INTRODUCTION
à la
THÉORIE DES PROBABILITÉS

Omar Boukhadra

Dép. Mathématiques
Faculté des Sciences Exactes
Université de Constantine 1
boukhadra@umc.edu.dz

27 septembre 2022

Le présent document réunit des notes de cours dérivées de [4] et constitue une structure générale d'une introduction à la théorie des probabilités pour les L3-M1. Par construction, celle-ci fait appel à la théorie de la mesure qui lui confère le cadre mathématique adéquat pour élaborer son axiomatique. Ainsi, notre théorie doit impérativement débuter par la définition de l'espace des évènements aléatoires mesurables et la mesure appropriée en tant que fonction de ces derniers, dite ***mesure de probabilité*** *; d'où, on obtient un* ***espace probabilisé****. On définit après les* ***variables aléatoires*** *(v.a.) sur ledit espace comme applications mesurables en ce sens qu'elles y sont reconnaissables, ce qui constitue l'objet essentiel de notre étude. Ensuite, les principales notions abordées sont donc l'indépendance, fonction caractéristique, modes de convergence de v.a., et pour applications, on jette un premier coup d'œil sur les lois des grands nombres et le théorème central limite. Toutes les preuves des résultats donnés sont dans [4] ou dans une des références signalées.*

Contenu

1 Espace probabilisé & variables aléatoires

La théorie des probabilités fait appel à la théorie de la mesure pour trouver le cadre mathématique adéquat pour élaborer son axiomatique. Dans ce premier chapitre qui requière donc une connaissance préalable de cette dernière théorie, nous allons établir les notions basiques de la théorie des probabilités, à savoir l'espace probabilisé, les variables aléatoires et l'espérance mathématique. Ensuite, nous allons voir des résultats fondamentaux comme des inégalités et des théorèmes de convergence, ce qui nous sera utile pour la suite.

1.1 Espace probabilisé

Il s'agit en premier d'établir notre espace d'étude de l'aléatoire : on définit un **espace probabilisé** comme étant un *espace mesuré* $(\Omega, \mathcal{F}, P)$ avec la particularité que la masse totale est 1, i.e. $P(\Omega) = 1$. Ledit espace sert à modéliser un *phénomène aléatoire* ou une *expérience aléatoire*. En langage probabiliste, Ω est un ensemble non vide dit **fondamental** constitué d'éléments appelés **issues** (ou *réalisations* de l'expérience en étude). $\mathcal{F}$ est une $\sigma-$**algèbre** (une **tribu**) de Ω, i.e. une collection de parties de Ω, dits **évènements**, telle que

$$\begin{aligned}
&(i)\, \Omega \in \mathcal{F}, \\
&(ii)\, A \in \mathcal{F} \Longrightarrow A^c \in \mathcal{F}, \\
&(iii)\, (A_n) \subset \mathcal{F} \Longrightarrow \cup_n A_n \in \mathcal{F}.
\end{aligned}$$

Le couple $(\Omega, \mathcal{F})$ est donc par définition un espace mesurable qu'on appelle **espace probabilisable**. Et la mesure P, dite **mesure de probabilité** ou probabilité ou aussi **loi de probabilité**, laquelle, en tant que mesure, est une application réelle sur $\mathcal{F}$, vérifiant les trois axiomes de probabilité : une application non négative

$P(A) \geq 0$ avec une **masse totale normalisée** $P(\Omega) = 1$, et ledit **axiome des probabilités totales**, autrement dit la *σ-additivité* de P : pour toute suite $(A_n)_{n\geq 1}$ d'*évènements disjoints 2 à 2*, i.e.

$$\forall i \neq j : \; A_i \cap A_j = \emptyset$$

on a

$$P(\cup_n A_n) = \sum_n P(A_n) \tag{APT}$$

Nous conviendrons pour la suite que les ensembles manipulés sont dans $\mathcal{F}$, sauf mention contraire. Ω est dit l'*évènement certain*. Nous disons aussi qu'un évènement A différent de Ω se réalise **presque surement (p.s.)** si $P(A) = 1$, laquelle situation est possible comme nous pouvons le voir dans les exemples ci-dessous. À l'opposé, nous avons le vide $\emptyset$, dit *évènement impossible*, lequel vérifie en vertu des axiomes de probabilité

$$P(\emptyset) = 0 \tag{1.1}$$

En effet, il suffit de choisir une suite d'évènements (A_n) telle que $A_1 = \Omega$ et $A_n = \emptyset$ pour tout $n \geq 2$, et voir que (APT) donne

$$P(\Omega) = \sum_{n=1}^{\infty} P(A_n) = P(\Omega) + \sum_{n=2}^{\infty} P(\emptyset)$$

sachant que $P(\Omega) = 1$. Toutefois, notons que $P(A) = 0$ ne signifie pas “impossible” !

Le fait (1.1) combiné avec (APT) entraîne immédiatement l'*additivité* de P : si $A_1, \ldots, A_n$ sont des évènements disjoints 2 à 2, alors on a

$$P\left(\bigcup_{i=1}^{n} A_n\right) = \sum_{i=1}^{n} P(A_n) \tag{APT+}$$

Ladite additivité de la probabilité entraîne des conséquences directes qu'il est impératif de connaître. Remarquons en premier que si nous prenons $A_1 = A, A_2 = A^c$ et pour tout $n \geq 3, A_n = \emptyset$, la définition même d'une loi de probabilité impliquent clairement que

$$P(A^c) = 1 - P(A) \tag{1.2}$$

Nous avons mieux que (1.2), la propriété fondamentale ci-après et une généralisation dans l'exercice suivant.

Théorème 1.1 *Pour tous évènements A et B, on a*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1.3}$$

Exercice 1.2 (Formule du crible) Montrer que

$$P(\cup_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \tag{1.4}$$

En plus, nous donnons le théorème suivant qui réunit des conséquences de la définition d'une probabilité, qui seront utiles pour la suite.

Théorème 1.3 *Une mesure de probabilité est monotone, sous-additive et continue, i.e.*

$$(i)\, A \subset B \Longrightarrow P(A) \leq P(B).$$

$$(ii)\, A \subset \cup_{n=1}^{\infty} A_n \Longrightarrow P(A) \leq \sum_{n=1}^{\infty} P(A_n).$$

$$(iii)\, A_n \downarrow A \text{ ou } A_n \uparrow A \Longrightarrow \lim_{n} P(A_n) = P(A).$$

Nous avons deux exemples de loi de probabilité, les suivants, qui restent des *constantes* de la théorie des probabilités !

Exemple **1.4** (**Loi uniforme discrète**) Soit Ω un ensemble dénombrable fini et appelons $\mathscr{P}(\Omega)$ l'ensemble de toutes ses parties. Il est clair que ce dernier est une tribu, d'où $(\Omega, \mathscr{P}(\Omega))$ forme un espace probabilisable. Maintenant définissons sur $\mathscr{P}(\Omega)$ l'application P telle que pour tout $A \in \mathscr{P}(\Omega)$,

$$P(A) = \frac{|A|}{|\Omega|} \tag{1.5}$$

Il vient de la définition même que

$$P(\Omega) = 1, \quad P(\{\omega\}) = \frac{1}{|\Omega|}$$

Par conséquent, il est clair que P définit une mesure de probabilités sur $\mathscr{P}(\Omega)$. Nous voyons alors que toutes les issues possibles ont la même probabilité d'occurrence, autrement dit une probabilité uniforme sur tout Ω. Cette loi de probabilité est alors appelée à juste titre la **loi uniforme discrète**. Ainsi, $(\Omega, \mathscr{P}(\Omega), P)$ constitue bien un espace probabilisé. Cet espace sert à modéliser des expériences scientifiques diverses comme il peut aussi modéliser le jeu du lancer de dé à 6 faces équilibrées, c'est-à-dire qu'il n'y a techniquement aucune raison qu'une face apparaisse plus qu'une autre. Dans ce cas, nous pouvons poser $\Omega = \{1, 2, 3, 4, 5, 6\}$ et pour chaque $i \in \Omega$, nous avons $P(\{i\}) = 1/6$. ○

<u>Exemple</u> **1.5** (**Loi normale**) Soit $\Omega = \mathbb{R}$ et $\mathcal{F} = \mathcal{B}$, la tribu de Borel sur $\mathbb{R}$. Nous avons alors un espace probabilisable $(\mathbb{R}, \mathcal{B})$. Soit l'application $\mu : \mathcal{B} \longrightarrow \mathbb{R}$ définie par

$$\mu(A) = \int_A \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, \mathrm{d}x$$

Il est clair que $\mu(A) \geq 0$. De plus, on vérifie par un changement de variables en coordonnées polaires que nous avons bien

$$\mu(\mathbb{R}) = \int \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, \mathrm{d}x = 1$$

En effet, nous avons

$$\begin{aligned}\left(\int e^{-x^2/2} \, \mathrm{d}x\right)^2 &= \int\int e^{-(x^2+y^2)/2} \, \mathrm{d}x\mathrm{d}y \\ &= \int_0^{2\pi} \left(\int_0^{\infty} r e^{-r^2/2} \, \mathrm{d}r\right) \mathrm{d}\theta \\ &= 2\pi\end{aligned}$$

Ensuite, par la σ-additivité de la mesure de Lebesgue, l'axiome des probabilités totales est bien vérifié. Ainsi, $(\mathbb{R}, \mathcal{B}, \mu)$ est bien un espace probabilisé. La mesure μ est appelée **loi normale standard** ou **loi de Gauss**. Ce caractère de normalité s'expliquera plus tard ! ○

1.2 Variables aléatoires et distributions

En second lieu, la question porte sur les réalisations aléatoires d'une quantité en étude qui, si l'on veut la définir formellement, doit être une application mesurable en ce sens qu'elle est reconnaissable dans notre espace probabilisé, laquelle application est appelée *variable aléatoire*, ce qui constitue la deuxième notion fondamentale de notre théorie.

Soit $(\Omega, \mathcal{F})$ un espace probabilisable. Une application X définie sur $(\Omega, \mathcal{F})$ et à valeurs dans un espace mesurable $(S, \mathcal{S})$, est dite **variable aléatoire**, en abrégé **v.a.**, si elle est mesurable, i.e. pour tout $A \in \mathcal{S}$, on a

$$X^{-1}(A) = \{X \in A\} \in \mathcal{F}$$

Autrement dit, une v.a. est une application mesurable au sens mathématique du terme sauf qu'elle sert à modéliser une quantité *imprévisible*. Quand nous avons besoin de préciser la tribu $\mathcal{F}$, nous disons que X est $\mathcal{F}$-mesurable ou nous écrivons $X \in \mathcal{F}$.

Nous nous intéresserons particulièrement aux v.a. dans $\mathbb{R}$ dites v.a. réelles, ou dans $\mathbb{R}^n$, appelées **vecteurs aléatoires** ou dans $\mathbb{R}^{\mathbb{N}}$, que l'on appelle **processus stochastiques** à temps discret. Notons que l'appellation variable aléatoire est souvent utilisée pour les variables aléatoires réelles.

Supposons maintenant que l'espace $(\Omega, \mathcal{F})$ soit muni d'une mesure de probabilité P. Soit X une v.a. définie sur $(\Omega, \mathcal{F})$ et à valeurs dans $(S, \mathcal{S})$. Définissons sur $\mathcal{S}$ l'application

$$A \longmapsto \mu(A) = P(X \in A) = P \circ X^{-1}(A)$$

Il est facile de voir que μ satisfait les deux premiers axiomes d'une mesure de probabilité. Pour le dernier axiome (APT), nous observons que si (A_n) est une famille d'éléments totalement disjoints de $\mathcal{S}$, alors les $\{X \in A_n\}$ sont aussi disjoints 2 à 2 et nous avons

$$\mu(\cup_n A_n) = P(\cup_n \{X \in A_n\}) = \sum_n P(X \in A_n) = \sum_n \mu(A_n)$$

Ainsi, X induit une probabilité **image** sur S appelée **loi** ou **distribution** de X que l'on peut noter aussi par μ_X pour préciser la v.a. associée.

En pratique, l'espace $(\Omega, \mathcal{F}, P)$ n'est généralement pas explicite et pour les v.a., seule compte la mesure image. Si l'on se donne une loi μ sur S, on peut toujours l'écrire comme une loi image par une application mesurable (prendre par exemple l'identité sur $(S, \mathcal{S}, \mu)$). Par conséquent, toute mesure de probabilité peut être considérée comme la loi d'une v.a. Et par abus de notation, on peut directement écrire $(\Omega, \mathcal{F}, P)$ pour noter l'espace des états ou des valeurs d'une v.a. X. Nous pouvons écrire par exemple, sans parler de l'espace de départ $(\Omega, \mathcal{F}, P)$, soit X une v.a. telle que

$$P(X = 1) = 1 - P(X = 0) = p \in [0, 1]$$

Nous pouvons aussi voir cette même variable comme étant l'application identique de l'espace $(\{0, 1\}, \mathscr{P}(\{0, 1\}), P)$ où P donne la mesure p à l'évènement $\{1\}$ et $1 - p$ à l'évènement $\{0\}$. Comme nous pouvons aussi considérer cette v.a. comme l'application identique de $\mathbb{R}$ muni de la tribu de Borel et de loi donnée par

$$p\delta_1 + (1 - p)\delta_0$$

où δ_1 et δ_0 sont les masses de Dirac aux points 1 et 0. Ce bel exemple simple présente en fait la très célèbre **v.a. de Bernoulli** qui sert à modéliser une expérience aléatoire qui consiste à observer un phénomène en particulier qui apparaîtrait avec une probabilité p.

Si X suit une distribution μ, on écrit $X \sim \mu$. Et si X et Y ont la même loi μ, on dit que X et Y sont égales en distribution, ce que l'on note par $X \overset{d}{=} Y$. Si X et Y sont deux v.a. sur $(\Omega, \mathcal{F}, P)$, alors $\{X = Y\}$ est bien entendu mesurable, et si $P(X = Y) = 1$, on dit que X et Y sont égales **presque sûrement** (en abrégé **p.s.**) et on note $X \overset{\text{p.s.}}{=} Y$.

Il existe deux classes fondamentales de v.a. ou de manière équivalente de distributions de probabilité auxquelles nous nous intéresserons. En premier, nous avons les v.a. dites **variables discrètes** qui ne prennent (p.s.) qu'un nombre dénombrable de valeurs $(x_n)_{n\in I} \subset S, I \subset \mathbb{N}$. Sa loi de probabilité est donnée en général par la formule

$$\mu = \sum_{n\in I} p_n \delta_{x_n} \tag{1.6}$$

Observons alors que μ est en l'occurrence absolument continue par rapport à la mesure de comptage, i.e. si un ensemble A ne compte aucune des valeurs de la v.a. X alors $\mu(A) = 0$.

<u>Exemple</u> **1.6** (**Variable binomiale**) Une v.a. X qui prend ses valeurs dans $\{0, 1, \ldots, n\}$ avec la distribution

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}$$

est appelée une v.a. **binomiale** de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$, notée par $\mathcal{B}(n, p)$. Les quantités données ci-dessus définissent bien une distribution de probabilité en vertu de la formule du binôme :

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

○

<u>Exemple</u> **1.7** (**Variable de Poisson**) On dit que X suit une loi de **Poisson** de paramètre $\lambda > 0$, notée par $\mathcal{P}(\lambda)$, si la distribution de probabilité est donnée par

$$P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

Nous avons bien une distribution de probabilité car

$$\sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1$$

○

D'autre part, il y a les **v.a. continues** dont la distribution est donnée par la formule

$$\mu(A) = \int_A f(x)\,\mathrm{d}\lambda(x), \tag{1.7}$$

où f est une fonction mesurable réelle et non négative, appelée **densité de probabilité** (en abrégé **d.p.**). par rapport à la mesure de Lebesgue que l'on notera souvent comme l'intégrale de Riemann $\mathrm{d}x$. Observons que dans ce cas, μ est absolument continue par rapport à la mesure de Lebesgue, i.e. $\lambda(A) = 0 \Longrightarrow \mu(A) = 0$.

Notons toutefois qu'il est possible d'avoir des v.a. mixtes en ce sens qu'elles prennent des valeurs discrètes et continument distribuées sur des intervalles.

Nous avons ci-après des exemples de distributions continues mais nous en découvrirons de nouvelles au fur à mesure des besoins.

<u>Exemple</u> **1.8** (**Variable uniforme continue**) Une v.a. X est dite **uniforme** ou uniformément distribuée sur $[0, 1]$ si elle admet la densité de probabilité donnée par

$$f(x) = \mathbb{1}_{[0,1]}(x)$$

On la note par $U_{[0,1]} = U$. ○

<u>Exemple</u> **1.9** (**Variable normale**) On appelle v.a. **normale centrée réduite** ou standard celle qui suit la distribution donnée dans l'Exemple 1.5, i.e. sa densité est sous la forme

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

On représente une telle distribution par $\mathcal{N}(0, 1)$ ou simplement par $\mathcal{N}$. ○

<u>Exemple</u> **1.10** (**Variable exponentielle**) On dit que X est de distribution **exponentielle** de paramètre $\lambda > 0$ si elle admet la densité de probabilité

$$f(x) = \lambda e^{-\lambda x}\mathbb{1}_{\mathbb{R}_+}(x)$$

Ladite distribution est notée $\mathcal{E}(\lambda)$ et si $\lambda = 1$, on écrit simplement $\mathcal{E}$. ○

1.3 Mesurabilité des v.a.

Il est impératif que les quantités que nous manipulons soient bien mesurables. Dans cette section, nous allons voir quelques résultats qui permettent de prouver que nous avons bien affaire à des variables aléatoires. Nous considérons généralement que X est une v.a. dans un espace mesurable $(S, \mathcal{S})$ qui peut être en particulier $\mathbb{R}$ ou $\mathbb{R}^n$, lesquels sont munis de leurs tribus de Borel respectives, $\mathcal{B}$ et $\mathcal{B}^n$.

Pour prouver qu'une application (aléatoire) est mesurable, nous avons le résultat basique suivant.

Théorème 1.11 *Soit X une v.a. dans S muni d'une tribu $\mathcal{S}$. Si $\{X \in A\} \in \mathcal{F}$ pour tout $A \in \mathcal{A}$, et que $\mathcal{A}$ est une collection qui génère $\mathcal{S}$, alors X est mesurable.*

Nous pouvons voir que la famille $\{\{X \in A\} : A \in \mathcal{S}\}$ est une tribu, c'est en fait la plus petite tribu qui rend X mesurable. On l'appelle la **tribu générée par** X et on la note par $\sigma(X)$.

<u>Exemples</u> **1.12** Si $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B})$, alors des choix possibles pour $\mathcal{A}$ sont

$$\{(-\infty, x], x \in \mathbb{R}\} \text{ ou } \{(-\infty, q], q \in \mathbb{Q}\}$$

Et si $(S, \mathcal{S}) = (\mathbb{R}^n, \mathcal{B}^n)$, on peut choisir

$$\mathcal{A} = \{(a_1, b_1) \times \ldots \times (a_n, b_n) : -\infty < a_i < b_i < +\infty\}$$

○

Le résultat fondamental suivant donne une représentation générale des v.a. que nous serons amenés à manipuler.

Théorème 1.13 *Si X est une v.a. et que $f : S \to \mathbb{R}$ est une application mesurable, alors $f(X)$ est une v.a.*

Ces deux derniers théorèmes entraînent des conséquences importantes démontrant la mesurabilité de quantités fréquemment utilisées.

Corollaire 1.14 *Un vecteur $X = (X_1, \ldots, X_n)$ dans $\mathbb{R}^n$ constitue une v.a. ssi les coordonnées X_i sont des v.a. réelles.*

Corollaire 1.15 *Si $X_1, X_2, \ldots, X_n$ sont des v.a., alors $X_1 + \cdots + X_n$ est une v.a.*

En outre, nous avons les cas suivants qui seront présents dans beaucoup de situations à venir.

Théorème 1.16 *Si $X_1, X_2, \ldots$ sont des v.a., alors les applications suivantes le sont aussi :* $\inf X_n, \sup X_n, \liminf X_n$ *et* $\limsup X_n$.

