesurable $f : (\Omega, \mathscr{G}) \to [0, \infty[$ telle que pour tout $B \in \mathscr{G}$, on ait

$$\nu(B) = \int_B f \, \mathrm{d}\mu.$$

La fonction f s'appelle la *dérivée de Radon–Nikodym* de ν par rapport à μ. Soulignons que le théorème assure que cette fonction ne prend pas la valeur ∞.

Ce théorème nous permet de démontrer l'existence d'une espérance conditionnelle pour des variables aléatoires d'espérance finie, ce qui constitue la première étape de notre première démonstration du théorème 2.1.

Proposition 2.3. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Pour toute variable aléatoire positive $X : \Omega \to [0, \infty]$ telle que $\mathbf{E}[X] < \infty$ et toute sous-tribu $\mathscr{G}$ de $\mathscr{F}$, il existe une espérance conditionnelle de X sachant $\mathscr{G}$. De plus, cette espérance conditionnelle est finie presque sûrement.*

Démonstration. Soit $X : \Omega \to [0, \infty]$ une variable aléatoire positive telle que $\mathbf{E}[X] < \infty$. Pour tout $B \in \mathscr{G}$, posons

$$\mathbf{Q}(B) = \mathbf{E}[X \mathbf{1}_B].$$

Alors $\mathbf{Q}$ est une mesure finie sur $(\Omega, \mathscr{G})$. En effet, pour tout $B \in \mathscr{G}$,

$$\mathbf{Q}(B) = \mathbf{E}[X\mathbf{1}_B] \geqslant 0 \text{ et } \mathbf{Q}(\Omega) = \mathbf{E}[X] < \infty.$$

De plus, $\mathbf{Q}(\varnothing) = 0$ et pour toute suite $(B_n)_{n\geqslant 0}$ d'éléments deux à deux disjoints de $\mathscr{G}$, on a

$$\begin{aligned}
\mathbf{Q}\Big(\bigcup_{n\geqslant 0} B_n\Big) &= \mathbf{E}\Big[X \sum_{n\geqslant 0} \mathbf{1}_{B_n}\Big] = \mathbf{E}\Big[\lim_{N\to\infty} X \sum_{n=0}^{N} \mathbf{1}_{B_n}\Big] \\
&= \lim_{N\to\infty} \mathbf{E}\Big[X \sum_{n=0}^{N} \mathbf{1}_{B_n}\Big] && \text{(convergence monotone)} \\
&= \lim_{N\to\infty} \sum_{n=0}^{N} \mathbf{E}\big[X\mathbf{1}_{B_n}\big] && \text{(additivité de l'espérance)} \\
&= \lim_{N\to\infty} \sum_{n=0}^{N} \mathbf{Q}(B_n) = \sum_{n\geqslant 0} \mathbf{Q}(B_n),
\end{aligned}$$

ce qui montre la σ-additivité de $\mathbf{Q}$.

Nous avons maintenant deux mesures finies $\mathbf{P}$ et $\mathbf{Q}$ sur $(\Omega, \mathscr{G})$. Montrons que $\mathbf{Q}$ est absolument continue par rapport à $\mathbf{P}$. Soit $B \in \mathscr{G}$ tel que $\mathbf{P}(B) = 0$. Alors la variable aléatoire $X\mathbf{1}_B$ est nulle $\mathbf{P}$-presque sûrement, donc $\mathbf{Q}(B) = \mathbf{E}[X\mathbf{1}_B] = 0$.

Le théorème de Radon–Nikodym s'applique donc à $\mathbf{P}$ et $\mathbf{Q}$ sur $(\Omega, \mathscr{G})$ et nous assure l'existence d'une variable aléatoire $Y : (\Omega, \mathscr{G}) \to [0, \infty[$, finie presque sûrement, qui est donc en particulier $\mathscr{G}$-mesurable, telle que pour tout $B \in \mathscr{G}$ on ait $\mathbf{Q}(B) = \mathbf{E}[Y\mathbf{1}_B]$, c'est-à-dire $\mathbf{E}[X\mathbf{1}_B] = \mathbf{E}[Y\mathbf{1}_B]$. La variable aléatoire Y est donc une espérance conditionnelle de X sachant $\mathscr{G}$. □

Il nous faut maintenant parvenir à nous passer de l'hypothèse que l'espérance de X est finie. Pour cela, étant donné une variable aléatoire positive X, nous allons appliquer la proposition 2.3 aux variables aléatoires de la forme $\min(X, n)$, où n est un entier que nous ferons tendre vers l'infini.

Nous allons énoncer et démontrer cette procédure d'approximation dans le résultat suivant, qui nous servira à nouveau dans la prochaine section.

Proposition 2.4. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Si toute variable aléatoire positive* bornée *sur $(\Omega, \mathscr{F}, \mathbf{P})$ admet une espérance conditionnelle sachant $\mathscr{G}$, alors toute variable aléatoire positive sur $(\Omega, \mathscr{F}, \mathbf{P})$ admet une espérance conditionnelle sachant $\mathscr{G}$.*

Démonstration. Soit $X : (\Omega, \mathscr{F}) \to [0, \infty]$ une variable aléatoire positive. Pour tout entier $n \geqslant 0$, la variable aléatoire $\min(X, n)$ est bornée, par n, et admet donc, par hypothèse, une espérance conditionnelle sachant $\mathscr{G}$. Nous en considérons une, que nous notons Y_n.

La suite $(\min(X, n))_{n\geqslant 0}$ est presque sûrement croissante. Fixons un entier $n \geqslant 0$. Pour tout événement $B \in \mathscr{G}$, nous avons

$$\mathbf{E}[Y_n\mathbf{1}_B] = \mathbf{E}[\min(X, n)\mathbf{1}_B] \leqslant \mathbf{E}[\min(X, n+1)\mathbf{1}_B] = \mathbf{E}[Y_{n+1}\mathbf{1}_B].$$

D'après la proposition 1.2, on a donc $Y_n \leqslant Y_{n+1}$ presque sûrement. La suite $(Y_n)_{n\geqslant 0}$ converge donc presque sûrement vers une variable aléatoire $\mathscr{G}$-mesurable positive que nous notons Y. Montrons que Y est une espérance conditionnelle de X sachant $\mathscr{G}$.

Nous savons déjà que Y est $\mathscr{G}$-mesurable, comme limite presque sûre d'une suite de variables aléatoires $\mathscr{G}$-mesurables.[1] Donnons-nous maintenant $B \in \mathscr{G}$. Nos hypothèses, et deux applications du théorème de convergence montone, nous donnent

$$\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[\lim_{n\to\infty} Y_n\mathbf{1}_B] = \lim_{n\to\infty} \mathbf{E}[Y_n\mathbf{1}_B] = \lim_{n\to\infty} \mathbf{E}[\min(X,n)\mathbf{1}_B] = \mathbf{E}[\lim_{n\to\infty} \min(X,n)\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B],$$

ce qui achève la démonstration. □

Le théorème 2.1 découle immédiatement de la proposition 2.3, qui entraîne que toute variable aléatoire positive bornée admet une espérance conditionnelle sachant $\mathscr{G}$, et de la proposition 2.4, qui permet d'en déduire que toute variable aléatoire positive admet une espérance conditionnelle sachant $\mathscr{G}$.

Il faut bien avouer que cette démonstration de l'existence de l'espérance conditionnelle nous laisse — en tout cas me laisse[2] — un peu démuni(s) quant à ce qu'est cette notion. Nous avons appliqué un théorème puissant et un peu mystérieux, et au moins en première approche, il me semble que nous n'avons pas beaucoup avancé dans notre compréhension de ce qu'est l'espérance conditionnelle. La deuxième démonstration de son existence va au contraire nous donner un point de vue assez différent, et très important, sur l'espérance conditionnelle : celui d'une *meilleure approximation*.

2.2 L'espérance conditionnelle comme projeté orthogonal

La deuxième démonstration de l'existence de l'espérance conditionnelle que nous allons donner est, au moins à première vue, assez différente de la précédente.[3] Elle repose sur le résultat suivant de géométrie des espaces de Hilbert. Rappelons qu'un espace de Hilbert (réel, dans notre cas) est un espace vectoriel réel muni d'un produit scalaire et qui est complet vis-à-vis de la norme induite par ce produit scalaire.

Théorème 2.5 (Projection orthogonale sur un sous-espace fermé)**.** *Soit H un espace de Hilbert. Soit F un sous-espace vectoriel fermé de H. Soit $F^\perp$ l'orthogonal de F. Alors $H = F \oplus F^\perp$.*

Pour tout sous-espace fermé F de H, la projection $p : H \to F$ sur F parallèlement à $F^\perp$ est donc bien définie et associe à tout élément de H l'unique élément de F qui lui est le plus proche. En symboles, pour tous $h \in H$ et $f \in F$, on a

$$\|h - p(h)\| \leqslant \|h - f\|,$$

avec égalité seulement si $f = p(h)$.

Nous allons appliquer ce théorème dans l'espace de Hilbert $L^2(\Omega, \mathscr{F}, \mathbf{P})$, au sous-espace fermé $L^2(\Omega, \mathscr{G}, \mathbf{P})$. Ceci nous fait sortir du cadre des variables aléatoires positives, ce qui n'est pas grave, et demande par ailleurs un peu de soin. Le paragraphe qui suit peut être laissé de

1. Si ce point vous inquiète, parce que Y n'a pas été définie partout, vous pouvez vérifier que quitte à remplacer, pour chaque $n \geqslant 0$, la variable aléatoire Y_n par $\max(Y_0, \dots, Y_n)$, qui est également une espérance conditionnelle de $\min(X, n)$ sachant $\mathscr{G}$, on peut supposer la suite $(Y_n)_{n\geqslant 0}$ croissante partout, si bien que la variable aléatoire Y est définie sans ambiguité en tout point de Ω, et est bien $\mathscr{G}$-mesurable.

2. Chacun construit sa compréhension des mathématiques de manière différente, et il est fort possible que vous trouviez cette première démonstration plus éclairante que la deuxième. Quoi qu'il en soit, les deux méritent d'être étudiées et méditées.

3. Un examen de la démonstration du théorème de Radon–Nikodym montre qu'il repose en fait lui-même sur un argument de projection orthogonale sur un sous-espace fermé d'un espace de Hilbert.

côté si vous le trouvez difficile. Si vous choisissez de ne pas l'étudier, lisez la version courte de la démonstration de la proposition 2.6

Expliquons d'abord comment $L^2(\Omega, \mathscr{G}, \mathbf{P})$ peut être considéré comme un sous-espace vectoriel de $L^2(\Omega, \mathscr{F}, \mathbf{P})$. Pour cela, rappelons que $L^2(\Omega, \mathscr{F}, \mathbf{P})$ est le quotient de l'espace vectoriel $\mathscr{L}^2(\Omega, \mathscr{F}, \mathbf{P})$, formé des variables aléatoires réelles $\mathscr{F}$-mesurables sur Ω qui sont de carré $\mathbf{P}$-intégrable, par le sous-espace $\mathscr{N}_{\mathscr{F}}$ formé des variables aléatoires $\mathscr{F}$-mesurables nulles $\mathbf{P}$-presque sûrement.

Une fonction réelle sur Ω qui est $\mathscr{G}$-mesurable est $\mathscr{F}$-mesurable. Il y a donc une application linéaire d'inclusion $\mathscr{L}^2(\Omega, \mathscr{G}, \mathbf{P}) \hookrightarrow \mathscr{L}^2(\Omega, \mathscr{F}, \mathbf{P})$. En composant cette inclusion par l'application quotient, on obtient une application

$$\mathscr{L}^2(\Omega, \mathscr{G}, \mathbf{P}) \hookrightarrow \mathscr{L}^2(\Omega, \mathscr{F}, \mathbf{P}) \to L^2(\Omega, \mathscr{F}, \mathbf{P}).$$

Le noyau de cette application est l'ensemble des fonctions réelles $\mathscr{G}$-mesurables sur Ω qui appartiennent à $\mathscr{N}_{\mathscr{F}}$, c'est-à-dire qui sont $\mathscr{F}$-mesurables (mais elles le sont toutes, puisqu'elles sont $\mathscr{G}$-mesurables) et qui sont nulles $\mathbf{P}$-presque sûrement. Ce noyau est donc exactement l'ensemble $\mathscr{N}_{\mathscr{G}}$ des variables aléatoires $\mathscr{G}$-mesurables sur Ω nulles $\mathbf{P}$-presque sûrement. En passant au quotient, on obtient donc une application injective

$$L^2(\Omega, \mathscr{G}, \mathbf{P}) \hookrightarrow L^2(\Omega, \mathscr{F}, \mathbf{P}).$$

On identifie $L^2(\Omega, \mathscr{G}, \mathbf{P})$ avec son image par cette application, qui est le sous-espace de $L^2(\Omega, \mathscr{F}, \mathbf{P})$ formé des classes d'équivalence qui contiennent une fonction $\mathscr{G}$-mesurable.

Proposition 2.6. *Le sous-espace $L^2(\Omega, \mathscr{G}, \mathbf{P})$ est fermé dans $L^2(\Omega, \mathscr{F}, \mathbf{P})$.*

Démonstration. Soit $(Y_n)_{n\geqslant 0}$ une suite d'éléments de $L^2(\Omega, \mathscr{G}, \mathbf{P})$ qui converge dans $L^2(\Omega, \mathscr{F}, \mathbf{P})$ vers une variable aléatoire Y. Nous allons montrer que Y appartient à $L^2(\Omega, \mathscr{G}, \mathbf{P})$.

Version courte de l'argument — De la suite $(Y_n)_{n\geqslant 0}$, qui est une suite de variables aléatoires qui converge dans L^2, on peut extraire une sous-suite qui converge presque sûrement. La limite Y est donc limite de cette sous-suite, qui est une suite de variables aléatoires $\mathscr{G}$-mesurables. Elle est donc elle-même $\mathscr{G}$-mesurable.

Version précise de l'argument — Choisissons pour chaque $n \geqslant 0$ un représentant $\mathscr{G}$-mesurable de Y_n. La suite de ces représentants converge dans L^2, donc on peut en extraire une sous-suite qui converge presque sûrement. La variable aléatoire définie comme égale à la limite de cette sous-suite sur l'événement où cette limite existe, et comme égale à 0 sur l'événement complémentaire, est d'une part $\mathscr{G}$-mesurable, et d'autre part limite presque sûre d'une sous-suite de la suite $(Y_n)_{n\geqslant 0}$, donc un représentant de Y. Ainsi, Y admet un représentant $\mathscr{G}$-mesurable, donc appartient à $L^2(\Omega, \mathscr{G}, \mathbf{P})$. □

Nous pouvons maintenant utiliser l'existence d'une projection sur un sous-espace vectoriel fermé d'un espace de Hilbert pour démontrer l'existence de l'espérance conditionnelle.

Proposition 2.7. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Pour toute variable aléatoire positive X sur Ω telle que $\mathbf{E}[X^2] < \infty$ et toute sous-tribu $\mathscr{G}$ de $\mathscr{F}$, il existe une espérance conditionnelle de X sachant $\mathscr{G}$.*

Démonstration. Notons p la projection orthogonale de $L^2(\Omega, \mathscr{F}, \mathbf{P})$ sur le sous-espace fermé $L^2(\Omega, \mathscr{G}, \mathbf{P})$. Notons $Y = p(X)$. Alors Y est $\mathscr{G}$-mesurable. De plus, pour tout $B \in \mathscr{G}$, l'indicatrice $\mathbf{1}_B$ appartient à $L^2(\Omega, \mathscr{G}, \mathbf{P})$, donc $X - Y$ est orthogonal à $\mathbf{1}_B$, ce qui s'écrit

$$\mathbf{E}[(X - Y)\mathbf{1}_B] = 0, \text{ c'est-à-dire } \mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B].$$

Tout représentant $\mathscr{G}$-mesurable de Y est donc une espérance conditionnelle de X sachant $\mathscr{G}$. □

Cette proposition entraîne que toute variable aléatoire positive bornée admet une espérance conditionnelle sachant $\mathscr{G}$, et la proposition 2.4 permet d'en déduire que toute variable aléatoire positive admet une espérance conditionnelle sachant $\mathscr{G}$.

2.2.1 Commentaires

De ce qui précède se dégage l'idée fondamentale que l'espérance conditionnelle de X sachant $\mathscr{G}$ est, au sens de l'erreur quadratique, la meilleure approximation de X par une variable aléatoire $\mathscr{G}$-mesurable.

Comme toujours, cette assertion simple ne s'applique pas en toute généralité, ne serait-ce que parce que certaines variables aléatoires positives ne sont pas de carré intégrable. Cela n'enlève rien à son importance, et au fait qu'elle constitue une des facettes de la notion d'espérance conditionnelle.

Exercice 2.1. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Soit X une variable aléatoire positive sur $(\Omega, \mathscr{F}, \mathbf{P})$. Soit Y une espérance conditionnelle de X sachant $\mathscr{G}$.*

Montrer que X est nulle sur l'événement $\{Y = 0\}$.

$$\mathbf{E}[X\mathbf{1}_{Y=0}] = \mathbf{E}[Y\mathbf{1}_{Y=0}] = 0.$$

Que peut-on dire de l'événement $\{Y = 0\}$?

Exercice 2.2. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Soit $A \in \mathscr{F}$ un événement. On pose $X = \infty\mathbf{1}_A$. Soit Y une espérance conditionnelle de X sachant $\mathscr{G}$.*

Montrer que pour tout $B \in \mathscr{G}$, l'espérance $\mathbf{E}[Y\mathbf{1}_B]$ vaut 0 ou ∞. En déduire qu'il existe $C \in \mathscr{G}$ tel que $Y = \infty\mathbf{1}_C$ presque sûrement.

Peut-on décrire C en fonction de A ?

2.3 Notations

Nous noterons désormais $\mathbf{E}[X|\mathscr{G}]$ l'espérance conditionnelle d'une variable aléatoire positive X sachant une sous-tribu $\mathscr{G}$, étant entendu que cette variable aléatoire n'est définie que modulo la relation d'égalité presque sûre.

Nous noterons $\mathbf{E}[X|Z]$ l'espérance conditionnelle $\mathbf{E}[X|\sigma(Z)]$ de X sachant la tribu engendrée par une variable aléatoire Z.

Pour faire un lien avec la notation "élémentaire" $\mathbf{P}(A|B)$ de la probabilité conditionnelle d'un événement A sachant un événement B, introduisons dès maintenant des notations dont nous ne nous servirons que plus tard, dans l'étude des chaînes de Markov.

Si A est un événement (un élément de $\mathscr{F}$, avec nos notations), nous noterons $\mathbf{P}(A|\mathscr{G})$ l'espérance conditionnelle $\mathbf{E}[\mathbf{1}_A|\mathscr{G}]$, qui est donc une variable aléatoire.

Dans le cas où B est un événement de probabilité ni nulle ni égale à 1, si on pose $\mathscr{G} = \sigma(\{B\}) = \{\varnothing, B, B^c, \Omega\}$, on a, d'après la proposition 1.4,

$$\mathbf{P}(A|\mathscr{G}) = \mathbf{E}[\mathbf{1}_A|\mathscr{G}] = \underbrace{\frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}}_{\mathbf{P}(A|B)} \mathbf{1}_B + \underbrace{\frac{\mathbf{P}(A \cap B^c)}{\mathbf{P}(B^c)}}_{\mathbf{P}(A|B^c)} \mathbf{1}_{B^c}.$$

Le nombre habituellement noté $\mathbf{P}(A|B)$ est donc, dans notre langage, la valeur sur l'événement B de l'espérance conditionnelle de l'indicatrice de A sachant la tribu engendrée par B.

On utilise parfois la notation $\mathbf{E}[X|A]$ pour la valeur sur l'événement A de l'espérance conditionnelle de X sachant $\sigma(\{A\})$. Ainsi comprise, la notation $\mathbf{E}[X|A]$ désigne un nombre, qui est la valeur moyenne de X sur A. Je ne ferai pas usage de cette notation dans ce cours.

3 Propriétés dans le cas positif

Nous avons maintenant défini ce qu'est l'espérance conditionnelle d'une variable aléatoire positive sachant une sous-tribu, démontré qu'elle existe, et qu'elle est unique modulo la relation d'égalité presque sûre.

Nous allons maintenant étudier les propriétés de l'espérance conditionnelle : dans ce chapitre, nous allons rassembler les outils qui nous permettront de calculer *avec* des espérances conditionnelles. Nous reviendrons plus tard à des méthodes qui permettent de *calculer des espérances conditionnelles*, dans les cas "concrets" où c'est possible.

3.1 Propriétés fondamentales

Le théorème suivant rassemble les propriétés qui sont d'usage constant lorsqu'on calcule avec des espérances conditionnelles. Il est donc la base d'une grande partie de ce que nous ferons dans la suite de ce cours, et très bientôt, nous l'utiliserons sans même le citer.

