Cours de probabilités, master 1

Yves Coudène

19 janvier 2015

Table des matières

Introduction

Ces notes proviennent d'un cours de Master première année donné à l'université de Brest sur la période 2011-2014. Le cours était composé de douze séances de deux heures et portait sur les théorèmes de convergence en théorie des probabilités. La frappe du texte a bénéficié de l'aide de Sabine Chanthara, je l'en remercie vivement.

Ce cours est destiné à des étudiants ayant déjà suivi un cours d'intégrale de Lebesgue. Une annexe en fin d'ouvrage rappelle les résultats d'intégration qui sont utilisés dans le corps de ce texte. Il est aussi important d'être familier avec les bases de la théorie hilbertienne que nous appliquons à l'espace L^2 à plusieurs reprises, par exemple pour définir la notion d'espérance conditionnelle. Enfin, un minimum de familiarité avec la théorie des probabilités discrètes, comme on peut la voir au lycée, est fortement conseillé.

On s'est concentré sur les théorèmes de convergence classiques, essentiellement dans le cadre indépendant : loi faible et forte des grands nombres, théorème de la limite centrée, convergence des martingales bornées dans L^2, théorème des trois séries, loi du 0-1 de Kolmogorov. Un résumé des théorèmes et des formules présentés dans le cours se trouve en annexe.

Le texte est organisé de façon à parvenir assez rapidement à la preuve de la loi forte des grands nombres, au chapitre 4, qui est faite pour des variables de carré intégrable. Le cas intégrable est traité plus tard, dans le chapitre concernant les martingales, comme corollaire des théorèmes de convergence pour ces martingales. Le second objectif est le théorème de la limite centrée, atteint au chapitre 6. Il faut pour cela étudier en détail les différents types de convergence et les relations qui s'établissent entre eux. On termine par la notion de martingale, qui permet de démontrer quelques résultats classiques de convergence, comme le théorème des trois séries et la loi du 0-1 de Kolmogorov. La théorie des martingales et des temps d'arrêt est illustrée par une marche aléatoire symétrique sur l'espace des entiers.

La théorie des chaînes de Markov n'est pas abordée dans ce texte. De même, on ne parle pas de sous et de sur-martingales, et la notion de convergence étroite est étudiée pour des mesures de probabilité définies sur $\mathbf{R}$ uniquement, ce qui permet quelques simplifications dans les preuves.

Notations

Les ensembles des nombres entiers, entiers relatifs, rationnels, réels et complexes sont notés respectivement $\mathbf{N}, \mathbf{Z}, \mathbf{Q}, \mathbf{R}, \mathbf{C}$.

On travaille en genéral sur un espace probabilisé $(\Omega, \mathcal{T}, P)$.

$\mathbf{1}_A$ .... fonction indicatrice de A
$B(x,r)$ .... boule ouverte de centre x de rayon r
C^∞ .... ensemble des fonctions indéfiniment différentiables
$Cov(X,Y)$ .... covariance de X et Y
δ_ω .... mesure de Dirac au point ω
$E(X)$ .... espérance de X
$E(X \mid \mathcal{F})$ .... espérance conditionnelle de X sachant $\mathcal{F}$
$\mathcal{F}$ .... tribu
F_X .... fonction de répartition
L^p .... espace des classes de fonctions L^p
$\overline{\lim}$ .... limite supérieure
$\underline{\lim}$ .... limite inférieure
μ .... mesure
$\mathbf{N}^*$ .... nombres entiers non nuls
Ω .... ensemble de résultats
$\circ$ .... composition
$\emptyset$ .... ensemble vide
P .... mesure de probabilité
P_X .... loi de la variable aléatoire X
$P_{(X,Y)}$ .... loi du couple (X,Y)
$P(A \mid \mathcal{F})$ .... probabilité conditionnelle de A sachant $\mathcal{F}$
$P \otimes Q$ .... produit des probabilités P et Q
$p.s.$ .... presque partout
S_n .... somme de X_1 à X_n
$\sigma(X)$ .... écart-type de X
$\mathcal{T}$ .... tribu
$\mathcal{T}_1 \otimes \mathcal{T}_2$ .... produit des tribus $\mathcal{T}_1$ et $\mathcal{T}_2$
$m \wedge n$ .... minimum de m et n
$V(X)$ .... variance de X
X .... variable aléatoire
$\langle X, Y\rangle$ .... produit scalaire dans L^2
$\|X\|_p$ .... norme L^p de X

Chapitre 1

Formalisme de Kolmogorov

Nous désignons par *épreuve* une expérience ou une observation réalisée dans des conditions bien définies (protocole expérimental) reproductible, et dont le résultat est l'un des éléments d'un ensemble déterminé (univers). Le but de la théorie des probabilités est d'associer à certains sous-ensembles de cet univers, appelés événements, un nombre réel compris entre 0 et 1, qui reflète notre degré de confiance dans la réalisation de l'événement une fois que l'épreuve a eu lieu.

La théorie moderne des probabilités est formalisée par Kolmogorov en 1933, en se basant sur la théorie de la mesure. La notion clef est celle d'espace probabilisé.

Définition 1 *Un espace probabilisé* $(\Omega, \mathcal{T}, P)$ *est la donnée :*
– *d'un ensemble* Ω *appelé univers, dont les éléments sont appelés résultats,*
– *d'une tribu* $\mathcal{T}$ *de parties de* Ω*, dont les éléments sont appelés événements,*
– *d'une mesure* P *définie sur la tribu* $\mathcal{T}$*, qui satisfait* $P(\Omega) = 1$*.*

Commençons par décrire trois exemples importants d'espaces probabilisés.

1.1 Le cas discret : Ω fini ou dénombrable

Pour $\mathcal{T}$, on prend l'ensemble des parties de Ω : $\mathcal{T} = \mathcal{P}(\Omega)$. Se donner une probabilité $P : \mathcal{T} \to [0,1]$ revient à se donner une famille de nombres réels p_ω, $\omega \in \Omega$, qui satisfait

– $0 \leq p_\omega \leq 1$ pour tout $\omega \in \Omega$,
– $\sum\limits_{\omega \in \Omega} p_\omega = 1$.

La correspondance entre P et les p_ω est donnée par

$$p_\omega = P(\{\omega\}), \quad P(A) = \sum_{\omega \in A} p_\omega.$$

Notons δ_ω la mesure de Dirac au point ω :

$$\forall A \in \mathcal{T},\ \delta_\omega(A) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$$

On peut exprimer la probabilité P comme une somme de Dirac : $P = \sum_{\omega\in\Omega} p_\omega \delta_\omega$. On a alors, pour $g : \Omega \to \mathbf{R}$ mesurable, positif ou P-intégrable,

$$\int_\Omega g\, dP = \sum_{\omega\in\Omega} p_\omega\, g(\omega).$$

Exemples

• Loi uniforme sur $\Omega = \{1, 2, ..., n\}$:

$$p_\omega = 1/n, \qquad P(A) = \frac{\#A}{\#\Omega}.$$

Le lancé d'un dé à 6 faces bien équilibré est modélisé par un tel espace probabilisé ($n = 6$).

• Loi binomiale de paramètres $n \in \mathbf{N}^*$, $p \in [0, 1]$, sur $\Omega = \{0, ..., n\}$:

$$p_k = P(\{k\}) = C_n^k\, p^k(1-p)^{n-k} \qquad \text{pour } k \in \{0, ..., n\}.$$

p_k est la probabilité d'obtenir k succès exactement au cours de n tirages indépendants, sachant que la probabilité de succès lors d'un tirage est égale à p.

• Loi de Poisson sur $\Omega = \mathbf{N}$ de paramètre $\lambda > 0$:

$$p_k = P(\{k\}) = \frac{\lambda^k}{k!}\, e^{-\lambda} \text{ pour } k \in \mathbf{N}.$$

1.2 Le cas continu : $\Omega = \mathbf{R}$ ou $\mathbf{R}^d$

Ici $\mathcal{T}$ est la tribu engendrée par les intervalles de $\mathbf{R}$ ou les rectangles de $\mathbf{R}^d$. Ses éléments sont appelés boréliens. On peut définir une mesure de probabilité sur Ω à partir d'une densité $f : \Omega \to \mathbf{R}_+$ satisfaisant les conditions suivantes :

– f est borélienne,

– $\forall \omega \in \Omega, \quad f(\omega) \geq 0$,

– $\int_\Omega f d\lambda = 1$.

On a noté la mesure de Lebesgue sur Ω avec un λ. La mesure de probabilité P associée à la densité f est donnée par

$$P(A) = \int_A f d\lambda = \int_A f(x)dx.$$

On a alors pour toute fonction mesurable $g : \Omega \to \mathbf{R}$ positive ou P-intégrable,

$$\int_\Omega g\, dP = \int_\Omega g(x) f(x)\, dx.$$

Exemples
– Probabilité uniforme sur $[a, b]$, avec $a, b \in \mathbf{R}$, $a < b$:

$$f = \frac{1}{b-a}\ \mathbf{1}_{[a,b]}$$

– Loi de Laplace-Gauss ou loi normale de paramètres $m \in \mathbf{R}$, $\sigma > 0$:

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

Elle est dite centrée si $m = 0$ et $\sigma = 1$.
– Probabilité exponentielle de paramètre $l > 0$:

$$f(x) = l\ e^{lx}\ \mathbf{1}_{\mathbf{R}_+}(x)$$

1.3 Le cas des espaces produits

On s'intéresse à une épreuve modélisée par un espace probabilisé $(\Omega, \mathcal{T}, P)$ et on veut répéter cette épreuve plusieurs fois de manière indépendante, disons n fois, $n \in \mathbf{N}^*$. Pour cela, on considère :

• l'univers $\Omega^n = \Omega \times \Omega \times ... \times \Omega$, ses éléments sont des multiplets $(\omega_1, \omega_2, ..., \omega_n) \in \Omega^n$. L'élément ω_1 est le résultat obtenu lors de la première épreuve, ω_2 lors de la seconde épreuve etc.

• La tribu produit $\mathcal{T} \otimes \mathcal{T} \otimes ...\mathcal{T} = \mathcal{T}^{\otimes n}$. C'est la tribu engendrée par les parties de Ω^n de la forme $A_1 \times A_2 \times ... \times A_n$, avec $A_i \in \mathcal{T}$ pour tout i.

• Dans le cas indépendant, la mesure produit $P \otimes ... \otimes P$ sur cette tribu. Cette mesure $P^{\otimes n}$ est l'unique mesure vérifiant

$$P^{\otimes n}(A_1 \times A_2 \times ... \times A_n) = P(A_1)P(A_2)...P(A_n)$$

pour tout $A_1, ..., A_n \in \mathcal{T}$.

On va chercher à étudier le comportement asymptotique d'une répétition d'épreuves, effectuées de manière indépendante, quand n tend vers l'infini. Pour cela, nous introduisons un nouvel espace probabilisé.

• L'univers $\Omega^{\mathbf{N}}$ est l'ensemble de toutes les suites d'éléments de Ω.

• On se place sur la tribu produit $\mathcal{T}^{\otimes \mathbf{N}}$. C'est la tribu de parties de $\Omega^{\mathbf{N}}$ engendrée par les cylindres de la forme

$$C_{A_0,...,A_n} = \{(\omega_i)_{i\in\mathbf{N}} \mid \forall i = 0...n,\ \omega_i \in A_i\}$$

avec $n \in \mathbf{N}$ et $A_0, ..., A_n \in \mathcal{T}$.

• Dans le cas indépendant, on considère sur $\mathcal{T}^{\otimes \mathbf{N}}$ la mesure produit $P^{\otimes \mathbf{N}}$, caractérisée de la façon suivante :

Théorème 1 (Kolmogorov) *Soit* $(\Omega, \mathcal{T}, P)$ *un espace probabilisé. Alors il existe une unique mesure de probabilité sur* $\mathcal{T}^{\otimes \mathbf{N}}$ *, notée* $P^{\otimes \mathbf{N}}$*, qui satisfait*

$$P(C_{A_0,...,A_n}) = P(A_0)P(A_1)...P(A_n)$$

pour tout $n \in \mathbf{N}$ *et* $A_0, ..., A_n \in \mathcal{T}$.

L'exemple le plus simple est donné par la répétition un nombre arbitrairement grand de fois du lancer d'une pièce de monnaie. L'univers est donné par l'ensemble de toutes les suites de pile ou face : $\{pile, face\}^{\mathbf{N}}$. Cet ensemble est muni de la tribu engendrée par tous les sous-ensembles de la forme

$$\{(\omega_i)_{i\in\mathbf{N}} \mid \omega_0 \in A_0, ..., \omega_m \in A_m\}$$

avec $m \in \mathbf{N}$ et $A_i \in \{pile, face\}$ pour i allant de 0 à m. Si la pièce est bien équilibrée, on peut prendre comme probabilité le produit $P^{\otimes \mathbf{N}}$, où $P(\{face\}) = P(\{pile\}) = 1/2$.

1.4 Exercices

exercice 1
Soit $(A_n)_{n\in\mathbf{N}}$ une suite d'événements deux à deux disjoints sur un espace probabilisé $(\Omega, \mathcal{T}, P)$. Montrer que $\lim_{n\to\infty} P(A_n) = 0$.

exercice 2
Soit A_1, ..., A_n des événements tels que $P(A_k) > 1 - \frac{1}{n}$ pour tout k de 1 à n. Montrer qu'il existe un résultat $\omega \in \Omega$ qui appartient à tous les A_k.

Soit $(A_k)_{k\in\mathbf{N}}$ une suite d'événements. Montrer que si $P(A_k)$ ne converge pas vers 0, alors il existe un résultat qui appartient à une infinité de A_k.

exercice 3
Soit $c > 0$ et A_1, ..., $A_n \subset \Omega$ des événements tels que $P(A_i) \geq c$ pour tout i. Montrer qu'il existe deux indices j, k distincts tels que

$$P(A_j \cap A_k) \geq \frac{nc^2 - c}{n-1}.$$

Indication : s'intéresser à l'intégrale de $(\sum \mathbf{1}_{A_i})^2$ *et utiliser une inégalité fameuse.*

exercice 4
Soit P une mesure de probabilité définie sur la tribu des boréliens de $\mathbf{R}^d$ et $B(x, r)$ la boule euclidienne de rayon r centrée en $x \in \mathbf{R}^d$.

– Montrer que $P(B(0, r))$ converge vers 1 quand r tend vers l'infini.

– Soit $r_0 \geq 0$ et $x_n \in \mathbf{R}^d$ une suite telle que $\|x_n\| \xrightarrow[n\to\infty]{} \infty$. Montrer que

$$P(B(x_n, r_0)) \xrightarrow[n\to\infty]{} 0.$$

exercice 5
Parmi ces fonctions, quelles sont celles qui sont des densités de probabilité ?

$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$ $\quad f(x) = e^{-x}\ \mathbf{1}_{[0,\infty[}(x)$ $\quad f(x) = x^2\ \mathbf{1}_{[0,1]}(x)$

$f(x) = \frac{1}{2}e^{-|x|}$ $\quad f(x) = \sin(x)\ \mathbf{1}_{[0,3\pi/2]}(x)$ $\quad f(x) = (1 + \cos(x))\ \mathbf{1}_{[0,\pi]}(x)$

exercice 6
Pour tout borélien $A \subset \mathbf{R}$, on pose

$$\mu(A) = \frac{1}{2\sqrt{\pi}} \int_A e^{-x^2}\ dx + \mathbf{1}_A(0)/2$$

– Montrer que μ est une mesure de probabilité.
On vérifiera en particulier qu'elle est bien σ-additive.

– Soit $f : \mathbf{R} \to \mathbf{R}$ une fonction borélienne bornée. Donner une formule pour $\int f\, d\mu$.

Chapitre 2

Variables aléatoires

En pratique, on s'intéresse à certaines quantités numériques attachées aux résultats obtenus à l'issue de notre épreuve. Pour modéliser cela, on introduit la notion de variable aléatoire.

2.1 Définition d'une variable aléatoire

Définition 2 *Soit* $(\Omega, \mathcal{T}, P)$ *un espace probabilisé. Par définition, une variable aléatoire* $X : \Omega \to \mathbf{R}$ *est une fonction mesurable définie sur* Ω*, à valeurs réelles : pour tout intervalle* $I \subset \mathbf{R}$*, l'image réciproque* $X^{-1}(I)$ *de cet intervalle est dans* $\mathcal{T}$*.*

Pour $A \subset \mathbf{R}$ borélien, on pose

$$\begin{array}{rcl} X^{-1}(A) & = & \{\omega \in \Omega \mid X(\omega) \in A\} = (X \in A) \\ X^{-1}([a,b]) & = & (a \leq X \leq b) \\ X^{-1}([a,\infty[) & = & (a \leq X) = (X \geq a) \end{array}$$

On a alors

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

C'est la probabilité d'obtenir, à l'issue de l'épreuve, un résultat pour lequel la valeur de X est dans A. La quantité $P(X \in A)$ est bien définie dès que A est borélien car l'image réciproque d'un borélien par une application mesurable est mesurable (c'est-à-dire est dans $\mathcal{T}$).

2.2 Espérance et variance

Définition 3 *Une variable aléatoire* X *est dite* intégrable *si* $\int_\Omega |X|\, dP < +\infty$*. Dans ce cas l'intégrale de* X *est bien définie, c'est* l'espérance *de* X*.*

$$E(X) = \int_\Omega X\, dP.$$

La variable aléatoire X est dite de carré intégrable *si* $\int_\Omega X^2\,dP < +\infty$.
Dans ce cas X est intégrable et on définit la variance *de X par la formule*

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

On remarque qu'une variable aléatoire de carré intégrable est intégrable en intégrant l'inégalité $2X \leq 1 + X^2$. On peut aussi faire appel à l'inégalité de *Cauchy-Schwarz* : pour toutes variables aléatoires $X, Y : \Omega \to \mathbf{R}$,

$$\int_\Omega |XY|\,dP \leq \sqrt{\int_\Omega X^2\,dP}\,\sqrt{\int_\Omega Y^2\,dP}.$$

En prenant $Y = \mathbf{1}$ dans cette formule, on obtient la majoration $E(|X|) \leq \sqrt{E(X^2)}$.

Développons le carré qui apparaît dans la définition de la variance.

$$E\Big((X - E(X))^2\Big) = E\Big(X^2 - 2XE(X) + E(X)^2\Big) = E(X^2) - 2E(X)^2 + E(X)^2.$$

Nous avons obtenu la formule suivante, très utile pour calculer $V(X)$:

Proposition 1 *$V(X) = E(X^2) - E(X)^2$.*

Dans le cas discret, la variable aléatoire X est intégrable si $\sum_{\omega\in\Omega} p_\omega |X(\omega)| < +\infty$ et dans ce cas

$$E(X) = \sum_{\omega\in\Omega} p_\omega X(\omega).$$

Dans le cas continu, en notant f la densité de P, la variable aléatoire X est intégrable si $\int_{\mathbf{R}^d} |X(\omega)| f(\omega) d\omega < +\infty$ et dans ce cas

$$E(X) = \int_{\mathbf{R}^d} X(\omega) f(\omega) d\omega.$$

Propriétés :
Soit $\lambda \in \mathbf{R}$ et X, Y deux variables aléatoires intégrables.

– $E(\lambda X + Y) = \lambda E(X) + E(Y)$. (linéarité)

– Si $X \leq Y$, c'est-à-dire si pour tout $\omega \in \Omega$, $X(\omega) \leq Y(\omega)$, alors

$$E(X) \leq E(Y). \qquad \text{(monotonie)}$$

– Pour tout événement $A \in \mathcal{T}$, $P(A) = E(\mathbf{1}_A)$.

– Soit (X_n) une suite de variables aléatoires qui converge de manière croissante vers X : pour presque tout $\omega \in \Omega$, $(X_n(\omega))_{n\in\mathbf{N}}$ est croissante et $X_n(\omega) \to X(\omega)$. Alors

$$E(X_n) \xrightarrow[n\to\infty]{} E(X).$$

– Soit (X_n) une suite de variable aléatoires qui converge vers X presque partout. On suppose qu'il existe Y intégrable telle que $|X_n| \leq Y$ pour tout $n \in \mathbf{N}$. Alors

$$E(X_n) \xrightarrow[n\to\infty]{} E(X).$$

– $V(\lambda X) = \lambda^2 V(X)$.

– $V(X+Y) = V(X) + V(Y) + 2Cov(X,Y)$
en notant

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

la covariance de X, Y qui est bien définie dès que X, Y sont de carrés intégrables.

Ces propriétés sont des conséquences immédiates des définitions. Les deux théorèmes de passage à la limite découlent du théorème de convergence croissante et du théorème de convergence dominée.

2.3 Inégalités

On s'intéresse maintenant à deux inégalités classiques qui donnent quelques informations sur la manière dont les valeurs d'une variable aléatoire se répartissent.

Théorème 2 (Inégalité de Markov) *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé, $Y : \Omega \to \mathbf{R}_+$ une variable aléatoire positive. Alors, pour tout $\lambda > 0$,*

$$P(Y \geq \lambda) \leq \frac{E(Y)}{\lambda}.$$

Preuve
On a l'inégalité $\lambda \mathbf{1}_{(Y\geq\lambda)} \leq Y$ ce qui donne, par monotonie,

$$E(\lambda \mathbf{1}_{(Y\geq\lambda)}) \leq E(Y).$$

On conclut en remarquant que

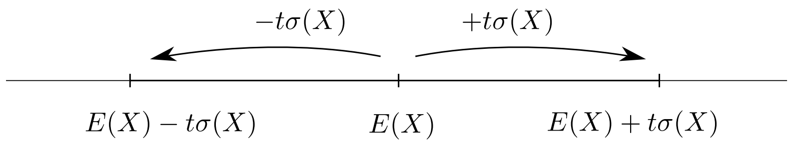
$$E(\lambda \mathbf{1}_{(Y\geq\lambda)}) = \lambda\, E(\mathbf{1}_{(Y\geq\lambda)}) = \lambda\, P(Y \geq \lambda).$$

Théorème 3 (Inégalité de Bienaymé-Tchebichev) *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé, $X : \Omega \to \mathbf{R}$ une variable aléatoire de carré intégrable. Alors, pour tout $t > 0$,*

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

Cette inégalité peut se récrire à l'aide de *l'écart-type* $\sigma(X) = \sqrt{V(X)}$ comme suit :

$$P\Big(X \notin \,]\, E(X) - t\sigma(X), E(X) + t\sigma(X)\,[\Big) \;\leq \frac{1}{t^2}$$



Application
Si X est de carré intégrable, la probabilité d'obtenir à l'issue de l'épreuve une valeur à plus de 10 fois l'écart-type de l'espérance est inférieure à $1/100$.

Preuve
L'égalité de Bienaymé-Tchebichev se déduit de l'inégalité de Markov en prenant $Y = (X - E(X))^2$ et $\lambda = t^2$ dans cette inégalité. On a alors

$$P(Y \geq \lambda) = P((X - E(X))^2 \geq t^2) = P(|X - E(X)| \geq t),$$

$$\frac{E(Y)}{\lambda} = \frac{E((X - E(X))^2)}{t^2} = \frac{V(X)}{t^2}.$$

La formule est démontrée.

2.4 Loi d'une variable aléatoire

À chaque variable aléatoire X définie sur un espace probabilisé $(\Omega, \mathcal{T}, P)$, on peut associer une probabilité P_X qui rend compte de la distribution de ses valeurs, en procédant de la façon suivante.

Définition 4 *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé, $X : \Omega \to \mathbf{R}$ une variable aléatoire. La loi de X est la probabilité définie sur la tribu des boréliens de $\mathbf{R}$ par la formule :*

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

pour tout $A \subset \mathbf{R}$ borélien.

La variable aléatoire X est dite *discrète* si sa loi P_X est discrète : il existe un ensemble fini ou dénombrable $\mathcal{D} \subset \mathbf{R}$ tel que $P_X(\mathcal{D}) = 1$. Indiçons ses éléments par un ensemble $I \subset \mathbf{N}$: $\mathcal{D} = \{x_i\}_{x \in I}$. On est presque sûr d'obtenir un résultat qui se trouve dans cet ensemble de valeurs $\{x_i\}_{i \in I}$ et on peut écrire

$$P_X = \sum_{i \in I} p_{x_i}\, \delta_{x_i}$$

où p_{x_i} est la probabilité d'obtenir la valeur x_i : $P(X = x_i) = p_{x_i}$.

La variable aléatoire X est dite *continue* si P_X est une loi continue, auquel cas sa densité est notée f_X. C'est une fonction borélienne positive dont l'intégrale vaut un. On a alors

$$P_X(A) = P(X \in A) = \int_A f_X(x)\, dx$$

pour tout $A \subset \mathbf{R}$ borélien. Dans ce cas, la probabilité $P(X = x)$ est bien sûr nulle pour tout $x \in \mathbf{R}$.

L'espérance et la variance d'une variable aléatoire peuvent s'exprimer en fonction de sa loi uniquement. En conséquence, deux variables qui ont même loi ont même espérance et même variance.

Proposition 2 *Si X est intégrable,*

$$E(X) = \int_{\mathbf{R}} x \ dP_X(x).$$

Si X est de carré intégrable,

$$V(X) = \int_{\mathbf{R}} (x - E(X))^2 \ dP_X(x).$$

Cette proposition se déduit de la *formule de transfert.*

Proposition 3 (formule de transfert) *Soit $g : \mathbf{R} \to \mathbf{R}$ borélienne, positive ou P_X-intégrable. Alors*

$$\int_\Omega g(X) \ dP = \int_{\mathbf{R}} g(x) \ dP_X(x).$$

Preuve de la formule de transfert

– C'est vrai pour $g = \mathbf{1}_A$, A borélien de $\mathbf{R}$:

$$\int \mathbf{1}_A(X) \ dP = P(X \in A) = P(X^{-1}(A)),$$

$$\int \mathbf{1}_A(x) \ dP_X(x) = P_X(A) = P(X \in A).$$

– C'est vrai pour les combinaisons linéaires de fonctions indicatrices $g = \sum c_i \mathbf{1}_{A_i}$ par linéarité de l'intégrale.

– Une combinaison linéaire de fonctions indicatrices s'appelle une *fonction étagée.* Toute fonction positive mesurable peut être approchée de manière croissante par une suite de fonctions étagées. Pour $g \geq 0$, borélienne, on prend $g_n \to g$, g_n étagées, et on passe à la limite

$$\int g_n(X) \ dP \xrightarrow[n\to\infty]{} \int g(X) \ dP$$

$$\int g_n(x)\ dP_X(x) \xrightarrow[n\to\infty]{} \int g(x)\ dP_X(x)$$

en appliquant le théorème de convergence croissante, si bien que $\int g(X)\ dP = \int g(x)\ dP_X(x)$.

– Pour g intégrable, on écrit g comme la différence de deux fonctions positives intégrables et on utilise la linéarité de l'intégrale pour conclure.

Exemple
Si X est variable aléatoire obéissant à une loi exponentielle de paramètre $l > 0$, P_X est associée à la densité $f_X(x) = le^{-lx}\mathbf{1}_{\mathbf{R}_+}(x)$ et on a :

$$P_X([a,b]) = P(X \in [a,b]) = \int_{[a,b]} f_X(x)\,dx = \int_a^b le^{lx}dx$$

dès que $0 \leq a \leq b$.

Il est parfois plus pratique de travailler avec des fonctions plutôt qu'avec des lois de probabilité. Ceci nous amène à la notion de fonction de répartition.

Définition 5 *La fonction de répartition de X est définie par*

$$F_X(x) = P(X \leq x).$$

On a alors l'égalité, pour tout $a, b \in \mathbf{R}$,

$$P(X \in]a,b]) = F_X(b) - F_X(a).$$

Comme une mesure de probabilité définie sur la tribu des boréliens de $\mathbf{R}$ est uniquement déterminée par ses valeurs sur les intervalles, la fonction de répartition caractérise la loi de X de manière unique : si deux variables aléatoires ont même fonction de répartition, elles ont même loi.

$$F_X = F_Y \Leftrightarrow P_X = P_Y.$$

La fonction de répartition possède les propriétés suivantes :

– elle est croissante, à valeur dans l'intervalle $[0,1]$,

– $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to+\infty} F_X(x) = 1$,

– elle est continue à droite et possède une limite à gauche en tout point,

– l'ensemble des points de discontinuité de F_X est composé des $x \in \mathbf{R}$ tels que $P(X = x) > 0$, il est donc dénombrable.

2.5 Loi d'un multiplet de variables aléatoires

Les considérations précédentes se généralisent à des couples et des multiplets de variables aléatoires. Soient X_1, ...X_n des variables aléatoires à valeurs réelles. On peut considérer ces variables comme une unique variable aléatoire à valeurs dans $\mathbf{R}^n$.

$$(X_1, X_2, ..., X_n) : \Omega \to \mathbf{R}^n.$$

On pose, pour A borélien de $\mathbf{R}^n$ et A_1, ..., A_n des boréliens de $\mathbf{R}$,

$$((X_1, ..., X_n) \in A) = \{\omega \in \Omega \mid (X_1(\omega), X_2(\omega), ..., X_n(\omega)) \in A\},$$

$$\begin{array}{rcl}(X_1 \in A_1, ..., X_n \in A_n) & = & ((X_1, ..., X_n) \in A_1 \times ... \times A_n) \\ & = & \{\omega \in \Omega \mid X_1(\omega) \in A_1, ..., X_n(\omega) \in A_n\} \\ & = & \bigcap\limits_{1\leq i\leq n} (X_i \in A_i).\end{array}$$

Définition 6 *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé, X_1, ..., X_n des variables aléatoires. La loi du multiplet $(X_1, ..., X_n)$ est la mesure de probabilité définie sur la tribu des boréliens de $\mathbf{R}^n$ par la formule :*

$$P_{(X_1,...,X_n)}(A) = P((X_1, ..., X_n) \in A) = P(\{\omega \in \Omega \mid (X_1(\omega), ..., X_n(\omega)) \in A\})$$

pour tout $A \subset \mathbf{R}^n$ borélien.

La loi du multiplet est *discrète* si la loi de $P_{(X_1,...,X_n)}$ est discrète : il existe un ensemble fini ou dénombrable $\mathcal{D} \subset \mathbf{R}^n$ tel que $P_{(X_1,...,X_n)}(\mathcal{D}) = 1$. Elle est dite *continue* si $P_{(X_1,...,X_n)}$ est une loi continue, auquel cas sa densité est notée $f_{X_1,...,X_n}$. Cette densité est une fonction borélienne, définie de $\mathbf{R}^n$ dans $\mathbf{R}_+$, positive, d'intégrale 1, et nous avons la relation

$$P((X_1, ..., X_n) \in A) = P_{(X_1,...,X_n)}(A) = \int_A f_{X_1,...,X_n}(x_1, ..., x_n)\ dx_1...dx_n.$$

La formule de transfert se généralise à n variables.

Proposition 4 (Formule de transfert) *Soit $g : \mathbf{R}^n \to \mathbf{R}$ borélienne, positive ou $P_{(X_1,...,X_n)}$-intégrable. Alors*

$$\int_\Omega g(X_1, ..., X_n)\ dP = \int_{\mathbf{R}^n} g(x_1, ..., x_n)\ dP_{X_1,...,X_n}(x_1, ..., x_n).$$

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

Les lois individuelles des X_i peuvent se déduire de la loi du multiplet en remarquant que

$$\begin{array}{rcl}P(X_i \in A) & = & P(X_1 \in \Omega, ..., X_{i-1} \in \Omega, X_i \in A, X_{i+1} \in \Omega, ..., X_n \in \Omega) \\ & = & P_{(X_1,...,X_n)}(\Omega \times ... \times \Omega \times A \times \Omega \times ... \times \Omega).\end{array}$$

On dit que les lois P_{X_i} sont les *lois marginales* de la distribution $(X_1, ..., X_n)$.

Dans le cas continu, la densité des X_i se déduisent de celle du multiplet grâce à la formule suivante :

$$P_{X_i}(I) = \int_I \Big(\int_{\mathbf{R}^{n-1}} f_{X_1,\ldots,X_n}(x_1, ..., x_n)\ dx_1...dx_{i-1}dx_{i+1}...dx_n) \Big) dx_i$$

où I est un intervalle ou un borélien de $\mathbf{R}$, si bien que

$$f_{X_i}(x_i) = \int_{\mathbf{R}^{n-1}} f_{X_1,\ldots,X_n}(x_1, ..., x_n)\ dx_1...dx_{i-1}\, dx_{i+1}...dx_n.$$

2.6 Exercices

exercice 1
Soit X une variable aléatoire strictement positive telle que X et $1/X$ sont intégrables. Montrer que $E(X)E(1/X) \geq 1$.

exercice 2
Soit X une variable aléatoire de carré intégrable. Montrer que pour tout $c \in \mathbf{R}$,

$$V(X) \leq E((X-c)^2).$$

Soit a, $b \in \mathbf{R}$. Montrer que

$$P(X \geq a) \leq \frac{E((X+b)^2)}{(a+b)^2}$$

En déduire l'inégalité de Cantelli, valide pour tout $t > 0$:

$$P\Big(X - E(X) \geq t\,\sigma(X)\Big) \leq \frac{1}{1+t^2}.$$

exercice 3
Soit X une variable aléatoire telle que $0 \leq X \leq 1$. Montrer que $V(X) \leq 1/4$.
À quelle condition a-t-on égalité ?

exercice 4
Soit X une variable aléatoire satisfaisant $P(X = x) = 0$ pour tout $x \in \mathbf{R}$ et F_X sa fonction de répartition. Montrer que la variable aléatoire $F_X(X)$ obéit à la loi uniforme sur $]0,1[$.

Étant donnée une variable aléatoire Y de loi uniforme sur $]0,1[$ et ν une mesure de probabilité sur la tribu des boréliens de $\mathbf{R}$, construire une variable aléatoire X de loi ν en composant Y par une fonction bien choisie.

exercice 5
Soit X une variable aléatoire intégrable. Montrer les assertions suivantes.
– $\displaystyle\lim_{N\to\infty} \int_{(|X|>N)} |X|\,dP = 0$.
– Pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que pour tout $A \in \mathcal{T}$,

$$P(A) \leq \delta \text{ implique } \int_A |X|\,dP \leq \varepsilon.$$

exercice 6
Soit X une variable aléatoire réelle positive, de fonction de répartition F_X. Montrer que pour $0 < p < \infty$,

$$E(X^p) = \int_0^\infty p\,t^{p-1}\,P(X > t)\,dt = \int_0^\infty p\,t^{p-1}\,(1 - F_X(t))\,dt.$$

Soit $X_1, X_2, ...X_n$ des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. Calculez $E(min(X_1, X_2, ...X_n))$ et $E(max(X_1, X_2, ...X_n))$.

Chapitre 3

Indépendance

On a vu comment modéliser une épreuve répétée un nombre fini ou infini de fois de manière indépendante, en prenant pour univers un espace produit et pour probabilité une probabilité produit. On va préciser cette notion d'indépendance en l'appliquant à des événements, des tribus ou des variables aléatoires.

3.1 Indépendance d'événements et de variables aléatoires

On commence par définir la notion d'événements indépendants.

Définition 7 *Soit* $(\Omega, \mathcal{T}, P)$ *un espace probabilisé. Deux événements* $A, B \in \mathcal{T}$ *sont dits indépendants entre eux si*

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

Soit $(A_i)_{i \in I}$ *une famille d'événements. Ces événements sont dits indépendants dans leur ensemble si*

$$\forall S \subset I \text{ fini, } P\Big(\bigcap_{i \in S} A_i\Big) = \prod_{i \in S} P(A_i).$$

Exemple
Pour une famille de trois événements $\{A_1, A_2, A_3\}$, $I = \{1, 2, 3\}$,

$S = \{1\}$	$P(A_1) = P(A_1)$
$S = \{2\}$	$P(A_2) = P(A_2)$
$S = \{3\}$	$P(A_3) = P(A_3)$
$S = \{1, 2\}$	$P(A_1 \cap A_2) = P(A_1)P(A_2)$
$S = \{1, 3\}$	$P(A_1 \cap A_3) = P(A_1)P(A_3)$
$S = \{2, 3\}$	$P(A_2 \cap A_3) = P(A_2)P(A_3)$
$S = \{1, 2, 3\}$	$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$

sont les conditions à vérifier pour avoir l'indépendance de A_1, A_2 et A_3 dans leur ensemble.