Dans le même ordre d'idées, on a les deux cas particuliers suivants. Soit (A_n) une suite d'évènements de Ω. On définit l'évènement

$$\limsup A_n = \cap_n \cup_{k \geq n} A_k = \lim_{n \to \infty} \cup_{k \geq n} A_k$$

En fait, c'est l'ensemble des ω qui sont dans une infinité de A_n, autrement dit les A_n se produisent **infiniment souvent**, d'où la notation

$$\limsup A_n = \{A_n \text{ i.s.}\}$$

C'est un évènement qui existe (mesurable) car c'est la limite (intersection dénombrable) d'une suite décroissante d'évènements, soit $\cup_{k\geq n} A_k$. Notons que

$$\limsup \mathbb{1}_{A_n} = \mathbb{1}_{\{\limsup A_n\}}$$

En parallèle, on a aussi l'évènement

$$\liminf A_n = \cup_n \cap_{k\geq n} A_k = \lim_{n\to\infty} \cap_{k\geq n} A_k$$

qui est l'évènement des ω qui sont dans tous les A_n sauf, peut-être, un nombre fini des A_n. Cet évènement existe, étant une limite d'évènements croissants, soit $\cap_{k\geq n} A_k$. Et on a

$$\liminf \mathbb{1}_{A_n} = \mathbb{1}_{\{\liminf A_n\}}$$

Enfin, on dit qu'**une suite d'évènements** (A_n) **converge vers** A ssi

$$\liminf A_n = A = \limsup A_n$$

Et en vertu de la continuité de la probabilité (cf. Théorème 1.3), nous avons dans ce cas que

$$\lim P(A_n) = P(\liminf A_n) = P(\limsup A_n) = P(A)$$

1.4 Fonction de répartition

Soit X une v.a. réelle et considérons la fonction suivante

$$F(x) = P(X \leq x) \tag{1.8}$$

C'est là une fonction de distribution de X que l'on appelle **fonction de répartition** (en abrégé **f.r.**). Pour spécifier la v.a. X, nous utiliserons aussi dans la suite la notation F_X pour désigner une telle fonction.

<u>Exemple</u> **1.17** (**Masse de Dirac**) La f.r. de la masse de Dirac en a, δ_a, est donnée par

$$F(x) = \mathbb{1}_{[a,\infty)}(x)$$

○

<u>Exemple</u> **1.18** (**Loi géométrique**) On définit la **loi géométrique** de paramètre $p = 1 - q \in [0, 1]$, notée $\mathcal{G}(p)$, par

$$\mu = \sum_{n \in \mathbb{N}^*} q^{n-1} p \, \delta_n$$

La f.r. associée est

$$F(x) = \sum_{n \in \mathbb{N}^*} q^{n-1} p \, \mathbb{1}_{[n,+\infty)}(x) = \sum_{n=1}^{\lfloor x \rfloor} = q^{n-1} p = 1 - q^{\lfloor x \rfloor}$$

○

<u>Exemple</u> **1.19** (**Loi uniforme sur** $[0, 1]$) La f.r. de la variable uniforme sur $[0, 1]$ est donnée par

$$F(x) = \int_{\infty}^{x} \mathbb{1}_{[0,1]}(u) \, \mathrm{d}u = x 1_{[0,1]}(x) + 1_{(1,\infty)}(x)$$

○

<u>Exemple</u> **1.20** (**Distribution exponentielle**) La f.r. de loi exponentielle de paramètre α est

$$F(x) = \mathbb{1}_{\mathbb{R}_+}(x) \int_0^x \alpha e^{-\alpha u} \, \mathrm{d}u = (1 - e^{-\alpha x}) \, \mathbb{1}_{\mathbb{R}_+}(x)$$

○

<u>Exemple</u> **1.21** (**f.r. normale**) Il n'existe pas d'expression explicite pour la f.r. de la loi normale standard $\mathcal{N}(0, 1)$ qui est

$$F(x) = \int_{-\infty}^{x} \frac{e^{-u^2/2}}{\sqrt{2\pi}} \, \mathrm{d}u$$

Cependant, sa valeur peut être approchée comme dans le résultat suivant qui peut être utile pour de grandes valeurs de x.

Théorème 1.22 *Si $X \sim \mathcal{N}$, alors pour $x > 0$, on a*

$$(1 - x^{-2}) \frac{e^{-x^2/2}}{x\sqrt{2\pi}} \leq P(X > x) \leq \frac{e^{-x^2/2}}{x\sqrt{2\pi}} \tag{1.9}$$

○

Propriétés

L'intérêt premier de la définition de la f.r. est qu'en fait, celle-ci est suffisante pour connaître la loi d'une v.a. :

Proposition 1.23 *La f.r. caractérise la loi d'une v.a., i.e.*

$$F_X = F_Y \iff \mu_X = \mu_Y \tag{1.10}$$

Intéressons-nous maintenant aux caractéristiques de la f.r.

Théorème 1.24 *Toute f.r. possède les propriétés suivantes :*

(i) *F est non décroissante.*

(ii) $\lim_{x\to\infty} F(x) = 1, \quad \lim_{x\to-\infty} F(x) = 0.$

(iii) *F est continue à droite (c-à-d), i.e.* $\lim_{y\downarrow x} F(y) = F(x)$.

La réciproque de ce dernier théorème est vraie :

Théorème 1.25 *Si F est une fonction qui vérifie (i), (ii) et (iii) du Théorème 1.24, alors c'est une f.r. d'une certaine v.a.*

Cela est vrai grâce au résultat suivant (voir par exemple [7]) :

Théorème 1.26 (Mesure de Stieltjes) *Soit F est une fonction non décroissante et continue à droite. Alors, il existe une mesure unique μ sur $(\mathbb{R}, \mathcal{B})$ telle que*

$$\mu((a, b]) = F(b) - F(a) \tag{1.11}$$

Remarque 1.27 *Quand on choisit $F(x) = x$, nous obtenons la mesure de Lebesgue* $\mathrm{d}\lambda(x) = \mathrm{d}x$.

Au vu de la preuve du Théorème 1.25, nous en déduisons un résultat important d'une grande utilité dans la *simulation* de v.a. :

Corollaire 1.28 *Soit une X une v.a. de f.r. F et U la v.a. uniforme sur $[0, 1]$. Soit*

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}$$

Alors, on a

$$X \stackrel{d}{=} F^{-1}(U) \tag{1.12}$$

Remarque 1.29 *La v.a. uniforme sur* $[0,1]$ *peut actuellement être simulée par des algorithmes dits* ***générateurs de nombres aléatoires****. Ainsi, cette conséquence importante permet d'obtenir des valeurs suivant la distribution d'une v.a. donnée* X *en prenant les réalisations de la v.a.* $F^{-1}(U)$*.*

EXEMPLE **1.30** (**Simulation d'une** $\mathcal{E}$) La f.r. de la distribution exponentielle $\mathcal{E}$ est $F(x) = 1 - e^{-x}$ sur $\mathbb{R}_+$. D'où, nous obtenons que pour $u \in (0,1)$,

$$F^{-1}(u) = -\log(1-u)$$

Alors, si U est la v.a. uniforme sur $(0,1)$, nous avons que

$$-\log(1-U) \stackrel{d}{=} \mathcal{E}$$

Cependant, il est clair que $1 - U \stackrel{d}{=} U$. De ce fait, il vient que

$$-\log U \stackrel{d}{=} \mathcal{E}$$

○

Une autre propriété essentielle de la f.r. est la suivante.

Proposition 1.31 *Une f.r. admet au plus un nombre dénombrable de points de discontinuité.*

L'exemple suivant montre à quel point une distribution peut être étrange, difficile à concevoir.

EXEMPLE **1.32** (**Distribution sur** $\mathbb{Q}$) Soit $q_1, q_2, \ldots$, une énumération des rationnels. Soit $\sum_{n=1}^{\infty} p_n = 1$. Posons

$$F(x) = \sum_{n=1}^{\infty} p_n \mathbb{1}_{[q_n,\infty)}(x)$$

Nous avons là une f.r. avec un ensemble de discontinuités dense ! ○

Vecteurs aléatoires

Considérons maintenant les vecteurs aléatoires. On appelle **f.r. d'un vecteur aléatoire** $X \in \mathbb{R}^n$, la fonction

$$x = (x_1, \ldots, x_n) \longmapsto F_X(x) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) \tag{1.13}$$

Notons alors en premier que la propriété de la Proposition 1.23 se généralise sans difficulté aux vecteurs aléatoires.

Proposition 1.33 *Soit X et Y deux vecteurs aléatoires de lois respectives μ_X et μ_Y. Alors, on a*

$$F_X = F_Y \Longleftrightarrow \mu_X = \mu_Y \tag{1.14}$$

À partir d'un vecteur aléatoire $X \in \mathbb{R}^n$, nous obtenons le loi de la v.a. X_i, dite i-ème **loi marginale** de X, en intégrant par rapport au reste des coordonnées, i.e.

$$\begin{aligned} F_{X_i}(x_i) &= P(X_i \leq x_i) \\ &= P(X_1 \in \mathbb{R}, \ldots, X_{i-1} \in \mathbb{R}, X_i \leq x_i, X_{i+1} \in \mathbb{R}, \ldots, X_n \in \mathbb{R}) \\ &= \lim_{\forall j \neq i, x_j \to \infty} F_X(x) \end{aligned}$$

Ainsi, la loi de X détermine chacune des lois marginales. Cependant, dans l'exemple ci-dessous, nous allons montrer que la réciproque est fausse.

Dans le résultat suivant, nous avons les distributions marginales dans deux cas importants :

Proposition 1.34 *Soit $X = (X_1, \ldots, X_n) \in S_1 \times \cdots \times S_n$ un vecteur aléatoire discret suivant une distribution μ. Alors, les distributions marginales sont données par*

$$\mu_i(x_i) = \sum_{\forall j \neq i:\, x_j \in S_j} \mu(\{(x_1, \ldots, x_n)\}) \tag{1.15}$$

Si X est continue de loi commune de densité $f(x_1, \ldots, x_n)$, alors les densités marginales sont données par

$$f_{X_i}(x_i) = \int f(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1}\, \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n \tag{1.16}$$

<u>Exemple</u> **1.35** Soit $X = (X_1, X_2)$ un vecteur aléatoire discret et concentré sur les points $(-1, 0), (1, 0), (0, -1), (0, 1)$ avec les mêmes probabilités, i.e. la loi de X est donnée par

$$P_X = \frac{1}{4}\delta_{(-1,0)} + \frac{1}{4}\delta_{(1,0)} + \frac{1}{4}\delta_{(0,-1)} + \frac{1}{4}\delta_{(0,1)}$$

Les lois marginales P_1, P_2 de X_1 et X_2 sont alors égales et sont données par

$$P_i = \frac{1}{4}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_1$$

On peut produire un autre vecteur $Y = (Y_1, Y_2)$, ayant les mêmes lois marginales. En effet, la distribution représentée sur le Tableau 1 donne les mêmes lois marginales. ○

TABLE 1 –

	-1	0	1
-1	$1/16$	$1/8$	$1/16$
0	$1/8$	$1/4$	$1/8$
1	$1/16$	$1/8$	$1/16$

EXEMPLE **1.36** (**Couple normal standard**) Soit (X, Y) un vecteur aléatoire qui admet la densité suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$.

$$f(x, y) = (2\pi)^{-1}\, e^{-(x^2+y^2)/2}$$

Comme c'est une fonction continue, alors sa f.r. est donnée par l'intégrale de Riemann double

$$F(x, y) = \int^x \int^y (2\pi)^{-1}\, e^{-(u^2+v^2)/2}\, \mathrm{d}u\mathrm{d}v$$

Les f.r. marginales sont alors par continuité de la probabilité et le théorème de Fubini :

$$\begin{aligned} F_X(x) &= \lim_{y\to\infty} \int^x \int^y (2\pi)^{-1}\, e^{-(u^2+v^2)/2}\, \mathrm{d}v\mathrm{d}u \\ &= \int^x \frac{e^{-u^2/2}}{\sqrt{2\pi}} \left(\lim_{y\to\infty} \int^y \frac{e^{-v^2/2}}{\sqrt{2\pi}}\, \mathrm{d}v \right) \mathrm{d}u \\ &= \int^x \frac{e^{-u^2/2}}{\sqrt{2\pi}}\, \mathrm{d}u \end{aligned}$$

idem pour Y,

$$F_Y(y) = \int^y \frac{e^{-v^2/2}}{\sqrt{2\pi}}\, \mathrm{d}v$$

Ainsi, nous remarquons que les lois marginales ne sont rien d'autre que la loi normale standard. En fait, la loi du couple (X, Y) est la *loi normale standard* sur $\mathbb{R}^2$. ○

Nous terminons cette section en donnant la généralisation du Théorème 1.25 à des dimension supérieures. Soit le rectangle fini $A = \prod_{i=1}^n (a_i, b_i]$ et appelons

$S = \prod_{i=1}^{n}\{a_i, b_i\}$, l'ensemble de ses sommets. Si $s \in S$, définissons

$$\operatorname{sgn}(s) = (-1)^{\#\text{ de } a_i \text{ dans } s}$$
$$\Delta F(A) = \sum_{s\in R} \operatorname{sgn}(s) F(s)$$

Alors, nous avons en premier la généralisation du Théorème 1.26 suivante. Nous rappelons que dans l'ordre de $\mathbb{R}^n$, on écrit

$$x \leq y \Leftrightarrow \forall i : x_i \leq y_i$$

Théorème 1.37 *Soit une fonction $F : \mathbb{R}^n \to \mathbb{R}_+$. Supposons que F vérifie les conditions suivantes :*

(i) F est non décroissante, i.e. $x \leq y \Longrightarrow F(x) \leq F(y)$.

(ii) F est continues à droite, i.e. $\lim_{y\downarrow x} F(y) = F(x)$. $(y \downarrow x \Leftrightarrow \forall i : y_i \downarrow x_i.)$

(iii) $\Delta F(A) \geq 0$ pour tout rectangle fini A.

et qu'en plus, on a

$$\lim_{\min x_i \to \infty} F(x) = 1, \qquad \lim_{\max x_i \to -\infty} F(x) = 0$$

Alors, il existe une unique mesure de probabilité μ sur $(\mathbb{R}^d, \mathcal{B}^d)$ telle que pour tout rectangle fini A, $\mu(A) = \Delta F(A)$.

Nous arrivons ainsi à la généralisation du Théorème 1.25.

Théorème 1.38 *Si F est une fonction qui satisfait les conditions $(i), (ii)$ et (iii) du Théorème 1.37 et qu'en plus, on a*

$$\lim_{\min x_i \to \infty} F(x) = 1, \qquad \lim_{\max x_i \to -\infty} F(x) = 0$$

alors, il existe une unique mesure de probabilité μ telle que $\mu(A) = \Delta F(A)$.

1.5 Espérance mathématique

Maintenant, il est question de connaître la moyenne des valeurs possibles de ladite variable aléatoire par intégration suivant une mesure de probabilité, laquelle quantité est appelée *espérance mathématique*, la troisième notion fondamentale qui est une caractérisation essentielle de celle-là et entraînent quelques propriétés principales, notamment des inégalités et des convergences.

Soit X une v.a. non négative sur l'espace $(\Omega, \mathcal{F}, P)$. On définit son **espérance mathématique** (ou sa **moyenne**) par

$$E(X) = \int X \mathrm{d}P \tag{1.17}$$

Pour des v.a. réelles quelconques, l'espérance est donnée par

$$E(X) = \int X \mathrm{d}P(\omega) = E(X^+) - E(X^-)$$

où $X^+ = \max\{X, 0\}$ et $X^- = \max\{-X, 0\}$ sont respectivement les parties positive et négative de X. L'intégrale $E(X)$ existe si $E(X^+) < \infty$ ou $E(X^-) < \infty$. D'où, l'espérance peut être infinie. X est dite **intégrable** si

$$E(|X|) = E(X^+) + E(X^-) < \infty$$

De manière équivalente, on dit que la distribution μ de X est intégrable ssi

$$\int |x| \, \mathrm{d}\mu(x) < \infty$$

On définit l'intégrale de X sur A par

$$E(X; A) = \int_A X \mathrm{d}P = \int X \mathbb{1}_A \, \mathrm{d}P$$

L'espérance mathématique hérite par conséquent toutes les propriétés de l'intégrale de Lebesgue dont nous résumons les plus essentielles ci-après.

Théorème 1.39 (Propriétés essentielles) *Supposons que $X \geq 0$ et $Y \geq 0$ ou que $E(|X|) < \infty$ et $E(|Y|) < \infty$. Alors, on a*

$$\begin{aligned} &(i)\, E(X+Y) = E(X) + E(Y). \\ &(ii)\, Y \geq X \ p.s. \Longrightarrow E(Y) \geq E(X) \\ &(iii)\, E(|X|) < \infty \Longrightarrow |X| < \infty \ p.s. \end{aligned}$$

La définition de l'espérance mathématique en tant qu'intégrale de Lebesgue conduit à établir les quantités suivantes. Pour $p \geq 1$, posons

$$\|X\|_p = \left(\int |X|^p \, \mathrm{d}P \right)^{1/p}$$

Si $p = 1$, on écrit $\|X\|_1 = \|X\|$. En plus, on définit

$$\|X\|_\infty = \inf\{M : P(|X| > M) = 0\}$$

Exercice 1.40 Montrer que

$$\|X\|_\infty = \lim_{p\to\infty} \|X\|_p$$

Soit

$$\mathcal{L}^p = \mathcal{L}^p(\Omega, \mathcal{F}, P) = \{X : \|X\|_p < \infty\}$$

Les éléments de $\mathcal{L}^p$ sont les v.a. de **puissance p-ème intégrable**. Les éléments de $\mathcal{L}^\infty$ sont les v.a. dites **essentiellement bornées.**

Dans $\mathcal{L}^p$, on peut voir facilement que l'égalité $X \overset{\text{p.s.}}{=} Y$ p.s. constitue une relation d'équivalence. Ainsi, on regroupe les v.a. de $\mathcal{L}^p$ en classes d'équivalence, ce qui donne l'espace L^p, i.e.

$$L^p = \mathcal{L}^p / \overset{\text{p.s.}}{=}$$

On montre alors que l'application $X \longmapsto \|X\|_p$ constitue une norme pour laquelle L^p est un espace de Banach.

L'intégration dans l'espace $(\Omega, \mathcal{F}, P)$ est belle en théorie mais nous devons passer à l'espace des états pour pouvoir faire des calculs. Dans la plupart des cas, nous appliquerons le résultat suivant, souvent appelé **théorème de transport**.

Théorème 1.41 *Soit X une v.a. de $(\Omega, \mathcal{F}, P)$ dans l'espace $(S, \mathcal{S})$ de distribution μ, i.e. $\mu(A) = P(X \in A)$. Si f est une fonction mesurable de $(S, \mathcal{S})$ dans $(\mathbb{R}, \mathcal{B})$ telle que $f \geq 0$ ou $E(|f(X)|) < \infty$, alors on a la **formule de changement de variable**,*

$$E(f(X)) = \int f(x)\mathrm{d}\mu(x) \tag{1.18}$$

Remarque 1.42 *Notons que si μ admet une densité ϕ par rapport à la mesure de Lebesgue i.e. $\mathrm{d}\mu/\mathrm{d}x = \phi(x)$, alors on montre par la méthode standard (voir la preuve ci-après) que la formule (1.18) devient*

$$E(g(X)) = \int f(x)\phi(x)\mathrm{d}x \tag{1.19}$$

Le Théorème 1.41 nous permet donc d'effectuer l'intégration sur la droite réel quand $S = \mathbb{R}$. Dans le cas particulier où $f(x) = x^n, n \in \mathbb{N}^*$, nous définissons le **moment d'ordre** n de X ou de sa distribution par $E(X^n)$. Remarquons que le moment d'ordre 1 n'est autre que la moyenne de X.

