PROBABILITÉS ÉLÉMENTAIRES

Édition 2012–2013

par
Anthony Phan

Département de Mathématiques
Boulevard Marie et Pierre Curie,
Téléport 2,
BP 30179, F-86962 Chasseneuil–Futuroscope cedex.

Les notes de cours qui suivent correspondent à un cours magistral de 22 heures en licence de Mathématiques avec pour orientation d'origine la préparation au Capes (L3 au département de Mathématiques de l'université de Poitiers). Ce cours est resté cependant préparatoire à des cours plus avancés de maîtrise où les convergences, les théorèmes limites et les notions de processus stochastiques peuvent être abordés. Le seul regret concerne la théorie de l'intégration selon Lebesgue qui n'est ici qu'éfleurée.

Ces notes sont complétées par un polycopié sur les lois de probabilité usuelles et diverses compilations d'exercices.

INTRODUCTION

A. **Quelques repères**

Des mots courants du Calcul des Probabilités sont « aléatoire » et « hasard ». Leurs origines sont assez semblables :

Alea [latin]. Dés, jeu de dés.

Az-Zahr [arabe]. Jeu de dés, chance.

On retient, parfois arbitrairement, quelques dates importantes :

1654. Échanges et correspondances entre le Chevalier Méré, Blaise Pascal (1623–1662) et Pierre de Fermat (1601–1665) autour de différents problèmes de jeux de dés ou de cartes.

1713. Jacques Bernoulli (1654–1705) publie, de manière posthume et édité par son neveu Nicolas Bernoulli (1687–1759) — et dont le père, frère de Jacques, s'appelait aussi Nicolas —, *Ars Conjectandi*, un premier traité de Dénombrement et de Calcul Combinatoire, notamment.

1812. Pierre-Simon Laplace (1749–1827) publie *Théorie analytique de probabilités*.

1900. Louis Bachelier (1870–1946) soutient sa thèse *Théorie de la spéculation* sous la direction d'Henri Poincaré.

1905. Albert Einstein (1879–1955) publie ses articles sur le mouvement brownien.

1933. Andrei Nikolaevich Kolmogorov (1903–1987) donne la fondation mathématique du Calcul des Probabilités.

Ce n'est qu'avec Kolmogorov que le Calcul des Probabilités sort réellement du simple dénombrement. Le formalisme qu'il propose provient de celui de l'intégration inventé par Henri Lebesgue (1875–1941) au début du XX^{e} siècle.

B. **Exemples d'expériences aléatoires**

Voici une petite liste d'expériences aléatoires très communes.

- On lance une pièce de monnaie et regarde le résultat obtenu.
- On lance un dé et regarde le résultat obtenu.
- On pioche sans remise n boules dans une urne qui en comporte $N \geqslant n$.
- On pioche avec remise n boules dans une urne qui en comporte au moins une.
- On suit la trajectoire d'une molécule de gaz dans une enceinte fermée.
- On suit un électron à ses passages au travers de différents détecteurs.
- On suit les recommandations de son conseiller financier.
- Etc.

Les aspects communs de ces différents exemples ne sautent pas nécessairement aux yeux. Il faut un formalisme général pour unifier ces situations.

Un premier formalisme aura été ébauché dans le secondaire, c'est ce qu'il convient d'appeler celui des *modèles probabilistes classiques*. L'aléa est appelé épreuve et modélisé par un élément

Ω d'un certain ensemble Ω appelé univers. Une collection d'épreuves, c'est-à–dire une partie de Ω, est un *événement*. La probabilité d'un événement $A \subset \Omega$ est

$$\mathbb{P}(A) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{\operatorname{Card} A}{\operatorname{Card} \Omega}, \dots$$

Ce cadre est rapidement trop limité (Ω infini par exemple) ou inadapté (épreuves non équiprobables).

C. **Expériences aléatoires, tentative de définition**

Considérons une *expérience* — ou plus précisément un protocole expérimental. Nous appelons *épreuve* toute réalisation de cette expérience, *événement* toute proposition portant sur le déroulement ou le résultat de cette expérience. Une épreuve peut ou non *réaliser* un événement donné.

Une expérience est *aléatoire* si :

— il existe au moins un événement qui peut être réalisé par certaines épreuves et pas par d'autres ;

— pour tout événement $\widehat{A}$, la fréquence de réalisation de $\widehat{A}$ au cours de n épreuves indépendantes converge vers un nombre réel compris entre 0 et 1 ne dépendant pas de la suite d'épreuves indépendantes choisie ; ce nombre est alors la *probabilité* de réalisation de l'événement $\widehat{A}$.

Remarque. — Le premier point stipule qu'une expérience aléatoire doit comporter un certain degré d'imprédictibilité. Le second point impose une condition de régularité qui est plus subtile. L'indépendance doit s'entendre pour l'instant dans son sens usuel : des épreuves sont indépendantes si aucune d'entre elles est influencée par les autres, ou les précédentes dans le cas d'une suite. La probabilité de réalisation d'un événement s'interprète alors comme un taux de réalisation *a priori* de cet événement avant la réalisation d'une épreuve particulière — qui elle réalisera, ou non, l'événement considéré.

Exemple. — Considérons pour expérience le lancer d'un dé à 6 faces équilibré. Une épreuve est un lancer effectif de ce dé. Des événements ayant trait à cette expérience sont par exemple : « obtenir 1 », « obtenir un nombre pair », « obtenir un nombre impair », « obtenir un nombre inférieur ou égal à 6 », « obtenir un nombre strictement supérieur à 6 ». Dans cet exemple, le second événement est l'événement *c*ontraire du troisième événement — et réciproquement — ; le quatrième événement est toujours réalisé, c'est un événement *certain* ; le dernier n'est jamais réalisé, c'est un événement impossible. Notons que si le premier événement est réalisé alors le troisième l'est aussi : le premier événement implique le troisième.

Étant donnée une expérience aléatoire, on appelle *variable aléatoire* toute quantité qu'on peut mesurer lors d'une réalisation de cette expérience, c'est-à-dire sur une épreuve. Les valeurs prises par une variable aléatoire peuvent dépendre de l'épreuve, ce sont les valeurs possibles de cette variable aléatoire, et chacune d'entre elles a une probabilité (éventuellement infinitésimale) d'être réalisée qui est celle de l'événement « la variable prend telle valeur ». La collection de ces probabilités est la *loi de la variable aléatoire*.