Définition 8 *Deux variables aléatoires $X, Y : \Omega \to \mathbf{R}$ sont indépendantes entre elles si pour tous boréliens $A, B \subset \mathbf{R}$, les événements $(X \in A)$ et $(Y \in B)$ sont indépendants entre eux :*

$$P\Big((X \in A) \cap (Y \in B)\Big) = P(X \in A)\, P(Y \in B).$$

Soit $(X_i)_{i\in I}$ une famille de variables aléatoires. Elle sont dites indépendantes entre elles si pour tout sous-ensemble $S \subset I$ fini et $(A_i)_{i\in S}$ des boréliens de $\mathbf{R}$, les événements $(X_i \in A_i)$ sont indépendants dans leur ensemble :

$$P\Big(\bigcap_{i\in S}(X_i \in A_i)\Big) = \prod_{i\in S} P(X_i \in A_i).$$

Introduisons les notations suivantes pour alléger les formules :

$$\begin{aligned}(X \in A, Y \in B) &= (X \in A) \cap (Y \in B)\\ &= \{\omega \in \Omega \mid X(\omega) \in A \text{ et } Y(\omega) \in B\}\end{aligned}$$

$$\Big(X_i \in A_i, i \in S\Big) = \bigcap_{i\in S}\, (X_i \in A_i)$$

$$(X_1 \in A_1, X_2 \in A_2, ..., X_n \in A_n) = (X_1 \in A_1) \cap (X_2 \in A_2) ... \cap (X_n \in A_n)$$

Définition 9 *Deux tribus $\mathcal{T}_1 \subset \mathcal{T}$, $\mathcal{T}_2 \subset \mathcal{T}$ sont indépendantes entre elles si pour tout $A \in \mathcal{T}_1$ et $B \in \mathcal{T}_2$, A et B sont indépendants entre eux.*
Soit $(\mathcal{T}_i)_{i\in I}$ une famille de tribus incluses dans $\mathcal{T}$. Elle sont dites indépendantes entre elles si pour tout sous-ensemble $S \subset I$ fini et toute famille $(A_i)_{i\in S}$ satisfaisant $A_i \in \mathcal{T}_i$ pour tout $i \in S$, les événements A_i sont indépendants dans leur ensemble :

$$P\Big(\bigcap_{i\in S} A_i\Big) = \prod_{i\in S} P(A_i).$$

3.2 Lemme de Borel-Cantelli

Voici une première application de la notion d'indépendance d'événements.

Lemme 1 (Borel-Cantelli) *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et $(A_i)_{i\in\mathbf{N}}$ une suite d'événements.*

Si $\displaystyle\sum_{i\in\mathbf{N}} P(A_i) < +\infty$, presque tout $\omega \in \Omega$ n'appartient qu'à un nombre fini de A_i.

Si $\displaystyle\sum_{i\in\mathbf{N}} P(A_i) = +\infty$, et si les A_i sont indépendants dans leur ensemble, alors presque tout $\omega \in \Omega$ appartient à une infinité de A_i.

On définit la limite supérieure de la suite d'ensembles A_i comme suit :

$$\overline{\lim_{i\in\mathbf{N}}}\, A_i = \bigcap_{N\in\mathbf{N}} \Big(\bigcup_{i\geq N} A_i\Big) = \{\omega \in \Omega \mid \omega \text{ appartient à une infinité de } A_i\}.$$

Le lemme se reformule alors de la façon suivante :

$$\sum_{i\in\mathbf{N}} P(A_i) < +\infty \text{ implique } P\Big(\overline{\lim_{i\in\mathbf{N}}}\ A_i\Big) = 0.$$

$$\sum_{i\in\mathbf{N}} P(A_i) = +\infty \text{ et } (A_i)_{i\in\mathbf{N}} \text{ indépendants implique } P\Big(\overline{\lim_{i\in\mathbf{N}}}\ A_i\Big) = 1.$$

Preuve du lemme
Nous avons la relation $\#\{i \in \mathbf{N} \mid \omega \in A_i\} = \sum_{i\in\mathbf{N}} \mathbf{1}_{A_i}(\omega)$. Intégrons cette égalité :

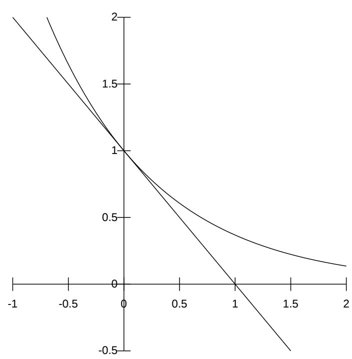
$$\int \#\{i \in \mathbf{N} \mid \omega \in A_i\}\ dP(\omega) = \int \sum_{i\in\mathbf{N}} \mathbf{1}_{A_i}\ dP = \sum_{i\in\mathbf{N}} \int \mathbf{1}_{A_i}\ dP = \sum_{i\in\mathbf{N}} P(A_i) < +\infty.$$

La fonction $\omega \mapsto \#\{i \in \mathbf{N} \mid \omega \in A_i\}$ est intégrable, donc finie presque partout ; pour presque tout $\omega \in \Omega$, $\#\{i \in \mathbf{N} \mid \omega \in A_i\} < +\infty$.

Supposons à présent les (A_i) indépendants et $M, N \in \mathbf{N}$, $N \leq M$.

$$P\Big(\bigcap_{i=N}^{M} A_i^c\Big) = \prod_{i=N}^{M} P(A_i^c) = \prod_{i=N}^{M} (1 - P(A_i)) \leq e^{-\sum_{i=N}^{M} P(A_i)}$$

d'après la majoration $1 - x \leq e^{-x}$, valide pour tout $x \in \mathbf{R}$. Nous avons donc



$$P\Big(\bigcap_{i=N}^{M} A_i^c\Big) \leq e^{-\sum_{i=N}^{M} P(A_i)}$$

et en passant à la limite sur M,

$$P\Big(\bigcap_{i\geq N} A_i^c\Big) = 0.$$

Ceci entraîne $P(\bigcup_{N\in\mathbf{N}} \bigcap_{i\geq N} A_i^c) = 0$ puis en passant au complémentaire,

$$P(\overline{\lim}\ A_i) = P\Big(\bigcap_{N\in\mathbf{N}} \bigcup_{i\geq N} A_i\Big) = 1.$$

3.3 Loi d'un multiplet de variables indépendantes

Calculons l'espérance d'un produit de variables aléatoires indépendantes.

Proposition 5 *Soit* $(\Omega, \mathcal{T}, P)$ *un espace probabilisé,* $X, Y : \Omega \to \mathbf{R}$ *deux variables aléatoires. On se donne* $f, g : \mathbf{R} \to \mathbf{R}$ *des fonctions boréliennes telles que*

$f(X)$ et $g(Y)$ soient intégrables. On suppose X et Y indépendantes entre elles. Alors

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

Ceci se généralise à un nombre quelconque de variables aléatoires $(X_i)_{i=1\ldots n}$ indépendantes entre elles :

$$E\Big(\prod_{i=1}^{n} f_i(X_i)\Big) = \prod_{i=1}^{n} E(f_i(X_i))$$

où les $f_i : \mathbf{R} \to \mathbf{R}$ sont des fonctions boréliennes telles que les $f_i(X_i)$ sont intégrables.

Preuve
Si f et g sont des fonctions indicatrices, $f = \mathbf{1}_A$, $g = \mathbf{1}_B$,

$$\begin{aligned}
E(f(X)g(Y)) &= E(\mathbf{1}_A(X)\mathbf{1}_B(Y)) = E(\mathbf{1}_{(X\in A)}\mathbf{1}_{(Y\in B)})\\
&= E(\mathbf{1}_{(X\in A)\cap(Y\in B)}) = E(\mathbf{1}_{(X\in A,\ Y\in B)})\\
&= P(X \in A, Y \in B)\\
&= P(X \in A)P(Y \in B) \quad \text{par indépendance,}\\
&= E(\mathbf{1}_A(X))E(\mathbf{1}_B(Y))\\
&= E(f(X))E(g(Y)).
\end{aligned}$$

On procède ensuite comme pour la preuve de la formule de transfert : on vérifie la formule pour les fonctions étagées, par linéarité, puis on vérifie la formule pour $f, g \geq 0$ en les approchant de manière croissante par des fonctions étagées, et enfin pourf, g intégrables en les écrivant comme différence de fonctions positives intégrables.

Le cas d'un nombre quelconque de variables indépendantes s'en déduit de la même façon. Rappelons que la covariance de deux variables aléatoires est égale à l'espérance du produit des variables moins le produit des espérances. On obtient le corollaire suivant :

Corollaire 1 *Soit X_1,..., X_n des variables aléatoires de carré intégrable, indépendantes entre elles. Alors*

$$Cov(X_i, X_j) = 0 \text{ si } i \neq j,$$

$$V\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} V(X_i).$$

Complément
On montre que la loi d'un couple ou d'un multiplet de variables aléatoires indépendantes entre elles est égale au produit des lois de chacune des variables aléatoires.

Proposition 6 *Soit X, Y deux variables aléatoires indépendantes entre elles. Alors*

$$P_{(X,Y)} = P_X \otimes P_Y,$$

$$E(h(X,Y)) = \int_\Omega h(X,Y)\, dP = \int_{\mathbf{R}^2} h(x,y)\, dP_X(x)\, dP_Y(y)$$

pour toute fonction $h : \mathbf{R}^2 \to \mathbf{R}$ borélienne, positive ou $P_{(X,Y)}$-intégrable.
Soit X_1,..., X_n des variables aléatoires indépendantes entre elles. Alors

$$P_{(X_1,\ldots,X_n)} = P_{X_1} \otimes P_{X_2} \otimes \ldots \otimes P_{X_n},$$

$$E(h(X_1,\ldots,X_n)) = \int_{\mathbf{R}^n} h(x_1,\ldots,x_n)\, dP_{X_1}(x_1)\ldots dP_{X_n}(x_n)$$

pour toute fonction $h : \mathbf{R}^n \to \mathbf{R}$ borélienne, positive ou $P_{(X_1,\ldots,X_n)}$-intégrable.

La preuve se ramène à celle de la proposition précédente en utilisant le fait que toute fonction borélienne bornée $h : \mathbf{R}^2 \to \mathbf{R}$ peut s'approcher en norme $L^1(\mathbf{R}^2, P_{(X,Y)} + P_X \otimes P_Y)$ par une combinaison linéaire de fonctions de la forme $(x,y) \mapsto f(x)g(y)$, avec f et g boréliennes bornées. On généralise ensuite aux fonctions boréliennes positives en utilisant le théorème de convergence croissante puis aux fonctions intégrables. Le raisonnement est le même pour un multiplet de variables aléatoires.

3.4 Exercices

exercice 1
Soit X une variable aléatoire intégrable. Montrer que si X est indépendante d'elle-même, elle est constante p.s. : il existe un réel C tel que $P(X = C) = 1$.

exercice 2
Soit A_n, $n \in \mathbf{N}$, des événements. Montrer que

$$\overline{\lim}\ \mathbf{1}_{A_n} = \mathbf{1}_{\overline{\lim} A_n}, \quad \overline{\lim}\ P(A_n) \leq P\Big(\overline{\lim} A_n\Big).$$

exercice 3
Soit X une variable aléatoire ; la tribu associée à X est définie par

$$\mathcal{T}_X = \{X^{-1}(A) \mid A \subset \mathbf{R} \text{ borélien}\}.$$

Montrer que deux variables aléatoires X et Y sont indépendantes si et seulement si $\mathcal{T}_X$ et $\mathcal{T}_Y$ sont indépendantes.

exercice 4
Montrer que si X et Y sont deux variables aléatoires indépendantes de densités f_X et f_Y, le couple (X, Y) admet pour densité la fonction $(x, y) \mapsto f_X(x) f_Y(y)$.

exercice 5
Soient X et Y deux variables aléatoires indépendantes. On suppose que X admet une densité f_X. Montrer que la somme $X + Y$ admet aussi une densité qui est égale à

$$f_{X+Y}(t) = \int_{\mathbf{R}} f_X(t - y)\, dP_Y(y).$$

Montrer que si X et Y admettent des densités f_X et f_Y, la densité de $X + Y$ est égale à la convolée de f_X et f_Y : $f_{X+Y}(t) = \int_{\mathbf{R}} f_X(t - y)\, f_Y(y)\, dy$.

exercice 6
Soit $X_1, ...X_n$ des variables aléatoires centrées, indépendantes entre elles, de cubes intégrables. On pose $S_n = \sum_{k=1}^n X_k$. Montrer que $E(S_n^3) = \sum_{k=1}^n X_k^3$.

exercice 7
Soient $(X_i)_{i \in \mathbf{N}}$ une suite de variables aléatoires indépendantes. On suppose que $P_{X_i} = P_{X_j}$ pour $i, j \in \mathbf{N}$ et $P(X_k = x) = 0$ pour $x \in \mathbf{R}$ et $k \in \mathbf{N}$. On pose

$$A_n = \{\omega \in \Omega \mid \forall\, k < n,\ X_k(\omega) < X_n(\omega)\}.$$

Si A_n est réalisé, on dit qu'on a obtenu un record à l'étape n.
– Montrer que $P(X_i = X_j) = 0$ si $i \neq j$.
– Montrer que les A_n sont indépendants entre eux et que $P(A_n) = \frac{1}{n}$.
– En déduire que presque sûrement, on observe une infinité de records.

Chapitre 4

Loi des grands nombres

On va s'intéresser au comportement asymptotique d'une suite de variables aléatoires. On se donne un espace probabilisé $(\Omega, \mathcal{T}, P)$ et pour chaque entier $n \in \mathbf{N}$ une variable aléatoire $X_n : \Omega \to \mathbf{R}$.

Définition 10 *La suite de variables aléatoires* $(X_i)_{i \in \mathbf{N}}$ *est dite* identiquement distribuée *si tous les* X_i *ont même loi :*

$$\forall i, j \in \mathbf{N},\ P_{X_i} = P_{X_j}.$$

En d'autres termes, pour tout borélien $A \subset \mathbf{R}$,

$$P(X_i \in A) = P(X_j \in A)$$

et pour toute $f : \mathbf{R} \to \mathbf{R}$ borélienne positive ou intégrable par rapport à P_{X_0},

$$E(f(X_i)) = E(f(X_j)).$$

En particulier, $E(X_i) = E(X_j)$ si les X_i sont intégrables, $E(X_i^2) = E(X_j^2)$ et $V(X_i) = V(X_j)$ si les X_i sont de carrés intégrables.

4.1 Loi faible des grands nombres

Soit $(X_i)_{i \in \mathbf{N}}$ une suite de variables aléatoires indépendantes entre elles, identiquement distribuées (v.a i.i.d). On pose

$$S_n = \sum_{i=1}^{n} X_i = X_1 + X_2 + ... + X_n.$$

Pour $\omega \in \Omega$, la quantité $\frac{S_n}{n}(\omega) = \frac{X_1(\omega)+X_2(\omega)+...+X_n(\omega)}{n}$ est la *moyenne empirique* calculée sur l'échantillon donné par le résultat $\omega \in \Omega$. On cherche à étudier le comportement asymptotique de la moyenne $\frac{S_n}{n}$.

Théorème 4 (loi faible des grands nombres) *Soit* $(X_i)_{i\in\mathbf{N}}$ *une suite de variables aléatoires indépendantes entre elles, identiquement distribuées, de carrés intégrables. Alors pour tout* $\varepsilon > 0$,

$$P\Big(\Big|\frac{S_n}{n} - E(X_0)\Big| > \varepsilon\Big) \xrightarrow[n\to\infty]{} 0.$$

La preuve du théorème repose sur le lemme suivant :

Lemme 2 *Soit* $(X_i)_{i\in\mathbf{N}}$ *une suite de variables aléatoires indépendantes, identiquement distribuées. Alors* $E(S_n) = nE(X_0), V(S_n) = nV(X_0)$.

Preuve du lemme
$E(S_n) = E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$ par linéarité.

$V(S_n) = V(X_1 + ... + X_n) = V(X_1) + ... + V(X_n) + 2\sum_{i<j} Cov(X_i, X_j)$,
où $Cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j)$.

La covariance de deux variables aléatoires est nulle dans le cas indépendant. D'où $V(S_n) = V(X_1) + ... + V(X_n) = n\,V(X_0)$.

Preuve du théorème
D'après le lemme,

$$E(S_n/n) = E(X_0), \quad V(S_n/n) = V(X_0)/n, \quad \sigma(S_n/n) = \sigma(X_0)/\sqrt{n}.$$

On applique alors l'inégalité de Bienaymé-Tchebichev :

$$P\Big(\Big|\frac{S_n}{n} - E(\frac{S_n}{n})\Big| > \varepsilon\Big) \leq \frac{V(S_n/n)}{\varepsilon^2} = \frac{\sigma(X_0)^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

Remarque On peut montrer que la loi faible des grands nombres est encore vraie pour des variables aléatoires indépendantes, identiquement distribuées, intégrables.

4.2 Loi forte des grands nombres

Théorème 5 (loi forte des grands nombres) *Soit* $(X_i)_{i\in\mathbf{N}}$ *une suite de variables aléatoires intégrables, indépendantes entre elles, identiquement distribuées et* $S_n = \sum_{i=1}^n X_i$. *Alors pour presque tout* $\omega \in \Omega$,

$$\frac{S_n}{n}(\omega) \xrightarrow[n\to\infty]{} E(X_0).$$

En d'autres termes, l'ensemble $\{\omega \in \Omega \mid \frac{S_n}{n}(\omega) \xrightarrow[n\to\infty]{} E(X_0)\}$ est un ensemble dont la probabilité vaut 1.

Vocabulaire On dit qu'une propriété est vraie *presque sûrement* si elle est satisfaite pour presque tout $\omega \in \Omega$. Nous utiliserons dans la suite l'abréviation *p.s.* pour le terme *presque sûrement.*

Énonçons un premier corollaire de la loi forte des grands nombres, qui sera démontré dans la suite. Ce corollaire montre que la probabilité d'un événement est presque sûrement égale à la limite du nombre de fois où il est réalisé sur le nombre total de fois où l'épreuve est répétée, lorsque le nombre de répétitions tend vers l'infini.

Corollaire 2 *Soit $(X_i)_{i\in\mathbf{N}}$ une suite de variables aléatoires indépendantes identiquement distribuées et soit A un borélien de $\mathbf{R}$. Alors*

$$\frac{\#\{i \leq n \mid X_i(\omega) \in A\}}{n} \xrightarrow[n\to\infty]{} P(X_0 \in A) \quad \textit{presque sûrement.}$$

Illustrons la loi forte des grands nombres sur un exemple avant de la démontrer.

Exemple
On lance une pièce de monnaie bien équilibrée un grand nombre de fois de manière indépendante. Pour modéliser ces épreuves, on commence par considérer la probabilité $\tilde{P}$ définie sur $\mathcal{P}(\{pile, face\})$ par $\tilde{P}(\{face\}) = \tilde{P}(\{pile\}) = 1/2$ et on pose :

- $\Omega = \{pile, face\}^{\mathbf{N}}$,
- $\mathcal{T} = \mathcal{P}(\{pile, face\})^{\otimes\mathbf{N}}$,
- $P = \tilde{P}^{\otimes\mathbf{N}}$.

Les éléments de Ω sont des suites infinies de pile ou face.

On définit maintenant une variable aléatoire $X : \{pile, face\} \to \mathbf{R}$ par $X(pile) = 0$, $X(face) = 1$ et on pose pour tout $i \in \mathbf{N}$:

$$X_i((\omega_k)_{k\in\mathbf{N}}) = X(\omega_i) = \begin{cases} 1 & si \quad \omega_i = pile \\ 0 & si \quad \omega_i = face \end{cases}$$

Soit $\omega = (\omega_k)_{k\in\mathbf{N}} \in \Omega$. L'élément ω_k de la suite ω est le résultat obtenu au $k^{\text{ième}}$ lancer. La quantité $X_k(\omega)$ vaut 1 si ce résultat est face, 0 si il est égal à pile. Définissons également

$$\frac{S_n}{n}(\omega) = \frac{X_1(\omega) + ... + X_n(\omega)}{n} = \frac{X(\omega_1) + X(\omega_2) + ... + X(\omega_n)}{n}.$$

C'est la moyenne des valeurs prises par X au cours des n premières épreuves. C'est le nombre moyen de fois où Face a été obtenu au cours des n premiers lancers.

Proposition 7 *Les variables aléatoires X_i sont indépendantes dans leur ensemble, identiquement distribuées, intégrables.*

Preuve

$$\begin{array}{rcl} P(X_0 \in A_0, ..., X_n \in A_n) & = & \tilde{P}^{\otimes \mathbf{N}}(\{\omega \in \Omega \mid X(\omega_0) \in A_0, ..., X(\omega_n) \in A_n\}) \\ & = & \tilde{P}^{\otimes \mathbf{N}}(C_{X^{-1}(A_0),...,X^{-1}(A_n)}) \\ & = & \displaystyle\prod_{i=0}^{n} \tilde{P}(X^{-1}(A_i)). \end{array}$$

De plus, $P(X_0 \in A_0) = P(C_{X^{-1}(A_0)}) = \tilde{P}(X^{-1}(A_0))$. Nous avons également

$$\begin{array}{rcl} P(X_i \in A_i) & = & P(\{\omega \in \Omega \mid \omega_i \in X^{-1}(A_i)\}) \\ & = & P(C_{\Omega,...,\Omega,X^{-1}(A_i)}) \\ & = & \tilde{P}^{\otimes \mathbf{N}}(C_{\Omega,...,\Omega,X^{-1}(A_i)}) \\ & = & \tilde{P}(\Omega)...\tilde{P}(\Omega)\tilde{P}(X^{-1}(A_i)) \\ & = & \tilde{P}(X^{-1}(A_i)) \end{array}$$

D'où $P(X_0 \in A_0, ..., X_n \in A_n) = P(X_0 \in A_0)...P(X_n \in A_n)$ pour tout $n \in \mathbf{N}$. On vient de démontrer que les X_i sont indépendants.

On a aussi vu que $P(X_i \in A) = \tilde{P}(X^{-1}(A))$. La loi P_{X_i} ne dépend donc pas de i et $P_{X_i} = P_{X_j}$ pour tout i, j. Ceci termine la démonstration de la proposition.

Dans notre exemple, nous avons $P_{X_i} = \frac{1}{2}(\delta_0 + \delta_1)$ ce qui implique l'égalité

$$E(X_1) = \int_{\mathbf{R}} x \, dP_{X_1}(x) = 0 \times 1/2 + 1 \times 1/2 = 1/2.$$

On peut maintenant appliquer la loi forte des grands nombres : pour presque tout $\omega \in \Omega$,

$$\frac{\#\{i \leq n \mid \omega_i = face\}}{n} \xrightarrow[n \to \infty]{} 1/2$$

ou encore

$$\tilde{P}^{\otimes \mathbf{N}}\Big(\Big\{(\omega_i)_{i \in \mathbf{N}} \in \{pile,\ face\}^{\mathbf{N}} \mid \frac{1}{n}\#\{i \leq n \mid \omega_i = face\} \xrightarrow[n \to \infty]{} 1/2\Big\}\Big) = 1.$$

La fréquence d'apparition de *face* au cours d'une infinité de lancers est égale à 1/2 presque sûrement, lorsque la pièce est bien équilibrée.

Nous allons démontrer la loi forte des grands nombres à partir du lemme suivant :

Lemme 3 *Soit (Y_i) une suite de variables aléatoires. Si pour tout $\varepsilon > 0$,*

$$\sum_{i=1}^{\infty} P(|Y_i| > \varepsilon) < \infty$$

alors la suite $(Y_i)_{i \in \mathbf{N}}$ converge presque sûrement vers 0 :

$$\textit{pour presque tout } \omega \in \Omega, \quad Y_i(\omega) \xrightarrow[i \to \infty]{} 0.$$

Le lemme montre que $P(\{\omega \in \Omega \mid Y_i(\omega) \xrightarrow[i\to\infty]{} 0\}) = 1$.

Preuve du lemme
On applique le lemme de Borel-Cantelli. La quantité ε étant fixée, on pose

$$A_i = (|Y_i| > \varepsilon).$$

Comme $\sum P(A_i) < \infty$, presque tout $\omega \in \Omega$ n'appartient qu'à un nombre fini de A_i. Notons par C_ε cet ensemble. Nous avons $P(C_\varepsilon) = 1$.

$$\forall \omega \in C_\varepsilon,\ \exists N \in \mathbf{N},\ \forall n \geq N,\ \ \omega \notin A_n \text{ et } |Y_n(\omega)| < \varepsilon.$$

On prend $\varepsilon = 1/k$, $k \in \mathbf{N}^*$ et on considère l'intersection des $C_{1/k}$.

$$C = \bigcap_{k \in \mathbf{N}^*} C_{1/k}, \qquad P(C) = 1.$$

Pour tout $\omega \in C$ et tout $k \in \mathbf{N}^*$, le point ω est dans $C_{1/k}$, si bien qu'il existe $N \in \mathbf{N}$ tel que pour tout $n \geq N$, $|Y_n(\omega)| < 1/k$. Ceci montre que $\lim\limits_{n\to\infty} Y_n(\omega) = 0$, comme souhaité.

Preuve de la loi forte des grands nombres
Pour simplifier, nous allons supposer que les X_i sont de carré intégrable dans la preuve. On donnera une preuve dans le cas intégrable plus tard, dans le chapitre consacré à la convergence de séries de variables aléatoires.

Nous avons, pour tout i, $E(X_i) = E(X_1)$. Quitte à remplacer les X_i par $X_i - E(X_i)$, on peut supposer $E(X_i) = 0$. On dit qu'on *centre* les variables aléatoires. On veut montrer que $\frac{S_n}{n}$ converge presque sûrement vers 0. Essayons d'appliquer le lemme précédent. Rappelons l'égalité $E(\frac{S_i}{i}) = E(X_1) = 0$.

$$P\Big(\Big|\frac{S_i}{i}\Big| > \varepsilon\Big) \leq \frac{V(S_i/i)}{\varepsilon^2} = \frac{V(X_1)}{i\,\varepsilon^2} \quad \text{par Bienaymé-Tchebichev.}$$

$$\sum_{n=1}^{\infty} P\Big(\Big|\frac{S_i}{i}\Big| > \varepsilon\Big) \ \leq\ \frac{V(X_1)}{\varepsilon^2} \sum_{i=1}^{\infty} \frac{1}{i} = +\infty.$$

La condition du lemme, avec $Y_i = S_i/i$, n'est pas vérifiée. Remplaçons i par i^2 : $Y_i = \frac{S_{i^2}}{i^2}$. Nous avons maintenant

$$\sum_i P\Big(\Big|\frac{S_{i^2}}{i^2}\Big| > \varepsilon\Big) \leq \frac{V(X_1)}{\varepsilon} \sum_i \frac{1}{i^2}.$$

La série $\sum \frac{1}{i^2}$ est convergente (sa limite vaut $\pi^2/6$). Le lemme précédent donne la convergence de la suite $\frac{S_{i^2}}{i^2}$:

$$\frac{S_{i^2}}{i^2} \xrightarrow[i\to\infty]{} 0 \quad \text{p.s.}$$

Pour chaque $n \in \mathbf{N}^*$, on prend $i \in \mathbf{N}$ le plus grand possible, tel que $i^2 \leq n$. L'entier i est égal à la partie entière de $\sqrt{n}$ et on a les encadrements :

$$i^2 \leq n \leq (i+1)^2 - 1, \quad i^2 \leq n \leq i^2 + 2i, \quad 0 \leq n - i^2 \leq 2i \leq 2\sqrt{n}.$$

$$S_n = \sum_{k=1}^{n} X_k = \sum_{k=1}^{i^2} X_k + \sum_{k=i^2+1}^{n} X_k$$

$$\Big|\frac{S_n}{n}\Big| \leq \Big|\frac{S_{i^2}}{i^2}\Big| + \frac{1}{n}\,\Big|\sum_{k=i^2+1}^{n} X_k\Big|.$$

Pour majorer le dernier terme, on raisonne comme précédemment :

$$P\left(\frac{1}{n}\,\Big|\sum_{k=i^2+1}^{n} X_k\Big| > \varepsilon\right) \leq \frac{1}{n^2\varepsilon^2} V\Big(\sum_{k=i^2+1}^{n} X_k\Big) \leq \frac{n-i^2}{n^2\varepsilon^2}\, V(X_1) \leq \frac{2}{n^{3/2}}\frac{V(X_1)}{\varepsilon^2}.$$

La série $\sum \frac{1}{n^{3/2}}$ est convergente. D'après le lemme,

$$\frac{1}{n}\sum_{k=i^2+1}^{n} X_k \xrightarrow[i\to\infty]{} 0 \quad \text{p.s.}$$

Le résultat est démontré.

Preuve du corollaire
On applique la loi des grands nombres à la suite $(\mathbf{1}_A \circ X_i)$.

$$E(\mathbf{1}_A \circ X_1) = E(\mathbf{1}_{X_1^{-1}(A)}) = E(\mathbf{1}_{(X_1 \in A)}) = P(X_1 \in A).$$

$$\frac{S_n}{n} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_A(X_k(\omega)) = \frac{1}{n}\ \#\{k \in \{1,...,n\} \mid X_k(\omega) \in A\}$$

Cette quantité converge vers $E(\mathbf{1}_A \circ X_1)$ d'après la loi forte des grands nombres.

Complément
Donnons une généralisation aisée de la loi des grands nombres qui s'avère utile en pratique.

Proposition 8 *Soit $(X_i)_{i\in\mathbf{N}}$ une suite de variables aléatoires indépendantes identiquement distribuées. Soit $m \in \mathbf{N}^*$ et $f : \mathbf{R}^m \to \mathbf{R}$ une fonction qui est $P_{(X_1,...,X_m)}$-intégrable. Alors pour presque tout $\omega \in \Omega$,*

$$\frac{1}{N}\sum_{k=1}^{N} f(X_k, ..., X_{k+m-1}) \xrightarrow[N\to\infty]{} E(f(X_1, ..., X_m)).$$

Preuve
On pose $Y_k = f(X_k, ...X_{k+m-1})$; les variables Y_k ne sont pas indépendantes dans leur ensemble. Par contre, les variables $Y_1, Y_{m+1}, Y_{2m+1}, Y_{3m+1}...$ sont indépendantes entre elles. Plus généralement, pour chaque $r \in \{1, ..., m\}$, les variables $(Y_{mk+r})_{k\in\mathbf{N}}$ sont intégrables, indépendantes et identiquement distribuées. On peut donc appliquer la loi des grands nombres à ces m suites de variables aléatoires et faire la somme des résultats, ce qui donne

$$\frac{1}{n}\sum_{k=1}^{mn} Y_k \xrightarrow[n\to\infty]{} \sum_{i=1}^{m} E(Y_i).$$

Comme les variables X_i sont indépendantes identiquement distribuées, nous avons

$$\begin{aligned} E(Y_i) = E(f(X_i, ..., X_{m+i-1})) &= \int f(x_1, ..., x_m)\, dP_{X_i}(x_1)...dP_{X_{m+i-1}}(x_m) \\ &= \int f(x_1, ..., x_m)\, dP_{X_1}(x_1)...dP_{X_1}(x_m) \\ &= E(f(X_1, ..., X_m)). \end{aligned}$$

Ceci montre le résultat pour N multiple de m. Si N n'est pas multiple de m, on peut l'écrire sous la forme $N = mn + i$ avec $0 < i < m$. On remarque alors que chacun des termes Y_{mn+i}/n converge vers 0 presque sûrement quand n tend vers l'infini, d'après la loi des grands nombres :

$$\frac{Y_{mn+i}}{n} = \Big(\frac{1}{n}\sum_{k=1}^{n} Y_{mk+i}\Big) - \frac{n-1}{n}\Big(\frac{1}{n-1}\sum_{k=1}^{n-1} Y_{mk+i}\Big) \xrightarrow[n\to\infty]{} E(Y_i) - E(Y_i) = 0.$$

La proposition s'ensuit.

Application
On revient à l'exemple de pile ou face. Prenons
- $\Omega = \{\text{pile}, \text{face}\}^{\otimes\mathbf{N}}$,
- $\mathcal{T} = \mathcal{P}(\{\text{pile}, \text{face}\})^{\otimes\mathbf{N}}$,
- $P = (\frac{1}{2}\delta_{pile} + \frac{1}{2}\delta_{face})^{\otimes\mathbf{N}}$.

D'après la loi des grands nombres,

$$\frac{1}{n}\ \#\{k \in \{1, ..., n\} \mid \omega_k = face\} \xrightarrow[n\to\infty]{} \frac{1}{2} \quad \text{pour presque tout } \omega \in \Omega.$$

En particulier, pour presque tout $\omega \in \Omega$, face apparaît une infinité de fois dans la suite Ω. Soit $(a_1, ..., a_m) \in \{pile, face\}^m$. Prenons $f = \mathbf{1}_{\{(a_1,...,a_m)\}}$. Nous obtenons

$$\begin{aligned} E(f(X_1, ..., X_m)) &= P(X_1 = a_1, ..., X_m = a_m) \\ &= P(X_1 = a_1)...P(X_n = a_m) \\ &= 1/2^m. \end{aligned}$$

Appliquons la proposition précédente.

$$\frac{1}{n}\ \#\{k \in \{1, ..., n\} \mid (\omega_k, ..., \omega_{k+m-1}) = (a_1, ..., a_m)\} \xrightarrow[n\to\infty]{} \frac{1}{2^m} \quad \text{p.s.}$$

Notons par $\Omega_{(a_1,...,a_m)}$ l'ensemble des $\omega \in \Omega$ pour lesquels on a cette convergence. Cet ensemble est de probabilité 1. On en déduit

$$P\Big(\bigcap_{m\in\mathbf{N}^*}\ \bigcap_{(a_1,...,a_m)\in\{pile,face\}^m} \Omega_{(a_1,...,a_m)}\Big) = 1$$

Presque tout $\omega \in \Omega$ appartient à tous les $\Omega_{(a_1,...,a_m)}$. Cela signifie que dans presque toute suite $\omega \in \Omega$, tous les mots $(a_1, ..., a_m)$ apparaissent une infinité de fois dans la suite ω avec fréquence $1/2^m$, pour tout $m \in \mathbf{N}^*$.

4.3 Illustration numérique

Pour illustrer la loi des grands nombres, on considère plusieurs suites numériques, chacune consistant en mille chiffres obtenus de plusieurs façons.

La première a été obtenue en lançant mille fois un dé à dix faces.

