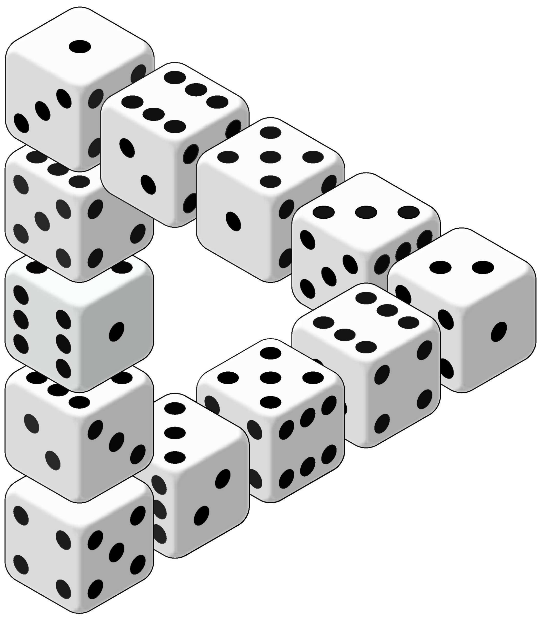
Probabilités II

3M290

Yves Coudène, 11 janvier 2022

Licence de mathématiques, Sorbonne Université

Version 1.1



2021-2022

Table des matières

Introduction

Ces notes accompagnent le cours de probabilités 3M290 *probabilités II* donné au second semestre des années universitaire 2020-2021 et 2021-2022.

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. 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, 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 séries aléatoires, comme corollaire des théorèmes de convergence pour ces séries. 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 étudie d'abord le cas unidimensionnel avant de passer à l'étude des vecteurs gaussiens et à la version multidimensionnelle du TCL. On termine par l'étude des séries de variables aléatoires indépendantes.

Le second appendice contient des rappels d'intégration qui couvrent les résultats utilisés dans le cours, énoncés avec le vocabulaire des probabilités.

Yves Coudène, le 11 janvier 2022.

Notations

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

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
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)
$P \otimes Q$ .... produit des probabilités P et Q
$p.s.$ .... presque sûrement
S_n .... somme de X_1 à X_n
$\sigma(X)$ .... écart-type de X
Σ .... matrice de covariance
$\mathcal{T}$ .... tribu
$\mathcal{T}_1 \otimes \mathcal{T}_2$ .... produit des tribus $\mathcal{T}_1$ et $\mathcal{T}_2$
$V(X)$ .... 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 évènements, un nombre réel compris entre 0 et 1, qui reflète notre degré de confiance dans la réalisation de l'évènement une fois que l'épreuve a eu lieu.

La théorie moderne des probabilités est formalisée par Kolmogorov en 1933, en faisant appel à 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* évènements*,*
– *d'une mesure* P *définie sur la tribu* $\mathcal{T}$*, qui satisfait* $P(\Omega) = 1$*.*

Une mesure P satisfaisant $P(\Omega) = 1$ est appelée *mesure de probabilité*. 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 $\omega \in \Omega$,
– $\sum_{\omega \in \Omega} p_\omega = 1$.

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 :

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

Cette mesure est supportée par le singleton $\{\omega\}$.

$$\delta_\omega(\{\omega\}) = 1, \quad \delta_\omega(\{\omega\}^c) = 0.$$

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 :

– la fonction f est borélienne,

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

– $\int_\Omega f d\lambda = 1$.

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 $\sigma = 1$. On l'appelle aussi loi gaussienne.

– 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 $\Omega^n = \Omega \times \Omega \times ... \times \Omega$, contenant 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 $\mathcal{T} \otimes \mathcal{T} \otimes ... \otimes \mathcal{T} = \mathcal{T}^{\otimes n}$. 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

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

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*

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

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 $P(\{face\}) = P(\{pile\}) = 1/2$.

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 son carré est intégrable :*

$$E(X^2) = \int_\Omega X^2\,dP < +\infty.$$

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 $E(|X|) \leq \sqrt{E(X^2)}$.

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 obtenons 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 elle satisfait

$$\sum_{\omega\in\Omega} p_\omega |X(\omega)| < +\infty, \text{ auquel cas } E(X) = \sum_{\omega\in\Omega} p_\omega X(\omega).$$

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 évènement $A \in \mathcal{T}$, $P(A) = E(\mathbf{1}_A)$.

– Soit (X_n) une suite de variables aléatoires positives 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 $E(X_n) \xrightarrow[n\to\infty]{} E(X)$.

– $V(\lambda X) = \lambda^2 V(X)$.

– $V(X+Y) = V(X) + V(Y) + 2\,\mathrm{Cov}(X,Y)$

La covariance de X, Y a été notée

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y),$$

elle 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 des 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

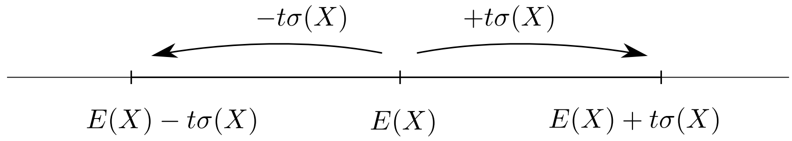
$$E(\lambda \mathbf{1}_{(Y\geq\lambda)}) = \lambda\, E(\mathbf{1}_{(Y\geq\lambda)}) = \lambda\, P(Y \geq \lambda).$$

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* $\sigma(X) = \sqrt{V(X)}$.

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



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 répartition 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 $P_X(\mathcal{D}) = 1$. Indexons 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 $a, b \in \mathbf{R}$,

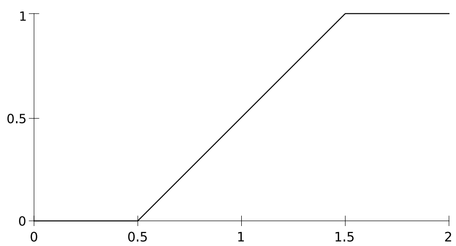
$$P(X \in]a,b]) = F_X(b) - F_X(a).$$

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.

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

Exemple
La fonction de répartition de la loi uniforme sur l'intervalle $[a,b]$, $a < b$, est donnée par

$$\begin{aligned} F_X(x) &= \frac{1}{b-a}\int_{-\infty}^{x} \mathbf{1}_{[a,b]}(x)\,dx \\ &= \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases} \end{aligned}$$



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

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

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

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 de $(X_1, ..., X_n)$ en remarquant que pour tout borélien $A \subset \mathbf{R}$,

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

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 de $(X_1, ..., X_n)$ grâce à la formule suivante :

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

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

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

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 évènements, des tribus ou des variables aléatoires.

3.1 Indépendance d'évènements et de variables aléatoires

On commence par définir la notion d'évènements indépendants.

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

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

Soit $(A_i)_{i \in I}$ *une famille d'évènements. Ces évènements 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 évènements $\{A_1, A_2, A_3\}$, $I = \{1, 2, 3\}$, ces conditions s'écrivent comme suit :

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

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 évènements $(X \in A)$ 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 évènements $(X_i \in A_i)$ 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 évènements 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'évènements.

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

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

Si $\sum_{i \in \mathbf{N}} P(A_i) = +\infty$, 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.

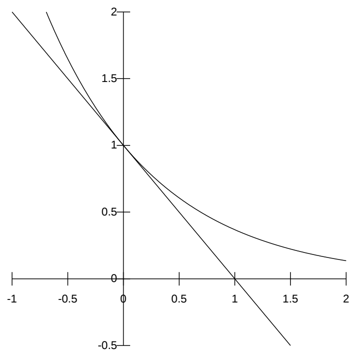
$$\int \#\{i\in\mathbf{N}\mid\omega\in A_i\}\,dP(\omega) = \int\sum_{i\in\mathbf{N}}\mathbf{1}_{A_i}\,dP = \sum_{i\in\mathbf{N}}P(A_i) < +\infty.$$

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,

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

Exemple
On considère une suite de variables aléatoires $(X_n)_{n\in\mathbf{N}}$ indépendantes entre elles et $x\in\mathbf{R}$ tel qu'il existe $\delta > 0$ pour lequel $P(X_n = x)\geq\delta$ pour tout n.

Comme application du lemme de Borel-Cantelli, montrons que pour presque tout $\omega \in \Omega$, le réel x apparaît consécutivement un nombre arbitrairement grand de fois dans la suite $(X_n)_{n\in\mathbf{N}}$. Pour cela, on se donne $N \in \mathbf{N}^*$ et on pose

$$A_i = (X_{Ni+1} = x, X_{Ni+2} = x, ..., X_{Ni+N} = x).$$

Les A_i sont indépendants entre eux et $P(A_i) \geq \delta^N$ pour tout i. D'après le lemme, il existe un ensemble Ω_N de complémentaire négligeable, tel que tout $\omega \in \Omega_N$ appartient à une infinité de A_i, si bien que x apparaît consécutivement N fois dans la suite $(X_i(\omega))$. Pour les ω appartenant à l'intersection des Ω_N, $N \in \mathbf{N}^*$, le nombre x apparaît consécutivement un nombre arbitrairement grand de fois dans la suite $(X_i(\omega))$.