Théorème 3.1 (Boîte à outils)**.** *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soient X, Y des variables aléatoires positives sur cet espace. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$.*

1. *Pour tous $a, b \in [0, \infty]$, on a $\mathbf{E}[aX + bY|\mathscr{G}] = a\mathbf{E}[X|\mathscr{G}] + b\mathbf{E}[Y|\mathscr{G}]$ p.s.*
2. *Si $X \leqslant Y$ p.s., alors $\mathbf{E}[X|\mathscr{G}] \leqslant \mathbf{E}[Y|\mathscr{G}]$ p.s.*
3. $\mathbf{E}[\mathbf{E}[X|\mathscr{G}]] = \mathbf{E}[X]$
4. *Si $\mathscr{H}$ est une sous-tribu de $\mathscr{G}$, alors $\mathbf{E}\big[\mathbf{E}[X|\mathscr{G}]\big|\mathscr{H}\big] = \mathbf{E}[X|\mathscr{H}]$ p.s.*
5. *Si X est indépendante de $\mathscr{G}$, alors $\mathbf{E}[X|\mathscr{G}] = \mathbf{E}[X]$ p.s.*
6. *Si X est $\mathscr{G}$-mesurable, alors $\mathbf{E}[X|\mathscr{G}] = X$ p.s.*
7. *Si X est $\mathscr{G}$-mesurable, alors $\mathbf{E}[XY|\mathscr{G}] = X\mathbf{E}[Y|\mathscr{G}]$ p.s.*

Les seules propriétés peut-être un peu nouvelles sont les propriétés 4 et 7. La propriété 7 est particulièrement importante : on peut la paraphraser en disant que, conditionnellement à $\mathscr{G}$, les variables aléatoires $\mathscr{G}$-mesurables se comportent comme des constantes. Plus encore que pour toutes les autres, il est particulièrement important de réfléchir à ce que cette propriété exprime lorsque la sous-tribu $\mathscr{G}$ est engendrée par une partition.

Démonstration. 1. Montrons que $a\mathbf{E}[X|\mathscr{G}] + b\mathbf{E}[Y|\mathscr{G}]$ est une espérance conditionnelle de $aX + bY$ sachant $\mathscr{G}$. D'abord, c'est la somme de deux variables aléatoires $\mathscr{G}$-mesurable, donc c'est une variable aléatoire $\mathscr{G}$-mesurable. Ensuite, pour tout événement $B \in \mathscr{G}$, on a

$$\mathbf{E}\big[(a\mathbf{E}[X|\mathscr{G}] + b\mathbf{E}[Y|\mathscr{G}])\mathbf{1}_B\big] = a\mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B] + b\mathbf{E}[\mathbf{E}[Y|\mathscr{G}]\mathbf{1}_B],$$

par additivité de l'intégrale, puis

$$a\mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B] + b\mathbf{E}[\mathbf{E}[Y|\mathscr{G}]\mathbf{1}_B] = a\mathbf{E}[X\mathbf{1}_B] + b\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[(aX + bY)\mathbf{1}_B],$$

par définition des espérances conditionnelles de X et Y, puis à nouveau par additivité de l'intégrale. C'est l'égalité souhaitée, qui termine la démonstration.

2. Les variables aléatoires $\mathbf{E}[X|\mathscr{G}]$ et $\mathbf{E}[Y|\mathscr{G}]$ sont $\mathscr{G}$-mesurables, et pour tout $B \in \mathscr{G}$, on a

$$\mathbf{E}\big[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B\big] = \mathbf{E}[X\mathbf{1}_B] \leqslant \mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}\big[\mathbf{E}[Y|\mathscr{G}]\mathbf{1}_B\big].$$

La proposition 1.2 nous assure donc que $\mathbf{E}[X|\mathscr{G}] \leqslant \mathbf{E}[Y|\mathscr{G}]$ presque sûrement.

3. Il suffit d'appliquer la définition de l'espérance conditionnelle avec l'événement $B = \Omega$ de $\mathscr{G}$.

4. On peut soit chercher à montrer que $\mathbf{E}\big[\mathbf{E}[X|\mathscr{G}]\big|\mathscr{H}\big]$ est une espérance conditionnelle de X sachant $\mathscr{H}$, soit chercher à montrer que $\mathbf{E}[X|\mathscr{H}]$ est une espérance conditionnelle de $\mathbf{E}[X|\mathscr{G}]$ sachant $\mathscr{H}$. Essayez les deux ! Nous allons montrer que $\mathbf{E}\big[\mathbf{E}[X|\mathscr{G}]\big|\mathscr{H}\big]$ est une espérance conditionnelle de X sachant $\mathscr{H}$.

D'abord, $\mathbf{E}\big[\mathbf{E}[X|\mathscr{G}]\big|\mathscr{H}\big]$ est $\mathscr{H}$-mesurable. Ensuite, soit $B \in \mathscr{H}$. On a

$$\mathbf{E}\Big[\mathbf{E}\big[\mathbf{E}[X|\mathscr{G}]\big|\mathscr{H}\big]\mathbf{1}_B\Big] \overset{B\in\mathscr{H}}{=} \mathbf{E}\Big[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B\Big] \overset{B\in\mathscr{G}}{=} \mathbf{E}\Big[X\mathbf{1}_B\Big],$$

ce qui est l'égalité espérée.

5. Nous avons établi cette propriété à la section 1.3.2.

6. Nous avons établi cette propriété à la section 1.3.1.

7. La démonstration de cette propriété est un peu moins immédiate que la démonstration de celles qui précèdent.

Nous allons montrer que $X\mathbf{E}[Y|\mathscr{G}]$ est une espérance conditionnelle de XY sachant $\mathscr{G}$. Tout d'abord, c'est une variable aléatoire $\mathscr{G}$-mesurable, comme produit de deux telles variables aléatoires. Donnons-nous maintenant $B \in \mathscr{G}$ et montrons que

$$\mathbf{E}\big[X\mathbf{E}[Y|\mathscr{G}]\mathbf{1}_B\big] = \mathbf{E}[XY\mathbf{1}_B]. \tag{3.1}$$

Pour cela, nous allons fixer B et Y, et montrer que l'égalité ci-dessus a lieu pour toute variable aléatoire positive $\mathscr{G}$-mesurable X.

Commençons par supposer que X est l'indicatrice d'un événement de $\mathscr{G}$, disons $X = \mathbf{1}_C$, avec $C \in \mathscr{G}$. Alors

$$\mathbf{E}\big[X\mathbf{E}[Y|\mathscr{G}]\mathbf{1}_B\big] = \mathbf{E}\big[\mathbf{1}_C\mathbf{E}[Y|\mathscr{G}]\mathbf{1}_B\big] = \mathbf{E}\big[\mathbf{E}[Y|\mathscr{G}]\mathbf{1}_{B\cap C}\big] = \mathbf{E}[Y\mathbf{1}_{B\cap C}] = \mathbf{E}[\mathbf{1}_C Y\mathbf{1}_B] = \mathbf{E}[XY\mathbf{1}_B].$$

Si l'égalité (3.1), vue comme une propriété de X, est vraie pour deux variables aléatoires X et X', alors elle est vraie pour $aX + a'X'$ pour tous $a, a' \in [0, \infty]$. Comme elle est vraie pour les indicatrices d'événements de $\mathscr{G}$, elle est vraie pour toutes les fonctions $\mathscr{G}$-mesurables étagées positives.

Puisque toute variable aléatoire positive est la limite d'une suite croissante de variables aléatoires étagées positives, le théorème de convergence monotone nous assure que (3.1) est vraie pour toute variable aléatoire positive X. □

3.2 Beppo Levi, Fatou, Hölder

Nous allons maintenant énoncer et démontrer les analogues pour l'espérance conditionnelle de trois résultats fondamentaux concernant l'intégrale des fonctions mesurables positives, qui sont le théorème de convergence monotone (ou de Beppo Levi), le lemme de Fatou, et l'inégalité de Hölder.

Théorème 3.2. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires positives sur $(\Omega, \mathscr{F}, \mathbf{P})$.*

1. (Théorème de convergence monotone conditionnel) *Si la suite $(X_n)_{n\geqslant 0}$ est croissante presque sûrement, alors en notant X sa limite presque sure, on a $\mathbf{E}[X|\mathscr{G}] = \lim \mathbf{E}[X_n|\mathscr{G}]$ p.s.*
2. (Lemme de Fatou conditionnel) *On a $\mathbf{E}[\liminf X_n|\mathscr{G}] \leqslant \liminf \mathbf{E}[X_n|\mathscr{G}]$ p.s.*

Démonstration. 1. Nous allons déduire le théorème de convergence monotone conditionnel du théorème de convergence monotone classique.

Nous allons montrer l'égalité en utilisant la proposition 1.2 : nous allons montrer que pour tout événement $B \in \mathscr{G}$, l'égalité

$$\mathbf{E}\big[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B\big] = \mathbf{E}\big[\lim \mathbf{E}[X_n|\mathscr{G}]\mathbf{1}_B\big]$$

a lieu. Choisissons donc $B \in \mathscr{G}$ et calculons les deux membres. D'une part,

$$\mathbf{E}\big[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B\big] = \mathbf{E}\big[\mathbf{E}[X\mathbf{1}_B|\mathscr{G}]\big] = \mathbf{E}[X\mathbf{1}_B].$$

D'autre part,

$$\begin{aligned}
\mathbf{E}\big[\lim \mathbf{E}[X_n|\mathscr{G}]\mathbf{1}_B\big] &= \mathbf{E}\big[\lim \mathbf{E}[X_n\mathbf{1}_B|\mathscr{G}]\big] && \text{(propriété 7)}\\
&= \lim \mathbf{E}\big[\mathbf{E}[X_n\mathbf{1}_B|\mathscr{G}]\big] && \text{(propriété 2 et convergence monotone)}\\
&= \lim \mathbf{E}[X_n\mathbf{1}_B] && \text{(propriété 3)}\\
&= \mathbf{E}[\lim X_n\mathbf{1}_B] && \text{(convergence monotone)}\\
&= \mathbf{E}[X\mathbf{1}_B],
\end{aligned}$$

ce qui achève la démonstration.

2. Nous allons déduire le lemme de Fatou conditionnel du théorème de convergence monotone conditionnel, comme on déduit le lemme de Fatou du théorème de convergence monotone.

Posons donc, pour tout $n \geqslant 0$,

$$Y_n = \inf\{X_k : k \geqslant n\}.$$

La suite $(Y_n)_{n\geqslant 0}$ est croissante et vérifie $Y_n \leqslant X_n$ pour tout $n \geqslant 0$, c'est d'ailleurs la plus grande suite (terme à terme) qui ait cette propriété.

La suite $(Y_n)_{n\geqslant 0}$ converge presque sûrement en croissant vers $\liminf X_n$. Ainsi, par convergence monotone conditionnelle,

$$\mathbf{E}[\liminf X_n|\mathscr{G}] = \mathbf{E}[\lim Y_n|\mathscr{G}] = \lim \mathbf{E}[Y_n|\mathscr{G}].$$

De plus, $Y_n \leqslant X_n$ pour tout $n \geqslant 0$, donc

$$\lim \mathbf{E}[Y_n|\mathscr{G}] = \liminf \mathbf{E}[Y_n|\mathscr{G}] \leqslant \liminf \mathbf{E}[X_n|\mathscr{G}].$$

En mettant les deux dernières équations bout à bout, on obtient l'inégalité souhaitée. □

Théorème 3.3 (Inégalité de Hölder conditionnelle)**.** *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soient X, Y des variables aléatoires positives sur cet espace. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Soient enfin*

p et q des réels positifs tels que $\frac{1}{p}+\frac{1}{q}=1$. Alors $\mathbf{E}[XY|\mathscr{G}] \leqslant \mathbf{E}[X^p|\mathscr{G}]^{\frac{1}{p}}\mathbf{E}[Y^q|\mathscr{G}]^{\frac{1}{q}}$ p.s.

La démonstration de ce théorème ne sera pas étudiée en cours, et fera l'objet d'un devoir à la maison à destinsation des étudiant(e)s à distance.

Exercice 3.1. *Soit X une variable aléatoire positive. Montrer que*

$$Y = \lim_{n\to\infty} \mathbf{E}[X^n|\mathscr{G}]^{\frac{1}{n}}$$

est la plus petite variable aléatoire $\mathscr{G}$-mesurable telle que $X \leqslant Y$.

3.3 Grossissement de la sous-tribu

La dernière propriété générale de l'espérance conditionnelle que nous allons étudier dans ce chapitre concerne une situation où l'on fait grossir la sous-tribu par rapport à laquelle on conditionne. Son énoncé est l'occasion d'une belle démonstration basée sur le théorème de la classe monotone, et que nous donnons en détail.

Proposition 3.4. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabiltiés. Soit X une variable aléatoire positive sur cet espace. Soient $\mathscr{G}$ et $\mathscr{H}$ des sous-tribus de $\mathscr{F}$. Si $\mathscr{H}$ est indépendante de $\sigma(\sigma(X)\cup\mathscr{G})$, alors*

$$\mathbf{E}[X|\sigma(\mathscr{G}\cup\mathscr{H})] = \mathbf{E}[X|\mathscr{G}] \text{ p.s.}$$

Il faut être très attentif à l'hypothèse qui est ici faite : on suppose que $\mathscr{H}$ est indépendante de $\sigma(\sigma(X)\cup\mathscr{G})$, qui est la plus petite tribu qui contient $\mathscr{G}$ et qui rend X mesurable. Cette hypothèse entraîne que X est indépendante de $\mathscr{H}$ et que $\mathscr{G}$ et $\mathscr{H}$ sont indépendantes. Mais, comme le montre par exemple l'exercice 3.2, la réciproque est fausse :

$$\left(X \perp\!\!\!\perp \mathscr{H} \text{ et } \mathscr{G} \perp\!\!\!\perp \mathscr{H}\right) \not\Rightarrow \mathscr{H} \perp\!\!\!\perp \sigma(\sigma(X)\cup\mathscr{G}).$$

Indiquons au passage qu'une notation courante pour $\mathbf{E}[X|\sigma(\mathscr{G}\cup\mathscr{H})]$ est $\mathbf{E}[X|\mathscr{G},\mathscr{H}]$.

Démonstration. Nous allons commencer par démontrer le résultat en faisant l'hypothèse supplémentaire que $\mathbf{E}[X] < \infty$. Une fois le résultat acquis sous cette hypothèse, et étant donné une variable aléatoire positive X quelconque, nous pourrons affirmer que pour tout entier $n \geqslant 0$, la variable aléatoire positive $\min(X,n)$ étant bornée, elle est d'espérance finie, et on a

$$\mathbf{E}[\min(X,n)|\sigma(\mathscr{G}\cup\mathscr{H})] = \mathbf{E}[\min(X,n)|\mathscr{G}] \text{ p.s.}$$

On pourra alors conclure en faisant tendre n vers l'infini et en appliquant le théorème de convergence monotone conditionnel (théorème 3.2) à chaque membre de l'égalité.

Il reste donc à nous donner une variable aléatoire positive X telle que $\mathbf{E}[X] < \infty$ et à démontrer l'égalité. Comme pour la propriété 4 du théorème 3.1, il faut choisir dans quel sens nous démontrons cette égalité : on peut chercher à montrer que $\mathbf{E}[X|\sigma(\mathscr{G}\cup\mathscr{H})]$ est une espérance conditionnelle de X sachant $\mathscr{G}$, ou à montrer que $\mathbf{E}[X|\mathscr{G}]$ est une espérance conditionnelle de X sachant $\sigma(\mathscr{G}\cup\mathscr{H})$. Essayez les deux !

Nous allons montrer que $\mathbf{E}[X|\mathscr{G}]$ est une espérance conditionnelle de X sachant $\sigma(\mathscr{G}\cup\mathscr{H})$, en vérifiant les deux propriétés de la définition.

Tout d'abord, la variable aléatoire positive $\mathbf{E}[X|\mathscr{G}]$ est mesurable par rapport à la tribu $\mathscr{G}$ sur son ensemble de départ Ω, donc par rapport à la tribu plus grande $\sigma(\mathscr{G}\cup\mathscr{H})$.

Il nous faut ensuite montrer que pour tout $B \in \sigma(\mathscr{G} \cup \mathscr{H})$, on a l'égalité

$$\mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B]. \tag{3.2}$$

Nous allons pour cela utiliser le théorème de la classe monotone (voir l'annexe B). Considérons la classe de parties de Ω

$$\mathscr{M} = \{B \in \sigma(\mathscr{G} \cup \mathscr{H}) : \text{ l'égalité (3.2) est vraie}\}$$

et montrons que c'est une classe monotone. Il y a trois choses à vérifier.

— D'abord, Ω appartient à $\mathscr{M}$, car lorsque $B = \Omega$, les deux membres de (3.2) valent $\mathbf{E}[X]$.

— Considérons A et B appartenant à $\mathscr{M}$ tels que $A \subseteq B$. Alors $\mathbf{1}_B = \mathbf{1}_A + \mathbf{1}_{B\setminus A}$, donc

$$\mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_{B\setminus A}] + \mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_A] = \mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B] \quad \text{et} \quad \mathbf{E}[X\mathbf{1}_{B\setminus A}] + \mathbf{E}[X\mathbf{1}_A] = \mathbf{E}[X\mathbf{1}_B].$$

C'est maintenant que nous allons utiliser l'hypothèse que $\mathbf{E}[X] < \infty$. Cette hypothèse nous garantit que les six nombres écrits ci-dessus sont finis, si bien que nous pouvons écrire [1]

$$\mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_{B\setminus A}] = \mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_A] - \mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_A] - \mathbf{E}[X\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_{B\setminus A}]$$

et conclure que $B \setminus A$ appartient à $\mathscr{M}$.

— Donnons-nous maintenant une suite croissante $(B_n)_{n\geqslant 0}$ d'éléments de $\mathscr{M}$, et notons $B = \bigcup_{n\geqslant 0} B_n$. Alors pour tout $n \geqslant 0$, on a $\mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_{B_n}] = \mathbf{E}[X\mathbf{1}_{B_n}]$.
Comme la suite $(\mathbf{1}_{B_n})_{n\geqslant 0}$ converge presque sûrement (et même partout) en croissant vers $\mathbf{1}_B$, les suites $(X\mathbf{1}_{B_n})_{n\geqslant 0}$ et $(\mathbf{E}[X|\mathscr{G}]\mathbf{1}_{B_n})_{n\geqslant 0}$ convergent respectivement en croissant vers $X\mathbf{1}_B$ et $\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B$. Le théorème de convergence monotone usuel nous permet donc de faire tendre n vers l'infini dans l'égalité précédente et d'affirmer que $\mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B]$, ce qui signifie que B appartient à $\mathscr{M}$.

Pour appliquer le théorème de la classe monotone, il nous faut montrer que $\mathscr{M}$ contient un π-système qui engendre la tribu $\sigma(\mathscr{G} \cup \mathscr{H})$. Un tel π-système ne nous est pas directement donné par le contexte, et il est normal, à cette étape du raisonnement, de tâtonner un peu.

Toutefois, le suspense ne faisant pas partie du registre normal du discours scientifique ou pédagogique, considérons la classe de parties

$$\mathscr{I} = \{G \cap H : G \in \mathscr{G}, H \in \mathscr{H}\}.$$

Vérifions que c'est un π-système sur Ω.

— Tout d'abord, $\Omega = \Omega \cap \Omega$ appartient à $\mathscr{I}$.

— Ensuite, si $G_1 \cap H_1$ et $G_2 \cap H_2$ sont deux éléments de $\mathscr{I}$, alors leur intersection s'écrit $(G_1 \cap G_2) \cap (H_1 \cap H_2)$ et appartient encore à $\mathscr{I}$.

Calculons maintenant $\sigma(\mathscr{I})$, la tribu engendrée par $\mathscr{I}$. La première observation est que $\mathscr{G} \subseteq \mathscr{I}$ et $\mathscr{H} \subseteq \mathscr{I}$, comme on le voit en prenant G ou H égal à Ω. Ainsi, $\sigma(\mathscr{I})$ est une tribu qui contient $\mathscr{G} \cup \mathscr{H}$. On a donc l'inclusion $\sigma(\mathscr{I}) \supseteq \sigma(\mathscr{G} \cup \mathscr{H})$.

Par ailleurs, $\sigma(\mathscr{G} \cup \mathscr{H})$ est une tribu qui contient $\mathscr{G}$ et $\mathscr{H}$, donc qui contient toutes les parties de la forme $G \cap H$ avec $G \in \mathscr{G}$ et $H \in \mathscr{H}$. Ainsi, on a l'autre inclusion $\sigma(\mathscr{G} \cup \mathscr{H}) \supseteq \sigma(\mathscr{I})$.

1. Quel aurait été le problème s'ils n'avait pas tous été finis ? Eh bien, des égalités $1 + \infty = \infty$ et $2 + \infty = \infty$, qui sont vraies avec l'arithmétique étendue que nous utilisons sur $[0, \infty]$, on ne peut (fort heureusement) pas déduire que $1 = 2$.