9 6 3 9 1 7 8 0 7 1 9 5 5 3 5 7 8 4 9 9 2 1 0 2 5 1 0 1 2 7 3 3 0 0 4 8 0 8 0 1 0 4 6 3 5 6 4 3 1 1 2 6 1 3 2 7 7 7 5
0 8 7 1 0 5 6 6 0 2 5 3 3 7 0 5 2 0 0 4 0 4 4 6 5 8 2 2 7 3 2 8 7 2 0 2 4 5 4 9 6 4 2 5 1 0 6 0 4 5 5 5 8 9 1 2 3 5 9
7 3 4 4 2 8 6 1 4 6 5 8 4 0 3 2 2 8 6 7 6 9 3 9 4 0 0 8 2 4 8 9 5 5 5 7 8 0 5 4 5 8 2 8 5 6 6 8 4 5 5 7 0 7 0 1 0 9 3
6 8 6 3 0 3 4 3 6 6 5 8 9 3 4 2 7 3 2 4 3 5 4 1 7 8 6 0 3 8 2 0 0 0 6 5 1 0 7 1 6 8 6 4 3 1 3 6 8 5 5 2 8 5 7 4 0 9 7
1 7 1 2 3 9 5 7 6 0 7 9 8 3 8 3 1 6 9 5 8 5 5 7 1 9 0 9 0 3 2 8 8 6 9 5 1 5 9 6 0 9 1 9 6 5 7 8 7 7 7 9 2 4 8 6 1 9 2
3 0 8 1 6 6 2 2 5 2 3 8 4 8 2 3 0 2 7 7 3 4 7 0 3 3 7 6 9 4 2 6 9 9 6 1 5 3 7 6 4 8 3 6 5 7 7 9 8 6 8 1 2 5 4 2 4 9 8
9 1 7 5 7 2 1 7 2 8 3 3 2 1 1 8 6 7 0 1 2 0 1 1 1 8 7 7 2 4 2 2 8 5 4 6 9 3 4 6 8 0 5 8 9 5 5 8 0 2 0 2 4 9 8 2 7 3 1
1 2 4 0 0 0 2 5 2 6 8 2 9 2 4 8 3 8 9 9 2 3 3 1 5 3 1 5 3 8 6 5 5 7 9 6 6 6 8 9 9 3 8 1 7 2 8 4 4 3 4 7 1 7 0 3 3 2 1
6 3 0 2 3 7 2 6 3 3 6 3 7 7 0 5 6 0 3 1 8 8 9 3 2 6 0 3 0 4 3 2 5 3 6 5 5 1 7 4 0 6 3 3 9 3 9 7 9 8 2 7 0 2 1 4 8 8 6
5 8 3 9 9 7 0 4 3 4 6 9 3 7 9 6 0 0 4 9 9 2 4 6 1 2 7 8 7 5 1 0 2 6 2 2 3 3 5 4 2 8 4 2 7 9 4 0 8 8 5 5 1 1 2 5 3 1 2
1 2 2 6 8 0 1 1 5 9 4 5 4 8 8 7 0 4 4 9 9 4 5 0 0 2 4 1 6 0 1 8 5 7 6 0 7 6 5 2 1 6 6 8 6 8 3 5 2 5 0 1 5 1 6 9 8 0 8
9 9 9 1 0 1 7 0 7 9 5 3 7 7 9 0 7 7 2 6 0 8 1 0 5 4 5 7 0 9 1 9 0 6 7 9 5 9 4 1 2 2 8 1 9 0 5 4 3 3 2 0 1 6 0 5 3 8 3
5 6 3 9 1 9 7 4 3 8 7 4 5 0 9 7 7 3 1 4 2 2 5 6 2 0 8 4 4 6 7 9 3 5 8 2 0 6 9 5 9 1 3 9 8 2 9 5 5 3 4 8 1 7 6 6 9 5 2
0 5 8 4 7 8 1 0 0 8 6 6 8 0 5 4 4 5 7 5 1 3 8 5 0 7 3 4 4 4 1 3 5 8 9 8 0 5 3 7 2 2 7 7 3 5 1 8 6 5 7 5 9 9 2 3 8 7 7
3 9 1 7 9 3 3 8 9 6 5 8 2 3 4 5 7 4 6 9 2 7 0 3 9 1 3 0 7 6 4 5 1 1 7 0 8 2 5 7 7 7 0 1 3 5 3 6 7 4 2 1 5 9 5 7 7 9 0
1 9 2 8 5 6 6 1 9 0 7 9 7 3 7 0 9 5 4 9 7 6 7 3 6 1 6 5 4 1 8 6 4 5 8 9 0 2 3 4 9 6 3 4 0 3 9 2 5 0 6 3 9 2 1 0 5 2 7
8 0 4 8 3 3 5 7 6 5 6 3 7 3 2 3 7 5 6 1 3 5 6 7 2 0 5 0 7 2 0 9 3 1 6 6 9 4 6 7 9 1 0 9 4 4 1 4 6 7 8 8 0 6 8 6

La seconde est obtenue en utilisant un ordinateur et un générateur de nombres aléatoires.

0 8 4 8 3 7 0 2 4 5 9 2 0 3 3 2 9 4 1 1 3 0 7 2 0 5 5 0 0 9 3 1 3 4 2 5 6 4 2 4 6 3 2 9 2 1 6 8 9 9 7 4 7 6 1 9 0 2 2
4 9 1 7 5 3 4 7 7 2 6 7 2 3 8 8 9 7 6 2 5 7 0 7 1 2 2 9 7 6 4 9 0 6 8 4 0 6 5 3 0 2 2 9 3 3 8 4 8 5 4 1 3 9 1 9 2 7 6
9 2 7 1 7 6 8 1 1 3 9 7 4 6 6 6 0 0 3 2 7 6 8 0 4 0 2 7 7 9 1 7 4 0 5 8 7 4 8 6 9 0 3 8 7 5 9 7 7 0 9 7 5 2 5 0 1 6 2
8 8 6 3 7 5 5 2 5 0 5 9 8 2 3 4 3 5 1 0 7 7 6 8 2 0 9 5 1 8 0 5 7 8 8 9 8 3 7 0 3 4 4 6 6 2 1 3 5 4 4 7 5 1 5 9 2 9 3
1 2 4 5 7 2 7 9 6 8 1 7 1 1 3 4 9 2 4 2 1 9 2 1 9 2 3 8 4 0 6 3 9 5 0 4 7 3 6 1 9 4 0 3 4 3 1 4 4 3 1 8 7 3 6 0 0 7 3
8 9 0 5 9 1 6 6 0 5 8 4 2 2 6 5 3 2 4 3 9 0 4 3 4 5 5 3 0 0 5 7 9 7 3 4 1 6 9 7 9 9 2 3 8 9 2 5 0 1 2 5 5 9 2 3 6 5 6
5 7 5 5 6 2 6 6 2 0 6 1 2 1 3 7 2 2 3 2 8 9 8 6 8 4 1 7 2 1 7 9 8 5 2 4 6 5 6 3 4 0 4 6 4 2 9 3 3 4 6 0 9 8 8 2 2 4 4
0 8 4 3 9 6 0 5 1 8 4 0 8 0 8 4 8 2 4 3 5 0 2 5 8 7 2 4 0 4 0 2 1 5 9 2 3 8 5 4 6 7 8 7 9 0 6 6 5 9 5 3 9 6 6 8 7 1 3
4 0 7 0 2 5 3 2 1 7 6 2 7 8 9 9 9 4 9 7 9 8 7 5 9 5 4 9 0 5 5 4 8 8 3 3 5 7 7 6 9 8 1 1 1 7 0 0 0 4 9 9 5 7 4 6 8 0 9
6 0 2 8 3 9 7 2 2 9 7 2 3 0 3 3 0 3 7 5 3 3 5 7 7 7 3 3 9 1 9 7 4 5 5 1 6 3 6 6 4 7 3 4 3 3 0 0 8 2 0 0 3 6 8 8 5 7 1
8 4 7 5 5 1 0 6 9 7 2 7 1 6 8 2 8 3 3 9 1 6 9 9 7 3 0 9 8 6 2 0 4 7 8 1 7 6 7 2 0 2 8 2 2 2 0 4 9 3 0 1 3 6 3 3 5 9 5
9 9 6 1 0 1 4 8 4 4 3 7 3 6 4 2 2 8 7 2 4 8 8 6 0 1 6 0 2 1 6 1 4 3 2 2 7 0 5 7 0 2 3 3 3 8 1 1 8 0 8 5 5 3 0 2 0 8 8
3 1 8 9 1 4 8 6 5 3 1 5 5 8 0 5 3 4 7 4 4 0 5 5 2 2 0 0 3 0 5 8 8 6 9 1 3 9 1 2 5 9 9 4 1 5 6 2 0 9 2 6 0 7 3 0 0 7 9
7 9 7 9 1 2 1 9 1 1 4 3 1 1 8 4 2 9 9 0 7 2 1 8 9 3 8 6 2 4 9 0 3 3 7 6 1 0 9 8 8 2 5 4 0 7 0 1 5 2 1 5 1 1 7 7 4 6 3
6 9 1 4 6 0 6 3 0 0 9 1 4 0 7 1 9 4 8 4 4 7 1 4 1 7 7 2 7 7 0 5 9 1 9 4 3 1 9 5 1 0 9 7 5 7 3 8 9 2 1 3 7 7 9 0 4 5 6
6 4 3 2 4 3 3 2 2 7 2 1 0 7 7 2 4 6 6 8 2 4 1 7 8 1 5 4 7 2 1 4 3 5 2 5 5 3 8 0 0 2 0 6 4 9 1 9 5 7 5 0 7 4 9 1 1 5 6
0 1 0 2 6 3 3 9 8 7 8 6 7 5 5 7 1 3 6 7 8 7 9 8 3 8 2 3 9 8 0 9 3 5 4 8 0 6 9 5 4 1 7 4 2 1 2 9 6 6 0 7 7 0 2 4

La troisième est constituée des mille premières décimales de π.

1 4 1 5 9 2 6 5 3 5 8 9 7 9 3 2 3 8 4 6 2 6 4 3 3 8 3 2 7 9 5 0 2 8 8 4 1 9 7 1 6 9 3 9 9 3 7 5 1 0 5 8 2 0 9 7 4 9 4
4 5 9 2 3 0 7 8 1 6 4 0 6 2 8 6 2 0 8 9 9 8 6 2 8 0 3 4 8 2 5 3 4 2 1 1 7 0 6 7 9 8 2 1 4 8 0 8 6 5 1 3 2 8 2 3 0 6 6
4 7 0 9 3 8 4 4 6 0 9 5 5 0 5 8 2 2 3 1 7 2 5 3 5 9 4 0 8 1 2 8 4 8 1 1 1 7 4 5 0 2 8 4 1 0 2 7 0 1 9 3 8 5 2 1 1 0 5
5 5 9 6 4 4 6 2 2 9 4 8 9 5 4 9 3 0 3 8 1 9 6 4 4 2 8 8 1 0 9 7 5 6 6 5 9 3 3 4 4 6 1 2 8 4 7 5 6 4 8 2 3 3 7 8 6 7 8
3 1 6 5 2 7 1 2 0 1 9 0 9 1 4 5 6 4 8 5 6 6 9 2 3 4 6 0 3 4 8 6 1 0 4 5 4 3 2 6 6 4 8 2 1 3 3 9 3 6 0 7 2 6 0 2 4 9 1
4 1 2 7 3 7 2 4 5 8 7 0 0 6 6 0 6 3 1 5 5 8 8 1 7 4 8 8 1 5 2 0 9 2 0 9 6 2 8 2 9 2 5 4 0 9 1 7 1 5 3 6 4 3 6 7 8 9 2
5 9 0 3 6 0 0 1 1 3 3 0 5 3 0 5 4 8 8 2 0 4 6 6 5 2 1 3 8 4 1 4 6 9 5 1 9 4 1 5 1 1 6 0 9 4 3 3 0 5 7 2 7 0 3 6 5 7 5
9 5 9 1 9 5 3 0 9 2 1 8 6 1 1 7 3 8 1 9 3 2 6 1 1 7 9 3 1 0 5 1 1 8 5 4 8 0 7 4 4 6 2 3 7 9 9 6 2 7 4 9 5 6 7 3 5 1 8
8 5 7 5 2 7 2 4 8 9 1 2 2 7 9 3 8 1 8 3 0 1 1 9 4 9 1 2 9 8 3 3 6 7 3 3 6 2 4 4 0 6 5 6 6 4 3 0 8 6 0 2 1 3 9 4 9 4 6
3 9 5 2 2 4 7 3 7 1 9 0 7 0 2 1 7 9 8 6 0 9 4 3 7 0 2 7 7 0 5 3 9 2 1 7 1 7 6 2 9 3 1 7 6 7 5 2 3 8 4 6 7 4 8 1 8 4 6
7 6 6 9 4 0 5 1 3 2 0 0 0 5 6 8 1 2 7 1 4 5 2 6 3 5 6 0 8 2 7 7 8 5 7 7 1 3 4 2 7 5 7 7 8 9 6 0 9 1 7 3 6 3 7 1 7 8 7
2 1 4 6 8 4 4 0 9 0 1 2 2 4 9 5 3 4 3 0 1 4 6 5 4 9 5 8 5 3 7 1 0 5 0 7 9 2 2 7 9 6 8 9 2 5 8 9 2 3 5 4 2 0 1 9 9 5 6
1 1 2 1 2 9 0 2 1 9 6 0 8 6 4 0 3 4 4 1 8 1 5 9 8 1 3 6 2 9 7 7 4 7 7 1 3 0 9 9 6 0 5 1 8 7 0 7 2 1 1 3 4 9 9 9 9 9 9
8 3 7 2 9 7 8 0 4 9 9 5 1 0 5 9 7 3 1 7 3 2 8 1 6 0 9 6 3 1 8 5 9 5 0 2 4 4 5 9 4 5 5 3 4 6 9 0 8 3 0 2 6 4 2 5 2 2 3
0 8 2 5 3 3 4 4 6 8 5 0 3 5 2 6 1 9 3 1 1 8 8 1 7 1 0 1 0 0 0 3 1 3 7 8 3 8 7 5 2 8 8 6 5 8 7 5 3 3 2 0 8 3 8 1 4 2 0
6 1 7 1 7 7 6 6 9 1 4 7 3 0 3 5 9 8 2 5 3 4 9 0 4 2 8 7 5 5 4 6 8 7 3 1 1 5 9 5 6 2 8 6 3 8 8 2 3 5 3 7 8 7 5 9 3 7 5
1 9 5 7 7 8 1 8 5 7 7 8 0 5 3 2 1 7 1 2 2 6 8 0 6 6 1 3 0 0 1 9 2 7 8 7 6 6 1 1 1 9 5 9 0 9 2 1 6 4 2 0 1 9 8 9

La quatrième est obtenue en conservant les cinq derniers chiffres de deux cents numéros de téléphone successifs d'un annuaire téléphonique.

4 1 0 1 4 9 1 4 0 1 5 7 0 1 6 2 3 7 4 8 9 6 1 7 8 5 1 2 2 9 1 2 7 5 3 5 1 0 3 5 4 4 9 4 1 9 1 3 3 0 8 7 9 0 7 8 8 4 2
3 0 8 0 2 0 5 9 1 5 2 5 1 7 9 1 9 8 3 3 8 9 0 1 6 9 6 4 2 7 9 9 1 3 0 9 5 2 0 4 6 8 2 3 1 6 9 7 4 4 0 5 3 3 9 2 9 1 9
4 0 5 9 6 1 9 0 2 7 0 8 5 7 7 7 7 5 4 8 4 2 3 1 8 2 6 4 6 0 3 5 2 1 1 7 8 5 6 4 9 2 5 0 9 6 7 5 8 5 6 6 5 1 3 6 9 1 7
9 8 0 5 7 6 5 9 4 1 0 6 8 5 1 1 3 8 5 9 6 5 4 8 5 8 4 5 8 9 5 5 7 4 0 9 8 5 5 1 0 0 0 7 2 8 5 5 9 4 2 5 5 8 9 9 0 5 5
0 7 3 9 7 9 8 5 2 9 8 0 2 9 0 4 0 9 8 5 1 0 0 9 3 8 3 7 3 8 4 0 6 4 5 9 1 4 4 9 3 8 5 1 5 9 5 6 1 5 9 1 6 2 8 5 6 7 6
9 4 1 5 4 5 8 2 5 8 2 1 7 2 2 5 0 9 8 4 2 7 7 1 1 5 0 0 8 6 5 4 5 9 0 8 2 7 7 1 5 7 5 7 8 6 4 2 4 5 2 5 0 1 5 5 6 9 1
5 8 5 0 4 7 9 7 4 3 4 5 9 7 1 0 5 9 8 9 5 9 2 8 5 3 5 9 3 1 5 4 2 0 6 8 5 9 2 2 7 3 9 2 2 7 9 4 0 4 5 5 9 0 9 5 9 4 8
4 9 4 0 1 1 5 3 8 9 5 8 2 6 2 8 7 5 4 9 5 8 5 9 2 0 6 6 8 3 6 4 5 9 0 0 3 5 7 3 0 2 0 1 8 7 6 5 6 8 4 6 8 5 4 3 7 5 6
2 1 7 5 9 2 3 9 8 4 5 0 9 9 6 9 8 8 5 6 4 4 2 8 9 9 5 4 4 1 5 1 5 2 7 3 4 4 8 9 9 5 8 7 4 0 6 1 9 2 8 3 9 5 2 6 8 4 5
7 6 2 9 5 9 3 6 4 5 8 0 3 9 5 8 0 3 3 5 9 3 5 6 0 2 0 8 1 8 9 1 2 9 5 9 3 8 6 8 5 3 8 5 5 0 4 3 6 8 0 3 1 9 6 4 2 3 9
5 7 8 5 6 5 9 6 1 3 0 5 0 5 8 5 7 5 6 7 5 8 9 7 1 9 4 5 8 9 8 9 5 6 2 5 4 0 6 9 8 5 4 3 4 5 6 7 9 4 5 9 2 6 3 7 9 2 6
3 8 4 5 3 3 5 5 3 4 4 9 8 0 3 4 3 7 7 4 3 8 9 3 6 7 3 3 5 6 8 6 7 8 3 0 5 2 8 5 7 9 2 4 5 9 5 6 0 8 6 5 9 3 8 1 5 4 1
5 5 8 2 0 0 4 1 1 0 1 9 0 7 5 7 5 5 0 1 4 9 9 7 4 0 4 5 1 4 3 9 1 0 6 4 9 5 2 6 3 9 9 2 4 5 7 5 9 7 4 4 8 2 8 2 1 3 8
5 7 7 5 7 3 1 3 4 7 3 1 3 0 0 9 2 0 8 9 9 7 3 5 4 3 6 8 5 4 3 7 8 9 8 9 2 0 9 8 9 2 9 9 3 6 8 6 6 1 4 1 0 2 5 0 5 7 4
1 0 5 6 3 6 5 8 7 4 3 9 1 1 4 1 5 3 0 8 5 4 6 8 5 9 5 3 7 8 4 3 3 2 6 5 8 8 2 2 2 3 2 8 9 2 3 6 2 8 4 3 1 8 9 2 7 5 8
5 9 1 6 0 0 4 9 9 7 9 4 8 6 9 8 2 2 0 5 5 4 4 8 1 5 7 6 3 1 7 6 1 8 5 8 1 6 5 8 9 4 3 7 6 8 5 7 7 6 1 3 7 2 8 5 3 2 4
0 6 0 1 9 2 0 8 9 5 8 5 5 3 7 8 2 6 6 5 7 5 2 5 3 2 8 3 0 6 7 9 1 4 3 7 5 5 7 1 9 5 9 3 0 5 9 5 8 7 6 5 9 1 2 8

La cinquième s'obtient en concaténant les nombres entiers dans l'ordre croissant en partant de un.

1 2 3 4 5 6 7 8 9 1 0 1 1 1 2 1 3 1 4 1 5 1 6 1 7 1 8 1 9 2 0 2 1 2 2 2 3 2 4 2 5 2 6 2 7 2 8 2 9 3 0 3 1 3 2 3 3 3 4
3 5 3 6 3 7 3 8 3 9 4 0 4 1 4 2 4 3 4 4 4 5 4 6 4 7 4 8 4 9 5 0 5 1 5 2 5 3 5 4 5 5 5 6 5 7 5 8 5 9 6 0 6 1 6 2 6 3 6
4 6 5 6 6 6 7 6 8 6 9 7 0 7 1 7 2 7 3 7 4 7 5 7 6 7 7 7 8 7 9 8 0 8 1 8 2 8 3 8 4 8 5 8 6 8 7 8 8 8 9 9 0 9 1 9 2 9 3
9 4 9 5 9 6 9 7 9 8 9 9 1 0 0 1 0 1 1 0 2 1 0 3 1 0 4 1 0 5 1 0 6 1 0 7 1 0 8 1 0 9 1 1 0 1 1 1 1 1 2 1 1 3 1 1 4 1 1
5 1 1 6 1 1 7 1 1 8 1 1 9 1 2 0 1 2 1 1 2 2 1 2 3 1 2 4 1 2 5 1 2 6 1 2 7 1 2 8 1 2 9 1 3 0 1 3 1 1 3 2 1 3 3 1 3 4 1
3 5 1 3 6 1 3 7 1 3 8 1 3 9 1 4 0 1 4 1 1 4 2 1 4 3 1 4 4 1 4 5 1 4 6 1 4 7 1 4 8 1 4 9 1 5 0 1 5 1 1 5 2 1 5 3 1 5 4
1 5 5 1 5 6 1 5 7 1 5 8 1 5 9 1 6 0 1 6 1 1 6 2 1 6 3 1 6 4 1 6 5 1 6 6 1 6 7 1 6 8 1 6 9 1 7 0 1 7 1 1 7 2 1 7 3 1 7
4 1 7 5 1 7 6 1 7 7 1 7 8 1 7 9 1 8 0 1 8 1 1 8 2 1 8 3 1 8 4 1 8 5 1 8 6 1 8 7 1 8 8 1 8 9 1 9 0 1 9 1 1 9 2 1 9 3 1
9 4 1 9 5 1 9 6 1 9 7 1 9 8 1 9 9 2 0 0 2 0 1 2 0 2 2 0 3 2 0 4 2 0 5 2 0 6 2 0 7 2 0 8 2 0 9 2 1 0 2 1 1 2 1 2 2 1 3
2 1 4 2 1 5 2 1 6 2 1 7 2 1 8 2 1 9 2 2 0 2 2 1 2 2 2 2 2 3 2 2 4 2 2 5 2 2 6 2 2 7 2 2 8 2 2 9 2 3 0 2 3 1 2 3 2 2 3
3 2 3 4 2 3 5 2 3 6 2 3 7 2 3 8 2 3 9 2 4 0 2 4 1 2 4 2 2 4 3 2 4 4 2 4 5 2 4 6 2 4 7 2 4 8 2 4 9 2 5 0 2 5 1 2 5 2 2
5 3 2 5 4 2 5 5 2 5 6 2 5 7 2 5 8 2 5 9 2 6 0 2 6 1 2 6 2 2 6 3 2 6 4 2 6 5 2 6 6 2 6 7 2 6 8 2 6 9 2 7 0 2 7 1 2 7 2
2 7 3 2 7 4 2 7 5 2 7 6 2 7 7 2 7 8 2 7 9 2 8 0 2 8 1 2 8 2 2 8 3 2 8 4 2 8 5 2 8 6 2 8 7 2 8 8 2 8 9 2 9 0 2 9 1 2 9
2 2 9 3 2 9 4 2 9 5 2 9 6 2 9 7 2 9 8 2 9 9 3 0 0 3 0 1 3 0 2 3 0 3 3 0 4 3 0 5 3 0 6 3 0 7 3 0 8 3 0 9 3 1 0 3 1 1 3
1 2 3 1 3 3 1 4 3 1 5 3 1 6 3 1 7 3 1 8 3 1 9 3 2 0 3 2 1 3 2 2 3 2 3 3 2 4 3 2 5 3 2 6 3 2 7 3 2 8 3 2 9 3 3 0 3 3 1
3 3 2 3 3 3 3 3 4 3 3 5 3 3 6 3 3 7 3 3 8 3 3 9 3 4 0 3 4 1 3 4 2 3 4 3 3 4 4 3 4 5 3 4 6 3 4 7 3 4 8 3 4 9 3 5 0 3 5
1 3 5 2 3 5 3 3 5 4 3 5 5 3 5 6 3 5 7 3 5 8 3 5 9 3 6 0 3 6 1 3 6 2 3 6 3 3 6 4 3 6 5 3 6 6 3 6 7 3 6 8 3 6 9 3

La sixième est obtenue en concaténant le nombre d'habitants de chacune des communes de l'Ain, ordonnées par ordre alphabétique (2012, Abergement-Clémenciat $\rightarrow$ Vonnas) et en conservant les mille premiers chiffres.

7 9 1 2 3 9 1 4 7 9 6 1 6 6 0 1 1 6 2 5 5 7 7 5 8 3 4 7 1 0 8 7 3 9 3 3 1 9 1 6 5 3 5 6 6 4 9 7 4 2 2 4 3 2 2 1 1 4 0
4 1 2 1 8 7 5 3 1 5 9 3 2 1 1 8 9 0 2 8 9 0 5 6 8 6 8 0 8 9 2 3 8 4 5 3 2 5 9 1 1 9 3 6 9 2 9 8 3 5 0 5 4 4 2 8 1 7 3
8 4 4 1 4 7 4 7 5 5 2 7 1 4 6 4 3 5 5 6 2 7 4 8 3 1 9 3 0 2 9 6 2 3 0 9 9 5 3 1 3 4 2 1 4 6 1 2 7 9 3 7 2 4 9 9 9 1 4
9 4 5 6 1 1 2 4 8 8 0 9 5 2 5 9 3 2 1 7 6 2 7 4 2 7 1 7 4 8 1 4 8 5 4 4 8 6 3 0 2 0 1 0 1 2 2 4 4 2 1 1 9 1 1 4 5 1 8
6 1 4 2 0 8 1 2 4 8 0 7 0 9 6 4 4 8 3 6 7 3 7 3 6 8 2 8 9 9 4 2 1 9 3 3 3 5 3 3 2 1 4 1 6 6 7 5 1 6 6 7 3 0 6 5 0 9 8
5 2 2 1 8 1 9 1 5 9 7 1 5 9 1 6 7 9 5 5 1 3 8 1 4 6 5 1 4 7 2 1 7 7 1 4 3 1 1 8 7 2 0 9 3 1 2 0 1 0 9 4 5 2 7 9 5 5 7
0 1 2 6 7 5 4 6 1 5 3 1 2 2 4 2 5 0 1 2 2 8 8 7 1 0 9 8 6 7 9 6 2 5 5 1 4 2 7 5 2 7 2 1 3 8 4 3 7 9 1 7 6 6 2 0 4 2 8
5 1 3 0 0 1 4 5 8 7 3 4 4 2 7 4 2 2 4 8 8 6 9 8 8 2 1 2 2 6 3 9 5 4 7 6 1 9 2 9 1 0 2 7 2 2 1 1 1 8 1 7 9 9 1 8 9 3 8
1 5 1 4 8 7 5 4 2 2 4 2 9 4 3 3 2 5 8 9 0 5 3 6 5 2 0 1 2 1 4 7 4 2 0 1 9 6 8 4 6 0 8 1 0 0 4 9 5 9 1 4 9 1 0 9 7 9 1
7 5 8 3 6 1 2 2 1 7 6 1 9 9 9 8 0 6 1 1 9 5 3 6 2 1 4 2 0 8 1 7 4 0 6 9 8 5 3 1 4 6 0 5 1 1 5 0 1 1 9 1 7 4 2 3 8 3 2
1 3 6 4 2 2 1 7 1 1 1 1 0 3 3 5 3 2 5 5 2 1 8 9 2 0 3 6 9 6 0 1 1 7 9 2 6 7 1 0 8 1 2 7 1 3 0 2 1 5 6 6 2 7 2 3 1 1 4
1 2 6 9 1 1 2 7 3 1 5 5 5 3 8 8 8 1 0 1 1 5 8 1 2 6 3 0 3 4 4 0 6 4 3 2 6 8 2 1 0 9 4 3 1 2 1 9 9 3 2 2 5 7 1 3 2 1 6
2 1 0 0 5 1 0 2 6 1 6 5 0 2 4 5 9 6 4 0 6 7 4 1 3 1 5 1 3 3 1 2 4 8 7 5 1 6 8 8 2 2 1 4 7 3 9 2 2 1 5 5 9 2 7 4 2 0 0
2 7 8 0 1 8 1 6 5 4 8 3 4 1 1 2 4 6 6 6 8 2 4 7 6 7 1 7 2 4 5 3 9 1 0 1 0 0 3 3 5 1 3 2 4 6 2 1 0 6 6 2 5 3 3 7 4 7 5
8 5 1 5 0 6 1 6 3 8 6 5 3 2 5 4 8 1 5 4 2 1 2 8 9 5 1 5 4 4 1 3 8 2 6 3 2 3 0 5 3 6 5 9 1 1 6 7 3 7 1 2 3 7 1 6 4 3 1
1 2 4 3 8 1 0 1 4 4 1 7 8 8 5 0 5 9 0 3 8 6 7 8 0 1 3 8 5 4 5 4 5 0 2 4 3 2 1 6 4 1 7 3 4 2 7 2 2 2 3 1 5 1 7 0 2 8 8
0 8 2 3 9 2 2 2 7 4 3 2 4 0 7 6 5 9 1 5 7 8 1 8 0 2 7 6 6 8 1 4 8 3 3 7 1 1 8 9 8 4 4 4 2 9 9 5 9 7 7 6 0 1 1 4

La dernière est “faite maison”. On a demandé à une personne de réciter mille chiffres successivement sans réfléchir. Voilà le résultat.

1 4 2 9 5 7 8 4 1 6 0 1 4 5 3 3 3 2 8 7 8 4 5 2 4 4 4 4 4 2 1 4 5 5 4 1 2 4 0 0 0 0 0 0 0 0 0 1 0 0 2 0 0 3 0 0 4 0 1
4 5 0 1 5 9 4 5 7 8 5 9 1 6 7 4 0 4 0 4 4 0 5 5 6 7 8 8 1 4 5 7 9 1 4 7 9 5 3 3 2 5 2 4 5 4 2 5 4 4 4 2 2 4 4 2 5 4 4
2 5 2 8 9 5 6 7 5 4 2 1 1 5 7 2 4 0 1 3 0 4 0 2 4 6 9 5 1 4 5 2 3 4 2 5 1 0 2 4 5 6 7 9 8 5 5 2 4 2 5 4 5 6 5 1 4 5 2
0 3 5 4 2 3 5 4 2 4 5 6 8 9 1 0 5 1 4 5 6 1 0 5 0 1 2 0 1 4 5 2 4 1 4 2 7 9 8 3 1 2 1 4 2 4 1 2 4 3 9 1 1 4 5 1 2 1 2
4 5 4 2 2 1 4 9 8 7 4 9 7 8 4 2 5 1 2 9 8 5 7 6 4 2 1 1 4 0 1 0 1 4 2 5 4 1 4 1 6 9 9 9 9 1 0 5 2 4 1 4 2 4 5 2 4 1 4
2 9 5 7 8 4 1 6 0 1 4 5 3 3 3 2 8 7 8 4 2 5 6 8 9 7 4 2 1 2 4 5 4 4 8 7 8 2 3 3 3 5 4 1 0 6 1 4 8 7 5 9 2 4 1 0 1 4 2
0 2 5 5 4 3 0 2 1 6 8 9 1 0 7 1 5 4 5 0 1 4 1 2 4 5 1 2 2 0 1 5 2 0 1 6 2 0 1 7 2 0 1 8 2 0 1 9 2 0 2 4 5 8 1 0 1 0 5
4 2 7 4 5 6 2 1 4 9 8 7 4 1 4 5 2 1 0 0 2 5 0 4 1 4 2 5 2 4 9 8 7 6 5 2 1 4 5 2 1 0 4 2 1 2 5 5 2 1 5 1 6 9 8 7 9 7 5
4 3 2 1 0 5 9 0 5 4 2 0 0 0 0 0 0 0 0 0 0 1 4 2 9 7 9 5 4 2 1 7 9 8 5 4 3 1 2 0 4 2 1 7 5 0 2 4 1 5 4 9 7 9 4 1 1 4 7
2 1 4 2 4 3 4 4 4 5 0 3 2 1 0 2 1 4 4 2 4 5 1 8 9 5 2 1 1 4 9 8 7 2 4 1 9 2 4 9 5 1 2 9 8 5 6 7 1 9 2 4 9 5 2 6 1 4 2
5 1 1 1 2 1 3 1 4 5 6 7 8 9 1 0 4 2 1 2 3 4 1 2 5 9 8 7 6 4 1 4 2 4 3 4 4 4 5 4 6 4 7 4 8 4 9 4 1 0 4 2 4 6 8 0 1 3 5
7 9 6 9 8 1 2 4 5 2 1 1 0 0 0 2 4 1 5 2 1 7 2 1 0 1 9 8 7 6 5 4 3 2 1 0 9 7 9 5 9 8 9 5 2 4 1 5 2 5 1 0 0 2 4 1 6 2 8
9 3 4 1 1 1 4 2 5 1 3 2 1 8 6 7 5 4 1 2 3 4 5 6 7 9 9 1 1 4 2 5 0 5 2 4 2 1 5 7 9 4 2 1 5 2 0 5 2 2 4 1 2 5 4 1 5 2 5
1 9 6 7 5 1 2 2 4 1 5 2 4 5 6 8 7 1 0 2 4 5 2 7 1 4 1 2 4 5 2 9 8 7 5 2 1 2 5 1 7 9 8 7 9 1 5 2 4 5 2 1 6 8 9 1 0 1 1
1 2 5 4 3 5 2 1 6 2 4 8 4 4 4 2 0 0 7 0 4 2 5 1 7 9 2 4 2 1 7 8 9 1 2 5 4 3 2 1 0 2 8 0 3 2 0 3 8 0 2 6 4 8 5 6 7 8 9
1 0 4 2 1 7 2 4 1 2 5 7 9 4 1 5 2 1 4 9 6 7 9 8 4 7 5 2 1 4 1 4 2 5 1 5 1 4 9 7 6 2 1 5 2 4 2 1 5 2 4 9 4 5 5 2 1 4 2
1 4 2 4 5 2 3 4 2 4 1 2 6 8 9 3 4 6 7 2 4 1 6 5 2 2 4 1 6 2 4 2 5 9 2 3 2 4 1 6 5 0 0 3 4 7 2 1 6 4 2 1 9 6 7 1

Le tableau qui suit donne, pour chacune des suites qui viennent d'être présentées, le nombre d'occurrences de chacun des dix chiffres dans la suite ainsi que de quelques nombres à deux chiffres pris au hasard : 00, 11, 32, 66, 69 et 77.

	0	1	2	3	4	5	6	7	8	9	00	11	32	66	69	77
1	104	89	98	107	86	112	100	108	99	97	10	8	11	11	12	16
2	107	94	108	106	99	92	87	112	89	106	12	10	9	11	9	13
3	93	116	103	102	93	97	94	95	101	106	7	16	9	11	6	9
4	85	82	80	87	96	164	80	83	113	130	8	6	4	4	8	8
5	66	177	177	148	77	77	77	67	67	67	3	25	25	5	5	4
6	73	171	132	95	104	93	83	83	84	82	5	25	16	10	5	3
7	92	161	167	39	183	131	45	61	51	70	26	13	10	0	4	0

1 dé à dix faces
2 générateur de nombres aléatoires
3 décimales de π
4 numéros de téléphone
5 nombres entiers par ordre croissant
6 nombre d'habitants par commune
7 récitation

Pour les trois premières suites, les occurrences sont proches des valeurs asymptotiques produites par une suite indépendante identiquement distribuée. Chaque chiffre apparaît avec une fréquence proche du dixième, tandis que les mots de deux lettres ont une fréquence proche du centième. On n'est pas surpris que les deux premières suites se comportent conformément à la loi des grands nombres. La question reste ouverte de démontrer qu'il en va vraiment de même pour la troisième suite constituée par les décimales de π. On ne sait même pas si tous les chiffres apparaissent une infinité de fois dans le développement décimal de π.

Les chiffres 5 et 9 sont sur-représentés dans la quatrième suite, sans qu'il soit possible d'en déterminer la raison. On pourrait s'attendre à ce que l'annuaire produise des valeurs aléatoires uniformément distribuées mais cet exemple ne permet pas de confirmer cette intuition. Il faudrait une analyse plus fine pour déterminer si c'est l'échantillon qui est particulier ou si un ordre se cache derrière la répartition des numéros.

La cinquième suite présente des disparités importantes, avec le chiffre 1 très largement représenté tandis que le 0 est peu fréquent. On n'est pas surpris que le chiffre 1 apparaisse souvent dans la liste des premiers entiers naturels. Le nombre dont les décimales sont obtenues en faisant la liste de tous les entiers par ordre croissant s'appelle la *constante de Champernowne*. On peut montrer que la fréquence de chacun des chiffres finit par converger vers un dixième, contrairement à ce que pourrait laisser penser les premiers termes de la suite. De manière étonnante, on peut même montrer que la constante de Champernowne est un nombre normal : pour tout entier $n > 0$, tous les mots constitués de n chiffres apparaissent dans la suite de ses décimales avec une fréquence égale à 10^{-n}.