Comme corollaire, considérons une suite (X_i) de variables indépendantes telles que $P(X_i = 1) = P(X_i = -1) = 1/2$ pour tout i. D'après ce qui précède, presque sûrement, il existe des suites arbitrairement longues de 1 consécutives dans la suite (X_i), qui ne peut donc pas être bornée. Dans un jeu de Pile ou Face, où la fortune des joueurs est bornée, on aboutit donc presque sûrement à la ruine d'un des deux joueurs.

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 $(X_i)_{i=1...n}$ 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, $f = \mathbf{1}_A$, $g = \mathbf{1}_B$,

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

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

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 pour f, 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

$$P_{(X_1,...,X_n)} = P_{X_1} \otimes P_{X_2} \otimes ... \otimes P_{X_n},$$
$$E(h(X_1, ..., X_n)) = \int_{\mathbf{R}^n} h(x_1, ..., x_n)\,dP_{X_1}(x_1)...dP_{X_n}(x_n)$$

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

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 fonction $f : \mathbf{R} \to \mathbf{R}$ 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.

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 évènement 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

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

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 $\lim_{n\to\infty} Y_n(\omega) = 0$, comme souhaité.

Preuve de la loi forte des grands nombres
Pour simplifier, nous allons supposer que les X_i sont de carrés intégrables 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 $E(\frac{S_i}{i}) = E(X_1) = 0$.

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

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

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,

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

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

$$\Omega = \{\text{pile}, \text{face}\}^{\otimes\mathbf{N}}, \quad \mathcal{T} = \mathcal{P}(\{\text{pile}, \text{face}\})^{\otimes\mathbf{N}}, \quad P = \Big(\frac{1}{2}\,\delta_{\text{pile}} + \frac{1}{2}\,\delta_{\text{face}}\Big)^{\otimes\mathbf{N}}.$$

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

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

Presque tout $\omega \in \Omega$ appartient à tous les $\Omega_{(a_1,\dots,a_m)}$. Cela signifie que dans presque toute suite $\omega \in \Omega$, tous les mots $(a_1, \dots, a_m)$ 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 → 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.

Chapitre 5

Convergence de suites 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*

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

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* $\mathbf{R}$ *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

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

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

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 $P_X = P_Y$.*

Cette proposition sera démontrée à la fin du chapitre. 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 $\varphi_Y(t) = \sum_{k=1}^n e^{itk} P(Y = k) = \sum_{k=1}^n \frac{1}{n}(e^{it})^k = \frac{1}{n}\, e^{it} \sum_{k=0}^{n-1} (e^{it})^k$.

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

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}, $t \in \mathbf{R}$.*

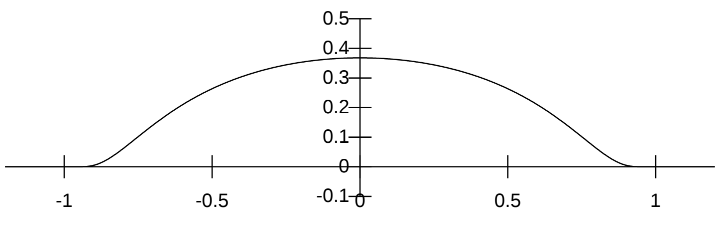
Le premier point correspond à la convergence étroite des μ_n vers μ. Le dernier point correspond à la convergence des transformées de Fourier des μ_n. On en déduit le corollaire suivant.

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.*

Considérons une suite de variables aléatoires (Y_n). 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 $\hat{f}$ est intégrable et*

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

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

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

Ici on a utilisé le théorème de Fubini pour intervertir les deux intégrales. L'emploi de ce théorème est justifié car 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 la convergence $\int f d\mu_n \longrightarrow \int f d\mu$ pour toute fonction f de classe 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 pour tout $t \in \mathbf{R}$, $\widehat{\mu}_n(t) \xrightarrow[n\to\infty]{} \widehat{\mu}(t)$ 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.

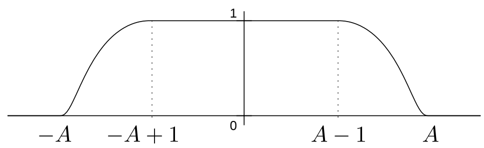
Lemme 4 *Pour tout $\varepsilon > 0$, il existe $A > 0$ tel que pour tout $n \in \mathbf{N}$,*

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

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 $n \geq n_0, \mu_n([-A_0, A_0]) \geq 1 - \varepsilon$.

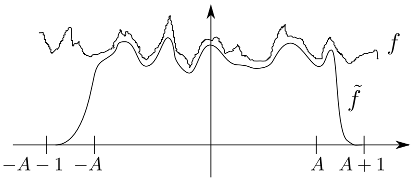
De plus, pour chaque $k \in \{0, ..., n_0\}$, on peut trouver un ensemble A_k tel que $\mu_k([-A_k, A_k]) \geq 1 - \varepsilon$. Pour tout A supérieur à $\max\{A_0, ..., A_{n_0}\}$, on a

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

Le lemme est démontré.

Soit f continue bornée. Sur $[-A-1, A+1]$, 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 $n \geq \mathbf{N}$, $|\int \tilde{f} d\mu_n - \int \tilde{f} d\mu| < \varepsilon$.

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

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 $P(X = a) = P(X = b) = 0$,*

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

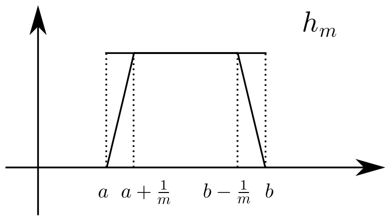
De plus, les fonctions de répartition des X_n convergent vers la fonction de répartition de X en tout point $x \in \mathbf{R}$ tel que $P(X = x) = 0$:

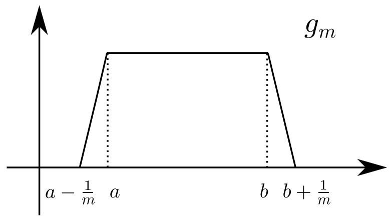
$$F_{X_n}(x) \xrightarrow[n\to\infty]{} F_X(x) \quad si \quad P(X = x) = 0.$$

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 $m \in \mathbf{N}$ 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,

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

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

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

ce qui donne le résultat recherché :

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

On termine ce chapitre par la preuve de deux résultats énoncés précédemment.

Proposition 12 *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 $n \longrightarrow \infty$.

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

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.

Proposition 13 *Deux variables aléatoires qui ont même fonction caractéristique ont même loi.*

Preuve
Notons μ et ν les lois des variables aléatoires et considérons la suite constante $\mu_n = \nu$. La suite $\int f\, d\mu_n$ est constante égale à $\int f\, d\nu$, les convergences dans le théorème 6 deviennent des égalités. On a donc équivalence entre l'égalité $\int f\, d\nu = \int f\, d\mu$ pour toute fonction f de la forme $f(x) = e^{itx}$ et la même égalité pour toute fonction f continue bornée. On en déduit que les deux mesures sont égales dès qu'elles ont même fonction caractéristique.

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 8 *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.

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

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 la relation

$$\begin{array}{rcl} I_{2l+2} & = & \int y^{2l+1} y e^{-y^2/2} dy \\ & = & [y^{2l+1}(-e^{-y^2/2})]_{-\infty}^{+\infty} + \int (2l+1) y^{2l} e^{-y^2/2} dy \\ & = & (2l+1) I_{2l}. \end{array}$$

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

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

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

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 9 *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*

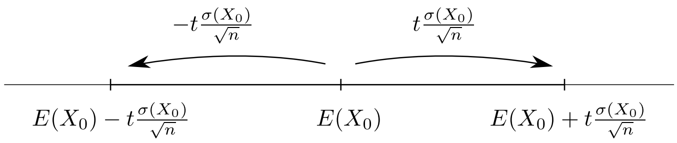
$$\frac{\sqrt{n}}{\sigma(X_0)} \Big(\frac{S_n}{n} - E(X_0) \Big)$$

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'évènement 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

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

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

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

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

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 :

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

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

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

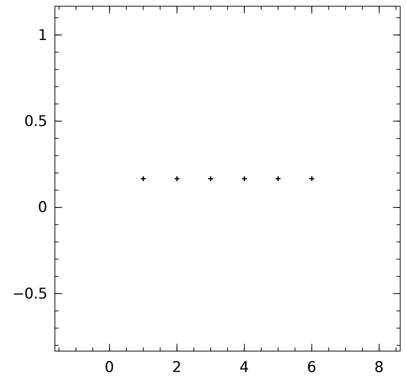
Le théorème de la limite centrée est démontré.