Il ne reste plus qu'à montrer que $\mathscr{I} \subseteq \mathscr{M}$, c'est-à-dire que l'égalité 3.2 est vraie lorsque $B \in \mathscr{I}$. Considérons donc $B = G \cap H$ avec $G \in \mathscr{G}$ et $H \in \mathscr{H}$. Les égalités

$$\begin{aligned}
\mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B] &= \mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_G\mathbf{1}_H] = \mathbf{E}[\mathbf{E}[X\mathbf{1}_G|\mathscr{G}]\mathbf{1}_H] && \text{(propriété 7)}\\
&= \mathbf{E}[\mathbf{E}[X\mathbf{1}_G|\mathscr{G}]]\mathbf{E}[\mathbf{1}_H] && (\mathscr{G} \perp\!\!\!\perp \mathscr{H})\\
&= \mathbf{E}[X\mathbf{1}_G]\mathbf{E}[\mathbf{1}_H] && \text{(propriété 3)}\\
&= \mathbf{E}[X\mathbf{1}_G\mathbf{1}_H] && (\mathscr{H} \perp\!\!\!\perp \sigma(\sigma(X) \cup \mathscr{G}))\\
&= \mathbf{E}[X\mathbf{1}_B]
\end{aligned}$$

montrent que $B \in \mathscr{M}$.

Résumons. Nous avons montré que la classe $\mathscr{M}$ des parties B pour lesquelles l'égalité (3.2) a lieu est une classe monotone, que cette classe monotone contient le π-système $\mathscr{I}$, et que la tribu engendrée par ce π-système est $\sigma(\mathscr{G} \cup \mathscr{H})$. La classe monotone $\mathscr{M}$ contient la classe monotone $\lambda(\mathscr{I})$ engendrée par $\mathscr{I}$, et e théorème de classe monotone nous dit que $\lambda(\mathscr{I}) = \sigma(\mathscr{I})$. Nous avons donc établi que la classe $\mathscr{M}$ contient $\sigma(\mathscr{G} \cup \mathscr{H})$, ce qui est exactement le résultat souhaité. □

Une manière de comprendre cette proposition est d'adopter le point de vue selon lequel l'espérance conditionnelle de X sachant $\mathscr{G}$ est la meilleure approximation de X par une variable aléatoire $\mathscr{G}$-mesurable. On pourrait dire, sans donner de sens précis à cette expression : la meilleure approximation *dans le langage de la tribu* $\mathscr{G}$. On décide alors de se donner un langage plus riche, et d'ajouter $\mathscr{H}$ à $\mathscr{G}$. La proposition nous dit que si $\mathscr{H}$ n'a rien à voir avec le langage "engendré" par X et $\mathscr{G}$, au sens de l'indépendance et avec toutes les précautions nécessaires déjà évoquées, alors l'enrichissement du langage, le passage de $\mathscr{G}$ à $\sigma(\mathscr{G} \cup \mathscr{H})$, ne nous permettra pas de donner une meilleure approximation de X.

Exercice 3.2. *Sur un espace de probabilité $(\Omega, \mathscr{F}, \mathbf{P})$, soient X et Y deux variables aléatoires indépendantes de loi uniforme sur l'ensemble $\{-1, 1\}$. On définit la variable aléatoire $Z = XY$.*

Montrer que $\sigma(Z) \perp\!\!\!\perp \sigma(X)$ et $\sigma(Z) \perp\!\!\!\perp \sigma(Y)$, mais $\sigma(Z) \not\perp\!\!\!\perp \sigma(\sigma(X) \cup \sigma(Y))$.

On pose $\mathscr{G} = \sigma(X)$ et $\mathscr{H} = \sigma(Y)$. On a donc $X \perp\!\!\!\perp \mathscr{H}$ et $\mathscr{G} \perp\!\!\!\perp \mathscr{H}$.

Montrer que $\mathbf{E}[X|\mathscr{G}] = 0$ et $\mathbf{E}[X|\sigma(\mathscr{G} \cup \mathscr{H})] = X$ p.s.

4 Le cas intégrable

Comme pour la théorie de l'intégration, il y a une théorie positive et une théorie intégrable de l'espérance conditionnelle, dont aucune ne contient l'autre : il y a des variables aléatoires positives qui ne sont pas intégrables et des variables aléatoires intégrables qui ne sont pas positives. J'ai choisi dans ce cours de donner une importance un peu plus grande au cas positif qu'il n'est d'usage, mais il est temps de passer au cas des variables aléatoires intégrables. Nous étudierons les méthodes concrètes de calcul d'espérance conditionnelle en parallèle pour les deux cas, après ce chapitre.

4.1 Définition

Commençons par une définition, tout à fait analogue à la définition 1.1.

Définition 4.1. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soit X une variable aléatoire intégrable sur cet espace. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. On appelle* espérance conditionnelle de X sachant $\mathscr{G}$ *une variable aléatoire intégrable Y qui vérifie les deux propriétés suivantes :*

1. *Y est $\mathscr{G}$-mesurable,*
2. *pour tout $B \in \mathscr{G}$, on a $\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B]$.*

4.1.1 Deux définitions pour une notion

On utilise exactement le même vocabulaire pour le cas intégrable que pour le cas positif : on ne dit pas *espérance conditionnelle intégrable* ni *espérance conditionnelle positive*, de même qu'on ne dit pas *intégrale (ou espérance) intégrable* ni *intégrale positive*.

Ceci soulève une question : si une variable aléatoire est à la fois positive et intégrable, c'est-à-dire qu'elle est positive et d'espérance finie, les deux notions d'espérance conditionnelle que nous avons définies coïncident-elles ? Nous répondrons à cette question un peu plus tard (positivement, comme vous vous en doutez).

4.1.2 Une remarque sur les hypothèses

Une autre question, que soulève souvent la définition 4.1, est la suivante : est-il nécessaire de supposer Y intégrable dans la définition, et ne pourrait-on pas le déduire des propriétés qu'on impose à Y ?

Un premier élément de réponse est qu'il faut tout de même, pour pouvoir écrire la définition, supposer que pour tout $B \in \mathscr{G}$, l'espérance $\mathbf{E}[Y\mathbf{1}_B]$ ait un sens. Pour que ce soit le cas, il faut qu'au moins une des deux variables aléatoires

$$(Y\mathbf{1}_B)^+ = \max(Y\mathbf{1}_B, 0) \text{ et } (Y\mathbf{1}_B)^- = \max(-Y\mathbf{1}_B, 0),$$

qu'on appelle *partie positive* et *partie négative* de $Y\mathbf{1}_B$, ait une espérance finie, auquel cas on pose

$$\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[(Y\mathbf{1}_B)^+] - \mathbf{E}[(Y\mathbf{1}_B)^-],$$

qui est éventuellement égal à $\pm\infty$.

La réponse à la question va donc nous être donnée par l'exercice suivant.

Exercice 4.1. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soit X une variable aléatoire intégrable sur cet espace. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Soit Z une variable aléatoire $\mathscr{G}$-mesurable. On suppose que pour tout $B \in \mathscr{G}$, on a*

$$\mathbf{E}[(Z\mathbf{1}_B)^+] < \infty \text{ ou } \mathbf{E}[(Z\mathbf{1}_B)^-] < \infty, \quad \text{et} \quad \mathbf{E}[Z\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B].$$

Montrer que Z est intégrable.

Il découle de cet exercice que si dans la définition 4.1 on enlève l'hypothèse que Y est intégrable et on remplace la propriété 2 par

> *2'. pour tout $B \in \mathscr{G}$, au moins une des deux variables aléatoires $(Y\mathbf{1}_B)^+$ et $(Y\mathbf{1}_B)^-$ a une espérance finie et $\mathbf{E}[(Y\mathbf{1}_B)^+] - \mathbf{E}[(Y\mathbf{1}_B)^-] = \mathbf{E}[X\mathbf{1}_B]$,*

alors on obtient une définition équivalente — mais bien plus compliquée, raison pour laquelle on garde la propriété 2.

4.1.3 L'exemple fondamental

La discussion de la section 1.4, et en particulier la proposition 1.4, se transposent presque mot pour mot au cas intégrable, en remplaçant "variable aléatoire positive" par "variable aléatoire intégrable".

Il est donc vrai, et vu l'importance de ce fait, je le réécris entièrement, que lorsque la sous-tribu $\mathscr{G}$ est engendrée par une partition (finie ou infinie dénombrable, par des éléments de $\mathscr{F}$ de probabilité strictement positive) de Ω, alors l'espérance conditionnelle d'une variable aléatoire intégrable se calcule en remplaçant, sur chaque bloc de la partition, la variable aléatoire par sa valeur moyenne sur ce bloc.

4.1.4 Cas d'une sous-tribu engendrée par une variable aléatoire

De même, la discussion de la section 1.5 reste valable dans le cas des variables aléatoires intégrables. Il nous faut simplement adapter la proposition 1.5 au cas intégrable, où les variables aléatoires sont signées et finies.

Proposition 4.2. *Soit $Z : (\Omega, \mathscr{F}) \to (E, \mathscr{E})$ une application mesurable. Soit $Y : (\Omega, \sigma(Z)) \to (\mathbb{R}, \mathscr{B}_\mathbb{R})$ une fonction mesurable. Alors il existe une fonction mesurable $h : (E, \mathscr{E}) \to (\mathbb{R}, \mathscr{B}_\mathbb{R})$ telle que $Y = h(Z)$.*

Démonstration. Appliquons la proposition 1.5 aux variables aléatoires positives Y^+ et Y^-, les parties positive et négative de Y. Puisque ce sont des variables aléatoires finies, la proposition nous donne deux fonctions h_+ et h_- de Ω dans $[0, \infty[$ telles que $Y_+ = h_+(Z)$ et $Y_- = h_-(Z)$. Il suffit maintenant de poser $h = h_+ - h_-$. □

4.2 Existence et unicité

Il n'est, heureusement, pas nécessaire de refaire dans le cas intégrable le travail déjà assez long que nous avons fait dans le cas positif : nous pouvons nous appuyer dessus. Commençons par traiter la question de l'existence et de l'unicité.

Théorème 4.3. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Pour toute variable aléatoire intégrable X sur Ω et toute sous-tribu $\mathscr{G}$ de $\mathscr{F}$, il existe une espérance conditionnelle de X sachant $\mathscr{G}$. De plus, deux espérances conditionnelles de X sachant $\mathscr{G}$ sont égales $\mathbf{P}$-presque sûrement.*

Dans le cas positif, nous nous sommes beaucoup appuyés sur la proposition 1.2, qui concernait les variables positives. Il nous faut énoncer et démontrer son analogue pour les variables aléatoires intégrables.

Proposition 4.4. *Soient $X, Y : (\Omega, \mathscr{F}, \mathbf{P}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ deux variables aléatoires intégrables.*

1. *Si pour tout événement $A \in \mathscr{F}$ on a $\mathbf{E}[X\mathbf{1}_A] \geqslant 0$, alors $X \geqslant 0$ p.s.*
2. *Si pour tout événement $A \in \mathscr{F}$ on a $\mathbf{E}[X\mathbf{1}_A] \leqslant \mathbf{E}[Y\mathbf{1}_A]$, alors $X \leqslant Y$ p.s.*
3. *Si pour tout événement $A \in \mathscr{F}$ on a $\mathbf{E}[X\mathbf{1}_A] = \mathbf{E}[Y\mathbf{1}_A]$, alors $X = Y$ p.s.*

Démonstration. 1. Pour tout entier $n \geqslant 1$, nous avons

$$0 \leqslant \mathbf{E}[X\mathbf{1}_{\{X \leqslant -\frac{1}{n}\}}] \leqslant -\tfrac{1}{n}\mathbf{P}(X \leqslant -\tfrac{1}{n}),$$

ce qui impose que $\mathbf{P}(X \leqslant -\frac{1}{n}) = 0$. Ainsi,

$$\mathbf{P}(X < 0) = \mathbf{P}\Big(\bigcup_{n \geqslant 1} \{X \leqslant -\tfrac{1}{n}\}\Big) = \lim_{n \to \infty} \mathbf{P}(X \leqslant -\tfrac{1}{n}) = 0,$$

et $X \geqslant 0$ presque sûrement.

2. On applique l'assertion 1 à la variable aléatoire intégrable $Y - X$.

3. D'après l'assertion 2, l'hypothèse implique d'une part que $X \leqslant Y$ presque sûrement et d'autre part que $X \geqslant Y$ presque sûrement. □

Nous pouvons maintenant démontrer le théorème.

Démonstration du théorème 4.3. L'unicité découle de la troisième assertion de la proposition 4.4. Pour démontrer l'existence, considérons une variable aléatoire intégrable X. Notons $X^+ = \max(X, 0)$ et $X^- = \max(-X, 0)$ les parties positive et négative de X. Ce sont deux variables aléatoires positives sur Ω, qui en vertu du théorème 2.1, admettent donc des espérances conditionnelles selon $\mathscr{G}$. Choisissons-en, que nous notons Y_+ et Y_-. Posons

$$Y = Y_+ - Y_-.$$

On a $\mathbf{E}[Y_+] = \mathbf{E}[Y_+\mathbf{1}_\Omega] = \mathbf{E}[X_+\mathbf{1}_\Omega] \leqslant \mathbf{E}[|X|] < \infty$ et, de même, $\mathbf{E}[Y_-] < \infty$. On a donc

$$\mathbf{E}[|Y|] = \mathbf{E}[Y_+ + Y_-] < \infty,$$

donc Y est intégrable.

La variable aléatoire est par ailleurs $\mathscr{G}$-mesurable, comme différence de deux fonctions réelles $\mathscr{G}$-mesurables. De plus, pour tout $B \in \mathscr{G}$, on a

$$\begin{aligned}\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[(Y_+ - Y_-)\mathbf{1}_B] &= \mathbf{E}[Y_+\mathbf{1}_B] - \mathbf{E}[Y_-\mathbf{1}_B] \\ &= \mathbf{E}[X_+\mathbf{1}_B] - \mathbf{E}[X_-\mathbf{1}_B] = \mathbf{E}[(X_+ - X_-)\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B],\end{aligned}$$

ce qui achève de montrer que Y est une espérance conditionnelle de X sachant $\mathscr{G}$. □

On notera désormais $\mathbf{E}[X|\mathscr{G}]$ l'espérance conditionnelle de X sachant $\mathscr{G}$, étant entendu que cette variable aléatoire n'est définie que modulo la relation d'égalité presque sûre.

4.2.1 Le cas des variables aléatoires positives et intégrables

Revenons à la question soulevée à la section 4.1.1. La proposition suivante apporte la réponse positive que nous avions annoncée.

Proposition 4.5. *Pour une variable aléatoire positive et intégrable, les espérances conditionnelles dont l'existence et l'unicité sont garanties respectivement par les théorèmes 2.1 et 4.3 sont égales presque sûrement.*

Démonstration. Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Soit X une variable aléatoire positive et intégrable sur $(\Omega, \mathscr{F}, \mathbf{P})$. Soit Y l'espérance conditionnelle sachant $\mathscr{G}$ de X vue comme variable aléatoire intégrable (donc au sens de la définition 4.1).

La variable aléatoire Y est $\mathscr{G}$-mesurable et pour tout $B \in \mathscr{G}$, on a

$$\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B] \geqslant 0,$$

car X est positive. La première assertion de la proposition 4.4 nous assure donc que Y est positive presque sûrement. C'est donc une espérance conditionnelle sachant $\mathscr{G}$ de X au sens de la définition 1.1. □

4.3 Propriétés

4.3.1 Propriétés fondamentales

Nous allons maintenant énoncer et démontrer un théorème très semblable au théorème 3.1, en utilisant autant que possible la même numérotation des propriétés.

Théorème 4.6. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soient X, Y des variables aléatoires intégrables sur cet espace. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$.*

1. *Pour tous $a, b \in \mathbb{R}$, on a $\mathbf{E}[aX + bY|\mathscr{G}] = a\mathbf{E}[X|\mathscr{G}] + b\mathbf{E}[Y|\mathscr{G}]$ p.s.*
2. *Si $X \leqslant Y$ p.s., alors $\mathbf{E}[X|\mathscr{G}] \leqslant \mathbf{E}[Y|\mathscr{G}]$ p.s.*

2'. *$\big|\mathbf{E}[X|\mathscr{G}]\big| \leqslant \mathbf{E}\big[|X|\big|\,\mathscr{G}\big]$ p.s.*

3. *$\mathbf{E}[\mathbf{E}[X|\mathscr{G}]] = \mathbf{E}[X]$*
4. *Si $\mathscr{H}$ est une sous-tribu de $\mathscr{G}$, alors $\mathbf{E}\big[\mathbf{E}[X|\mathscr{G}]\big|\mathscr{H}\big] = \mathbf{E}[X|\mathscr{H}]$ p.s.*
5. *Si X est indépendante de $\mathscr{G}$, alors $\mathbf{E}[X|\mathscr{G}] = \mathbf{E}[X]$ p.s.*
6. *Si X est $\mathscr{G}$-mesurable, alors $\mathbf{E}[X|\mathscr{G}] = X$ p.s.*
7. *Si XY est intégrable et si X est $\mathscr{G}$-mesurable, alors $\mathbf{E}[XY|\mathscr{G}] = X\mathbf{E}[Y|\mathscr{G}]$ p.s.*

Démonstration. La démonstration est presque mot pour mot celle du théorème 3.1. Soulignons simplement que la propriété 2 se déduit maintenant de la deuxième assertion de la proposition 4.4, et disons un mot de la manière dont 2' se déduit de 2.

On a d'une part $X \leqslant |X|$, donc $\mathbf{E}[X|\mathscr{G}] \leqslant \mathbf{E}\big[|X|\big|\,\mathscr{G}\big]$, et d'autre part $-X \leqslant |X|$, donc $-\mathbf{E}[X|\mathscr{G}] \leqslant \mathbf{E}\big[|X|\big|\,\mathscr{G}\big]$. On a donc

$$\big|\mathbf{E}[X|\mathscr{G}]\big| = \max\big(\mathbf{E}[X|\mathscr{G}], -\mathbf{E}[X|\mathscr{G}]\big) \leqslant \mathbf{E}\big[|X|\big|\,\mathscr{G}\big].$$

La vérification du détail des autres propriétés peut constituer un bon exercice. □

4.3.2 Convergence dominée, Jensen

Théorème 4.7 (Théorème de convergence dominée conditionnel). *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires qui converge presque sûrement vers une variable aléatoire X. On suppose qu'il existe une variable aléatoire intégrable Y sur $(\Omega, \mathscr{F}, \mathbf{P})$ telle que pour tout $n \geqslant 0$ on ait $|X_n| \leqslant Y$ presque sûrement.*
Alors $\lim \mathbf{E}\big[|X_n - X|\big|\mathscr{G}\big] = 0$ *p.s.*
En particulier, $\lim \mathbf{E}[X_n|\mathscr{G}] = \mathbf{E}[X|\mathscr{G}]$ *p.s.*

Démonstration. On déduit ce théorème du lemme de Fatou conditionnel comme on déduit le théorème de convergence dominée usuel du lemme de Fatou usuel.

Pour tout $n \geqslant 0$, définissons la variable aléatoire réelle positive $Z_n = 2Y - |X_n - X|$. La suite $(Z_n)_{n\geqslant 0}$ converge presque sûrement vers $2Y$, donc le lemme de Fatou assure que

$$\mathbf{E}[2Y|\mathscr{G}] \leqslant \liminf \mathbf{E}[2Y - |X_n - X||\mathscr{G}] = \mathbf{E}[2Y|\mathscr{G}] - \limsup \mathbf{E}\big[|X_n - X|\big|\mathscr{G}\big].$$

Ainsi,

$$\limsup \mathbf{E}\big[|X_n - X|\big|\mathscr{G}\big] \leqslant 0$$

et comme les variables aléatoires $|X_n - X|$ sont positives, cette limite supérieure est nulle, et c'est la limite presque sûre de la suite $\big(\mathbf{E}\big[|X_n - X|\big|\mathscr{G}\big]\big)_{n\geqslant 0}$.

L'inégalité

$$\big|\mathbf{E}[X_n|\mathscr{G}] - \mathbf{E}[X|\mathscr{G}]\big| \leqslant \mathbf{E}\big[|X_n - X|\big|\mathscr{G}\big]$$

permet de déduire la deuxième assertion de la première. □

Théorème 4.8 (Inégalité de Jensen conditionnelle). *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Soit X une variable aléatoire intégrable sur $(\Omega, \mathscr{F}, \mathbf{P})$. Soit $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction convexe. On suppose que $\varphi(X)$ est intégrable.*
Alors $\varphi\big(\mathbf{E}[X|\mathscr{G}]\big) \leqslant \mathbf{E}[\varphi(X)|\mathscr{G}]$ *p.s.*

Démonstration. On utilise le fait qu'une fonction convexe est le supremum d'une famille de fonction affines. On écrit donc $\varphi = \sup\{f_i : i \in I\}$, où I est un ensemble quelconque et chaque f_i est une fonction affine sur $\mathbb{R}$. On a alors, pour tout $i \in I$,

$$f_i\big(\mathbf{E}[X|\mathscr{G}]\big) = \mathbf{E}[f_i(X)|\mathscr{G}] \leqslant \mathbf{E}[\varphi(X)|\mathscr{G}].$$

En prenant le sup sur i, on obtient l'inégalité de Jensen. □

4.3.3 Grossissement de la sous-tribu

Nous concluons ce chapitre par le pendant, dans le cas intégrable, de la proposition 3.4.

