

Sabin Lessard

Les rudiments du calcul stochastique

Cours et exercices corrigés



Chapitre 1

Rappels sur la théorie des probabilités

1.1 Introduction

Dans ce chapitre, on présente les fondements du calcul des probabilités. On commence avec la définition même d'une probabilité comme d'une mesure des chances de réalisation d'événements et on termine avec les principaux théorèmes de convergence qui permettent d'approfondir les concepts. Cela passe par un rappel des notions d'indépendance, de variable aléatoire et surtout d'espérance, qui correspond à une intégrale, sans oublier la notion plus abstraite d'espérance conditionnelle. Les différents modes de convergence et les inégalités les plus importantes sont également passés en revue. Quant aux résultats de convergence, ils concernent principalement les moyennes de variables aléatoires indépendantes de même loi de probabilité, ainsi que les sommes partielles de variables aléatoires indépendantes d'espérance nulle qui représentent l'exemple type d'une martingale à temps discret.

1.2 Espace de probabilité

Un espace de probabilité, aussi appelé espace probabilisé, est décrit par un triplet $(\Omega, \mathcal{F}, \mathbb{P})$, où Ω est un ensemble fondamental, \mathcal{F} une collection d'événements et \mathbb{P} une mesure de probabilité.

L'ensemble fondamental Ω est l'ensemble de tous les résultats élémentaires possibles d'une expérience aléatoire. Un événement est un sous-ensemble de l'ensemble fondamental Ω . Cependant, tous les sous-ensembles ne sont pas nécessairement des événements. Étant donné un résultat élémentaire $\omega \in \Omega$, on dit qu'un événement $A \subseteq \Omega$ se réalise si $\omega \in A$, et qu'il ne se réalise pas sinon.

La collection des événements \mathcal{F} vérifie les conditions suivantes :

- (1) $\emptyset \in \mathcal{F}$;
- (2) $A^c \in \mathcal{F}$ si $A \in \mathcal{F}$;
- (3) $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{F}$ si $A_i \in \mathcal{F}$ pour tout entier $i \geq 1$.

Sous ces conditions, on a aussi :

- (4) $\bigcup_{i=1}^n A_i \in \mathcal{F}$ si $A_i \in \mathcal{F}$ pour tout entier $i = 1, \dots, n$.

Les conditions ci-dessus définissent une σ -algèbre, appelée aussi tribu. Sans la condition (3), on a seulement une algèbre.

La probabilité est une application $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ qui associe à chaque événement une mesure des chances de réalisation de cet événement. Cette application vérifie les conditions suivantes :

- (a) $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$;
- (b) $\mathbb{P}(\bigcup_{i=1}^{+\infty} A_i) = \sum_{i=1}^{+\infty} \mathbb{P}(A_i)$ si $A_i \in \mathcal{F}$ pour tout entier $i \geq 1$ et $A_i \cap A_j = \emptyset$ dès que $i \neq j$.

Sous ces conditions, on a aussi :

- (c) $\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$ si $A_i \in \mathcal{F}$ pour tout entier $i = 1, \dots, n$ avec $A_i \cap A_j = \emptyset$ dès que $i \neq j$;
- (d) $\mathbb{P}(\bigcup_{i=1}^{+\infty} A_i) = \lim_{n \rightarrow +\infty} \mathbb{P}(\bigcup_{i=1}^n A_i)$, $\mathbb{P}(\bigcap_{i=1}^{+\infty} A_i) = \lim_{n \rightarrow +\infty} \mathbb{P}(\bigcap_{i=1}^n A_i)$, si $A_i \in \mathcal{F}$ pour tout entier $i \geq 1$.

La condition (b) est la propriété de σ -additivité. Avec la condition $\mathbb{P}(\emptyset) = 0$, cette propriété fait de \mathbb{P} une mesure. En y ajoutant la condition $\mathbb{P}(\Omega) = 1$, on obtient une mesure de probabilité.

La condition (c) est la propriété d'additivité et la condition (d) la propriété de continuité. Ces conditions sont équivalentes à la condition de σ -additivité.