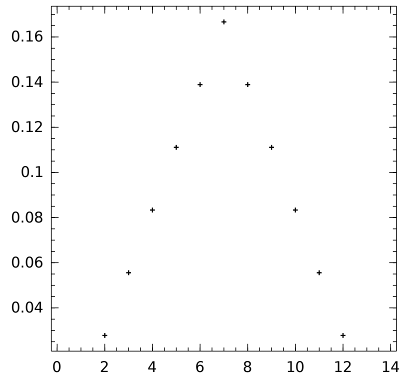
6.3 Illustration numérique

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 $S_2 = X_1 + X_2$.



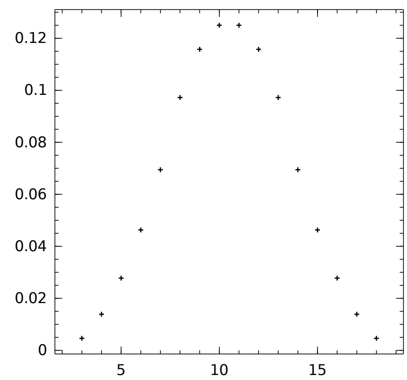


Graphe des fréquences de X_0 Graphe des fréquences de $S_2 = X_1 + X_2$

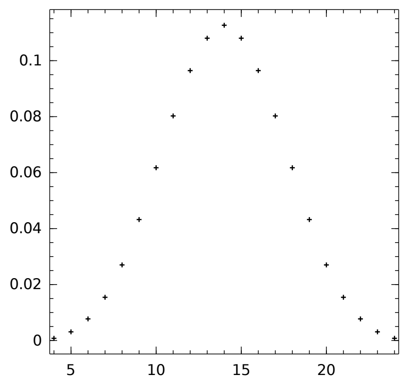
On calcule le graphe des fréquences de S_n pour tout n par récurrence en utilisant la formule

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

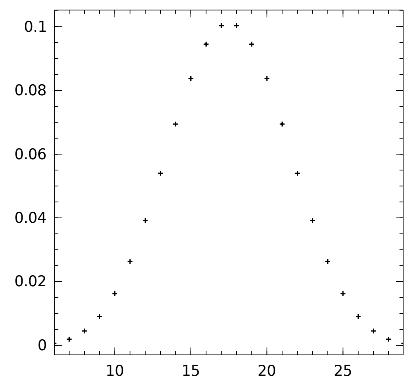
où la somme porte sur l'ensemble des valeurs l que prend X_{n+1}. Si n est suffisamment grand, le graphe des fréquences devrait se rapprocher d'une gaussienne, une fois renormalisé. On s'est restreint ci-dessous à des valeurs de x à moins de trois fois l'écart-type de l'espérance de S_n.



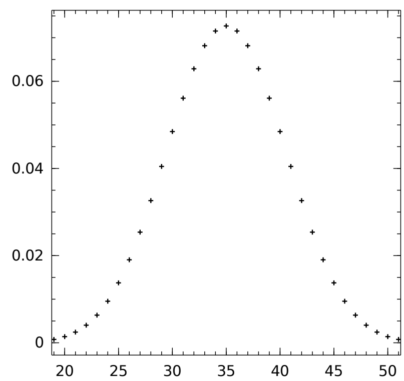
Graphe de S_3



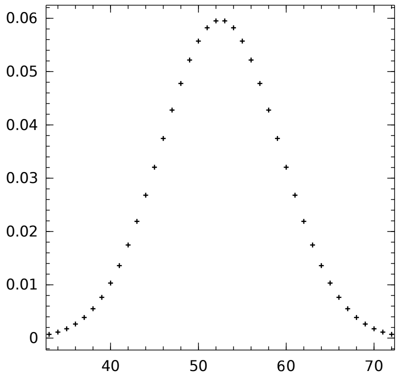
Graphe de S_4



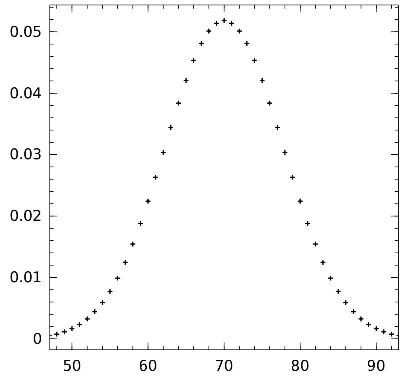
Graphe de S_5



Graphe de S_{10}



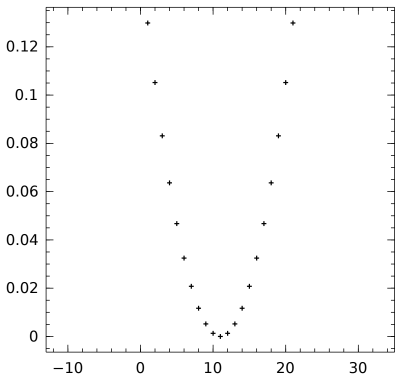
Graphe de S_{15}



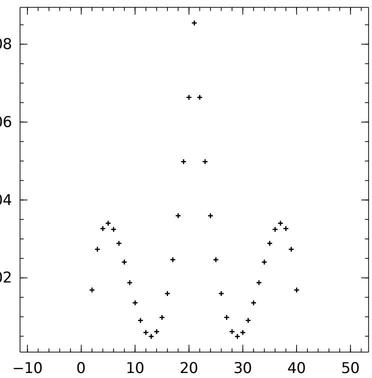
Graphe de S_{20}

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

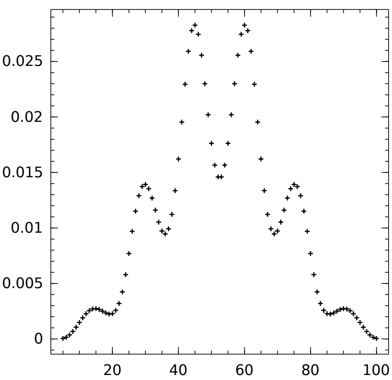
Il est intéressant de regarder ce qu'on obtient lorsqu'on part d'une loi qui présente plusieurs maxima. Prenons pour X_0 la loi $P(X_0 = k) = \frac{(k-11)^2}{770}$ pour k compris entre 1 et 21. Le graphe des fréquences de X_0 est ci-dessous.



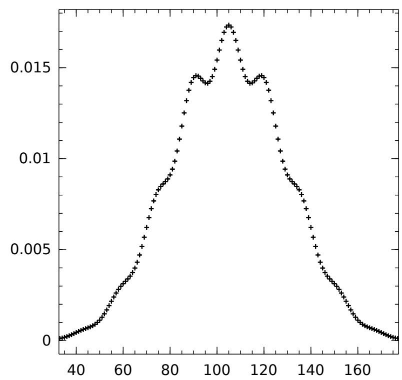
Graphe de X_0



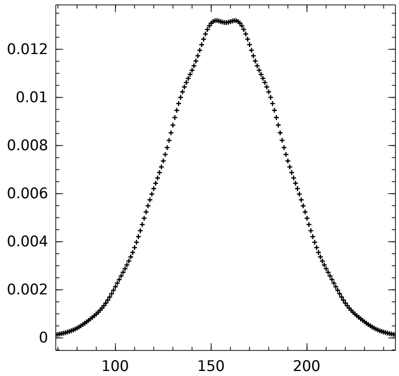
Graphe de S_2



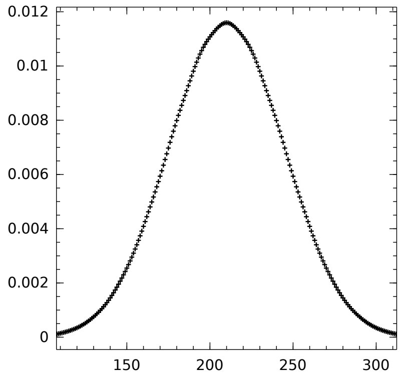
Graphe de S_5



Graphe de S_{10}



Graphe de S_{15}



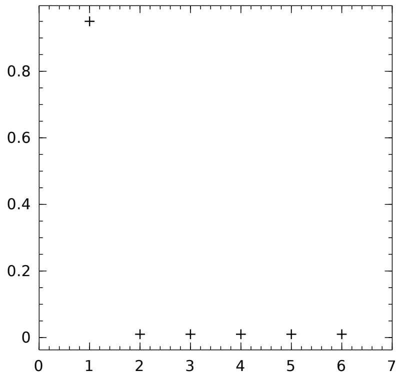
Graphe de S_{20}

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

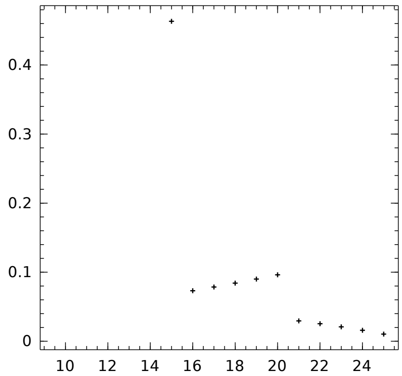
Un autre cas intéressant est donné par une loi fortement dissymétrique. Considérons un X_0 pour lequel