Proposition 4.9. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabiltiés. Soit X une variable aléatoire intégrable sur cet espace. Soient $\mathscr{G}$ et $\mathscr{H}$ des sous-tribus de $\mathscr{F}$. Si $\mathscr{H}$ est indépendante de $\sigma(\sigma(X) \cup \mathscr{G})$, alors*

$$\mathbf{E}[X|\sigma(\mathscr{G} \cup \mathscr{H})] = \mathbf{E}[X|\mathscr{G}] \text{ p.s.}$$

Démonstration. Il suffit d'appliquer la proposition 3.4 aux parties positive et négative de X, et de faire la différence des deux égalités obtenues. □

5 Calcul d'espérances conditionnelles

Nous avons désormais défini, montré l'existence et l'unicité presque sûre, et étudié les propriétés fondamentales de l'espérance conditionnelle des variables aléatoires positives et des variables aléatoires intégrables. Cette distinction entre cas positif et cas intégrable va devenir moins essentielle qu'elle ne l'a été jusqu'ici : les deux cas sont bel et bien distincts mais nous avons constaté qu'ils étaient très similaires, et ils continuerons à l'être. Nous les traiterons désormais de manière parallèle.

Nous avons étudié le cas fondamental dans lequel la sous-tribu est engendrée par une partition, cas dans lequel il est possible d'écrire une formule pour l'espérance conditionnelle.

Nous avons aussi commencé à étudier le cas où la sous-tribu est engendrée par une variable aléatoire, et montré que dans ce cas, l'espérance conditionnelle est une fonction de cette variable aléatoire. C'est ici que nous reprenons notre étude.

5.1 Sous-tribu engendrée par une variable aléatoire

Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités, X une variable aléatoire positive ou intégrable sur cet espace et $Z : (\Omega, \mathscr{F}) \to (E, \mathscr{E})$ une variable aléatoire quelconque sur le même espace. Nous avons démontré aux propositions 1.5 et 4.2 qu'il existe une fonction h sur $(E, \mathscr{E})$ telle que

$$\mathbf{E}[X|Z] = h(Z) p.s.$$

Commençons par déterminer à quel point une telle fonction h est unique. Nous savons que deux espérances conditionnelles de X sachant $\sigma(Z)$ sont égales presque sûrement. Dans le lemme qui suit, nous ne précisons pas l'espace d'arrivée de h_1 et h_2, qui peut être $[0, \infty]$ ou $\mathbb{R}$ suivant que nous sommes dans le cas positif ou le cas intégrable.

Lemme 5.1. *Soit $Z : (\Omega, \mathscr{F}, \mathbf{P}) \to (E, \mathscr{E})$ une application mesurable. Soient h_1, h_2 deux fonctions mesurables sur $(\Omega, \mathscr{F})$ à valeurs dans le même espace mesurable.*
On a $h_1(Z) = h_2(Z)$ $\mathbf{P}$-presque sûrement si et seulement si $h_1 = h_2$ $(\mathbf{P} \circ Z^{-1})$-presque sûrement.

Démonstration. L'égalité

$$\mathbf{P}(h_1(Z) \neq h_2(Z)) = \mathbf{P}\big(Z^{-1}(\{h_1 \neq h_2\})\big) = (\mathbf{P} \circ Z^{-1})(h_1 \neq h_2)$$

entraîne l'équivalence annoncée. □

Ce lemme nous dit donc qu'une fonction h telle que $h(Z)$ soit une espérance conditionnelle de X sachant $\sigma(Z)$ est unique modulo l'égalité presque sûre relativement à la mesure $\mathbf{P} \circ Z^{-1}$, qui est la loi de la variable aléatoire Z.

La proposition suivante précise de quoi cette fonction h dépend. Elle est écrite dans le cas intégrable, mais elle est également vraie dans le cas positif.

Proposition 5.2. *Soit d'une part*

— $(\Omega, \mathscr{F}, \mathbf{P})$ *un espace de probabilités,*

— $X : (\Omega, \mathscr{F}, \mathbf{P}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ *une variable aléatoire intégrable,*

— $Z : (\Omega, \mathscr{F}, \mathbf{P}) \to (E, \mathscr{E})$ *une variable aléatoire.*

Soit d'autre part

— $(\Omega', \mathscr{F}', \mathbf{P}')$ *un espace de probabilités,*

— $X' : (\Omega', \mathscr{F}', \mathbf{P}') \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ *une variable aléatoire intégrable,*

— $Z' : (\Omega', \mathscr{F}', \mathbf{P}') \to (E, \mathscr{E})$ *une variable aléatoire.*

On suppose que le couple (X', Z') a même loi que le couple (X, Z).

Soit enfin $h : (E, \mathscr{E}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ une fonction mesurable telle que $\mathbf{E}[X|Z] = h(Z)$ p.s. Alors

$$\mathbf{E}[X'|Z'] = h(Z') \ \ p.s.$$

En français, cette proposition dit que la fonction h ne dépend que de la loi du couple (X, Z), ce qui est un tout petit peu moins précis, mais somme toute beaucoup plus clair.

Soulignons que l'hypothèse est que les couples (X, Z) et (X', Z') ont mêmes lois. Ces lois sont des mesures de probabilités sur l'espace mesurable $(\mathbb{R} \times E, \mathscr{B}_{\mathbb{R}} \otimes \mathscr{E})$. Cette hypothèse est strictement plus forte que l'hypothèse qui consisterait à demander que X ait même loi que X' et Z même loi que Z'.

Démonstration. Montrons que $h(Z')$ est une espérance conditionnelle de X sachant $\sigma(Z')$. D'abord, $h(Z')$ est une variable aléatoire $\sigma(Z')$ mesurable. Ensuite, elle est intégrable. En effet, l'hypothèse entraîne que Z' a même loi que Z, donc

$$\mathbf{E}'[|h(Z')|] = \int_E |h| \, \mathrm{d}(\mathbf{P} \circ (Z')^{-1}) = \int_E |h| \, \mathrm{d}(\mathbf{P} \circ Z^{-1}) = \mathbf{E}[|h(Z)|] < \infty,$$

la dernière inégalité venant du fait que $h(Z)$, qui est l'espérance conditionnelle de X sachant $\sigma(Z)$, est intégrable.

Soit maintenant B' un événement de $\sigma(Z')$. Soit $C \in \mathscr{E}$ tel que $B' = (Z')^{-1}(C)$. Alors

$$\begin{aligned}\mathbf{E}'[h(Z')\mathbf{1}_{B'}] &= \mathbf{E}'[h(Z')\mathbf{1}_C(Z')] \\ &= \mathbf{E}[h(Z)\mathbf{1}_C(Z)] = \mathbf{E}[h(Z)\mathbf{1}_{Z^{-1}(C)}] = \mathbf{E}[X\mathbf{1}_{Z^{-1}(C)}] = \mathbf{E}[X\mathbf{1}_C(Z)] \\ &= \mathbf{E}'[X'\mathbf{1}_C(Z')] = \mathbf{E}[X'\mathbf{1}_{B'}],\end{aligned}$$

où nous avons utilisé successivement : la définition de C, le fait que Z et Z' ont même loi, l'égalité $\mathbf{1}_C(Z) = \mathbf{1}_{Z^{-1}(C)}$, le fait que $h(Z)$ est une espérance conditionnelle de X sachant $\sigma(Z)$, à nouveau l'égalité $\mathbf{1}_C(Z) = \mathbf{1}_{Z^{-1}(C)}$ mais dans l'autre sens, puis le fait que (X, Z) et (X', Z') ont même loi, et enfin à nouveau la définition de C.

La démonstration est complète. □

Ceci ne nous dit toujours pas comment calculer la fameuse fonction h. Nous allons enfin étudier une méthode générale qui permet de la chercher, et parfois de la trouver.

5.2 La méthode de la fonction muette

Donnons la méthode sous la forme d'un énoncé.

Proposition 5.3 (Méthode de la fonction muette). *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilités, X une variable aléatoire positive ou intégrable sur cet espace et $Z : (\Omega, \mathscr{F}) \to (E, \mathscr{E})$ une variable aléatoire quelconque sur le même espace. Soit h une fonction sur $(E, \mathscr{E})$, positive (dans le cas positif) ou telle que $h(Z)$ soit intégrable (dans le cas intégrable). Les assertions suivantes sont équivalentes.*

1. *On a $\mathbf{E}[X|Z] = h(Z)$ p.s.*
2. *Pour toute fonction mesurable positive bornée g sur $(E, \mathscr{E})$, on a l'égalité*

$$\mathbf{E}[Xg(Z)] = \mathbf{E}[h(Z)g(Z)].$$

Nous verrons après la démonstration en quoi cette proposition fournit effectivement une méthode pour calculer des espérances conditionnelles, ce qui pour l'instant n'est pas tout fait évident.

Comme dans l'énoncé, nous menons en parallèle les cas positif et intégrable.

Démonstration. Montrons que $2 \Rightarrow 1$. La variable aléatoire $h(Z)$ est positive (ou intégrable) et mesurable par rapport à $\sigma(Z)$. Soit maintenant $B \in \sigma(Z)$. Alors il existe $C \in \mathscr{E}$ tel que $B = Z^{-1}(C)$. En appliquant 2 à la fonction $g = \mathbf{1}_C$, qui est positive et bornée, on trouve

$$\mathbf{E}[X\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_C(Z)] = \mathbf{E}[h(Z)\mathbf{1}_C(Z)] = \mathbf{E}[h(Z)\mathbf{1}_B],$$

ce qui montre que $h(Z)$ est une espérance conditionnelle de X sachant $\sigma(Z)$.

Montrons maintenant que $1 \Rightarrow 2$, sans trop entrer dans les détails car ce n'est pas l'implication qui nous intéresse le plus. Supposons donc l'assertion 1. Le même raisonnement que nous venons de faire mais lu à l'envers montre que l'assertion 2 est vraie pour toute fonction g de la forme $\mathbf{1}_C$ avec $C \in \mathscr{E}$. Par linéarité, l'assertion 2 est donc vraie pour toute fonction g étagée positive. Par convergence monotone dans le cas positif, ou par convergence dominée dans le cas intégrable, on en déduit que l'assertion est vraie pour toute fonction mesurable positive bornée. □

La méthode de la fonction muette, qui dérive de cette proposition, est la suivante. On dispose d'un couple de variables aléatoires (X, Z) et on cherche à calculer $\mathbf{E}[X|Z]$. On se donne une fonction (muette) g mesurable positive bornée sur l'espace où Z prend ses valeurs. On écrit alors

$$\mathbf{E}[Xg(Z)] = \dots$$

et on fait ce qu'on peut, ce qu'il faut, pour aboutir à une expression de la forme

$$\dots = \mathbf{E}[h(Z)g(Z)],$$

sans jamais, bien entendu, rien supposer de plus sur g que le fait qu'elle est mesurable positive et bornée. Cette fonction g ne sert que de support pour mener le calcul de manière agréable, dans la mesure du possible.

Comment fait-on pour passer de $\mathbf{E}[Xg(Z)]$ à $\mathbf{E}[h(Z)g(Z)]$? Cela dépend du contexte, de ce que l'on sait sur X et Z. En tout cas, la proposition 5.2 nous dit qu'une information importante pour calculer $\mathbf{E}[X|Z]$ soit la loi du couple (X, Z), et c'est exactement l'information qui permet de calculer $\mathbf{E}[Xg(Z)]$.

Apprendre à faire fonctionner cette méthode dans des cas concrets, explicites, un peu variés, est l'une des tâches qui vous incombent dans la partie de ce cours consacrée à l'espérance conditionnelle. Même avec les meilleurs conseils du monde, ceci ne pourra se faire qu'avec du temps, par la pratique personnelle, en cherchant des exercices, ceux des TD, ceux des sujets d'examen des années passées, ou ceux que vous inventerez.

5.3 Un exemple : le cas des vecteurs aléatoires à densité

Sur un espace de probabilités $(\Omega, \mathscr{F}, \mathbf{P})$, soit (X, Z) un couple de variables aléatoires réelles. Supposons que X est intégrable, et supposons que la loi de (X, Z) admet par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$ une densité qui est en tout point strictement positive. Notons $f_{(X,Z)} : \mathbb{R}^2 \to]0, \infty[$ cette densité. Cherchons une expression de l'espérance conditionnelle de X sachant Z, en appliquant la méthode de la fonction muette.

Donnons-nous donc une fonction $g : \mathbb{R} \to \mathbb{R}$ mesurable positive et bornée, et calculons $\mathbf{E}[Xg(Z)]$. Pour reconnaître, à la fin de notre calcul, que nous aurons abouti à une expression de la forme $\mathbf{E}[h(Z)g(Z)]$, pour une certaine fonction h, il nous faut connaître la loi de Z.

Puisque la loi du vecteur aléatoire (X, Z) admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$, la loi de Z admet une densité f_Z qui s'obtient à partir de la densité du couple (X, Z) en intégrant par rapport à la variable correspondant à X :

$$\forall z \in \mathbb{R},\ f_Z(z) = \int_{\mathbb{R}} f_{(X,Z)}(x, z)\ \mathrm{d}x.$$

Faisons maintenant le calcul. Il n'y a qu'une manière de commencer :

$$\mathbf{E}[Xg(Z)] = \int_{\mathbb{R}^2} x g(z)\, f_{(X,Z)}(x, z)\ \mathrm{d}x\mathrm{d}z.$$

Nous voulons faire apparaître une intégrale par rapport à la variable z, faisons-le, en utilisant le théorème de Fubini :

$$\mathbf{E}[Xg(Z)] = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x f_{(X,Z)}(x, z)\ \mathrm{d}x \right) g(z)\ \mathrm{d}z.$$

Nous cherchons, plus précisément, à aboutir à une intégrale de la forme $\int_{\mathbb{R}} h(z)g(z)\ f_Z(z)\ \mathrm{d}z$. Nous n'en sommes pas loin, mais il nous manque la densité de la loi de Z. Forçons la à apparaître, en tirant parti du fait que nos hypothèses impliquent qu'elle est strictement positive en tout point :

$$\mathbf{E}[Xg(Z)] = \int_{\mathbb{R}} \frac{\int_{\mathbb{R}} x f_{(X,Z)}(x, z)\ \mathrm{d}x}{f_Z(z)}\, g(z) f_Z(z)\ \mathrm{d}z.$$

La fonction h est maintenant écrite devant nous : il suffit de poser

$$h(x) = \frac{\int_{\mathbb{R}} x f_{(X,Z)}(x, z)\ \mathrm{d}x}{f_Z(z)} = \frac{\int_{\mathbb{R}} x f_{(X,Z)}(x, z)\ \mathrm{d}x}{\int_{\mathbb{R}} f_{(X,Z)}(x, z)\ \mathrm{d}x}$$

pour avoir

$$\mathbf{E}[Xg(Z)] = \int_{\mathbb{R}} h(z)g(z)\, f_Z(z)\ \mathrm{d}z = \mathbf{E}[h(Z)g(Z)].$$

Nous avons réussi, et démontré que $\mathbf{E}[X|Z] = h(Z)$.

5.4 Mélange d'indépendance et de mesurabilité

Cette section est consacrée à un résultat qui concerne une situation qui se présente souvent dans des applications au monde réel de la notion d'espérance conditionnelle (il y en a). Cette situation est celle où la variable dont on cherche l'espérance conditionnelle sachant une sous-tribu $\mathscr{G}$ est exprimée comme une fonction de deux variables aléatoires, l'une mesurable par rapport à $\mathscr{G}$, l'autre indépendante de $\mathscr{G}$.

Proposition 5.4. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabiltiés. Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$.*
Soient $X : (\Omega, \mathscr{F}) \to (A, \mathscr{A})$ et $Y : (\Omega, \mathscr{F}) \to (B, \mathscr{B})$ des variables aléatoires. Soit f une fonction mesurable sur $(A \times B, \mathscr{A} \otimes \mathscr{B})$, positive ou telle que $f(X, Y)$ soit intégrable.
On suppose que X est mesurable par rapport à $\mathscr{G}$ et que Y est indépendante de $\mathscr{G}$. On définit h en posant

$$\forall a \in A,\ h(a) = \mathbf{E}[f(a, Y)].$$

Alors $\mathbf{E}[f(X, Y)|\mathscr{G}] = h(X)$ p.s.

Explicitons la définition de h : c'est une fonction numérique (positive ou à valeurs réelles) sur $(A, \mathscr{A})$, l'espace mesurable dans lequel X prend ses valeurs. Pour tout $a \in A$, on a

$$h(a) = \mathbf{E}[f(a, Y)] = \int_B f(a, b)\ \mathrm{d}(\mathbf{P} \circ Y^{-1})(b).$$

Certains auteurs écrivent le résultat sous la forme

$$\mathbf{E}[f(X, Y)|\mathscr{G}] = \int_B f(X, b)\ \mathrm{d}(\mathbf{P} \circ Y^{-1})(b),$$

ce qui est également juste. J'attire toutefois votre attention sur le fait que dans la formulation de l'énoncé, on lit $f(a, Y)$, et que dans l'équation ci-dessus, on lit $f(X, b)$. Il n'y a pas de contradiction, tout est correct, mais il y a là une source de confusion possible, dont j'ai parfois constaté les conséquences néfastes.

Démonstration. Écrivons la démonstration dans le cas intégrable, le cas positif étant analogue et plus simple.

Convenons de noter P_X la loi de X, P_Y celle de Y, et $P_{(X,Y)}$ la loi du couple. Remarquons que nos hypothèses impliquent que X est indépendant de Y, si bien que $P_{(X,Y)} = P_X \otimes P_Y$.

Montrons que $h(X)$ est une espérance conditionnelle de $f(X, Y)$ sachant $\mathscr{G}$. D'abord, c'est une fonction de X, donc une variable aléatoire mesurable par rapport à $\sigma(X)$, qui est une sous-tribu de $\mathscr{G}$, puisque X est $\mathscr{G}$-mesurable. Ainsi, $h(X)$ est bien $\mathscr{G}$-mesurable.

Vérifions maintenant que $h(X)$ est intégrable. Nous avons, grâce au théorème de Fubini,

$$\begin{aligned}
\mathbf{E}[|h(X)|] &= \int_A \left| \int_B f(a, b)\ \mathrm{d}P_Y(b) \right| \mathrm{d}P_X(a). \\
&\leqslant \int_{A \times B} |f(a, b)|\ \mathrm{d}\big(P_X \otimes P_Y\big)(a, b) \\
&= \int_{A \times B} |f(a, b)|\ \mathrm{d}P_{(X,Y)}(a, b) = \mathbf{E}[|f(X, Y)|] < \infty
\end{aligned}$$

si bien que $h(X)$ est intégrable.

Soit enfin B un élément de $\mathscr{G}$. Calculons $\mathbf{E}[h(X)\mathbf{1}_B]$. Pour cela, notons $P_{(X,\mathbf{1}_B,Y)}$ la loi du triplet $(X, \mathbf{1}_B, Y)$ et $P_{(X,\mathbf{1}_B)}$ celle du couple $(X, \mathbf{1}_B)$. Puisque Y est indépendante de $\mathscr{G}$, qui contient $\sigma(X, \mathbf{1}_B)$, on a $P_{(X,\mathbf{1}_B,Y)} = P_{(X,\mathbf{1}_B)} \otimes P_Y$. Nous avons donc

$$\begin{aligned}
\mathbf{E}[h(X)\mathbf{1}_B] = \int_{A \times \mathbb{R}} h(a)u\ \mathrm{d}P_{(X,\mathbf{1}_B)}(a, u) &= \int_{A \times \mathbb{R}} \left(\int_B f(a, b)\ \mathrm{d}P_Y(b) \right) u\ \mathrm{d}P_{(X,\mathbf{1}_B)}(a, u) \\
&= \int_{A \times \mathbb{R} \times B} f(a, b)u\ \mathrm{d}P_{(X,\mathbf{1}_B,Y)}(a, u, b) = \mathbf{E}[f(X, Y)\mathbf{1}_B],
\end{aligned}$$

ce qui conclut la démonstration. □

5.5 Le cas des vecteurs gaussiens

Les vecteurs aléatoires gaussiens ont vis-à-vis de l'espérance conditionnelle des propriétés particulières, et ce n'est pas surprenant si l'on pense d'une part au fait que l'espérance conditionnelle est une projection orthogonale dans L^2, et d'autre part au fait que la loi gaussienne (multivariée) est liée dans sa définition même à la géométrie euclidienne.