La sixième suite présente aussi des variations importantes avec le chiffre 1 qui apparaît le plus fréquemment. Ce phénomène est parfois observé quand on étudie des données statistiques concernant des populations humaines et provient de la croissance exponentielle de ces populations. Il est relié à la *loi de Benford*. Cette loi est bien vérifiée par le nombre d'habitants des trente six mille communes de France et on l'observe déjà sur l'échantillon que nous avons considéré.

Finalement, la septième suite est loin d'être uniformément répartie, avec le chiffre 3 sous-représenté tandis que le 4 revient fréquemment. Elle montre à quel point il est difficile pour un être humain de simuler le hasard. L'absence de certains mots de longueur deux est typique dans ce genre d'expérimentation et permet de repérer aisément les suites qui sont le produit d'une intervention humaine plutôt que d'un procédé aléatoire.

4.4 Exercices

Dans les exercices qui suivent, $(X_n)_{n\in\mathbf{N}}$ est une suite de variables aléatoires indépendantes identiquement distribuées et $S_n = \sum_{k=1}^n X_k$.

exercice 1
On suppose les (X_n) intégrables. Montrer que $\dfrac{X_n}{n} \xrightarrow[n\to\infty]{} 0 \quad p.s.$

Si de plus l'espérance des X_k est strictement positive, alors $S_n \xrightarrow[n\to\infty]{p.s.} +\infty$.

exercice 2
On suppose les (X_n) de carrés intégrables. Montrer que les suites suivantes convergent presque sûrement vers une limite qu'on calculera.

$$\frac{1}{n}\sum_{k=1}^n X_k^2, \qquad \frac{1}{n}\sum_{k=1}^n X_k X_{k+1}, \qquad \frac{2}{n(n-1)}\sum_{1\le i<j\le n} X_i X_j, \qquad \frac{\sum_{k=1}^n X_k}{\sum_{k=1}^n X_k^2}.$$

exercice 3
Soit $x \in \mathbf{R}$. Montrer la convergence $\dfrac{1}{n}\#\{1 \le k \le n \mid X_k(\omega) \le x\} \xrightarrow[n\to\infty]{p.s.} F_{X_0}(x)$.

exercice 4
Soit $f : [0,1] \to \mathbf{R}$ continue. Montrer que la suite

$$u_n = \int_0^1 \dots \int_0^1 f\Big(\frac{x_1 + \dots + x_n}{n}\Big)\, dx_1 \dots dx_n$$

admet une limite que l'on précisera.

exercice 5
Soit $(p_k)_{k=1..r}$ des nombres réels strictement positifs tels que $\sum p_k = 1$. On suppose que la loi des X_n est donnée par $P(X_i = k) = p_k$. On considère les variables aléatoires $\Pi_n(\omega) = p_{X_1(\omega)} p_{X_2(\omega)} \dots p_{X_n(\omega)}$. Montrer que

$$\frac{1}{n}\ln(\Pi_n) \xrightarrow[n\to\infty]{} \sum_{k=1}^r p_k \ln(p_k) \quad p.s.$$

exercice 6
On suppose les X_n centrés de carrés intégrables. En inspectant la preuve de la loi forte des grands nombres, montrer qu'on peut trouver $\alpha > 0$ tel que

$$\frac{S_n}{n^{1-\alpha}} \xrightarrow[n\to\infty]{p.s.} 0.$$

Montrer l'égalité $\displaystyle\sum_{k=1}^n \frac{X_k}{k} = \sum_{k=1}^n \Big(\frac{1}{k} - \frac{1}{k+1}\Big) S_k + \frac{1}{n+1} S_n.$

En déduire que la série $\sum \frac{X_k}{k}$ converge presque sûrement.

Chapitre 5

Convergence de suites de variables aléatoires

5.1 Les différents types de convergence.

Les résultats précédents font appel à différentes notions de convergence. On va préciser ces notions et étudier les relations qu'elles entretiennent entre elles. Rappelons la définition des normes L^p, $p \geq 1$.

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé. Pour $p \in [1, +\infty[$, la norme L^p de la variable aléatoire $Y : \Omega \to \mathbf{R}$ est définie par

$$\|Y\|_p = (\int |Y|^p \, dP)^{1/p}.$$

La norme L^∞ de Y est définie par

$$\|Y\|_\infty = \inf\{C > 0 \mid \exists\ \Omega' \text{ tel que } P(\Omega') = 1 \text{ et } |Y(\omega)| \leq C \text{ pour tout } \omega \in \Omega'\}$$

Définition 11 *Soient Y_n, Y des variables aléatoires définies sur $(\Omega, \mathcal{T}, P)$ et $p \in [1, +\infty]$.*

– *La suite Y_n converge* en norme L^p *vers Y si*

$$\|Y_n - Y\|_p \xrightarrow[n\to\infty]{} 0.$$

– *La suite Y_n converge* en probabilité *vers Y si*

$$\forall\ \varepsilon > 0, \quad P(|Y_n - Y| > \varepsilon) \xrightarrow[n\to\infty]{} 0.$$

– *La suite Y_n converge* presque sûrement *vers Y si*

$$\textit{pour presque tout } \omega \in \Omega,\ Y_n(\omega) \xrightarrow[n\to\infty]{} Y(\omega).$$

– *La suite Y_n converge* en loi *vers Y si*

$$\textit{pour toute fonction } f : \mathbf{R} \to \mathbf{R} \textit{ continue bornée,} \qquad \int f \, dP_{Y_n} \xrightarrow[n\to\infty]{} \int f \, dP_Y.$$

Proposition 9 *Soient $p, q \in \mathbf{R}$ tels que $1 \leq p \leq q \leq \infty$. On a les implications*

$$CV\ L^\infty \Rightarrow CV\ L^q \Rightarrow CV\ L^p \Rightarrow CV\ L^1 \Rightarrow CV\ \text{en proba} \Rightarrow CV\ \text{en loi.}$$

$$CV\ L^\infty \Rightarrow CV\ p.s. \Rightarrow CV\ \text{en proba.}$$

$$CV\ L^\infty \Rightarrow CV\ \text{en proba} \Rightarrow CV\ p.s.\ \text{d'une sous-suite.}$$

Remarque
La convergence L^2 implique la convergence en probabilité. C'est comme cela que nous avons démontré la loi faible des grands nombres. Celle-ci affirme que $\frac{S_n}{n}$ converge vers $E(X_0)$ en probabilité, si les (X_i) sont indépendantes, identiquement distribuées. On avait obtenu ce résultat en montrant que $V(\frac{S_n}{n}) \xrightarrow[n\to\infty]{} 0$. D'après la relation suivante, cela est équivalent à la convergence L^2 :

$$V\Big(\frac{S_n}{n}\Big) = E\Big(\Big|\frac{S_n}{n} - E\Big(\frac{S_n}{n}\Big)\Big|^2\Big) = E\Big(\Big|\frac{S_n}{n} - E(X_0)\Big|^2\Big) = \Big\|\frac{S_n}{n} - E(X_0)\Big\|_2^2.$$

Démonstration de la proposition

• $CV\ L^q \Rightarrow CV\ L^p$ si $p \leq q$.

Démontrons l'égalité

$$\|Y\|_p \leq \|Y\|_q$$

en utilisant l'inégalité de Hölder : pour tout $p, q \geq 1$ tels que $1/p + 1/q = 1$,

$$\int |YZ|\, dP \leq \|Y\|_p\ \|Z\|_q.$$

On prend Y constant égal à 1 dans cette inégalité, auquel cas $\|Y\|_p = 1$ et $\|Z\|_1 \leq \|Z\|_q$. Ceci démontre le résultat pour $p = 1$. Pour p général, on remplace q par q/p et Z par Y^p, ce qui donne :

$$\int Y^p\, dP \leq \Big(\int Y^{pq/p} dP\Big)^{p/q},$$

$$\|Y\|_p \leq \|Y\|_q.$$

• $CV\ L^\infty \Rightarrow CV\ L^p$.

On a pour presque tout $\omega \in \Omega$, $|Y(\omega)| \leq \|Y\|_\infty$. En intégrant, on obtient

$$\|Y\|_p^p = \int |Y(\omega)|^p\ dP(\omega) \leq \int \|Y\|_\infty^p\ dP = \|Y\|_\infty^p.$$

• $CV\ L^1 \Rightarrow CV$ en proba

C'est une conséquence de l'inégalité de Markov. Si $Y_n \xrightarrow[n\to\infty]{L^1} Y$,

$$P(|Y_n - Y| > \varepsilon) \leq \frac{E(|Y_n - Y|)}{\varepsilon} = \frac{\|Y_n - Y\|_1}{\varepsilon} \xrightarrow[n\to\infty]{} 0$$

• CV $L^\infty \Rightarrow CV$ p.s.

Si $Y_n \xrightarrow[n\to\infty]{L^\infty} Y$, il existe $\Omega' \subset \Omega$ de probabilité 1 tel que

$$\sup_{\omega\in\Omega'} |Y_n(\omega) - Y(\omega)| \xrightarrow[n\to\infty]{} 0.$$

On en déduit, pour tout $\omega \in \Omega'$, $Y_n(\omega) \xrightarrow[n\to\infty]{} Y(\omega)$.

• CV en proba $\Rightarrow CV$ p.s. d'une sous-suite

Nous savons que pour tout $\varepsilon > 0$, $P(|Y_n - Y| > \varepsilon) \xrightarrow[n\to\infty]{} 0$.
Pour tout $k \in \mathbf{N}$, on peut donc trouver $n_k \in \mathbf{N}$ aussi grand qu'on veut, tel que $P(|Y_{n_k} - Y| > 1/k) \leq 1/2^k$. On a alors

$$\sum_{k=0}^{\infty} P(|Y_{n_k} - Y| > 1/k) < \infty$$

On applique le lemme de Borel-Cantelli : pour presque tout $\omega \in \Omega$, hormis pour un nombre fini d'indices k, $|Y_{n_k}(\omega) - Y(\omega)| < 1/k$. La suite Y_{n_k} converge vers Y presque sûrement.

• CV p.s. $\Rightarrow CV$ en proba

Nous avons les deux conditions suivantes :

– $\mathbf{1}_{(|Y_n - Y| > \varepsilon)}(\omega) \xrightarrow[n\to\infty]{} 0$ pour presque tout $\omega \in \Omega$ car $|Y_n(\omega) - Y(\omega)| \xrightarrow[n\to\infty]{} 0$.

– $|\mathbf{1}_{(|Y_n - Y| > \varepsilon)}| \leq \mathbf{1}_\Omega$ et $\mathbf{1}_\Omega$ est intégrable, ne dépend pas de n.

On peut appliquer le théorème de convergence dominée :

$$\lim_{n\to\infty} P(|Y_n - Y| > \varepsilon) = \lim_{n\to\infty} \int \mathbf{1}_{(|Y_n - Y| > \varepsilon)}\, dP = \int \lim_{n\to\infty} \mathbf{1}_{(|Y_n - Y| > \varepsilon)}\, dP = 0$$

L'implication *CV en proba* $\Rightarrow$ *CV en loi* sera démontrée dans la suite.

5.2 Fonction caractéristique et transformée de Fourier

Pour étudier plus en détail la convergence en loi, on va utiliser la notion de fonction caractéristique d'une variable aléatoire et de transformée de Fourier d'une mesure de probabilité.

Définition 12 *La* fonction caractéristique *d'une variable aléatoire* $Y : \Omega \to \mathbf{R}$ *est définie par*

$$\varphi_Y(t) = E(e^{itY}) = \int_\Omega e^{itY}\, dP = \int_{\mathbf{R}} e^{ity}\, dP_Y(y).$$

La transformée de Fourier *d'une mesure de probabilité μ définie sur la tribu des boréliens de* **R** *est définie par*

$$\widehat{\mu}(t) = \int_{\mathbf{R}} e^{itx}\, d\mu(x).$$

On a donc l'égalité $\varphi_Y(t) = \widehat{P}_Y(t)$.

Propriétés

– $|\varphi_Y(t)| \leq 1$ pour tout $t \in \mathbf{R}$,
– $\varphi_Y(0) = 1$,
– $t \mapsto \varphi_Y(t)$ est continue sur $\mathbf{R}$,
– si Y est intégrable, alors $t \mapsto \varphi_Y(t)$ est dérivable et $\varphi'_Y(0) = iE(Y)$,
– si Y est de carré intégrable, $t \mapsto \varphi_Y(t)$ est de classe C^2 et $\varphi''_Y(0) = -E(Y^2)$.

La continuité et la dérivabilité découlent des théorèmes de continuité et de dérivabilité sous le signe intégrable. Par exemple, si X est intégrable, on a la majoration

$$\Big|\frac{\partial}{\partial t} e^{itY}\Big| = |iY\, e^{itY}| \leq |Y|$$

ce qui implique $\varphi'_Y(t) = \frac{d}{dt}\Big(\int_\Omega e^{itY}\, dP\Big) = \int_\Omega \frac{\partial}{\partial t} e^{itY}\, dP = \int_\Omega iYe^{itY}\, dP$.

La loi d'une variable aléatoire est complètement caractérisée par sa fonction caractéristique.

Proposition 10 *Deux variables aléatoires qui ont même fonction caractéristique ont même loi : $\varphi_X = \varphi_Y$ implique $P_X = P_Y$.*

Ce théorème est admis et ne sera pas utilisé dans la suite. On passe maintenant à quelques calculs explicites de fonctions caractéristiques.

Cas discret

La variable aléatoire Y prend un nombre fini ou dénombrable de valeurs y_k, $k \in I$, avec $I = \{1, ..., n\}$ ou $I = \mathbf{N}$.

$$\varphi_Y(t) = E(e^{itY}) = \sum_{k \in I} e^{ity_k} P(Y = y_k).$$

• Loi de Bernoulli de paramètre $p \in [0, 1]$

Si Y obéit à une telle loi, $P(Y = 0) = 1 - p$, $P(Y = 1) = p$. On a alors
$\varphi_Y(t) = e^{it\times 0} P(Y = 0) + e^{it\times 1} P(Y = 1)$,

$$\varphi_Y(t) = 1 - p + pe^{it}.$$

• Loi uniforme sur $\{1, ..., n\}$, $n \in \mathbf{N}^*$

Si Y obéit à une telle loi, $P(Y=k)=1/n$ pour $k\in\{1,...,n\}$, ce qui implique $\varphi_Y(t)=\sum_{k=1}^n e^{itk}P(Y=k)=\sum_{k=1}^n \frac{1}{n}(e^{it})^k=\frac{1}{n}\,e^{it}\,\sum_{k=0}^{n-1}(e^{it})^k$.

$$\varphi_Y(t)=\frac{1}{n}\;e^{it}\;\frac{1-e^{itn}}{1-e^{it}}\qquad \text{si } t\notin 2\pi\mathbf{Z}.$$

Cas continu
La variable aléatoire Y est associée à la densité $f_Y:\mathbf{R}\to\mathbf{R}_+$ si bien que $P(Y\in A)=\int_A f_Y(y)\,dy$.

$$\varphi_Y(t)=E(e^{itY})=\int e^{ity}\,dP_Y(y)=\int e^{ity}f_Y(y)\,dy.$$

• Loi uniforme sur $[a,b]$, $a<b$.

$\varphi_Y(t)=\int_{\mathbf{R}} e^{ity}\frac{1}{b-a}\mathbf{1}_{[a,b]}(y)\,dy=\frac{1}{b-a}\int_a^b e^{ity}\,dy=\frac{1}{b-a}[\frac{e^{ity}}{it}]_a^b$.

$$\varphi_Y(t)=\frac{e^{itb}-e^{ita}}{it\,(b-a)}\qquad \text{si } t\neq 0.$$

• Loi exponentielle de paramètre $l>0$

$\varphi_Y(t)=\int_{\mathbf{R}} e^{ity}\;le^{-ly}\;\mathbf{1}_{\mathbf{R}}(y)\;dy=\int_0^{+\infty} le^{(it-l)y}\;dy=[le^{(it-l)y}/(it-l)]_0^{+\infty}$.

$$\varphi_Y(t)=\frac{l}{l-it}.$$

Remarque on utilise parfois à la place de la fonction caractéristique la notion de fonction génératrice.

Définition 13 *On considère l'ensemble des $z\in\mathbf{C}$ pour lesquels la fonction z^Y est intégrable. La fonction génératrice d'une variable aléatoire Y est définie sur cet ensemble par l'expression*

$$z\mapsto E(z^Y)$$

Attention, elle n'est pas forcément définie pour tout $z\in\mathbf{C}$, la fonction $z\mapsto z^Y$ n'étant pas forcément intégrable. Lorsque $z=e^{it}$, elle est bien intégrable et on retrouve la fonction caractéristique de la variable Y.

5.3 Convergence en loi

Rappelons que Y_n converge en loi vers Y si pour toute fonction $f:\mathbf{R}\to\mathbf{R}$ continue bornée,

$$\int f\,dP_{Y_n}\xrightarrow[n\to\infty]{}\int f\,dP_Y.$$

Définition 14 *Soit μ_n et μ des mesures de probabilité définies sur la tribu des boréliens de $\mathbf{R}$. Nous dirons que μ_n converge étroitement vers μ si pour toute fonction $f : \mathbf{R} \to \mathbf{R}$ continue bornée, $\int f\, d\mu_n \xrightarrow[n\to\infty]{} \int f\, d\mu$.*

La suite Y_n converge en loi vers Y si et seulement si P_{Y_n} converge étroitement vers P_Y. On va relier la convergence en loi à la convergence simple des fonctions caractéristiques, dans le but de démontrer le théorème de la limite centrée.

Théorème 6 *Soit μ, μ_n, $n \in \mathbf{N}$, des mesures de probabilité définies sur la tribu des boréliens de $\mathbf{R}$. Les propriétés suivantes sont équivalentes :*

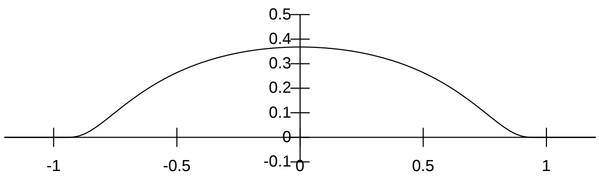
- $\int f d\mu_n \xrightarrow[n\to\infty]{} \int f d\mu$ *pour toute fonction f continue bornée,*
- $\int f d\mu_n \xrightarrow[n\to\infty]{} \int f d\mu$ *pour toute fonction f C^∞ à support compact,*
- $\int f d\mu_n \xrightarrow[n\to\infty]{} \int f d\mu$ *pour toute fonction f de la forme e^{itx}, $t \in \mathbf{R}$.*

Le premier point correspond à la convergence étroite des μ_n vers μ. Le dernier point correspond à la convergence des transformées de Fourier des μ_n. On en déduit le corollaire suivant.

Corollaire 3 *Soit μ, μ_n des mesures de probabilité définies sur la tribu des boréliens de $\mathbf{R}$. Si pour tout $t \in \mathbf{R}$, $\widehat{\mu}_n(t) \xrightarrow[n\to\infty]{} \widehat{\mu}(t)$, alors μ_n converge vers μ étroitement.*
Soit (Y_n) une suite de variables aléatoires. Si pour tout $t \in \mathbf{R}$, $\varphi_{Y_n}(t) \xrightarrow[n\to\infty]{} \varphi_Y(t)$, alors Y_n converge vers Y en loi.

Rappelons que $f : \mathbf{R} \to \mathbf{R}$ est à *support compact* s'il existe $A > 0$ tel que f est nulle hors de $[-A, A]$. Un exemple de fonction C^∞ à support compact est donné par

$$f(x) = e^{-\frac{1}{1-x^2}}\ \mathbf{1}_{[-1,1]}(x).$$



Pour démontrer le théorème, nous allons avoir besoin de la formule d'inversion de Fourier. Soit $f : \mathbf{R} \to \mathbf{R}$ une fonction intégrable par rapport à la mesure de Lebesgue. Sa transformée de Fourier est définie par

$$\widehat{f}(t) = \int_{\mathbf{R}} e^{-itx} f(x)\, dx.$$

On montre que cette fonction est continue en appliquant le théorème de continuité sous le signe intégral.

Théorème 7 (formule d'inversion de Fourier) *Soit f une fonction C^∞ à support compact. Alors*

$$f(x) = \frac{1}{2\pi} \int_{\mathbf{R}} e^{itx} \widehat{f}(t)\, dt \quad \textit{pour tout } x \in \mathbf{R}.$$

La preuve du théorème est donnée en annexe, sous des hypothèses un peu plus générales. La formule d'inversion de Fourier implique la relation suivante entre μ et sa transformée de Fourier.

Corollaire 4 *Soit μ une mesure de probabilité définie sur la tribu des boréliens de $\mathbf{R}$ et $f : \mathbf{R} \to \mathbf{R}$ une fonction C^∞ à support compact. Alors*

$$\int f(x)\, d\mu(x) = \frac{1}{2\pi} \int \widehat{f}(t)\, \widehat{\mu}(t)\, dt.$$

Preuve du corollaire

$$\begin{aligned}
\int_{\mathbf{R}} f(x)\, d\mu(x) &= \int_{\mathbf{R}} \frac{1}{2\pi} \int_{\mathbf{R}} e^{itx} \widehat{f}(t)\, dt \quad d\mu(x) \\
&= \frac{1}{2\pi} \int_{\mathbf{R}} \int_{\mathbf{R}} e^{itx} \widehat{f}(t)\, d\mu(x)\, dt \\
&= \frac{1}{2\pi} \int_{\mathbf{R}} \widehat{f}(t) \Big(\int_{\mathbf{R}} e^{itx} d\mu(x) \Big)\, dt \\
&= \frac{1}{2\pi} \int_{\mathbf{R}} \widehat{f}(t)\, \widehat{\mu}_n(t)\, dt.
\end{aligned}$$

Ici on a utilisé le théorème de Fubini pour intervertir les deux intégrales. Pour justifier l'emploi de ce théorème, remarquons que l'intégrale $\iint_{\mathbf{R}^2} |e^{itx} \widehat{f}(t)|\, d\mu(x)\, dt$ est finie :

$$\int_{\mathbf{R}} \int_{\mathbf{R}} |e^{itx} \widehat{f}(t)|\, d\mu(x)\, dt = \int_{\mathbf{R}} d\mu(x) \int_{\mathbf{R}} |\widehat{f}(t)|\, dt = \int_{\mathbf{R}} |\widehat{f}(t)|\, dt < \infty.$$

La preuve est terminée.

On commence par démontrer que $\int f d\mu_n \longrightarrow \int f d\mu$ pour toute f C^∞ à support compact si $\widehat{\mu}_n(t) \longrightarrow \widehat{\mu}(t)$ pour tout $t \in \mathbf{R}$.

D'après le corollaire précédent,

$$\int f(x)\, d\mu_n(x) = \frac{1}{2\pi} \int \widehat{f}(t)\, \widehat{\mu}_n(t)\, dt,$$

$$\int f(x)\, d\mu(x) = \frac{1}{2\pi} \int \widehat{f}(t)\, \widehat{\mu}(t)\, dt.$$

Il suffit d'appliquer le théorème de convergence dominée pour conclure :

$$\int \widehat{f}(t)\, \widehat{\mu}_n(t)\, dt \xrightarrow[n\to\infty]{} \int \widehat{f}(t)\, \widehat{\mu}(t)\, dt.$$

L'emploi du théorème de convergence dominée est justifié ici car $\widehat{\mu}_n(t) \xrightarrow[n\to\infty]{} \widehat{\mu}(t)$ pour tout $t \in \mathbf{R}$ par hypothèse et $\widehat{f}\,\widehat{\mu}_n$ est majorée par $\widehat{f}$ qui est intégrable.

On cherche à présent à démontrer que si $\int f d\mu_n \longrightarrow \int f d\mu$ pour toute fonction C^∞ à support compact, il en va de même pour toute fonction continue bornée.

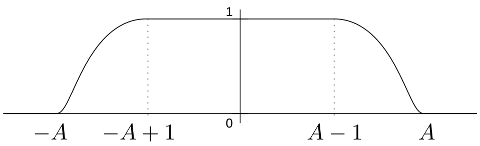
Lemme 4 *Pour tout $\varepsilon > 0$, il existe $A > 0$ tel que pour tout $n \in \mathbf{N}$,*

$$\mu_n([-A, A]) \geq 1 - \varepsilon.$$

Une suite de mesures de probabilité qui vérifie cette propriété est dite *tendue*.

Preuve du lemme
Soit g une fonction C^∞ telle que

- $0 \leq g \leq 1$,
- $g = 1$ sur $[-A+1, A-1]$,
- $g = 0$ sur $[-A, A]^c$.



Fixons $\varepsilon > 0$. Comme $\mu([-A+1, A-1]) \xrightarrow[A\to\infty]{} \mu(\mathbf{R}) = 1$, on peut choisir A_0 tel que $\mu[-A_0+1, A_0-1] > 1-\varepsilon$.

$$\mu_n([-A_0, A_0]) \geq \int g d\mu_n \xrightarrow[n\to\infty]{} \int g d\mu \geq \mu([-A_0+1, A_0-1]) > 1-\varepsilon.$$

On peut donc trouver $n_0 \in \mathbf{N}$ tel que pour tout $n \geq n_0$, $\mu_n([-A_0, A_0]) \geq 1-\varepsilon$.

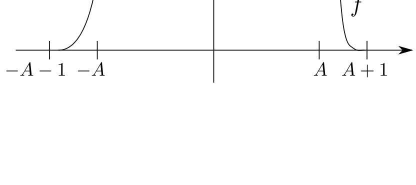
De plus, pour chaque $k \in \{0, ..., n_0\}$, on peut trouver un ensemble A_k tel que $\mu_k([-A_k, A_k]) \geq 1-\varepsilon$. Pour tout A supérieur à $max\{A_0, ..., A_{n_0}\}$, on a

$$\forall n \in \mathbf{N}, \quad \mu_n([-A, A]) \geq 1-\varepsilon.$$

Le lemme est démontré.

Soit f continue bornée. Sur $[-A-2, A+2]$, on peut approcher f uniformément par une fonction C^∞ en faisant appel au théorème de Stone-Weierstraß ou en convolant avec une fonction C^∞. Cette approximation peut être prolongée en une fonction C^∞ à support compact définie sur $\mathbf{R}$ tout entier en la multipliant par une fonction de classe C^∞, comprise entre 0 et 1, qui vaut 1 sur $[-A, A]$ et 0 hors de $[-A-1, A+1]$. Pour tout $\varepsilon > 0$ on peut donc trouver $\tilde{f}$ C^∞ à support compact telle que

$$\sup_{x\in[-A,A]} |f(x) - \tilde{f}(x)| < \varepsilon.$$



On veut montrer que $|\int f d\mu_n - \int f d\mu|$ est inférieur à ε pour tout n suffisamment grand. On décompose comme suit :

$$\Big|\int f d\mu_n - \int f d\mu\Big| \leq \Big|\int f d\mu_n - \int \tilde{f} d\mu_n\Big| + \Big|\int \tilde{f} d\mu_n - \int \tilde{f} d\mu\Big| + \Big|\int \tilde{f} d\mu - \int f d\mu\Big|$$

• Comme $\tilde{f}$ est C^∞ à support compact, $\int \tilde{f}\, d\mu_n \longrightarrow \int \tilde{f}\, d\mu$. On peut trouver $N \in \mathbf{N}$ tel que pour tout $n \geq \mathbf{N}$, $|\int \tilde{f} d\mu_n - \int \tilde{f} d\mu| < \varepsilon$.

• $$\begin{aligned} |\int f d\mu_n - \tilde{f} d\mu_n| &\leq \int_{[-A,A]} |f - \tilde{f}|\, d\mu_n + \int_{[-A,A]^c} |f - \tilde{f}|\, d\mu_n \\ &\leq \varepsilon\, \mu_n([-A,A]) + (\sup_{\mathbf{R}} |f| + \sup_{\mathbf{R}} |\tilde{f}|)\, \mu_n([-A,A]^c) \\ &\leq \varepsilon + (\sup_{\mathbf{R}} |f| + \sup_{\mathbf{R}} |\tilde{f}|)\, \varepsilon. \end{aligned}$$

Cette majoration est valide pour tout $n \in \mathbf{N}$.

• Le terme $|\int f - \tilde{f}\, d\mu\,|$ se majore de la même façon.

Finalement, on remarque que $\sup |\tilde{f}| \leq \sup |f| + \varepsilon \leq \sup |f| + 1$ sur $\mathbf{R}$ par construction. On a donc, pour tout $n \geq N$,

$$\Big|\int f\, d\mu_n - \int f\, d\mu\Big| \leq (4 + 2 \sup |f|)\, \varepsilon.$$

Le théorème est démontré.

Proposition 11 *Soient μ_n, μ des mesures de probabilités définies sur la tribu des boréliens de $\mathbf{R}$. On suppose que μ_n converge étroitement vers μ. Alors pour tout a, $b \in \mathbf{R}$ tels que $\mu(\{a\}) = 0$ et $\mu(\{b\}) = 0$, on a*

$$\mu_n([a,b]) \xrightarrow[n\to\infty]{} \mu([a,b]).$$

De même, pour tout $x \in \mathbf{R}$ tel que $\mu(\{x\}) = 0$,

$$\mu_n([x,+\infty[) \xrightarrow[n\to\infty]{} \mu([x,+\infty[),$$

$$\mu_n(]-\infty,x]) \xrightarrow[n\to\infty]{} \mu(]-\infty,x]).$$

Appliquons cette proposition à une suite de variables aléatoires.

Corollaire 5 *Soient $(X_n)_{n\in\mathbf{N}}$ et X des variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{T}, P)$ telles que X_n converge en loi vers X. Alors pour tout $a, b \in \mathbf{R}$ tels que $P(X = a) = P(X = b) = 0$,*

$$P(a \leq X_n \leq b) \xrightarrow[n\to\infty]{} P(a \leq X \leq b).$$

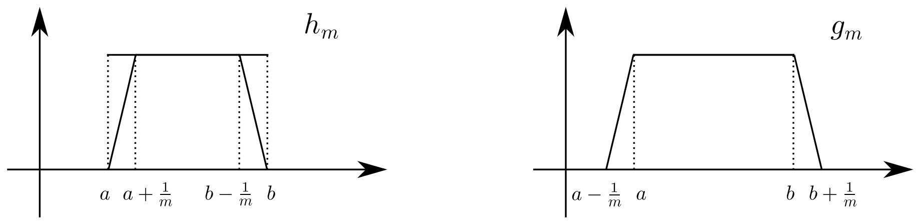
De plus, les fonctions de répartition des X_n convergent vers la fonction de répartition de X en tout point $x \in \mathbf{R}$ tel que $P(X = x) = 0$:

$$F_{X_n}(x) \xrightarrow[n\to\infty]{} F_X(x) \quad si \quad P(X = x) = 0.$$

Remarque On peut démontrer que la convergence des fonctions de répartition en tout point x tel que $P(X = x) = 0$ est en fait équivalente à la convergence en loi de la suite X_n vers X.

Preuve
Il s'agit d'approcher $\mathbf{1}_{[a,b]}$ par des fonctions continues bornées. Soit h_m la fonction continue bornée, affine par morceaux telle que :
• $h_m = 1$ sur $[a + \frac{1}{m}, b - \frac{1}{m}]$,
• $h_m = 0$ hors de $[a, b]$.
• la pente de h_m vaut m sur $[a, a + \frac{1}{m}]$ et $-m$ sur $[b - \frac{1}{m}, b]$.

Soit g_m la fonction continue bornée, affine par morceaux, telle que
• $g_m = 1$ sur $[a, b]$,
• $g_m = 0$ hors de $[a - \frac{1}{m}, b + \frac{1}{m}]$,
• la pente de g_m vaut m sur $[a - \frac{1}{m}, a]$ et $-m$ sur $[b, b + \frac{1}{m}]$.



Nous avons la majoration $0 \le g_m - h_m \le \mathbf{1}_{[a-\frac{1}{m}, a+\frac{1}{m}]} + \mathbf{1}_{[b-\frac{1}{m}, b+\frac{1}{m}]}$ si bien que

$$0 \le \int g_m - h_m \, d\mu \le \mu\left(\left[a - \frac{1}{m}, a + \frac{1}{m}\right]\right) + \mu\left(\left[b - \frac{1}{m}, b + \frac{1}{m}\right]\right)$$

Ce dernier terme converge vers $\mu(\{a\}) + \mu(\{b\})$, quantité qui est nulle par hypothèse. Fixons $\varepsilon > 0$ et choisissons $n \in \mathbf{N}$ tel que $\int g_m - h_m \, d\mu \le \varepsilon$.

La suite μ_n converge vers μ étroitement et $h_m \le \mathbf{1}_{[a,b]} \le g_m$, nous avons donc pour tout n suffisamment grand,

$$\int h_m \, d\mu - \varepsilon \le \int h_m \, d\mu_n \le \mu_n([a, b]) \le \int g_m \, d\mu_n \le \int g_m \, d\mu + \varepsilon.$$

Nous avons aussi, en vertu des inégalités $h_m \le \mathbf{1}_{[a,b]} \le g_m$ et $\int g_m - h_m \, d\mu \le \varepsilon$,

$$\int g_m \, d\mu - \varepsilon \le \int h_m \, d\mu \le \mu([a, b]) \le \int g_m \, d\mu \le \int h_m \, d\mu + \varepsilon,$$

ce qui donne le résultat recherché :

$$\mu([a, b]) - 2\varepsilon \le \mu_n([a, b]) \le \mu([a, b]) + 2\varepsilon.$$

On termine ce chapitre par la preuve d'un résultat énoncé précédemment.

Théorème 8 *Soit X_n, X des variables aléatoires. Si X_n converge vers X en probabilité, alors X_n converge vers X en loi.*

Preuve
Soit $f : \mathbf{R} \to \mathbf{R}$ C^∞ à support compact. Par le théorème des valeurs intermédiaires, pour tout $x, y \in \mathbf{R}$,

$$|f(x) - f(y)| \leq \sup_{\mathbf{R}} |f'| \quad |x-y|.$$

On veut montrer que la différence $\int f dP_{X_n} - \int f dP_X$ tend vers 0 quand $n \longrightarrow \infty$.

$$\begin{aligned}
&| \textstyle\int f\, dP_{X_n} - \int f\, dP_X| \\
&\qquad = \quad | \textstyle\int f(X_n)\, dP - \int f(X)\, dP| \\
&\qquad \leq \quad \textstyle\int |f(X_n) - f(X)|\, dP \\
&\qquad \leq \quad \textstyle\int_{|X_n - X| > \delta} |f(X_n) - f(X)|\, dP + \int_{|X_n - X| < \delta} |f(X_n) - f(X)|\, dP \\
&\qquad \leq \quad 2 \sup_{\mathbf{R}} |f|\, P(|X_n - X| > \delta) + \sup_{\mathbf{R}} |f'|\ \delta.
\end{aligned}$$

Comme X_n converge vers X en probabilité, $P(|X_n - X| > \delta) \xrightarrow[n\to\infty]{} 0$.
Pour tout $\varepsilon > 0$, on choisit δ telle que $\sup |f'|\, \delta < \varepsilon/2$. Il existe alors $N \in \mathbf{N}$ tel que pour tout $n \geq N$,

$$P(|X_n - X| > \delta) \leq \frac{\varepsilon}{4 \sup_{\mathbf{R}} |f|}$$

ce qui implique $| \int f dP_{X_n} - \int f dP_X| < \varepsilon$. Le théorème est démontré.