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

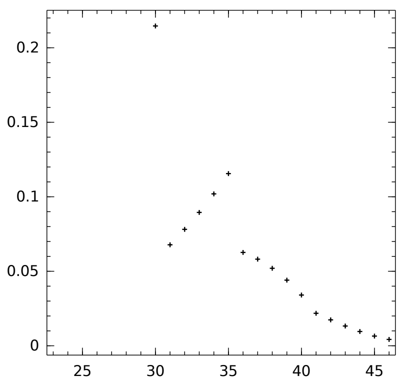
$$P(X_0 = 2) = P(X_0 = 3) = P(X_0 = 4) = P(X_0 = 5) = P(X_0 = 6) = 0,01.$$



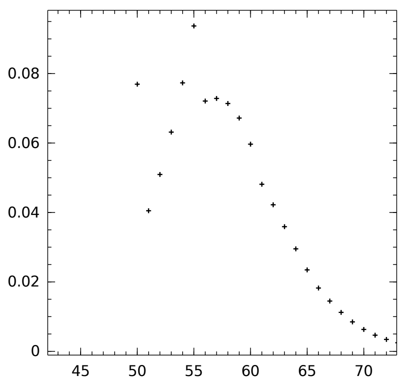
Graphe de X_0



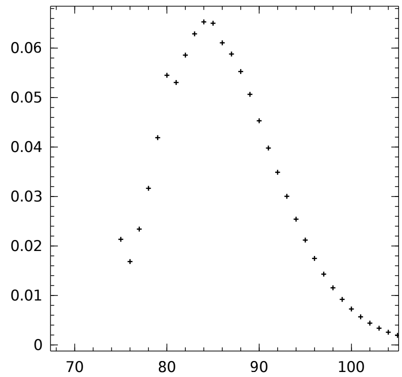
Graphe de S_{15}



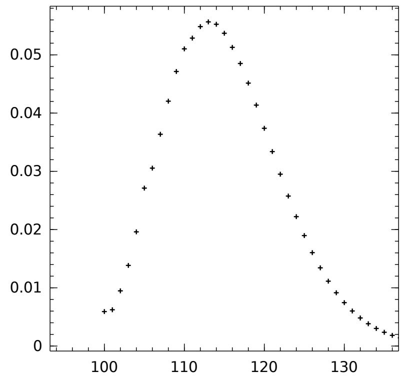
Graphe de S_{30}



Graphe de S_{50}



Graphe de S_{75}



Graphe de S_{100}

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.

Chapitre 7

Vecteurs aléatoires

Dans ce chapitre, nous introduisons la notion de vecteur gaussien afin de généraliser le théorème de la limite centrée au cas multidimensionnel.

7.1 Variables aléatoires à valeurs vectorielles

Définition 15 *Soit* $(\Omega, \mathcal{T}, P)$ *un espace probabilisé et* X_1, ..., X_d *des variables aléatoires définies sur* Ω *à valeurs réelles. L'application définie sur* Ω *à valeurs dans* $\mathbf{R}^d$

$$\omega \mapsto \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_d(\omega) \end{pmatrix}$$

est appelée vecteur aléatoire.

Les notions d'espérance, de covariance et de fonction caractéristique se généralisent aux vecteurs aléatoires. L'*espérance* de $(X_1, ..., X_d)$ est maintenant un vecteur donné par

$$(E(X_1), ..., E(X_d)).$$

Sa *matrice de covariance*, parfois notée Σ, est de taille $d \times d$ et vaut

$$V(X) = \{Cov(X_i, X_j)\}_{i,j}.$$

Sa *fonction caractéristique* est définie sur $\mathbf{R}^d$ par la formule

$$\forall\, u \in \mathbf{R}^d,\ \varphi_X(u) = E(e^{iu.X}) = E(\prod_k e^{iu_k X_k}).$$

On emploie dans la suite la notation $u.X = \sum u_i X_i$, $u \in \mathbf{R}^d$. Les vecteurs u et X sont considérés comme des vecteurs colonnes.

Proposition 14 *Soit X un vecteur aléatoire à valeurs dans $\mathbf{R}^d$ et $u = (u_1, ..., u_d)$ un vecteur de $\mathbf{R}^d$. Alors*

$$E(u.X) = u.E(X), \quad V(u.X) = {}^t uV(X)\, u.$$

Preuve
Ces formules sont une conséquence directe des propriétés de l'espérance et de la variance.

$$E(u.X) = E(\sum u_i X_i) = \sum u_i E(X_i) = u.E(X).$$

$$V(u.X) = V(\sum u_i X_i) = \sum_{i,j} u_i u_j Cov(X_i, X_j) = {}^t u\, V(X)\, u.$$

Notons que l'application $u \mapsto V(u.X)$ est une forme quadratique définie sur $\mathbf{R}^d$. Cette forme quadratique est positive : pour tout $u \in \mathbf{R}^d$,

$${}^t u\, V(X)\, u = V(u.X) \geq 0.$$

La plupart des notions concernant les variables aléatoires admettent un analogue pour les vecteurs aléatoires. Nous avons vu dans un chapitre précédent comment définir la loi d'un multiplet de variables aléatoires. Un vecteur gaussien étant un multiplet, ces définitions s'appliquent ici. La *loi* d'un vecteur $X = (X_1, ..., X_n)$ est donc une mesure de probabilité définie sur la tribu des boréliens de $\mathbf{R}^d$ par la formule

$$P_X(A) = P(X \in A), \text{ pour tout } A \subset \mathbf{R}^d \text{ borélien.}$$

On parle de vecteurs discret ou à densité suivant que cette loi est discrète ou absolument continue par rapport à la mesure de Lebesgue sur $\mathbf{R}^d$. On dit également que deux vecteurs $X = (X_1, ..., X_d)$ et $Y = (Y_1, ..., Y_d)$ sont *indépendants entre eux* si $P_{(X,Y)} = P_X \otimes P_Y$, c'est-à-dire si

$$P_{(X_1,...,X_n,Y_1,...,Y_n)} = P_{(X_1,...,X_n)} \otimes P_{(Y_1,...,Y_n)}.$$

La notion d'intégrabilité se généralise aussi sans difficulté aux vecteurs gaussiens. Le vecteur aléatoire X est *intégrable* si $E(\|X\|) < \infty$, où $\|.\|$ désigne une norme sur $\mathbf{R}^d$, par exemple la norme euclidienne. Il est de carré intégrable si $E(\|X\|^2) < \infty$ et ainsi de suite.

7.2 Définition des vecteurs gaussiens

On s'intéresse à la généralisation de la loi normale au cas multidimensionnel.

Définition 16 *Un vecteur aléatoire* $(X_1, ..., X_d)$ *est dit* gaussien *si pour tout* u_1,..., $u_d \in \mathbf{R}$, *la somme*

$$u_1X_1 + ... + u_dX_d$$

suit une loi normale ou est constante.

On convient ici de considérer la masse de Dirac δ_m comme une loi normale de variance nulle et de moyenne égale à $m \in \mathbf{R}$. Avec cette convention, une variables aléatoire constante suit une loi normale d'écart-type nul.

Remarquons que les composantes X_i d'un vecteur gaussien suivent une loi normale. Il suffit de prendre tous les u_i nuls sauf un pour s'en convaincre. Donnons un premier exemple de vecteur gaussien.

Proposition 15 *Soit* X_1,..., X_d *des variables aléatoires indépendantes entre elles, suivant chacune une loi normale. Alors le vecteur* $(X_1, ..., X_d)$ *est gaussien et sa matrice de covariance est diagonale.*

Lemme 5 *Soit* a, b, c *trois nombres réels et soit* Y_1, Y_2 *deux variables aléatoires indépendantes entre elles suivant une loi normale centrée réduite :*

$$E(Y_1) = E(Y_2) = 0, \quad V(Y_1) = V(Y_2) = 1.$$

Alors la variable aléatoire $aY_1 + bY_2 + c$ *suit une loi normale d'espérance égale à* c *et de variance égale à* $a^2 + b^2$.

Preuve du lemme
Posons $Z = aY_1 + bY_2 + c$ et soit $g : \mathbf{R} \to \mathbf{R}$ une fonction mesurable bornée.

$$\begin{aligned} E(g(Z)) &= \iint g(ay_1 + by_2 + c)\, dP_{Y_1}(y_1)dP_{Y_2}(y_2) \\ &= \frac{1}{2\pi} \iint g(ay_1 + by_2 + c)\, e^{-\frac{y_1^2+y_2^2}{2}}\, dy_1dy_2. \end{aligned}$$

Effectuons le changement de variables

$$\begin{cases} z_1 = ay_1 + by_2 + c, \\ z_2 = by_1 - ay_2. \end{cases}$$

Un calcul direct donne les relations

$$(z_1 - c)^2 + z_2^2 = (a^2 + b^2)(y_1^2 + y_2^2), \quad dz_1dz_2 = (a^2 + b^2)\, dy_1dy_2$$

ce qui implique

$$\begin{aligned} E(g(Z_1)) &= \frac{1}{2\pi(a^2 + b^2)} \iint g(z_1)\, e^{-\frac{(z_1-c)^2}{2(a^2+b^2)}}\, e^{-\frac{z_2^2}{2(a^2+b^2)}}\, dz_2dz_1 \\ &= \frac{1}{\sqrt{2\pi(a^2 + b^2)}} \int g(z_1)\, e^{-\frac{(z_1-c)^2}{2(a^2+b^2)}}\, dz_1. \end{aligned}$$

Il s'agit bien d'une loi normale de paramètres c et $a^2 + b^2$.