La notion de *moment* conduit à s'intéresser à des quantités qui informent sur la distribution des v.a. : si $E(X^2) < \infty$, on définit la **variance de** X par

$$V(X) = E\big((X - E(X))^2\big)$$

Nous voyons alors que cette quantité mesure les écarts entre les valeurs aléatoires de X et de sa moyenne sauf que pour revenir à l'*unité* de X, il faudrait prendre la racine de la dite quantité, cela donne alors ce que l'on appelle l'**écart-type** :

$$\sigma(X) = \sqrt{V(X)} \tag{1.20}$$

Ce sont des notions basiques en statistiques mathématiques.

Cependant, pour calculer la variance, il peut être plus facile d'utiliser la formule suivante qui découle des propriétés essentielles de l'espérance mathématique (cf. Théorème 1.39).

$$V(X) = E\big(X^2 - 2XE(X) + E(X)^2\big) = E(X^2) - E(X)^2 \tag{1.21}$$

Ce qui entraîne que

$$V(X) \leq E(X^2)$$

Pour calculer un moment d'une v.a. donnée ou une de ses valeurs normées, nous avons une formule commode qui fait intervenir la f.r. :

Théorème 1.43 *Soit X une v.a. non négative de f.r. F. Alors pour tout $p > 1$,*

$$E(X^p) = p\int_0^\infty x^{p-1} P(X > x)\,\mathrm{d}x$$

De plus, on a

$$E(X) < \infty \Longleftrightarrow \sum_{n\in\mathbb{N}} P(X > n) < \infty$$

Exemples

<u>Exemple</u> **1.44** (**Moyenne arithmétique**) Soit X une v.a. qui prend les valeurs $x_1, \ldots, x_n$ avec la même probabilité $1/n$. La distribution de X est alors

$$\mu = \frac{1}{n}\sum_{i=1}^n \delta_i$$

et son espérance est donc

$$E(X) = \sum_{i=1}^n x_i\, P(X = i) = \frac{\sum_{i=1}^n x_i}{n}$$

Ce n'est rien d'autre que la *moyenne arithmétique* des quantités $x_1, \ldots, x_n$. ○

<u>Exemple</u> **1.45** (**Loi de Bernoulli**) Soit X une v.a. de Bernoulli de paramètre p, i.e. $\mu_X = p\delta_1 + (1-p)\delta_0$. Clairement,

$$E(X) = 1\,p + 0\,(1-p) = p, \quad E(X^2) = E(X) = p$$

et par conséquent,

$$V(X) = p - p^2 = p(1-p)$$

○

<u>Exemple</u> **1.46** (**Loi binomiale**) Si X est une loi binomiale de paramètres n et $p = 1-q$, alors son espérance est

$$E(X) \;=\; \sum_{k=0}^{n} k C_n^k p^k (1-p)^{n-k} = np$$

Quant à sa variance, nous avons d'abord que

$$E(X^2) \;=\; \sum_{k=0}^{n} k^2 C_n^k p^k (1-p)^{n-k} = n(n-1)p^2 + np$$

Ainsi, il vient que

$$V(X) = npq$$

○

<u>Exemple</u> **1.47** (**Distribution de Poisson**) Pour une v.a. X qui suit une loi de Poisson de paramètre $\lambda > 0$, on a

$$E(X) = e^{-\lambda} \sum_{n \geq 0} n \frac{\lambda^n}{n!} = \lambda e^{-\lambda} \sum_{n \geq 1} (n-1) \frac{\lambda^{n-1}}{(n-1)!} = \lambda$$

Et nous avons

$$E(X^2) \;=\; e^{-\lambda} \sum_{n \geq 0} n^2 \frac{\lambda^n}{n!} = \lambda^2 + \lambda$$

Nous en déduisons que

$$V(X) = \lambda$$

○

<u>Exemple</u> **1.48 (v.a. normale)** Si X est une variable normale standard (cf. Exemple 1.9), il vient par symétrie que

$$E(X) = \int x\,(2\pi)^{-1/2}\,e^{-x^2/2}\,\mathrm{d}x = 0$$

Et par intégration par parties, on a

$$V(X) = E(X^2) = \int x^2\,(2\pi)^{-1/2}\,e^{-x^2/2}\,\mathrm{d}x = 1$$

Si l'on fait une translation de valeur $\mu \in \mathbb{R}$ de la variable X, i.e. $X + \mu$, on obtient par changement de variable une v.a. de densité,

$$\frac{1}{\sqrt{2\pi}}\,e^{-(x-\mu)^2/2}$$

Et si l'on multiplie X par une valeur positive σ i.e. σX, cela par changement de variable une v.a. de densité

$$\frac{1}{\sigma\sqrt{2\pi}}\,e^{-x^2/2\sigma^2}$$

Il en découle alors que de manière générale, la loi de $\sigma X + \mu$ est de densité

$$\frac{1}{\sigma\sqrt{2\pi}}\,e^{-(x-\mu)^2/(2\sigma^2)}$$

avec

$$E(\sigma X + \mu) = \mu, \quad V(\sigma X + \mu) = \sigma^2 V(X) = \sigma^2$$

$\sigma X + \mu$ est appelée v.a. normale de moyenne μ et de variance σ^2. Elle est représentée par une *cloche gaussienne*, symétrique par rapport à μ, d'autant plus pointue que σ est petit. Et on note

$$\sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$$

○

<u>Exemple</u> **1.49 (v.a. exponentielle)** Si X suit une loi exponentielle de paramètre 1, alors on obtient par intégration par parties,

$$E(X^n) = \int_0^\infty x^n e^{-x}\,\mathrm{d}x = n!$$

Ainsi, la moyenne de X est 1 et sa variance et $E(X^2) - E(X)^2 = 1$. ○

Vecteurs

Quant aux vecteurs aléatoires dans $\mathbb{R}^n$, on définit leur **espérance mathématique** par l'extension naturelle suivante

$$E(X) = (E(X_1), \dots, E(X_n))$$

On dit que X est de puissance p-ème intégrable ($p \geq 1$) si chacune de ses composantes l'est, i.e. $X_i \in L^p$. En particulier, si X est de carré intégrable, on généralise la notion de variance par celle de **covariances** qui sont données par :

$$C_{ij} = \mathrm{Cov}(X_i, X_j) = E\big((X_i - E(X_i))(X_j - E(X_j))\big)$$

Observons que alors que $C_{ii} = V(X_i)$. La matrice $(C_{ij})_{1\leq i,j\leq n}$ ainsi constituée est appelée la **matrice de covariance** de X. Elle est clairement symétrique et de plus, semi-définie positive : pour tout $a = (a_1, \dots, a_n)$,

$${}^t aCa = \sum_{1\leq i,j\leq n} a_i a_j C_{ij} = E\Big(\big(\sum_{i=1}^{n} a_i(X_i - E(X_i))\big)^2\Big) \geq 0$$

<u>Exemple</u> **1.50** (**Vecteur gaussien**) Si X un vecteur aléatoire dans $\mathbb{R}^n$ qui admet par rapport à la mesure de Lebesgue la densité

$$f_X(x_1, \dots, x_n) = \frac{e^{-(x_1^2+\cdots+x_n^2)/2}}{(\sqrt{2\pi})^n} = \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{1}{2}\,{}^t x I_n x\right)$$

où, dans l'écrire matricielle, I_n est la matrice identité d'ordre n, i.e. $(\delta_{ij})_{1\leq i,j\leq n}$. Alors, nous avons $E(X) = 0_{\mathbb{R}^n}$ et la matrice de covariance est I_n. Il est facile de voir que les lois marginales sont normales $\mathcal{N}(0,1)$. X est dit vecteur *normal* ou *gaussien standard* et on écrit $X \sim \mathcal{N}(0, I_n)$.

On généralise la distribution normale standard de la façon suivante : soit $m \in \mathbb{R}^n$ et Σ une matrice symétrique définie (strictement) positive, c'est-à-dire inversible. Considérons le vecteur

$$Y = m + \Sigma X$$

Alors, il est facile par changement de variables de voir que la d.p. de Y est donnée par

$$f_Y(y) = \frac{1}{(\sqrt{2\pi})^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}\,{}^t(y-m)\Sigma^{-1}(y-m)\right)$$

Ainsi, on a une distribution normale de moyenne m et de matrice de covariance Σ, notée $\mathcal{N}(m, \Sigma)$.

Notons que l'on peut se donner une matrice symétrique Σ semi-définie positive de *rang* inférieur à n, et là, on a ce qui est appelé une distribution normale *dégénérée*. (Voir par exemple [3].) ○

1.6 Inégalités

Nous donnons dans cette section quelques inégalités fondamentales qui nous seront utiles dans les démonstrations futures. Nous avons en premier

Théorème 1.51 (Inégalité de Jensen) *Soit X une v.a. et φ une fonction convexe, i.e., pour tous $\lambda \in [0,1], x, y \in \mathbb{R}$,*

$$\lambda\varphi(x) + (1-\lambda)\varphi(y) \geq \varphi(\lambda x + (1-\lambda)y) \tag{1.22}$$

Si X et $\varphi(X)$ sont intégrables, alors, on a

$$E(\varphi(X)) \geq \varphi(E(X)) \tag{IJ}$$

Nous avons des cas usuels de (IJ), à savoir que

Corollaire 1.52 *Quand les intégrales existent, et pour $p \geq 1$, on a*

$$E(|X|^p) \geq |E(X)|^p$$

En second lieu, nous avons l'incontournable inégalité suivante.

Théorème 1.53 (Inégalité de Cauchy-Schwarz) *Soit $X, Y \in L^2$. Alors, on a*

$$\|XY\| \leq \|X\|_2 \, \|Y\|_2 \tag{ICS}$$

L'inégalité de Cauchy-Schwarz se généralise pour donner la célèbre inégalité suivante.

Théorème 1.54 (Inégalité de Hölder) *Soit $X \in L^p$ et $Y \in L^q$. Si p et q sont conjugués, i.e. $1/p + 1/q = 1$, alors on a*

$$\|XY\| \leq \|X\|_p \|Y\|_q \tag{IH}$$

Une application de l'inégalité de Hölder (IH) donne une autre inégalité célèbre.

Théorème 1.55 (Inégalité de Minkowski) *Soit $p \in [1, \infty]$ et $X, Y \in L^p$. Alors, on a*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p \tag{IMK}$$

En outre, nous avons l'inégalité très utile suivante.

Théorème 1.56 (Inégalité de Markov) *Soit X une v.a. non négative. Alors, pour $\lambda > 0$, on a*

$$P(X > \lambda) \leq \lambda^{-1} \, E(X) \tag{IM}$$

L'inégalité de Markov (IM) entraîne une autre inégalité importante.

Théorème 1.57 (Inégalité de Tchebychev) *Soit X une v.a. de carré intégrable. Alors, on a pour tout $\lambda > 0$,*

$$P(|X - E(X)| > \lambda) \leq \frac{V(X)}{\lambda^2} \tag{IT}$$

<u>Exemple</u> **1.58** (**Théorème de Stone-Weierstrass**) Soit f une fonction continue sur $[0,1]$ et soit (X_n) une suite de v.a. telles que $X_n \sim \mathcal{B}(n,x)$. Posons $Y_n = X_n/n$. Alors, on a que

$$\lim_{n\to\infty} \sup_{x\in[0,1]} |f(x) - E(f(Y_n))| = 0 \tag{1.23}$$

Remarquons auparavant que $E(f(Y_n))$ est un polynôme en x, dit **polynôme de Bernstein** associé à f :

$$E(f(Y_n)) = \sum_{k=0}^{n} f(k/n) C_n^k x^k (1-x)^{n-k}$$

Ainsi, (1.23) montre en fait la densité des fonctions polynomiales dans l'espace des fonctions continues sur $[0,1]$, soit le célèbre théorème de Stone-Weierstrass.

Soit $\delta > 0$ et posons

$$\Delta(f,\delta) = \sup_{|x-y|<\delta} \{|f(x) - f(y)|\}$$

Ce module de continuité est bien finie puisque f est uniformément continue sur $[0,1]$, étant elle-même supposée continue sur $[0,1]$. Notons que $\inf_{\delta>0} \Delta(f,\delta) = 0$.

Observons alors que

$$E\big(\,|f(x) - f(Y_n)|\,\big) \leq \Delta(f,\delta) P(|Y_n - x| < \delta) + 2\|f\|_\infty P(|Y_n - x| \geq \delta)$$

ce qui, par (IT) est inférieur à

$$\Delta(f,\delta) P(|Y_n - x| < \delta) + 2\|f\|_\infty \frac{V(X_n)}{\delta^2 n^2} \quad \leq \quad \Delta(f,\delta) + 2\|f\|_\infty \frac{x(1-x)}{\delta^2 n}$$

Or, $x(1-x) \leq 1$ quand $x \in [0,1]$. D'où, il résulte finalement que

$$E\left(|f(x) - E(f(Y_n))|\right) \leq \Delta(f,\delta) + \frac{2\|f\|_\infty}{\delta^2 n}$$

Le résultat s'ensuit. ○

À l'opposé de l'inégalité de Markov, nous avons l'inégalité suivante.

Théorème 1.59 (Inégalité de Paley-Zygmund) *Soit X une v.a. non négative de carré intégrable. Alors, on a pour tout $\lambda \in (0,1)$,*

$$P(X \geq \lambda E(X)) \geq (1-\lambda)^2 \frac{E(X)^2}{E(X^2)} \tag{IPZ}$$

1.7 Théorèmes de convergence

Cette section réunit quelques résultats fondamentaux en théories des probabilités qui reformulent des théorèmes classiques en théorie de la mesure. La question est d'étudier la possibilité d'inter-changer la limite d'une suite de v.a. et l'espérance.

D'abord, soit (X_n) une suite de v.s. et posons

$$\Omega_0 := \{\lim X_n \, \text{existe}\} \tag{1.24}$$

Autrement dit,

$$\Omega_0 = \{\liminf X_n = \limsup X_n\}$$

De ce fait, et par le Théorème 1.16, on voit que Ω_0 est un ensemble mesurable. Si $P(\Omega_0) = 1$, on dit alors que X_n **converge presque sûrement** (en abrégé **p.s.**). Cette convergence qui est qualifiée de **convergence forte** n'est autre que la convergence *presque partout* en théorie de la mesure. Ainsi, par définition, une suite (X_n) de v.a. converge p.s. vers une v.a. X si

$$\forall \varepsilon > 0 : \quad P(\cup_{l\geq 1} \cap_{n \geq l} \{|X_n - X| < \varepsilon\}) = 1 \tag{1.25}$$

ce qui, en passant au complémentaire, équivaut aussi à

$$\forall \varepsilon > 0 : \quad P(\{|X_n - X| \geq \varepsilon \text{ i.s.}\}) = 0$$

Par continuité (cf. Théorème 1.3), ceci est encore équivalent à

$$\forall \varepsilon > 0 : \quad \lim_{l \to \infty} P(\sup_{n \geq l} |X_n - X| \geq \varepsilon) = 0 \tag{1.26}$$

On écrit alors que

$$X_n \xrightarrow[n\to\infty]{\text{p.s.}} X$$

Le premier résultat fondamental sur la convergence des moyennes de v.a. est le suivant.

Théorème 1.60 (Convergence bornée) *Soit (X_n) une suite de v.a. bornées convergente p.s. Alors, on a*

$$\lim E(X_n) = E(\lim X_n) \tag{TCB}$$

Il découle de (TCB) l'inégalité importante suivante.

Théorème 1.61 (Lemme de Fatou) *Pour une suite de v.a. non négatives (X_n), on a*

$$\liminf E(X_n) \geq E(\liminf X_n) \tag{LF}$$

Le résultat important suivant est une conséquence immédiate de (LF).

Théorème 1.62 (Convergence monotone) *Soit (X_n) une suite de v.a. non négatives et non décroissante p.s. Alors, on a*

$$\lim E(X_n) = E(\lim X_n) \tag{TCM}$$

Enfin, nous obtenons :

Théorème 1.63 (Convergence dominée) *Si (X_n) est une suite de v.a. convergente p.s., $|X_n| \leq Y$ pour tout n et que $E(Y) < \infty$, alors on a*

$$\lim_{n\to\infty} E(X_n) = E(\lim X_n) \tag{TCD}$$

2 Probabilité conditionnelle & Indépendance

L'objet d'étude du présent chapitre est la notion d'indépendance, autrement dit l'absence d'influence des évènements aléatoires les uns sur les autres. Pour établir la définition de celle-ci, nous introduisons d'abord la probabilité conditionnelle. Ensuite, on peut donner une définition formelle à l'indépendance des évènements, laquelle est étendue à des tribus et du coup, à des v.a. Nous verrons quelques propriétés fondamentales relatives à l'indépendance, notamment l'espérance d'un produit et d'une somme finis de v.a. indépendantes. Après, nous donnerons la preuve d'un résultat fondamental, dit extension de Kolmogorov, ce qui démontre l'existence des processus stochastiques, et en conséquence, d'une infinité d'évènements indépendants. Ce qui, par la suite, nous permettra de connaître un résultat essentiel traduisant un phénomène étrange de l'aléatoire, appelé loi du $0-1$ *de Kolmogorov, conséquence de la notion d'indépendance.*

2.1 Probabilité conditionnelle

Avant de quantifier l'indépendance par des probabilités, il est intuitivement plus facile de concevoir l'influence entre les évènements aléatoires en définissant des probabilités *relatives* ou *conditionnelles*.

Ici et dans la suite, nous travaillerons dans un espace probabilisé $(\Omega, \mathcal{F}, P)$ et les v.a. prennent généralement leurs valeurs dans un espace mesurable $(S, \mathcal{S})$, en particulier $(\mathbb{R}, \mathcal{B})$.

Commençons par considérer l'exemple suivant. Jetons un dé équilibré et supposons que nous sachions que le résultat obtenu soit impair. Quelle est alors la probabilité que la face obtenue soit 5 ? Comme toutes les issues sont supposées équiprobables, il est logiquement cohérent de penser que la probabilité d'obtenir la face 5 parmi 3 faces impaires est de 1/3. En même temps, nous avons que la probabilité d'obtenir la face impaire 5 est de 1/6 ; celle d'obtenir une face impaire est de $3/6 = 1/2$. Alors, le rapport de la probabilité d'obtenir la face impaire 5 à la probabilité d'obtenir une face impaire est

$$\frac{1/6}{3/6} = \frac{1}{3}$$

Formellement, soit A et B deux évènements tels que $P(A) \neq 0$. Pour mesurer l'influence entre les évènements, on définit la **probabilité conditionnelle** de B par rapport à A par la quantité *relative* :

$$P(B \mid A) \coloneqq \frac{P(B \cap A)}{P(A)} \tag{2.1}$$

Ce qu'on peut lire aussi *probabilité de B sachant A ou conditionné à A*. Autrement dit, c'est la probabilité de la réalisation de B dans A qui est considérée comme l'ensemble fondamental relatif. Si $A = \Omega$, on retrouve alors la probabilité de B non conditionnée.

L'intérêt premier de la quantité (2.1) est le suivant :

Proposition 2.1 *L'application $B \mapsto P(B \mid A)$ définit bien une mesure de probabilité sur $\mathcal{F}$.*

<u>Exemple</u> **2.2** Répétons deux fois une expérience de Bernoulli simple telle que

$$P(\{1\}) = 1/2 = P(\{0\})$$

Les issues possibles d'une telle expérience sont $\{(0,0),(0,1),(1,0),(1,1)\}$. Quelle est la probabilité que le phénomène en question se produise deux fois sachant qu'il s'est réalisé ?

Formellement, appelons A l'évènement que ledit phénomène se réalise et B est celui qu'il se produise deux fois. Alors, nous avons

$$A = \{(1,1),(1,0),(0,1)\}, \quad B = \{(1,1)\}$$

Ce qui, par équiprobabilité, donne d'abord

$$P(A) = \frac{3}{4}, \quad P(B) = \frac{1}{4}$$

Mais la probabilité recherchée est

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Maintenant, si C est l'évènement que le phénomène se produise à la première observation, i.e. $C = \{(1,1),(1,0)\}$, alors

$$P(C) = \frac{2}{4} = \frac{1}{2}$$

D'où,

$$P(B \mid C) = \frac{P(C \cap B)}{P(C)} = \frac{1/4}{1/2} = \frac{1}{2}$$

Notons ici l'importance de la formulation mathématique correcte de l'expérience pour arriver aux bonnes probabilités. ○

La formule (2.1) peut être présentée sous la forme

$$P(A \cap B) = P(A)P(B \mid A) \tag{2.2}$$

Ce que l'on appelle *formule des probabilités composées* ou *Théorème de Bayes*. L'utilité d'une telle formule vient du fait que dans bien des situations, il est plus facile de déterminer la probabilité conditionnelle $P(B \mid A)$ pour certains B et certains A, A étant une information disponible, ce qui permet par la suite de déterminer $P(A \cap B)$.