5.4 Exercices

exercice 1
Montrer qu'une suite de variables aléatoires (X_n) converge en loi vers zéro si et seulement si les lois P_{X_n} convergent étroitement vers la mesure de Dirac en zéro.

exercice 2
Soit g une fonction C^1 à support compact et X une variable aléatoire. Montrer que

$$E(g(X)) = -\int_{\mathbf{R}} F_X(t)\, g'(t)\, dt.$$

Soit (X_n) une suite de variables aléatoires telle que $F_{X_n}(x)$ converge vers $F_X(x)$ pour tout x satisfaisant $P(X = x) = 0$. Montrer que X_n converge en loi vers X.

exercice 3
Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé. On considère sur l'ensemble des couples de fonctions mesurables de Ω dans $\mathbf{R}$ l'expression $d(X,Y) = E\Big(\frac{|X-Y|}{1+|X-Y|}\Big)$.
Montrer que cela définit une distance si on identifie les fonctions qui coïncident presque partout. Montrer que la convergence relativement à cette distance est équivalente à la convergence en loi : $X_n \xrightarrow[n\to\infty]{proba} X \;\Leftrightarrow\; d(X_n, X) \xrightarrow[n\to\infty]{} 0.$

exercice 4
Soit (A_i) une suite d'événements indépendants dans leur ensemble.
– Montrer que $\mathbf{1}_{A_i} \xrightarrow[i\to\infty]{proba} 0$ si et seulement si $P(A_i) \to 0$.
– Montrer que $\mathbf{1}_{A_i} \xrightarrow[i\to\infty]{p.s.} 0$ si et seulement si la série $\sum_{i\in\mathbf{N}} P(A_i)$ est convergente.

exercice 5
Étudier la convergence en loi des suites $(X_n)_{n\in\mathbf{N}^*}$ suivantes :
– $P(X_n = 1 + \frac{1}{n}) = \frac{1}{2} = P(X_n = 1 - \frac{1}{n})$.
– $P(X_n = 0) = 1 - \frac{1}{n}$ et $P(X_n = 1) = \frac{1}{n}$.
– X_n de loi uniforme sur $\{0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}, 1\}$.

exercice 6
Soit $S_{n,p}$ une variable aléatoire qui obéit à une loi binomiale de paramètres n, p.
– Calculer la fonction caractéristique de $S_{n,p}$. Cette fonction est notée $\varphi_{n,p}$.
– La suite $\varphi_{n,\frac{1}{n}}$ converge-t-elle simplement lorsque n tend vers $+\infty$?
– Calculer la fonction caractéristique de la loi de Poisson de paramètre $\lambda > 0$.
– Que peut-on dire de la suite $S_{n,\frac{1}{n}}$?

exercice 7
Soit (X_n) et (Y_n) deux suites de variables aléatoires et $a \in \mathbf{R}$. Montrer que

$$X_n \xrightarrow[n\to\infty]{loi} a \text{ si et seulement si } X_n \xrightarrow[n\to\infty]{proba} a.$$

$$X_n \xrightarrow[n\to\infty]{loi} X \text{ et } Y_n \xrightarrow[n\to\infty]{loi} 0 \text{ implique } X_n + Y_n \xrightarrow[n\to\infty]{loi} X.$$

Chapitre 6

Théorème de la limite centrée

Pour démontrer le théorème de la limite centrée, nous allons utiliser la caractérisation de la convergence en loi par le biais des fonctions caractéristiques. On commence par calculer la fonction caractéristique de la loi normale.

6.1 Fonction caractéristique de la loi normale

Théorème 9 *Soit Y une variable aléatoire qui obéit à une loi normale centrée normalisée $(m = 0, \sigma = 1)$. Sa densité est donnée par $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ et sa fonction caractéristique vaut*

$$\varphi_Y(t) = e^{-t^2/2}.$$

Preuve
Par définition, $\varphi_Y(t) = \int_{\mathbf{R}} e^{ity} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$.

On sait que $e^{ity} = \sum_{k=0}^{+\infty} \frac{(ity)^k}{k!}$ pour $y \in \mathbf{R}$. Remplaçons dans l'intégrale précédente.

$$\begin{aligned}
\int_{\mathbf{R}} e^{ity} e^{-y^2/2}\, dy &= \int_{\mathbf{R}} \sum_{k=0}^{+\infty} \frac{(ity)^k}{k!}\, e^{-y^2/2}\, dy \\
&= \sum_{k=0}^{+\infty} \int_{\mathbf{R}} \frac{(it)^k}{k!}\, y^k\, e^{-y^2/2}\, dy \\
&= \sum_{k=0}^{+\infty} \frac{(it)^k}{k!} \int_{\mathbf{R}} y^k\, e^{-y^2/2}\, dy.
\end{aligned}$$

Pour justifier l'interversion signe somme intégrale, il faut vérifier que la quantité $\int_{\mathbf{R}} \sum_0^\infty |\frac{(ity)^k}{k!} e^{y^2/2}|\, dy$ est finie.

$$\int_{-\infty}^{+\infty} e^{-y^2/2} \sum_0^\infty \frac{|ty|^k}{k!}\, dy = \int_{-\infty}^{+\infty} e^{-y^2/2} e^{|ty|}\, dy < +\infty.$$

Il faut maintenant calculer $I_k = \int_{\mathbf{R}} y^k e^{-y^2/2}\, dy$.

Lorsque k est impair, la fonction $y \mapsto y^k e^- y^2/2$ est une fonction impaire, si bien que son intégrale est nulle : $I_{2l+1} = 0$ pour tout $l \in \mathbf{N}$. Pour k pair, $k = 2l$, on fait une intégration par partie pour obtenir une relation de récurrence.

$$I_{2l+2} = \int y^{2l+1} y e^{-y^2/2} dy = [y^{2l+1}(-e^{-y^2/2})]_{-\infty}^{+\infty} + \int (2l+1) y^{2l} e^{-y^2/2} dy = (2l+1) I_{2l}$$

Nous savons que $I_0 = \int_{\mathbf{R}} e^{-y^2/2} dy = \sqrt{2\pi}$, si bien que

$$I_{2l} = (2l-1)(2l-3)...3 \times 1 \times \sqrt{2\pi}.$$

La quantité I_{2l} est divisée par $(2l)!$ dans l'expression de la fonction caractéristique.

$$\begin{aligned} \frac{I_{2l}}{(2l)!} &= \frac{(2l-1)(2l-3)...1}{(2l)(2l-1)(2l-2)(2l-3)...1} \sqrt{2\pi} \\ &= \frac{1}{(2l)(2l-2)(2l-4)...2} \sqrt{2\pi} \\ &= \frac{1}{2^l} \frac{1}{l(l-1)(l-2)...1} \sqrt{2\pi} \\ &= \frac{1}{2^l l!} \sqrt{2\pi}. \end{aligned}$$

Nous pouvons calculer φ_Y :

$$\varphi_Y(t) = \frac{1}{\sqrt{2\pi}} \int e^{ity} e^{-y^2/2} dy = \sum_{l=0}^{+\infty} (it)^{2l} \frac{I_{2l}}{(2l)! \sqrt{2\pi}} = \sum_{l=0}^{\infty} \frac{(-t^2)^l}{2^l\, l!} = e^{-t^2/2}.$$

La formule est démontrée.

6.2 Théorème de la limite centrée

Théorème 10 *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilité, $(X_n)_{n\in\mathbf{N}}$ une suite de variables aléatoires indépendantes, identiquement distribuées, de carrés intégrables et de variance non nulle. On pose $S_n = \sum_{i=1}^n X_i$. Alors la loi de la variable aléatoire*

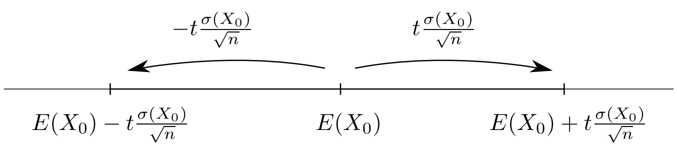
$$\frac{\sqrt{n}}{\sigma(X_0)} \Big(\frac{S_n}{n} - E(X_0) \Big)$$

converge étroitement vers une loi normale d'espérance nulle et d'écart-type 1. En particulier, pour tout intervalle $[a, b] \in \mathbf{R}$,

$$P\Big(a \leq \frac{\sqrt{n}}{\sigma(X_0)} \Big(\frac{S_n}{n} - E(X_0) \Big) \leq b \Big) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Remarque
L'événement ci-dessus peut s'écrire comme suit :

$$\begin{aligned}\Big(a \leq \tfrac{\sqrt{n}}{\sigma(X_0)}(\tfrac{S_n}{n} - E(X_0)) \leq b\Big) &= \Big(\tfrac{\sqrt{n}}{\sigma(X_0)}(\tfrac{S_n}{n} - E(X_0)) \in [a,b]\Big) \\ &= \Big(E(X_0) + a\,\tfrac{\sigma(X_0)}{\sqrt{n}} \leq \tfrac{S_n}{n} \leq E(X_0) + b\,\tfrac{\sigma(X_0)}{\sqrt{n}}\Big)\end{aligned}$$



Lorsque n est grand, la probabilité que $\frac{S_n}{n}$ soit dans l'intervalle

$$\left[E(X_0) - t\,\frac{\sigma(X_0)}{\sqrt{n}}, E(X_0) + t\,\frac{\sigma(X_0)}{\sqrt{n}}\right]$$

est proche de $\frac{1}{\sqrt{2\pi}}\int_{-t}^{t} e^{-x^2/2}dx$.

- Pour $t = 1,96$, $\frac{1}{\sqrt{2\pi}}\int_{-t}^{t} e^{-x^2/2}dx = 0,95$.
- Pour $t = 2,58$, $\frac{1}{\sqrt{2\pi}}\int_{-t}^{t} e^{-x^2/2}dx = 0,99$.

Il y a donc à peu près 99% de chance, lorsque n est grand, d'avoir une moyenne empirique $\frac{S_n}{n}$ dans l'intervalle $\Big[E(X_0) - 2,58\,\frac{\sigma(X_0)}{\sqrt{n}}, E(X_0) + 2,58\,\frac{\sigma(X_0)}{\sqrt{n}}\Big]$.

Il est d'usage de noter la convergence des lois d'une suite de variables aléatoires Y_n vers la loi normale de paramètres m, σ comme suit :

$$Y_n \xrightarrow[n\to\infty]{loi} \mathcal{N}(m, \sigma^2)$$

Dans le cas où les X_i sont indépendantes identiquement distribuées d'espérance nulle et d'écart-type égal à un, le théorème de la limite centrée peut se résumer comme suit :

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{loi} \mathcal{N}(0,1).$$

Preuve du théorème
Quitte à remplacer les X_i par $X_i - E(X_i)$, on peut supposer que les X_i sont centrées : $E(X_i) = 0$. Quitte à diviser par $\sigma(X_i)$, on peut aussi supposer que $\sigma(X_i) = 1$. On veut montrer que la loi de $\frac{S_n}{\sqrt{n}}$ converge vers la loi normale. Il suffit donc de montrer que

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) \xrightarrow[n\to\infty]{} e^{-t^2/2} \quad \text{pour tout } t \in \mathbf{R}.$$

Calculons :

$$\begin{aligned}
\varphi_{\frac{S_n}{\sqrt{n}}}(t) &= E\Big(e^{it\frac{S_n}{\sqrt{n}}}\Big) \\
&= E\Big(e^{\frac{it}{\sqrt{n}}\sum_1^n X_k}\Big) \\
&= E\Big(\prod_{k=1}^n e^{\frac{it}{\sqrt{n}}X_k}\Big) \\
&= \prod_{k=1}^n E\Big(e^{\frac{it}{\sqrt{n}}X_k}\Big) && \text{par indépendance,} \\
&= E\Big(e^{\frac{it}{\sqrt{n}}X_0}\Big)^n && \text{car les } X_i \text{ sont de même loi,} \\
&= \varphi_{X_0}\Big(\frac{t}{\sqrt{n}}\Big)^n.
\end{aligned}$$

Pour calculer la limite de cette expression quand n tend vers $+\infty$, on fait un développement limité. Comme X_0 est de carré intégrable, φ_{X_0} est C^2 et on a :

$$\varphi_{X_0}(t) = \int e^{itX_0}dP, \quad \varphi'_{X_0}(t) = \int iX_0\, e^{itX_0}dP, \quad \varphi''_{X_0}(t) = \int -X_0^2\, e^{itX_0}dP,$$

$$\varphi_{X_0}(0) = 1, \quad \varphi'_{X_0}(0) = iE(X_0) = 0, \quad \varphi''_{X_0}(0) = -E(X_0^2) = -1.$$

D'après la formule de Taylor, $\varphi_{X_0}(x) = 1 - \frac{x^2}{2} + x^2\,\varepsilon_0(x)$, avec $\varepsilon_0(x) \longrightarrow 0$ lorsque $x \to 0$. Ceci implique :

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \varphi_{X_0}\Big(\frac{t}{\sqrt{n}}\Big)^n = \Big(1 - \frac{t^2}{2n} + \frac{t^2}{2n}\,\varepsilon_0(\frac{t}{\sqrt{n}})\Big)^n,$$

$$n\,\ln\Big(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon_0(\frac{t}{\sqrt{n}})\Big) = n\Big(-\frac{t^2}{2n} + \frac{t^2}{n}\varepsilon_0(\frac{t}{\sqrt{n}}) + \frac{1}{n}\varepsilon_1(\frac{1}{\sqrt{n}})\Big) = -\frac{t^2}{2} + \varepsilon_2(\frac{1}{\sqrt{n}}),$$

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \Big(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon_0(\frac{t}{\sqrt{n}})\Big)^n = e^{-\frac{t^2}{2}+\varepsilon_2(\frac{1}{\sqrt{n}})} \xrightarrow[n\to\infty]{} e^{-\frac{t^2}{2}}.$$

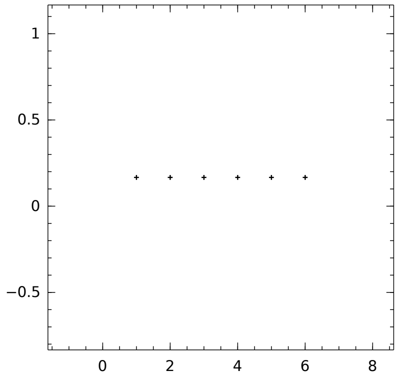
Le théorème de la limite centrée est démontré.

6.3 Illustration numérique

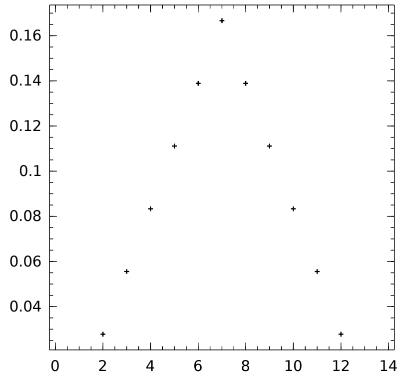
Nous allons illustrer le théorème de la limite centrée à l'aide des graphes des fréquences de la suite S_n.

Soit X une variable aléatoire discrète. Le *graphe des fréquences* de X correspond au graphe de la fonction $x \mapsto P(X = x)$, où x varie parmi les nombres réels tels que $P(X = x) > 0$.

Considérons le lancer d'un dé à six faces, modélisé par une variable aléatoire X_0 qui suit une loi uniforme sur l'ensemble $\{1, 2, 3, 4, 5, 6\}$: $P(X_0 = k) = 1/6$ pour k entier compris entre 1 et 6. On répète le lancer n fois, $n \in \mathbf{N}^*$, ce qui se décrit par une suite de variables aléatoires X_1, ... X_n indépendantes entre elles et ayant même loi que X_0. On pose $S_n = \sum_{k=1}^n X_k$. Voici les graphes des fréquences de X_0 et $S_2 = X_1 + X_2$.



Graphe des fréquences de X_0

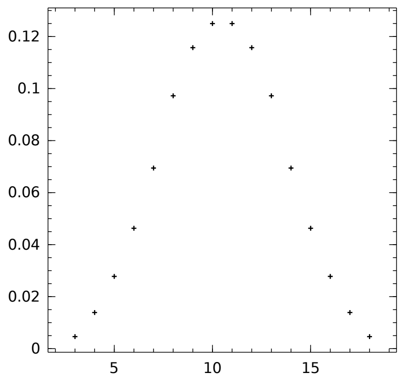


Graphe des fréquences de $S_2 = X_1 + X_2$

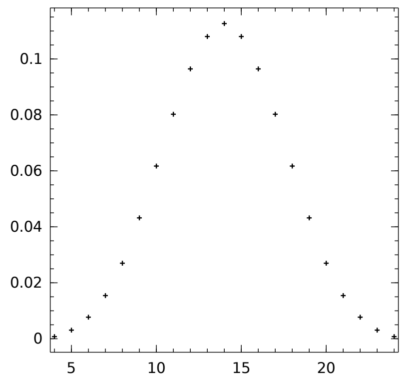
On calcule le graphe des fréquences de S_n pour tout n par récurrence en utilisant la formule

$$P(S_{n+1} = k) = \sum_j P(X_{n+1} = l) P(S_n = k - l)$$

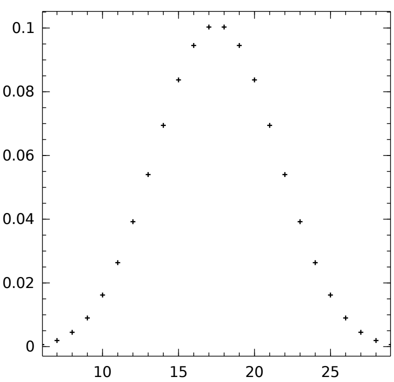
où la somme porte sur l'ensemble des valeurs l que prend X_{n+1}. Si n est suffisamment grand, le graphe des fréquences devrait se rapprocher d'une gaussienne, pour peu qu'il soit convenablement renormalisé. Dans les figures qui suivent, on s'est restreint sur l'axe des abscisses aux valeurs de k qui sont à moins de trois fois l'écart-type de l'espérance de S_n.



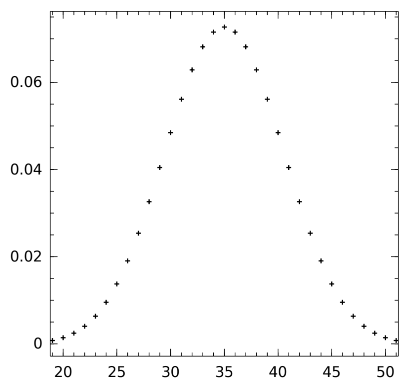
Graphe de S_3



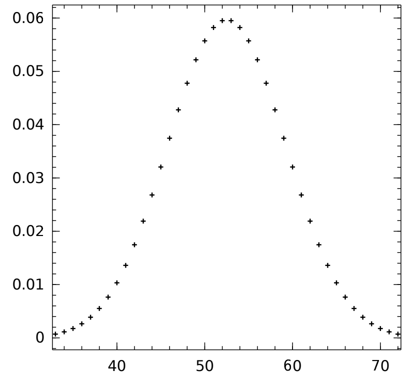
Graphe de S_4



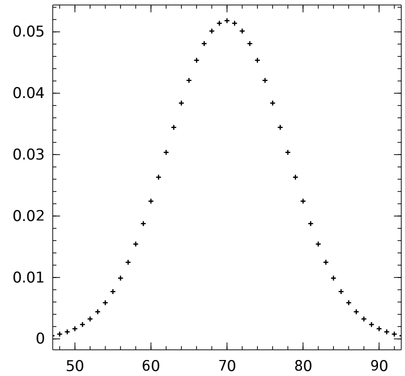
Graphe de S_5



Graphe de S_{10}



Graphe de S_{15}

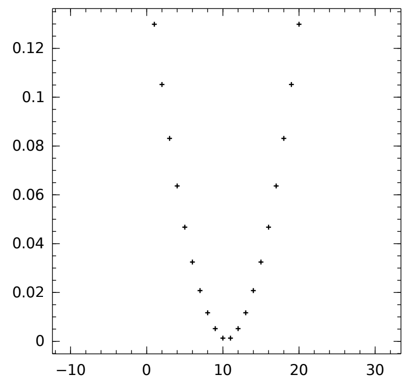


Graphe de S_{20}

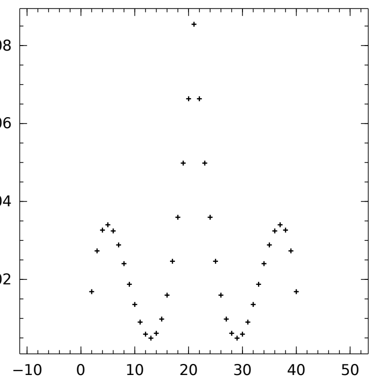
Dès $n = 5$, on voit les probabilités s'ordonner selon la fameuse courbe en cloche, dont la densité est donnée par la gaussienne.

Il est intéressant de regarder ce qu'on obtient lorsqu'on part d'une loi qui présente plusieurs maxima. Prenons pour X_0 la loi $P(X_0 = k) = \frac{k^2}{770}$ pour k compris entre -10 et 10. Le graphe des fréquences de X_0 est ci-dessous.

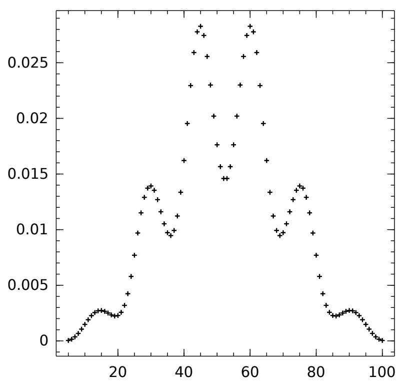
$$P(X_0 = k) = \frac{k^2}{770}, \quad k \in \{-10, -9, .., 9, 10\}$$



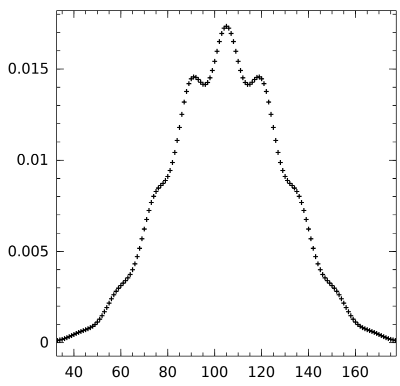
Graphe de X_0



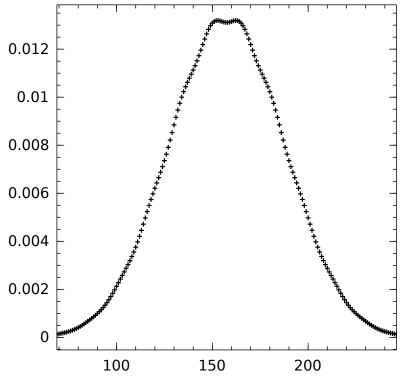
Graphe de S_2



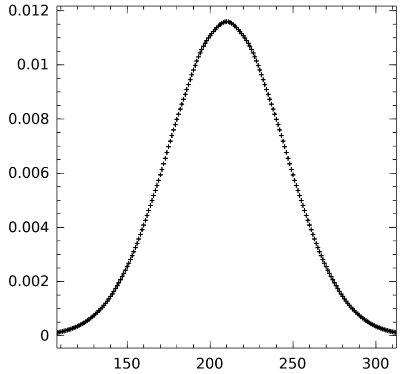
Graphe de S_5



Graphe de S_{10}



Graphe de S_{15}



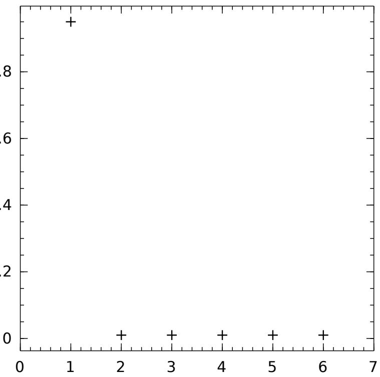
Graphe de S_{20}

La gaussienne met plus de temps à apparaître. Les premiers graphes présentent des oscillations qui s'amortissent quand n devient grand.

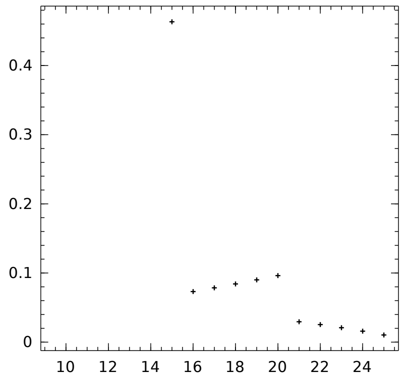
Un autre cas intéressant est donné par une loi fortement dissymétrique. Considérons un X_0 pour lequel

$$P(X_0 = 1) = 0,95$$

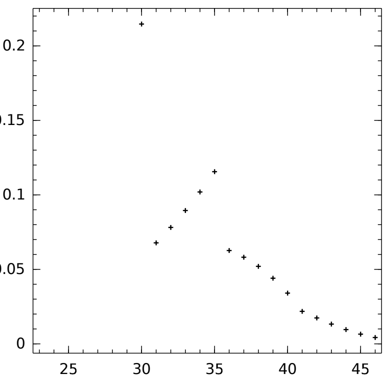
$$P(X_0 = 2) = P(X_0 = 3) = P(X_0 = 4) = P(X_0 = 5) = P(X_0 = 6) = 0,01.$$



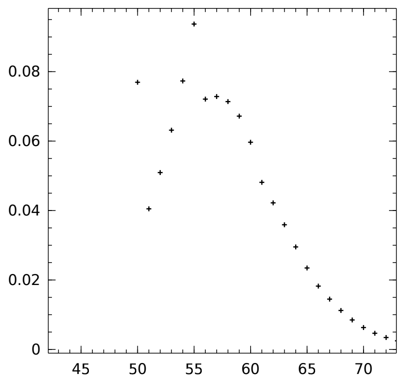
Graphe de X_0



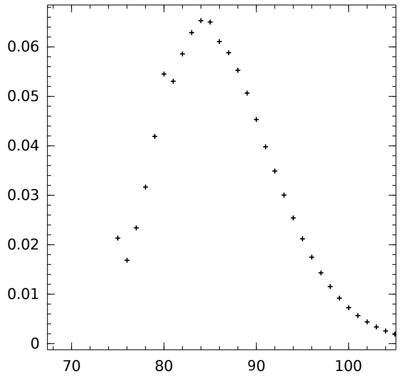
Graphe de S_2



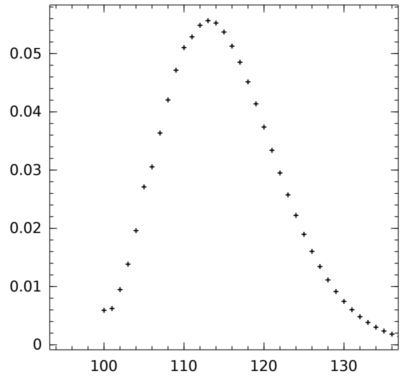
Graphe de S_5



Graphe de S_{10}



Graphe de S_{15}



Graphe de S_{20}

Comme nous pouvons le voir sur ces graphiques, la dissymétrie est encore présente pour $n = 100$. Cet exemple doit donc inciter à la prudence quant aux valeurs de n pour lesquelles l'approximation donnée par la loi normale est pertinente. Il est d'usage en statistique de faire cette approximation dès que $n = 30$, mais cela n'est pas toujours valide en pratique.

6.4 Exercices

exercice 1
Soit X_k une suite de variables aléatoires indépendantes suivant chacune une loi de Poisson de paramètre un. Quelle est la loi de la somme $\sum_{1\le k\le n} X_k$?

Considérons pour chaque $n \in \mathbf{N}^*$ une variable aléatoire Y_n suivant une loi de Poisson de paramètre $n \in \mathbf{N}^*$. Montrer que
$$\frac{Y_n - n}{\sqrt{n}} \xrightarrow[n\to\infty]{loi} \mathcal{N}(0,1).$$

exercice 2
Soit Y, Y_n et Z_n des variables aléatoires telles que Y_n converge en loi vers Y et Z_n converge presque sûrement vers une constante $c \in \mathbf{R}$. Montrer que Y_nZ_n converge en loi vers cY. *Indication : Y_n et Z_n sont tendues.*

Soit $(X_n)_{n\ge1}$ une suite de variables aléatoires indépendantes identiquement distribuées, de carrés intégrables, centrées. Montrer que
$$\sum_{k=1}^n X_k \Big/ \Big(\sum_{k=1}^n X_k^2\Big)^{1/2} \xrightarrow[n\to\infty]{} \mathcal{N}(0,1).$$

exercice 3
Soit $(X_n)_{n\ge1}$ une suite de variables aléatoires indépendantes identiquement distribuées, positives, de carrés intégrables et telles que $E(X_n) = 1$, $\sigma(X_n) = 1$. Montrer que
$$2\Big(\sqrt{S_n} - \sqrt{n}\Big) \xrightarrow[n\to\infty]{loi} \mathcal{N}(0,1).$$

exercice 4
Soit (X_n) une suite de variables aléatoires indépendantes, $S_n = \sum_{k=1}^n X_k$. Étudier la convergence en loi de $S_n/\sigma(S_n)$ dans les cas suivants :
– $P(X_k = \sqrt{k}\ ) = P(X_k = -\sqrt{k}\ ) = \frac{1}{2}$
– $P(X_k = k) = P(X_k = -k) = \frac{1}{2}$
– $P(X_k = \frac{1}{\sqrt{k}}) = P(X_k = -\frac{1}{\sqrt{k}}) = \frac{1}{2}$
– $P(X_k = a_k) = P(X_k = -a_k) = \frac{1}{2}$, où $(a_k)_{k\in\mathbf{N}^*}$ est bornée et $\displaystyle\sum_{k=1}^{\infty} a_k^2 = +\infty$.

exercice 5
Soit φ la fonction caractéristique d'une variable aléatoire réelle X non constante. On suppose que φ est réelle et positive.
– Montrer que $\varphi(t) = \int_{\mathbf{R}} \cos(tx)\, dP_X(x)$.
– Montrer que pour tout $u \in \mathbf{R}$, $\cos(u) \le 1 - \frac{u^2}{2} + \frac{u^4}{24}$.
– Montrer la minoration pour $t \in \mathbf{R}$: $\frac{1-\varphi(t)}{t^2} \ \ge\ \frac{1}{2} \int_{[-\frac{1}{t},\frac{1}{t}]} x^2 \left(1 - \frac{t^2x^2}{12}\right) dP_X(x)$.
– Montrer qu'il existe $\delta > 0$ et $\epsilon > 0$ tels que pour tout $t \in]-\delta,\delta[$, $|\varphi(t)| \le 1-\varepsilon t^2$.
– En déduire que pour tout $p > 2$, la fonction $t \mapsto e^{-|t|^p}$ n'est la fonction caractéristique d'aucune mesure de probabilité.

Chapitre 7

Espérance conditionnelle

Nous allons introduire un nouveau concept pour mesurer l'indépendance de deux variables aléatoires ou d'une variable aléatoire et d'une tribu.

7.1 Définition de la notion d'espérance conditionnelle

Définition 15 *Soit* $(\Omega, \mathcal{T}, P)$ *un espace probabilisé,* $X : \Omega \to \mathbf{R}$ *une variable aléatoire intégrable et* $\mathcal{F} \subset \mathcal{T}$ *une tribu. Alors il existe une unique fonction intégrable* $E(X \mid \mathcal{F})$*, appelée l'espérance conditionnelle de* X *sachant* $\mathcal{F}$*, qui satisfait les deux conditions suivantes :*

– $E(X \mid \mathcal{F})$ *est* $\mathcal{F}$*-mesurable,*

– *pour toute variable aléatoire* Y *qui est* $\mathcal{F}$*-mesurable bornée,*

$$E(E(X \mid \mathcal{F})\, Y) = E(XY).$$

L'unicité est à comprendre "presque partout". Deux fonctions qui satisfont les conditions précédentes sont égales presque partout.

Définition 16 *La probabilité conditionnelle d'un événement* $A \subset \Omega$ *relativement à la tribu* $\mathcal{F}$ *est la fonction définie par*

$$P(A \mid \mathcal{F}) = E(\mathbf{1}_A \mid \mathcal{F}).$$

Attention, $P(A \mid \mathcal{F})$ est une fonction définie presque partout sur Ω. Il n'est pas garanti qu'il y ait, pour presque tout $\omega \in \Omega$, une mesure de probabilité μ_ω définie sur $\mathcal{T}$ telle que $\mu_\omega(A) = P(A \mid \mathcal{F})(\omega)$.

Preuve de l'existence de $E(X \mid \mathcal{F})$
On commence par le cas où X est de carré intégrable. L'espace $L^2(\Omega, \mathcal{T}, P)$ est muni d'un produit scalaire :

$$\langle X, Y \rangle = \int X\, Y dP = E(XY).$$

On a donc une notion d'orthogonalité et de projection orthogonale. Considérons le sous-espace fermé $L^2(\Omega, \mathcal{F}, P)$ constitué des fonctions de carré intégrable qui sont $\mathcal{F}$-mesurables. L'espérance conditionnelle de X est la projection orthogonale de X sur ce sous-espace. Cette projection orthogonale appartient au sous-espace si bien que $E(X \mid \mathcal{F})$ est $\mathcal{F}$-mesurable. C'est la première condition.

La variable aléatoire $X - E(X \mid \mathcal{F})$ est orthogonale à tous les éléments Y de $L^2(\Omega, \mathcal{F}, P)$:

$$\langle X - E(X \mid \mathcal{F}), Y \rangle = 0$$

Autrement dit, $E((X - E(X \mid \mathcal{F}))\ Y) = 0$. C'est la seconde condition. Ces deux conditions caractérisent de manière unique la projection orthogonale.

On peut généraliser cela aux variables aléatoires positives. Soit $X \geq 0$, on pose $X_l = \min(X, l)$. La variable aléatoire X_l est borné donc de carré intégrable et X_l converge en croissant vers X. On va vérifier dans la suite que $E(X_l \mid \mathcal{F})$ est une fonction positive et que la suite $E(X_l \mid \mathcal{F})$ est croissante. On définit alors

$$E(X \mid \mathcal{F}) = \lim_{l \to \infty} E(X_l \mid \mathcal{F}) \ \in [0, +\infty].$$

En utilisant le théorème de convergence croissante, on montre que $E(X \mid \mathcal{F})$ satisfait les conditions voulues pour tout $Y \geq 0$.

Pour les variables aléatoires intégrables, on les écrit comme différence de variables aléatoires intégrables positives, pour lesquelles on a vu comment définir l'espérance conditionnelle.

Le passage du cadre L^2 au cadre des variables aléatoires positives puis intégrables est justifié par les propriétés de l'espérance conditionnelle qui vont être détaillées après les exemples qui suivent.

7.2 Exemples

Tribu associée à une partition finie

On considère une partition finie $(E_i)_{i=1..n}$ de Ω :

– $\Omega = \bigcup_{i=1}^{n} E_i$

– $E_i \cap E_j = \emptyset$ pour tout i, j tels que $i \neq j$.

– $P(E_i) \neq 0$ pour tout i.

On peut associer à cette partition une tribu

$$\mathcal{F} = \Big\{ \bigcup_{i \in S} E_i \mid S \subset \{1, ..., n\} \Big\}.$$

C'est l'ensemble des parties de Ω qui peuvent s'écrire comme union d'éléments de la partition.

Proposition 12

$$E(X \mid \mathcal{F})(\omega) = \sum_{i=1}^{n} \frac{E(X\,\mathbf{1}_{E_i})}{P(E_i)}\,\mathbf{1}_{E_i}(\omega) = \sum_{i=1}^{n} \Big(\frac{1}{P(E_i)} \int_{E_i} X\,dP\Big)\,\mathbf{1}_{E_i}(\omega)$$

$$P(A \mid \mathcal{F})(\omega) = \sum_{i=1}^{n} P(A \mid E_i)\,\mathbf{1}_{E_i}(\omega)$$

avec la notation $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$.

On a donc $P(A \mid \mathcal{F})(\omega) = P(A \mid E_i)$ si $\omega \in E_i$.

Preuve de la proposition

On montre que la variable aléatoire $\sum \frac{E(X\,\mathbf{1}_{E_i})}{P(E_i)}\mathbf{1}_{E_i}$ satisfait les deux propriétés qui définissent l'espérance conditionnelle.

Commençons par vérifier que toute fonction $\mathcal{F}$-mesurable est combinaison linéaire des fonctions indicatrices des E_i. Pour cela, montrons qu'une telle fonction f est constante sur chacun des E_i. Soit $x, y \in E_i$, l'ensemble $f^{-1}(\{f(x)\})$ est dans $\mathcal{F}$ et contient x, il contient donc E_i. Le point $y \in E_i$ est dans cet ensemble, par conséquent $f(y) = f(x)$.