Preuve de la proposition
On commence par le cas de deux variables aléatoires X_1, X_2 d'espérances m_1, m_2 et de variances σ_1^2 et σ_2^2. Supposons ces variances non nulles et renormalisons ces variables en posant $Y_i = \frac{X_i - m_i}{\sigma_i}$. Le lemme affirme que toute combinaison linéaire des variables Y_1, Y_2 suit une loi normale. Il en va donc de même pour la variable

$$u_1X_1 + u_2X_2 = u_1m_1 + u_2m_2 + u_1\sigma_1Y_1 + u_2\sigma_2Y_2.$$

Le cas de n variables aléatoires s'en déduit par une récurrence immédiate sur n. La proposition est démontrée.

On crée de nouveaux vecteurs gaussiens en appliquant une transformation affine à un vecteur gaussien.

Proposition 16 *Soit $X = (X_1, ..., X_d)$ un vecteur gaussien, A une matrice de taille $d' \times d$ et B un vecteur de $R^{d'}$. Alors le vecteur $AX + B$ est gaussien à valeurs dans $\mathbf{R}^{d'}$ et*

$$E(AX + B) = AE(X) + B, \quad V(AX + B) = A\,V(X)\,{}^tA.$$

Preuve
Toute combinaison linéaire des coordonnées du vecteur $AX + B$ est combinaison linéaire des coordonnées des X_i et du vecteur constant égal à 1, on obtient bien une loi normale. Notons $a_{i,j}$ les coefficients de A, b_i ceux de B.

$$E((AX + B)_i) = E(\sum_j a_{i,j}X_j + b_i) = \sum_j a_{i,j}E(X_j) + b_i = (AE(X) + B)_i,$$

$$\begin{aligned} Cov((AX + B)_i, (AX + B)_j) &= Cov\Big(\sum_k a_{i,k}X_k + B_k, \sum_l a_{j,l}X_l + B_l\Big) \\ &= \sum_{k,l} a_{i,k}a_{j,l}\,Cov(X_k, X_l) \end{aligned}.$$

C'est le coefficient i, j de la matrice $A\,V(X)\,{}^tA$. La proposition est démontrée.

7.3 Loi des vecteurs gaussiens

Nous allons montrer que la loi d'un vecteur gaussien ne dépend que de son vecteur espérance et de sa matrice de covariance et nous allons déterminer explicitement sa densité. Commençons par calculer la fonction caractéristique d'un vecteur gaussien. Rappelons que la fonction caractéristique d'une loi normale de paramètres m, σ^2 est égale à

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

Proposition 17 *Soit $X = (X_1, ..., X_d)$ un vecteur gaussien d'espérance m et de matrice de covariance Σ. Alors*

$$\varphi_X(u) = e^{iu.m - \frac{1}{2}{}^t u \Sigma u}.$$

Preuve
Nous savons que la variable $u.X$ obéit à une loi normale et nous avons calculé son espérance et sa variance.

$$E(u.X) = u.E(X) = u.m, \quad V(u.X) = {}^t u V(X) u = {}^t u \Sigma u.$$

On en déduit que $E(e^{itu.X}) = e^{itu.m - \frac{1}{2}{}^t u \Sigma u\, t}$ pour tout t et le résultat s'ensuit.

Tout comme dans le cas d'une variable aléatoire, on peut montrer que la fonction caractéristique d'un vecteur aléatoire caractérise de manière unique la loi d'un vecteur aléatoire. La loi d'un vecteur gaussien est donc uniquement déterminée par m et Σ.

Pour calculer la densité d'un vecteur gaussien, nous allons avoir besoin de quelques propriétés des matrices symétriques. Rappelons qu'une matrice symétrique Σ est positive si ${}^t u \Sigma u \geq 0$ pour tout $u \in \mathbf{R}^d$, et définie positive si ${}^t u \Sigma u > 0$ pour tout u non nul. Une matrice symétrique positive est définie positive si et seulement si elle est inversible, c'est-à-dire de déterminant non nul.

Théorème 10 *Tout vecteur gaussien X a même loi qu'un vecteur gaussien de la forme $AY + B$, où Y est un vecteur gaussien dont les coordonnées sont indépendantes, identiquement distribuées et obéissent à la loi normale $\mathcal{N}(0,1)$.*

Preuve
La preuve repose sur le résultat suivant : toute matrice symétrique S positive est de la forme TD^tT, où T est une matrice inversible et D une matrice diagonale dont les coefficients valent 0 ou 1. La matrice T est obtenue par l'algorithme de réduction de Gauss appliqué à la forme quadratique $u \mapsto {}^t uSu$. On peut aussi la construire en diagonalisant S en base orthonormée. Prenons pour matrice symétrique la matrice $V(X)$ qui est bien positive :

$${}^t u V(X) u = V(u.X) \geq 0.$$

Considérons un vecteur gaussien $Y = (Y_1, ..., Y_d)$ dont les composantes sont indépendantes entre elles, centrées et dont la matrice de covariance est égale à D. Posons $Z = TDY + E(X)$, calculons son espérance et sa variance.

$$E(Z) = TDE(Y) + E(X) = E(X),$$

$$V(Z) = V(TDY) = TDV(Y)\,{}^t D\,{}^t T = TD\,{}^t D\,{}^t T = T\,D\,{}^t T = V(X).$$

Les vecteurs Z et X sont gaussiens, ils ont même espérance et même variance. Ils ont donc même fonction caractéristique et même loi. La proposition est démontrée.

Remarquons que la preuve précédente montre que toute matrice symétrique positive est la matrice de covariance d'un vecteur gaussien. Nous sommes maintenant en mesure de déterminer la densité des vecteurs gaussiens dont la matrice de covariance est définie positive, ou de manière équivalente, inversible, ou encore de déterminant non nul. On dit qu'elle est *non dégénérée.*

Théorème 11 *Soit Σ une matrice symétrique $d \times d$ définie positive et m un vecteur de $\mathbf{R}^d$. Le vecteur $X = (X_1, ..., X_d)$ de densité*

$$f_X(x) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}{}^t(x-m)\Sigma^{-1}(x-m)}, \ x \in \mathbf{R}^d,$$

est un vecteur gaussien d'espérance m et de covariance Σ. Réciproquement, tout vecteur gaussien dont la matrice de covariance Σ est de déterminant non nul a pour densité la fonction f précédente.

Preuve
Employons les décompositions $X = TY + E(X)$ et $\Sigma = TD^tT$ vues dans la proposition précédente. Comme Σ est définie positive, la matrice D est égale à l'identité. Utilisons les notations $x = (x_1, ..., x_d) \in \mathbf{R}^d$, $dx = dx_1...dx_d$ et posons $m = E(X)$.

$$\begin{aligned} E(g(X)) &= E(g(TY + E(X))) \\ &= \frac{1}{(2\pi)^{d/2}} \int g(Ty + m)\, e^{-\frac{1}{2}{}^t yy}\, dy, \\ &= \frac{1}{(2\pi)^{d/2}} \int g(x)\, e^{{}^t(x-m)(T^tT)^{-1}(x-m)}\, \frac{dx}{\det(T)} \end{aligned}$$

grâce au changement de variables $x = Ty + m$, $dx = \det(T)\, dy$. Il suffit de remarquer que $\det(T)^2 = \det(\Sigma)$ pour conclure.

Comme application du théorème précédent, voici un critère concernant l'indépendance des composantes d'un vecteur gaussien.

Corollaire 6 *Soit X un vecteur gaussien. Ses composantes $(X_1, ..., X_d)$ sont indépendantes entre elles si et seulement si la matrice de covariance est diagonale : $Cov(X_i, X_j) = 0$ pour tout i, j distincts.*

En effet, si la matrice de covariance est diagonale, nous voyons sur l'expression que nous avons obtenue que la densité de X est égale au produit des densités des X_i, ce qui montre l'indépendance.

7.4 Théorèmes limites pour les vecteurs aléatoires

Les théorèmes limites pour les variables aléatoires admettent des analogues dans le cas des vecteurs aléatoires.