Exemple. — L'expérience considérée est toujours le lancer d'un dé à 6 faces équilibré. Nous ne nous intéressons qu'aux propositions portant sur le résultat du lancer. On convient alors qu'il s'agit là d'une expérience aléatoire : selon les lancers, on pourra voir apparaître 1, 2, …, ou encore 6 ; si on s'intéresse à la fréquence d'apparition d'un nombre donné dans

$\{1,\dots,6\}$ lors d'une suite de lancers indépendants, celle-ci doit tendre vers 1/6 puisque le dé est supposé équilibré ; chaque événement correspond ici à un sous-ensemble, éventuellement vide, de $\{1,\dots,6\}$, sa probabilité est alors égale à son cardinal divisé par 6. Une variable aléatoire qu'on définit naturellement est celle qui pour chaque épreuve prend pour valeur le résultat du lancer ; ses valeurs possibles sont $\{1,\dots,6\}$ et sa loi est la séquence indexée par $\{1,\dots,6\}$ constante, égale à 1/6 : la variable aléatoire considérée prend chacune des valeurs dans $\{1,\dots,6\}$ avec une probabilité égale à 1/6. D'autres variables aléatoires peuvent être définies, par exemple celle qui indique si le résultat est pair ou non et, dont les valeurs possibles sont {vrai, faux}, ou encore $\{1,0\}$, et dont la loi est la séquence indexée par l'un de ces ensembles, constante, égale à 1/2.

D. **Modélisation mathématique**

Nous donnons la correspondance entre le vocabulaire des expériences aléatoires et celui des modèles mathématiques — qui seuls nous intéresseront par la suite. Chacun des termes mathématiques nouveaux est défini précisément dans le cours.

Expérience aléatoire	Modèle mathématique
Collection des épreuves	Ensemble (mathématique) noté le plus souvent Ω et qualifié d'« ensemble fondamental », parfois d'« univers » ou encore simplement « grand Oméga »
Épreuve	Élément ω de Ω ($\omega \in \Omega$)
Collection des événements	Famille de parties de Ω, notée $\mathcal{A}$, possédant la structure de tribu (*voir* plus loin)
Événement $\widehat{A}$	$A \in \mathcal{A}$ une partie de Ω, ensemble des « épreuves » $\omega \in \Omega$ réalisant cet événement, $A = \{\omega \in \Omega : \widehat{A}(\omega) = \text{vrai}\}$
Opérations sur les événements	*Opérations sur les ensembles*
non $\widehat{A}$	$A^c = \complement_\Omega A = \{\omega \in \Omega : \widehat{A}(\omega) = \text{faux}\}$ (complémentaire de A dans Ω)
$\widehat{A}$ et $\widehat{B}$	$A \cap B = \{\omega \in \Omega : (\widehat{A}(\omega) \text{ et } \widehat{B}(\omega)) = \text{vrai}\}$ (épreuves réalisant $\widehat{A}$ et $\widehat{B}$)
$\widehat{A}$ ou $\widehat{B}$	$A \cup B = \{\omega \in \Omega : (\widehat{A}(\omega) \text{ ou } \widehat{B}(\omega)) = \text{vrai}\}$ (épreuves réalisant $\widehat{A}$ ou $\widehat{B}$)
$\widehat{A}$ et (non $\widehat{B}$)	$A \backslash B = A \cap (B^c) = \{\omega \in \Omega : (\widehat{A}(\omega) \text{ et non } \widehat{B}(\omega)) = \text{vrai}\}$ (A privé de B)
$\widehat{A} \Rightarrow \widehat{B}$ ((non $\widehat{A}$) ou $\widehat{B}$ = vrai) ..	$A \subset B$ ($A^c \cup B = \Omega$)
Probabilités des événements	*Mesure de probabilité* (*voir* plus loin) $\mathbb{P} : \mathcal{A} \to [0,1]$, $A \mapsto \mathbb{P}(A)$
Événement certain	$\Omega \in \mathcal{A}$, $\mathbb{P}(\Omega) = 1$
Événement impossible	$\varnothing \in \mathcal{A}$, $\mathbb{P}(\varnothing) = 0$

Événements incompatibles ..... ($\widehat{A}$ et $\widehat{B}$ = faux)	A, $B \in \mathcal{A}$, tels que $A \cap B = \varnothing$, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
Variables aléatoires	*Applications mesurables* (*voir* plus loin)
Variable aléatoire X ..... à valeurs dans un ensemble E	Application mesurable $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$, où $(E, \mathcal{E})$ est un espace mesurable (*voir* plus loin)
Loi d'une variable aléatoire X ...	Mesure de probabilité sur $(E, \mathcal{E})$ image par X de la mesure de probabilité $\mathbb{P}$

E. **Bibliographie sommaire**

Quelques livres *au hasard* :

1. Barbé (Ph.), Ledoux (M.), *Probabilité*, EDP Sciences (2007). De la licence à l'agrégation comme le disait le nom de la collection...

2. Foata (D.), Fuchs (A.), *Calcul des probabilités : Cours, exercices et problèmes corrigés*, Dunod (2003). Assez bien pour la licence.

3. Foata (D.), Fuchs (A.), *Processus stochastiques : Processus de Poisson, chaînes de Markov et martingales*, Dunod (2004). Peut-être trop orienté et pas assez généraliste.

4. Ouvrard (J.-Y.), *probabilités 1*, Cassini (1999).

5. Ouvrard (J.-Y.), *probabilités 2*, Cassini (2004). Assez technique et complet. Convient bien au cours 1m02.

6. Revuz (D.), Mesure et intégration, Hermann (1994). Un compagnon honnête pour l'intégration.

De nombreux autres ouvrages existent et sont très bien dans le cadre d'un L3 de Mathématiques. Il faut néanmoins garder à l'esprit que le cours est la première base à maîtriser avant de se lancer dans une bibliographie qui n'a parfois aucun effet bénéfique.

Chapitre premier

FORMALISME GÉNÉRAL

1. **Définitions générales**

Définition. — Un *espace mesurable* est un couple $(E, \mathcal{E})$ où :

(i) E est un ensemble ;

(ii) $\mathcal{E}$ est une tribu sur E, c'est-à-dire une classe de parties de E ($\mathcal{E} \subset \mathcal{P}(E)$) telle que

a) $\varnothing \in \mathcal{E}$ et $E \in \mathcal{E}$,

b) si $B \in \mathcal{E}$ alors $B^c \in \mathcal{E}$ (stabilité par passage au complémentaire),

c) si $(B_n)_n \subset \mathcal{E}$ est une suite finie ou dénombrable d'éléments de $\mathcal{E}$, alors $\bigcup_n B_n \in \mathcal{E}$ et $\bigcap_n B_n \in \mathcal{E}$ (stabilité par réunion et intersection finies ou dénombrables).