L'espérance conditionnelle $E(X \mid \mathcal{F})$ est $\mathcal{F}$-mesurable, nous venons de voir qu'elle est de la forme $E(X \mid \mathcal{F}) = \sum c_k \mathbf{1}_{E_k}$, où les c_k restent à déterminer. On utilise la deuxième propriété définissant l'espérance conditionnelle en prenant $Y = \mathbf{1}_{E_i}$ pour un certain i.

$$\begin{array}{rcl} E(E(X \mid \mathcal{F})\,Y) & = & E((\sum c_k \mathbf{1}_{E_k})\mathbf{1}_{E_i}) \\ & = & E(c_i \mathbf{1}_{E_i}) \\ & = & c_i\,P(E_i) \\ E(XY) & = & E(X\,\mathbf{1}_{E_i}). \end{array}$$

Ces deux expressions sont égales, ce qui donne $c_i = \frac{E(X\,\mathbf{1}_{E_i})}{P(E_i)}$ comme souhaité.

Le cas d'un espace produit

On considère l'espace produit $\Omega = \Omega_1 \times \Omega_2$ muni de la probabilité $P = P_1 \otimes P_2$ et la tribu $\mathcal{F} = \{A \times \Omega_2 \mid A \subset \Omega_1\}$. Vérifions qu'une fonction f est mesurable par rapport à $\mathcal{F}$ si et seulement si elle ne dépend que de la première coordonnée.

Soit $(\omega_1, \omega_2) \in \Omega$; posons $f(\omega_1, \omega_2) = t$. Nous avons $(\omega_1, \omega_2) \in f^{-1}(\{t\}) \in \mathcal{F}$. Il existe donc $A \subset \Omega_1$ tel que $f^{-1}(\{t\}) = A \times \Omega_2$. Le résultat (ω_1, ω_2) est dans $A \times \Omega_2$ et ω_1 est dans A. Pour tout $\omega \in \Omega_2$, (ω_1, ω) appartient à $A \times \Omega_2 = f^{-1}(\{t\})$, si bien que $f(\omega_1, \omega) = t = f(\omega_1, \omega_2)$.

Calculons l'espérance conditionnelle d'une variable aléatoire X relativement à la tribu $\mathcal{F}$.

Proposition 13 $E(X \mid \mathcal{F})(\omega_1) = \int X(\omega_1, \omega_2)\, dP_2(\omega_2)$.

Preuve
Nous savons que $E(X \mid \mathcal{F})(\omega_1, \omega_2)$ est de la forme $f(\omega_1)$ pour une certaine fonction f par $\mathcal{F}$-mesurabilité. On doit avoir de plus pour tout $Y : \Omega_1 \to \mathbf{R}$

$$\begin{aligned} E(E(X \mid \mathcal{F})\ Y) &= \int f(\omega_1) Y(\omega_1)\, dP_1(\omega_1)\, dP_2(\omega_2) \\ &= \int f(\omega_1) Y(\omega_1)\, dP_1(\omega_1). \end{aligned}$$

Ceci doit être égal à $E(XY) = \int X(\omega_1, \omega_2) Y(\omega_1)\, dP_1(\omega_1)\, dP_2(\omega_2)$. Nous avons donc l'égalité, pour tout Y,

$$\int f(\omega_1) Y(\omega_1)\, dP_1(\omega_1) = \int \Big(\int X(\omega_1, \omega_2) dP_2(\omega_2) \Big) Y(\omega_1)\, dP_1(\omega_1)$$

ce qui implique $f(\omega_1) = \int X(\omega_1, \omega_2)\, dP_2(\omega_2)$.

7.3 Propriétés de l'espérance conditionnelle

L'espérance conditionnelle jouit d'un certain nombre de propriétés proches de celles de l'intégrale. Soient X, Y deux variables aléatoires intégrables ou positives et $\mathcal{F}$ une tribu.

Proposition 14
- $E(X \mid \{\emptyset, \Omega\}) = E(X)$, $\quad E(X \mid \mathcal{T}) = X$, $\quad E(\mathbf{1} \mid \mathcal{F}) = \mathbf{1}$.
- $E(E(X \mid \mathcal{F})) = E(X)$.
- *Si X est $\mathcal{F}$-mesurable, alors $E(X \mid \mathcal{F}) = X$.*
- *Si X est indépendante de $\mathcal{F}$, alors $E(X \mid \mathcal{F}) = E(X)$.*
- Linéarité *: pour tout $\lambda \in \mathbf{R}$, $E(\lambda X + Y \mid \mathcal{F}) = \lambda E(X \mid \mathcal{F}) + E(Y \mid \mathcal{F})$.*
- Positivité *: si $X \geq 0$ alors $E(X \mid \mathcal{F}) \geq 0$.*
- Monotonie *: si $X \leq Y$ alors $E(X \mid \mathcal{F}) \leq E(Y \mid \mathcal{F})$.*
- *Pour tout Y $\mathcal{F}$-mesurable, positive si X est positive, bornée si X est intégrable,*

$$E(YX \mid \mathcal{F}) = Y E(X \mid \mathcal{F}).$$

– *Si* X *est dans* L^p *et* $1 \leq p \leq \infty$, *alors* $\|E(X \mid \mathcal{F})\|_p \leq \|X\|_p$.

– *Si* $\mathcal{F}_1 \subset \mathcal{F}_2$, *alors*

$$E(E(X \mid \mathcal{F}_1) \mid \mathcal{F}_2) = E(X \mid \mathcal{F}_1),$$

$$E(E(X \mid \mathcal{F}_2) \mid \mathcal{F}_1) = E(X \mid \mathcal{F}_1).$$

Preuve
Toutes ces propriétés sont d'abord démontrées pour des variables aléatoires de carré intégrables en utilisant le fait que l'espérance conditionnelle est une projection orthogonale. Elles sont ensuite généralisées aux variables aléatoires positives par un argument de convergence croissante, et enfin aux variables aléatoires intégrables en les décomposant comme différence de deux fonctions croissantes intégrables.

– La variable $E(X \mid \mathcal{T})$ est la projection de X sur $L^2(\Omega, \mathcal{T}, P)$, cette projection est égale à l'identité. Pour $E(X \mid \{\emptyset, \Omega\})$, toute fonction mesurable relativement à la tribu $\{\emptyset, \Omega\}$ est constante. La variable $E(X \mid \{\emptyset, \Omega\})$ est donc constante. Prenons $Y = \mathbf{1}$, nous avons $E(E(X \mid \mathcal{F})) = E(X)$, cette constante vaut donc $E(X)$.

– Si X est $\mathcal{F}$-mesurable, X est dans $L^2(\Omega, \mathcal{F}, P)$, elle est donc égale à sa projection sur $L^2(\Omega, \mathcal{F}, P)$: $E(X \mid \mathcal{F}) = X$.

– Si X est indépendante de $\mathcal{F}$, elle est indépendante de toute variable aléatoire Y $\mathcal{F}$-mesurable. Nous avons alors

$$E(E(X \mid \mathcal{F})\, Y) = E(XY) = E(X)E(Y) = E(E(X \mid \mathcal{F}))\, E(Y).$$

Prenons $Y = E(X \mid \mathcal{F})$ qui est bien $\mathcal{F}$-mesurable. Nous avons

$$E(E(X \mid \mathcal{F})^2) = E(E(X \mid \mathcal{F}))^2$$

ou encore $V(E(X \mid \mathcal{F})) = 0$. Ceci montre que $E(X \mid \mathcal{F})$ est constante p.s. En conséquence, $E(X \mid \mathcal{F}) = E(X)$.

– *Linéarité* : c'est la linéarité de la projection orthogonale.

– *Positivité* : soit $X \geq 0$; on prend $Y = \mathbf{1}_{(E(X|\mathcal{F})<0)}$.

$$E(E(X \mid \mathcal{F})\, Y) = E(E(X \mid \mathcal{F})\, \mathbf{1}_{(E(X|\mathcal{F})<0)}) \leq 0$$

$$E(XY) = E(X\mathbf{1}_{\left(E(X|\mathcal{F})<0\right)}) \geq 0$$

Ces deux expressions sont égales, elles sont donc toutes les deux nulles, ce qui implique $E(X \mid \mathcal{F})\, \mathbf{1}_{\left(E(X|\mathcal{F})<0\right)} = 0$ p.s. et par conséquent $E(X \mid \mathcal{F}) \geq 0$ p.s.

– *Monotonie* : soit Y, Z tels que $Y \geq Z$. D'après le point précédent,

$$E(Y \mid \mathcal{F}) - E(Z \mid \mathcal{F}) = E(Y - Z \mid \mathcal{F}) \geq 0.$$

– Soit Y $\mathcal{F}$-mesurable. La fonction $YE(X \mid \mathcal{F})$ est aussi $\mathcal{F}$-mesurable. De plus pour tout Z $\mathcal{F}$-mesurable,

$$E(ZYE(X \mid \mathcal{F})) = E(ZYX)$$

car ZY est $\mathcal{F}$-mesurable. La fonction $YE(X \mid \mathcal{F})$ satisfait donc les deux propriétés caractérisant $E(YX \mid \mathcal{F})$ et nous avons l'égalité

$$E(YX \mid \mathcal{F}) = YE(X \mid \mathcal{F}) \quad p.s.$$

– Soit $\mathcal{F}_1 \subset \mathcal{F}_2$. La fonction $E(X \mid \mathcal{F}_1)$ est $\mathcal{F}_2$-mesurable si bien que

$$E\Big(E(X \mid \mathcal{F}_1) \mid \mathcal{F}_2\Big) = E(X \mid \mathcal{F}_1).$$

Démontrons l'égalité $E\Big(E(X \mid \mathcal{F}_2) \mid \mathcal{F}_1\Big) = E(X \mid \mathcal{F}_1)$.
La fonction $E\Big(E(X \mid \mathcal{F}_2) \mid \mathcal{F}_1\Big)$ est $\mathcal{F}_1$-mesurable. De plus, pour tout $Y\mathcal{F}_1$-mesurable, Y est à la fois $\mathcal{F}_1$ et $\mathcal{F}_2$-mesurable et d'après la propriété précédente,

$$\begin{aligned} E\Big(YE\Big(E(X \mid \mathcal{F}_2) \mid \mathcal{F}_1\Big)\Big) &= E\Big(E\Big(E(YX \mid \mathcal{F}_2) \mid \mathcal{F}_1\Big)\Big) \\ &= E\Big(E(YX \mid \mathcal{F}_2)\Big) \\ &= E(YX) \end{aligned}$$

Ce sont les deux propriétés qui caractérisent $E(X \mid \mathcal{F}_1)$.

La majoration de la norme L^p de $E(X \mid \mathcal{F})$ découle de l'inégalité de Jensen, qui fait l'objet de la proposition qui suit.

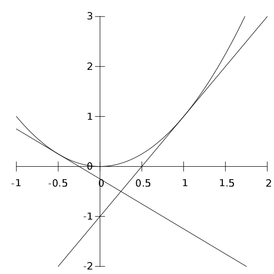
Proposition 15 (Inégalité de Jensen) *Soit* $(\Omega, \mathcal{T}, P)$ *un espace probabilisé,* $X : \Omega \to \mathbf{R}$ *une variable aléatoire et* $\varphi : \mathbf{R} \to \mathbf{R}$ *une fonction convexe. On suppose que* X *et* $\varphi(X)$ *sont intégrables ou positives. Alors*

$$\varphi\Big(E(X \mid \mathcal{F})\Big) \leq E\Big(\varphi(X) \mid \mathcal{F}\Big) \quad \textit{presque sûrement.}$$

Preuve
Une fonction convexe est continue, admet des dérivées à gauche et à droite en chaque point et on a l'inégalité pour tout $x, x_0 \in \mathbf{R}$,

$$\varphi(x) \geq \varphi(x_0) + \varphi'_+(x_0)(x - x_0)$$

où $\varphi'_+(x_0)$est la dérivée à droite en x_0. Le graphe de φ se trouve au-dessus de ses tangentes.



Appliquons cette inégalité pour $x = X(\omega)$ et $x_0 = E(X \mid \mathcal{F})(\omega)$.

$$\varphi(X) \geq \varphi(E(X \mid \mathcal{F})) + \varphi'_+(E(X \mid \mathcal{F}))(X - E(X \mid \mathcal{F}))$$

On prend maintenant l'espérance conditionnelle des deux termes de cette inégalité.

$$E\Big(\varphi(X) \mid \mathcal{F}\Big) \geq E\Big(\varphi\Big(E(X \mid \mathcal{F})\Big) \mid \mathcal{F}\Big) + E\Big(\Big(X - E(X \mid \mathcal{F})\Big)\varphi'_+\Big(E(X \mid \mathcal{F})\Big) \mid \mathcal{F}\Big)$$

Les fonctions $\varphi\Big(E(X \mid \mathcal{F})\Big)$ et $\varphi'_+\Big(E(X \mid \mathcal{F})\Big)$ sont $\mathcal{F}$-mesurables si bien que

$$E\Big(\varphi\Big(E(X \mid \mathcal{F}) \mid \mathcal{F}\Big) = \varphi\Big(E(X \mid \mathcal{F})\Big),$$

$$E\Big(\Big(X - E(X \mid \mathcal{F})\Big)\varphi'_+\Big(E(X \mid \mathcal{F})\Big)\Big) = \varphi'_+\Big(E(X \mid \mathcal{F})\Big)E\Big(X - E(X \mid \mathcal{F}) \mid \mathcal{F}\Big) = 0.$$

L'inégalité est démontrée.

Exemple
Les fonctions $x \mapsto |x|$, $x \mapsto e^x$, $x \mapsto |x|^p$ pour $p \geq 1$ sont convexes. Appliquons l'inégalité de Jensen à la fonction $x \mapsto |x|^p$.

$$|E(X \mid \mathcal{F})|^p \leq E(|X|^p \mid \mathcal{F}).$$

Utilisons la monotonie de l'espérance :

$$\|E(X \mid \mathcal{F})\|_p^p = E\Big(|E(X \mid \mathcal{F})|^p\Big) \leq E\Big(E(|X|^p \mid \mathcal{F})\Big) = E(|X|^p) = \|X\|_p^p.$$

Nous avons démontré l'inégalité recherchée entre la norme L^p d'une variable aléatoire et celle de son espérance conditionnelle.

7.4 Conditionnement relativement à une variable aléatoire

On commence par associer à toute variable aléatoire X une tribu dont les éléments sont composés d'union de lignes de niveau de X.

Définition 17 *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé, $X : \Omega \to \mathbf{R}$ une variable aléatoire. La tribu associée à X, notée $\mathcal{T}_X$ ou $\mathfrak{S}(X)$, est définie par :*

$$\mathcal{T}_X = \mathfrak{S}(X) = \{X^{-1}(A) \in \mathcal{T} \mid A \textit{ borélien de } \mathbf{R}\}$$

Proposition 16 *Une variable aléatoire $Y : \Omega \to \mathbf{R}$ est $\mathcal{T}_X$-mesurable si et seulement si il existe une fonction $f : \mathbf{R} \to \mathbf{R}$ borélienne telle que*

$$Y = f \circ X.$$

Preuve
Si $Y = f(X)$, pour tout A borélien, l'ensemble $Y^{-1}(A) = X^{-1}(f^{-1}(A))$ est bien dans $\mathcal{T}_X$.

Soit Y une variable aléatoire $\mathcal{T}_X$-mesurable. Montrons que Y est de la forme $f(X)$ pour une certaine fonction f borélienne.

• Traitons le cas où Y est étagée :

$$Y = \sum_{i=1}^{n} c_i \mathbf{1}_{A_i}, \ A_i \in \mathcal{T}_X.$$

Pour chaque i, on peut trouver un borélien $B_i \subset \mathbf{R}$ tel que $A_i = X^{-1}(B_i)$.

$$Y = \sum c_i \mathbf{1}_{X^{-1}(B_i)} = \sum c_i \mathbf{1}_{B_i} \circ X = \left(\sum c_i \mathbf{1}_{B_i}\right) \circ X$$

La variable aléatoire Y est bien de la forme $h \circ X$ avec $h = \sum c_i \mathbf{1}_{B_i}$.

• Passons au cas où $Y \geq 0$. On peut l'approcher de manière croissante par une suite de fonctions étagées $(Y_n)_{n \in \mathbf{N}}$ qui sont $\mathcal{T}_X$-mesurables, donc de la forme $h_n \circ X$. Nous avons :

$$Y_n = h_n \circ X \leq (\sup_n h_n) \circ X = \sup_n (h_n \circ X) = \sup Y_n = Y$$

Comme Y_n converge vers Y, on en déduit $Y = (\sup_n h_n) \circ X$ et Y est bien une fonction de X.

• Toute variable aléatoire $\mathcal{T}_X$-mesurable peut s'écrire comme différence de deux fonctions $\mathcal{T}_X$-mesurables positives de la forme $h \circ X$. La proposition est démontrée.

Nous somme en mesure de définir l'espérance conditionnelle d'une variable aléatoire relativement à une autre variable aléatoire.

Définition 18 *Soit X, Y deux variables aléatoires définies sur Ω. L'espérance conditionnelle de Y relativement à X, notée $E(Y \mid X)$, est définie par*

$$E(Y \mid X) = E\Big(Y \mid \mathcal{T}_X\Big).$$

Soit $A \in \mathcal{T}$. La probabilité conditionnelle de A relativement à X est définie par

$$P(A \mid X) = E(\mathbf{1}_A \mid X) = E(\mathbf{1}_A \mid \mathcal{T}_X).$$

La fonction $E(Y \mid X)$ est $\mathcal{T}_X$-mesurable. Elle peut donc s'exprimer comme une fonction de X. Cette fonction est notée $E(Y \mid X = x)$. C'est l'espérance de Y sachant que X a pris la valeur x et on a pour P-presque tout $\omega \in \Omega$,

$$E(Y \mid X)(\omega) = E\Big(Y \mid X = X(\omega)\Big),$$

$$P(A \mid X)(\omega) = P\Big(A \mid X = X(\omega)\Big).$$

Notons que l'espérance conditionnelle $E(Y \mid X = x)$ est une fonction de $\mathbf{R}$ dans $\mathbf{R}$ qui est définie P_X-presque partout. De fait, deux fonctions f et g sont égales P_X-presque partout si et seulement si $f \circ X$ et $g \circ X$ sont égales P-presque partout, d'après la formule de transfert.

On peut exprimer $E(Y \mid X)$ en fonction de la loi du couple (X, Y). Rappelons que cette loi est donnée par la formule

$$P_{(X,Y)}(A) = P\Big((X, Y) \in A\Big) = P\Big(\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in A\}\Big)$$

pour tout sous-ensemble borélien $A \subset \mathbf{R}^2$. La mesure $P_{(X,Y)}$ est une mesure de probabilité définie sur la tribu des boréliens de $\mathbf{R}^2$. Elle est dite à densité si on peut trouver $f_{X,Y} : \mathbf{R}^2 \to \mathbf{R}_+$, d'intégrale un sur $\mathbf{R}^2$, telle que

$$P_{(X,Y)}(A) = \int_A f_{X,Y}(x, y)\, dxdy, \quad A \subset \mathbf{R}^2.$$

Soit $g : \mathbf{R} \to \mathbf{R}$ une fonction borélienne telle que $g(y)$ est positive ou intégrable. L'espérance conditionnelle $E(g(Y) \mid X)$ est uniquement déterminée par le fait qu'elle est $\mathcal{T}_X$-mesurable et qu'elle satisfait l'égalité

$$E\Big(E(g(Y) \mid X)\ Z\Big) = E\Big(g(Y)\ Z\Big)$$

pour tout Z $\mathcal{T}_X$-mesurable, c'est-à-dire pour tout Z de la forme $f(X)$. La fonction $x \mapsto E(g(Y) \mid X = x)$ est donc uniquement déterminée par la relation :

$$\int f(x)\, E(g(Y) \mid X = x)\, dP_X(x) = \int f(X)\, g(Y) dP = \int f(x)\, g(y)\, dP_{(X,Y)}(x, y).$$

Cas d'une loi $P_{(X,Y)}$ à densité

Cherchons à donner une expression pour l'espérance conditionnelle dans le cas à densité. Lorsque $P_{(X,Y)}$ admet une densité $f_{X,Y}$, la loi P_X admet la densité

$$f_X(x) = \int_{\mathbf{R}} f_{X,Y}(x, y)\, dy.$$

En remplaçant dans l'expression précédente, on doit avoir pour tout $f : \mathbf{R} \to \mathbf{R}$ borélienne bornée,

$$\int f(x)\, E(g(Y) \mid X = x) \Big(\int f_{X,Y}(x, y) dy\Big)\, dx = \int f(x) \Big(\int g(y)\, f_{X,Y}(x, y)\, dy\Big)\, dx.$$

Comme cette relation doit être vérifiée pour toute fonction f, on en déduit la proposition suivante.

Proposition 17 *Soit X, Y deux variables aléatoires telles que le couple (X, Y) admette une densité $f_{X,Y}$ et soit $g : \mathbf{R} \to \mathbf{R}$ une fonction borélienne positive ou P_Y-intégrable. Alors, pour P_X-presque tout $x \in \mathbf{R}$,*

$$E(g(Y) \mid X = x) = \frac{\int_{\mathbf{R}} g(y) f_{X,Y}(x, y)\, dy}{\int_{\mathbf{R}} f_{X,Y}(x, y)\, dy}.$$

Cas d'une loi P_X discrète
Étudions maintenant le cas où $P(X = x_0) > 0$ pour un certain $x_0 \in \mathbf{R}$, ce qui se produit par exemple si X est une variable aléatoire discrète. Prenons pour f la fonction indicatrice du singleton $\{x_0\}$. Nous avons la relation :

$$\begin{aligned} P(X = x_0) E(g(Y) \mid X = x_0) &= \textstyle\int_{\{x_0\}} E(g(Y) \mid X = x)\, dP_X(x) \\ &= \textstyle\int_\Omega g(Y(\omega))\, \mathbf{1}_{(X=x_0)}(\omega)\, dP(\omega) \\ &= \textstyle\int_{(X=x_0)} g(Y(\omega))\, dP(\omega). \end{aligned}$$

On retrouve l'expression habituelle de l'espérance conditionnelle pour les variables aléatoires discrètes.

Proposition 18 *Soit X, Y deux variables aléatoires et $g : \mathbf{R} \to \mathbf{R}$ une fonction borélienne positive ou P_Y-intégrable. Pour tout $x \in \mathbf{R}$ tel que $P(X = x) > 0$,*

$$E(g(Y) \mid X = x) = \frac{E(g(Y)\, \mathbf{1}_{(X=x)})}{P(X = x)} = \frac{1}{P(X = x)} \int_{(X=x)} g(Y)\, dP.$$

On termine par une formule qui permet de calculer la probabilité d'un événement à partir des probabilités conditionnelles.

Proposition 19 *Soit X une variable aléatoire, $A \in \mathcal{T}$ et $I \subset \mathbf{R}$ un intervalle ou un borélien de $\mathbf{R}$. Alors*

$$P\Big(A \cap X^{-1}(I)\Big) = \int_I P(A \mid X = x)\, dP_X(x),$$

$$P(A) = \int_{\mathbf{R}} P(A \mid X = x)\, dP_X(x).$$

Ces formules découlent des égalités

$$\begin{aligned} \textstyle\int_I E(\mathbf{1}_A \mid X = x)\, dP_X(x) &= E(\mathbf{1}_{(X\in I)} E(\mathbf{1}_A \mid X)) \\ &= E(\mathbf{1}_{(X\in I)}\, \mathbf{1}_A) \\ &= P(A \cap (X \in I)). \end{aligned}$$

Généralisation à n variables
Les considérations précédentes se généralisent à un nombre quelconque de variables aléatoires.

Définition 19 *Soient $X_1, ..., X_n$ des variables aléatoires définies sur Ω. La tribu engendrée par les X_i est notée $\mathcal{T}_{X_1,...,X_n}$ ou $\mathfrak{S}(X_1, ..., X_n)$.*

$$\mathcal{T}_{X_1,...,X_n} = \mathfrak{S}(X_1, ..., X_n) = \{(X_1, ..., X_n)^{-1}(A) \mid A \subset \mathbf{R}^n \text{ } \textit{borélien}\}$$

On pose

$$E(Y \mid X_1, X_2, ..., X_n) = E\Big(Y \mid \mathcal{T}_{X_1,X_2,...,X_n}\Big).$$

On démontre comme précédemment que toute fonction $\mathcal{T}_{X_1,X_2,...,X_n}$-mesurable est de la forme $f(X_1, X_2, ..., X_n)$ où $f : \mathbf{R}^n \to \mathbf{R}$ est borélienne. On a donc une fonction $(x_1, ..., x_n) \mapsto E(Y \mid X_1 = x_1, ..., X_n = x_n)$ définie de $\mathbf{R}^n$ dans $\mathbf{R}$ qui, composée avec $(X_1, X_2, ..., X_n)$, redonne $E(Y \mid X_1, X_2, ..., X_n)$.

7.5 Exercices

exercice 1
On considère un espace probabilisé $(\Omega, \mathcal{T}, P)$, une tribu $\mathcal{F} \subset \mathcal{T}$ et deux événements $A \in \mathcal{T}$, $F \in \mathcal{F}$. Montrer la formule suivante :

$$P(F \mid A) = \frac{\int_F P(A \mid \mathcal{F})\, dP}{\int_\Omega P(A \mid \mathcal{F})\, dP}.$$

Quelle formule obtient-on lorsque $\mathcal{F}$ est associée à une partition $\coprod E_i = \Omega$?

exercice 2
Soit X une variable aléatoire de carré intégrable. Montrer que $E(X^2 \mid X+1) = X^2$.

exercice 3
Soit $X : \Omega \to \mathbf{R}_+$ une variable aléatoire positive définie sur $(\Omega, \mathcal{T}, P)$ et $\mathcal{F} \subset \mathcal{T}$ une tribu. Montrer que presque sûrement, $X > 0$ implique $E(X \mid \mathcal{F}) > 0$.

exercice 4
Soit X, Y deux variables aléatoires telle que le couple (X, Y) suit une loi uniforme sur le disque unité : $f_{X,Y}(x, y) = \frac{1}{\pi}\, \mathbf{1}_{[0,1]}(x^2 + y^2)$. Calculer $E(X^2 \mid Y)$.

exercice 5
Soit X une variable aléatoire de densité f_X. On considère sa valeur absolue $|X|$.

– Montrer que $|X|$ admet la densité $f_{|X|}(x) = \Big(f_X(x) + f_X(-x)\Big)\mathbf{1}_{\mathbf{R}_+}(x)$.

– Montrer que $E\Big(g(X) \mid |X| = x\Big) = \dfrac{g(x)f_X(x) + g(-x)f_X(-x)}{f_X(x) + f_X(-x)}$ $p.s.$

exercice 6
On considère une variable aléatoire suivant une loi exponentielle de paramètre λ. Sa densité est donc donnée par : $f(x) = \lambda\, e^{-\lambda x}\, \mathbf{1}_{[0,\infty[}(x)$

– Pour $s, t > 0$, calculez $P(X \geq t)$ et $P(X \geq t + s \mid X \geq s)$. On dit que la loi exponentielle est sans mémoire.

– Montrer que cette propriété est aussi vraie pour une variable aléatoire discrète suivant une loi géométrique de paramètre p.

exercice 7
Soit X une variable aléatoire de carré intégrable définie sur $(\Omega, \mathcal{T}, P)$, $\mathcal{F} \subset \mathcal{T}$ une tribu et un nombre réel $t > 0$. Montrer que $P(|X| \geq t \mid \mathcal{F}) \leq \frac{E(X^2|\mathcal{F})}{t^2}$.

exercice 8
Soit X une variable aléatoire de carré intégrable définie sur $(\Omega, \mathcal{T}, P)$ et $\mathcal{F} \subset \mathcal{T}$ une tribu. On définit la variance conditionnelle de X relativement à $\mathcal{F}$ comme suit :

$$V(X \mid \mathcal{F}) = E(X^2 \mid \mathcal{F}) - E(X \mid \mathcal{F})^2$$

Montrer que $V(X) = E(V(X \mid \mathcal{F})) + V(E(X \mid \mathcal{F}))$.

Chapitre 8

Théorie des martingales

Pour obtenir des théorèmes de convergence pour les séries de variables aléatoires indépendantes, nous allons introduire le concept de martingale.

8.1 La notion de martingale

Définition 20 *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé, $\mathcal{F}_n \subset \mathcal{T}$ des tribus, $n \in \mathbf{N}$. On dit que les $\mathcal{F}_n$ forment une* filtration *si elles constituent une suite croissante pour l'inclusion : $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ pour tout $n \in \mathbf{N}$.*

Une suite de variables aléatoires (M_n) est dite adaptée *à la filtration $\mathcal{F}_n$ si M_n est $\mathcal{F}_n$-mesurable pour tout $n \in \mathbf{N}$.*

La suite $(M_n)_{n\in\mathbf{N}}$ est une martingale *relativement à la filtration $\mathcal{F}_n$ si*

– les M_n sont intégrables,

– la suite (M_n) est adaptée à $\mathcal{F}_n$,

– $E(M_{n+1} \mid \mathcal{F}_n) = M_n$ pour tout $n \in \mathbf{N}$.

Si la filtration $\mathcal{F}_n$ n'est pas spécifiée, on prend $\mathcal{F}_n = \mathcal{T}_{M_1,\ldots,M_n}$.

Lorsque la suite est composée de fonctions de carrés intégrables, la variable M_n s'obtient en projetant orthogonalement M_{n+1} sur $L^2(\Omega, \mathcal{F}_n)$. Montrons que la projection de M_{n+k}, $k > 0$, sur $L^2(\Omega, \mathcal{F}_n)$ est aussi égale à M_n.

Proposition 20 *Pour tout $k > 0$, $E(M_{n+k} \mid \mathcal{F}_n) = M_n$.*

Cette proposition se démontre par récurrence sur k. Supposons là vérifiée pour $k > 0$ et démontrons là pour $k + 1$. Comme $\mathcal{F}_n \subset \mathcal{F}_{n+k}$,

$$E(M_{n+k+1} \mid \mathcal{F}_n) = E(E(M_{n+k+1} \mid \mathcal{F}_{n+k}) \mid \mathcal{F}_n) = E(M_{n+k} \mid \mathcal{F}_n) = M_n.$$

La proposition est démontrée.

Donnons quelques exemples de martingales.

Premier exemple
Soit $(X_k)_{k\in\mathbf{N}}$ une suite de variables aléatoires indépendantes entre elles et centrées ($E(X_n) = 0$ pour tout $n \in \mathbf{N}$). Posons $S_n = X_1 + X_2 + ... + X_n$. Alors la suite $(S_n)_{n\in\mathbf{N}}$ est une martingale relativement à la filtration $\mathcal{F}_n = \mathcal{T}_{X_1,X_2,...,X_n}$.

Preuve
La variable aléatoire S_n est bien intégrable et $\mathcal{F}_n$-mesurable.

$$\begin{aligned} E(S_{n+1} \mid \mathcal{F}_n) &= E(S_{n+1} \mid X_1, ..., X_n) \\ &= E(X_{n+1} + S_n \mid X_1, ..., X_n) \\ &= E(X_{n+1} \mid X_1, ..., X_n) + E(S_n \mid X_1, ..., X_n) \\ &= E(X_{n+1}) + S_n \\ &= S_n \end{aligned}$$

C'est la relation voulue.

Remarquons qu'on a l'égalité $\mathcal{T}_{X_1,X_2,...,X_n} = \mathcal{T}_{S_1,S_2,...,S_n}$ car les X_i s'expriment en fonction des S_i et réciproquement. La suite (S_n) est une martingale aussi bien par rapport aux (X_i) qu'aux (S_i).

Second exemple
Soit M une variable aléatoire intégrable et $(\mathcal{F}_n)$ une filtration. La suite

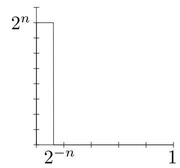
$$M_n = E(M \mid \mathcal{F}_n)$$

est une martingale.

Preuve
$E(M_{n+1} \mid \mathcal{F}_n) = E\Big(E(M \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n\Big) = E(M \mid \mathcal{F}_n) = M_n$.

Troisième exemple
On considère sur l'espace $\Omega = [0, 1[$, muni de la mesure de Lebesgue, les fonctions

$$M_n(x) = 2^n\, \mathbf{1}_{[0,1/2^n[}(x).$$



On prend pour $\mathcal{F}_n$ la tribu associée à la partition $[0, 1[= \coprod_{k=0}^{2^n-1} \Big[\dfrac{k}{2^n}, \dfrac{k+1}{2^n}\Big[$.

La suite (M_n) est une martingale relativement à la filtration $\mathcal{F}_n$.

Preuve
Les fonctions M_n sont intégrables : $\displaystyle\int_0^1 |M_n(x)|\, dx = 1$.
La variable aléatoire M_n est mesurable relativement à la tribu $\mathcal{F}_n$ car elle est constante sur chacun des éléments de la partition associée à $\mathcal{F}_n$. On a vu précédemment une formule explicite pour l'espérance conditionnelle relativement aux $\mathcal{F}_n$.

Notons $I_k = \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[\in \mathcal{F}_n$.

$$E(M_{n+1} \mid \mathcal{F}_n) = \sum_k \frac{E(M_{n+1}\mathbf{1}_{I_k})}{P(I_k)} \mathbf{1}_{I_k} = 2^n E\Big(2^{n+1}\mathbf{1}_{[0,\frac{1}{2^{n+1}}[}\Big) \mathbf{1}_{I_0} = 2^n \mathbf{1}_{I_0} = M_n.$$

La suite M_n satisfait bien les propriétés qui définissent les martingales.

Quatrième exemple
Voici un autre exemple en lien avec les jeux de hasard.

Proposition 21 *Soit $(M_n)_{n\in\mathbf{N}}$ une martingale et $(H_n)_{n\in\mathbf{N}}$ une suite de variables aléatoires telle que H_n est $\mathcal{F}_{n-1}$-mesurable pour tout $n \geq 1$. On suppose que H_n est bornée pour chaque $n \in \mathbf{N}$. On pose $G_0 = 0$ et*

$$G_n = \sum_{m=1}^{n} H_m(M_m - M_{m-1})$$

Alors $(G_n)_{n\in\mathbf{N}}$ est une martingale.

Preuve

$$\begin{aligned} E(G_{n+1} \mid \mathcal{F}_n) &= E(\textstyle\sum_1^{n+1} H_m(M_m - M_{m-1} \mid \mathcal{F}_n) \\ &= E(H_{n+1}(M_{n+1} - M_n) \mid \mathcal{F}_n) + E(\textstyle\sum_1^n H_m(M_m - M_{m-1}) \mid \mathcal{F}_n) \\ &= H_{n+1}E(M_{n+1} - M_n \mid \mathcal{F}_n) + E(G_n \mid \mathcal{F}_n) \\ &= G_n. \end{aligned}$$

De plus G_n est bien $\mathcal{F}_n$-mesurable et G_n est intégrable.

Prenons $(X_n)_{n\in\mathbf{N}^*}$ une suite de variables aléatoires indépendantes identiquement distribuées telle que $P(X_n = -1) = P(X_n = 1) = 1/2$ et définissons

$$M_0 = 0, \quad M_n = \sum_{i=1}^{n} X_i, \quad H_1 = 1,$$

$$H_n = \begin{cases} 2H_{n-1} & \text{si } X_{n-1} = -1 \\ 1 & \text{sinon.} \end{cases}$$

Il s'agit ici de modéliser une succession de mises à la roulette. On suppose que rouge et noir sortent chacun avec probabilité 1/2. La variable X_n vaut 1 si le joueur gagne au tirage n et vaut -1 sinon. La quantité H_n est la mise engagée par le joueur au $n^{\text{ième}}$ tirage, elle est fonction des résultats observés au cours des précédents tirages. Dans la martingale classique, le joueur double sa mise s'il a perdu au tirage précédent ; sinon il mise 1 euro. La valeur G_n représente le gain (ou la perte) à l'étape n. Notons F_0 la fortune initiale du joueur. Après avoir joué n fois, sa fortune vaut $F_n = G_n + F_0$. C'est une martingale.

Proposition 22 *Soit M_n une martingale. Alors*

$$E(M_n) = E(M_{n-1}) = ... = E(M_1) = E(M_0).$$

Preuve
$E(M_n) = E(E(M_{n+1} \mid \mathcal{F}_n)) = E(M_{n+1})$ pour tout $n \in \mathbf{N}$.