Lorsque nous parlons de théorème limite, nous voulons étudier le comportement asymptotique d'une suite $\{X_k\}_{k\in\mathbf{N}}$ de vecteurs aléatoires de $\mathbf{R}^d$. Pour éviter toute confusion, nous noterons dans cette section les composantes du vecteur X_i par

$$X_k = (X_k^{(1)}, ..., X_k^{(d)}).$$

La notation X_k représente un terme d'une suite de vecteurs aléatoires de $\mathbf{R}^d$ et non une composante d'un vecteur aléatoire.

Loi des grands nombres

On se place sur un espace probabilisé $(\Omega, \mathcal{T}, P)$. Rappelons que deux vecteurs aléatoires $X = (X^{(1)}, ..., X^{(d)})$ et $Y = (Y^{(1)}, ..., Y^{(d)})$ sont indépendants entre eux si $P_{(X,Y)} = P_X \otimes P_Y$, c'est-à-dire si

$$P_{(X^{(1)},...,X^{(d)},Y^{(1)},...,Y^{(d)})} = P_{(X^{(1)},...,X^{(d)})} \otimes P_{(Y^{(1)},...,Y^{(d)})}.$$

De même, une suite (X_k) de vecteurs aléatoires est indépendante dans son ensemble si pour tout $n \in \mathbf{N}$,

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

Les lois faible et forte des grands nombres se généralisent sans difficulté au cas des vecteurs aléatoires, il suffit de travailler composante par composante. Posons $S_n = X_1 + ... + X_n$, il s'agit d'un vecteur aléatoire qui à tout résultat $\omega \in \Omega$ associe un vecteur de $\mathbf{R}^d$. Notons $S_n^{(i)}$ sa *ième* composante. $S_n^{(i)} = X_1^{(i)} + ... + X_n^{(i)}$.

Théorème 12 *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et $(X_n)_{n\in\mathbf{N}}$ une suite de vecteurs aléatoires à valeurs dans $\mathbf{R}^d$, indépendants, identiquement distribués, intégrables. Alors*

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

La convergence précédente est bien sûr équivalente aux convergences

$$\frac{S_n^{(i)}}{n} \xrightarrow[n\to\infty]{} E(X_0^{(i)}) \text{ presque sûrement}$$

qui découlent du cas unidimensionnel appliqué aux suites $(X_n^{(i)})_{n\in\mathbf{N}}$.

Théorème de la limite centrée multidimensionnel

La limite dans le théorème de la limite centrée pour les vecteurs aléatoires fait intervenir une loi normale multidimensionnelle. Notons $\mathcal{N}(0, \Sigma)$ la loi d'un vecteur gaussien de $\mathbf{R}^d$ dont la matrice de covariance est égale à Σ et d'espérance égale au vecteur nul.

Théorème 13 *Soit* $(\Omega, \mathcal{T}, P)$ *un espace probabilisé et* $(X_n)_{n\in\mathbf{N}}$ *une suite de vecteurs aléatoires à valeurs dans* $\mathbf{R}^d$*, indépendants, identiquement distribués, de carrés intégrables. Soit* m *le vecteur espérance de chacun des* X_n *et* Σ *leur matrice de covariance. On suppose* Σ *inversible. Alors*

$$\frac{S_n - nm}{\sqrt{n}} \longrightarrow \mathcal{N}(0, \Sigma) \text{ en loi.}$$

Cela entraîne les convergences

$$P\Big(nm_i + a_i\sqrt{n} \leq S_n^{(i)} \leq nm_i + b_i\sqrt{n} \text{ pour tout } i \in \{1, ..., d\}\Big)$$

$$\xrightarrow[n\to\infty]{} \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \int_{a_1}^{b_1} ... \int_{a_d}^{b_d} e^{-\frac{1}{2}{}^t x \Sigma^{-1} x}\, dx.$$

La preuve procède comme dans le cas unidimensionnel. On caractérise la convergence en loi grâce aux fonctions caractéristiques.

Proposition 18 *Considérons une suite de vecteurs aléatoires* (Y_n) *définis sur un espace probabilisé* $(\Omega, \mathcal{T}, P)$*. Si pour tout* $u \in \mathbf{R}^d$*,*

$$\varphi_{Y_n}(u) \xrightarrow[n\to\infty]{} \varphi_Y(u)$$

alors Y_n *converge vers* Y *en loi.*

Puis on fait un développement limité de la fonction caractéristique de $\frac{S_n - m}{\sqrt{n}}$. Les calculs que nous avons fait sur $\mathbf{R}$ se généralisent à $\mathbf{R}^d$ sans difficulté.

Chapitre 8

Séries de variables aléatoires indépendantes

Considérons une suite de variables aléatoires $(X_n)_{n\in\mathbf{N}}$ indépendantes entre elles. Que peut-on dire de la convergence de la série

$$S_n = \sum_{k=1}^{n} X_k \text{ ?}$$

Si les variables aléatoires ont même loi, la loi des grands nombres affirme que cette série diverge presque partout dès que leur espérance est non nulle.

Nous allons nous intéresser au cas où les variables aléatoires (X_n) sont indépendantes entre elles mais n'ont pas forcément même loi.

8.1 Loi du 0-1 de Kolmogorov

Notons par $\mathcal{T}_{(X_k, m\leq k\leq n)}$ la tribu engendrée par tous les évènements de la forme $X_k^{-1}(B)$, pour tout k compris entre m et n et tout borélien $B \subset \mathbf{R}$. Considérons également la tribu $\mathcal{T}_{(X_k, k\geq m)}$ associée à tous les indices $k \geq m$. Cette tribu est engendrée par toutes les tribus $\mathcal{T}_{(X_k, m\leq k\leq n)}$, avec $n \geq m$.

Rappelons un résultat d'approximation classique de théorie de la mesure qui s'applique à ces tribus.

Proposition 19 *Soit* $(\Omega, \mathcal{T}, P)$ *un espace probabilisé,* $\mathcal{S}$*,* $\mathcal{S}_n$*,* $n \in \mathbf{N}$ *des tribus incluses dans* $\mathcal{T}$ *telles que* $\mathcal{S}$ *soit engendrée par les* $\mathcal{S}_n$ *et telles que les* $(\mathcal{S}_n)$ *forment une suite croissante pour l'inclusion :*

$$\mathcal{S}_n \subset \mathcal{S}_{n+1} \text{ pour tout } n.$$

Soit $A \in \mathcal{S}$*. Alors, pour tout* $\varepsilon > 0$*, il existe* $n \in \mathbf{N}$ *et* $A' \in \mathcal{S}_n$ *tels que*

$$P(A\Delta A') < \varepsilon.$$

Nous avons noté $A\Delta A'$ la différence symétrique de A et A'.

$$A\Delta A' = (A \cup A') \setminus (A \cap A').$$

Un évènement est dit *asymptotique* s'il appartient aux tribus $\mathcal{T}_{(X_k, k\geq m)}$ pour tout $m \in \mathbf{N}$. L'exemple le plus simple d'évènement asymptotique est donné par l'ensemble des $\omega \in \Omega$ pour laquelle la série $\sum X_k$ converge :

$$\Big\{\omega \in \Omega \,\Big|\, \sum_{k\geq 1} X_k(\omega) \text{ converge}\Big\}.$$

En effet, le caractère convergent ne dépend pas des valeurs prises par les m premières valeurs de la suite $X_k(\omega)$. Pour tout $m \in \mathbf{N}$,

$$\Big\{\omega \in \Omega \,\Big|\, \sum_{k\geq 1} X_k(\omega) \text{ converge}\Big\} = \Big\{\omega \in \Omega \,\Big|\, \sum_{k\geq m} X_k(\omega) \text{ converge}\Big\} \in \mathcal{T}_{(X_k, k\geq m)}.$$

La loi du 0-1 de Kolmogorov affirme que les évènements asymptotiques ont pour probabilité 0 ou 1 si les variables aléatoires sont indépendantes.

Théorème 14 (loi du 0-1 de Kolmogorov) *Soit* $(\Omega, \mathcal{T}, P)$ *un espace probabilisé,* $(X_k)_{k\in\mathbf{N}}$ *une suite de variables aléatoires définies sur* Ω *et indépendantes entre elles. Considérons un évènement* $A \in \mathcal{T}$ *tel que*

$$A \in \bigcap_{m\in\mathbf{N}} \mathcal{T}_{(X_k, k\geq m)}.$$

Alors $P(A)$ *vaut* 0 *ou* 1.

Preuve
Soit $\varepsilon > 0$. Comme A est dans $\mathcal{T}_{(X_k, k\geq 1)}$, il existe $m \in \mathbf{N}$ et $A' \in \mathcal{T}_{(X_1,\ldots,X_m)}$ tels que

$$P(A\Delta A') < \varepsilon.$$

Comme A est aussi dans $\mathcal{T}_{(X_k, k\geq m+1)}$, il existe $n \geq m+1$ et $A'' \in \mathcal{T}_{(X_{m+1},\ldots,X_n)}$ tels que

$$P(A\Delta A'') < \varepsilon.$$

Les ensembles A' et A'' sont indépendants, ce qui montre que

$$P(A' \cap A'') = P(A')P(A'') < (P(A) + \varepsilon)^2.$$

Nous avons aussi, en vertu de l'inclusion $A\Delta(A' \cap A'') \subset (A\Delta A') \cup (A\Delta A'')$,

$$P(A' \cap A'') \geq P(A) - P(A\Delta(A' \cap A'')) > P(A) - 2\varepsilon.$$

Cela implique l'inégalité $P(A) \leq P(A)^2$. L'inégalité inverse découlant du fait que $0 \leq P(A) \leq 1$, nous en déduisons que $P(A) = P(A)^2$ puis que $P(A) = 0$ ou 1. Le théorème est démontré.