<u>Exemple</u> **2.3** (**Tirage sans remise**) Considérons une urne contenant n boules dont k sont bleues. L'expérience consiste à tirer une première boule de l'urne, ensuite, sans remettre la première boule dans l'urne, on tire une deuxième boule. Une formalisation possible de cette expérience est la suivante. On considère que Ω est l'ensemble fini des couples ordonnés (a, b) où a représente la première boule tirée et b la seconde. Par le principe fondamental de dénombrement, $|\Omega| = n(n-1)$. Si l'on suppose que les issues sont équiprobables, alors

$$P(\{(a,b)\}) = \frac{1}{n(n-1)}$$

Quelle est la probabilité de tirer deux boules bleus ? Appelons E cet évènement. Pour calculer directement la probabilité en question, i.e. $P(E)$, il suffit de connaître le nombre de cas favorables à sa réalisation. Or, le nombre de possibilités pour obtenir une boule bleue au premier tirage est de k, et il va rester $k-1$ chances pour le deuxième tirage. D'où, $|E| = k(k-1)$, et donc

$$P(E) = \frac{k(k-1)}{n(n-1)}$$

Cependant, si A est l'évènement que a soit bleue, et B est celui que b soit bleue. Alors, $E = A \cap B$ et la probabilité recherchée est

$$P(E) = P(A)P(B \mid A)$$

Mais, après le premier tirage il reste $n-1$ boules dont $k-1$ bleues, donc nous avons clairement

$$P(A) = \frac{k}{n}, \quad P(B \mid A) = \frac{k-1}{n-1}$$

Ce qui donne

$$P(E) = \frac{k}{n}\frac{k-1}{n-1} = \frac{k(k-1)}{n(n-1)}$$

○

La formule (2.2) se généralise de la manière suivante.

Exercice 2.4 Soit $A_1, \ldots, A_n$ des évènements aléatoires tels que

$$P(A_1 \cap \cdots \cap A_{n-1}) \neq 0$$

Poser $A_0 = \Omega$ et montrer que

$$P(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^{n} P(A_i \mid \cap_{j=0}^{i-1} A_j)$$

Une application importante de (2.2) consiste à calculer une probabilité d'un évènement donné sur une partition : on appelle **système constituant** toute famille d'évènements (A_n), finie ou dénombrable, deux à deux disjoints, tels que $\Omega = \cup A_n$.

Proposition 2.5 *Soit (A_n) un système constituant. Alors, pour tout évènement B, on a*

$$P(B) = \sum P(A_n)P(B \mid A_n) \tag{2.3}$$

<u>Exemple</u> **2.6 (Pas de fille !)** Considérons le problème suivant : un père n'a pas de fille ! Quelle est la probabilité pour que cela se produise ? Appelons E cet évènement et Ω l'ensemble des compositions possibles des enfants d'un père quelconque en supposant avec un peu de machisme que le nombre d'enfants peut être infini. Soit N la v.a. qui compte le nombre d'enfants ; si $N = n$, alors le nombre d'issues possibles est de 2^n. Posons $p_n = P(N = n)$, donc $\sum p_n = 1$.

Si nous supposons que les filles et les garçons naissent avec la même chance, alors par équiprobabilité des naissances, nous avons

$$P(E \mid N = n) = \frac{1}{2^n}$$

Et par la formule de Bayes,

$$P(E \cap \{N = n\}) = P(N = n)P(E \mid N = n) = \frac{p_n}{2^n}$$

Mais, $\Omega = \cup_{n=0}^{\infty}\{N = n\}$, les $\{N = n\}$ étant disjoints. Par conséquent, il vient par application de la Proposition 2.5 que

$$P(E) = \sum_{n=0}^{\infty} P(E \cap \{N = n\}) = \sum_{n=0}^{\infty} \frac{p_n}{2^n}$$

$\bigcirc$

Dans l'Exemple 2.6, on peut se poser la question de savoir quelle est la probabilité qu'un père ait 10 enfants sachant qu'il n'a pas de fille. Si l'on a 10 enfants, la probabilité que des garçons est de $1/2^{10}$. Il en résulte que la probabilité en question est

$$P(N = 10 \mid E) = \frac{P(N = 10)P(E \mid N = 10)}{\sum_{n=0}^{\infty} P(E \cap \{N = n\})} = \frac{p_{10}}{\sum_{n=0}^{\infty} p_n 2^{-n+10}}$$

Généralement, nous avons

Proposition 2.7 (Formule de Bayes) *Soit (A_n) un système constituant. Alors, pour tout évènement B de probabilité non nulle, pour tout k, on a*

$$P(A_k \mid B) = \frac{P(A_k)P(B \mid A_k)}{\sum P(A_n)P(B \mid A_n)} \tag{FB}$$

2.2 Définition de l'indépendance

Soit A et B deux évènements de probabilités non nulles. En toute logique, s'il n'y a aucune influence de la réalisation de A sur B, autrement dit B est indépendant de l'occurrence de A, on considère que

$$P(B \mid A) = P(B)$$

Ce qui, par application de (2.1), implique que

$$P(A \cap B) = P(A)P(B) \tag{2.4}$$

Inversement, si la réalisation de B n'affecte pas A, nous avons que

$$P(A \mid B) = P(A)$$

ce qui donne aussi la même formule (2.4). Ceci conduit alors à la définition de l'indépendance suivante : deux évènements A et B sont dits **indépendants** si

$$P(A \cap B) = P(A)P(B) \tag{2.5}$$

Cette relation est souvent notée par le signe de l'orthogonalité : $A \perp B$.

Remarquons déjà la propriété basique suivante :

Proposition 2.8 *Si A est indépendant de B alors il est indépendant de B^c.*

On étend la définition de l'indépendance entre deux évènements à une famille d'évènements $\{A_1, \ldots, A_n\}$ en disant qu'ils sont (mutuellement) indépendants ssi

$$\forall I \subset \{1, \ldots, n\} : P\Big(\bigcap_{i \in I} A_i\Big) = \prod_{i \in I} P(A_i)$$

Cependant, on dit que les A_i sont **indépendants deux à deux** ssi

$$\forall i \neq j : P(A_i \cap A_j) = P(A_i)P(A_j)$$

Cette propriété est moins forte que l'indépendance. En effet, il ne suffit pas d'avoir l'indépendance deux à deux des A_i pour avoir l'indépendance de tous les A_i comme nous pouvons le voir dans l'exemple ci-dessous.

Plus généralement, une collection de familles d'évènements, $\mathcal{A}_1, \ldots, \mathcal{A}_n \subset \mathcal{F}$, sont indépendantes si pour tout $I \subset \{1, \ldots, n\}$, et pour tous $A_i \in \mathcal{A}_i$, on a

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

Une séquence de tribus $\mathcal{F}_1, \ldots, \mathcal{F}_n$ sont dites indépendantes si pour tous $A_i \in \mathcal{F}_i, i = 1, \ldots, n$, on a

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n)$$

Observons ici qu'il est possible que $A_i = \Omega$.

L'indépendance des tribus conduit machinalement à la définition de l'indépendance des v.a. On dit que $X_1, \ldots, X_n$ sont des **v.a. indépendantes** si leurs tribus associées respectives, $\sigma(X_1), \ldots, \sigma(X_n)$, sont indépendantes. Ce qui est équivalent à ce que pour tous $A_i, i = 1, \ldots, n$, on ait

$$P(X_1 \in A_1, \cdots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n)$$

<u>EXEMPLE</u> **2.9** (**Tribus triviales**) Si A et B sont deux évènements indépendants, alors on a en vertu de la Proposition 2.8 l'indépendance des tribus triviales générées par A et B, respectivement, i.e.

$$\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}, \quad \sigma(\{B\}) = \{\emptyset, B, B^c, \Omega\}$$

○

<u>EXEMPLE</u> **2.10** (**v.a. indicatrices**) $\mathbb{1}_A$ et $\mathbb{1}_B$ sont des v.a. indépendantes ssi A et B sont indépendants. En effet, si l'on suppose que $\mathbb{1}_A$ et $\mathbb{1}_B$ soient indépendantes ou que A et B soient indépendants, on a

$$P(\mathbb{1}_A = 1, \mathbb{1}_B = 1) = P(A \cap B) = P(A)P(B) = P(\mathbb{1}_A = 1)P(\mathbb{1}_B = 1)$$

○

<u>EXEMPLE</u> **2.11** (**Indépendance 2 à 2**) Soit l'expérience aléatoire qui consiste à jeter deux dés, un blanc et un noir, que nous supposons équilibrés, i.e. toutes les faces ont la même chance d'apparition. L'ensemble des issues possibles $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ peut être munie de la probabilité uniforme, i.e. si $A \subset \Omega$,

$$P(A) = \frac{|A|}{|\Omega|}$$

Appelons X et Y les v.a. qui représentent respectivement les chiffres donnés par le dé blanc et le dé noir. Et considérons les évènements

$$\begin{aligned} A &= \{X \in \{1, 3, 5\}\} \\ B &= \{Y \in \{1, 3, 5\}\} \\ C &= \{X + Y \in \{3, 5, 7, 9, 11\}\} \end{aligned}$$

Il est facile de voir que A, B et C sont deux à deux indépendants. En effet, nous obtenons par un calcul trivial de dénombrement que

$$P(A) = P(B) = \frac{3}{6} = \frac{1}{2} = \frac{2 \times 9}{36} = P(C)$$

Nous pouvons voir alors que

$$P(A \cap B) = \frac{9}{36} = \frac{1}{4} = P(A)P(B)$$

et que

$$P(A \cap C) = P(B \cap C) = \frac{9}{36} = \frac{1}{4} = P(A)PC) = P(B)P(C).$$

Cependant, $A \cap B \cap C = \emptyset$, d'où

$$P(A \cap B \cap C) = 0 \neq \frac{1}{8} = P(A)P(B)P(C)$$

Ainsi, A, B et C ne sont pas indépendants. $\circ$

2.3 Critères d'indépendance

Nous allons connaître des critères basiques pour l'indépendance des v.a. En premier, nous avons une condition suffisante pour l'indépendance des tribus pour laquelle on rappelle deux définitions de la théorie de la mesure.

On dit qu'une collection $\mathcal{A}$ de parties d'un ensemble donné Ω non vide est un π-**système** si elle est stable par intersection finie, i.e.

$$\forall A, B \in \mathcal{A} : \; A \cap B \in \mathcal{A}$$

Et on dit qu'une collection $\mathcal{L}$ de parties de Ω est un λ-**système** si :

$$\begin{aligned}
&(i)\, \Omega \in \mathcal{L} \\
&(ii)\, A, B \in \mathcal{L} : \; A \subset B \Longrightarrow B \setminus A \in \mathcal{L} \\
&(iii)\, (A_n) \subset \mathcal{L} : \; A_n \uparrow A \Longrightarrow A \in \mathcal{L}
\end{aligned}$$

Le résultat en théorie des probabilités qui nous intéresse est alors

Théorème 2.12 *Soit $\mathcal{A}_1, \ldots, \mathcal{A}_{n-1}$ et $\mathcal{A}_n$ des π-systèmes indépendants. Alors, les tribus $\sigma(\mathcal{A}_1), \ldots,$ et $\sigma(\mathcal{A}_n)$ sont indépendantes.*

Remarque 2.13 *Clairement, une algèbre est un $\pi-$système, donc nous pouvons déduire de ce théorème que si nous avons une famille d'algèbres indépendantes, les tribus générées par celles-ci sont indépendantes.*

Le dernier théorème entraîne une conséquence importante qui prouve l'existence d'une séquence finie de v.a. indépendantes.

Théorème 2.14 *Si $X = (X_1, \ldots, X_n)$ est un vecteur aléatoire tel que pour tous $x_1, \ldots, x_n \in (-\infty, \infty]$,*

$$F_X(x_1, \ldots, x_1) \;=\; F_{X_1}(x_1) \cdots F_{X_n}(x_n) \tag{2.6}$$

alors $X_1, \ldots, X_n$ sont des v.a. indépendantes.

Remarque 2.15 *La formule* (2.6) *montre que*

$$\mu_X\Big(\prod_{i=1}^n(-\infty,x_i]\Big)=\prod_{i=1}^n\mu_{X_i}((-\infty,x_i])=\mu_1\otimes\cdots\otimes\mu_n\Big(\prod_{i=1}^n(-\infty,x_i]\Big).$$

où $\mu_1\otimes\cdots\otimes\mu_n$ *désigne la mesure produit des* μ_i*. De ce fait et par unicité de l'extension de Carathéodory, la distribution commune est la mesure de probabilité produit des* μ_i*. Ainsi, nous pouvons construire un nombre fini de v.a. indépendantes en prenant la mesure de probabilité produit de leurs distributions. Si, en plus, les* μ_i *sont égales, on parle alors de v.a. indépendantes et identiquement distribuées, ce qui est noté en abrégé par* ***i.i.d.***

Réciproquement, nous avons

Théorème 2.16 *Si* $X_1,\dots,X_n$ *sont des v.a. indépendantes de distributions respectives* $\mu_1,\dots,\mu_n$*, alors* $(X_1,\dots,X_n)$ *possède la mesure de probabilité* $\mu_1\otimes\cdots\otimes\mu_n$*.*

Le corollaire suivant est une conséquence directe du Théorème 2.16.

Corollaire 2.17 *Soit* $X_1,\dots,X_n$ *des v.a. qui prennent leurs valeurs dans des ensembles dénombrables respectifs* $S_1,\dots,S_n$*. Pour avoir l'indépendance des* X_i*, il faut et il suffit que pour tous* $x_i\in S_i$*, on ait*

$$P(X_1=x_1,\dots,X_n=x_n)=\prod_{i=1}^n P(X_i=x_i)$$

En parallèle, si $X_1,\dots,X_n$ *sont des v.a. de densités respectives* $f_1,\dots,f_n$*. Alors, ces v.a. sont indépendantes ssi la distribution (conjointe) du vecteur* $(X_1,\dots,X_n)$ *admet une densité* f *donnée par*

$$f(x_1,\dots,x_n)=f_1(x_1)\cdots f_n(x_n)$$

<u>Exemple</u> **2.18** (**Statistique d'ordre**) Soit $U_1,\dots,U_{2n+1}$ des v.a. i.i.d. de distribution uniforme sur $[0,1]$. Ordonnons les U_i de manière croissante et appelons V_{n+1} la valeur d'ordre $n+1$. Alors, V_{n+1} possède la densité

$$f_{n+1}(x)=(2n+1)C_{2n}^n x^n(1-x)^n\,\mathbb{1}_{[0,1]}(x) \tag{2.7}$$

En effet, observons en premier que la probabilité d'avoir $U_i\le x-\varepsilon$ pour $i=1,\dots,n$, $U_{n+1}\in(x-\varepsilon,x+\varepsilon]$, et $U_i>x+\varepsilon$ pour $i=n+1,\dots,2n+1$, est par indépendance égale à

$$2\varepsilon\,(x-\varepsilon)^n(1-(x+\varepsilon))^n$$

En dénombrant toutes les configurations possibles, $2n+1$ pour jouer le rôle de U_n et C_{2n}^n pour choisir les n premières v.a. (et du coup, les n dernières v.a.), nous obtenons une probabilité de

$$P_\varepsilon(x) = (2n+1)C_{2n}^n \, 2\varepsilon \, (x-\varepsilon)^n(1-(x+\varepsilon))^n$$

Il résulte alors que

$$f_{n+1}(x) = \lim_{\varepsilon \to 0} \frac{P_\varepsilon(x)}{2\varepsilon} = (2n+1)C_{2n}^n \, x^n(1-x)^n$$

○

Exercice 2.19 Soit le vecteur aléatoire $X = (X_1, \dots, X_n)$ possédant la densité f. Supposer que f s'écrive sous la forme

$$f(x) = f_1(x_1) \cdots f_n(x_n)$$

où f_i sont des fonctions non négatives. Montrer que les X_i sont indépendantes.

Le résultat suivant montre que l'indépendance des v.a. est conservée sous des transformations déterministes.

Théorème 2.20 *Si $X_1, X_2, \dots, X_n$ sont des v.a. indépendantes et que $f_1, \dots, f_n$ sont des fonctions réelles mesurables, alors $f_1(X_1), \dots, f_n(X_n)$ sont des v.a. indépendantes.*

2.4 Espérance du produit

Nous savons déjà que la loi de v.a. indépendantes est le produit de leurs lois (cf. Théorème 2.16). Ceci entraîne alors la propriété fondamentale suivante.

Théorème 2.21 *Si X et Y sont des v.a. indépendantes non négatives ou intégrables, alors*

$$E(XY) = E(X)E(Y) \tag{2.8}$$

Remarque 2.22 *La propriété (2.8) n'est pas suffisante à l'indépendance de X et Y comme l'atteste l'exemple ci-dessous. Cependant, si cette égalité est vraie, ceci peut traduire une dépendance faible entre X et Y, ce que l'on peut qualifier aussi d'une absence de corrélation ; un terme utilisé en statistique mathématique. On peut alors donner une mesure à cette corrélation à l'aide de la quantité suivante, dite*

coefficient de corrélation :

$$\rho = \rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)V(Y)}}$$

Observons alors que $|\rho| \in [0,1]$*. D'où on peut dire que la corrélation est forte quand* $|\rho| = 1$ *et nulle si* $\rho = 0$*.*

Par ailleurs, il est clair que cette même propriété peut être généralisée à plus de deux v.a. : si $X_1, \ldots, X_n$ *sont des v.a. indépendantes, non négatives ou intégrables, alors on a*

$$E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n) \tag{2.9}$$

<u>EXEMPLE</u> **2.23** Cet exemple montre que la réciproque du dernier théorème est fausse, autrement dit, il est possible d'avoir $E(XY) = E(X)E(Y)$ sans que X et Y soient des v.a. indépendantes. En effet, supposons que la distribution de (X,Y) est donnée dans le Tableau 2 où $a, b, c \geq 0$ et $2a + 2b + c = 1$.

$X\backslash Y$	1	0	-1
1	0	a	0
0	b	c	b
-1	0	a	0

TABLE 2 –

Observons alors que

$$E(XY) = 0 = E(X) = E(Y)$$

Néanmoins, X et Y ne sont par indépendantes puisque nous avons

$$P(X=1, Y=1) = 0 \neq ab = P(X=1)P(Y=1)$$

○

Au vu du Théorème 2.8 et de l'Exemple 2.23, on est amené à établir une définition moins forte que l'indépendance, laquelle toutefois traduit une *corrélation* faible entre les v.a. : on dit que les v.a., $X_1, \ldots, X_n$ sont **non corrélées** si

$$\forall i \neq j : \; E(X_i X_j) = E(X_i)E(X_j) \tag{2.10}$$

Sous l'hypothèse de la non corrélation d'une famille de v.a., il est utile de savoir que dans L^2, la variance de la somme de v.a. non corrélées est égale à la somme des variances ; a fortiori, le résultat suivant est vrai dans le cas de l'indépendance :

Proposition 2.24 *Si $X_1, \ldots, X_n$ sont des v.a. non corrélées choisies dans L^2, on a*

$$V(X_1 + \cdots + X_n) = V(X_1) + \cdots + V(X_n) \tag{2.11}$$

2.5 Produit de convolution

Passons maintenant à la somme de v.a. indépendantes. Il s'agit ici de la question suivante. Si X et Y sont deux v.a. indépendantes, quelles est la distribution de leur somme $X + Y$? Le résultat principal suivant donne une réponse générale à l'aide des fonctions de répartition.