Dans l'exemple précédent, nous avons $E(F_n) = E(F_0)$. Le joueur ne peut pas augmenter l'espérance de sa fortune par une stratégie H_n judicieuse.

8.2 Convergence des martingales

On va montrer qu'une martingale est convergente dès qu'elle est bornée en norme L^2.

Définition 21 *Une martingale (M_n) est dite bornée dans L^2 si il existe $C > 0$ telle que pour tout $n \in \mathbf{N}$, $\|M_n\|_2 \leq C$.*

Théorème 11 (convergence L^2) *Toute martingale bornée dans L^2 converge au sens de la norme L^2 et presque sûrement : il existe une variable aléatoire M de carré intégrable telle que*

$$M_n \xrightarrow[n\to\infty]{L^2} M,$$
$$M_n \xrightarrow[n\to\infty]{p.s.} M.$$

Remarque On a alors $M_n = E(M \mid \mathcal{F}_n)$. On peut montrer que ce théorème se généralise à L^p pour $p \in]1, \infty[$.

Preuve de la convergence L^2
Posons $Y_i = M_i - M_{i-1}$ si bien que

$$M_n = \sum_{i=1}^{n} Y_i \ + M_0.$$

Vérifions que la famille des Y_i est orthogonale. Comme M_n est une martingale,

$$E(Y_i \mid \mathcal{F}_{i-1}) = E(M_i \mid \mathcal{F}_{i-1}) + E(M_{i-1} \mid \mathcal{F}_{i-1}) = M_{i-1} - M_{i-1} = 0$$

La variable Y_j est $\mathcal{F}_{i-1}$-mesurable pour tout $j < i$, ce qui donne

$$\langle Y_i, Y_j \rangle = E(Y_i Y_j) = E(E(Y_i \mid \mathcal{F}_{i-1})\ Y_j) = E(0) = 0.$$

On a donc :

$$\sum_{i=1}^{n} E(Y_i^2) = E\Big(\sum_{i=1}^{n} Y_i^2\Big) = E\Big((M_n - M_0)^2\Big) = E(M_n^2) - E(M_0^2) \leq C.$$

Cette série à termes positifs est bornée, elle converge. Elle est donc de Cauchy. On en déduit que la suite M_i est aussi de Cauchy pour la norme L^2 :

$$\|M_n - M_m\|_2^2 = E((M_n - M_m)^2) = E\Big(\Big(\sum_{i=m+1}^{n} Y_i\Big)^2\Big) = \sum_{m+1}^{n} E(Y_i^2)$$

pour $n \in \mathbf{N}$ et $m \geq n$. La série $\sum E(Y_i^2)$ étant de Cauchy, la dernière somme est inférieure à ε dès que m et n sont suffisamment grands. La suite (M_n) est bien de Cauchy pour la norme L^2, elle converge donc.

Pour démontrer la convergence presque sûre, on utilise le lemme suivant :

Lemme 5 (inégalité maximale) *Soit $(M_n)_{n\in\mathbf{N}}$ une martingale. Alors pour tout $N \in \mathbf{N}$ et tout $\lambda > 0$,*

$$P\Big(\max_{0\leq i\leq N} |M_i| \geq \lambda\Big) \leq \frac{E(M_N^2)}{\lambda^2}.$$

Preuve du lemme
On s'intéresse au premier indice pour lequel la martingale dépasse λ.

$A_0 = (|M_0| \geq \lambda)$
$A_1 = (|M_0| < \lambda, |M_1| \geq \lambda)$
$A_j = (|M_0| < \lambda, ..., |M_{j-1}| < \lambda, |M_j| \geq \lambda)$

Nous avons alors

$$\Big(\max_{0\leq i\leq N} |M_i| \geq \lambda\Big) = \coprod_{j=0}^{N} A_j.$$

Cherchons à minorer $E(M_N^2 \mathbf{1}_{A_j})$ en insérant le terme M_j dans le carré.

$$E(M_N^2 \mathbf{1}_{A_j}) = E((M_N - M_j)^2\, \mathbf{1}_{A_j}) + E(M_j^2\, \mathbf{1}_{A_j}) + 2E((M_N - M_j)M_j \mathbf{1}_{A_j}).$$

Le premier terme à droite de l'égalité est positif, tandis que le second terme est supérieur à $E(\lambda^2\, \mathbf{1}_{A_j})$ car M_j est supérieur à λ sur A_j. Vérifions que le dernier terme est nul en utilisant le fait que M_n est une martingale :

$$E(M_N M_j \mathbf{1}_{A_j}) = E(E(M_N M_j \mathbf{1}_{A_j} \mid \mathcal{F}_j)) = E(M_j \mathbf{1}_{A_j} E(M_N \mid \mathcal{F}_j)) = E(M_j \mathbf{1}_{A_j} M_j)$$

car M_j et $\mathbf{1}_{A_j}$ sont $\mathcal{F}_j$-mesurables. On en déduit $E((M_N - M_j)M_j \mathbf{1}_{A_j}) = 0$ comme souhaité. Au final,

$$E(M_N^2 \mathbf{1}_{A_j}) \geq \lambda^2 E(\mathbf{1}_{A_j}) = \lambda^2 P(A_j).$$

On conclut en faisant la somme pour j allant de 0 à n.

$$E(M_N^2) \geq E(M_N^2 \sum \mathbf{1}_{A_j}) \geq \lambda^2 P(\coprod A_j) = \lambda^2 P(\max |M_i| \geq \lambda).$$

Preuve de la convergence presque sûre
On a vu que pour tout $\varepsilon > 0$, il existe N tel que pour tout $m, n \geq N$

$$E((M_n - M_m)^2) < \varepsilon.$$

On peut donc construire par récurrence une suite croissante n_i telle que $n_0 = 0$ et

$$E((M_{n_i} - M_{n_{i-1}})^2) \leq 1/2^i.$$

Soit $n \in \mathbf{N}$ et i le plus grand entier tel que $n_i \leq n$. Nous avons l'encadrement $n_i \leq n < n_{i+1}$. Décomposons la martingale M_n comme suit :

$$M_n = (M_n - M_{n_i}) + \sum_{j=1}^{i} (M_{n_j} - M_{n_{j-1}}) \ + M_0.$$

Appliquons l'inégalité maximale à la martingale $\tilde{M}_n = M_{n+n_i}$, $N = n_{i+1} - n_i$ et $\lambda = \frac{1}{i^2}$:

$$P\Big(\max_{n_i \leq n \leq n_{i+1}} |M_n - M_{n_i}| \geq 1/i^2\Big) \leq \frac{E((M_{n_{i+1}} - M_{n_i})^2)}{1/i^2} \leq \frac{i^2}{2^i}.$$

La série de terme général $i^2/2^i$ converge, on peut appliquer le lemme de Borel-Cantelli : presque sûrement, pour i assez grand,

$$\max_{n_i \leq n \leq n_{i+1}} |M_n - M_{n_i}| \leq \frac{1}{i^2} \xrightarrow[i \to \infty]{} 0.$$

En particulier, pour $n = n_{i+1}$ et i supérieur à un certain $i_0 \in \mathbf{N}$ dépendant de ω,

$$|M_{n_{i+1}} - M_{n_i}| \leq \frac{1}{i^2}.$$

La série $\sum_{i \geq i_0} |M_{n_{i+1}} - M_{n_i}| \leq \sum_{i_0} 1/i^2$ est de nature convergente, ce qui montre que la série $\sum M_{n_{i+1}} - M_{n_i}$ est absolument convergente. Le théorème est démontré.

8.3 Séries de variables aléatoires indépendantes

Le premier résultat que nous pouvons déduire du théorème de convergence des martingales bornées dans L^2 est un critère pour la convergence presque sûre d'une série de variables aléatoires indépendantes.

Corollaire 6 *Soit (X_i) une suite de variables aléatoires indépendantes centrées ($E(X_i) = 0$ pour tout i) et de carrés intégrables. On suppose que*

$$\sum_{i \in \mathbf{N}} V(X_i) < \infty.$$

Alors la série $\sum X_i$ converge presque sûrement et en norme L^2.

Preuve
Nous avons vu précédemment que sous les hypothèses du corollaire, la suite $S_n = \sum_{i=1}^n X_i$ est une martingale. Remarquons que l'espérance de S_n est nulle :

$$E(S_n) = \sum_1^n E(X_i) = 0.$$

Sa variance est donc égale au carré de sa norme L^2 que nous évaluons comme suit :

$$\|S_n\|_2^2 = E(S_n^2) = V(S_n) = V(\sum_1^n X_i) = \sum_1^n V(X_i) \leq \sum_1^\infty V(X_i).$$

La martingale S_n est bornée dans L^2, elle converge presque sûrement et dans L^2.

Exemple
La série harmonique $\sum \frac{1}{k}$ est divergente. La série alternée $\sum \frac{(-1)^k}{k}$ est convergente. Qu'en est-il lorsque nous choisissons les signes des termes de la série de manière aléatoire ?

Proposition 23 *Soit (ε_k) une suite de variables aléatoires indépendantes identiquement distribuées telles que*

$$P(\varepsilon_k = 1) = P(\varepsilon_k = -1) = 1/2.$$

Alors la série $\displaystyle\sum_k \frac{\varepsilon_k}{k}$ converge presque sûrement.

Cela se déduit du corollaire. Il suffit de remarquer d'abord que $E(\frac{\varepsilon_i}{i}) = 0$ puis que

$$\sum_{k=1}^\infty V\Big(\frac{\varepsilon_k}{k}\Big) = \sum_{k=1}^\infty \frac{1}{k^2}\, V(\varepsilon_k) = \sum_{k=1}^\infty \frac{1}{k^2} < \infty.$$

Comme exemple d'une telle suite de variables aléatoires, on peut prendre

$$\Omega = \{-1, 1\}^{\otimes \mathbf{N}}, \quad P = \Big(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1\Big)^{\otimes \mathbf{N}}, \quad \varepsilon_i\Big((x_k)_{k\in\mathbf{N}}\Big) = x_i.$$

On a alors

$$P\Big(\Big\{(x_i)_{i\in\mathbf{N}} \in \Omega \;\Big|\; \sum_{k\geq 1} \frac{x_k}{k} \text{ converge } \Big\}\Big) = 1.$$

On se pose maintenant la question générale de la convergence d'une série $\sum X_i$ lorsque les X_i sont indépendantes entre elles. Le théorème suivant est dû à A. Kolmogorov.

Théorème 12 (théorème des trois séries) *Soit* $(X_i)_{i\in\mathbf{N}}$ *une suite de variables aléatoires indépendantes entre elles. Posons* $Y_i = X_i\mathbf{1}_{(|X_i|\leq 1)}$. *Alors la série* $\sum X_i$ *converge presque sûrement si et seulement si les trois séries suivantes convergent :*

- $\sum P(|X_i| \geq 1)$,
- $\sum E(Y_i)$,
- $\sum V(Y_i)$.

Preuve
Ce théorème se déduit du corollaire précédent. On se contente de démontrer que la convergence des trois séries implique la convergence presque sûre de $\sum X_i$. Comme $\sum P(|X_i| \geq 1)$ converge, nous pouvons appliquer le lemme de Borel-Cantelli : pour presque tout ω, il existe i_0 tel que pour tout $i \geq i_0$, $|X_i(\omega)| \leq 1$. On a alors $Y_i(\omega) = X_i(\omega)$. Les séries $\sum X_i$ et $\sum Y_i$ sont donc de même nature.

Posons $\tilde{Y}_i = Y_i - E(Y_i)$. Comme $\sum E(Y_i)$ converge, il suffit de démontrer la convergence presque sûre de $\sum \tilde{Y}_i$. Les $\tilde{Y}_i$ sont centrées et leur variance est égale à celle des Y_i :

$$\|\tilde{Y}_i\|_2^2 = V(\tilde{Y}_i) = V(Y_i).$$

On sait que la série $\sum V(\tilde{Y}_i) = \sum V(Y_i)$ converge. Le corollaire s'applique, la série $\sum \tilde{Y}_i$ est convergente presque sûrement et le théorème est démontré.

Complément
Donnons une preuve de la loi des grands nombres dérivée des théorèmes précédents et valide pour toute suite de variables aléatoires $(X_n)_{n\in\mathbf{N}}$ indépendantes identiquement distribuées intégrables.

On considère les variables $Y_k = X_k\mathbf{1}_{(|X_k|\leq k)}$. Montrons que la série $\sum \frac{V(Y_k)}{k^2}$ est convergente.

$$\sum_{k\geq 1} \frac{E(Y_k^2)}{k^2} = \sum_{k\geq 1} \frac{1}{k^2}\int_0^\infty x^2\mathbf{1}_{\{x\leq k\}} dP_{|X_k|}(x) = \int_0^\infty \Big(\sum_{k\geq 1} \frac{x}{k^2}\,\mathbf{1}_{\{k\geq x\}}\Big)\, x\, dP_{|X_0|}(x).$$

La somme qui apparaît sous l'intégrale dans le dernier terme est majorée par $2\sum_{k\geq 1}\frac{1}{k^2}$ pour $x \in [0,2]$. Pour $x \geq 2$, on effectue une comparaison série-intégrale.

$$\sum_{k\geq 1} \frac{x}{k^2}\,\mathbf{1}_{\{k\geq x\}} \leq \sum_{k\geq x}\int_{k-1}^{k} \frac{x}{t^2}\,dt \leq \int_{x-1}^{\infty} \frac{x}{t^2}\,dt \leq \frac{x}{x-1} \leq 2.$$

Nous avons de plus $V(Y_k - E(Y_k)) = V(Y_k) \leq E(Y_k^2)$, si bien que la série $\sum V(\frac{Y_k - E(Y_k)}{k})$ est convergente. La série $\sum \frac{Y_k - E(Y_k)}{k}$ converge donc presque sûrement, en vertu du théorème vu précédemment.

De manière générale, pour toute suite (x_k) telle que $\sum \frac{x_i}{i}$ converge, la moyenne $\frac{1}{n}\sum x_k$ converge vers 0. Cela découle de la formule suivante

$$\frac{1}{n}\sum_{k=1}^{n}\Big(\sum_{i=1}^{k}\frac{x_i}{i}\Big)=\sum_{i=1}^{n}\frac{x_i}{i}-\frac{1}{n}\sum_{i=1}^{n}x_i$$

qui se démontre en intervertissant les deux signes sommes. On en déduit

$$\frac{1}{n}\ \sum_{k=1}^{n}(Y_k-E(Y_k))\ \xrightarrow[n\to\infty]{p.s.}\ 0.$$

Par convergence dominée, la suite $E(Y_k) = \int x\mathbf{1}_{\{|x|\leq k\}}dP_{X_0}(x)$ converge vers $E(X_0)$. Il en va donc de même pour $\frac{1}{n}\sum E(Y_k)$. Il reste à remarquer que

$$\sum_{k=1}^{\infty}P(Y_k\neq X_k)=\sum_{k=1}^{\infty}P(|X_0|\geq k)=\int\sum_{k=1}^{\infty}\mathbf{1}_{\{k\leq x\}}\ dP_{|X_0|}(x)\leq\int x\,dP_{|X_0|}(x)<\infty$$

D'après le lemme de Borel-Cantelli, pour presque tout ω, les suites $X_k(\omega)$ et $Y_k(\omega)$ coïncident à partir d'un certain rang et la différence $\frac{1}{n}\sum X_k(\omega)-\frac{1}{n}\sum Y_k(\omega)$ tend vers 0. Le résultat est démontré.

8.4 Convergence des espérances conditionnelles

Théorème 13 *Soit $(\Omega,\mathcal{T},P)$ un espace probabilisé et $\mathcal{F}_n\subset\mathcal{T}$ une suite de tribus croissante pour l'inclusion telle que $\mathcal{T}$ soit engendrée par tous les $\mathcal{F}_n$. Alors pour tout $X:\Omega\to\mathbf{R}$ de carré intégrable,*

$$E(X\mid\mathcal{F}_n)\xrightarrow[n\to\infty]{}X \textit{ presque sûrement et en norme } L^2.$$

Remarque Ce théorème est encore vrai si X est juste intégrable, on a alors convergence presque sûrement et en norme L^1.

Preuve
On a vu précédemment que la suite $E(X|\mathcal{F}_n)$ est une martingale. On sait aussi qu'elle est bornée dans L^2 : $\|E(X\mid\mathcal{F}_n)\|_2\leq\|X\|_2$.
Elle converge donc presque sûrement et en norme L^2, notons $\tilde{X}$ sa limite et montrons que $\tilde{X}=X$. Soit $n_0\in\mathbf{N}$ et $n\geq n_0$.

$$\|E(X-\tilde{X}\mid\mathcal{F}_{n_0})\|_2=\|E(E(X\mid\mathcal{F}_n)-\tilde{X}\mid\mathcal{F}_{n_0})\|_2\leq\|E(X\mid\mathcal{F}_n)-\tilde{X}\|_2\xrightarrow[n\to\infty]{}0$$

On a donc pour tout $n_0\in\mathbf{N}$ et $A\in\mathcal{F}_{n_0}$, $E(X-\tilde{X}\mid\mathcal{F}_{n_0})=0$. Ceci implique :

$$\int_A X-\tilde{X}\,dP=E(\mathbf{1}_A(X-\tilde{X}))=E(\mathbf{1}_AE(X-\tilde{X}\mid\mathcal{F}_{n_0}))=0.$$

On en déduit que $X=\tilde{X}$ en posant $\mathcal{A}=\bigcup_{n\in\mathbf{N}}\mathcal{F}_n$ et en appliquant le résultat d'intégration suivant.

Proposition 24 *Soit* $(\Omega, \mathcal{T}, P)$ *un espace probabilisé,* $Y : \Omega \to \mathbf{R}$ *intégrable,* $\mathcal{A} \subset \mathcal{T}$ *une algèbre de parties de* $\mathcal{T}$*. Si* $\int_A Y dP = 0$ *pour tout* $A \in \mathcal{A}$*, alors* $Y = 0$ *presque sûrement.*

Donnons deux applications du théorème de convergence précédent. Notons $\mathcal{T}_{(X_k,\, k\geq N)}$ ou $\mathfrak{S}(X_k, k \geq N)$ la tribu engendrée par toutes les variables X_k, $k \geq N$.

Théorème 14 (Loi du 0-1 de Kolmogorov) *Soit* $(X_k)_{k\in\mathbf{N}}$ *une suite de variables aléatoires indépendantes et*

$$A \in \bigcap_{N\in\mathbf{N}} \mathcal{T}_{(X_k,\, k\geq N)}.$$

Alors la probabilité de l'événement A *vaut* 0 *ou* 1 *:* $P(A) \in \{0, 1\}$*.*

Preuve
Pour $n \in \mathbf{N}$, l'événement A appartient à la tribu $\mathcal{T}_{(X_k, k\geq n+1)}$. Il est donc indépendant de $X_0, X_1, ..., X_n$. On en déduit

$$P(A) = P(A \mid X_0, ..., X_n).$$

Les tribus $\mathcal{F}_n = \mathcal{T}_{X_0,...,X_n}$ sont croissantes et engendrent $\mathcal{F} = \mathcal{T}_{(X_k,\, k\geq 0)}$. On peut appliquer le théorème précédent en se plaçant sur l'espace $(\Omega, \mathcal{F}, P)$:

$$P(A) = P(A \mid X_0, ..., X_n) = E(\mathbf{1}_A \mid \mathcal{F}_n) \xrightarrow[n\to\infty]{} \mathbf{1}_A \quad p.s.$$

On peut donc trouver $\omega \in \Omega$ tel que $P(A) = \mathbf{1}_A(\omega) \in \{0, 1\}$.

Exemple
Soit $(X_n)_{n\in\mathbf{N}}$ une suite de variables aléatoires indépendantes et $A = \{\omega \in \Omega \mid \sum_{k\geq 0} X_k \text{ converge}\}$. On peut écrire A sous la forme suivante, pour tout $N \in \mathbf{N}$:

$$A = \{\omega \in \Omega \mid \sum_{k\geq N} X_k \text{ converge}\} \in \mathcal{T}_{(X_k, k\geq N)}.$$

D'après la loi du 0-1, l'événement A a pour probabilité 0 ou 1. Par conséquent, ou bien pour presque tout $\omega \in \Omega$, la série $\sum X_k(\omega)$ converge, ou bien pour presque tout $\omega \in \Omega$, la série $\sum X_k(\omega)$ diverge.

Posons $S_n = \sum_{k=1}^n X_k$. Par le même raisonnement, l'ensemble

$$\{\omega \in \Omega \mid \text{la suite } (S_n(\omega))_{n\in\mathbf{N}^*} \text{ est bornée}\}$$

a comme probabilité 0 ou 1. De fait, N étant donné, la suite $(S_n(\omega))_{n\geq 1}$ est bornée si et seulement si la suite $(\sum_{k=N}^n X_k(\omega))_{n\geq N}$ est bornée et cette dernière suite ne dépend que de X_k pour $k \geq N$.

Comme autre application, montrons qu'on peut approcher presque partout toute fonction de carré intégrable définie sur $[0, 1[$ par des fonctions en escalier explicites.

Proposition 25 *Soit $f \in L^2([0,1[,dx)$. Alors*

$$\sum_{k=0}^{2^n-1} \Big(2^n \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f(x)\,dx\Big)\, \mathbf{1}_{[\frac{k}{2^n},\frac{k+1}{2^n}[} \xrightarrow[n\to\infty]{p.s.} f$$

Preuve
On prend pour $\mathcal{F}_n$ la tribu associée à la partition $[0,1[= \coprod_{k=0}^{2^n-1} [\frac{k}{2^n},\frac{k+1}{2^n}[$. On a vu précédemment que

$$E(f \mid \mathcal{F}_n) = \sum_{k=0}^{2^n-1} \Big(2^n \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f(x)\,dx\Big)\, \mathbf{1}_{[\frac{k}{2^n},\frac{k+1}{2^n}[}.$$

De plus, les intervalles de la forme $[\frac{k}{2^n},\frac{k+1}{2^n}[$, $n \in \mathbf{N}$ et $k \in \{0,...,2^n-1\}$, engendrent la tribu des boréliens. Le théorème précédent montre que $E(f \mid \mathcal{F}_n) \to f$ p.s. et en norme L^2.

Complément
Une martingale (M_n) est dite bornée dans L^1 s'il existe $C \geq 0$ tel que pour tout $n \in \mathbf{N}$, $\|M_n\|_1 \leq C$. On peut montrer qu'une telle martingale converge presque sûrement mais on n'a pas forcément la convergence en norme L^1.

Cette convergence a lieu en norme L^1 si et seulement si (M_n) satisfait la condition suivante :

$$\sup_{n\in\mathbf{N}} E(|M_n|\,\mathbf{1}_{|M_n|\geq\lambda)}) \xrightarrow[\lambda\to\infty]{} 0.$$

Une suite de variable aléatoires (M_n) qui satisfait cette condition est dite *équi-intégrable.*

Voici un exemple d'une telle suite. Considérons une variable aléatoire X intégrable et $\mathcal{F}_n$ une filtration. Alors la suite $E(X \mid \mathcal{F}_n)$ est une martingale équi-intégrable. C'est une conséquence du calcul suivant.

$$\begin{array}{rcl}
\int_{(|E(X|\mathcal{F}_n)|\geq\lambda)} |E(X \mid \mathcal{F}_n)|\,dP & \leq & \int_{(E(|X|\,|\,\mathcal{F}_n)\geq\lambda)} E(|X| \mid \mathcal{F}_n)\,dP \\
& \leq & \int_{(E(|X|\,|\,\mathcal{F}_n)\geq\lambda)} |X|\,dP \\
& \leq & \int_{(|X|\geq\sqrt{\lambda})} |X|\,dP + \sqrt{\lambda}\,P\Big(E(|X| \mid \mathcal{F}_n) > \lambda\Big) \\
& \leq & \int_{(|X|\geq\sqrt{\lambda})} |X|\,dP + \frac{E(|X|)}{\sqrt{\lambda}}
\end{array}$$

La première inégalité provient de la majoration $|E(X \mid \mathcal{F})| \leq E(|X| \mid \mathcal{F})$ et de la croissance de la fonction $x \mapsto x\mathbf{1}_{x\geq\lambda}$. La dernière majoration provient de l'inégalité de Markov.

Comme application, on peut montrer que la suite $E(X \mid \mathcal{F}_n)$ converge vers X aussi bien en norme L^1 que presque sûrement lorsque les $\mathcal{F}_n$ engendrent toute la tribu $\mathcal{T}$ et que X est intégrable.

8.5 Temps d'arrêt

Définition 22 *Soit* $(\mathcal{F}_n)$ *une filtration. Une variable aléatoire* $\tau : \Omega \to \mathbf{N} \cup \{\infty\}$ *est un* temps d'arrêt *relativement à la filtration* $\mathcal{F}_n$ *si l'événement* $(\tau = n)$ *est dans* $\mathcal{F}_n$ *pour tout* $n \in \mathbf{N}$.

Remarque
On a alors $(\tau \leq n) \in \mathcal{F}_n$ pour tout $n \in \mathbf{N}$ car

$$(\tau \leq n) = \coprod_{0 \leq k \leq n} (\tau = k) \in \bigcup_{0 \leq k \leq n} \mathcal{F}_k = \mathcal{F}_n.$$

Théorème 15 *Soit* (M_n) *une martingale et* τ *un temps d'arrêt. On pose*

$$(n \wedge \tau)(\omega) = \min(n, \tau(\omega)),$$

$$(M_{n\wedge\tau})(\omega) = M_{(n\wedge\tau)(\omega)}(\omega).$$

Alors $(M_{n\wedge\tau})$ *est une martingale.*

Remarque
Considérons un résultat $\omega \in \Omega$ pour lequel $\tau(\omega)$ est fini. Nous avons

$$M_{n\wedge\tau}(\omega) = \begin{cases} M_n(\omega) & \text{si } n \leq \tau(\omega) \\ M_{\tau(\omega)}(\omega) & \text{si } n \geq \tau(\omega) \end{cases}$$

Nous voyons que si $\tau(\omega) < \infty$, la suite $(M_{n\wedge\tau})(\omega)$ est stationnaire, c'est-à-dire constante à partir d'un certain rang. La suite $(M_{n\wedge\tau})_{n\in\mathbf{N}}$ converge sur l'ensemble $(\tau < \infty)$ vers la variable aléatoire $\omega \mapsto M_{\tau(\omega)}(\omega)$, notée M_τ. Sur l'ensemble $(\tau = \infty)$, elle coïncide avec la suite (M_n).

Exemple
On revient à l'exemple de la fortune F_n d'un joueur qui joue au casino. On a vu que c'est une martingale et que $E(F_n) = E(F_0)$. Le joueur peut-il augmenter l'espérance de sa fortune en s'arrêtant de jouer au bon moment ?

Reprenons l'exemple de la martingale consistant à doubler la mise si on est perdant. Plutôt que de jouer un nombre de fois n fixé, le joueur décide de s'arrêter au premier gain s'il y a effectivement un gain qui se produit au cours des n tirages. La variable

$$\tau(\omega) = \min\{k \in \mathbf{N}^* \mid X_k(\omega) = 1\}$$

est un temps d'arrêt :

$$(\tau = n) = (X_1 = -1, ..., X_{n-1} = -1, X_n = 1) \in \mathcal{T}_{X_1,...,X_n}.$$

La fortune du joueur est maintenant égale à $F_{n\wedge\tau}$. C'est une martingale, nous avons donc $E(F_{n\wedge\tau}) = E(F_{0\wedge\tau}) = E(F_0) = F_0$. On ne peut pas améliorer les gains par cette méthode.

Preuve du théorème
On pose $H_n = \mathbf{1}_{(\tau \geq n)}$. Cette fonction est $\mathcal{F}_{n-1}$-mesurable :

$$(\tau \geq n) = (\tau < n)^c = (\tau \leq n-1)^c \in \mathcal{F}_{n-1}.$$

Posons $G_n = \sum_{m=1}^{n} H_m(M_m - M_{m-1}) = \sum_{m=1}^{n} \mathbf{1}_{(m \leq \tau)}(M_m - M_{m-1})$.
Nous avons

$$\begin{array}{rcll} G_n(\omega) & = & M_{\tau(\omega)}(\omega) - M_0(\omega) & si \quad n \geq \tau(\omega), \\ & = & M_n(\omega) - M_0(\omega) & si \quad n \leq \tau(\omega). \end{array}$$

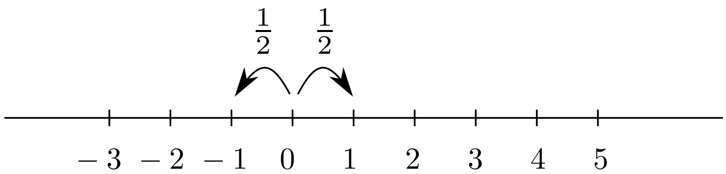
ce qui montre que $G_n = M_{n \wedge \tau} - M_0$ et nous avons vu plus haut que G_n est une martingale.

8.6 Illustration par une marche aléatoire

On se donne une suite de variables aléatoires $(X_k)_{k \in \mathbf{N}^*}$ indépendantes identiquement distribuées telle que

$$P(X_k = 1) = P(X_k = -1) = 1/2.$$

Posons $S_n = \sum_{k=1}^{n} X_k$. La suite $(S_n(\omega))_{n \in \mathbf{N}}$ est ici interprétée comme un déplacement sur l'ensemble des entiers $\mathbf{Z}$. L'entier $S_n(\omega)$ correspond à la position de la marche au temps n, en considérant que nous sommes à l'origine au temps 0.

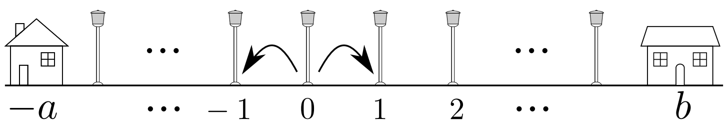


Si nous sommes à la position S_n au temps n, nous avons une chance sur deux de nous déplacer d'un pas vers la droite au temps $n+1$, et une chance sur deux de nous déplacer d'un pas vers la gauche :

$$P(S_{n+1} = S_n + 1) = P(X_{n+1} = 1) = 1/2,$$

$$P(S_{n+1} = S_n - 1) = P(X_{n+1} = -1) = 1/2.$$

On parle ici d'une marche aléatoire symétrique sur $\mathbf{Z}$.



Voici une présentation imagée de cette marche aléatoire et des problématiques associées. Un individu passablement éméché se déplace sur un chemin, de réverbères en réverbères, de manière aléatoire, avec probabilité 1/2 de partir vers la gauche ou la droite à chaque étape. Sa maison se trouve en $-a$, le bar en b avec $a, b \in \mathbf{N}^*$. Deux questions se posent :

- Parviendra-t-il à atteindre sa maison ou à retourner au bar ?

- Quelle est la probabilité qu'il atteigne la maison avant le bar ?

Théorème 16 *Presque sûrement, la marche atteint $-a$ ou b :*

$$P(\{\omega \in \Omega \mid \exists\, n \in \mathbf{N}^* \text{ tel que } S_n = -a \text{ ou } S_n = b\}) = 1.$$

La probabilité que la marche atteigne $-a$ avant b est égale à $\frac{b}{a+b}$:

$$P(\{\omega \mid \exists\, n \in \mathbf{N}^* \text{ tel que } -a < S_k(\omega) < b \text{ pour tout } k < n \text{ et } S_n(\omega) = -a\}) = \frac{b}{a+b}.$$

preuve
Posons $\tau(\omega) = \inf\{k \in \mathbf{N}^* \mid S_k = -a \text{ ou } S_k = b\}$ si la suite $(S_n(\omega))$ atteint effectivement $-a$ ou b et $\tau(\omega) = +\infty$ sinon. Cette variable aléatoire τ est un temps d'arrêt à valeurs dans $\mathbf{N} \cup \{\infty\}$. La preuve du théorème se fait en deux temps. On commence par vérifier que τ est fini presque sûrement :

$$P(\tau < \infty) = 1.$$

La variable aléatoire $\omega \mapsto S_{\tau(\omega)}(\omega)$ est alors bien définie, on la note S_τ, elle ne prend que les deux valeurs $-a$ et b. On montre ensuite que cette variable est d'espérance nulle :

$$E(S_\tau) = 0.$$

On a alors le système de deux équations

$$\begin{cases} P(S_\tau = -a) & + & P(S_\tau = b) & = & 1 \\ -a\, P(S_\tau = -a) & + & b\, P(S_\tau = b) & = & E(S_\tau) = 0 \end{cases}$$

qu'il suffit de résoudre pour trouver la probabilité $P(S_\tau = -a)$.

Montrons que τ est fini presque sûrement. On a vu précédemment que l'événement

$$\{\omega \in \Omega \mid \text{ la suite } (S_n(\omega))_{n \in \mathbf{N}^*} \text{ est bornée}\}$$

a pour probabilité 0 ou 1. C'est une conséquence de la loi du 0-1. On veut montrer que cette probabilité vaut 0. Si cela n'est pas le cas, pour presque tout $\omega \in \Omega$, la suite $(S_n(\omega))$ est bornée, ce qui implique

$$\frac{S_n(\omega)}{\sqrt{n}} \xrightarrow[n \to \infty]{} 0,$$

et pour toute f continue bornée, par convergence dominée,

$$\int_{\mathbf{R}} f\, dP_{\frac{S_n}{\sqrt{n}}} = \int_\Omega f\Big(\frac{S_n}{\sqrt{n}}(\omega)\Big) dP(\omega) \xrightarrow[n \to \infty]{} \int f(0)\, dP = f(0) = \int f(\omega) d\delta_0(\omega).$$

La suite $\frac{S_n}{\sqrt{n}}$ converge en loi vers la mesure de Dirac en zéro, ce qui contredit le théorème de la limite centrée. La suite n'est pas bornée, elle atteint $-a$ ou b presque sûrement.

Nous savons maintenant que la variable S_τ est bien définie, montrons que son espérance est nulle.

$$S_{n\wedge\tau} \xrightarrow[n\to\infty]{p.s.} S_\tau \quad \text{et} \quad -a \leq S_{n\wedge\tau} \leq b \text{ pour tout } n \in \mathbf{N}^*.$$

Appliquons le théorème de convergence dominée :

$$E(S_\tau) = \lim_{n\to\infty} E(S_{n\wedge\tau}).$$

Comme $(S_{n\wedge\tau})$ est une martingale, nous avons $E(S_{n\wedge\tau}) = E(S_{1\wedge\tau}) = E(S_1) = 0$, ce qui montre que $E(S_\tau) = 0$. Le théorème est démontré.

Complément
Donnons une autre preuve des deux résultats $P(\tau < \infty) = 1$ et $E(S_\tau) = 0$ en utilisant le théorème de convergence des martingales plutôt que la loi du 0-1. Nous savons que la suite $(S_{n\wedge\tau})$ est une martingale dont l'espérance est nulle et qui est toujours comprise entre $-a$ et b. Elle est donc bornée dans L^2 et converge en norme L^2 vers une certaine variable aléatoire Y. Par l'inégalité de Cauchy-Schwarz, la convergence L^2 implique la convergence des espérances :

$$E(Y) = \lim_{n\to\infty} E(S_{n\wedge\tau}) = 0.$$

Sur l'ensemble $(\tau = \infty)$, nous avons pour presque tout ω et pour tout n,

$$S_{n\wedge\tau}(\omega) = S_n(\omega), \quad -a < S_n(\omega) < b.$$

La suite S_n converge donc vers Y en norme L^2 sur $(\tau = \infty)$ et la suite $|S_{n+1} - S_n|$ doit converger vers 0 sur cet ensemble. Comme elle est constante égale à un, on en déduit que l'ensemble $(\tau = \infty)$ est négligeable et $Y = S_\tau$ presque sûrement.