1.3 Espace de mesure

Plus généralement, un espace de mesure est un triplet $(\mathbb{X}, \mathcal{M}, \mu)$, où \mathbb{X} est un ensemble, \mathcal{M} est une σ -algèbre de sous-ensembles de \mathbb{X} , appelés ensembles

mesurables, et enfin $\mu : \mathcal{M} \rightarrow [0, +\infty]$ est une mesure, c'est-à-dire une application σ -additive qui vérifie la condition $\mu(\emptyset) = 0$. La mesure est finie si $\mu(\mathbb{X}) < +\infty$. Elle est σ -finie si $\mathbb{X} = \bigcup_{n=1}^{+\infty} \mathbb{X}_n$ avec $\mathbb{X}_n \subseteq \mathbb{X}_{n+1}$ et $\mu(\mathbb{X}_n) < +\infty$ pour tout entier $n \geq 1$.

Un ensemble négligeable est un ensemble de mesure 0 ou une partie d'un tel ensemble. La σ -algèbre complétée d'une σ -algèbre \mathcal{M} de sous-ensembles de \mathbb{X} est définie par

$$\overline{\mathcal{M}} = \{E \subseteq \mathbb{X} : A \subseteq E \subseteq B \text{ avec } A, B \in \mathcal{M} \text{ et } B \setminus A = B \cap A^c \in \mathcal{N}\},$$

où \mathcal{N} est la collection des ensembles négligeables, c'est-à-dire

$$\mathcal{N} = \{N \subseteq \mathbb{X} : N \subseteq M \in \mathcal{M} \text{ avec } \mu(M) = 0\}.$$

La σ -algèbre complétée de \mathcal{M} est la plus petite σ -algèbre qui contient \mathcal{M} et \mathcal{N} , notée $\sigma(\mathcal{M} \cup \mathcal{N})$.

Une mesure est complète si tous les ensembles négligeables sont mesurables, c'est-à-dire $\mathcal{N} \subseteq \mathcal{M}$. La mesure complétée d'une mesure μ sur \mathcal{M} est définie par

$$\overline{\mu}(E) = \mu(A)$$

si $A \subseteq E \subseteq B$ avec $A, B \in \mathcal{M}$ et $B \setminus A \in \mathcal{N}$. L'espace de mesure $(\mathbb{X}, \overline{\mathcal{M}}, \overline{\mu})$ est dit complété.

La tribu de Borel sur un espace topologique est définie comme la plus petite σ -algèbre sur cet espace qui contient tous les ensembles ouverts. Un ensemble appartenant à cette σ -algèbre est appelé un ensemble borélien ou, plus simplement, un borélien.

La tribu de Borel sur l'ensemble des nombres réels \mathbb{R} , notée $\mathcal{B}(\mathbb{R})$, correspond à la plus petite σ -algèbre sur \mathbb{R} contenant tous les intervalles de forme $(a, b]$ pour des nombres réels a et b vérifiant $a < b$.

La mesure de Lebesgue λ sur \mathbb{R} est l'unique mesure par rapport à la tribu $\mathcal{B}(\mathbb{R})$ qui est telle que

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

pour $-\infty < a < b < +\infty$. Cette mesure est σ -finie, car $\lambda((-n, n]) = 2n < +\infty$ pour tout entier $n \geq 1$ et $\bigcup_{n=1}^{+\infty} (-n, n] = \mathbb{R}$. D'autre part, la mesure de Lebesgue λ restreinte aux sous-ensembles de $(0, 1]$ dans $\mathcal{B}(\mathbb{R})$ est une probabilité.

La σ -algèbre complétée de $\mathcal{B}(\mathbb{R})$, notée $\mathcal{L}(\mathbb{R})$, est appelée la tribu de Lebesgue. Cette σ -algèbre est la plus petite par rapport à laquelle la mesure de Lebesgue λ sur \mathbb{R} est complète, et celle qui est utilisée dans la définition générale de cette mesure.

De même, la mesure de Lebesgue λ_n sur \mathbb{R}^n muni de la tribu de Borel $\mathcal{B}(\mathbb{R}^n)$, qui correspond à la plus petite σ -algèbre sur \mathbb{R}^n contenant tout pavé

de forme $\prod_{i=1}^n (a_i, b_i]$, où $a_i < b_i$ avec $a_i, b_i \in \mathbb{R}$ pour $i = 1, \dots, n$, est l'unique mesure sur cette tribu telle que

$$\lambda_n \left(\prod_{i=1}^n (a_i, b_i] \right) = \prod_{i=1}^n (b_i - a_i)$$

pour tout pavé de cette forme. De plus, $\mathcal{L}(\mathbb{R}^n) = \overline{\mathcal{B}(\mathbb{R}^n)}$ est la plus petite σ -algèbre par rapport à laquelle λ_n est complète, et celle utilisée dans la définition générale de cette mesure.