Proposition 5.5. *Soit $(X, Z_1, \dots, Z_n)$ un vecteur aléatoire gaussien. Alors il existe des réels $a_0, a_1, \dots, a_n$ tels que*

$$\mathbf{E}[X|Z_1, \dots, Z_n] = a_0 + a_1 Z_1 + \dots + a_n Z_n \text{ p.s.}$$

On savait, de manière générale, que cette espérance conditionnelle était une fonction de $Z_1, \dots, Z_n$, et cette proposition nous dit que c'est une fonction *affine*.

Rappelons que l'hypothèse que le vecteur $(X, Z_1, \dots, Z_n)$ est gaussien signifie que toute combinaison linéaire de $X, Z_1, \dots, Z_n$, est une variable aléatoire réelle gaussienne, étant entendu que nous considérons les variables aléatoires constantes comme gaussiennes.

Explicitons ensuite la notation $\mathbf{E}[X|Z_1, \dots, Z_n]$: c'est bien entendu l'espérance conditionnelle de X sachant la tribu engendrée par $(Z_1, \dots, Z_n)$, qu'on peut de manière équivalente considérer comme une famille de n variables aléatoires réelles, ou comme un vecteur aléatoire de dimension n.

La propriété des vecteur gaussiens dont nous allons avoir besoin est celle qui lie indépendance et orthogonalité dans L^2. Plus précisément, si $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ est un vecteur gaussien, et si pour tout $i \in \{1, \dots, n\}$ et tout $j \in \{1, \dots, m\}$ on a $\mathrm{Cov}(X_i, Y_j) = 0$, alors $(X_1, \dots, X_n)$ est indépendant de $(Y_1, \dots, Y_m)$. Dans le cas où toutes les variables aléatoires sont centrées (c'est-à-dire d'espérance nulle), la condition $\mathrm{Cov}(X_i, Y_j) = 0$ s'écrit $\mathbf{E}[X_i Y_j] = 0$, et c'est exactement la condition que X_i et Y_j sont orthogonaux dans l'espace L^2 de notre espace de probabilité sous-jacent.

Démonstration. Le cœur de la preuve consiste à démontrer qu'il existe des réels $a_1, \dots, a_n$ tels que la variable aléatoire $X - a_1 Z_1 - \dots - a_n Z_n$ soit indépendante de $(Z_1, \dots, Z_n)$. Ceci peut se démontrer de plusieurs manières, et une fois que c'est fait, on calcule de deux manières l'espérance conditionnelle de $X - a_1 Z_1 - \dots - a_n Z_n$ sachant $(Z_1, \dots, Z_n)$. D'une part, par indépendance, c'est une constante, que nous notons a_0 :

$$\mathbf{E}[X - a_1 Z_1 - \dots - a_n Z_n | Z_1, \dots, Z_n] = \mathbf{E}[X - a_1 Z_1 - \dots - a_n Z_n] = a_0.$$

D'autre part, par linéarité, et par le fait que chacune des variables aléatoires $Z_1, \dots, Z_n$ est mesurable par rapport à $\sigma(Z_1, \dots, Z_n)$, elle vaut

$$\mathbf{E}[X - a_1 Z_1 - \dots - a_n Z_n | Z_1, \dots, Z_n] = \mathbf{E}[X|Z_1, \dots, Z_n] - aZ_1 - \dots - a_n Z_n.$$

En comparant les deux expressions, on obtient le résultat.

Venons-en à l'existence de réels $a_1, \dots, a_n$ tels que $X - a_1 Z_1 - \dots - a_n Z_n$ soit indépendante de $(Z_1, \dots, Z_n)$. Commençons par observer que cette propriété d'indépendance n'est pas affectée par l'ajout de constantes aux variables aléatoires $X, Z_1, \dots, Z_n$. En effet, cet ajout ne change ni la tribu $\sigma(Z_1, \dots, Z_n)$, ni la tribu engendrée par la variable aléatoire $X - a_1 Z_1 - \dots - a_n Z_n$, qui sera elle-même modifiée par l'ajout d'une constante. Nous pouvons donc, sans perdre de généralité, supposer nos variables aléatoire centrées.

Supposons donc nos variables aléatoires centrées. Considérons, dans l'espace L^2, le projeté orthogonal de X sur le sous-espace vectoriel $\mathrm{Vect}(Z_1, \dots, Z_n)$. C'est une combinaison linéaire

de $Z_1, \ldots, Z_n$, que nous notons $Y = a_1 Z_1 + \ldots + a_n Z_n$. Alors $X - Y$ est orthogonal au sous-espace vectoriel $\mathrm{Vect}(Z_1, \ldots, Z_n)$. Ceci signifie exactement que $\mathrm{Cov}(X - Y, Z_i) = 0$ pour tout $i \in \{1, \ldots, n\}$. Or le vecteur aléatoire $(X - Y, Z_1, \ldots, Z_n)$ est gaussien. Donc $X - Y = X - a_1 Z_1 - \ldots - a_n Z_n$ est indépendant de $Z_1, \ldots, Z_n$. □

En pratique, on détermine les réels $a_1, \ldots, a_n$ en résolvant le système d'équations linéaires

$$\mathrm{Cov}(Z_1, Z_i)\, a_1 + \ldots + \mathrm{Cov}(Z_n, Z_i)\, a_n = \mathrm{Cov}(X, Z_i), \;\; i \in \{1, \ldots, n\}.$$

Le cas où $n = 1$ est bien sûr plus simple que le cas général. Dans ce cas, on se donne un vecteur gaussien (X, Z). On cherche un réel a tel que $X - aZ$ soit indépendant de Z. On résout pour cela l'équation

$$\mathrm{Cov}(X - aZ, Z) = 0.$$

Si $\mathrm{Cov}(Z, Z) = \mathrm{Var}(Z) = 0$, ce qui signifie que Z est constante, alors X est indépendante de Z. Dans ce cas, on pose $a = 0$ et $b = \mathbf{E}[X]$, et on a $\mathbf{E}[X|Z] = aZ + b$ p.s.

Si $\mathrm{Var}(Z) > 0$, on pose $a = \mathrm{Cov}(X, Z)/\mathrm{Var}(Z)$. On pose ensuite $b = \mathbf{E}[X - aZ]$, et on aboutit à $\mathbf{E}[X|Z] = aZ + b$.

Deuxième partie

Compléments

A Tribus

Cette annexe est consacrée à des rappels et des compléments sur la notion de tribu. Il s'agit en effet d'une des notions fondamentales de la théorie de la mesure, mais qui, dans la pratique de la théorie de l'intégration et d'un premier cours de probabilités, reste au second plan. Dans ce cours, elle va au contraire jouer un rôle essentiel.

A.1 Tribus et tribus engendrées

Définition A.1. *Soit E un ensemble. On appelle* tribu sur E *une classe $\mathscr{E}$ de parties de E, c'est-à-dire un sous-ensemble $\mathscr{E} \subseteq \mathscr{P}(E)$ de l'ensemble des parties de E, qui vérifie les trois propriétés suivantes :*

1. $\varnothing \in \mathscr{E}$,
2. *pour tout $A \in \mathscr{E}$, on a $E \setminus A \in \mathscr{E}$,*
3. *pour toute suite $(A_n)_{n\geqslant 0}$ d'éléments de $\mathscr{E}$, on a $\bigcup_{n\geqslant 0} A_n \in \mathscr{E}$.*

Une tribu sur E contient la partie vide $\varnothing$, la partie pleine $E = E \setminus \varnothing$. Pour toute tribu $\mathscr{E}$ sur E, on a donc

$$\{\varnothing, E\} \subseteq \mathscr{E} \subseteq \mathscr{P}(E).$$

De plus, $\{\varnothing, E\}$ et $\mathscr{P}(E)$ sont bien des tribus sur E.

Une tribu sur E est stable par toutes les opérations ensemblistes : union, intersection, différence symétrique. Elle est stable par unions et intersections dénombrables.

Un fait d'importance fondamentale est le suivant.

Proposition A.2. *Soit E un ensemble. Soit $(\mathscr{E}_i)_{i\in I}$ une famille de tribus sur E. Alors $\bigcap_{i\in I} \mathscr{E}_i$ est une tribu sur E.*

Dans cette proposition, I est un ensemble quelconque, et pour chaque $i \in I$, on se donne une tribu $\mathscr{E}_i$ sur E. La proposition affirme que la classe de parties

$$\bigcap_{i\in I} \mathscr{E}_i = \{A \subseteq E : \forall i \in I, A \in \mathscr{E}_i\}$$

est une tribu sur E. C'est facile à démontrer, faites-le !

Cette proposition permet de définir la tribu engendrée par une classe de parties quelconque de E.

Proposition A.3. *Soit $\mathscr{C} \subseteq \mathscr{P}(E)$ une classe de parties sur E. Il existe une unique tribu sur E qui contient $\mathscr{C}$ et qui est incluse dans toute tribu sur E qui contient $\mathscr{C}$.*

On appelle cette tribu la *tribu engendrée par* $\mathscr{C}$ et on la note $\sigma(\mathscr{C})$. En d'autres termes, c'est la plus petite tribu (au sens de l'inclusion) qui contienne $\mathscr{C}$.

Démonstration. Considérons l'intersection de toutes les tribus sur E qui contiennent $\mathscr{C}$. C'est, d'après la proposition précédente, une tribu sur E. Elle contient $\mathscr{C}$, et elle est, par construction, incluse dans toute tribu qui contient $\mathscr{C}$. Ceci prouve l'existence de la tribu engendrée.

Son unicité découle du fait que si deux tribus ont les propriétés demandées, alors chacune est incluse dans l'autre, donc elles sont égales. □

Exercice A.1. *Soit $\mathscr{C}$ une classe de parties de E. Que vaut $\sigma(\sigma(\mathscr{C}))$?*

Exercice A.2. *Soit E un ensemble.*

1. *Soit A une partie de E qui n'est ni vide ni égale à E. Calculer la tribu engendrée par la classe $\{A\}$.*
2. *Soient A et B deux parties de E telles qu'aucune des parties $(A \cup B)^c$, $A \setminus B$, $B \setminus A$, $A \cap B$ ne soient vides. Calculer $\sigma(\{A, B\})$.*

Exercice A.3. *Déterminer, sur un ensemble E, la tribu engendrée par la classe des singletons. Autrement dit, calculer*

$$\sigma\big(\{\{x\} : x \in E\}\big).$$

A.2 Exemples

A.2.1 Tribu engendrée par une partition

Un exemple fondamental pour nous de tribu engendrée sur un ensemble E est celui de la tribu engendrée par une partition de E. La section 1.4.1 est consacrée à cet exemple, ainsi que l'annexe D. Nous n'y revenons pas ici et renvoyons à ces paragraphes.

A.2.2 Tribu borélienne de $\mathbb{R}$ et de $\mathbb{R}^n$

Un autre exemple est celui de la tribu borélienne d'un espace topologique, dont nous allons dire quelques mots dans le cas de $\mathbb{R}$.

Sur l'ensemble $\mathbb{R}$ des nombres réels, nous disposons d'une classe de parties, qui est la classe, que nous allons noter $\mathscr{O}$, des parties ouvertes. Cette classe n'est pas une tribu, parce qu'elle n'est pas stable par complémentaire. On pourrait penser qu'en lui adjoignant les parties fermées (qui sont les complémentaires des parties ouvertes), on obtient une tribu.

Exercice A.4. *Montrer que la classe $\{A \subseteq \mathbb{R} : A$ ouverte ou fermée$\}$ n'est pas une tribu sur $\mathbb{R}$.*

Définition A.4. *On appelle* tribu borélienne *de $\mathbb{R}$ la tribu sur $\mathbb{R}$ engendrée par la classe des parties ouvertes.*

On note $\mathscr{B}_{\mathbb{R}}$ la tribu borélienne de $\mathbb{R}$.

Exercice A.5. *Montrer que la tribu borélienne de $\mathbb{R}$ est engendrée par chacune des classes de parties suivantes :*

1. *les intervalles ouverts,*
2. *les intervalles fermés,*
3. *les intervelles fermés bornés,*

4. *les intervalles de la forme* $[a, b[$,
5. *les intervalles de la forme* $]-\infty, a[$,
6. *chacune des classes précédentes où l'on impose en plus aux intervalles d'avoir leurs extrémités dans une partie dense fixée de* $\mathbb{R}$, *par exemple l'ensemble* $\mathbb{Q}$ *des nombres rationnels.*

La définition de la tribu borélienne s'étend à tout espace topologique, c'est-à-dire à tout ensemble sur lequel on a une classe de parties, dites ouvertes, qui satisfait les axiomes d'une topologie. Par exemple, c'est le cas de $\mathbb{R}^n$ pour tout $n \geqslant 1$. Ainsi, la tribu borélienne de $\mathbb{R}^2$ est-elle la tribu $\mathscr{B}_{\mathbb{R}^2}$ sur $\mathbb{R}^2$ engendrée par la classe des parties ouvertes de $\mathbb{R}^2$.

A.2.3 Tribu produit

La notion de tribu engendrée permet de munir le produit cartésien de deux espaces mesurables d'une tribu.

Définition A.5. *Soient* $(E, \mathscr{E})$ *et* $(F, \mathscr{F})$ *deux espaces mesurables. Sur le produit cartésien* $E \times F$ *de* E *et* F, *on appelle* tribu produit de $\mathscr{E}$ et $\mathscr{F}$ *la tribu*

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

La tribu produit est donc la tribu engendrée par les parties de la forme $A \times B$, où A et B appartiennent aux tribus respectives sur les deux facteurs du produit cartésien — on appelle de telles parties des *pavés mesurables*.

Cette définition nous amène à la question suivante : sur l'ensemble $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, nous avons maintenant les deux tribus $\mathscr{B}_{\mathbb{R}^2}$ et $\mathscr{B}_{\mathbb{R}} \otimes \mathscr{B}_{\mathbb{R}}$. Sont-elles égales ?

Exercice A.6. *Montrer que toute partie ouverte de* $\mathbb{R}^2$ *est une union de rectangles ouverts* $]a, b[\times]c, d[$ *avec* a, b, c, d *rationnels. En déduire une inclusion entre les tribus* $\mathscr{B}_{\mathbb{R}^2}$ *et* $\mathscr{B}_{\mathbb{R}} \otimes \mathscr{B}_{\mathbb{R}}$.

A.3 Applications mesurables

Une paire $(E, \mathscr{E})$ formée d'un ensemble E et d'une tribu $\mathscr{E}$ sur E s'appelle un *espace mesurable*.

Définition A.6. *Soient* $(E, \mathscr{E})$ *et* $(F, \mathscr{F})$ *deux espaces mesurables. On dit qu'une application* $f : E \to F$ *est mesurable par rapport aux tribus* $\mathscr{E}$ *et* $\mathscr{F}$ *si l'image réciproque par* f *de tout élément de* $\mathscr{F}$ *appartient à* $\mathscr{E}$ *:*

$$\forall B \in \mathscr{F},\ f^{-1}(B) \in \mathscr{E}.$$

Rappelons que $f^{-1}(B)$ est l'ensemble des éléments de E qui sont envoyés par f dans B :

$$f^{-1}(B) = \{x \in E : f(x) \in B\}$$

et que cette définition ne suppose rien sur l'application f. En particulier, f^{-1} ne désigne *pas* ici une application réciproque de f, qui n'a aucune raison d'exister.

L'application identité d'un espace mesurable dans lui-même est mesurable, et la composée de deux applications mesurables est mesurable : si $f : (E, \mathscr{E}) \to (F, \mathscr{F})$ et $g : (F, \mathscr{F}) \to (G, \mathscr{G})$ sont mesurables, alors $g \circ f : (E, \mathscr{E}) \to (G, \mathscr{G})$ est mesurable.

Exercice A.7. *Soient* $\mathscr{E}$ *et* $\mathscr{F}$ *deux tribus sur le même ensemble* E. *À quelle condition l'application identique* $\mathrm{id}_E : (E, \mathscr{E}) \to (E, \mathscr{F})$ *est-elle mesurable ?*

Exercice A.8. *Soit $f : (E, \mathscr{E}) \to (F, \mathscr{F})$ une application mesurable.*

1. *Si $\mathscr{F}'$ est une tribu sur F et si $\mathscr{F}' \subseteq \mathscr{F}$, alors $f : (E, \mathscr{E}) \to (F, \mathscr{F}')$ est encore mesurable.*
2. *Si $\mathscr{E}'$ est une tribu sur E et si $\mathscr{E} \subseteq \mathscr{E}'$, alors $f : (E, \mathscr{E}') \to (F, \mathscr{F})$ est encore mesurable.*

Proposition A.7. *Soient E un ensemble et $(F, \mathscr{F})$ un espace mesurable. Soit $f : E \to F$ une application. La classe de parties*

$$\sigma(f) = \{f^{-1}(B) : B \in \mathscr{F}\}$$

est une tribu sur E et l'application $f : (E, \sigma(f)) \to (F, \mathscr{F})$ est mesurable. De plus, si $\mathscr{E}$ est une tribu sur E telle que $f : (E, \mathscr{E}) \to (F, \mathscr{F})$ soit mesurable, alors $\sigma(f) \subseteq \mathscr{E}$.

Ainsi, la tribu $\mathscr{F}$ sur F étant fixée, $\sigma(f)$ est la plus petite tribu sur E qui rende f mesurable. On l'appelle la *tribu sur E engendrée par f*.

Exercice A.9. *Démontrer la proposition ci-dessus.*

La tribu $\sigma(f)$ peut être décrite comme l'intersection de toutes les tribus sur E qui rendent f mesurable (la tribu $\mathscr{F}$ étant toujours fixée sur F). L'existence de cette tribu est donc liée à la proposition A.2.

Il se trouve qu'il est également possible de faire une construction analogue en échangeant les rôles de E et F. Pour justifier la notation $f_*\mathscr{E}$ qui apparaît ci-dessous, indiquons qu'on note parfois $f^*\mathscr{F}$ la tribu $\sigma(f)$.

Proposition A.8. *Soient $(E, \mathscr{E})$ un espace mesurable et F un ensemble. Soit $f : E \to F$ une application. La classe de parties*

$$f_*\mathscr{E} = \{B \subseteq F : f^{-1}(B) \in \mathscr{E}\}$$

est une tribu sur F et l'application $f : (E, \mathscr{E}) \to (F, f_*\mathscr{E})$ est mesurable. De plus, si $\mathscr{F}$ est une tribu sur F telle que $f : (E, \mathscr{E}) \to (F, \mathscr{F})$ soit mesurable, alors $\mathscr{F} \subseteq f_*\mathscr{E}$.

Autrement dit, $f_*\mathscr{E}$ est la plus grande tribu sur F qui rende E mesurable. Cette tribu s'appelle la *tribu image* de $\mathscr{E}$ par f.

Exercice A.10. *Soient $(E, \mathscr{E})$ et $(F, \mathscr{F})$ deux espaces mesurables. Soit $f : E \to F$ une application. Montrer que f est mesurable par rapport à $\mathscr{E}$ et $\mathscr{F}$ si et seulement si $f^*\mathscr{F} \subseteq \mathscr{E}$, si et seulement si $f_*\mathscr{E} \supseteq \mathscr{F}$.*

Montrons comment la notion de tribu image peut être utilisée.

Proposition A.9. *Soit $f : \mathbb{R}^n \to \mathbb{R}^m$ une application continue. Alors f est mesurable par rapport aux tribus $\mathscr{B}_{\mathbb{R}^n}$ et $\mathscr{B}_{\mathbb{R}^m}$.*

Démonstration. En effet, soit U une partie ouverte de $\mathbb{R}^m$. Alors puisque f est continue, $f^{-1}(U)$ est ouverte, donc appartient à $\mathscr{B}_{\mathbb{R}^n}$. Ainsi, la tribu image de $\mathscr{B}_{\mathbb{R}^n}$ par f contient toutes les parties ouvertes de $\mathbb{R}^m$. Comme c'est une tribu sur $\mathbb{R}^m$, elle contient la tribu borélienne de $\mathbb{R}^m$. Il s'ensuit que f est mesurable. □

A.4 Produits cartésiens

Revenons à la situation d'un produit cartésien. Si E et F sont deux ensembles, leur produit cartésien $E \times F$ est muni de deux applications $p_E : E \times F \to E$ et $p_F : E \times F \to F$ de projection sur les coordonnées.

Proposition A.10. *Soient $(E, \mathscr{E})$ et $(F, \mathscr{F})$ deux espaces mesurables. Sur $E \times F$, la tribu produit $\mathscr{E} \otimes \mathscr{F}$ est la plus petite tribu sur $E \times F$ qui rende mesurables les applications p_E et p_F.*

Démonstration. Toute tribu sur $E \times F$ qui rend p_E et p_F mesurables contient les parties de la forme

$$p_E^{-1}(A) \cap p_F^{-1}(B) = A \times B,\ A \in \mathscr{E}, B \in \mathscr{F}$$

donc elle contient la tribu $\mathscr{E} \otimes \mathscr{F}$. □

Nous pouvons utiliser cette proposition pour déterminer quand une application à valeurs dans un espace produit est mesurable. Étant donné trois ensembles E, F et G, et une application $f : E \to G \times H$, nous notons $g = p_G \circ f$ et $h = p_H \circ f$, de sorte que $f = (g, h)$.