Comme corollaire, on voit qu'une série ne peut simultanément converger pour un ensemble de résultats de probabilité non nulle et diverger pour un ensemble de résultats de probabilité non nulle.

Corollaire 7 *Soit* $(X_k)_{k\in\mathbf{N}}$ *une suite de variables aléatoires indépendantes. Alors la série* $\sum X_k$ *converge presque sûrement ou diverge presque sûrement.*

Il s'agit maintenant de donner des critères permettant de conclure à la convergence ou à la divergence de la série presque sûrement.

8.2 Convergence des séries aléatoires

Voici un premier critère de convergence dans le cas de variables aléatoires de carrés intégrables.

Proposition 20 *Soit* (X_k) *une suite de variables aléatoires indépendantes centrées et de carrés intégrables :* $E(X_k) = 0$ *et* $V(X_k) < \infty$ *pour tout* $k \in \mathbf{N}$*. On suppose que*

$$\sum_{k\in\mathbf{N}} V(X_k) < \infty.$$

Alors la série $\sum X_k$ *converge en norme* L^2 *et presque sûrement.*

Preuve de la convergence L^2
Pour montrer la convergence de S_n en norme L^2, montrons qu'elle est de Cauchy. Soit m, $n \in \mathbf{N}$ avec $m < n$.

$$\|S_n - S_m\|_2^2 = V(S_n - S_m) = V\Big(\sum_{k=m+1}^{n} X_k\Big) = \sum_{k=m+1}^{n} V(X_k).$$

La série $\sum V(X_k)$ est convergente donc de Cauchy. Pour tout $\varepsilon > 0$, il existe $N \in \mathbf{N}$ tel que pour tout m, n satisfaisant $N < m < n$,

$$\|S_n - S_m\|_2^2 = \sum_{k=m+1}^{n} V(X_k) < \varepsilon.$$

La suite S_n est de Cauchy et converge en norme L^2.

La démonstration de la convergence presque sûre dans la proposition précédente repose sur l'inégalité maximale de Kolmogorov.

Lemme 6 (inégalité maximale) *Soit* $(X_n)_{n\in\mathbf{N}}$ *une suite de variables aléatoires indépendantes entre elles, centrées et de variance finie :* $E(X_i) = 0$, $V(X_i) < \infty$*. Alors pour tout* $n \in \mathbf{N}$ *et tout* $\lambda > 0$,

$$P\Big(\max_{1\le i\le n} |S_i| \ge \lambda\Big) \le \frac{E(S_n^2)}{\lambda^2}.$$

Preuve du lemme
On s'intéresse au premier indice pour lequel la série dépasse λ.

$A_1 = (|S_1| \geq \lambda)$,
$A_2 = (|S_1| < \lambda, |S_2| \geq \lambda)$,
$A_j = (|S_1| < \lambda, ..., |S_{j-1}| < \lambda, |S_j| \geq \lambda)$.

Nous avons alors

$$\Big(\max_{1 \leq i \leq N} |S_i| \geq \lambda\Big) = \coprod_{j=1}^{N} A_j.$$

Cherchons à minorer $E(S_N^2 \mathbf{1}_{A_j})$ en insérant le terme S_j dans le carré.

$$S_N^2 = (S_N - S_j + S_j)^2 = (S_N - S_j)^2 + S_j^2 + 2(S_N - S_j)S_j$$

$$E(S_N^2 \mathbf{1}_{A_j}) = E((S_N - S_j)^2\, \mathbf{1}_{A_j}) + E(S_j^2\, \mathbf{1}_{A_j}) + 2E((S_N - S_j)S_j \mathbf{1}_{A_j}).$$

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 S_j est supérieur à λ sur A_j. Vérifions que le dernier terme est nul.

$$S_N - S_j = X_{j+1} + ... + X_N, \quad E(S_N - S_j) = 0.$$

La variable $S_j \mathbf{1}_{A_j}$ ne dépend que de X_1,..., X_j, elle est donc indépendante de $S_N - S_j$, ce qui implique

$$E((S_N - S_j)S_j \mathbf{1}_{A_j}) = E(S_N - S_j)E(S_j \mathbf{1}_{A_j}) = 0.$$

Au final,

$$E(S_N^2 \mathbf{1}_{A_j}) \geq \lambda^2 E(\mathbf{1}_{A_j}) = \lambda^2 P(A_j).$$

On conclut en faisant la somme pour j allant de 1 à n.

$$E(S_N^2) \geq E(S_N^2 \sum \mathbf{1}_{A_j}) \geq \lambda^2 P(\coprod A_j) = \lambda^2 P(\max |S_i| \geq \lambda).$$

Preuve de la convergence presque sûre
Considérons la variable aléatoire $R_N = \sup\{|S_n - S_N| \mid n \geq N\}$. D'après l'inégalité maximale appliquée à la suite $(X_{n-N})_{n \geq N}$,

$$P\Big(\max_{N \leq i \leq N'} |S_i - S_N| \geq \varepsilon\Big) \leq \frac{1}{\varepsilon^2} \sum_{i=N+1}^{N'} V(X_i) \leq \frac{1}{\varepsilon^2} \sum_{i=N+1}^{\infty} V(X_i)$$

et en passant à la limite quand N' tend vers l'infini,

$$P(R_N \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{i=N+1}^{\infty} V(X_i).$$

Par hypothèse, la série $\sum V(X_i)$ converge, son reste tend vers 0. La suite R_N converge en probabilité vers 0. Elle admet donc une sous-suite qui converge vers 0 presque sûrement et comme elle est décroissante, elle converge vers 0 presque sûrement. Au final, pour presque tout $\omega \in \Omega$, pour tout $\varepsilon > 0$, il existe $N \in \mathbf{N}$ tel que pour tout $m, n \geq N$,

$$|S_m(\omega) - S_n(\omega)| \leq |S_m(\omega) - S_N(\omega)| + |S_N(\omega) - S_n(\omega)| \leq 2R_N(\omega) \leq 2\varepsilon.$$

La suite $S_n(\omega)$ est de Cauchy et converge. La proposition est démontrée.

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, par exemple en les tirant à pile ou face ?

Proposition 21 *Soit $(\varepsilon_k)_{k\in\mathbf{N}^*}$ 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.

Ce résultat se déduit de la proposition 8.2. Il suffit de remarquer d'abord que $E(\frac{\varepsilon_k}{k}) = 0$ puis que

$$\sum_{k=1}^{\infty} V\Big(\frac{\varepsilon_k}{k}\Big) = \sum_{k=1}^{\infty} \frac{1}{k^2}\ V(\varepsilon_k) = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

Le critère de convergence est bien satisfait.

Comme exemple d'une telle suite de variables aléatoires, on peut prendre

$$\Omega = \{-1, 1\}^{\otimes \mathbf{N}}, \quad P = \Big(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1\Big)^{\otimes \mathbf{N}}, \quad \varepsilon_k\Big((x_n)_{n\in\mathbf{N}}\Big) = x_k.$$

On a alors

$$P\Big(\Big\{(x_k)_{k\in\mathbf{N}} \in \Omega \ \Big|\ \sum_{k\geq 1} \frac{x_k}{k} \text{ converge } \Big\}\Big) = 1.$$

On se pose maintenant la question générale de la convergence d'une série $\sum X_k$ lorsque les X_k sont indépendantes entre elles. Le théorème suivant ramène ce problème à la convergence de trois séries réelles, il est dû à Andreï Kolmogorov (1903-1987).

Théorème 15 (théorème des trois séries) *Soit* $(X_k)_{k\in\mathbf{N}}$ *une suite de variables aléatoires indépendantes entre elles. Posons* $Y_k = X_k\,\mathbf{1}_{(|X_k|\leq 1)}$. *Alors la série* $\sum X_k$ *converge presque sûrement si et seulement si les trois séries suivantes convergent :*

- $\sum P(|X_k| > 1)$,
- $\sum E(Y_k)$,
- $\sum V(Y_k)$.

Preuve
On se contente de démontrer que la convergence des trois séries implique la convergence presque sûre de $\sum X_k$. Comme $\sum P(|X_k| > 1)$ converge, nous pouvons appliquer le lemme de Borel-Cantelli : pour presque tout ω, il existe k_0 tel que pour tout $k \geq k_0$, $|X_k(\omega)| \leq 1$. On a alors $Y_k(\omega) = X_k(\omega)$. Les séries $\sum X_k$ et $\sum Y_k$ sont donc de même nature.