En fait, l'espace de mesure $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda_n)$ est l'espace produit de n espaces de mesure $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. Plus généralement, on définit l'espace produit de n espaces de mesure $(\mathbb{X}_i, \mathcal{M}_i, \mu_i)$ pour $i = 1, \dots, n$ par le triplet $(\mathbb{X}, \mathcal{M}, \mu)$, où $\mathbb{X} = \prod_{i=1}^n \mathbb{X}_i$, $\mathcal{M} = \bigotimes_{i=1}^n \mathcal{M}_i$, qui est la σ -algèbre produit définie comme la plus petite σ -algèbre contenant tout ensemble de forme $\prod_{i=1}^n A_i$ avec $A_i \in \mathcal{M}_i$ pour $i = 1, \dots, n$, et

$$\mu \left(\prod_{i=1}^n A_i \right) = \prod_{i=1}^n \mu_i(A_i)$$

pour tout ensemble de cette forme. De plus, la mesure produit $\mu = \bigotimes_{i=1}^n \mu_i$ est unique si les mesures μ_i pour $i = 1, \dots, n$ sont σ -finies.

L'existence et l'unicité de mesures vérifiant les propriétés ci-dessus sont garanties par les résultats fondamentaux ci-dessous.

Proposition (Théorème d'extension de Carathéodory). *Soit*

$$\mu : \mathcal{A} \rightarrow [0, +\infty]$$

une application σ -additive sur une algèbre \mathcal{A} de sous-ensembles de S avec $\mu(\emptyset) = 0$. Alors, l'application

$$\mu^* : \sigma(\mathcal{A}) \rightarrow [0, +\infty]$$

définie pour tout A dans $\sigma(\mathcal{A})$, la plus petite σ -algèbre contenant \mathcal{A} , par

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^{+\infty} \mu(A_n) : A_n \in \mathcal{A} \text{ pour } n \geq 1, A \subseteq \bigcup_{n=1}^{+\infty} A_n \right\}$$

est une mesure, appelée mesure extérieure, qui est telle que $\mu^ = \mu$ sur \mathcal{A} .*

Proposition (Théorème d'unicité de Hahn). *Soient*

$$\mu_1, \mu_2 : \sigma(\Pi) \rightarrow [0, +\infty]$$

deux mesures σ -finies définies sur $\sigma(\Pi)$, la plus petite σ -algèbre contenant un π -système Π de sous-ensembles de S , c'est-à-dire un système tel que $I \cap J \in \Pi$ si $I, J \in \Pi$. Si $\mu_1(I) = \mu_2(I)$ pour tout $I \in \Pi$, alors $\mu_1 = \mu_2$ sur $\sigma(\Pi)$.

Proposition (Théorème λ - π de Dynkin). *Soit \mathcal{C} une collection de sous-ensembles de S qui est un λ -système, c'est-à-dire une collection qui vérifie :*

- (i) $\emptyset \in \mathcal{C}$;
 - (ii) $B^c \in \mathcal{C}$ si $B \in \mathcal{C}$;
 - (iii) $\bigcup_{n=1}^{+\infty} B_n \in \mathcal{C}$ si $B_n \in \mathcal{C}$ pour $n \geq 1$ avec $B_n \cap B_m = \emptyset$ dès que $n \neq m$.
- Si \mathcal{C} contient un π -système Π , alors \mathcal{C} contient nécessairement $\sigma(\Pi)$, la plus petite σ -algèbre contenant Π .*

1.4 Variable aléatoire

Une application $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire (v.a.) par rapport à une espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ si l'image réciproque de tout borélien de \mathbb{R} est un événement, c'est-à-dire

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

pour tout $B \in \mathcal{B}(\mathbb{R})$. En fait, il suffit que cela soit vrai pour tout B de forme $(-\infty, x]$ pour $x \in \mathbb{R}$. La σ -algèbre engendrée par la variable aléatoire X définie par

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

est alors une sous-tribu de \mathcal{F} .