Proposition A.11. *Soient $(E, \mathscr{E})$, $(G, \mathscr{G})$ et $(H, \mathscr{H})$ trois espaces mesurables. Soit $f = (g, h) : E \to G \times H$ une application. Alors f est mesurable par rapport à $\mathscr{E}$ et $\mathscr{G} \otimes \mathscr{H}$ si et seulement si $g : (E, \mathscr{E}) \to (G, \mathscr{G})$ et $h : (E, \mathscr{E}) \to (H, \mathscr{H})$ sont mesurables.*

Démonstration. Si f est mesurable, alors $g = p_G \circ f$ et $h = p_H \circ f$, qui sont des composées d'applications mesurables, le sont aussi.

Réciproquement, supposons que g et h soient mesurables. Alors pour tous $A \in \mathscr{G}$ et $B \in \mathscr{H}$,

$$f^{-1}(A \times B) = g^{-1}(A) \cap h^{-1}(B) \in \mathscr{E},$$

si bien que la tribu image de $\mathscr{E}$ par f contient la classe des pavés mesurables sur $G \times H$. Elle contient donc la tribu produit $\mathscr{G} \otimes \mathscr{H}$, et f est mesurable. □

Concluons par un point de notation. Étant donné deux applications (ou plus, mais nous nous sommes restreints au cas de produits de deux facteurs) $g : E \to (G, \mathscr{G})$ et $h : E \to (H, \mathscr{H})$ entre un ensemble et des espaces mesurables, on note $\sigma(g, h)$ la plus petite tribu sur E qui rende g et h mesurables. C'est donc, d'une part, la tribu

$$\sigma(g, h) = \sigma(\sigma(g) \cup \sigma(h)),$$

et d'autre part la plus petite tribu sur l'ensemble E qui rende mesurable l'application $(g, h) : E \to (G \times H, \mathscr{G} \otimes \mathscr{H})$.

B
Théorème de la classe monotone

Cette annexe est consacrée, comme son titre l'indique, au théorème de la classe monotone. Il s'agit d'un résultat de théorie de la mesure, et plus précisément d'un résultat qui concerne les tribus et d'autres classes de parties d'un ensemble (il n'y est pas question de mesure). Il s'agit d'un résultat technique élémentaire, ce qui ne veut pas dire facile, à comprendre, à démontrer, ou à utiliser, mais ce qui veut dire que son énoncé et sa démonstration ne font presque intervenir que les définitions de quelques classes de parties d'un ensemble.

Le but de cette annexe est de donner un énoncé, et une démonstration de ce résultat, dont certaines parties sont laissées en exercice.

B.1 π-systèmes et λ-systèmes

Soit E un ensemble, fixé une fois pour toutes. On appellera *classe de parties* de E une partie de l'ensemble des parties de E.

Définition B.1 (π-système)**.** *On dit qu'une classe de parties $\mathscr{I}$ de E est un* π-système *si*

1. $E \in \mathscr{I}$,
2. *pour tous* $A, B \in \mathscr{I}$, *on a* $A \cap B \in \mathscr{I}$.

Certains auteurs n'exigent pas qu'un π-système contienne l'ensemble E. Parmi ceux-ci, certains exigent qu'un π-système soit non vide, mais d'autres permettent qu'il soit vide. Ces différences ne sont pas essentielles et ne changent rien à la substance des arguments.

On dit qu'une suite $(A_n)_{n\geqslant 0}$ de parties de E est *croissante* si pour tout $n \geqslant 0$, on a l'inclusion $A_n \subseteq A_{n+1}$.

Définition B.2 (λ-système)**.** *On dit qu'une classe de parties $\mathscr{M}$ de E est un* λ-système, *ou une* classe monotone, *si*

1. $E \in \mathscr{M}$,
2. *pour tous* $A, B \in \mathscr{M}$ *tels que* $A \subseteq B$, *on a* $B \setminus A \in \mathscr{M}$,
3. *pour toute suite croissante* $(A_n)_{n\geqslant 0}$ *d'éléments de* $\mathscr{M}$, *on a* $\bigcup_{n\geqslant 0} A_n \in \mathscr{M}$.

Exercice B.1. *Montrer qu'une intersection quelconque de λ-systèmes sur E est encore un λ-système.*

Le résultat de cet exercice permet de définir, comme pour les tribus, le λ-système engendré par une classe de parties $\mathscr{C}$ de E : c'est l'intersection de tous les λ-systèmes qui contiennent $\mathscr{C}$, et c'est aussi *le plus petit λ-système* qui contient $\mathscr{C}$. On le note $\lambda(\mathscr{C})$.

On observe immédiatement qu'une tribu sur E est à la fois un π-système et un λ-système. La réciproque est moins évidente, mais vraie.

B.2 Lien avec les tribus

Proposition B.3. *Une classe de parties sur E est une tribu si et seulement si elle est à la fois un π-système et un λ-système.*

Démonstration. Soit $\mathscr{C}$ une classe de parties de E qui est à la fois un π-système et un λ-système. Montrons que $\mathscr{C}$ est une tribu.

Tout d'abord, $E \in \mathscr{C}$ et $\varnothing = E \setminus E \in \mathscr{C}$.

Ensuite, pour tout $A \in \mathscr{C}$, on a $A^c = E \setminus A \in \mathscr{C}$.

Montrons maintenant que $\mathscr{C}$ est stable par unions. Pour tous $A, B \in \mathscr{C}$, on a $A \cup B = (A^c \cap B^c)^c \in \mathscr{C}$.

Soit maintenant $(A_n)_{n\geqslant 0}$ une suite d'éléments de $\mathscr{C}$. Pour remédier au fait que cette suite n'est pas nécessairement croissante, posons, tout $n \geqslant 0$, $B_n = A_0 \cup \ldots \cup A_n$. Alors la suite $(B_n)_{n\geqslant 0}$ est une suite croissante d'éléments de $\mathscr{C}$. Ainsi, $\bigcup_{n\geqslant 0} B_n$ appartient à $\mathscr{C}$. Mais $\bigcup_{n\geqslant 0} B_n$ n'est autre que $\bigcup_{n\geqslant 0} A_n$. □

Remarquons qu'au cours de cette démonstration, nous avons démontré la chose suivante : un λ-système stable par unions finies est une tribu.

Proposition B.4. *Pour toute classe de parties de E, on a l'inclusion*

$$\lambda(\mathscr{C}) \subseteq \sigma(\mathscr{C}).$$

Démonstration. En effet, la classe de parties $\sigma(\mathscr{C})$ est une tribu qui contient $\mathscr{C}$, donc en particulier un λ-système qui contient $\mathscr{C}$. Elle contient donc le plus petit λ-système qui contient $\mathscr{C}$, c'est-à-dire $\lambda(\mathscr{C})$. □

B.3 Théorème de la classe monotone

Le théorème de la classe de monotone affirme qu'il y a égalité dans certains cas.

Théorème B.5 (Théorème de la classe monotone)**.** *Soit $\mathscr{I}$ un π-système sur E. Alors*

$$\lambda(\mathscr{I}) = \sigma(\mathscr{I}).$$

On utilise en général ce théorème en disant qu'un λ-système qui contient un π-système contient la tribu engendrée par ce π-système. Nous donnerons un exemple après la démonstration.

Démonstration. En vertu de la proposition précédente, il suffit de montrer que $\lambda(\mathscr{I}) \supseteq \sigma(\mathscr{I})$. Pour cela, il suffit de montrer que $\lambda(\mathscr{I})$ est une tribu : elle contiendra alors la plus petite tribu qui contient $\mathscr{I}$. Pour alléger la notation, posons $\mathscr{M} = \lambda(\mathscr{I})$.

On sait déjà que $E \in \mathscr{M}$ et que $\mathscr{M}$ est stable par passage au complémentaire. Si nous savions que $\mathscr{M}$ est stable par union, nous pourrions conclure cette démonstration comme nous avons conclu celle de la proposition B.3. Il ne reste donc qu'à montrer que $\mathscr{M}$ est stable par union, et puisque cette classe est stable par passage au complémentaire, il est équivalent de prouver qu'elle est stable par intersection.

Pour cela, considérons la classe

$$\mathscr{C}_1 = \{A \in \mathscr{M} : \forall B \in \mathscr{I}, A \cap B \in \mathscr{M}\}.$$

D'une part, $\mathscr{C}_1$ contient $\mathscr{I}$, puisque si $A \in \mathscr{I}$, alors pour tout $B \in \mathscr{I}$, on a $A \cap B \in \mathscr{I} \subseteq \mathscr{M}$.

D'autre part, $\mathscr{C}_1$ est un λ-système : c'est une vérification sans difficulté, qui est laissée en exercice.

Ainsi, $\mathscr{C}_1$ est un λ-système qui contient $\mathscr{I}$: ceci entraîne que $\mathscr{C}_1$ contient $\mathscr{M}$. Nous avons donc montré que l'intersection d'un élément de $\mathscr{M}$ et d'un élément de $\mathscr{I}$ appartient à $\mathscr{M}$. Posons maintenant

$$\mathscr{C}_2 = \{B \in \mathscr{M} : \forall A \in \mathscr{M}, A \cap B \in \mathscr{M}\}.$$

Comme précédemment, on vérifie que $\mathscr{C}_2$ contient $\mathscr{I}$, et que c'est une classe monotone : elle contient donc $\mathscr{M}$. Ceci achève de démontrer que $\mathscr{M}$ est stable par intersection. □

B.4 Utilisation du théorème

La manière habituelle dont on utilise ce théorème est la suivante. On souhaite montrer que tous les éléments d'une tribu $\mathscr{E}$ ont une certaine propriété (P).

— On identifie un π-système $\mathscr{I}$ qui engendre $\mathscr{E}$, au sens où $\sigma(\mathscr{I}) = \mathscr{E}$.

— On vérifie que tous les éléments de $\mathscr{I}$ ont la propriété (P) (pour cela, il faut avoir bien choisi $\mathscr{I}$, naturellement).

— On montre que la classe, qu'on appelle $\mathscr{M}$, de tous les éléments de $\mathscr{E}$ qui ont la propriété (P) est un λ-système.

On conclut en disant que la classe $\mathscr{M}$ est un λ-système qui contient $\mathscr{I}$: on a donc

$$\mathscr{E} \supseteq \mathscr{M} \supseteq \lambda(\mathscr{I}) = \sigma(\mathscr{I}) = \mathscr{E}$$

et en particulier, $\mathscr{M} = \mathscr{E}$.

C

Compléments sur la notion d'indépendance

Cette annexe est consacrée à une présentation de la notion d'indépendance et de ses premières propriétés. Il s'agit d'une notion qui fait partie d'un premier cours de probabilités, mais qui est un peu délicate, et sur laquelle nous n'avons pas le temps de revenir en détail pendant ce cours.

Dans tout ce texte, on travaille sur un espace de probabilité $(\Omega, \mathscr{F}, \mathbf{P})$.

C.1 Définition de l'indépendance

Commençons par donner les définitions de l'indépendance de tribus, de variables aléatoires, et d'événements. Nous nous intéressons pour l'instant à des familles finies.

Définition C.1. *Soit $n \geqslant 2$ un entier. Soient $\mathscr{G}_1, \ldots, \mathscr{G}_n$ des sous-tribus de $\mathscr{F}$. On dit que $\mathscr{G}_1, \ldots, \mathscr{G}_n$ sont* indépendantes *relativement à $\mathbf{P}$ si pour tous événements $G_1 \in \mathscr{G}_1, \ldots, G_n \in \mathscr{G}_n$, on a l'égalité $\mathbf{P}(G_1 \cap \ldots \cap G_n) = \mathbf{P}(G_1) \ldots \mathbf{P}(G_n)$.*

Définition C.2. *Soit $n \geqslant 2$ un entier. Soient $X_1, \ldots, X_n$ des variables aléatoires sur $(\Omega, \mathscr{F}, \mathbf{P})$. On dit que $X_1, \ldots, X_n$ sont* indépendantes *relativement à $\mathbf{P}$ si les tribus $\sigma(X_1), \ldots, \sigma(X_n)$ sont indépendantes au sens de la définition C.1.*

Soulignons que cette définition a un sens même si les variables aléatoires $X_1, \ldots, X_n$ sont à valeurs dans des espaces mesurables distincts.

Définition C.3. *Soit $n \geqslant 2$ un entier. Soient $A_1, \ldots, A_n \in \mathscr{F}$ des évélements. On dit que $A_1, \ldots, A_n$ sont* indépendants *relativement à $\mathbf{P}$ si les variables aléatoires $\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_n}$ sont indépendantes au sens de la définition C.2.*

On généralise la notion d'indépendance à des familles infinies de sous-tribus, de variables aléatoires, ou d'événements, de la manière la plus simple possible.

Définition C.4. *On dit qu'une famille quelconque de sous-tribus (resp. de variables aléatoires, resp. d'événements) est indépendante si chacune de ses sous-familles finies est indépendante au sens de la définition C.1 (resp. de la définition C.2, resp. de la définition C.3).*

Dans la suite de ce texte, on s'intéresse à des familles finies.

Exercice C.1. *La définition C.4 appliquée à une famille finie de sous-tribus donne-t-elle la même définition de son indépendance que la définition C.1 ?*

C.2 Événements indépendants

Nous allons donner plusieurs caractérisations de l'indépendance de n événements. Examinons d'abord ce que dit la définition. Pour la comprendre, il nous faut comprendre ce qu'est la tribu engendrée par l'indicatrice d'un événement.

Exercice C.2. *Soit $A \in \mathscr{F}$ un événement. Si $A = \varnothing$ ou $A = \Omega$, alors $\sigma(\mathbf{1}_A) = \{\varnothing, \Omega\}$. Sinon, $\sigma(\mathbf{1}_A) = \{\varnothing, A, A^c, \Omega\}$.*

Par définition, des événements $A_1, \dots, A_n$ de la tribu $\mathscr{F}$ sont donc indépendants si pour tous $G_1 \in \{\varnothing, A_1, A_1^c, \Omega\}, \dots, G_n \in \{\varnothing, A_n, A_n^c, \Omega\}$, on a $\mathbf{P}(G_1 \cap \dots \cap G_n) = \mathbf{P}(G_1) \dots \mathbf{P}(G_n)$. Cela fait 4^n égalités à vérifier. Un certain nombre d'entre elles sont automatiquement vérifiées, par exemple celles où l'un des G_i est vide, celle où tous les G_i sont égaux à Ω, et quelques autres encore.

Proposition C.5. *Soient $A_1, \dots, A_n \in \mathscr{F}$ des évenements. Alors les deux assertions suivantes sont équivalentes.*
1. $A_1, \dots, A_n$ sont indépendants (au sens de la définition C.3).
2. Pour tous $B_1 \in \{A_1, A_1^c\}, \dots, B_n \in \{A_n, A_n^c\}$, on a $\mathbf{P}(B_1 \cap \dots \cap B_n) = \mathbf{P}(B_1) \dots \mathbf{P}(B_n)$.

Si cette proposition est vraie (ce que nous allons démontrer), elle ramène la vérification de l'indépendance de n événements à la vérification de 2^n égalités.

Démonstration de la proposition C.5. Supposons la première assertion vérifiée et donnons-nous $B_1 \in \{A_1, A_1^c\}, \dots, B_n \in \{A_n, A_n^c\}$. Alors pour tout $i \in \{1, \dots, n\}$, l'événement B_i appartient à la tribu $\sigma(\mathbf{1}_{A_i})$, si bien que l'égalité $\mathbf{P}(B_1 \cap \dots \cap B_n) = \mathbf{P}(B_1) \dots \mathbf{P}(B_n)$ a lieu par définition de l'indépendance de $A_1, \dots, A_n$. Ainsi, la seconde assertion est vérifiée.

Supposons réciproquement que la seconde assertion soit vérifiée et montrons que la première l'est. Pour cela, donnons-nous des événements $G_1 \in \{\varnothing, A_1, A_1^c, \Omega\}, \dots, G_n \in \{\varnothing, A_n, A_n^c, \Omega\}$ et montrons qu'on a l'égalité $\mathbf{P}(G_1 \cap \dots \cap G_n) = \mathbf{P}(G_1) \dots \mathbf{P}(G_n)$.

Si l'un au moins des événements $G_1, \dots, G_n$ est vide, alors l'égalité est vraie. Supposons désormais qu'aucun d'entre eux n'est vide. Si aucun d'entre eux n'est égal à Ω, alors pour tout $i \in \{1, \dots, n\}$, on a $G_i \in \{A_i, A_i^c\}$, et l'égalité découle directement de notre hypothèse que la seconde assertion est vérifiée.

Il reste à traiter le cas où aucun des événements $G_1, \dots, G_n$ n'est vide, et où au moins l'un d'entre eux est égal à Ω. Quitte à changer la numérotation des événements $A_1, \dots, A_n$, nous pouvons supposer que, pour un certain $k \in \{0, \dots, n-1\}$, nous avons

$$G_1 \in \{A_1, A_1^c\}, \dots, G_k \in \{A_k, A_k^c\}, \text{ et } G_{k+1} = \dots = G_n = \Omega.$$

Nous allons calculer $\mathbf{P}(G_1 \cap \dots \cap G_n)$ en écrivant, pour chaque $i \in \{k+1, \dots, n\}$, l'événement Ω sous la forme $A_i \cup A_i^c$, puis utiliser la distributivité de l'union par rapport à l'intersection et enfin l'additivité de la mesure. Cela nous donne

$$\begin{aligned}
\mathbf{P}(G_1 \cap \dots \cap G_n) &= \mathbf{P}\big(G_1 \cap \dots \cap G_k \cap (A_{k+1} \cup A_{k+1}^c) \cap \dots \cap (A_n \cup A_n^c)\big) \\
&= \mathbf{P}\Bigg(\bigcup_{H_{k+1} \in \{A_{k+1}, A_{k+1}^c\}, \dots, H_n \in \{A_n, A_n^c\}} G_1 \cap \dots \cap G_k \cap H_{k+1} \cap \dots \cap H_n \Bigg) \\
&= \sum_{H_{k+1} \in \{A_{k+1}, A_{k+1}^c\}, \dots, H_n \in \{A_n, A_n^c\}} \mathbf{P}(G_1 \cap \dots \cap G_k \cap H_{k+1} \cap \dots \cap H_n).
\end{aligned}$$

L'hypothèse que la seconde assertion est vérifiée nous permet de calculer chacune des probabilités qui apparaît dans la dernière somme. Nous trouvons

$$\mathbf{P}(G_1 \cap \ldots \cap G_n) = \sum_{H_{k+1} \in \{A_{k+1}, A_{k+1}^c\}, \ldots, H_n \in \{A_n, A_n^c\}} \mathbf{P}(G_1) \ldots \mathbf{P}(G_k) \mathbf{P}(H_{k+1}) \ldots \mathbf{P}(H_n).$$

Les k premiers facteurs de chaque terme de la somme se factorisent, et il reste une somme qui se calcule facilement, à savoir

$$\begin{aligned}\mathbf{P}(G_1 \cap \ldots \cap G_n) &= \mathbf{P}(G_1) \ldots \mathbf{P}(G_k) \sum_{H_{k+1} \in \{A_{k+1}, A_{k+1}^c\}, \ldots, H_n \in \{A_n, A_n^c\}} \mathbf{P}(H_{k+1}) \ldots \mathbf{P}(H_n) \\ &= \mathbf{P}(G_1) \ldots \mathbf{P}(G_k) \big(\mathbf{P}(A_{k+1}) + \mathbf{P}(A_{k+1}^c)\big) \ldots \big(\mathbf{P}(A_n) + \mathbf{P}(A_n^c)\big) \\ &= \mathbf{P}(G_1) \ldots \mathbf{P}(G_k).\end{aligned}$$

Enfin, puisque $G_{k+1} = \ldots = G_n = \Omega$, cette dernière quantité n'est autre que $\mathbf{P}(G_1) \ldots \mathbf{P}(G_n)$, et l'égalité souhaitée est vérifiée. □

On peut donner un autre ensemble de 2^n égalités (en fait un peu moins) qui est équivalent à l'indépendance de n événements : c'est ce que fait la proposition suivante.

Proposition C.6. *Soient $A_1, \ldots, A_n \in \mathscr{F}$ des évenements. Alors les deux assertions suivantes sont équivalentes.*
1. $A_1, \ldots, A_n$ sont indépendants.
2. Pour tout entier $k \geqslant 2$ et tous entiers $i_1, \ldots, i_k$ tels que $1 \leqslant i_1 < \ldots < i_k \leqslant n$, on a $\mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbf{P}(A_{i_1}) \ldots \mathbf{P}(A_{i_k})$.