8.7 Exercices

exercice 1
Soit (X_n) une suite de variables aléatoires indépendantes identiquement distribuées centrées de carrés intégrables. On pose $S_n = \sum_{k=1}^n X_k$. Montrer que pour tout $\varepsilon > 0$, la série suivante converge presque sûrement :

$$\sum_{k\geq 2} \frac{|X_k|}{\sqrt{k}\ln(k)^{1/2+\varepsilon}}$$

En déduire que la suite $\frac{S_n}{\sqrt{n}\ln(n)^{1/2+\varepsilon}}$ est bornée, puis qu'elle converge vers zéro.

exercice 2
Soit $(M_k)_{k\in\mathbf{N}}$ une martingale relativement à une filtration $(\mathcal{F}_k)$. On suppose les M_k de carrés intégrables. Montrer que si les M_k sont indépendants entre eux, chacun des M_k est constant presque partout.

exercice 3
Soit $(X_i)_{i\in\mathbf{N}}$ une suite de variables aléatoires positives, indépendantes entre elles, telles que $E(X_i) = 1$ pour tout i. On pose $M_n = X_1X_2...X_n$.

– Montrer que M_n est une martingale.

– En déduire que M_n converge presque sûrement. La limite est notée M.

On suppose maintenant que $P(X_i = 1/2) = P(X_i = 3/2) = 1/2$. Posons

$$\tau_n(\omega) = Card\{1 \leq k \leq n \mid X_k(\omega) = 3/2\}.$$

– Montrer que $\frac{\tau_n}{n}$ converge presque sûrement vers une limite qu'on calculera.

– Vérifier que $M_n = 3^{\tau_n}/2^n$.

– En déduire que $M = 0$ presque sûrement puis que $E\left(\prod_{i=1}^{\infty} X_i\right) \neq \prod_{i=1}^{\infty} E(X_i)$.

exercice 4
Soit a un entier strictement positif, $(X_i)_{i\in\mathbf{N}^*}$ une suite de variables aléatoires indépendantes identiquement distribuées, telles que $P(X_i = 1) = P(X_i = -1) = 1/2$. On pose $S_n = \sum_{i=1}^n X_i$, $\mathcal{F}_n = \mathcal{T}_{X_1,\ldots,X_n}$ et on s'intéresse au premier entier τ tel que S_τ est égal à a en valeur absolue :

$$\tau(\omega) = \inf\{k \in \mathbf{N} \mid |S_k(\omega)| = a\}.$$

– Montrer que la suite $M_n = (S_n)^2 - n$ est une martingale relativement aux $(\mathcal{F}_n)$.

– Montrer que τ est un temps d'arrêt relativement aux $(\mathcal{F}_n)$.

– Montrer que $E(M_{n\wedge\tau}) = 0$ pour tout $n \in \mathbf{N}^*$.

– Montrer la convergence $E(n \wedge \tau) \xrightarrow[n\to\infty]{} E(\tau)$.

– Montrer que pour presque tout $\omega \in \Omega$, il existe $k \in \mathbf{N}^*$ tel que $|S_k(\omega)| = a$.
En déduire que $|S_{n\wedge\tau}| \leq a$ pour tout $n \in \mathbf{N}^*$ et $(S_{n\wedge\tau})^2 \xrightarrow[n\to\infty]{} a^2$ p.s.

– Déduire de ce qui précède que $E(\tau) = a^2$.

Annexe A

Rappels d'intégration

On rappelle dans cette annexe un certain nombre de résultats d'intégration utilisés dans le cours. Le cadre est l'intégrale de Lebesgue. On adopte les notations probabilistes : $(\Omega, \mathcal{T}, P)$ est un espace probabilisé, c'est-à-dire un espace mesuré pour lequel $P(\Omega) = 1$.

A.1 Théorèmes de convergence

Théorème 17 (convergence croissante) *Soit $f_n : \Omega \to \mathbf{R}_+$ une suite de fonctions mesurables positives. On suppose que pour presque tout $\omega \in \Omega$, la suite $(f_n(\omega))_{n\in\mathbf{N}}$ est croissante et on note $f(\omega)$ la limite de cette suite. Alors*

$$\lim_{n\to\infty} \int_\Omega f_n(\omega)\, dP(\omega) = \int_\Omega f(\omega)\, dP(\omega).$$

Commentaire : la valeur des intégrales peut être égale à $+\infty$.

Cas particulier : en appliquant ce théorème à une suite de fonctions indicatrices $\mathbf{1}_{A_n}$, où $(A_n)_{n\in\mathbf{N}}$ est une suite d'ensembles croissante pour l'inclusion, alors

$$P\Big(\bigcup_0^\infty A_n\Big) = \lim_n P(A_n).$$

Théorème 18 (lemme de Fatou) *Soit $f_n : \Omega \to \mathbf{R}_+$ une suite de fonctions mesurables positives. Alors*

$$\int_\Omega \liminf_{n\to\infty} f_n(\omega)\, dP(\omega) \leq \liminf_{n\to\infty} \int_\Omega f_n(\omega)\, dP(\omega).$$

Théorème 19 (convergence dominée) *Soit $f_n : \Omega \to \mathbf{R}$ une suite de fonctions mesurables qui converge presque partout vers une fonction f. On suppose que pour presque tout $\omega \in \Omega$, la suite $f_n(\omega)$ est dominée par une fonction $g : \Omega \to \mathbf{R}_+$* intégrable *:*

$$|f_n(\omega)| \leq g(\omega) \quad \textit{pour presque tout } \omega \in \Omega.$$

Alors

$$\lim_{n\to\infty}\int_\Omega f_n(\omega)\,dP(\omega)=\int_\Omega f(\omega)\,dP(\omega).$$

Commentaire : Nous avons supposé $P(\Omega)=1$ si bien que toute suite f_n bornée est dominée par une fonction constante, qui est intégrable. Le théorème s'applique donc à une telle suite.

Théorème 20 (interversion somme intégrale, cas positif) *Soit $f_n:\Omega\to\mathbf{R}_+$ une suite de fonctions mesurables positives. Alors*

$$\int_\Omega \sum_{n=0}^{\infty} f_n(\omega)\;dP(\omega)=\sum_{n=0}^{\infty}\;\int_\Omega f_n(\omega)\;dP(\omega).$$

Commentaire : la somme de la série peut être égale à $+\infty$.

Théorème 21 (interversion somme intégrale, cas intégrable) *Soit $f_n:\Omega\to\mathbf{R}$ une suite de fonctions mesurables. On suppose que*

$$\sum_{n=0}^{\infty}\;\int_\Omega |f_n(\omega)|\;dP(\omega)<+\infty.$$

Alors

$$\int_\Omega \sum_{n=0}^{\infty} f_n(\omega)\;dP(\omega)=\sum_{n=0}^{\infty}\;\int_\Omega f_n(\omega)\;dP(\omega).$$

Commentaire : la série qui apparaît dans le second terme est convergente.

A.2 Intégrales dépendant d'un paramètre

Théorème 22 (continuité sous le signe intégral) *Soit I un intervalle de* $\mathbf{R}$. *Soit $f:I\times\Omega\to\mathbf{R}$ une fonction mesurable telle que*

– *pour P-presque tout $\omega\in\Omega$, $t\mapsto f(t,\omega)$ est continue sur I,*

– *il existe une fonction intégrable $g:\Omega\to\mathbf{R}$ telle que pour tout $t\in I$,*

$$|f(t,\omega)|\le g(\omega)\qquad \textit{pour presque tout } \omega\in\Omega.$$

Alors la fonction $t\mapsto\int_\Omega f(t,\omega)\,dP(\omega)$ est continue sur I : pour tout $t_0\in I$

$$\lim_{t\to t_0}\int_\Omega f(t,\omega)\,dP(\omega)=\int_\Omega f(t_0,\omega)\,dP(\omega).$$

Théorème 23 (dérivée sous le signe intégral) *Soit I un intervalle de* $\mathbf{R}$ *et* $f : I \times \Omega \to \mathbf{R}$ *une fonction mesurable telle que*

– *pour tout $t \in I$, $\omega \mapsto f(t,\omega)$ est intégrable,*

– *pour P-presque tout $\omega \in \Omega$, $t \mapsto f(t,\omega)$ est dérivable en tout point $t \in I$,*

– *il existe une fonction intégrable $g : \Omega \to \mathbf{R}$ telle que pour tout $t \in I$,*

$$\left|\frac{\partial}{\partial t} f(t,\omega)\right| \leq g(\omega) \quad \text{pour presque tout } \omega \in \Omega.$$

Alors en tout point $t \in I$,

$$\frac{d}{dt}\int_\Omega f(t,\omega)\,dP(\omega) = \int_\Omega \frac{\partial}{\partial t} f(t,\omega)\,dP(\omega).$$

A.3 Intégrales multiples

Ici, $(\Omega_1, \mathcal{T}_1, P_1)$ et $(\Omega_2, \mathcal{T}_2, P_2)$ sont des espaces probabilisés.

Théorème 24 (Fubini, cas positif) *Soit $f : \Omega_1 \times \Omega_2 \to \mathbf{R}_+$ une fonction $\mathcal{T}_1 \otimes \mathcal{T}_2$-mesurable positive. Alors*

$$\begin{aligned}\iint_{\Omega_1\times\Omega_2} f(\omega_1,\omega_2)\ dP_1 \otimes P_2(\omega_1,\omega_2) &= \int_{\Omega_2}\Big(\int_{\Omega_1} f(\omega_1,\omega_2)\,dP_1(\omega_1)\Big)\,dP_2(\omega_2)\\ &= \int_{\Omega_1}\Big(\int_{\Omega_2} f(\omega_1,\omega_2)\,dP_2(\omega_2)\Big)\,dP_1(\omega_1)\end{aligned}$$

Commentaire : les intégrales peuvent valoir $+\infty$.

Théorème 25 (Fubini, cas intégrable) *Soit $f : \Omega_1 \times \Omega_2 \to \mathbf{R}$ une fonction $\mathcal{T}_1 \otimes \mathcal{T}_2$-mesurable. On suppose que*

$$\iint_{\Omega_1\times\Omega_2} |f(\omega_1,\omega_2)|\ dP_1 \otimes P_2(\omega_1,\omega_2) < +\infty.$$

Alors

$$\begin{aligned}\iint_{\Omega_1\times\Omega_2} f(\omega_1,\omega_2)\ dP_1 \otimes P_2(\omega_1,\omega_2) &= \int_{\Omega_2}\Big(\int_{\Omega_1} f(\omega_1,\omega_2) dP_1(\omega_1)\Big)\,dP_2(\omega_2)\\ &= \int_{\Omega_1}\Big(\int_{\Omega_2} f(\omega_1,\omega_2) dP_2(\omega_2)\Big)\,dP_1(\omega_1).\end{aligned}$$

Commentaire : la fonction f est dans $L^1(\Omega_1 \times \Omega_2)$.

Théorème 26 (changement de variables) *Soient U, V deux ouverts de $\mathbf{R}^d$, $\varphi : U \to V$ un difféomorphisme de classe C^1, $f : V \to \mathbf{R}$ une application mesurable relativement à la mesure de Lebesgue sur V. On suppose f positive ou intégrable. Alors*

$$\int_U f(\varphi(u))J\varphi(u)\ du = \int_V f(v)\,dv$$

où $J\varphi(u)$ est le jacobien de φ : $J\varphi(u) = |det(d_u\varphi)|$.

Commentaire : pour le changement de variables en coordonnées polaires, $u = (r,\theta)$, $v = \varphi(u) = \varphi(r,\theta) = (r\cos(\theta), r\sin(\theta))$, $du = drd\theta$ et $J\varphi(r,\theta) = r$.

A.4 Espaces L^p

Rappel :

$$||f||_p = \left(\int_\Omega |f|^p \, dP\right)^{1/p} \quad \text{pour } 1 \leq p < \infty.$$

$$||f||_\infty = \inf\{M \geq 0 \mid \text{ pour presque tout } \omega \in \Omega,\ |f(\omega)| \leq M\}.$$

Théorème 27 (convergence normale dans L^p) *Soit $p \in [1, \infty]$ et (f_n) une suite de fonctions dans $L^p(\Omega)$. On suppose que*

$$\sum_{n \in \mathbf{N}} ||f_n||_p < \infty.$$

Alors la série $\sum f_n$ converge presque partout et en norme L^p vers une certaine fonction $f \in L^p(\Omega)$.

Théorème 28 (inclusion des espaces L^p) *Soit $p, q \in \mathbf{R}$ tels que $1 \leq p \leq q \leq \infty$. Alors*

$$L^\infty(\Omega) \subset L^q(\Omega) \subset L^p(\Omega) \subset L^1(\Omega).$$

De plus, pour tout $f : \Omega \to \mathbf{R}$ mesurable,

$$||f||_1 \leq ||f||_p \leq ||f||_q \leq ||f||_\infty.$$

Commentaire : le cas $p = 2$ est important : $L^\infty(\Omega) \subset L^2(\Omega) \subset L^1(\Omega)$.

Théorème 29 (extraction de sous-suites) *Soit $(f_n)_{n \in \mathbf{N}}$ une suite de fonctions de $L^p(\Omega)$ qui converge au sens de la norme L^p vers une certaine fonction f dans $L^p(\Omega)$. Alors il existe une sous-suite n_k telle que f_{n_k} converge presque partout vers f.*

Commentaire : en général, la convergence L^p n'implique pas la convergence presque partout.

A.5 Inégalités

Théorème 30 (inégalité de Minkowski) *Soit $p \in [1, \infty]$ et $f, g \in L^p(\Omega)$. Alors*

$$||f + g||_p \leq ||f||_p + ||g||_p.$$

Commentaire : c'est l'inégalité triangulaire pour les normes L^p.

Théorème 31 (inégalité de Cauchy-Schwarz) *Soit $f, g \in L^2(\Omega)$. Alors fg est intégrable et*

$$\int_\Omega fg\,dP \leq ||f||_2 \; ||g||_2.$$

Commentaire : on a égalité si et seulement si f et g sont proportionnelles.

Théorème 32 (inégalité de Hölder) *Soit $p, q \in [1, \infty]$ tels que $1/p + 1/q = 1/r$ ainsi que $f \in L^p(\Omega)$, $g \in L^q(\Omega)$. Alors fg est dans $L^r(\Omega)$ et*

$$||fg||_r \leq ||f||_p \; ||g||_q.$$

Commentaire : l'inégalité de Cauchy-Schwarz correspond à $p = q = 2$, $r = 1$.

Théorème 33 (inégalité de Jensen) *Rappelons que $P(\Omega) = 1$. Soit $\varphi : \mathbf{R} \to \mathbf{R}$ une fonction convexe et $f : \Omega \to \mathbf{R}$ telle que f et $\varphi \circ f$ sont intégrables. Alors*

$$\varphi\Big(\int_\Omega f \; dP\Big) \leq \int_\Omega \varphi \circ f \; dP$$

A.6 Formule d'inversion de Fourier

Le théorème suivant est une version ponctuelle de la formule d'inversion de Fourier ; c'est l'analogue du théorème de Dirichlet pour les séries de Fourier. On donne un énoncé est un peu plus général que celui utilisé dans le cours. La convention utilisée pour la transformée de Fourier est la suivante :

$$\hat{f}(t) = \int_{\mathbf{R}} e^{-itx} f(x) dx.$$

Lorsque f est intégrable, sa transformée $\hat{f}$ est continue. Elle tend vers 0 en l'infini, en vertu du lemme de Riemann-Lebesgue.

Lemme 6 (Riemann-Lebesgue) *Soit $f \in L^1$. Alors $\lim_{t\to\infty} \int_{\mathbf{R}} e^{-itx} f(x) \; dx = 0$.*

Ce lemme se démontre par un calcul explicite lorsque f est la fonction indicatrice d'un intervalle. Dans le cas général, il suffit d'approcher en norme L^1 la fonction f par une combinaison linéaire de fonctions indicatrices.

Théorème 34 (formule d'inversion de Fourier) *Soit $f \in L^1(\mathbf{R})$ et $t \in \mathbf{R}$. On suppose que f admet une limite à gauche et une limite à droite en t, notées $f(t^-)$ et $f(t^+)$. On suppose également que f est dérivable à droite et à gauche en t. Alors,*

$$\frac{1}{2}\Big(f(t^-) + f(t^+)\Big) = \lim_{A\to\infty} \int_{-A}^{A} e^{itx} \hat{f}(x) \; \frac{dx}{2\pi}.$$

Lorsque f est intégrable de classe C^1 et que $\hat{f}$ est intégrable, la formule devient

$$f(t) = \frac{1}{2\pi}\int_{\mathbf{R}} e^{itx}\,\hat{f}(x)\,dx \quad \text{pour tout } t \in \mathbf{R}.$$

Remarquons que $\hat{f}$ est intégrable dès que f est C^2 et f'' est intégrable. En effet, $\hat{f}$ est alors continue et majorée par une constante multipliée par $\frac{1}{t^2}$, comme le montre l'égalité

$$\hat{f}(t) = -\frac{1}{t^2}\,\widehat{f''}(t), \quad t \in \mathbf{R}^*,$$

qui s'obtient par une intégration par partie. En particulier, la formule d'inversion est vraie pour toute fonction C^∞ à support compact.

Preuve de la formule d'inversion
Quitte à translater la variable, on peut supposer $t = 0$. On a

$$\int_{\mathbf{R}} \mathbf{1}_{[-A,A]}(x)\hat{f}(x)\,\frac{dx}{2\pi} = \int_{\mathbf{R}} \widehat{\mathbf{1}_{[-A,A]}}(x)f(x)\,\frac{dx}{2\pi} = \int_{\mathbf{R}} \frac{2\sin Ax}{x}\,f(x)\,\frac{dx}{2\pi}.$$

On va montrer que $\displaystyle \lim_{A\to\infty}\left(\int_0^\infty \frac{2\sin Ax}{x}\,f(x)\,\frac{dx}{2\pi} - \frac{1}{2}f(0^+)\right) = 0.$

Faisons le changement de variable $y = Ax$ et remarquons que

$$\int_0^\infty \frac{\sin Ax}{x}\,dx = \int_0^\infty \frac{\sin y}{y}\,dy = \frac{\pi}{2},$$

et qu'ainsi $\displaystyle \int_0^\infty \frac{2\sin Ax}{x}\,f(x)\,\frac{dx}{2\pi} - \frac{1}{2}f(0^+) = \int_0^\infty 2\sin(Ax)\frac{f(x)-f(0^+)}{x}\,\frac{dx}{2\pi}.$
Sans le facteur $1/x$, il suffirait d'appliquer le lemme de Riemann-Lebesgue. On découpe en deux l'intégrale pour analyser ce qui se passe près de 0 et loin de 0.

Près de 0, on utilise l'hypothèse suivante :

$$f(x) = f(0^+) + xf'(0^+) + x\,\varepsilon(x), \text{ avec } \lim_{x\to 0}\varepsilon(x) = 0.$$

Par conséquent, il existe $\delta > 0$ tel que $\frac{f(x)-f(0^+)}{x}$ est borné sur $]0,\delta]$. La fonction $\frac{f(x)-f(0^+)}{x}\,\mathbf{1}_{]0,\delta]}(x)$ est intégrable et par le lemme de Riemann-Lebesgue,

$$\lim_{A\to+\infty}\int_0^\delta \sin(Ax)\,\frac{f(x)-f(0^+)}{x}\,dx = 0.$$

Loin de 0, sur $[\delta,+\infty[$, on a $0 < 1/x < 1/\delta$, et la fonction $\frac{f(x)}{x}\,\mathbf{1}_{[\delta,\infty[}(x)$ est intégrable. Par Riemann-Lebesgue,

$$\lim_{A\to+\infty}\int_\delta^\infty \sin(Ax)\,\frac{f(x)}{x}\,dx = 0.$$

Enfin, par définition des intégrales généralisées, on a :

$$\lim_{A\to+\infty}\int_{\delta}^{\infty}\frac{\sin(Ax)}{x}\,f(0^+)\,dx=\lim_{A\to+\infty}\int_{A\delta}^{\infty}\frac{\sin y}{y}\,dy\,f(0^+)=0.$$

On démontre de même que $\lim_{A\to\infty}\int_{-\infty}^{0}\frac{2\sin Ax}{x}\,f(x)\,\frac{dx}{2\pi}=\frac{1}{2}f(0^-)$, ce qui termine la preuve.

A.7 Exercices

exercice 1
Calculer $\lim_{n\to\infty} \int_0^n \left(1-\frac{x}{n}\right)^n dx$.

exercice 2
Soit $f : [0,1] \to \mathbf{R}$ une fonction mesurée. Calculer $\lim_{n\to\infty} \int_0^1 \frac{1}{1+n\,f(x)^2}\, dx$.

exercice 3
Montrer l'égalité $\int_0^1 \int_0^1 \frac{1}{1-xy}\, dx\, dy = \sum_{n=1}^{\infty} \frac{1}{n^2}$.

exercice 4
Soit $t > 0$, calculer l'intégrale

$$\int_0^\infty \left(\frac{\sin x}{x}\right)^2 e^{-tx}\, dx.$$

En déduire la valeur de $\int_0^\infty \left(\frac{\sin x}{x}\right)^2 dx$.

exercice 5
Calculer de deux manières différentes l'intégrale $\int_0^\infty \int_0^\infty \frac{1}{(1+y)(1+x^2y)}\, dxdy$.
En déduire la valeur de l'intégrale $\int_0^\infty \frac{\ln(x)}{(x^2-1)}\, dx$.

exercice 6
On considère la série suivante, pour $x \in [0,1]$,

$$\sum_n \frac{1}{2^n \sqrt{|x-\sin(n)|}}.$$

Montrer qu'elle converge pour presque tout $x \in [0,1]$, mais qu'elle diverge pour un ensemble dense de $x \in [0,1]$.

exercice 7
Calculer l'intégrale $\int_0^1 \int_0^1 \frac{1}{1-xy}\, dx\, dy$ en effectuant le changement de variables

$$\begin{cases} x &= \cos\theta - t \\ y &= \cos\theta + t \end{cases}$$

exercice 8
On se place sur $\mathbf{R}_+$ muni de la mesure de Lebesgue et on considère

$$f(x) = \frac{1}{x\,(1+|ln(x)|)^2}.$$

Montrer que :
- $f \in L^1([1,\infty[)$ et $f \notin L^p([1,\infty[)$ si $p < 1$.
- $f \in L^1([0,1])$ et $f \notin L^p([0,1])$ si $p > 1$.
- $f \in L^1([0,\infty[)$ et $f \notin L^p([0,\infty[)$ si $p \neq 1$.

A.8 Contre-exemples

Vérifier les assertions suivantes en calculant les deux termes de chacune des inégalités. Expliquer pourquoi les théorèmes classiques ne s'appliquent pas dans chacun des cas.

Linéarité :

$$\int_0^1 \frac{1}{x(x+1)} - \frac{1}{x} \; dx \quad \neq \quad \int_0^1 \frac{1}{x(x+1)} \, dx \quad - \quad \int_0^1 \frac{1}{x} \, dx$$

Interversion limite intégrale :

$$\lim_{n\to\infty} \int_0^1 nx^n \; dx \quad \neq \quad \int_0^1 \lim_{n\to\infty} nx^n \quad dx$$

Interversion somme intégrale :

$$\int_0^1 \sum_{k=0}^{+\infty} x^k - x^{2k} - x^{2k+1} \quad dx \quad \neq \quad \sum_{k=0}^{+\infty} \int_0^1 x^k - x^{2k} - x^{2k+1} \, dx$$

Continuité sous le signe intégral :

$$\lim_{t\to 0} \int_0^\infty \frac{|t|}{1+t^2x^2} \; dx \quad \neq \quad \int_0^\infty \lim_{t\to 0} \frac{|t|}{1+t^2x^2} \quad dx$$

Intégrales multiples :

$$\int_1^\infty \left(\int_1^\infty \frac{x-y}{max\,(x^3,y^3)} \; dx \right) dy \quad \neq \quad \int_1^\infty \left(\int_1^\infty \frac{x-y}{max\,(x^3,y^3)} \; dy \right) dx$$

Annexe B

Formulaire

On collecte dans cette annexe les formules vues dans le cours.

B.1 Loi d'une variable aléatoire

Loi d'une variable aléatoire X

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

Espérance

$$E(X) = \int_\Omega X\,dP = \int_{\mathbf{R}} x\,dP_X(x)$$

Variance

$$V(X) = E\Big((X - E(X))^2\Big) = \int_\Omega (X - E(X))^2\,dP = \int_{\mathbf{R}} (x - E(X))^2\,dP_X(x)$$

$$V(X) = E(X^2) - E(X)^2 = \int_\Omega X^2\,dP - \Big(\int_\Omega X\,dP\Big)^2 = \int_{\mathbf{R}} x^2\,dP_X(x) - \Big(\int_{\mathbf{R}} x\,dP_X(x)\Big)^2$$

Formule de transfert

$$E(g(X)) = \int_\Omega g(X)\ dP = \int_{\mathbf{R}} g(x)\ dP_X(x)$$

Fonction de répartition

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} dP_X(x)$$

Fonction caractéristique

$$\varphi_X(t) = E(e^{itX}) = \int_\Omega e^{itX}\,dP = \int_{\mathbf{R}} e^{itx}\,dP_X(x)$$

Cas discret

$$P_X = \sum_{k\in I} p_{x_k}\delta_{x_k}, \quad P_X(A) = \sum_{x_k\in A} p_{x_k}$$

$$E(X) = \int_{\mathbf{R}} x\,dP_X(x) = \sum_{k\in I} x_k\,P(X = x_k)$$

$$V(X) = \sum_{k\in I}(x_k - E(X))^2 P(X = x_k) = \sum_{k\in I} x_k^2\,P(X = x_k) - \Big(\sum_{k\in I} x_k\,P(X = x_k)\Big)^2$$

$$E(g(X)) = \int_\Omega g(X)\,dP = \int_{\mathbf{R}} g(x)\,dP_X(x) = \sum_{k\in I} g(x_k)P(X = x_k)$$

$$F_X(x) = \sum_{x_k\le x} P(X = x_k)$$

$$\varphi_X(t) = \sum_k e^{itx_k} P(X = x_k)$$

Cas continu

$$dP_X(x) = f_X(x)\,dx, \quad P_X(A) = \int_A f_X(x)\,dx$$

$$E(X) = \int_{\mathbf{R}} x\,f_X(x)\,dx$$

$$V(X) = \int_{\mathbf{R}}(x - E(X))^2\,f_X(x)\,dx = \int_{\mathbf{R}} x^2\,f_X(x)\,dx - \Big(\int_{\mathbf{R}} x\,f_X(x)\,dx\Big)^2$$

$$E(g(X)) = \int_\Omega g(X)\,dP = \int_{\mathbf{R}} g(x)\,dP_X(x) = \int_{\mathbf{R}} g(x)f_X(x)\,dx$$

$$F_X(x) = \int_{-\infty}^x f_X(x)\,dx$$

$$\varphi_X(t) = \int_{\mathbf{R}} e^{itx} f_X(x)\,dx$$

B.2 Inégalités

Inégalité de Cauchy-Schwarz

$$E(|XY|) = \int_\Omega |XY|\,dP \le \sqrt{\int_\Omega X^2\,dP}\,\sqrt{\int_\Omega Y^2\,dP} = \|X\|_2\,\|Y\|_2$$

Inégalité de Markov

$$P(Y \ge \lambda) \le \frac{E(Y)}{\lambda} \qquad \text{si } \lambda > 0,\ Y \ge 0.$$

Inégalité de Bienaymé-Tchebichev

$$P(|X - E(X)| \geq t) \leq \frac{V(X)}{t^2} \qquad \text{si } t > 0,\ E(X^2) < \infty.$$

Majoration L^p, $1 \leq p \leq \infty$, pour l'espérance conditionnelle

$$\|E(X \mid \mathcal{F})\|_p \leq \|X\|_p \quad \text{si } X \text{ est dans } L^p.$$

Inégalité de Jensen conditionnelle

$$\varphi\Big(E(X \mid \mathcal{F})\Big) \leq E\Big(\varphi(X) \mid \mathcal{F}\Big) \quad p.s. \quad \text{si } \varphi \text{ est convexe et } X,\ \varphi(X) \text{ intégrables.}$$

Inégalité maximale pour les martingales

$$P\Big(\max_{0\leq i\leq N} |M_i| \geq \lambda\Big) \leq \frac{E(M_N^2)}{\lambda^2} \quad \text{si } (M_n) \text{ est une martingale.}$$

B.3 Couples de variables aléatoires

Soit X et Y deux variables aléatoires. La loi du couple (X, Y) est donnée par

$$P_{(X,Y)}(A) = P((X,Y) \in A) = P(\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in A\})$$

Covariance

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \int_\Omega XY\, dP - \Big(\int_\Omega X\, dP\Big)\Big(\int_\Omega Y\, dP\Big).$$

$$V(X+Y) = V(X) + V(Y) + 2Cov(X,Y)$$

Formule de transfert

$$E(g(X,Y)) = \int_\Omega g(X,Y)\, dP = \int_{\mathbf{R}^2} g(x,y)\, dP_{(X,Y)}(x,y)$$

Espérance d'un produit de variables indépendantes

$$E(f(X)g(Y)) = E(f(X))\, E(g(Y))$$

Loi d'un couple de variables indépendantes

$$E(g(X,Y)) = \int_{\mathbf{R}^2} g(x,y)\, dP_{(X,Y)}(x,y) = \int_{\mathbf{R}^2} g(x,y)\, dP_X(x)\, dP_Y(y)$$

Cas discret

$$P_{(X,Y)} = \sum_{i,j} p_{x_i,y_j}\, \delta_{(x_i,y_j)}$$

$$P_{(X,Y)}(A) = \sum_{i,j \text{ tels que } (x_i,y_j)\in A} P(X = x_i, Y = y_j)$$

$$E(g(X,Y)) = \int_{\mathbf{R}^2} g(x,y)\, dP_{(X,Y)}(x,y) = \sum_{i,j} g(x_i,y_j)\, P(X=x_i, Y=y_j)$$

$$P(X=x_i) = \sum_j P(X=x_i, Y=y_j)$$

Cas continu

$$dP_{(X,Y)}(x,y) = f_{X,Y}(x,y)\, dxdy, \quad P_{(X,Y)}(A) = \int_A f_{X,Y}(x,y)\, dxdy$$

$$E(g(X,Y)) = \int_{\mathbf{R}^2} g(x,y)\, dP_{(X,Y)}(x,y) = \int_{\mathbf{R}^2} g(x,y)\, f_{X,Y}(x,y)\, dxdy$$

$$f_{X_1}(x_1) = \int_{\mathbf{R}} f_{X_1,X_2}(x_1,x_2)\, dx_2$$

B.4 Convergence de variables aléatoires

Convergence presque sûre

$$X_n \xrightarrow[n\to\infty]{p.s.} X \quad \text{si} \quad P(\{\omega \in \Omega \mid X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)\}) = 1.$$

Convergence L^p

$$X_n \xrightarrow[n\to\infty]{L^p} X \quad \text{si} \quad \|X_n - X\|_p \xrightarrow[n\to\infty]{} 0.$$

Convergence en probabilité

$$X_n \xrightarrow[n\to\infty]{proba} X \quad \text{si} \quad P(|X_n - X| > \varepsilon) \xrightarrow[n\to\infty]{} 0 \text{ pour tout } \varepsilon > 0.$$

Convergence en loi

$$X_n \xrightarrow[n\to\infty]{loi} X \quad \text{si} \quad \int f\, dP_{X_n} \xrightarrow[n\to\infty]{} \int f\, dP_X \text{ pour toute } f \text{ continue bornée.}$$

B.5 Théorèmes limites

On pose $S_n = \sum_{k=1}^n X_k$. Alors $E(S_n) = \sum_{k=1}^n E(X_k)$. De plus,

$$V(S_n) = \sum_{k=1}^n V(X_k) \quad \text{si les } X_i \text{ sont indépendantes.}$$

Loi faible des grands nombres

$$\text{Si les } X_i \text{ sont i.i.d. intégrables,} \quad \frac{S_n}{n} \xrightarrow[n\to\infty]{proba} E(X_1)$$

Loi forte des grands nombres

$$\text{Si les } X_i \text{ sont i.i.d. intégrables,} \qquad \frac{S_n}{n} \xrightarrow[n\to\infty]{p.s.} E(X_1)$$

Théorème de la limite centrée

$$\text{Si les } X_i \text{ sont i.i.d. centrées telles que } 0 < \sigma(X_i)^2 < \infty, \qquad \frac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{loi} \mathcal{N}(0,\sigma^2)$$

Convergence des martingales bornées dans L^2

$$\text{Si } M_n \text{ est une martingale et } \|M_n\|_2 \le C,\ M_n \text{ converge p.s. et } L^2$$

Convergence de la somme dans le cas de variance bornée

$$\text{Si les } X_i \text{ sont indépendantes centrées et } \sum V(X_i) < \infty, S_n \text{ converge p.s. et } L^2$$

Théorème des trois séries

Soit $Y_i = X_i \mathbf{1}_{(|X_i|\le 1)}$. Si les X_i sont indépendantes,

$$S_n \text{ converge p.s. } \Leftrightarrow \sum P(|X_i| \ge 1),\ \sum E(Y_i),\ \sum V(Y_i) \text{ convergent.}$$

Convergence des espérances conditionnelles

Si X est de carré intégrable et les $\mathcal{F}_i$ sont croissantes et engendrent $\mathcal{T}$,

$$E(X \mid \mathcal{F}_n) \xrightarrow[n\to\infty]{p.s.\ et\ L^2} X$$

B.6 Espérance conditionnelle

Caractérisation de l'espérance conditionnelle

$$E(X \mid \mathcal{F}) \text{ est } \mathcal{F}\text{-mesurable}$$

$$E(E(X \mid \mathcal{F})\, Y) = E(XY) \text{ pour tout } Y\ \mathcal{F}\text{-mesurable}$$

Propriétés

$$E(E(X \mid \mathcal{F})) = E(X)$$

$$E(YX \mid \mathcal{F}) = Y E(X \mid \mathcal{F}) \quad \text{si } Y \text{ est } \mathcal{F}\text{-mesurable}$$

$$E(E(X \mid \mathcal{F}_1) \mid \mathcal{F}_2) = E(E(X \mid \mathcal{F}_2) \mid \mathcal{F}_1) = E(X \mid \mathcal{F}_1) \quad \text{si } \mathcal{F}_1 \subset \mathcal{F}_2$$

Conditionnement relativement à une variable aléatoire

$$\int f(x)\, E(g(Y) \mid X = x)\, dP_X(x) = \int f(x)\, g(y)\, dP_{(X,Y)}(x,y) \ \text{ si } f \text{ est bornée}$$

Cas discret

$$E(g(Y) \mid X = x) = \frac{E(g(Y)\, \mathbf{1}_{(X=x)})}{P(X=x)} = \frac{1}{P(X=x)} \int_{(X=x)} g(Y)\, dP$$

Cas continu

$$E(g(Y) \mid X = x) = \frac{\int_{\mathbf{R}} g(y) f_{X,Y}(x,y)\, dy}{\int_{\mathbf{R}} f_{X,Y}(x,y)\, dy}$$

B.7 Martingales

Caractérisation

$$M_n \textit{ est } \mathcal{F}_n - \textit{mesurable}, \quad E(M_{n+1} \mid \mathcal{F}_n) = M_n \;\; \textit{pour tout } n.$$

Propriétés

$$E(M_{n+k} \mid \mathcal{F}_n) = M_n \qquad \text{pour tout } k > 0.$$

$$E(M_n) = E(M_0) \textit{ pour tout } n.$$

Annexe C

Références

Références en français concernant les probabilités

Jean Jacod, Philip Protter
L'essentiel en théorie des probabilités
Cassini. ISBN 978-2842250508

Dominique Foata, Aimé Fuchs
Calcul des probabilités
Dunod. ISBN 978-2100574247

Références en anglais concernant les probabilités

Rick Durrett
Probability : theory and examples.
Cambridge University Press. ISBN 978-0-521-76539-8

Patrick Billingsley
Probability and measure.
John Wiley & Sons, Inc. ISBN 0-471-00710-2

Référence en français pour la théorie de l'intégration

Marc Briane, Gilles Pagès
Théorie de l'intégration, cours et exercices
Vuibert. ISBN 978-2311402261

Référence en anglais pour la théorie de l'intégration

Richard Mansfield Dudley
Real analysis and probability.
Cambridge University Press. ISBN 0-521-00754-2

Yves Coudène, Janvier 2015.