Les éléments de $\mathcal{E}$, qui sont des parties de E, sont appelés *sous-ensembles mesurables* de $(E, \mathcal{E})$.

Remarque. — Dans la définition d'une tribu ci-dessus certaines conditions sont redondantes : puisque $\varnothing^c = E$ et $E^c = \varnothing$ une seule condition suffit pour *a*) ; puisque $(\bigcup_n B_n)^c = \bigcap_n B_n^c$, une seule condition suffit pour *c*).

Définition. — Un *espace probabilisé* est un triplet $(E, \mathcal{E}, \mu)$ où :

(i) E est un ensemble non vide et $(E, \mathcal{E})$ est un espace mesurable ;

(ii) μ est une mesure de probabilité sur $(E, \mathcal{E})$, c'est-à-dire une application $\mu : \mathcal{E} \to [0, 1]$, $B \mapsto \mu(B)$ telle que

a) $\mu(\varnothing) = 0$ et $\mu(E) = 1$,

b) si $(B_n)_n \subset \mathcal{E}$ est suite finie ou dénombrable d'éléments de $\mathcal{E}$ deux à deux disjoints ($B_m \cap B_n = \varnothing$ dès que $m \neq n$) alors $\mu(\bigcup_n B_n) = \sum_n \mu(B_n)$ (additivité dénombrable).

Remarque. — La condition $\mu(\varnothing) = 0$ est une conséquence des autres conditions car $E = E \cup \varnothing$ et $E \cap \varnothing = \varnothing$ donc $1 = \mu(E) = \mu(E \cup \varnothing) = \mu(E) + \mu(\varnothing) = 1 + \mu(\varnothing)$, d'où $\mu(\varnothing) = 0$. Notamment, comme on a la condition $\mu(E) = 1$, il est nécessaire dans la définition précédente que $E \neq \varnothing$.

Définition. — Soient $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces mesurables. Une application $f : E \to F$ est *$\mathcal{E}/\mathcal{F}$-mesurable* si et seulement si :

$$\text{pour tout } C \in \mathcal{F}, \qquad f^{-1}(C) = \{x \in E : f(x) \in C\} \in \mathcal{E}.$$

On note alors $f : (E, \mathcal{E}) \to (F, \mathcal{F})$.

Lorsqu'il n'y a pas d'ambiguïté sur les tribus dont sont munis les ensembles E et F, on se contente de l'expression « f mesurable » sans préciser par rapport à quelles tribus.

Lorsqu'un espace probabilisé modélise une expérience ou un phénomène aléatoire, on le note le plus souvent $(\Omega, \mathcal{A}, \mathbb{P})$; les éléments de $\mathcal{A}$ sont appelés *événements* et les applications mesurables $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ sont appelées *variables aléatoires* ; on note alors $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$. Pour $B \in \mathcal{E}$, on a coutume de noter $\{X \in B\}$ le sous-ensemble (mesurable) de Ω égal à $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$.

Théorème et définition. — *Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ une variable aléatoire.*

(i) *On note $\sigma(X) = \{X^{-1}(B) : B \in \mathcal{E}\}$. C'est une sous-tribu de $\mathcal{A}$ appelée tribu des événements associés à X, ou tribu engendrée par X.*

(ii) *On définit $P_X : \mathcal{E} \to [0,1]$ par*

$$P_X(B) = \mathbb{P}\big(X^{-1}(B)\big) = \mathbb{P}\{\omega \in \Omega : X(\omega) \in B\} \overset{\text{notation}}{=} \mathbb{P}\{X \in B\}.$$

L'application P_X est une mesure de probabilité sur $(E, \mathcal{E})$ appelée loi (distribution, répartition) de la variable aléatoire X.

Démonstration. — (i) On a $\Omega = X^{-1}(E)$, donc $\Omega \in \sigma(X)$. Si $A = X^{-1}(B) \in \sigma(X)$ alors $A^c = (X^{-1}(B))^c = X^{-1}(B^c) \in \sigma(X)$. Enfin si $A_n = X^{-1}(B_n) \in \sigma(X)$ avec $B_n \in \mathcal{E}$, alors $\bigcup_n A_n = \bigcup_n X^{-1}(B_n) = X^{-1}(\bigcup_n B_n) \in \sigma(X)$ puisque $\bigcup_n B_n \in \mathcal{E}$. Ainsi $\sigma(X)$ est une tribu sur Ω (ceci ne repose que sur le fait que $\mathcal{E}$ est une tribu et sur les propriétés de l'image réciproque) qui par hypothèse de mesurabilité de X est incluse dans la tribu $\mathcal{A}$.

(ii) Si $B \in \mathcal{E}$ alors $X^{-1}(B) \in \mathcal{A}$ (mesurabilité) et $P_X(B) = \mathbb{P}(X^{-1}(B)) \in [0,1]$ est bien défini. On a $P_X(E) = \mathbb{P}(X^{-1}(E)) = \mathbb{P}(\Omega) = 1$; si $(B_n)_n \subset \mathcal{E}$ sont deux à deux disjoints alors $(X^{-1}(B_n))_n \subset \mathcal{A}$ sont deux à deux disjoints et on a :

$$P_X\Big(\bigcup_n B_n\Big) = \mathbb{P}\Big(X^{-1}\Big(\bigcup_n B_n\Big)\Big)$$
$$= \mathbb{P}\Big(\bigcup_n X^{-1}(B_n)\Big) \overset{\text{disjonction}}{=} \sum_n \mathbb{P}\big(X^{-1}(B_n)\big) = \sum_n P_X(B_n).$$

L'application P_X est donc bien une mesure de probabilité sur $(E, \mathcal{E})$. □

Remarque. — L'espace $(E, \mathcal{E}, \mu)$ est alors un espace probabilisé qui peut bien tenir le rôle de modèle probabiliste.

Exercice. — Soit E un ensemble. Montrer que $\mathcal{E}_{\text{triviale}} = \{\varnothing, E\}$ et $\mathcal{E}_{\text{discrète}} = \mathcal{P}(E)$ sont des tribus sur E (elles sont appelées respectivement tribu triviale, aussi appelée tribu grossière, et tribu discrète sur l'ensemble E).

Soient $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces mesurables. Constater que si $\mathcal{E} = \mathcal{P}(E)$ alors toute application $f : E \to F$ est mesurable pour ces structures mesurables.