Théorème 2.25 *Si X et Y sont deux v.a. indépendantes, alors on a*

$$F_{X+Y}(z) = \int F_X(z-y)\,\mathrm{d}F_Y(y) \tag{2.12}$$

Remarque 2.26 *L'intégrale (2.12) est appelée **produit de convolution** de F_X et F_Y que nous notons par $(F_X * F_Y)(z)$. Nous rappelons que la notation $\mathrm{d}F_Y(y)$ signifie que l'on intègre par rapport à la distribution associée à F_Y. Par symétrie, nous avons clairement que*

$$F_{X+Y}(z) = \int F_Y(z-x)\,\mathrm{d}F_X(x)$$

En outre, remarquons que la distribution de la somme $X + Y$ est le produit de convolution de μ_X et μ_Y, ce que l'on note par $\mu_X * \mu_Y$.

Exemple **2.27** (**Loi binomiale**) Une des applications importantes du Théorème 2.25 porte sur la définition de la distribution binomiale. Prenons d'abord des v.a. indépendantes telles que $X \sim B(n,p)$ et $Y \sim B(m,p)$ avec $q = 1 - q$. Alors, on a

$$X + Y \sim B(n+m, p)$$

En effet, soit $k \in \{0, 1, \ldots, n+m\} = (X+Y)(\Omega)$ et observons que

$$\begin{aligned} P(X+Y=k) &= \sum_{l=0}^{k} P(X=l)P(Y=k-l) \\ &= p^k q^{n+m-k} \sum_{l=0}^{k} C_n^l C_m^{k-l} \\ &= C_{n+m}^k p^k q^{n+m-k} \end{aligned}$$

où nous avons utilisé le fait suifant (laissé en exercice de dénombrement) :

$$\sum_{l=0}^{k} C_n^l C_m^{k-l} = C_{n+m}^k \tag{2.13}$$

De ce fait, si $\mathcal{B}_i(p)$ représentent des distributions de Bernoulli indépendantes, alors on a

$$\mathcal{B}(n,p) = \mathcal{B}_1(p) * \cdots * \mathcal{B}_n(p) =: \mathcal{B}(p)^{*n}$$

Ce fait est autrement plus clair : soit (X_i) une suite de v.a. i.i.d. de distribution de Bernoulli $\mathcal{B}(p)$. Posons

$$X = X_1 + \cdots + X_n \ (=: S_n)$$

Alors, X compte par construction le nombre de succès parmi les n v.a. de Bernoulli qui, par indépendance, donnent

$$P(X=k) = \#\{X=k\}\, p^k q^{n-k} = C_n^k\, p^k q^{n-k}$$

○

<u>Exemple</u> **2.28** (**Loi binomiale négative**) On définit la distribution binomiale négative de paramètres (n,p) (avec $q = 1-p$) comme étant la somme de n v.a. i.i.d. de distribution géométrique $\mathcal{G}(p)$. Ainsi, par définition, si X est une v.a. suivant une telle distribution, elle compte le nombre total d'échecs avant le n-ème succès dans une succession de v.a. de Bernoulli $\mathcal{B}(p)$ indépendantes, i.e.

$$P(X=k) = C_{n+k-1}^{n-1} p^n q^k$$

Il est alors facile de vérifier que

$$E(X) = n\,\frac{q}{p}, \quad V(X) = n\,\frac{q}{p^2}$$

○

Pour des exemples concrets de v.a. continues, nous avons la version spéciale du théorème précédent.

Théorème 2.29 *Soit X et Y deux v.a. continues et indépendantes. Alors, $X+Y$ possède la densité*

$$(f_X * f_Y)(z) = \int f_X(z-y)\, f_Y(y)\, \mathrm{d}y, \tag{2.14}$$

laquelle est appelée ***produit de convolution*** *de f_X et f_Y.*

Remarque 2.30 *Par symétrie, il est clair que*

$$(f_X * f_Y)(z) = \int f_X(x)\, f_Y(z-x)\, \mathrm{d}x$$

Nous avons un exemple standard d'application de ce dernier résultat (cf. Exemple 1.48).

Théorème 2.31 *On a que*

$$\mathcal{N}(\mu, \sigma^2) * \mathcal{N}(\nu, \theta^2) = \mathcal{N}(\mu+\nu, \sigma^2+\theta^2)$$

Exercice 2.32 (Loi gamma) La **densité gamma** de paramètres α et λ est définie par

$$f(x) = \frac{1}{\Gamma(\alpha)}\, \lambda^\alpha\, x^{\alpha-1}\, e^{-\lambda x}\, \mathbb{1}_{\mathbb{R}_+}(x), \tag{2.15}$$

où $\Gamma(\alpha)$, l'élément normalisateur, est la célèbre fonction gamma (d'Euler), i.e.

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\, e^{-x}\, \mathrm{d}x$$

Une v.a. X qui a une une telle densité de probabilité est signalée par $X \sim \Gamma(\alpha, \lambda)$. Remarquer que $\Gamma(1, \lambda) = \mathcal{E}(\lambda)$. Et par abus de notation, nous écrivons $\Gamma(\alpha, 1) = \Gamma(\alpha)$.

Montrer que

$$\Gamma(\alpha, \lambda) * \Gamma(\beta, \lambda) = \Gamma(\alpha+\beta, \lambda)$$

En déduire que le produit de convolution de n fois $\mathcal{E}$ avec elle-même donne

$$\Gamma(n) =: \mathcal{E}^{*n}$$

2.6 Extension de Kolmogorov

Avant d'aller plus loin dans l'étude des suites des v.a. indépendantes, il est impératif de se poser la question de leur existence. Quant à une collection finie de v.a. indépendantes, nous avons vu au Théorème 2.14 et sa remarque qu'il est facile de les construire en prenant la mesure de probabilité produit. En effet, étant donné une collection de fonctions de distributions $F_i, i = 1, \ldots, n$ que nous associons aux lois de probabilité $\mu_i, i = 1, \ldots, n$, il existe alors sur l'espace produit $(\mathbb{R}^n, \mathbb{R}^n)$ en vertu de ce dernier théorème une loi de probabilité unique, produit des μ_i, qui est définie par

$$P\left(\prod_{i=1}^{n}(a_i, b_i]\right) = \prod_{i=1}^{n}(F_i(b_i) - F_i(a_i)) = \prod_{i=1}^{n}\mu_i((a_i, b_i]).$$

Les coordonnées $(x_1, \ldots, x_n)$ dans $\mathbb{R}^n$ représentent alors respectivement les v.a. X_i qui sont ainsi indépendantes.

Nous voulons ici étendre cette construction et montrer qu'une suite infinie de v.a. indépendantes existe bien sur l'espace infini

$$\mathbb{R}^{\mathbb{N}^*} = \{x = (x_1, x_2, \ldots) : x_i \in \mathbb{R}\}.$$

Cet espace n'est autre que l'espace des suites numériques. Il est naturellement associé à $\mathbb{R}^{\mathbb{N}^*}$ la tribu produit $\mathcal{B}^{\mathbb{N}^*}$ générée par les pavés $\prod_{i=1}^{n}(a_i, b_i]$ où $-\infty \leq a_i < b_i \leq \infty$.

En fait, nous sommes sur le point de voir une condition suffisante à l'existence de processus stochastique, en particulier les suites de v.a. indépendantes.

Théorème 2.33 (Théorème d'extension de Kolmogorov) *Pour chaque $n \in \mathbb{N}^*$, soit μ_n une mesure de probabilité sur $\mathbb{R}^n$. Supposons que $(\mu_n)_{n\in\mathbb{N}^*}$ soit une **famille de mesures de probabilité consistantes**, i.e.*

$$\mu_{n+1}\Big(\prod_{i=1}^{n}(a_i, b_i] \times \mathbb{R}\Big) = \mu_n\Big(\prod_{i=1}^{n}(a_i, b_i]\Big) \tag{2.16}$$

il existe alors une unique mesure de probabilité μ sur $\mathbb{R}^{\mathbb{N}^*}$ telle que

$$\mu\Big(x : (x_1, \ldots, x_n) \in \prod_{i=1}^{n}(a_i, b_i]\Big) = \mu_n\Big(\prod_{i=1}^{n}(a_i, b_i]\Big) \tag{2.17}$$

Remarque 2.34 *Quand nous obtenons la mesure de probabilité μ de l'extension de Kolmogorov, nous pouvons alors définir notre suite de v.a. réelles (X_n) telle que la distribution de $X_n(\to x_n)$ est donnée par*

$$\nu_n(A) = \mu(x : x_n \in A).$$

Ainsi, nous avons construit une suite de v.a. (X_n), i.e. un processus stochastique à temps discret, lesquelles v.a. suivent les distributions désirées. (Une construction similaire est vraie en temps continue dans $\mathbb{R}^{\mathbb{R}_+}$.)

Si les les mesures μ_n sont définies par un produit de lois de probabilité sous la forme

$$\mu_n\Big(\prod_{i=1}^{n}(a_i, b_i]\Big) = \prod_{i=1}^{n}\nu_i((a_i, b_i])$$

les X_n sont alors indépendantes de distributions respectives ν_n. En plus, si les ν_n sont égales, on dit que les X_n sont ***i.i.d.****, indépendantes et identiquement distribuées.*

Le raisonnement ci-après peut être généralisé pour construire un processus stochastique dans l'espace $((\mathbb{R}^d)^{\mathbb{N}^*}, (\mathcal{B}^d)^{\mathbb{N}^*})$.

2.7 Loi du $0-1$ de Kolmogorov

L'indépendance implique un phénomène *asymptotiques* étrange qui fait l'objet de cette section. Pour l'énoncer, nous avons besoin d'établir quelques définitions.

Soit (X_n) une suite de v.a. et posons $S_n = X_1 + \cdots + X_n$. Définissons

$$\mathcal{T}_n = \sigma(X_n, X_{n+1}, \ldots) = \text{le future de la suite } (X_n) \text{ après } n.$$

C'est en fait la plus petite tribu qui rend les $X_m, m \geq n$ mesurables. Soit

$$\mathcal{T} = \cap_n \mathcal{T}_n = \textbf{tribu asymptotique}.$$

Comme les $\mathcal{T}_n$ sont décroissantes, l'on peut dire intuitivement que si un évènement A est dans $\mathcal{T}$, le fait de connaître ou non un nombre fini de $\sigma(X_n)$ n'affecte pas l'occurrence de A.

<u>Exemples</u> **2.35** Si $A_n \in \mathcal{B}$, alors

$$\limsup_n \{X_n \in A_n\} \in \mathcal{T}$$

D'autre part, il est facile de vérifier que

$$\{\lim S_n \text{ existe}\} \in \mathcal{T}, \tag{2.18}$$

et aussi que

$$\{\limsup S_n/c_n > x\} \in \mathcal{T} \text{ si } c_n \to \infty,$$

Mais

$$\{\limsup S_n > 0\} \notin \mathcal{T}$$

○

Le phénomène en question est le suivant :

Théorème 2.36 (loi du 0−1 de Kolmogorov) *Soit (X_n) une suite des v.a. indépendantes. Si $A \in \mathcal{T}$, alors on a*

$$P(A) = 0 \text{ ou } 1.$$

<u>EXEMPLE</u> **2.37** Au vu de (2.18), nous avons l'exemple fondamental suivant comme application de la loi du 0 − 1 de Kolmogorov :

$$P(\lim S_n \text{ existe}) = 0 \text{ ou } 1.$$

Nous allons voir plus tard des situations où cette probabilité est 1, autrement dit, la série $\sum_n X_n$ converge p.s. ○

3 Fonction caractéristique

Nous nous intéresserons dans ce chapitre à une fonction qui caractérise les distributions de probabilité en ce sens qu'elle peut leur servir comme définition, et en même temps elle donne quelques unes de leurs propriétés importantes. C'est aussi un puissant outil de calcul dont nous verrons des applications au Chapitre 4 ; son intérêt premier se voit sur la somme de v.a. indépendantes.

3.1 Définition et exemples

Avant de définir l'objet principal du présent chapitre, nous aurons besoin auparavant de faire une *extension* immédiate de la définition de l'espérance mathématique. Soit X une v.a. qui prend ses valeurs dans l'espace mesurable $(S, \mathcal{S})$ et f est une fonction complexe définie sur S :

$$f(x) = f_1(x) + i f_2(x)$$

où f_1 et f_2 sont des fonctions mesurables définies sur S et à valeurs dans $\mathbb{R}$. On dit que $f(X)$ est intégrable ssi $f_1(X)$ et $f_2(X)$ sont intégrables. On définit alors l'espérance de $f(X)$ par

$$E(f(X)) = E(f_1(x)) + iE(f_2(X))$$

Soit X une v.a. réelle définie sur $(\Omega, \mathcal{F}, P)$. On définit pour tout $t \in \mathbb{R}$ la **fonction caractéristique** de X (en abrégé **f.c.**) par

$$\varphi_X(t) = E(e^{itX}) = E(\cos(tX)) + iE(\sin(tX)). \tag{3.1}$$

Cette fonction est bien définie puisque $\cos(tX)$ et $\sin(tX)$ sont clairement intégrables.

Si μ est la distribution de X, on a alors par le théorème de transport

$$\varphi_X(t) = \int e^{itx} \,\mathrm{d}\mu(x) =: \phi_\mu(t)$$

Et si μ admet une densité de probabilité f, on note également cette fonction caractéristique par φ_f. Donc, on a dans ce cas

$$\phi_\mu(t) = \phi_f(t) = \int e^{itx} f(x)\,\mathrm{d}x$$

On voit ainsi que la f.c. d'une v.a. est le résultat d'une transformation de sa distribution, ou éventuellement de sa densité, que l'on appelle communément **transformation de Fourier**.

Clairement, la f.c. d'une v.a. X dépend uniquement de sa distribution, ainsi deux v.a. qui ont la même loi de probabilité auront la même f.c. Nous verrons plus bas (cf. Théorème 3.25) que la réciproque est vraie. Autrement dit, la f.c. détermine *uniquement* la loi de probabilité.

<u>Exemples</u> **3.1** (**Lois discrètes**) Si $X \sim \mathcal{B}(n,p)$, alors

$$\varphi_X(t) = \sum_{k=0}^{n} C_n^k\, p^k (1-p)^{n-k}\, e^{itk} = (1 + p(e^{it} - 1))^n$$

Maintenant, supposons que $X \sim \mathcal{P}(\lambda)$. Dans ce cas, on a que

$$\varphi_X(t) = \sum_{n \geq 0} e^{-\lambda} \frac{\lambda^n}{n!}\, e^{itn} = e^{-\lambda}\, e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}$$

○

<u>Exemple</u> **3.2** (**Loi exponentielle**) Si $X \sim \mathcal{E}$, alors on a par intégration simple,

$$\varphi_X(t) = \int_0^\infty e^{x(it-1)}\,\mathrm{d}x = \left[\frac{e^{x(it-1)}}{it-1}\right]_0^\infty = \frac{1}{1-it}$$

○

<u>Exemple</u> **3.3** (**Loi normale**) Soit $X \sim \mathcal{N}$. Nous avons

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int e^{itx - x^2/2}\,\mathrm{d}x = \frac{e^{-t^2/2}}{\sqrt{2\pi}} \int e^{-(x-it)^2/2}\,\mathrm{d}x \tag{3.2}$$

Désignons par $L = \{z : \mathrm{Im}(z) = -t\}$, la droite des nombres complexes dont la partie imaginaire est $-t$, parcourue de gauche à droite, nous obtenons

$$\int e^{-(x-it)^2/2}\,\mathrm{d}x = \int_L e^{-z^2/2}\,\mathrm{d}z$$

La fonction $z \longmapsto e^{-z^2/2}$ est analytique sur tout le plan complexe, donc son intégrale le long d'une courbe simple fermée est nulle. Choisissons en particulier le rectangle de sommets $A = (-a, -t), B = (a, -t), C = (a, 0)$ et $D = (-a, 0)$. Alors, nous obtenons que

$$\begin{aligned}\left|\int_{BC} e^{-z^2/2}\,\mathrm{d}z\right| &= \left|\int_{-t}^{0} e^{-(a+iu)^2/2}\,\mathrm{d}u\right| \\ &\leq e^{-a^2/2}\int_{-t}^{0} |e^{u^2/2+iua}|\,\mathrm{d}u \\ &\leq e^{-a^2/2}\int_{0}^{|t|} e^{u^2}\,\mathrm{d}u\end{aligned}$$

De même, nous avons

$$\left|\int_{DA} e^{-z^2/2}\,\mathrm{d}z\right| \leq e^{-a^2/2}\int_{0}^{|t|} e^{u^2}\,\mathrm{d}u$$

Il en résulte alors que

$$0 = \lim_{a\to+\infty}\frac{1}{\sqrt{2\pi}}\int_{ABCD} e^{-z^2/2}\,\mathrm{d}z = \frac{1}{\sqrt{2\pi}}\int_{L} e^{-z^2/2}\,\mathrm{d}z + \frac{1}{\sqrt{2\pi}}\int_{+\infty}^{-\infty} e^{-x^2/2}\,\mathrm{d}x$$

D'où,

$$\frac{1}{\sqrt{2\pi}}\int_{L} e^{-z^2/2}\,\mathrm{d}z = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-x^2/2}\,\mathrm{d}x = 1$$

Ainsi, nous obtenons que

$$\varphi_X(t) = e^{-t^2/2} \tag{3.3}$$

De manière générale, considérons $Y = \sigma X + \mu$ avec $\sigma > 0, \mu \in \mathbb{R}$; par définition, $Y \sim \mathcal{N}(\mu, \sigma^2)$. Alors, nous avons

$$\varphi_Y(t) = e^{it\mu}E\left(e^{it\sigma X}\right) = e^{it\mu}\varphi_X(\sigma t) = e^{it\mu-\sigma^2t^2/2} \tag{3.4}$$

○

Exercice 3.4 (Loi gamma) Soit $X \sim \Gamma(n, \lambda)$ avec $n \in \mathbb{N}$ (cf. Exercice 2.32). Montrer que

$$\varphi_X(t) = \frac{1}{(1 - it/\lambda)^n} \tag{3.5}$$

Ensuite, généraliser au cas général ; si $X \sim \Gamma(\alpha, \lambda)$ avec $\alpha > 0$, alors

$$\frac{1}{(1 - it/\lambda)^\alpha} \tag{3.6}$$

○

Enfin, on généralise la f.c. aux vecteurs aléatoires de la manière intuitive suivante. Si X est un vecteur aléatoire dans $\mathbb{R}^n$ muni de son produit scalaire naturel : $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$, on appelle f.c. de X la fonction de $\mathbb{R}^n$ dans $\mathbb{C}$ donnée par

$$t \longmapsto \varphi_X(t) = \int e^{i\,\langle t,x\rangle}\,\mathrm{d}\mu_X(x) \tag{3.7}$$

<u>Exemple</u> **3.5** Soit $X \in \mathbb{R}^n$ un vecteur aléatoire de distribution $\mathcal{N}(m, \Sigma)$ (cf. Exemple 1.50). Alors, la formule (3.8) se généralise facilement par

$$\varphi_X(t) = \exp\left(\langle t, m\rangle - \frac{1}{2}\,{}^t t \Sigma t\right)$$

Observons en même temps que si Σ est semi-définie positive, c'est-à-dire qu'elle est de rang $r < n$, alors il existe (voir [3]) une matrice D d'ordre (n, r) telle que

$$X \stackrel{d}{=} m + D\,\mathcal{N}(0, I_r)$$

En particulier, si $\Sigma = \mathrm{diag}(\sigma_i^2)$, alors la loi marginale de la coordonnée X_i est $\mathcal{N}(m_i, \sigma_i^2)$ si $\sigma_i \neq 0$, sinon $X \stackrel{\text{p.s.}}{=} m_i$. ○

3.2 Propriétés générales

En premier, nous réunissons ci-après les propriétés fondamentales de la f.c. :

Théorème 3.6 *Toute f.c. φ (d'une v.a. X) est uniformément continue sur $\mathbb{R}$ et vérifie les propriétés suivantes :*

$$\begin{array}{ll} (i) & \varphi(0) = 1. \\ (ii) & \varphi(-t) = \overline{\varphi(t)}. \\ (iii) & |\varphi(t)| \leq 1. \end{array}$$

En outre, si $Y = aX + b$ avec $a, b \in \mathbb{R}$, alors on a que

$$\varphi_Y(t) = e^{itb}\varphi_X(at)$$

<u>Exemple</u> **3.7 (Distribution normale)** Soit $X \sim \mathcal{N}$ et considérons $Y = \sigma X + \mu$ avec $\sigma > 0, \mu \in \mathbb{R}$. Alors, par la Proposition 3.6, il vient que

$$\varphi_Y(t) = e^{it\mu}\varphi_X(\sigma t) = e^{it\mu - \sigma^2 t^2/2} \tag{3.8}$$

○

Une autre propriété intéressante des f.c. dit que la moyenne de f.c. est une f.c.