Posons $\tilde{Y}_k = Y_k - E(Y_k)$. Comme $\sum E(Y_k)$ converge, il suffit de démontrer la convergence presque sûre de $\sum \tilde{Y}_k$. Les $\tilde{Y}_k$ sont centrées et leur variance est égale à celle des Y_k :

$$\|\tilde{Y}_k\|_2^2 = V(\tilde{Y}_k) = V(Y_k).$$

On sait que la série $\sum V(\tilde{Y}_k) = \sum V(Y_k)$ converge. La proposition 8.2 s'applique, la série $\sum \tilde{Y}_k$ est convergente presque sûrement et le théorème est démontré.

8.3 Retour sur la loi des grands nombres

Pour terminer ce chapitre, 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 entre parenthèses 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 de la proposition 8.2.

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

$$\sum_{k=1}^{\infty} P(Y_k \neq X_k) = \sum_{k=1}^{\infty} P(|X_0| \geq k) = \int \sum_{k=1}^{\infty} \mathbf{1}_{\{k \leq x\}} \, dP_{|X_0|}(x) \leq \int x \, dP_{|X_0|}(x)$$

est une somme finie. 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.

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 16 (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, on obtient

$$P\Big(\bigcup_0^\infty A_n\Big) = \lim_n P(A_n).$$

Théorème 17 (lemme de Fatou) *Soit $f_n : \Omega \to \mathbf{R}_+$ une suite de fonctions mesurables positives. Alors*

$$\int_\Omega \liminf_{n\to\infty} f_n(\omega)\, dP(\omega) \le \liminf_{n\to\infty} \int_\Omega f_n(\omega)\, dP(\omega).$$

Théorème 18 (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 la suite f_n est dominée par une fonction $g : \Omega \to \mathbf{R}_+$* intégrable *:*

$$|f_n(\omega)| \le g(\omega) \quad \text{pour presque tout } \omega \in \Omega.$$

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 19 (interversion somme intégrale, cas positif) *Soit $f_n:\Omega\to\mathbf{R}_+$ une suite de fonctions mesurables positives. Alors*

$$\int_\Omega\sum_{n=1}^\infty f_n(\omega)\;dP(\omega)=\sum_{n=1}^\infty\int_\Omega f_n(\omega)\;dP(\omega).$$

Commentaire : la somme de la série peut être égale à $+\infty$.

Théorème 20 (interversion somme intégrale, cas intégrable) *Soit $f_n:\Omega\to\mathbf{R}$ une suite de fonctions mesurables. On suppose que*

$$\sum_{n=1}^\infty\int_\Omega|f_n(\omega)|\;dP(\omega)<+\infty.$$

Alors

$$\int_\Omega\sum_{n=1}^\infty f_n(\omega)\;dP(\omega)=\sum_{n=1}^\infty\int_\Omega f_n(\omega)\;dP(\omega).$$

Commentaire : la série qui apparaît dans le second terme est convergente.

A.2 Intégrales dépendant d'un paramètre

Théorème 21 (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 $t\in I$,

$$|f(t,\omega)|\le g(\omega)\quad\text{pour presque tout }\omega\in\Omega.$$

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 22 (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 $t \in I$,*

$$\left|\frac{\partial}{\partial t} f(t,\omega)\right| \leq g(\omega) \qquad \textit{pour presque tout } \omega \in \Omega.$$

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 23 (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*

$$\begin{aligned} \textstyle\int\int_{\Omega_1\times\Omega_2} f(\omega_1,\omega_2)\ dP_1 \otimes P_2(\omega_1,\omega_2) &= \textstyle\int_{\Omega_2}\Big(\int_{\Omega_1} f(\omega_1,\omega_2)\,dP_1(\omega_1)\Big)\,dP_2(\omega_2) \\ &= \textstyle\int_{\Omega_1}\Big(\int_{\Omega_2} f(\omega_1,\omega_2)\,dP_2(\omega_2)\Big)\,dP_1(\omega_1) \end{aligned}$$

Commentaire : les intégrales peuvent valoir $+\infty$.

Théorème 24 (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*

$$\int\int_{\Omega_1\times\Omega_2} |f(\omega_1,\omega_2)|\ dP_1 \otimes P_2(\omega_1,\omega_2) < +\infty.$$

Alors

$$\begin{aligned} \textstyle\int\int_{\Omega_1\times\Omega_2} f(\omega_1,\omega_2)\ dP_1 \otimes P_2(\omega_1,\omega_2) &= \textstyle\int_{\Omega_2}\Big(\int_{\Omega_1} f(\omega_1,\omega_2) dP_1(\omega_1)\Big)\,dP_2(\omega_2) \\ &= \textstyle\int_{\Omega_1}\Big(\int_{\Omega_2} f(\omega_1,\omega_2) dP_2(\omega_2)\Big)\,dP_1(\omega_1). \end{aligned}$$

Commentaire : la fonction f est dans $L^1(\Omega_1 \times \Omega_2)$.

Théorème 25 (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, $u = (r,\theta)$, $v = \varphi(u) = \varphi(r,\theta) = (r\cos(\theta), r\sin(\theta))$, $du = drd\theta$, $J\varphi(r,\theta) = r$.

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 26 (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 27 (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 28 (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 29 (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 30 (inégalité de Cauchy-Schwarz) *Soit* $f, g \in L^2(\Omega)$*. Alors* fg *est intégrable et*

$$\Big| \int_\Omega fg\, dP \Big| \leq ||f||_2\ ||g||_2.$$

Commentaire : on a égalité si et seulement si f et g sont proportionnelles.

Théorème 31 (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 32 (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 7 (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 33 (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 :

$$f(x) = f(0^+) + x f'(0^+) + x\, \varepsilon(x), \text{ avec } \lim_{x\to 0} \varepsilon(x) = 0.$$

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 $\displaystyle\lim_{A \to \infty} \int_{-\infty}^{0} \frac{2 \sin Ax}{x} \ f(x) \ \frac{dx}{2\pi} = \frac{1}{2} f(0^-)$, ce qui termine la preuve.

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

$$P_X = \sum_{k \in I} p_{x_k} \delta_{x_k}, \quad P_X(A) = \sum_{x_k \in A} p_{x_k}.$$

$$E(X) = \int_{\mathbf{R}} x \, dP_X(x) = \sum_{k \in I} x_k \, P(X = x_k).$$

$$V(X) = \sum_{k \in I} (x_k - E(X))^2 P(X = x_k) = \sum_{k \in I} x_k^2 \, P(X = x_k) - \Big(\sum_{k \in I} x_k \, P(X = x_k) \Big)^2.$$

$$E(g(X)) = \int_{\Omega} g(X) \, dP = \int_{\mathbf{R}} g(x) \, dP_X(x) = \sum_{k \in I} g(x_k) P(X = x_k).$$

$$F_X(x) = \sum_{x_k \leq x} P(X = x_k).$$

$$\varphi_X(t) = \sum_k e^{itx_k} P(X = x_k).$$

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

$$E(|XY|) = \int_{\Omega} |XY| \, dP \leq \sqrt{\int_{\Omega} X^2 \, dP} \, \sqrt{\int_{\Omega} Y^2 \, dP} = \|X\|_2 \, \|Y\|_2.$$

Inégalité de Markov

$$P(Y \geq \lambda) \leq \frac{E(Y)}{\lambda} \qquad \text{si } \lambda > 0, \ Y \geq 0.$$

Inégalité de Bienaymé-Tchebichev

$$P(|X - E(X)| \geq t) \leq \frac{V(X)}{t^2} \qquad \text{si } t > 0,\ E(X^2) < \infty.$$

Inégalité maximale de Kolmogorov

$$P\Big(\max_{0 \leq i \leq n} |S_i| \geq \lambda\Big) \leq \frac{E(S_n^2)}{\lambda^2}.$$

B.3 Couples de variables aléatoires

Soit X, 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

$$\text{Si les } X_i \text{ sont i.i.d. intégrables,} \quad \frac{S_n}{n} \xrightarrow[n\to\infty]{p.s.} E(X_1).$$

Théorème de la limite centrée

$$\text{Si les } X_i \text{ sont i.i.d. centrées telles que } 0 < \sigma(X_i) < \infty, \quad \frac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{loi} \mathcal{N}(0, \sigma^2).$$

Convergence de la somme dans le cas de variance bornée

Si les X_i sont indépendantes centrées et $\sum V(X_i) < \infty$, S_n 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,

$$S_n \text{ converge p.s. } \Leftrightarrow \sum P(|X_i| \ge 1),\ \sum E(Y_i),\ \sum V(Y_i) \text{ convergent.}$$

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 l'intégration

Marc Briane, Gilles Pagès
Théorie de l'intégration, cours et exercices
Vuibert. ISBN 978-2311402261

Référence en anglais pour l'intégration

Richard Mansfield Dudley
Real analysis and probability.
Cambridge University Press. ISBN 0-521-00754-2

Index