Démonstration de la proposition C.6. Supposons la première assertion vérifiée. Donnons-nous $k \geqslant 2$ et $i_1 < \ldots < i_k$ entre 1 et n. Pour tout $j \in \{1, \ldots, n\}$, posons $G_j = A_{i_\ell}$ si $j = i_\ell$, et $G_j = \Omega$ si j n'appartient pas à $\{i_1, \ldots, i_k\}$. Alors l'égalité $\mathbf{P}(G_1 \cap \ldots \cap G_n) = \mathbf{P}(G_1) \ldots \mathbf{P}(G_n)$, qui a lieu par hypothèse, est exactement l'égalité $\mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbf{P}(A_{i_1}) \ldots \mathbf{P}(A_{i_k})$. Ainsi, la deuxième assertion est vérifiée.

Supposons maintenant la deuxième assertion vérifiée. Nous allons montrer que la deuxième assertion de la proposition C.5 est vérifiée, si bien que $A_1, \ldots, A_n$ sont indépendants.

Considérons donc $B_1 \in \{A_1, A_1^c\}, \ldots, B_n \in \{A_n, A_n^c\}$. Comme lors de la démonstration de la proposition précédente, nous ne perdons pas de généralité, quitte à changer la numérotation des événements, à supposer qu'il existe $k \in \{0, \ldots, n\}$ tel que $B_i = A_i$ pour $i \leqslant k$ et $B_i = A_i^c$ pour $i > k$.

Pour calculer la probabilité de l'intersection de $B_1, \ldots, B_n$, nous allons l'écrire sous la forme d'une espérance :

$$\mathbf{P}(B_1 \cap \ldots \cap B_n) = \mathbf{E}[\mathbf{1}_{B_1 \cap \ldots \cap B_n}] = \mathbf{E}[\mathbf{1}_{B_1} \ldots \mathbf{1}_{B_n}] = \mathbf{E}[\mathbf{1}_{A_1} \ldots \mathbf{1}_{A_k} \mathbf{1}_{A_{k+1}^c} \ldots \mathbf{1}_{A_n^c}].$$

Exprimons maintenant les indicatrices des complémentaires en fonction des indicatrices des ensembles :

$$\mathbf{P}(B_1 \cap \ldots \cap B_n) = \mathbf{E}[\mathbf{1}_{A_1} \ldots \mathbf{1}_{A_k} (1 - \mathbf{1}_{A_{k+1}}) \ldots (1 - \mathbf{1}_{A_n})]. \tag{C.1}$$

Développons maintenant le produit dans l'espérance. Nous trouvons

$$\begin{aligned}\mathbf{P}(B_1 \cap \ldots \cap B_n) &= \sum_{J \subseteq \{k+1, \ldots, n\}} (-1)^{|J|} \mathbf{E}\left[\mathbf{1}_{A_1} \ldots \mathbf{1}_{A_k} \prod_{i \in J} \mathbf{1}_{A_j}\right] \\ &= \sum_{J \subseteq \{k+1, \ldots, n\}} (-1)^{|J|} \mathbf{P}\left(A_1 \cap \ldots \cap A_k \cap \bigcap_{j \in J} A_j\right)\end{aligned}$$

L'hypothèse nous permet de calculer chacun des termes de cette somme, pour trouver

$$\mathbf{P}(B_1 \cap \ldots \cap B_n) = \sum_{J \subseteq \{k+1,\ldots,n\}} (-1)^{|J|}\, \mathbf{P}(A_1) \ldots \mathbf{P}(A_k) \prod_{i \in J} \mathbf{P}(A_j).$$

En reprenant, mais à l'envers, le calcul où nous avons développé le terme de droite de (C.1), nous trouvons

$$\mathbf{P}(B_1 \cap \ldots \cap B_n) = \mathbf{P}(A_1) \ldots \mathbf{P}(A_k)\big(1 - \mathbf{P}(A_{k+1})\big) \ldots \big(1 - \mathbf{P}(A_n)\big)$$

et nous aboutissons finalement à l'égalité

$$\mathbf{P}(B_1 \cap \ldots \cap B_n) = \mathbf{P}(A_1) \ldots \mathbf{P}(A_k)\mathbf{P}(A_{k+1}^c) \ldots \mathbf{P}(A_n^c) = \mathbf{P}(B_1) \ldots \mathbf{P}(B_n).$$

La deuxième assertion de la proposition C.5 est donc bien vérifiée, et les événements $A_1, \ldots, A_n$ sont bien indépendants. □

Par exemple, trois événements A, B, C sont indépendants si et seulement si les quatre égalités

$$\begin{aligned}
\mathbf{P}(A \cap B \cap C) &= \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C) \\
\mathbf{P}(A \cap B) &= \mathbf{P}(A)\mathbf{P}(B) \\
\mathbf{P}(B \cap C) &= \mathbf{P}(B)\mathbf{P}(C) \\
\mathbf{P}(C \cap A) &= \mathbf{P}(C)\mathbf{P}(A)
\end{aligned}$$

sont vérifiées.

Exercice C.3. *Montrer qu'il n'est pas possible d'enlever une des quatre conditions ci-dessus. Autrement dit, pour chacune de ces quatre conditions, il existe trois événements qui ne sont pas indépendants mais qui vérifient les trois autres conditions.*

En particulier, la condition $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$ ne suffit pas à assurer que les événements A, B, C sont indépendants.

Les assertions "A, B, C sont indépendants" et "A, B, C sont deux à deux indépendants" ne sont pas non plus équivalentes.

C.3 Variables aléatoires indépendantes

Convenons de noter $P_X = \mathbf{P} \circ X^{-1}$ la loi d'une variable aléatoire X.

Proposition C.7. *Soient $X_1, \ldots, X_n$ des variables aléatoires, à valeurs respectivements dans des espaces mesurables $(E_1, \mathscr{E}_1), \ldots, (E_n, \mathscr{E}_n)$. Les deux assertions suivantes sont équivalentes.*
1. Les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes.
2. Sur l'espace $(E_1 \times \ldots \times E_n, \mathscr{E}_1 \otimes \ldots \otimes \mathscr{E}_n)$, on a l'égalité $P_{(X_1,\ldots,X_n)} = P_{X_1} \otimes \ldots \otimes P_{X_n}$.

Démonstration. Supposons les variables aléatoires indépendantes. Pour montrer l'égalité des mesures de probabilité $P_{(X_1,\ldots,X_n)}$ et $P_{X_1} \otimes \ldots \otimes P_{X_n}$, il suffit de montrer leur égalité sur un π-système qui engendre la tribu produit $\mathscr{E}_1 \otimes \ldots \otimes \mathscr{E}_n$. Le choix naturel est celui du π-système constitué des pavés mesurables. Considérons donc $F_1 \in \mathscr{E}_1, \ldots, F_n \in \mathscr{E}_n$. On a

$$P_{(X_1,\ldots,X_n)}(F_1 \times \ldots \times F_n) = \mathbf{P}(X_1 \in F_1, \ldots, X_n \in F_n).$$

Les événements $\{X_1 \in F_1\}, \ldots, \{X_n \in F_n\}$ appartiennent respectivement à $\sigma(X_1), \ldots, \sigma(X_n)$, qui sont indépendantes par hypothèse. Ainsi,

$$\begin{aligned} P_{(X_1,\ldots,X_n)}(F_1 \times \ldots \times F_n) &= \mathbf{P}(X_1 \in F_1) \ldots \mathbf{P}(X_n \in F_n) \\ &= P_{X_1}(F_1) \ldots P_{X_n}(F_n) \\ &= (P_{X_1} \otimes \ldots \otimes P_{X_n})(F_1 \times \ldots \times F_n), \end{aligned}$$

ce qui montre l'égalité souhaitée.

Réciproquement, supposons que la deuxième assertion ait lieu et montrons que les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes. Par définition, nous devons montrer que les tribus $\sigma(X_1), \ldots, \sigma(X_n)$ sont indépendantes. Pour ce faire, et encore par définition, il nous faut choisir n événements $A_1 \in \sigma(X_1), \ldots A_n \in \sigma(X_n)$ et montrer que la probabilité de leur intersection est le produit de leurs probabilités.

Or la tribu $\sigma(X_1)$, par exemple, est l'ensemble des images réciproques par X_1 des événements de la tribu $\mathscr{E}_1$. Ainsi, il existe $F_1 \in \mathscr{E}_1$ tel que $A_1 = X_1^{-1}(A_1)$. De même, il existe, pour tout $i \in \{1, \ldots, n\}$, un événement $F_i \in \mathscr{E}_i$ tel que $A_i = X_i^{-1}(F_i)$. Calculons maintenant la probabilité de $A_1 \cap \ldots \cap A_n$. Nous trouvons

$$\begin{aligned} \mathbf{P}(A_1 \cap \ldots \cap A_n) &= \mathbf{P}(X_1^{-1}(F_1) \cap \ldots \cap X_n^{-1}(F_n)) \\ &= \mathbf{P}(X_1 \in F_1, \ldots, X_n \in F_n) \\ &= P_{(X_1,\ldots,X_n)}(F_1 \times \ldots \times F_n) \end{aligned}$$

et par hypothèse, ceci est égal à

$$P_{X_1}(F_1) \ldots P_{X_n}(F_n) = \mathbf{P}(X_1^{-1}(F_1)) \ldots \mathbf{P}(X_n^{-1}(F_n)) = \mathbf{P}(A_1) \ldots \mathbf{P}(A_n),$$

ce qui conclut la démonstration. □

Exercice C.4. *Soit $n \geqslant 2$ un entier. Soient $X_1, \ldots, X_n$ des variables aléatoires, à valeurs dans des ensembles finis que l'on note respectivement $E_1, \ldots, E_n$. Montrer que $X_1, \ldots, X_n$ sont indépendantes si et seulement si pour tous $x_1 \in E_1, \ldots, x_n \in E_n$, on a $\mathbf{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbf{P}(X_1 = x_1) \ldots \mathbf{P}(X_n = x_n)$.*

Exercice C.5. *Soit $n \geqslant 3$. Soient $X_1, \ldots, X_{n-1}$ des variables aléatoires indépendantes telles que pour tout $i \in \{1, \ldots, n-1\}$, on ait $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = -1) = \frac{1}{2}$. On pose $X_n = X_1 \ldots X_{n-1}$, le produit de $X_1, \ldots, X_n$. Vérifier que X_n a la même distribution que chacune des variables aléatoires $X_1, \ldots, X_{n-1}$. Montrer que les n variables aléatoires $X_1, \ldots, X_n$ ne sont pas indépendantes, mais que $n-1$ quelconques d'entre elles (et plus généralement k quelconques d'entre elles, pour tout $k \in \{2, \ldots, n-1\}$) sont indépendantes.*

C.4 Tribus indépendantes

La proposition suivante est un outil très utile pour démontrer que des tribus sont indépendantes.

Proposition C.8. *Soient $n \geqslant 2$ un entier et $\mathscr{G}_1, \ldots, \mathscr{G}_n$ des sous-tribus de $\mathscr{F}$. Soient $\mathscr{I}_1, \ldots, \mathscr{I}_n$ des π-systèmes sur Ω tels que pour tout $j \in \{1, \ldots, n\}$, on ait $\sigma(\mathscr{I}_j) = \mathscr{G}_j$. Les deux assertions sont équivalentes.*
1. $\mathscr{G}_1, \ldots, \mathscr{G}_n$ sont indépendantes.
2. pour tous $A_1 \in \mathscr{I}_1, \ldots, A_n \in \mathscr{I}_n$, on a $\mathbf{P}(A_1 \cap \ldots \cap A_n) = \mathbf{P}(A_1) \ldots \mathbf{P}(A_n)$.

Démonstration. La première assertion implique tautologiquement la seconde : en effet, pour tout $j \in \{1, \dots, n\}$, le π-système est inclus dans la tribu $\mathscr{G}_j$, et la seconde assertion est un cas particulier de la définition de l'indépendance de $\mathscr{G}_1, \dots, \mathscr{G}_n$.

C'est l'assertion réciproque qui est intéressante. Elle repose sur l'application du lemme de classe monotone, sous la forme de l'assertion suivante, que nous allons d'abord énoncer et démontrer.

À tous éléments $C_1, \dots, C_{n-1}$ de $\mathscr{F}$, associons la classe

$$\mathscr{M}(C_1, \dots, C_{n-1}) = \Big\{ G \in \mathscr{F} : \mathbf{P}(C_1 \cap \dots \cap C_{n-1} \cap G) = \mathbf{P}(C_1) \dots \mathbf{P}(C_{n-1}) \mathbf{P}(G) \Big\}.$$

Nous affirmons que si $\mathbf{P}(C_1 \cap \dots \cap C_{n-1}) = \mathbf{P}(C_1) \dots \mathbf{P}(C_{n-1})$, alors cette classe est une classe monotone.

Montrons-le. Il y a trois choses à vérifier. Tout d'abord, l'hypothèse que $\mathbf{P}(C_1 \cap \dots \cap C_{n-1}) = \mathbf{P}(C_1) \dots \mathbf{P}(C_{n-1})$ entraîne que Ω appartient à $\mathscr{M}(C_1, \dots, C_{n-1})$.

Supposons maintenant que deux événements G et H tels que $G \subseteq H$ appartiennent à $\mathscr{M}(C_1, \dots, C_{n-1})$. Alors

$$C_1 \cap \dots \cap C_{n-1} \cap (H \setminus G) = (C_1 \cap \dots \cap C_{n-1} \cap H) \setminus (C_1 \cap \dots \cap C_{n-1} \cap G),$$

donc

$$\begin{aligned} \mathbf{P}(C_1 \cap \dots \cap C_{n-1} \cap (H \setminus G)) &= \mathbf{P}(C_1 \cap \dots \cap C_{n-1} \cap H) - \mathbf{P}(C_1 \cap \dots \cap C_{n-1} \cap G) \\ &= \mathbf{P}(C_1) \dots \mathbf{P}(C_{n-1}) \big(\mathbf{P}(H) - \mathbf{P}(G) \big) \\ &= \mathbf{P}(C_1) \dots \mathbf{P}(C_{n-1}) \mathbf{P}(H \setminus G), \end{aligned}$$

ce qui montre que $H \setminus G$ appartient à $\mathscr{M}(C_1, \dots, C_{n-1})$.

Donnons-nous enfin une suite croissante $(G_k)_{k \geqslant 0}$ d'éléments de $\mathscr{M}(C_1, \dots, C_{n-1})$. Notons $G = \bigcup_{k \geqslant 0} G_k$. Alors la suite de terme général $C_1 \cap \dots \cap C_{n-1} \cap G_k$ est croissante, et l'union des termes de cette suite est $C_1 \cap \dots \cap C_{n-1} \cap G$. Alors les hypothèses et la propriété de convergence monotone pour les mesures nous permettent d'affirmer que

$$\begin{aligned} \mathbf{P}(C_1 \cap \dots \cap C_{n-1} \cap G) &= \lim_{k \to \infty} \mathbf{P}(C_1 \cap \dots \cap C_{n-1} \cap G_k) \\ &= \lim_{k \to \infty} \mathbf{P}(C_1) \dots \mathbf{P}(C_{n-1}) \mathbf{P}(G_k) \\ &= \mathbf{P}(C_1) \dots \mathbf{P}(C_{n-1}) \mathbf{P}(G), \end{aligned}$$

si bien que G appartient à $\mathscr{M}(C_1, \dots, C_{n-1})$.

Ceci conclut la démonstration du fait que $\mathscr{M}(C_1, \dots, C_{n-1})$ est une classe monotone et nous pouvons à présent utiliser ce fait pour démontrer la proposition.

Cette démonstration se fait par étapes. La première étape consiste à montrer que pour tous événements $G_1 \in \mathscr{G}_1, A_2 \in \mathscr{I}_2, \dots, A_n \in \mathscr{I}_n$, on a

$$\mathbf{P}(G_1 \cap A_2 \cap \dots \cap A_n) = \mathbf{P}(G_1) \mathbf{P}(A_2) \dots \mathbf{P}(A_n).$$

Pour cela, fixons $A_2 \in \mathscr{I}_2, \dots, A_n \in \mathscr{I}_n$. Montrer que l'égalité a lieu pour tout $G_1 \in \mathscr{G}_1$ revient à montrer que la classe $\mathscr{M}(A_2, \dots, A_n)$ contient $\mathscr{G}_1$. Or notre hypothèse que la deuxième assertion est vérifiée entraîne que $\mathbf{P}(A_2 \cap \dots \cap A_n) = \mathbf{P}(A_2) \dots \mathbf{P}(A_n)$, si bien que $\mathscr{M}(A_2, \dots, A_n)$ est une classe monotone. Or par hypothèse, elle contient le π-système $\mathscr{I}_1$. Donc, en vertu du lemme de classe monotone, elle contient $\sigma(\mathscr{I}_1) = \mathscr{G}_1$.

La première étape est achevée, nous pouvons maintenant passer à la deuxième, qui consiste à montrer que pour tous $G_1 \in \mathscr{G}_1, G_2 \in \mathscr{G}_2, A_3 \in \mathscr{I}_3, \dots, A_n \in \mathscr{I}_n$, on a

$$\mathbf{P}(G_1 \cap G_2 \cap A_3 \cap \dots \cap A_n) = \mathbf{P}(G_1)\mathbf{P}(G_2)\mathbf{P}(A_3)\dots\mathbf{P}(A_n).$$

On procède comme à la première étape : la classe $\mathscr{M}(G_1, A_3, \dots, A_n)$ est une classe monotone, qui contient le π-système $\mathscr{I}_2$, donc elle contient $\sigma(\mathscr{I}_2) = \mathscr{G}_2$, ce qu'on voulait démontrer.

On continue ensuite jusqu'à la n-ième étape, où on aboutit à la conclusion que pour tous $G_1 \in \mathscr{G}_1, \dots, G_n \in \mathscr{G}_n$, on a

$$\mathbf{P}(G_1 \cap \dots \cap G_n) = \mathbf{P}(G_1)\dots\mathbf{P}(G_n),$$

ce qui est la définition de l'indépendance des tribus $\mathscr{G}_1, \dots, \mathscr{G}_n$. □

La succession des étapes de la démonstration pourrait être rédigée comme une récurrence, mais cela la rendrait probablement encore plus difficile à lire, sans en éclairer mieux, à mon avis, le fonctionnement.

Une application de ce résultat est la suivante.

Proposition C.9. *Soient $\mathscr{G}_1, \dots, \mathscr{G}_n$ des sous-tribus de $\mathscr{F}$ indépendantes. Soient $\ell \geqslant 2$ un entier et $0 = m_0 < m_1 < \dots < m_\ell = n$ des entiers. Pour tout $k \in \{1, \dots, \ell\}$, posons $\mathscr{H}_k = \sigma(\mathscr{G}_{m_{k-1}+1}, \dots, \mathscr{G}_{m_k})$. Alors les tribus $\mathscr{H}_1, \dots, \mathscr{H}_\ell$ sont indépendantes.*

Par exemple, si $\mathscr{G}_1, \dots, \mathscr{G}_6$ sont indépendantes, la proposition affirme que les sous-tribus $\sigma(\mathscr{G}_1, \mathscr{G}_2), \mathscr{G}_3, \sigma(\mathscr{G}_4, \mathscr{G}_5, \mathscr{G}_6)$ sont indépendantes.

Exercice C.6. *Déterminer avec quelles valeurs de n, ℓ, $m_0, \dots, m_\ell$ on a appliqué la proposition pour obtenir l'assertion donnée ci-dessus en exemple.*

Exercice C.7. *Montrer que si $\mathscr{G}_1, \dots, \mathscr{G}_m$ sont des sous-tribus de $\mathscr{F}$, alors la classe*

$$\{G_1 \cap \dots \cap G_m : G_1 \in \mathscr{G}_1, \dots, G_m \in \mathscr{G}_m\}$$

est un π-système qui engendre $\sigma(\mathscr{G}_1, \dots, \mathscr{G}_m)$.

Exercice C.8. *Démontrer la proposition C.9 en utilisant l'exercice ci-dessus et la proposition C.8.*

C.5 La loi du $0-1$ de Kolmogorov

Pour conclure, énonçons et démontrons la loi du $0-1$ de Kolmogorov.

Théorème C.10. *Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilité. Soit $(\mathscr{G}_n)_{n\geqslant 0}$ une suite de sous-tribus indépendantes. Pour tout $n \geqslant 0$, on pose $\mathscr{H}_n = \sigma(\mathscr{G}_p : p \geqslant n)$. On pose $\mathscr{H}_\infty = \bigcap_{n\geqslant 0} \mathscr{H}_n$.*
Pour tout événement $A \in \mathscr{H}_\infty$, on a $\mathbf{P}(A) \in \{0, 1\}$.

Démonstration. La démonstration consiste à démontrer que la tribu $\mathscr{H}_\infty$ est indépendante d'elle-même. Ceci implique en effet que pour tout $A \in \mathscr{H}_\infty$, on a $\mathbf{P}(A) = \mathbf{P}(A \cap A) = \mathbf{P}(A)^2$, ce qui entraîne que $\mathbf{P}(A)$ ne peut valoir que 0 ou 1.