Proposition 3.8 *Si $\varphi_1, \varphi_2, \ldots, \varphi_n$ sont des f.c. associées respectivement aux f.r. $F_1, \ldots, F_n$, et que $\lambda_1 + \cdots + \lambda_n = 1$, une somme de valeurs non négatives, alors $\sum_{i=1}^n \lambda_i \varphi_i$ est la f.c. de la f.r. $\sum_{i=1}^n \lambda_i F_i$.*

Remarque 3.9 *Il est facile à vérifier que la présente assertion reste vraie si l'on prend une distribution de probabilité infinie $\sum_{n=1}^\infty \lambda_n = 1$.*

<u>Exemple</u> **3.10** (**Densité exponentielle bilatérale**) On appelle densité exponentielle bilatérale la fonction

$$f(x) = \frac{1}{2}\, e^{-|x|}$$

Sa f.c. est alors

$$\varphi(t) = \frac{1}{1+t^2}$$

En effet, nous pouvons vérifier ça en utilisant la Proposition 3.8. Remarquons que

$$f(x) = \frac{1}{2}\, e^{x}\, \mathbb{1}_{\mathbb{R}_-}(x) + \frac{1}{2}\, e^{-x}\, \mathbb{1}_{\mathbb{R}_+}(x)$$

De ce fait, nous voyons qu'une telle f.r. peut être exprimée sous la forme $\frac{1}{2}\, F_1 + \frac{1}{2}\, F_2$ où F_1 et F_2 sont respectivement les f.r. d'une exponentielle $-\mathcal{E}$ et de $\mathcal{E}$; les f.c. associées à F_1 et F_2 sont (cf. Exemple 3.2)

$$\varphi_1(t) = \frac{1}{1+it}, \quad \varphi_2(t) = \frac{1}{1-it}$$

En vertu de ladite proposition, nous obtenons que la f.c. en question est

$$\varphi(t) = \frac{1}{2}\, \varphi_1(t) + \frac{1}{2}\, \varphi_2(t) = \frac{1+it+1-it}{2(1+t^2)} = \frac{1}{1+t^2}$$

○

En outre, nous donnons le résultat crucial suivant qui exprime bien l'effet de l'indépendance des v.a. sur leur somme à l'aide des f.c. L'utilité de ce résultat sera visible dans l'application du théorème de convergence des f.c. (cf. Chapitre 4).

Théorème 3.11 *Si X et Y sont deux v.a. indépendantes, alors on a*

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) \tag{3.9}$$

Exemple **3.12** (**Loi binomiale**) Reprenons l'Exemple 2.27 dans lequel nous avons vu que la convolution de n v.a. X_i de Bernoulli $\mathcal{B}(p)$ indépendantes donne une binomiale $\mathcal{B}(p, n)$. Ici, grâce au Théorème 3.11, nous pouvons arriver à cette même conclusion plus facilement. En effet, si X représente une telle somme, nous avons par ce même théorème que

$$\varphi_X(t) = \varphi_{X_i}(t)^n = \big(1 + p(e^{it} - 1)\big)^n$$

○

Exemple **3.13** (**Loi de Poisson**) Si X est une somme de n v.a. i.i.d. X_i suivant une distribution de Poisson $\mathcal{P}(\lambda)$, alors il vient par indépendance et de l'Exemple 3.1-(iv) que

$$\varphi_X(t) = \varphi_{X_i}(t)^n = e^{n\lambda(e^{it}-1)}$$

D'où, il résulte que

$$X \sim \mathcal{P}(n\lambda)$$

○

Exemple **3.14** (**Loi triangulaire**) On définit (géométriquement) la loi triangulaire par la densité

$$f(x) = (1 - |x|)\mathbb{1}_{[-1,1]}(x)$$

Il est possible de calculer directement la f.c. de ladite distribution triangulaire sauf que nous avons utilisé le fait qu'une telle distribution est égale celle de la somme de deux v.a. i.i.d. de loi uniforme sur $[-1/2, 1/2]$. En effet, la formule (2.14) donne que la densité de cette somme est

$$\int_{-1/2}^{1/2} \mathbb{1}(x-y)_{[-1/2,1/2]}\,\mathrm{d}y = \lambda([-1/2, 1/2] \cap [x - 1/2, x + 1/2]) = f(x)$$

Ainsi, par le Théorème 3.11 et l'Exemple **??**, nous avons

$$\varphi_X(t) = \left(\frac{e^{it/2} - e^{-it/2}}{it}\right)^2 = \left(\frac{2\sin(t/2)}{t}\right)^2 = 2\,\frac{(1-\cos t)}{t^2}$$

où nous avons utilisé que $\cos t = 1 - 2\sin^2(t/2)$. ○

Exercice 3.15 Vérifier à l'aide du Théorème 3.11 que l'assertion du Théorème 2.31 est bien vraie.

3.3 Moments et Dérivées

Nous allons nous intéresser ici à la relation entre la f.c. d'une v.a. donnée X et ses moments. Remarquons d'abord que

$$e^{itX} = \sum_{n\geq 0} \frac{(it)^n}{n!} X^n$$

Si nous intégrons cette somme terme par terme (sans se soucier encore des conditions de sa faisabilité), nous obtenons

$$\varphi_X(t) = \sum_{n\geq 0} \frac{(it)^n}{n!} \mu_n$$

Ainsi, par comparaison avec la formule de Taylor, nous pouvons voir que nous devrions avoir

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n}\varphi_X(0) = \varphi_X^{(n)}(0) = i^n \mu_n$$

En fait, ce *calcul de physicien* est vrai et est repris formellement dans le théorème suivant.

Théorème 3.16 *Soit X une v.a. et φ sa f.c. Alors, si les n premiers moments de X existent, φ est n fois dérivable et ses dérivées sont données par*

$$\varphi^{(k)}(t) = i^k E\big(X^k\, e^{itX}\big) \tag{3.10}$$

D'où, en particulier, on a

$$\varphi^{(k)}(0) = i^k E\big(X^k\big) \tag{3.11}$$

Remarque 3.17 *L'existence de $\varphi^{(n)}(0)$ et la formule de Taylor-Young garantisse le développement suivant :*

$$\varphi(t) = \sum_{k=0}^{n} \frac{\varphi^{(k)}(0)}{k!}\, t^k + o(t^n) = \sum_{k=0}^{n} E\big(X^k\big)\, \frac{(it)^k}{k!} + o(t^n) \tag{3.12}$$

La proposition suivante donne une réciproque au Théorème 3.16.

Proposition 3.18 *Si la f.c. d'une v.a. est n fois dérivable en 0, alors celle-ci admet tous les moments jusqu'à l'ordre $2k$ avec $2k \leq n$.*

Le Théorème 3.16 avec l'existence de tous les moments conduit donc au résultat suivant :

Théorème 3.19 *Si tous les moments μ_n d'une v.a. X existent et que*

$$\limsup_{n\to\infty} \sqrt[n]{\frac{|\mu_n|}{n!}} = \frac{1}{\varrho} < \infty, \tag{3.13}$$

alors pour tout $|t| < \varrho$, on a

$$\varphi(t) = \sum_{n\geq 0} \mu_n \frac{(it)^n}{n!} \tag{3.14}$$

Remarque 3.20 *En vertu de (3.14), on en déduit que φ est analytique au voisinage de tout point de $\mathbb{R}$ et est prolongeable en une fonction analytique sur toute la bande du plan complexe $\{x + iy : |y| < \varrho\}$.*

La limite (3.13) donne une condition suffisante pour déterminer une f.c., et donc une distribution de probabilité. Notons qu'elle est vraie si X est bornée. Cependant, nous avons un exemple ci-dessous qui montre que les moments ne caractérisent généralement pas une mesure de probabilité.

Exemple **3.21** (**Distribution normale**) La f.c. de la loi normale standard est donnée par

$$e^{-t^2/2} = \sum_{k=0}^{\infty} (-1)^n \frac{t^{2k}}{2^k k!} = \sum_{n=0}^{\infty} \frac{\varphi^{(n)}(0)}{n!} t^n$$

Ce qui, par comparaison, donne que

$$\varphi^{(2k)}(0) = (-1)^k \frac{(2k)!}{2^k k!}, \quad \varphi^{(2k+1)}(0) = 0$$

D'où, en utilisant (3.11), nous obtenons que les moments d'ordre impair sont nuls et ceux d'ordre pair sont de la forme

$$\mu_{2k} = \frac{(2k)!}{2^k k!}$$

○

Maintenant, au vu de (3.14), la question est : une loi est-elle déterminée par ses moments ?! L'exemple suivant montre que la réponse est non !

Exemple **3.22** (**Loi** log-**normale**) Soit X une v.a. de loi normale standard. Posons $Y = e^X$ et montrons que Y est de densité

$$f_Y(y) = (2\pi)^{-1/2} y^{-1} e^{-(\log y)^2/2} \mathbb{1}_{\mathbb{R}_+^*}(y)$$

En effet, Y ne prend que des valeurs positives. Donc, nous avons

$$F_Y(y) = P(Y \leq y) = P(X \leq \log y) = F_X(\log y)\, \mathbb{1}_{\mathbb{R}_+^*}(y)$$

Ce qui est une fonction continue sur $\mathbb{R}$ et dérivable sur $\mathbb{R}^*$. D'où, il vient par dérivation que f_Y est bien la densité de Y. La loi de Y s'appelle la *loi* log-*normale*.

Ensuite, pour $a \in [-1, 1]$, soit

$$f_a(y) = f_Y(y)(1 + a \sin(2\pi \log y))$$

Remarquons que nous avons là une densité de probabilité. En effet, elle est non négative, de plus, observons que par la formule de changement de variable, nous avons

$$\int_0^\infty f_Y(y) \sin(2\pi \log y)\, \mathrm{d}y = E(\sin(2\pi \log Y)) = E(\sin(2\pi X)) = 0$$

La dernière égalité vient du fait que la densité de X est paire et sin est impaire. Par conséquent, nous avons $\int f_a \,\mathrm{d}x = 1$.

Maintenant, supposons que Y_a soit une v.a. de densité f_a. Et soit $n \in \mathbb{N}^*$. Alors, nous avons

$$E(Y_a^n) = \int y^n f_a(y)\, \mathrm{d}y = E(Y^n) + E(e^{nX}\, \sin(2\pi X))$$

Or, la dernière intégrale est nulle. En effet, par le théorème de transport,

$$\begin{aligned} E(e^{nX}\, \sin(2\pi X)) &= \frac{e^{n^2/2}}{\sqrt{2\pi}} \int e^{-(x-n)^2/2}\, \sin(2\pi x)\, \mathrm{d}x \\ &= \frac{e^{n^2/2}}{\sqrt{2\pi}} \int e^{-(x-n)^2/2}\, \sin(2\pi (x-n))\, \mathrm{d}x \\ &\overset{u=x-n}{=} \frac{e^{n^2/2}}{\sqrt{2\pi}} \int e^{-u^2/2}\, \sin(2\pi u)\, \mathrm{d}u \\ &= 0 \end{aligned}$$

Ainsi, Y_a et Y ont les mêmes moments sans avoir toutefois la même loi. On conclut que les moments d'une v.a. ne déterminent pas sa loi. ○

3.4 Formule d'inversion

Nous allons maintenant traiter la question inverse qui est de déterminer une mesure de probabilité à partir de sa f.c., et par la même occasion, répondre à la question de l'unicité de celle-ci. En premier, nous avons un résultat général qui donne une formule inverse permettant de connaître une distribution de probabilité à partir de la f.c. associée :

Théorème 3.23 (Formule d'inversion) *Soit μ une mesure de probabilité et $\varphi(t)$ la f.c. associée. Si $a < b$, alors on a*

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_a^b e^{-itx}\varphi(t)\,\mathrm{d}x\,\mathrm{d}t = \mu((a,b)) + \frac{1}{2}\mu(\{a,b\}) \tag{FI}$$

Remarque 3.24 *Si $\mu = \delta_0$, alors $\varphi(t) = 1$ et si $a = -1, b = 1$, alors*

$$\int_a^b e^{-itx}\,\mathrm{d}x = \frac{e^{-ita} - e^{-itb}}{it} = 2\frac{\sin t}{t}$$

L'intégrale dans (FI) ne converge pas absolument.

Soit F est la f.r. associée à μ. Si a et b sont deux de ses points de continuité tels que $a < b$, alors (FI) donne

$$F(b) - F(a) = \mu((a,b)) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_a^b e^{-itx}\varphi(t)\,\mathrm{d}x\,\mathrm{d}t \tag{3.15}$$

En conséquence, la formule d'inversion permet de prouver facilement que la f.c. détermine uniquement une distribution de probabilité ou de manière équivalente une f.r. :

Théorème 3.25 (Unicité) *La f.c. détermine uniquement une f.r.*

Exemple **3.26** (**Distribution de Cauchy**) On définit la distribution de Cauchy par la densité

$$f(x) = \frac{1}{\pi(1+x^2)} = \frac{1}{\pi}\frac{\mathrm{d}}{\mathrm{d}x}\arctan x$$

Nous allons vérifier par (FI) que sa f.c. est donnée par

$$\varphi(t) = e^{-|t|}$$

Remarquons que $|e^{-itx}\varphi(t)| = e^{-|t|}$, donc, si $x \in [a,b]$, nous avons là une fonction bornée et intégrable. De ce fait, nous pouvons appliquer Fubini et obtenir que

$$\begin{aligned}\int_{-T}^{T}\int_a^b e^{-itx}e^{-|t|}\,\mathrm{d}x\,\mathrm{d}t &= \int_a^b\int_{-T}^{0} e^{t(1-ix)}\,\mathrm{d}t\,\mathrm{d}x + \int_a^b\int_0^T e^{-t(1+ix)}\,\mathrm{d}t\,\mathrm{d}x \\ &= \int_a^b \frac{1-e^{-T(1-ix)}}{1-ix}\,\mathrm{d}x + \int_a^b \frac{1-e^{-T(1+ix)}}{1+ix}\,\mathrm{d}x\end{aligned}$$

Mais, nous avons pour T assez grand,

$$\left|\frac{1-e^{-T(1\pm ix)}}{1\pm ix}\right| \leq 2$$

Ainsi, par le (TCB) et quand $T \to \infty$, cette somme d'intégrales tend vers

$$\begin{aligned}\int_a^b \frac{1}{1+ix}\,\mathrm{d}x + \int_a^b \frac{1}{1-ix}\,\mathrm{d}x &= \int_a^b \frac{1+ix}{1+x^2}\,\mathrm{d}x + \int_a^b \frac{1-ix}{1+x^2}\,\mathrm{d}x \\ &= \int_a^b \frac{2}{1+x^2}\,\mathrm{d}x\end{aligned}$$

En divisant par 2π, le résultat s'ensuit. ○

On peut généraliser ce dernier exemple pour obtenir le cas particulier intéressant suivant.

Théorème 3.27 *Soit μ une mesure de probabilité et φ sa f.c.. Si $\int |\varphi(t)|\,\mathrm{d}t < \infty$, alors μ admet une densité continue donnée par*

$$f(x) = \frac{1}{2\pi}\int e^{-itx}\,\varphi(t)\,\mathrm{d}t \tag{3.16}$$

Exercice 3.28 Donner un exemple d'une f.c. φ d'une mesure de probabilité continue telle que $\int |\varphi(t)|\,\mathrm{d}t = \infty$, autrement dit, montrer que la réciproque du Théorème 3.27 est fausse. ○

3.5 Fonction génératrice ; $X \in \mathbb{N}$

Dans le cas particulier où la probabilité est définie sur $\mathbb{N}$, il s'offre à nous une autre transformation qui fait correspondre à celle-ci une fonction complexe analytique sur le disque unité d'origine 0.

Soit P une probabilité sur $\mathbb{N}$ définie par $P(\{n\}) = p_n$. En vertu du Lemme d'Abel et de l'hypothèse $\sum_{n\geq 0} p_n = 1$, la série $\sum_n p_n z^n$ est alors convergente dans le disque $D = \{z \in \mathbb{C} : |z| \leq 1\}$. Ainsi, on appelle **fonction génératrice** de P (en abrégé **f.g.**) l'application définie sur D par

$$G_P(z) = \sum_n p_n z^n$$

De manière équivalente, on définit la **f.g. d'une v.a.** X par

$$G_X(z) = E(z^X) = \sum_n P(X=n) z^n \tag{3.17}$$

Remarquons que $\varphi_P(t) = G_P(e^{it})$. La f.g. apparaît donc comme un prolongement de la f.c. à tout le disque unité centré, laquelle est définie sur son cercle.

Le résultat suivant montre qu'une f.g. caractérise entièrement la loi de probabilité associée en ce sens qu'il suffit de connaître une f.g. pour connaître la loi de probabilité associée.

Théorème 3.29 *Une distribution de probabilités P sur $\mathbb{N}$ est entièrement définie par sa fonction génératrice en ce sens qu'on a*

$$p_n = \frac{G_P^{(n)}(0)}{n!} \tag{3.18}$$

<u>Exemple</u> **3.30** ($\mathcal{B}(n,p)$) D'après la formule du binôme, la f.g. Soit d'une $\mathcal{B}(n,p)$.

$$G(z) = \sum_{i=0}^{n} C_n^i p^i (1-p)^{n-i} z^i = (pz + 1 - p)^n$$

○

<u>Exemple</u> **3.31** ($\mathcal{P}(\lambda)$) La f.g. d'une $\mathcal{P}(\lambda)$ est

$$G(z) = \sum_n e^{-\lambda} \frac{(\lambda z)^n}{n!} = e^{(z-1)\lambda}$$

○

Comme pour la f.c., la f.g. donne l'espérance d'une v.a. de la manière suivante.

Théorème 3.32 *Si X est une v.a. dans $\mathbb{N}$, alors on a*

$$E(X) = G'_X(1)$$

Il existe un résultat similaire au Théorème 3.11 pour les f.g.

Théorème 3.33 *Si X et Y sont deux v.a. indépendantes dans $\mathbb{N}$, alors on a*

$$G_{X+Y}(z) = G_X(z) G_Y(z) \tag{3.19}$$

4 Modes de convergence

Nous sommes concernés dans notre cours par quelques modes de convergence fondamentales des suites de v.a., à savoir les convergences presque sûre, en probabilité, dans L^p et en loi. Il s'agit ici de connaître quelques unes de leurs propriétés importantes et leurs interactions avec des applications principales.

4.1 Convergence p.s.

On rappelle de (1.25) qu'une suite de v.a. (X_n) est convergente **presque surement** (p.s.) vers une v.a. X si

$$P(\lim X_n = X) = 1$$

et l'on écrit

$$X \xrightarrow[n\to\infty]{\text{p.s.}} X$$

Ainsi, la convergence p.s. équivaut par définition à

$$\forall \varepsilon > 0 : \quad P(\cup_{l\geq 1} \cap_{n\geq l} \{|X_n - X| < \varepsilon\}) = 1$$

ou, en passant au complémentaire, est équivalente à

$$\forall \varepsilon > 0 : \quad \lim_{l\to\infty} P(\sup_{n\geq l} |X_n - X| \geq \varepsilon) = 0 \tag{4.1}$$

Remarquons que ceci équivaut aussi à

$$P\left(\cap_{k\geq 1} \cup_{l\geq 1} \cap_{n\geq l}\{|X_n - X| < 1/k\}\right) = 1$$

En effet, pour tout choix de $\varepsilon > 0$, nous pouvons trouver k tel que $\varepsilon \geq 1/k$.