2. Cas usuels

2.1. Espaces probabilisés discrets

Espaces probabilisés classiques. — L'ensemble Ω est un ensemble fini non vide, la tribu $\mathcal{A}$ est la tribu discrète $\mathcal{P}(\Omega)$ et la mesure de probabilité $\mathbb{P}$ est définie par :

$$\text{si } A \subset \Omega, \qquad \mathbb{P}(A) = \frac{\operatorname{Card} A}{\operatorname{Card} \Omega}\,.$$

Le triplet $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est alors qualifié d'*espace probabilisé classique* et la mesure de probabilité est dite *uniforme*. On omet souvent — et abusivement — de préciser la tribu des événements dans ce cas.

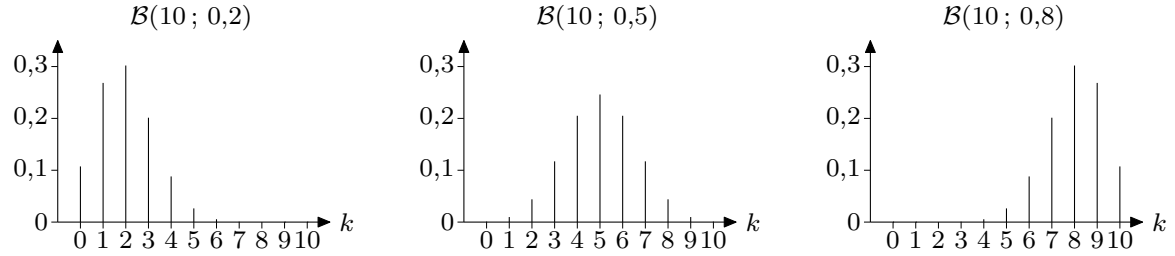
Ce type d'espace intervient dans la modélisation d'expériences aléatoires pour lesquelles il n'y a qu'un nombre fini de résultats possibles — résultats identifiés avec les épreuves — et qui ont une même probabilité d'être réalisés (tirages de boules dans une urne [avec ou sans remise], de cartes, lancers de pièces ou de dés équilibrés, ...).

Espaces probabilisés discrets. — Plus généralement, considérons Ω un ensemble non vide, fini ou dénombrable, muni de la tribu discrète $\mathcal{A} = \mathcal{P}(\Omega)$. Alors la donnée d'une mesure de probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{P}(\Omega))$ équivaut à la donnée d'une suite $(p_\omega)_{\omega\in\Omega} \subset [0,1]$ telle que $\sum_{\omega\in\Omega} p_\omega = 1$ simplement en posant $p_\omega = \mathbb{P}(\{\omega\})$ pour tout $\omega \in \Omega$. On a alors

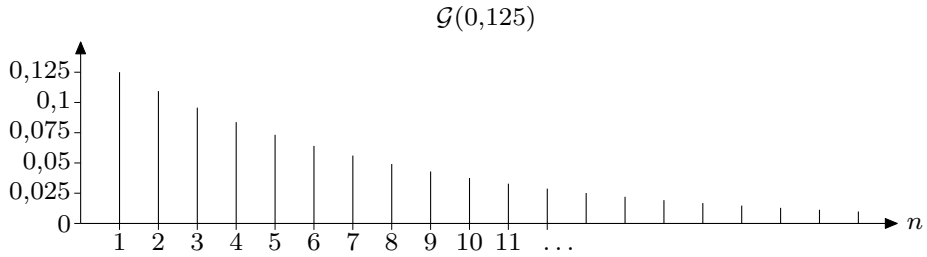
$$\text{si } A \subset \Omega, \qquad \mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}\{\omega\} = \sum_{\omega\in A} p_\omega.$$

Remarque. — Pour vérifier précisément l'affirmation précédente lorsque Ω est infini dénombrable, il est nécessaire d'avoir quelques connaissances sur les séries à termes positifs : le fait que la somme d'une telle série ne dépend pas de l'ordre de sommation (convergence commutative), et que de telles sommes peuvent s'écrire comme somme de sous-sommes (sommation par paquets). Nous n'insisterons pas sur ces questions et renvoyons le lecteur à tout cours de niveau universitaire sur les séries numériques.

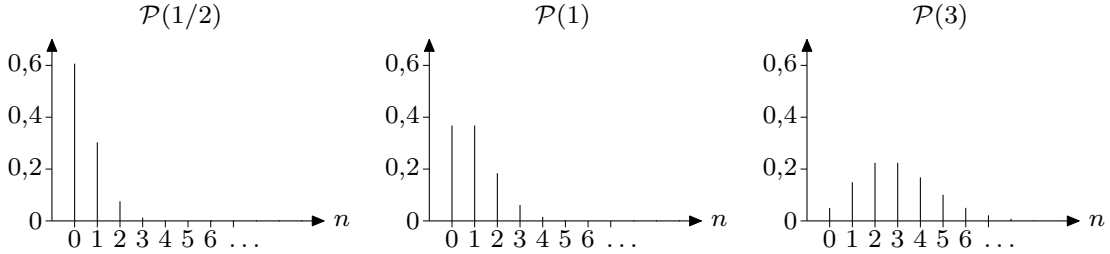
Exemples. — *a*) Soient $n \in \mathbb{N}^*$, $p \in [0,1]$, $\Omega = \{0, 1, \ldots, n\}$. En posant, pour $k \in \{0, 1, \ldots, n\}$, $p_k = C_n^k\, p^k (1-p)^{n-k}$, on définit une mesure de probabilité sur $\{0, 1, \ldots, n\}$ qui est appelée loi binomiale de paramètres n et p et notée $\mathcal{B}(n, p)$.



b) Soient $p \in \,]0,1]$, $\Omega = \mathbb{N}^* = \{1, 2, 3, \ldots\}$. En posant, pour $k \in \mathbb{N}^*$, $p_k = p(1-p)^{k-1}$, on définit une mesure de probabilité sur $\mathbb{N}^*$ qui est appelée loi géométrique de paramètre p et notée parfois $\mathcal{G}(p)$.



c) Soient $\lambda > 0$, $\Omega = \mathbb{N}$. En posant, pour $n \in \mathbb{N}^*$, $p_n = \mathrm{e}^{-\lambda}\, \lambda^n / n\,!$, on définit une mesure de probabilité sur $\mathbb{N}$ qui est appelée loi de Poisson — du nom de Siméon Denis Poisson, grand mathématicien et physicien français (1781–1840) — de paramètre λ et notée parfois $\mathcal{P}(\lambda)$.