Pour montrer que $\mathscr{H}_\infty$ est indépendante d'elle-même, introduisons, pour tout $n \geqslant 0$, la tribu $\mathscr{F}_n = \sigma(\mathscr{G}_p : 0 \leqslant p \leqslant n)$. Notons $\mathscr{F}_\infty = \sigma(\mathscr{G}_p : p \geqslant 0)$, qui est aussi égale à $\mathscr{H}_0$.

Tout d'abord, nous avons l'inclusion

$$\mathscr{H}_\infty \subseteq \mathscr{H}_0 = \mathscr{F}_\infty.$$

Pour montrer que $\mathscr{H}_\infty$ est indépendante d'elle-même, il suffit donc de montrer que $\mathscr{H}_\infty$ est indépendante de $\mathscr{F}_\infty$.

Observons que la classe de parties $\mathscr{I} = \bigcup_{n\geqslant 0} \mathscr{F}_n$, qui engendre la tribu $\mathscr{F}_\infty$, est un π-système. En effet, si B et C appartiennent à $\mathscr{I}$, alors il existe n et m entiers tels que $B \in \mathscr{F}_n$ et $C \in \mathscr{F}_m$. Or la suite $(\mathscr{F}_n)_{n\geqslant 0}$ de tribus est croissante pour l'inclusion, si bien qu'en posant $p = \max(n, m)$, on a $B, C \in \mathscr{F}_p$. En particulier, $B \cap C \in \mathscr{F}_p \subseteq \mathscr{I}$.

En vertu de la proposition C.8, il nous suffit donc de montrer que pour tout $H \in \mathscr{H}_\infty$ et tout $F \in \mathscr{I}$, on a $\mathbf{P}(H \cap F) = \mathbf{P}(H) \cap \mathbf{P}(F)$. Soient donc $H \in \mathscr{H}_\infty$ et $F \in \mathscr{I}$. Il existe un entier $n \geqslant 0$ tel que $F \in \mathscr{F}_n$. Or $\mathscr{H}_\infty \subseteq \mathscr{H}_{n+1}$ et, grâce à la proposition C.9, nous savons que les tribus $\mathscr{F}_n$ et $\mathscr{H}_{n+1}$ sont indépendantes. On a donc bien $\mathbf{P}(H \cap F) = \mathbf{P}(H) \cap \mathbf{P}(F)$.

On en déduit que la tribu $\mathscr{H}_\infty$ est indépendante d'elle-même, comme on l'a annoncé. □

Exercice C.9. *Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires indépendantes. On suppose que la convergence presque sûre*

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow[n\to\infty]{\text{p.s.}} Y$$

a lieu, vers une certaine variable aléatoire Y. Montrer que Y est constante presque sûrement.

Exercice C.10. *Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires indépendantes. On suppose que pour un certain réel ℓ, on a*

$$\mathbf{P}\left(\frac{X_1 + \ldots + X_n}{n} \xrightarrow[n\to\infty]{} \ell\right) > 0.$$

Montrer que la convergence a lieu avec probabilité 1.

Exercice C.11. *Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires indépendantes. On suppose que pour une certaine variable aléatoire Y, on a*

$$\mathbf{P}\left(\frac{X_1 + \ldots + X_n}{n} \xrightarrow[n\to\infty]{} Y\right) > 0.$$

Montrer qu'il existe une constante ℓ telle que

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow[n\to\infty]{\text{p.s.}} \ell.$$

D
Tribus et partitions

Dans cette annexe, je vous propose de réfléchir à la question de savoir si toute tribu est engendrée par une partition, et à quelques questions liées. Ce texte ne contient pas de démonstrations : libre à vous de le compléter.

D.1 Tribu engendrée par une partition

Soit E un ensemble. On appelle *partition* de E une classe de parties de E qui sont non vides, deux à deux disjointes, et dont l'union est E tout entier.

Soit $\mathscr{C} = \{C_i : i \in I\}$ une partition de E. On considère la classe de parties

$$\mathscr{T} = \left\{ \bigcup_{j\in J} C_j : J \subseteq I, J \text{ dénombrable ou } I \setminus J \text{ dénombrable} \right\}.$$

Les trois assertions suivantes :

1. On a l'inclusion $\mathscr{C} \subseteq \mathscr{T}$.
2. Pour toute tribu $\mathscr{E}$ sur E telle que $\mathscr{C} \subseteq \mathscr{E}$, on a l'inclusion $\mathscr{T} \subseteq \mathscr{E}$.
3. La classe $\mathscr{T}$ est une tribu sur E.

sont vraies et entraînent que $\mathscr{T}$ est la tribu engendrée par $\mathscr{C}$.

D.2 Une question

Nous voulons réfléchir à la question suivante.

Soit $\mathscr{E}$ une tribu sur E. Existe-t-il une partition $\mathscr{C}$ de E telle que $\mathscr{E} = \sigma(\mathscr{C})$? Sinon, quelles hypothèses sur $\mathscr{E}$ suffiraient à garantir l'existence d'une telle partition $\mathscr{C}$?

Pour répondre à cette question, il faut essayer, partant d'une tribu $\mathscr{E}$ sur la quelle nous ne faisons a priori aucune hypothèse, de construire une partition $\mathscr{C}$ de E, avec l'espoir que la tribu $\mathscr{E}$ soit, au moins dans certains cas, engendrée par cette partition $\mathscr{C}$.

Nous allons faire une première tentative, basée sur la notion d'élément minimal, ou atome, d'une tribu.

D.3 Atomes

Soit $\mathscr{E}$ une tribu sur E. On appelle *élément minimal*, ou *atome*, de $\mathscr{E}$ un élément de $\mathscr{E}$ qui n'est pas vide et ne contient (au sens de l'inclusion) aucun élément de $\mathscr{E}$ autre que l'ensemble vide et lui-même. En symboles, $A \in \mathscr{E}$ est un atome si A est non vide et si

$$\forall B \in \mathscr{E},\ B \subseteq A \Rightarrow (B = \varnothing \text{ ou } B = A).$$

Pour toute tribu $\mathscr{E}$ sur E, notons $\mathscr{A}(\mathscr{E})$ la classe des atomes de $\mathscr{E}$:

$$\mathscr{A}(\mathscr{E}) = \{A \in \mathscr{E} : A \text{ est un atome de } \mathscr{E}\}.$$

Proposition D.1. *Les éléments de $\mathscr{A}(\mathscr{E})$ sont deux à deux disjoints.*

C'est un premier fait encourageant. Un autre est le suivant.

Proposition D.2. *Soit $\mathscr{C}$ une partition de E. Les élément minimaux de la tribu $\sigma(\mathscr{C})$ sont exactement les éléments de $\mathscr{C}$. En symboles,*

$$\mathscr{A}(\sigma(\mathscr{C})) = \mathscr{C}.$$

Nous pouvons apporter une réponse partielle à notre question.

Proposition D.3. *Si une tribu $\mathscr{E}$ sur E est engendrée par une partition, alors cette partition est celles des atomes de $\mathscr{E}$. Dans ce cas,*

$$\mathscr{E} = \sigma(\mathscr{A}(\mathscr{E})).$$

En lisant un peu vite la dernière proposition, on pourrait se dire : "C'est gagné, toute tribu sur E est engendrée par la partition de E par ses atomes".

Si c'est ce que vous pensez, prenez un moment, avant de lire la suite, pour essayer de voir pourquoi cette conclusion était un peu trop rapide, et ce qui pourrait poser problème.

Il y a au moins deux problèmes que nous n'avons pas résolus.

1. Tout d'abord, nous avons vu que les atomes d'une tribu sont deux à deux disjoints, mais nous n'avons pas démontré que leur réunion est E tout entier. Tant que nous ne l'avons pas démontré, nous ne pouvons pas exclure que la réunion des atomes soit une partie de E strictement incluse dans E, auquel cas les atomes ne forment pas une partition de E, mais d'un sous-ensemble strict de E.

2. Ensuite, même en supposant que la classe $\mathscr{A}(\mathscr{E})$ des atomes de $\mathscr{E}$ est bien une partition de E, nous n'avons pas démontré que la tribu $\mathscr{E}$ est engendrée par cette partition. Nous avons certainement l'inclusion $\mathscr{E} \supseteq \sigma(\mathscr{A}(\mathscr{E}))$, mais nous n'avons pas montré l'égalité.

Pour chacun de ces deux problèmes, nous pouvons soit essayer de démontrer qu'il ne se produit jamais, soit, pour en avoir le cœur net, chercher un exemple où il se produit.

Il se trouve que ces deux problèmes *peuvent* se produire.

1. Il est possible de trouver un ensemble E et une tribu $\mathscr{E}$ sur E telle que la classe $\mathscr{A}(\mathscr{E})$ est une partition d'un sous-ensemble strict de E. En fait, il est même possible de trouver des exemples où la classe $\mathscr{A}(\mathscr{E})$ est vide ! Il existe des tribus qui n'ont *aucun* atome.

2. Il est possible de trouver un ensemble E et une tribu $\mathscr{E}$ sur E telle que la classe $\mathscr{A}(\mathscr{E})$ soit bien une partition de E, mais où pourtant l'inclusion $\mathscr{E} \supsetneq \sigma(\mathscr{A}(\mathscr{E}))$ est stricte.

Un exemple de la deuxième situation est plus facile à construire que pour la première.

Exemple D.4. Considérons $E = \mathbb{R}$ et $\mathscr{E} = \mathscr{P}(\mathbb{R})$, l'ensemble des parties de $\mathbb{R}$. Déterminer $\mathscr{A}(\mathscr{E})$, vérifier que c'est une partition de $\mathbb{R}$, et que la tribu $\sigma(\mathscr{A}(\mathscr{E}))$ est strictement incluse dans $\mathscr{E}$.

De cet exemple, nous déduisons que *la tribu $\mathscr{P}(\mathbb{R})$ sur $\mathbb{R}$ n'est pas engendrée par une partition.* Dans toute sa généralité, la réponse à notre question initiale est donc négative.

L'exemple qui suit est plus difficile à traiter et peut être laissé de côté en première lecture.

Exemple D.5. Soit X un ensemble infini non dénombrable. Posons $E = \{0,1\}^X$, le produit cartésien d'un nombre infini non dénombrable de copies de $\{0,1\}$ indexées par X. Pour chaque

élément $x \in X$, notons $p_x : E \to \{0,1\}$ la projection sur le facteur indexé par x. Munissons l'ensemble $\{0,1\}$ de la tribu $\mathscr{P}(\{0,1\})$, et munissons E de la tribu

$$\mathscr{E} = \sigma(p_x : x \in X),$$

la plus petite tribu qui rende chacune des applications p_x mesurable. Une autre notation pour $\mathscr{E}$, qui est la tribu produit sur E, est $\mathscr{P}(\{0,1\})^{\otimes X}$.

Nous allons montrer que la tribu $\mathscr{E}$ n'a aucun atome. Pour cela, nous allons montrer d'abord qu'un élément de $\mathscr{E}$ qui contient au moins deux points de E n'est pas minimal pour l'inclusion, donc pas un atome. Seuls les singletons de E pourraient donc être des atomes de $\mathscr{E}$, mais nous allons montrer par ailleurs que $\mathscr{E}$ ne contient aucun singleton. La conclusion est qu'aucun élément non vide de $\mathscr{E}$ n'est minimal pour l'inclusion.

Commençons par le premier point. Soit $A \in \mathscr{E}$ qui contient deux éléments distincts a et b de E. Écrivons $a = (a_x)_{x\in X}$ et $b = (b_x)_{x\in X}$, où a_x et b_x sont, pour tout $x \in X$, des éléments de $\{0,1\}$. Puisque $a \neq b$, il existe $z \in X$ tel que $a_z \neq b_z$. Quitte à échanger a et b, supposons que $a_z = 0$ et $b_z = 1$. L'ensemble $Z = p_z^{-1}(\{0\})$ des éléments de E dont la composante z vaut 0 contient donc a mais pas b. De plus, Z appartient à $\mathscr{E}$. On a donc les inclusions strictes

$$\varnothing \subsetneq A \cap Z \subsetneq A$$

qui montrent que la partie $A \cap Z$ de $\mathscr{E}$ est non vide et strictement incluse dans A, si bien que A n'est pas un atome de $\mathscr{E}$.

Nous avons montré qu'une partie de E qui contient au moins deux éléments n'est pas un atome. Puisqu'un atome est non vide par définition, ce ne peut être qu'un singleton. Nous allons maintenant montrer que $\mathscr{E}$ ne contient aucun singleton. C'est le point un peu plus délicat.

Pour toute partie Y de X, définissons sur E la sous-tribu

$$\mathscr{E}_Y = \sigma(p_x : x \in Y)$$

de $\mathscr{E}$. On a par définition $\mathscr{E}_X = \mathscr{E}$.

La clé de notre argument est l'égalité

$$\mathscr{E} = \bigcup_{\substack{Y \subset X \\ Y \text{ dénombrable}}} \mathscr{E}_Y.$$

La chose remarquable ici est que l'union de sous-tribus dans le membre de droite *est encore* une tribu. Contrairement à ce qui se passe en général, où une union de tribus n'est pas tribu, il n'y a pas ici besoin de considérer la tribu engendrée par cette union.

Soit maintenant a un élément de E. Supposons que le singleton $\{a\}$ appartienne à $\mathscr{E}$. En vertu de l'égalité ci-dessus, il existe une partie dénombrable Y de X telle que $\{a\}$ appartienne à la tribu $\mathscr{E}_Y$.

Nous voulons montrer que c'est impossible, que la tribu $\mathscr{E}_Y$ ne contient aucun singleton. Il n'est pas possible de décrire très explicitement la tribu $\mathscr{E}_Y$, mais nous allons contourner ce problème, de la façon suivante. Notons $p_Y : E \to \{0,1\}^Y$ la projection qui oublie toutes les composantes correspondant à des éléments de $X \setminus Y$. Posons

$$\mathscr{F}_Y = \{p_Y^{-1}(B) : B \subseteq \{0,1\}^Y\}.$$

Alors nous avons l'inclusion $\mathscr{E}_Y \subseteq \mathscr{F}_Y$. En effet, $\mathscr{F}_Y$ est une tribu sur E, qui rend mesurable l'application p_y pour tout $y \in Y$.

Nous affirmons maintenant que la tribu $\mathscr{F}_Y$ elle-même ne contient pas le singleton $\{a\}$ (ni aucun singleton). En effet, supposons que $\{a\}$ appartienne à $\mathscr{F}_Y$. Il existerait alors une partie $B \subseteq \{0,1\}^Y$ telle que $\{a\} = p_Y^{-1}(B)$.

Or, puisque X est infini non dénombrable et Y est dénombrable, il existe un point $x \in X$ qui n'est pas dans Y. Soit $b \in E$ l'élément obtenu en modifiant la composante a_x de a (de 0 à 1 ou de 1 à 0), et en laissant toutes les autres inchangées. Alors b est distinct de a, mais pour tout $y \in Y$, on a $b_y = a_y$. Ceci signifie exactement que $p_Y(b) = p_Y(a)$. Or a appartient à $p_Y^{-1}(B)$, ce qui signifie que $p_Y(a)$ appartient à B. Donc $p_Y(b)$, qui est égal à $p_Y(a)$, appartient aussi à B. Ceci signifie, finalement, que b appartient à $p_Y^{-1}(B)$. Ainsi, $b \in \{a\}$, ce qui est contradictoire.

Ceci achève la démonstration du fait que la tribu $\mathscr{E}$ n'a aucun atome.

D.4 Une relation d'équivalence

Nous allons faire un deuxième essai pour associer à toute tribu sur un ensemble une partition de cet ensemble.

Soit E un ensemble et $\mathscr{E}$ une tribu sur E. Nous allons définir une relation sur E. Pour tous $x, y \in E$, écrivons $x \sim y$ si tout élément de $\mathscr{E}$ qui contient x contient également y. En symboles,

$$(x \sim y) \Leftrightarrow (\forall A \in \mathscr{E},\ x \in A \Rightarrow y \in A).$$

Proposition D.6. *La relation $\sim$ est une relation d'équivalence sur E.*

L'ensemble des classes d'équivalence de la relation $\sim$ est une partition de $\mathscr{E}$, que nous notons $\mathscr{K}(\mathscr{E})$. Un atome de $\mathscr{E}$ est toujours une classe d'équivalence de la relation $\sim$. Autrement dit, on a toujours l'inclusion

$$\mathscr{A}(\mathscr{E}) \subseteq \mathscr{K}(\mathscr{E}).$$

En particulier, si $\mathscr{A}(\mathscr{E})$ est une partition de E, alors on a égalité : c'est par exemple le cas lorsque la tribu $\mathscr{E}$ est engendrée par une partition de E, ou lorsque $E = \mathbb{R}$ et $\mathscr{E} = \mathscr{P}(\mathbb{R})$.

Dans le cas traité dans le long exemple à la fin de la section précédente, de la tribu produit sur $\{0,1\}^X$ où X est infini non dénombrable, la relation $\sim$ est la relation d'égalité, et ses classes d'équivalence les singletons. Cet exemple montre que que *les classes de la relation $\sim$ n'appartiennent pas nécessairement à la tribu $\mathscr{E}$.*

Donnons une autre description de ces classes.

Proposition D.7. *Pour tout $x \in E$, la classe d'équivalence de x pour la relation $\sim$ est l'intersection de tous les éléments de $\mathscr{E}$ qui contiennent x. Autrement dit, l'unique élément de $\mathscr{K}(\mathscr{E})$ qui contient x est la partie*

$$K(x) = \bigcap_{\substack{A \in \mathscr{E} \\ x \in A}} A.$$

Le fait que la partie $K(x)$ n'appartienne pas nécessairement à $\mathscr{E}$ est dû au fait que l'intersection ci-dessus n'est pas nécessairement l'intersection d'une famille dénombrable d'éléments de $\mathscr{E}$.

Ceci nous met sur la piste d'une hypothèse qui garantit que les éléments de $\mathscr{K}(\mathscr{E})$ appartiennent à $\mathscr{E}$. Avant de l'explorer, tirons une conclusion importante du fait que dès qu'un élément A de $\mathscr{E}$ contient un élément x de E, la classe d'équivalence de x est incluse dans A.

Proposition D.8. *Tout élément de $\mathscr{E}$ est une réunion d'éléments de $\mathscr{K}(\mathscr{E})$.*

Dans le langage des relations d'équivalence, on peut dire que tout élément de $\mathscr{E}$ est *saturé* pour la relation $\sim$.

D.5 Le cas des tribus dénombrables

Soit $\mathscr{E}$ une tribu dénombrable sur l'ensemble E. Alors toute intersection d'éléments de $\mathscr{E}$ est une intersection dénombrable. En particulier, en vertu de la proposition D.7, la tribu $\mathscr{E}$ contient les classes d'équivalence de la relation $\sim$, c'est-à-dire la partition $\mathscr{K}(\mathscr{E})$ de E. On a donc

$$\sigma(\mathscr{K}(\mathscr{E})) \subseteq \mathscr{E}.$$

Mais la proposition D.8 nous assure que tout élément de $\mathscr{E}$ est une réunion d'éléments de $\mathscr{K}(\mathscr{E})$. Or, puisque la partition $\mathscr{K}(\mathscr{E})$, qui est incluse dans $\mathscr{E}$, est dénombrable, la tribu engendrée par $\mathscr{K}(\mathscr{E})$ est exactement l'ensemble des unions d'éléments de $\mathscr{K}(\mathscr{E})$. Ainsi,

$$\mathscr{E} \subseteq \sigma(\mathscr{K}(\mathscr{E})).$$

Nous avons donc l'égalité

$$\mathscr{E} = \sigma(\mathscr{K}(\mathscr{E})),$$

et nous avons finalement montré le résultat suivant.

Théorème D.9. *Toute tribu dénombrable sur un ensemble est engendrée par une partition de cet ensemble.*

Nous pouvons encore faire un pas. Considérons une tribu dénombrable $\mathscr{E}$ sur un ensemble E. Elle est, d'après notre théorème, engendrée par la partition $\mathscr{K}(\mathscr{E})$. Cette partition est incluse dans $\mathscr{E}$, elle est donc finie ou infinie dénombrable. Si elle était infinie dénombrable, la tribu $\mathscr{E}$ serait en bijection avec l'ensemble des parties de $\mathscr{K}(\mathscr{E})$, et serait infinie non dénombrable.

Il s'ensuit que la partition $\mathscr{K}(\mathscr{E})$ ne peut qu'être finie, et la tribu $\mathscr{E}$ elle-même est donc finie. Nous avons donc montré le résultat suivant.

Théorème D.10. *Toute tribu dénombrable sur un ensemble est finie.*

Le cardinal d'une tribu est donc soit de la forme 2^n pour un certain n entier naturel, soit infini non dénombrable. Il n'existe sur aucun ensemble de tribu dénombrable.