Propriétés

La convergence p.s. d'une suite de v.a. (X_n) en tant que convergence ponctuelle, est conservée par continuité. Explicitement, nous avons

Théorème 4.1 *Si f est une fonction continue et que $X_n \to X$ p.s., alors*

$$f(X_n) \xrightarrow[n\to\infty]{p.s.} f(X)$$

1^er^ lemme de Borel-Cantelli

Pour montrer la convergence p.s., il est un moyen usuel qui utilise l'incontournable résultat suivant qui va nous servir dans diverses démonstrations à venir, notamment dans la preuve de la *loi forte des grands nombres* (cf. Chapitre 5.2).

Théorème 4.2 (Premier lemme de Borel-Cantelli) *Pour toute suite d'évènements (A_n), on a*

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Longrightarrow P(\limsup_n A_n) = 0 \tag{LBC1}$$

<u>Exemple</u> **4.3** (**La réciproque de** (LBC1) **est fausse**) Dans l'espace probabilisé $([0,1], \mathcal{B} \cap [0,1], \lambda)$, soit les évènements $A_n = (0, 1/n)$. Nous avons clairement $\limsup A_n = \emptyset$ mais

$$\sum \lambda(A_n) \geq \sum 1/n = \infty$$

○

En conséquence de ce lemme, nous avons :

Proposition 4.4 *Soit (X_n) et X des v.a. Alors, on a que*

$$\forall \varepsilon > 0 : \sum_n P(|X_n - X| > \varepsilon) < \infty \Longrightarrow X_n \xrightarrow[n\to\infty]{p.s.} X \tag{4.2}$$

<u>Exemple</u> **4.5** Soit (X_n) une suite de v.a. i.i.d. de distribution exponentielle standard, i.e.

$$P(X_n > x) = e^{-x}, \quad x \geq 0$$

Posons

$$M_n = \max_{1\leq i\leq n} X_i$$

Alors, la distribution de M_n est clairement donnée par

$$P(M_n \leq x) = \prod_{i=1}^{n} P(X_i \leq x) = (1 - e^{-x})^n$$

Montrons que

$$\frac{M_n}{\log n} \xrightarrow[n\to\infty]{\text{p.s.}} 1 \tag{4.3}$$

Soit $\varepsilon \in (0,1)$ et observons en premier que

$$P(M_n \leq (1-\varepsilon)\log n) = (1 - n^{\varepsilon-1})^n = e^{n\log(1-n^{\varepsilon-1})} = e^{-n^{\varepsilon}(1+o(1))}$$

Par conséquent, nous obtenons que

$$\sum P(M_n \leq (1-\varepsilon)\log n) < \infty$$

D'où, par (LBC1), il vient que

$$\liminf \frac{M_n}{\log n} \geq 1 - \varepsilon, \quad \text{p.s.} \tag{4.4}$$

Inversement, nous allons procéder par un *argument de bloc*. Remarquons que

$$P(M_n > (1+\varepsilon)\log n) = 1 - P(M_n \leq (1+\varepsilon)\log n) = 1 - e^{-n^{-\varepsilon}(1+o(1))} = n^{-\varepsilon}(1+o(1))$$

Considérons alors ces probabilités sur la suite $n_k = [k^{\delta}]$ avec $\delta\varepsilon > 1$. Dans ce cas, nous avons

$$\sum P(M_{n_k} > (1+\varepsilon)\log n_k) < \infty$$

Ce qui, par (LBC1), montre que

$$\limsup \frac{M_{n_k}}{\log n_k} \leq 1 + \varepsilon, \quad \text{p.s.}$$

Il reste à s'assurer que cela est aussi vrai entre les n_k. En effet, comme M_n est croissante, nous avons pour $n \in [n_k, n_{k+1}]$ que

$$\frac{M_n}{\log n} \leq \frac{M_{n_{k+1}}}{\log n_{k+1}} \frac{\log n_{k+1}}{\log n_k} \xrightarrow[k\to\infty]{\text{p.s.}} 1 + \varepsilon$$

De ce fait, nous déduisons que

$$\limsup \frac{M_n}{\log n} \leq 1 + \varepsilon, \quad \text{p.s.} \tag{4.5}$$

Ainsi, comme ε est arbitraire, (4.4) et (4.5) donnent la limite recherchée. ○

4.2 Convergence en probabilité

Nous pouvons alléger la condition de la convergence p.s. dans (4.1) et définir une version **faible** de convergence de suite de v.a. : on dit que X_n converge vers X **en probabilité** si

$$\forall \varepsilon > 0 : \quad \lim_n P(|X_n - X| \geq \varepsilon) = 0 \tag{4.6}$$

Ce que l'on note par

$$X_n \xrightarrow[n\to\infty]{P} X$$

Ce n'est rien d'autre que la *convergence en mesure* en théorie de la mesure. Ainsi, par définition, la convergence p.s. implique la convergence en probabilité.

<u>Exemple</u> **4.6** (**Loi des grands nombres**) Nous allons voir ici un phénomène important en théorie des probabilités que nous développerons plus tard. Soit (X_n) une suite de v.a. dans L^2 non corrélées en ce sens que

$$\forall i \neq j : \quad E(X_i X_j) = E(X_i)E(X_j)$$

Supposons que $E(X_i) = 0$ et $V(X_i) = \sigma^2$. Alors, il vient que

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{P} 0$$

En effet, par (IT) et à cause de la non corrélation des X_i, nous avons

$$P(|X_1 + \cdots + X_n| > n\varepsilon) \leq \frac{V(\sum_{i=1}^n X_i)}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

Pour se convaincre que cette convergence a un sens tout à fait correct, il suffit de jeter une pièce de monnaie non truquée un bonne cinquantaine de fois pour constater que la proportion de piles se stabilise vers $1/2$, si tant est que l'on soit pas trop malchanceux ! ○

Propriétés

Intéressons-nous aux propriétés de la convergence en probabilité. Par définition, la convergence en probabilité est plus faible que la convergence p.s. Cependant, celle-là, par une application typique de LBC1, peut être caractérisée par cette dernière par une version *probabiliste* d'un résultat fondamental en topologie :

Théorème 4.7 *Une suite de v.a. (X_n) converge en probabilité vers X ssi de toute sous-suite de (X_n), on peut extraire une sous-suite qui converge p.s. vers X.*

La démonstration de ce résultat repose donc sur le résultat suivant de la topologie générale.

Lemme 4.8 *Soit (x_n) une suite d'éléments d'une espace topologique. Si toute sous-suite $(x_{n(k)})$ possède une sous-suite $(x_{n(k_\ell)})$ qui converge vers x, alors (x_n) converge vers x.*

Le Théorème 4.7 entraîne des conséquences importantes. En premier, nous avons :

Théorème 4.9 *Soit (X_n) une suite de v.a. et f une fonction réelle continue. Supposons que $X_n \to X$ en probabilité. Alors, on a*

$$f(X_n) \xrightarrow[n\to\infty]{P} f(X) \tag{4.7}$$

De plus, si f est bornée, alors on a

$$\lim E(f(X_n)) = E(f(X)) \tag{4.8}$$

Théorème 4.10 *Si $X_n \xrightarrow[n\to\infty]{P} X, Y_n \xrightarrow[n\to\infty]{P} Y$, on a que*

$$(i)\, X_n + Y_n \xrightarrow[n\to\infty]{P} X + Y$$
$$(ii)\, X_n Y_n \xrightarrow[n\to\infty]{P} XY$$

2ème lemme de Borel-Cantelli

Nous allons voir ici une version opposée au (LBC1) en ce sens que pour des évènements indépendants, la condition $P(\limsup_n A_n) > 0$ est suffisante pour $P(\limsup_n A_n) = 1$. C'est là un résultat incontournable en théorie des probabilité :

Théorème 4.11 (Second lemme de Borel-Cantelli) *Soit (A_n) est une suite d'évènements aléatoires indépendants. Alors, on a*

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Longrightarrow P(\limsup A_n) = 1 \tag{LBC2}$$

Maintenant, grâce au (LBC2), nous pouvons vérifier que la convergence en probabilité n'implique généralement pas la convergence p.s. comme le montre le (contre) exemple suivant.

<u>EXEMPLE</u> **4.12** Soit (X_n) une suite de v.a. i.i.d. définies par

$$P(X_n = 1) = \frac{1}{n} = 1 - P(X_n = 0)$$

Il est facile de voir que

$$\sum_n P(X_n = 1) = \infty$$

Il en résulte alors par (LBC2) que

$$\limsup X_n = 1, \quad \text{p.s.}$$

Autrement dit, p.s., une infinité de X_n est égale à 1. Or, pour tout $\varepsilon > 0$,

$$P(|X_n| \geq \varepsilon) = \frac{1}{n} \xrightarrow[n\to\infty]{} 0$$

Par conséquent, nous obtenons que

$$X_n \xrightarrow[n\to\infty]{P} 0$$

○

4.3 Convergence dans L^p

Nous allons reprendre ici la notion de convergence dans l'espace L^p dans le contexte particulier d'un espace probabilisé. Considérons $p \in [1, \infty]$. Alors, on dit qu'une suite (X_n) est convergente dans l'espace L^p vers X si

$$\lim_{n\to\infty} \|X_n - X\|_p = 0$$

autrement dit, si

$$\lim_n E(|X_n - X|^p) = 0$$

On écrit alors

$$X_n \xrightarrow[n\to 0]{L^p} X$$

Nous rappelons que L^p est un espace de Banach. Ainsi, si $X_1, X_2, \ldots,$ est une suite de v.a. dans L^p satisfaisant le critère de Cauchy

$$\lim_{n,m\to\infty} \|X_n - X_m\|_p = 0,$$

alors celle-ci converge dans L^p vers une v.a. X.

La convergence dans L^p est considérée comme **convergence forte**. On peut déjà voir qu'elle implique la convergence en probabilité :

Théorème 4.13 *Dans L^p, on a*

$$X_n \xrightarrow[n\to 0]{L^p} X \Longrightarrow X_n \xrightarrow[n\to 0]{P} X$$

Les exemples suivants montrent toutefois qu'en général, la convergence en probabilité, ou même presque sûre, n'implique pas la convergence dans L^p.

EXEMPLE **4.14** Soit $\Omega = (0, 1]$ muni de la tribu de Borel et soit P la loi uniforme sur Ω, autrement dit, la mesure de Lebesgue sur Ω. Soit $\alpha > 0$ et définissons sur Ω,

$$X_n(\omega) = \omega^{-\alpha} \mathbb{1}_{(1/n,2/n]}(\omega)$$

Remarquons que pour tout $\varepsilon > 0$, on a

$$P(|X_n| \geq \varepsilon) = 1/n$$

De ce fait, il vient que

$$X_n \xrightarrow[n\to\infty]{P} X$$

Cependant, on a aussi pour $\alpha p < 1$,

$$E(X_n^p) = \int_{1/n}^{2/n} \omega^{-\alpha p} \, \mathrm{d}P(\omega) = \left(\frac{2^{1-\alpha p} - 1}{1 - \alpha p} \right) n^{1-\alpha p} \xrightarrow[n\to\infty]{} \infty$$

○

EXEMPLE **4.15** Soit X_n une v.a. de distribution $(1 - n^{-p})\delta_0 + n^{-p}\delta_n$, où $p > 1$. Pour tout $\varepsilon > 0$ et n assez grand ($n > \varepsilon$), on a $P(X_n \geq \varepsilon) = n^{-p}$. Donc, il vient par (LBC1) que

$$X_n \xrightarrow[n\to\infty]{\text{p.s.}} 0$$

Cependant, on a aussi $E(|X_n|^p) = 1$, autrement dit (X_n) ne converge pas dans L^p.

○

Étudions la question inverse : sous quelle condition la convergence en probabilité implique la convergence dans L^p. Pour ce faire, on définit une collection de v.a. $(X_i)_{i\in I}$ **uniformément intégrable** par la propriété suivante

$$\lim_{M\to\infty} \sup_{i\in I} E(|X_i|; |X_i| \geq M) = 0 \tag{4.9}$$

En vertu du (TCD), un exemple élémentaire d'une famille de v.a. uniformément intégrable est une collection de v.a. dominées par une v.a. intégrable, i.e. $|X_i| \leq Y$ et $E(Y) < \infty$. Et une des techniques connues pour montrer l'intégrabilité uniforme d'une famille de v.a. est donnée dans l'exercice suivant.

Exercice 4.16 Soit φ une fonction telle que $\varphi(x)/x \longrightarrow 0$ quand $x \to \infty$, par exemple, $\varphi(x) = x^p$ pour $p > 1$. Alors, si $\sup_{i\in I} E(\varphi(|X_i|)) < \infty$, la famille $(X_i)_{i\in I}$ est uniformément intégrable.

L'ingrédient qui permet à la convergence en probabilité de passer à la convergence dans L^1 est exprimé dans le résultat suivant.

Théorème 4.17 *Si $X_n \xrightarrow{P} X$, alors les propositions suivantes sont équivalentes :*

$$(i)\ (X_n)\ \textit{est uniformément intégrable}$$
$$(ii)\ X_n \xrightarrow[n\to\infty]{L^1} X$$
$$(iii)\ E(|X_n|) \longrightarrow E(|X|) < \infty$$

Nous avons alors le résultat suivant qui donne une condition suffisante pour avoir la convergence dans L^p si nous avons déjà la converge en probabilité.

Proposition 4.18 *Soit $X_n \xrightarrow{P} X$. Supposons que $\sup_n E(|X_n|^q) < \infty$, pour un $q > 1$. Alors, pour tout $p < q$, on a*

$$X_n \xrightarrow[n\to\infty]{L^p} X$$

4.4 Convergence en loi

La présente section présente un quatrième mode de convergence, sans doute le plus important, qui apparaît dans nombre de résultats importants notamment le phénomène important dit *théorème central limite*.

Définition et exemples

Soit (F_n) suite de f.r. On dit que (F_n) **converge faiblement** vers une f.r. F si $F_n(x)$ converge vers $F(x)$ pour tout point de continuité x de F, et on écrit

$$F_n \Longrightarrow F \tag{4.10}$$

Si μ_n et μ sont les mesures de probabilité (de Lebesgue-Stieltjes), associées respectivement aux f.r. F_n et F, on écrit aussi

$$\mu_n \Longrightarrow \mu$$

De plus, si les F_n sont les f.r. des v.a. X_n respectives et que F est celle d'une v.a. X, alors on dit que la suite (X_n) est **converger en loi** ou **en distribution** vers X ssi $F_n \Longrightarrow F$, et on note

$$X_n \Longrightarrow X$$

Notons que la convergence aux points de continuité est suffisante pour déterminer la limite car la f.r. est continue à droite avec des limites à gauche et par la Proposition 1.31, elle admet au plus un nombre dénombrable de point de discontinuité.

<u>Exemple</u> **4.19** Soit X une v.a. de f.r. F. Montrons que

$$X - \frac{1}{n} \Longrightarrow X \Longleftarrow X + \frac{1}{n}$$

En effet, pour $n \in \mathbb{N}^*$, la v.a. $X + 1/n$ possède la distribution

$$F_n(x) = F(x - 1/n)$$

D'où, il vient que

$$F_n(x) \xrightarrow[n\to\infty]{} F(x_-)$$

Ce qui donne clairement que $X + 1/n \Longrightarrow X$. A fortiori, $X - 1/n \Longrightarrow X$, vu que F est continue à droite. ○

<u>Exemple</u> **4.20** (**Attendre un évènement rare**) Soit X la v.a. qui compte le nombre de répétitions indépendantes nécessaires pour avoir un succès d'une expérience de Bernoulli de paramètre $1/n$, autrement dit $X \sim \mathcal{G}(1/n)$. Alors, pour tout $k \geq 0$, nous avons

$$P(X > k) = (1 - 1/n)^k$$

Ce qui donne pour tout $x \geq 0$,

$$P(n^{-1}X > x) \xrightarrow[n\to\infty]{} e^{-x}$$

Ainsi, nous avons

$$\frac{X}{n} \Longrightarrow \mathcal{E}$$

○

<u>Exemple</u> **4.21** Reprenons l'Exemple 4.5 où nous avons appris que le maximum de n v.a. i.i.d. de distribution exponentielle se comporte comme $\log n$, i.e.

$$M_n = \max_{1\leq i\leq n} X_i = \log n + o(\log n)$$

Ici, nous allons connaître la vitesse de convergence de cette suite, autrement dit le contenu du terme $o(\log n)$. Posons $Z_n = M_n - \log n$ et remarquons que pour tout x, nous avons

$$\begin{aligned} F_{Z_n}(x) &= P(M_n \leq \log n + x) && (4.11)\\ &= \left(1 - e^{-x-\log n}\right)^n && (4.12)\\ &= \exp\left(n \log\left(1 - e^{-x}/n\right)\right) && (4.13)\\ &= e^{-e^{-x}+o(1)} && (4.14) \end{aligned}$$

D'où, Z_n converge en loi vers une v.a. de f.r. $e^{-e^{-x}}$. ○

Exercice 4.22 Soit F une f.r. quelconque. Montrer qu'il existe une suite de f.r. (F_n) définie par des *familles triangulaire* $(x_{ni})_{1\leq i\leq n}$ associées aux familles de probabilités (p_{ni}), i.e.

$$F_n(x) = \sum_{i=1}^{n} p_{ni} \mathbb{1}_{[x_{ni},\infty)}(x) \tag{4.15}$$

telle que

$$F_n \Longrightarrow F$$

Indication. Construire des fonctions étagées de la forme (4.15) qui converge vers F en tout point de continuité.

Propriétés fondamentales

Commençons par comparer la convergence en distribution avec les autres modes et voir en fait que c'est la plus faible d'entre elles, néanmoins la plus importante, peut-être.

Le résultat suivant compare les deux convergences dites *faibles*, en loi et en probabilité, ce qui donne du coup des comparaisons avec les convergences *fortes*, p.s. et en L^p.

Théorème 4.23 *On a*

$$X_n \xrightarrow[n\to\infty]{P} X \Longrightarrow X_n \Longrightarrow X$$

Remarque 4.24 *Le présent résultat et les Théorèmes 4.13–4.31 donnent alors le schéma comparatif suivant*

$$\begin{array}{ccccc} X_n \xrightarrow[n\to 0]{p.s.} X & \Longrightarrow & X_n \xrightarrow[n\to 0]{P} X & \Longleftarrow & X_n \xrightarrow[n\to 0]{L^p} X \\ & & \Downarrow & & \\ & & X_n \xrightarrow[n\to 0]{\mathcal{L}} X & & \end{array}$$

Revenons maintenant à la caractérisation théorique de la convergence en loi. Nous avons en premier :

Théorème 4.25 *Soit (X_n) une suite de v.a. Alors, (X_n) converge en distribution vers une v.a. X ssi pour toute fonction f continue et bornée sur $\mathbb{R}$, on a*

$$E(f(X_n)) \longrightarrow E(f(X)) \tag{4.16}$$

Dans le cas particulier où les distributions de probabilité en jeu admettent des densités, le Théorème 4.25 implique

Théorème 4.26 (Lemme de Sheffé) *Soit (X_n) des v.a. admettant des densités de probabilité respectives (f_n). Supposons qu'il existe une v.a. X de densité de probabilité f telle que pour tout x,*

$$f_n(x) \xrightarrow[n\to\infty]{} f(x)$$

Alors, on a

$$X_n \Longrightarrow X$$

Remarque 4.27 *Dans la preuve du Lemme de Sheffé, si nous appelons μ_n et μ les distributions respectives de X_n et X, nous observons que nous avons alors montré implicitement que*

$$\sup_B |\mu_n(B) - \mu(B)| =: \|\mu_n - \mu\| \Longrightarrow 0$$

C'est là en fait un résultat plus fort que la convergence en loi où la quantité $\|\mu_n - \mu\|$ s'appelle la ***norme de variation totale****.*

La convergence en loi peut être exprimée à l'aide de distances. En effet, si F et G sont deux f.r., définissons

$$\rho(F,G) = \inf\{\varepsilon :\ F(x-\varepsilon) - \varepsilon \le G(x) \le F(x+\varepsilon) + \varepsilon\} \tag{4.17}$$

On montre alors (faire en exercice) que cette quantité définit bien une distance dans l'espace des fonctions de répartition, appelée **distance de Lévy**, i.e.

$$\begin{aligned} &(i)\, d_L(F,G) = 0 \Longleftrightarrow F = G \\ &(ii)\, d_L(F,G) = d_L(F,G), \\ &(iii)\, d_L(F,G) \le d_L(F,H) + d_L(H,G) \end{aligned}$$

Laquelle vérifie

$$d_L(F_n, F) \longrightarrow 0 \Longleftrightarrow F_n \Longrightarrow F \tag{4.18}$$

Ce fait entraîne le résultat suivant.