2.2. Variables aléatoires discrètes

Définition. — Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(E, \mathcal{E})$ un espace mesurable. Une variable aléatoire $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ est *discrète* si et seulement si (quitte à restreindre l'espace d'arrivée, on peut supposer que) E est un ensemble fini ou dénombrable et $\mathcal{E} = \mathcal{P}(E)$. Dans ce cas les éléments $x \in E$ tels que $\mathbb{P}\{X = x\} > 0$ sont appelés *valeurs possibles* de la variable aléatoire X.

Lorsqu'une variable aléatoire $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ est discrète, c'est-à-dire si E est fini ou dénombrable et $\mathcal{E} = \mathcal{P}(E)$, on se dispense généralement de préciser la tribu sur l'espace image et on note notamment $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E$. En revanche, lorsque l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ n'est pas lui-même discret, la mention de la tribu sur l'espace de départ reste nécessaire.

Exemple. — Soit $A \in \mathcal{A}$. Définissons

$$\begin{aligned} \mathbb{1}_A : \Omega &\longrightarrow \{0,1\} \\ \omega &\longmapsto \begin{cases} 0 & \text{si } \omega \notin A \\ 1 & \text{si } \omega \in A \end{cases} \end{aligned}$$

La tribu discrète de $\{0,1\}$ est $\{\varnothing, \{0\}, \{1\}, \{0,1\}\}$, et on a $\mathbb{1}_A^{-1}(\varnothing) = \varnothing$, $\mathbb{1}_A^{-1}\{0\} = A^c$, $\mathbb{1}_A^{-1}\{1\} = A$ et $\mathbb{1}_A^{-1}\{0,1\} = \Omega$ qui sont tous des éléments de la tribu de départ $\mathcal{A}$. Ainsi l'application $\mathbb{1}_A$ est une variable aléatoire ; elle est appelée *indicatrice* de l'événement A.

Proposition. — *Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et E un ensemble fini ou dénombrable. Une application $X : \Omega \to E$ est une variable aléatoire discrète si et seulement si*

$$\text{pour tout } x \in E, \qquad \{X = x\} = X^{-1}\{x\} = \{\omega \in \Omega : X(\omega) = x\} \in \mathcal{A}.$$

Démonstration. — Si X est une variable aléatoire discrète, la condition précédente est évidemment satisfaite puisqu'alors toutes les pré-images de sous-ensembles de E par X sont des éléments de $\mathcal{A}$ et c'est donc *a fortiori* le cas des singletons de E.

Réciproquement, supposons que les pré-images des singletons de E par X soient des éléments de la tribu $\mathcal{A}$. Soit $B \subset E$. Le sous-ensemble B est au plus dénombrable et on a

$$X^{-1}(B) = X^{-1}\left(\bigcup_{x \in B} \{x\}\right) = \bigcup_{x \in B} X^{-1}\{x\}$$

qui est une réunion au plus dénombrable d'éléments de $\mathcal{A}$ et qui est donc encore dans $\mathcal{A}$. Puisque cela est vrai pour tout $B \subset E$, X est alors une variable aléatoire discrète. □

Remarques. — *a*) Si $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E$ est une variable aléatoire discrète, alors le triplet $(E, \mathcal{P}(E), P_X)$ est un espace probabilisé discret.

b) Deux généralisations de la notion de variable aléatoire discrète peuvent être données en deuxième et troisième lecture. La première est la suivante : soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $(E, \mathcal{E})$ un espace mesurable et $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ une variable aléatoire ; la variable aléatoire est discrète si $B = X(\Omega) \in \mathcal{E}$, si B est au plus dénombrable et si la tribu induite $\mathcal{E}_B$ de $\mathcal{E}$ sur B, $\mathcal{E}_B = \{B \cap C : C \in \mathcal{E}\}$ (*voir* fin de chapitre), est la tribu discrète de B, c'est-à-dire $\mathcal{E}_B = \mathcal{P}(B)$. La seconde généralisation consiste à remplacer la condition $(B = X(\Omega) \in \mathcal{E})$ par (il existe $B \in \mathcal{E}$ tel que $\mathbb{P}\{X \in B\} = 1$).

2.3. Mesures de probabilité absolument continues sur $\mathbb{R}$

Soit E un intervalle de $\mathbb{R}$, ou $\mathbb{R}$ tout entier. On appelle tribu borélienne de E, qu'on note $\mathcal{B}(E)$, la plus petite tribu sur E qui contient tous les sous-intervalles de E. C'est une tribu assez complexe incluse mais généralement distincte de $\mathcal{P}(E)$ — l'égalité ne pouvant avoir lieu

ici que si E est au plus réduit à un singleton, et si E n'est pas supposé être un intervalle, $\mathcal{B}(E) = \mathcal{P}(E)$ si et seulement si E est au plus dénombrable.

On peut montrer que toute application $f : E \to \mathbb{R}$ continue est mesurable lorsque les ensembles de départ et d'arrivée sont munis de leur tribu boréliennes. Il en est de même de toutes les applications de E dans $\mathbb{R}$ raisonnables (limites simples de fonctions continues). (*Voir* les compléments en fin de chapitre.)

Mesures de probabilité uniformes sur des intervalles bornés. — Soit E un intervalle borné de $\mathbb{R}$ de bornes $a < b$ ($[a,b]$, $[a,b[$, $]a,b]$, $]a,b[$) muni de la tribu $\mathcal{B}(E)$. Si $I \subset E$ de bornes $c \leqslant d$, on pose

$$\mu(I) = \frac{d-c}{b-a} \qquad \text{(longueur relative de l'intervalle } I \text{ dans } E\text{)}.$$

On montre que μ s'étend à $\mathcal{B}(E)$ en une unique mesure de probabilité qu'on nomme *mesure de probabilité uniforme* sur l'intervalle E. Cette mesure de probabilité s'étend à $\mathcal{B}(\mathbb{R})$ en posant $\mu(B) = \mu(B \cap E)$ pour tout $B \in \mathcal{B}(\mathbb{R})$.