La notion de variable aléatoire peut être étendue à une application X à valeurs dans $\mathbb{R} \cup \{-\infty, +\infty\}$ telle que $X^{-1}(B \cup I) \in \mathcal{F}$ pour tout $B \in \mathcal{B}(\mathbb{R})$ et tout $I \subseteq \{-\infty, +\infty\}$.

Une variable aléatoire indicatrice est définie par

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \in A^c, \end{cases}$$

où $A \in \mathcal{F}$. Une variable aléatoire étagée est une variable aléatoire de forme

$$X = \sum_{i=1}^n x_i \mathbb{1}_{A_i},$$

où $x_i \in \mathbb{R}$ et $A_i \in \mathcal{F}$ pour $i = 1, \dots, n$ avec $A_i \cap A_j = \emptyset$ dès que $i \neq j$ et $\bigcup_{i=1}^n A_i = \Omega$. Dans ce cas, on a

$$\sigma(X) = \left\{ \bigcup_{i \in I} A_i : I \subseteq \{1, \dots, n\} \right\}$$

comme σ -algèbre engendrée par X .

La loi de probabilité d'une variable aléatoire X par rapport à un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ est définie comme l'application $\mathbb{P}_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ donnée par

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$$

pour tout $B \in \mathcal{B}(\mathbb{R})$. Cette loi est entièrement déterminée par la fonction de distribution $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par

$$F_X(x) = \mathbb{P}(X^{-1}((-\infty, x])) = \mathbb{P}(X \leq x)$$

pour tout $x \in \mathbb{R}$.

La variable aléatoire X est dite à densité s'il existe une fonction $f_X \geq 0$ telle que

$$\mathbb{P}_X(B) = \int_B f_X(x) dx = \int_{\mathbb{R}} \mathbf{1}_B(x) f_X(x) dx$$

pour tout $B \in \mathcal{B}(\mathbb{R})$, où l'intégrale à droite est par rapport à la mesure de Lebesgue sur \mathbb{R} (voir la section 1.10 sur l'intégrale d'une fonction mesurable). En fait, il suffit que

$$F_X(x) = \int_{(-\infty, x]} f_X(y) dy = \int_{-\infty}^x f_X(y) dy$$

pour tout $x \in \mathbb{R}$.

Dans la plupart des applications, la fonction de densité f_X est continue par morceaux, auquel cas l'intégrale ci-dessus pour la fonction de distribution F_X coïncide avec l'intégrale au sens de Riemann. De plus, la fonction f_X en ses points de continuité est alors donnée par la dérivée de F_X en ces points. Dans tous les cas où f_X est une fonction intégrable sur \mathbb{R} au sens de Lebesgue, la fonction F_X définie ci-dessus est continue et on a

$$F'_X = f_X$$

presque partout, c'est-à-dire partout sauf peut-être sur un ensemble de mesure de Lebesgue nulle, donc un ensemble négligeable. C'est le théorème fondamental du calcul.

Une variable aléatoire X avec fonction de densité

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

pour tout $x \in \mathbb{R}$, où $-\infty < \mu < +\infty$ et $\sigma^2 > 0$, est dite de loi gaussienne de paramètres μ et σ^2 , appelée aussi loi normale et notée $N(\mu, \sigma^2)$. La loi normale est dite centrée lorsque $\mu = 0$ et réduite lorsque $\sigma^2 = 1$. La loi $N(0, 1)$ est appelée la loi normale standard.

1.5 Vecteur aléatoire

Un vecteur aléatoire par rapport à une espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ est une application

$$\mathbf{X} = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$$

telle que

$$\sigma(\mathbf{X}) = \mathbf{X}^{-1}(\mathcal{B}(\mathbb{R}^n)) \subseteq \mathcal{F}.$$

Les composantes X_1, \dots, X_n du vecteur \mathbf{X} sont alors des variables aléatoires et $\sigma(\mathbf{X})$, la σ -algèbre engendrée par \mathbf{X} , est la plus petite σ -algèbre qui contienne $\sigma(X_1), \dots, \sigma(X_n)$.

La loi de probabilité du vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$ définie par

$$\mathbb{P}_{\mathbf{X}}(B) = \mathbb{P}(\mathbf{X} \in B)$$

pour tout $B \in \mathcal{B}(\mathbb{R}^n)$, appelée aussi loi de probabilité conjointe des variables aléatoires X_1, \dots, X_n , est entièrement déterminée par la fonction de distribution de \mathbf{X} , appelée aussi fonction de distribution conjointe de X_1, \dots, X_n . Celle-ci est donnée par

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$$

pour tout $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$.