Théorème 4.28 *Si toute sous-suite de* (X_n) *admet une autre sous-suite qui converge en loi vers* X*, alors* $X_n \Longrightarrow X$*.*

D'autre par, nous avons

Théorème 4.29 (Théorème de Sélection de Helly) *Pour toute suite de f.r.* (F_n)*, il existe une sous-suite* $(F_{n(k)})$ *et une fonction non décroissante et continue à droite* F *telles que* $F_{n(k)} \to F$ *en tout point de continuité de* F*.*

Remarque 4.30 *La limite dans le Théorème 4.29 peut ne pas être une f.r. En effet, prenons l'exemple suivant :*

$$F_n(x) = a\mathbb{1}_{\{x \geq n\}}(x) + b\mathbb{1}_{\{x \geq -n\}}(x) + c\,G(x)$$

où $a + b + c = 1$ *et* G *est une f.r. On a alors* $F_n(x) \to F(x) = b + c\,G(x)$ *et*

$$\lim_{x \to -\infty} F(x) = b, \qquad \lim_{x \to \infty} F(x) = b + c = 1 - a$$

Autrement dit, il y a une masse a *qui s'échappe à* $-\infty$ *et une masse* b *qui s'échappe à* $+\infty$*.*

Enfin, notons qu'entre deux convergences faibles, la convergence en probabilité est plus forte que la convergence en loi :

Théorème 4.31 *On a*

$$X_n \xrightarrow[n \to \infty]{P} X \Longrightarrow X_n \Longrightarrow X$$

Suite tendue

Nous pouvons aussi connaître la converge en loi des suites de v.a. en étudiant une propriété essentielle associée à leurs f.r. Ceci est exprimé dans le résultat suivant.

Théorème 4.32 (Suite tendue) *Soit* (F_n) *une suite de f.r. Toute limite d'une sous-suite de* (F_n) *est une f.r. ssi* (F_n) *est* ***tendue****, i.e.*

$$\forall \varepsilon > 0, \exists M > 0 : \limsup_n 1 - F_n(M) + F_n(-M) \leq \varepsilon \tag{4.19}$$

Le théorème suivant donne une condition suffisante pour avoir des suites de f.r. tendues.

Théorème 4.33 *S'il existe une fonction $\phi \geq 0$ telle que $\phi(x) \to \infty$ quand $|x| \to \infty$ et que*

$$\sup_n \int \phi(x)\,\mathrm{d}F_n(x) < \infty \tag{4.20}$$

alors (F_n) est tendue.

Convergence des f.c.

La convergence en loi des v.a. peut être caractérisée par les f.c., ce que nous allons voir dans le théorème suivant. Nous verrons au chapitre suivant une application importante de ce résultat.

De manière équivalente à la tension des f.r. (cf. (4.19)), on dit qu'une suite (μ_n) est **tendue** si pour tout $\varepsilon > 0$, il existe un $M > 0$ tel que

$$\liminf_n \mu_n([-M, M]) \geq 1 - \varepsilon \tag{4.21}$$

Théorème 4.34 (Théorème de Paul Lévy) *Soit (μ_n) une suite de mesures de probabilité de f.c. associées φ_n. Alors, on a que*

- (i) *si (μ_n) converge faiblement vers une distribution μ, (φ_n) converge simplement sur tout $\mathbb{R}$ vers φ_μ.*
- (ii) *si (φ_n) converge simplement sur $\mathbb{R}$ vers une fonction φ continue en 0, (μ_n) est tendue et converge faiblement vers une mesure de probabilité dont la f.c. est φ.*

Remarque 4.35 *Pour voir la nécessité de la continuité de φ en 0 dans l'assertion (ii), soit μ_n la distribution normale de moyenne 0 et de variance n. On a $\varphi_n(t) = e^{-nt^2/2}$, ce qui donne que $\varphi_n(t) \to 0$ quand $t \neq 0$ et $\varphi_n(0) = 1$ pour tout n. Cependant, les mesures μ_n ne convergent vers aucune distribution vu que*

$$\forall x \in \mathbb{R} : \; \mu_n((-\infty, x]) \xrightarrow[n\to\infty]{} 1/2$$

<u>Exemple</u> **4.36 (Une binomiale vers une Poisson)** Soit une suite de v.a. (X_n) telle que $X_n \sim \mathcal{B}(n, p_n)$. Supposons que $np_n \to \lambda \in (0, \infty)$. Alors,

$$X_n \Longrightarrow \mathcal{P}(\lambda)$$

En effet, nous avons

$$\phi_{X_n}(t) = \big(1 + p_n(e^{it} - 1)\big)^n \xrightarrow[n\to\infty]{} e^{\lambda(e^{it}-1)}$$

où nous avons utilisé la limite usuelle pour les suites de nombres complexes

$$z_n \to z \Longrightarrow (1+n^{-1}z_n)^n \xrightarrow[n\to\infty]{} e^z$$

Or, $e^{\lambda(e^{it}-1)}$ est la f.c. de $\mathcal{P}(\lambda)$. Ainsi, nous avons bien par le Théorème 4.34 la convergence en loi en question. ○

<u>Exemple</u> **4.37** (**Une Poisson vers une Normale**) Considérons des v.a. X_n de distributions respectives $\mathcal{P}(\lambda_n)$. Si $\lambda_n \to \infty$, alors

$$\frac{X_n-\lambda_n}{\sqrt{\lambda_n}} \Longrightarrow \mathcal{N}(0,1)$$

En effet, soit φ_n la f.c. de $(X_n-\lambda_n)/\sqrt{\lambda_n}$. Nous connaissons déjà la f.c. d'une loi de Poisson (voir Exemple 3.1) et en utilisant les propriétés de base des f.c., nous obtenons que

$$\begin{aligned}\phi_n(t) &= e^{-it\sqrt{\lambda_n}}\exp\left(\lambda_n(e^{it/\sqrt{\lambda_n}}-1)\right)\\ &= \exp\left(\lambda_n\left(e^{it/\sqrt{\lambda_n}}-1-it/\sqrt{\lambda_n}\right)\right)\end{aligned}$$

Maintenant, observons que le développement de e^{ix} donne

$$\left|e^{ix}-1-ix-\frac{(ix)^2}{2}\right| \le \frac{|x|^3}{3!}$$

De ce fait,

$$\lambda_n\left(e^{it/\sqrt{\lambda_n}}-1-it/\sqrt{\lambda_n}\right) \xrightarrow[n\to\infty]{} e^{-t^2/2}$$

○

Exercice 4.38 (Convergence uniforme des f.c.) Soit (μ_n) une famille de mesures (de probabilités) que l'on suppose tendue, i.e.

$$\sup_n \mu_n\left(\mathbb{R}\setminus[-M,M]\right) \xrightarrow[M\to\infty]{} 0$$

(a) Montrer en utilisant la continuité uniforme des f.c. (cf. Théorème 3.6) que les φ_n sont équicontinues, i.e.

$$\forall\varepsilon>0, \exists\delta>0 : |h|<\delta \Longrightarrow |\varphi_n(t+h)-\varphi_n(t)|<\varepsilon \tag{4.22}$$

(b) Supposer que $\mu_n \Longrightarrow \mu$. Utiliser le Théorème 4.34 et l'équicontinuité pour montrer directement que $\varphi_n \longrightarrow \varphi$ uniformément sur tout ensemble compact. Et donner un exemple que la convergence uniforme sur tout $\mathbb{R}$ n'est pas nécessaire.

Convergence des f.g.

Nous terminons avec le cas des v.a. dans $\mathbb{N}$ pour lesquelles il est plus naturel de considérer la fonction génératrice définie dans (3.17). En vertu du Théorème 4.25 sur la convergence en loi des v.a., on peut déduire facilement qu'une suite (X_n) de v.a. dans $\mathbb{N}$ converge en loi vers une v.a. X à valeurs dans $\mathbb{N}$ ssi

$$\forall k \in \mathbb{N} : \lim_n P(X_n = k) = P(X = k) \tag{4.23}$$

Ce résultat peut être exprimé en vertu du Théorème de Paul Lévy par les fonctions génératrices comme suit.

Théorème 4.39 *Si (X_n) est une suite de v.a. dans $\mathbb{N}$ convergeant en loi vers X, alors, uniformément dans le disque unité $\{z : |z| \leq 1\}$, on a*

$$\lim_{n\to\infty} G_{X_n}(z) = G_X(z) \tag{4.24}$$

Réciproquement, si $G_{X_n}(z) \longrightarrow G(z)$ pour tout z tel que $|z| \leq 1$, G étant une fonction continue au point $z = 1$, alors G est la fonction génératrice d'une v.a. X et $X_n \Longrightarrow X$.

<u>Exemple</u> **4.40** (**Une binomiale vers une Poisson**) Comme application de ce dernier résultat, reprenons l'Exemple 4.36 de la convergence d'une binomiale vers une Poisson et considérons la loi binomiale de paramètres $(n, \lambda/n)$. La f.g. de celle-ci est

$$G(z) = (1 + \lambda(z-1)/n)^n$$

Nous avons

$$\lim_n G(z) = e^{\lambda(z-1)}$$

ce qui est la f.g. de la loi de $\mathcal{P}(\lambda)$ (cf. Exemples 3.31). Ainsi, on a

$$\mathcal{B}(n, \lambda/n) \Longrightarrow \mathcal{P}(\lambda)$$

○

5 LGN & TCL

Si (X_n) est une suite de v.a. i.i.d., nous sommes intéressés ici par la convergence de sa moyenne empirique

$$\frac{X_1 + \cdots + X_n}{n}$$

Ceci est d'autant plus intéressante que ladite moyenne définit statistiquement une valeur approximative de la moyenne des X_n. Cette étude constitue l'objet principal premier de ce chapitre et aboutit à deux résultats importants, à savoir, les lois des grands nombres (LGN), dites faible et forte.

Ensuite, nous nous intéresserons au même objet recentrée et rééchelonnée par $\sqrt{n}$, une opération que l'on appelle normalisation. Il s'avère que si l'on choisit les v.a. sommées dans L^2, la quantité en question se comporte asymptotiquement comme une loi normale, ce que l'on appelle Théorème Central Limite ou simplement (TCL).

5.1 Lois des grands nombres

Le premier résultat principal que nous voulons montrer ici est le suivant :

Théorème 5.1 (Loi faible des grands nombres) *Soit (X_n) une suite de v.a. i.i.d. telles que $E(|X_n|) < \infty$. Posons $\mu = E(X_n)$. Alors, on a*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{P} \mu \tag{LGN}$$

<u>Exemple</u> **5.2** (**Approximation polynomiale**) Nous allons voir ou plutôt revoir ici (cf. Exemple 1.58) qu'il est toujours possible d'approximer une fonction continue

par des polynômes. En effet, soit f une fonction continue sur $[0,1]$ et rappelons le polynôme de Bernstein de degré n associé à f :

$$f_n(x) = \sum_{k=0}^{n} f(k/n)\, C_n^k x^k (1-x)^{n-k}$$

Alors, nous avons grâce à (LGN) que

$$\lim_{n\to\infty} f_n(x) = f(x) \tag{5.1}$$

En fait, nous avons vu à l'Exemple 1.58 que cette convergence est uniforme.

Pour montrons (5.1), considérons une suite (X_n) de v.a. de Bernoulli indépendantes de même loi $\mathcal{B}(x)$, i.e.

$$P(X_n = 1) = x = 1 - P(X_n = 0)$$

Posons $S_n = X_1 + \cdots + X_n$. Alors, en vertu de (LGN), nous avons que

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{P} x$$

En même temps, remarquons que

$$E(f(S_n/n)) = f_n(x)$$

Alors, par le Théorème 4.9, il résulte que

$$f_n(x) = E(f(S_n/n)) \xrightarrow[n\to\infty]{} f(x)$$

○

En second, nous avons le résultat classique de la théorie des probabilités qui améliore le premier sous les mêmes conditions :

Théorème 5.3 (Loi forte des grands nombres) *Soit $(X_n)_{n\geq 1}$ une suite de v.a. identiquement distribuées, de moyenne finie notée μ et indépendantes deux à deux. Alors, on a*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{p.s.} \mu \tag{LFGN}$$

<u>Exemple</u> **5.4** (**Fonction de distribution empirique**) Soit $(X_n)_{n\geq 1}$ une suite de v.a. i.i.d. de f.r. F. Définissons la fonction (nelle) aléatoire

$$F_n(x) = n^{-1} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} \tag{5.2}$$

laquelle donne la fréquence des valeurs inférieures à x observées, d'où on l'appelle *fonction de distribution empirique.* Alors, on a le **Théorème de Glivenko-Cantelli** : uniformément sur x, on a

$$F_n(x) \xrightarrow[n\to\infty]{p.s.} F(x) \tag{5.3}$$

○

5.2 Théorème central limite

Commençons dans cette section par connaître le résultat élémentaire du phénomène en question, à savoir le TCL pour une suite de v.a. de Bernoulli. Nous rappelons que $\mathcal{N}$ représente la loi normale centrée réduite et par abus de notation, cela sert à représenter aussi une v.a. qui suit cette loi.

Théorème 5.5 (De Moivre-Laplace) *Soit $X_1, X_2, \ldots$, des v.a. i.i.d. de distribution de Bernoulli telles que*

$$P(X_i = -1) = P(X_i = 1) = 1/2$$

Soit $S_n = X_1 + \cdots + X_n$. Alors, on a

$$\frac{S_n}{\sqrt{n}} \Longrightarrow \mathcal{N} \tag{5.4}$$

La démonstration de ce résultat se révèle principalement être un calcul d'approximation du factoriel, laquelle demande en premier un résultat sur le comportement local du processus (S_n). Nous rappelons que la notation $u_n \sim v_n$ signifie que $u_n/v_n \to 1$.

Théorème 5.6 *Si $2k/\sqrt{2n} \to x$, alors*

$$P(S_{2n} = 2k) = C_{2n}^{n+k}\, 2^{-2n} \sim \frac{e^{-x^2/2}}{\sqrt{n\pi}} \tag{5.5}$$

Remarque 5.7 *La formule (5.5) peut être exprimée sous la forme*

$$\frac{\sqrt{2n}}{2} P\left(\frac{S_{2n}}{\sqrt{2n}} = \frac{2k}{\sqrt{2n}}\right) \xrightarrow[n\to\infty]{} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \tag{5.6}$$

Ceci est en fait la version basique de ce que l'on appelle <u>*théorème central limite local*</u> *que nous verrons plus généralement dans une section plus bas.*

<u>Exemple</u> **5.8** (**Statistique d'ordre centrale**) Soit $U_1, \ldots, U_{2n+1}$ des v.a. i.i.d. de distribution uniforme sur $[0,1]$. Ordonnons les U_i de manière croissante et appelons V_{n+1} la $n+1$-valeur. Alors, on a

$$X_n := \frac{V_{n+1} - 1/2}{\sqrt{n/2}} \Longrightarrow \mathcal{N} \tag{5.7}$$

Pour montrer ça, rappelons d'abord de l'Exemple 2.18 que V_{n+1} possède la densité

$$(2n+1)C_{2n}^n u^n (1-u)^n \, \mathbb{1}_{[0,1]}(u)$$

D'où, par un changement de variable simple,

$$u = \frac{1}{2} + \frac{x}{2\sqrt{2n}}, \quad x \leq \sqrt{2n}$$

il vient que la densité de X_n est donnée par

$$\begin{aligned} f_n(x) &= (2n+1)C_{2n}^n \left(\frac{1}{2} + \frac{x}{2\sqrt{2n}}\right)^n \left(\frac{1}{2} - \frac{x}{2\sqrt{2n}}\right)^n \frac{1}{2\sqrt{2n}} \\ &= C_{2n}^n \, 2^{-2n} \left(1 - \frac{x^2}{2n}\right)^n \frac{2n+1}{2n} \sqrt{\frac{n}{2}} \end{aligned}$$

où nous remarquons que la première quantité est la probabilité $P(S_{2n} = 0)$ dans (5.5). Ainsi, nous obtenons facilement à la fin que

$$f_n(x) \xrightarrow[n\to\infty]{} \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

Il reste à appliquer le Lemme de Sheffé (cf. Théorème 4.26) pour obtenir le résultat souhaité (5.7). ○

Maintenant, nous allons voir le TCL dans sa version la plus connue et qui généralise le résultat précédent, à savoir le Théorème de De Moivre-Laplace, en considérant une suite de v.a. i.i.d.

Théorème 5.9 (Théorème central limite) *Soit (X_n) une suite de v.a. i.i.d. de moyenne μ et de variances σ^2 finies. Posons $S_n = X_1 + \cdots + X_n$. Alors, on a*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Longrightarrow \mathcal{N} \tag{TCL}$$

La démonstration de ce résultat est basée essentiellement sur le Théorème 4.34 de continuité de Paul Levy et de la Remarque 3.17.

<u>Exemple</u> **5.10** (**Suite de Bernoulli**) Soit (X_n) une suite de v.a. i.i.d. de Bernoulli avec une probabilité de succès de $1/2$, i.e. $P(X_n = 0) = P(X_n = 1) = 1/2$. Si $X_n = 1$ désigne un succès, alors $S_n = X_1 + \cdots + X_n$ est la le nombre de succès observés durant n essais. Par exemple, n est le nombre d'apparition de pile durant les n premiers lancers d'une pièce de monnaie équilibrée. Nous avons

$$E(X_n) = \frac{1}{2}, \quad V(X_n) = \frac{1}{4}$$

En vertu du (TCL), nous avons

$$Y_n := \frac{S_n - n/2}{\sqrt{n/4}} \Longrightarrow \mathcal{N}$$

Et la table de la loi normale donne

$$P(|\mathcal{N}| \leq 2) \approx 0,9546$$

Ainsi, nous obtenons que

$$P(Y_n \in [-2, 2]) = P(S_n - n/2 \in [-\sqrt{n}, \sqrt{n}]) \approx 0,95$$

À titre d'illustration, si $n = 10000$, ceci nous apprend qu'avec 95 % de chances, le nombre de succès serait entre 4900 et 5100. ○

Références

[1] A. Bouziad et J. Calbrix (2007). *Théorie de la mesure et de l'intégration.* EDP Sciences.

[2] P. Barbe et M. Ledoux (2007). *Probabilité.* EDP Sciences.

[3] J-R. Barra (1971). *Notions fondamentales de statistique mathématique.* Dunod.

[4] Omar Boukhadra. (2021) *Théorie des probabilités.* Édition de l'Université de Constantine 1.

[5] Youcef B. Boukhadra. (1993) *Cours de probabilités et statistique* (SEM 340). Édition de l'Université de Constantine.

[6] G. Calot. (2010). *Cours de calculs des probabilités.* Cambridge University Press.

[7] R. Durrett. (1967). *Probability : Theory and Examples.* 4th edition. Cambridge University Press.

[8] M. Métivier. (1972). *Notions Fondamentales de la Théorie des Probabilités.* Dunod.

[9] J. Neveu. (1970) *Calcul des Probabilités.* Masson & Cie.

[10] Sheldon M. Ross and Erol A. Peköoz. (2007). *A second course in probability.* ProbabilityBookstore.com, Boston, MA.