Mesures de probabilité absolument continues sur $\mathbb{R}$. — Soit $p : \mathbb{R} \to \mathbb{R}_+$ une application positive et intégrable (au sens de Riemann par exemple) telle que

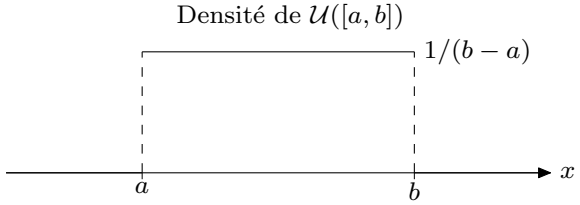
$$\int_{-\infty}^{+\infty} p(x)\,\mathrm{d}x = 1.$$

Si $I \subset \mathbb{R}$ est un intervalle de bornes $-\infty \leqslant a \leqslant b \leqslant +\infty$, on pose

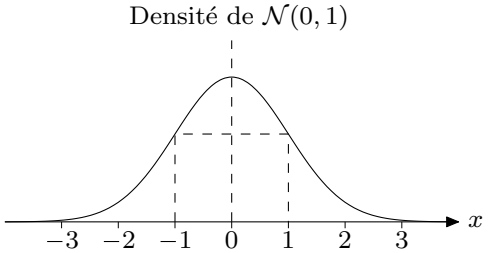
$$\mu(I) = \int_a^b p(x)\,\mathrm{d}x.$$

On montre que μ s'étend à $\mathcal{B}(\mathbb{R})$ en une unique mesure de probabilité qui est alors qualifiée d'*absolument continue*, ou *à densité*, de fonction de densité p.

Exemples. — *a*) Définissons pour $-\infty < a < b < +\infty$, $p : \mathbb{R} \to \mathbb{R}_+$, $x \mapsto \mathbb{1}_{[a,b]}(x)/(b-a)$. Alors p est une fonction de densité de la mesure de probabilité uniforme sur l'intervalle $[a,b]$ (que les bornes soient ouvertes ou fermées dans la définition de p ou de l'intervalle est inessentiel).



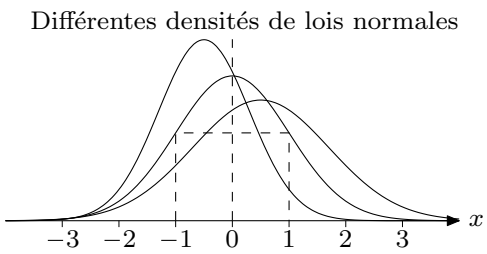
b) Définissons $p : \mathbb{R} \to \mathbb{R}_+$, $x \mapsto \mathrm{e}^{-x^2/2}/\sqrt{2\pi}$. Alors la fonction p est une fonction de densité d'une unique mesure de probabilité sur $\mathbb{R}$ appelée *loi Normale*, ou loi gaussienne standard, ou de Laplace–Gauss, ou encore de de Moivre–Laplace–Gauss, et qui est notée $\mathcal{N}(0,1)$. Cette fonction de densité est bien connue pour la forme en cloche de sa courbe représentative.

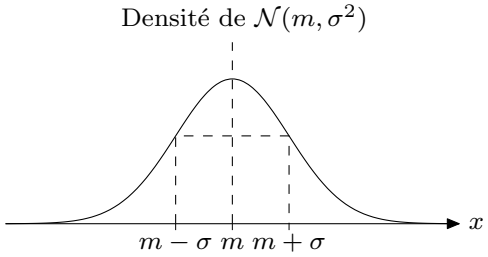


Plus généralement, pour $m \in \mathbb{R}$ et $\sigma > 0$, en posant pour fonction de densité

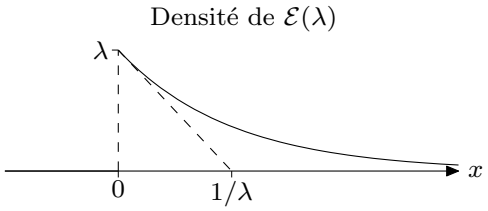
$$p : \mathbb{R} \longrightarrow \mathbb{R}_+$$
$$x \longmapsto \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right) \times \frac{1}{\sqrt{2\pi}\,\sigma}\,,$$

on définit une unique mesure de probabilité nommée *loi normale* (gaussienne) de moyenne m et de variance σ^2 (ou d'écart-type σ). Elle est généralement notée $\mathcal{N}(m, \sigma^2)$ (ou $\mathcal{N}(m, \sigma)$). La forme de la courbe représentative de cette fonction de densité s'obtient par une simple transformation affine de la précédente, ce qui a pour conséquence que toute transformation affine d'une variable aléatoire de loi normale est de loi normale avec pour paramètres les paramètres de la loi de la variable aléatoire de départ convenablement transformés.





c) Définissons pour $\lambda > 0$, $p : \mathbb{R} \to \mathbb{R}_+$, $x \mapsto \mathbb{1}_{\mathbb{R}_+}(x)\,\lambda\,\mathrm{e}^{-\lambda x}$. Alors p est une fonction de densité d'une unique mesure de probabilité sur $\mathbb{R}$ appelée *loi exponentielle* de paramètre λ et parfois notée $\mathcal{E}(\lambda)$. Ce type de loi intervient fréquemment dans la modélisation de temps aléatoires et on remplace alors souvent le paramètre λ par son inverse $\tau = 1/\lambda$ de sorte qu'on considère alors la fonction de densité $p : \mathbb{R}_+ \to \mathbb{R}_+$, $t \mapsto \mathrm{e}^{-t/\tau}/\tau$.



Remarques. — *a*) Ce qui précède admet une généralisation immédiate à $\mathbb{R}^d$.

b) Se limiter à la tribu borélienne plutôt que de considérer l'ensemble des parties dans le cas de $\mathbb{R}$ ou un intervalle de longueur strictement positive est une nécessité. Par exemple, dans le cas de $\mathbb{R}^3$ l'espace euclidien usuel, on a le théorème (souvent qualifié de paradoxe) de Banach–Tarski : *considérons une boule unité (solide) ; il est possible de découper cette boule en un nombre fini de morceaux (solides) et de les réassembler (sans déformation) en deux boules unités, et d'avoir ainsi un doublement de volume.* Le problème est que les morceaux en question ne sont pas mesurables, *i.e.*, boréliens, dans le contexte de la mesure de volume usuelle. Ainsi pour éviter ce genre de problèmes on est obligé de se limiter aux ensembles boréliens pour lesquels cette difficulté ne se pose pas.