Le vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$ est à densité s'il existe une fonction $f_{\mathbf{X}} \geq 0$, appelée fonction de densité conjointe de X_1, \dots, X_n , telle que

$$\mathbb{P}_{\mathbf{X}}(B) = \int_B f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^n} \mathbb{1}_B(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

pour tout $B \in \mathcal{B}(\mathbb{R}^n)$, où l'intégrale à droite est par rapport à la mesure de Lebesgue sur \mathbb{R}^n (voir la section 1.10 sur l'intégrale d'une fonction mesurable). En fait, il suffit que l'identité ci-dessus soit vraie pour tout B de la forme $\{X_1 \leq x_1, \dots, X_n \leq x_n\}$. Chaque composante X_i d'un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$ est une variable aléatoire de loi de probabilité dite marginale donnée par

$$\mathbb{P}_{X_i}(B_i) = \mathbb{P}_{\mathbf{X}}(B_{(i)}),$$

où $B_{(i)} = \mathbb{R}^{i-1} \times B_i \times \mathbb{R}^{n-i}$ pour $B_i \in \mathcal{B}(\mathbb{R})$ et $i = 1, \dots, n$. Si \mathbf{X} est un vecteur aléatoire avec fonction de densité $f_{\mathbf{X}}$, alors X_i est une variable aléatoire avec fonction de densité dite marginale donnée par

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(x_1, \dots, x_n) d(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

pour $i = 1, \dots, n$.

Un vecteur aléatoire à densité $\mathbf{X} = (X_1, \dots, X_n)$ est de loi gaussienne multidimensionnelle, ou loi normale multidimensionnelle, de paramètres donnés par le vecteur colonne $\boldsymbol{\mu} \in \mathbb{R}^n$ et la matrice symétrique définie positive $\boldsymbol{\Sigma}$ de taille n si la fonction de densité conjointe est de forme

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

pour tout vecteur colonne $\mathbf{x} \in \mathbb{R}^n$, où $|\boldsymbol{\Sigma}|$ représente le déterminant de $\boldsymbol{\Sigma}$ et T désigne la transposée. Cette loi est notée $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Une propriété importante de cette loi est que, si

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \boldsymbol{\nu},$$

où $\boldsymbol{\nu} \in \mathbb{R}^m$ et \mathbf{A} est une matrice de taille $m \times n$ dont le rang est m , ce qui est le cas si et seulement si $\mathbf{A}\mathbf{A}^T$ est définie positive, alors $\mathbf{Y} = (Y_1, \dots, Y_m)$ est un vecteur aléatoire à densité de loi $N(\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\nu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$.

1.6 Événements indépendants

Les événements A_1, \dots, A_n dans un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sont des événements indépendants si

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i)$$

pour toute partie $I \subseteq \{1, \dots, n\}$. Si cela est le cas pour tout entier $n \geq 1$, alors $(A_n)_{n \geq 1}$ est une suite d'événements indépendants.

Pour une suite d'événements $(A_n)_{n \geq 1}$, l'événement $\{A_n \text{ i.o.}\}$ avec *i.o.* pour *infinitely often* dans le sens pour une infinité de n est défini par

$$\{A_n \text{ i.o.}\} = \bigcap_{m=1}^{+\infty} \bigcup_{n=m}^{+\infty} A_n.$$

Son complémentaire est l'événement $\{A_n^c \text{ ev.}\}$ avec *ev.* pour *eventually* dans le sens pour tout n assez grand, c'est-à-dire

$$\{A_n^c \text{ ev.}\} = \bigcup_{m=1}^{+\infty} \bigcap_{n=m}^{+\infty} A_n^c.$$

La probabilité de ces événements est 0 ou 1 dans deux situations importantes.

Proposition (Lemme de Borel-Cantelli). *On a :*

- (i) $\mathbb{P}(A_n \text{ i.o.}) = 0$ si $\sum_{n=1}^{+\infty} \mathbb{P}(A_n) < +\infty$;
- (ii) $\mathbb{P}(A_n \text{ i.o.}) = 1$ si $\sum_{n=1}^{+\infty} \mathbb{P}(A_n) = +\infty$ et $(A_n)_{n \geq 1}$ est une suite d'événements indépendants.