3. **Propriétés essentielles**

Théorème. — *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé.*

(i) *Pour tout* $A \in \mathcal{A}$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(ii) *Si* $A, B \in \mathcal{A}$, $A \subset B$, *alors* $\mathbb{P}(A) \leqslant \mathbb{P}(B)$ *(monotonie).*

(iii) *Pour tous* $A, B \in \mathcal{A}$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

(iv) *Si $(A_n)_n \subset \mathcal{A}$ est une suite croissante d'événements (pour tout n, $A_n \subset A_{n+1}$), alors*

$$\mathbb{P}\left(\bigcup_n A_n\right) = \lim_N \mathbb{P}(A_N) \qquad \textit{(continuité suivant les suites croissantes).}$$

(iv-bis) *Si $(A_n)_n \subset \mathcal{A}$ est une suite décroissante d'événements (pour tout n, $A_n \supset A_{n+1}$), alors*

$$\mathbb{P}\left(\bigcap_n A_n\right) = \lim_N \mathbb{P}(A_N) \qquad \textit{(continuité suivant les suites décroissantes).}$$

(v) *Si $(A_n)_n \subset \mathcal{A}$ est une suite quelconque d'événements, alors*

$$\mathbb{P}\left(\bigcup_n A_n\right) \leqslant \sum_n \mathbb{P}(A_n). \qquad \textit{(inégalité de Boole).}$$

Démonstration. — (i) On a $A \cup A^c = \Omega$ et $A \cap A^c = \varnothing$, donc $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$ par disjonction de A et A^c, d'où $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(ii) Si $A \subset B$, $B = A \cup (B \backslash A)$ et $A \cap (B \backslash A) = \varnothing$. Précisons au passage que $B \backslash A = B \cap A^c$, écriture qui montre que $B \backslash A$ est dans $\mathcal{A}$. Donc $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \backslash A) \geqslant \mathbb{P}(A)$.

(iii) On a $A = (A \backslash (A \cap B)) \cup (A \cap B)$ union de deux événements disjoints et donc $\mathbb{P}(A) = \mathbb{P}(A \backslash (A \cap B)) + \mathbb{P}(A \cap B)$. De même, $B = (B \backslash (A \cap B)) \cup (A \cap B)$ et $\mathbb{P}(B) = \mathbb{P}(B \backslash (A \cap B)) + \mathbb{P}(A \cap B)$. Comme $A \cup B = (A \backslash (A \cap B)) \cup (B \backslash (A \cap B)) \cup (A \cap B)$ s'écrit ainsi comme la réunion de trois événements disjoints, on a $\mathbb{P}(A \cup B) = \mathbb{P}(A \backslash (A \cap B)) + \mathbb{P}(B \backslash (A \cap B)) + \mathbb{P}(A \cap B) = (\mathbb{P}(A) - \mathbb{P}(A \cap B)) + (\mathbb{P}(B) - \mathbb{P}(A \cap B)) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, ce qu'il fallait démontrer.

(iv) Posons $B_1 = A_1$ et $B_n = A_n \backslash A_{n-1}$ pour tout $n \geqslant 2$. On a alors $\bigcup_n A_n = \bigcup_n B_n$ et les $(B_n)_n$ sont disjoints. Donc $\mathbb{P}(\bigcup_n A_n) = \mathbb{P}(\bigcup_n B_n) = \sum_n \mathbb{P}(B_n)$ par additivité dénombrable. Or $\sum_n \mathbb{P}(B_n) = \lim_N \sum_{n=1}^N \mathbb{P}(B_n)$ par définition de la somme d'une série, mais $\sum_{n=1}^N \mathbb{P}(B_n) = \mathbb{P}(\bigcup_{n=1}^N B_n)$ par additivité dénombrable. Comme $A_N = \bigcup_{n=1}^N B_n$ par construction, on a alors $\mathbb{P}(\bigcup_n A_n) = \lim_N \mathbb{P}(\bigcup_{n=1}^N B_n) = \lim_N \mathbb{P}(A_N)$.

(iv-bis) Cette propriété s'obtient de manière immédiate à partir de la précédente par passage au complémentaire.

(v) Montrons par récurrence que la propriété suivante est vraie pour tout $N \geqslant 1$:

$$\text{pour tous } A_n, \ldots, A_N \in \mathcal{A}, \qquad \mathbb{P}\left(\bigcup_{n=1}^N A_n\right) \leqslant \sum_{n=1}^N \mathbb{P}(A_n). \qquad [\mathrm{HR}_N]$$

Pour $n = 1$, c'est immédiat : $\mathbb{P}(A_1) \leqslant \mathbb{P}(A_1)$ pour tout $A_1 \in \mathcal{A}$. Pour $n = 2$, soient A_1, $A_2 \in \mathcal{A}$; on a $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) \leqslant \mathbb{P}(A_1) + \mathbb{P}(A_2)$.

Supposons, pour un certain $N \geqslant 1$, que $[\mathrm{HR}_N]$ est vraie. Soient $A_1, \ldots, A_{N+1} \in \mathcal{A}$. On a

$$\mathbb{P}\left(\bigcup_{n=1}^{N+1} A_n\right) = \mathbb{P}\left(\left(\bigcup_{n=1}^{N} A_n\right) \cup A_{N+1}\right) \leqslant \mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right) + \mathbb{P}(A_{N+1})$$

d'après le cas de la réunion de deux événements, et en utilisant l'hypothèse de récurrence on obtient immédiatement

$$\mathbb{P}\left(\bigcup_{n=1}^{N+1} A_n\right) \leqslant \sum_{n=1}^{N+1} \mathbb{P}(A_n).$$

Ainsi, par récurrence, la propriété $[\mathrm{HR}_N]$ est vraie pour tout N.

Soit $(A_n)_{n\geqslant 1}$ une suite d'éléments de $\mathcal{A}$. Posons $B_n = \bigcup_{k=1}^{n} A_k$, $n \geqslant 1$. Les $(B_n)_{n\geqslant 1}$ forment une suite croissante d'événements et on a

$$\mathbb{P}\Big(\bigcup_n B_n\Big) = \mathbb{P}\Big(\bigcup_n A_n\Big) = \lim_N \mathbb{P}(B_N)$$

par continuité d'une mesure de probabilité le long des suites croissantes d'événements. Or,

$$\mathbb{P}(B_N) \leqslant \sum_{n=1}^{N} \mathbb{P}(A_n) \leqslant \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

En passant à la limite sur N,

$$\lim_N \mathbb{P}(B_N) \leqslant \sum_{n=1}^{\infty} \mathbb{P}(A_n), \qquad \text{c'est-à-dire} \qquad \mathbb{P}\Big(\bigcup_n A_n\Big) \leqslant \sum_{n=1}^{\infty} \mathbb{P}(A_n),$$

ce qu'il fallait démontrer. □

Corollaire. — *Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(A_n)_n$ une suite finie ou dénombrable d'événements.*

(i) *Si pour tout n, $\mathbb{P}(A_n) = 0$, alors $\mathbb{P}(\bigcup_n A_n) = 0$.*

(ii) *Si pour tout n, $\mathbb{P}(A_n) = 1$, alors $\mathbb{P}(\bigcap_n A_n) = 1$.*

Démonstration. — Dans le premier cas, par l'inégalité de Boole,

$$\mathbb{P}\Big(\bigcup_n A_n\Big) \leqslant \sum_n \mathbb{P}(A_n) = \sum_n 0 = 0.$$

Dans le second cas, on passe au complémentaire : pour tout n, $\mathbb{P}(A_n^c) = 1 - \mathbb{P}(A_n) = 0$, ainsi $\mathbb{P}(\bigcup_n A_n^c) = 0$, et alors

$$\mathbb{P}\Big(\bigcap_n A_n\Big) = 1 - \mathbb{P}\Big(\Big(\bigcap_n A_n\Big)^c\Big) = 1 - \mathbb{P}\Big(\bigcup_n A_n^c\Big) = 1 - 0 = 1.$$ □

Remarque et définition. — Le rôle des ensembles de mesure nulle, ou de manière complémentaire égale à 1, est important. Un événement de probabilité 1 est dit *presque-sûr*. Une propriété ayant lieu sur événement de probabilité 1 est dite *presque-sûre*. Un événement de probabilité 0 est dit *négligeable*.

Théorème. — *Soient $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ et $g : (F, \mathcal{F}) \to (G, \mathcal{G})$ deux applications mesurables. Alors $g \circ f : (E, \mathcal{E}) \to (G, \mathcal{G})$ est une application mesurable.*

En particulier si $X : (\Omega, \mathcal{A}, \mathbb{P})$ est une variable aléatoire, $Y = f \circ X : (\Omega, \mathcal{A}, \mathbb{P}) \to (F, \mathcal{F})$ est une variable aléatoire. De plus la loi de Y qui est l'image de la mesure de probabilité $\mathbb{P}$ par Y, est aussi l'image de la loi de X par l'application mesurable f.

Démonstration. — Soit $D \in \mathcal{G}$. Puisque g est mesurable, $C = g^{-1}(D) \in \mathcal{F}$. Puisque f est mesurable, $B = f^{-1}(C) = f^{-1}(g^{-1}(D)) = (g \circ f)^{-1}(D) \in \mathcal{E}$. Puisque c'est vrai pour tout $D \in \mathcal{G}$, $g \circ f$ est mesurable.

Soit $D \in \mathcal{G}$. On a $P_Y(D) = \mathbb{P}\{Y \in D\} = \mathbb{P}\{f(X) \in D\} = \mathbb{P}\{X \in f^{-1}(D)\} = P_X(f^{-1}(D))$ $= P_X\{f \in D\}$, ce qui justifie la seconde partie du théorème. □

Remarques. — *a*) Dans la plupart des cas courants (bon choix de tribus, applications pas trop pathologiques) les applications rencontrées sont mesurables.

b) Si $Y = f(X)$, on a $\sigma(Y) \subset \sigma(X)$. La réciproque est vraie dans les cas usuels, c'est-à-dire lorsque Y est à valeurs dans un espace métrisable séparable muni de sa tribu borélienne (lemme de Doob).

COMPLÉMENTS AU CHAPITRE I

A. **Séries à termes positifs**

B. **Génération de tribus**

THÉORÈME. — *Soient E un ensemble et $(\mathcal{B}_j)_{j\in J}$ une collection non vide de tribus sur E. Alors*

$$\bigcap_{j\in J} \mathcal{B}_j = \{B \subset E : B \in \mathcal{B}_j,\ \forall\, j \in J\}$$

est une tribu sur E.

Démonstration. — C'est immédiat : pour tout $j \in J$, $E \in \mathcal{B}_j$, donc $E \in \bigcap_{j\in J} \mathcal{B}_j$; si $B \in \bigcap_{j\in J} \mathcal{B}_j$, alors $B^c \in \mathcal{B}_j$ pour tout $j \in J$, donc $B^c \in \bigcap_{j\in J} \mathcal{B}_j$; si $(B_n)_n \subset \bigcap_{j\in J} \mathcal{B}_j$, alors pour tout $j \in J$, $(B_n)_n \subset \mathcal{B}_j$ et $\bigcap_n B_n \in \mathcal{B}_j$, donc $\bigcap_n B_n \in \bigcap_{j\in J} \mathcal{B}_j$. □

THÉORÈME ET DÉFINITION. — *Soient E un ensemble et $(B_i)_{i\in I}$ une collection quelconque de parties de E. Il existe une unique tribu sur E appelée tribu engendrée par $(B_i)_{i\in I}$ et notée $\sigma((B_i)_{i\in I})$ contenant $(B_i)_{i\in I}$ et telle que toute autre tribu sur E contenant elle aussi $(B_i)_{i\in I}$ contient la tribu $\sigma((B_i)_{i\in I})$.*

Démonstration. — Considérons l'ensemble des tribus sur E contenant $(B_i)_{i\in I}$, notons cet ensemble $(\mathcal{B}_j)_{j\in J}$. Cet ensemble est non vide puisqu'il contient $\mathcal{P}(E)$. Posons $\mathcal{B} = \bigcap_{j\in J} \mathcal{B}_j$. C'est une tribu sur E qui contient $(B_i)_{i\in I}$. Si $\mathcal{B}'$ est une tribu sur E contenant $(B_i)_{i\in I}$, alors $\mathcal{B}' \in (\mathcal{B}_j)_{j\in J}$ et donc $\mathcal{B} \subset \mathcal{B}'$ (minimalité). L'unicité est immédiate. □

Exemples. — *a*) Soit E un ensemble fini ou dénombrable. Alors la tribu $\mathcal{E} = \sigma\{\{x\} : x \in E\}$ engendrée par tous les singletons de E est égale à la tribu discrète $\mathcal{P}(E)$. En effet, toute partie $B \subset E$ est au plus dénombrable et est donc égale à la réunion au plus dénombrable des singletons qu'elle contient et ainsi $B \in \mathcal{E}$, d'où $\mathbb{P}(E) \subset \mathcal{E}$. Comme bien sûr $\mathcal{E} \subset \mathbb{P}(E)$, on a $\mathcal{E} = \mathcal{P}(E)$.

b) Lorsque E est un ensemble infini non dénombrable (penser à $\mathbb{R}$ par exemple), la tribu engendrée par les singletons consiste en les parties au plus dénombrables ou de complémentaires au plus dénombrables. Elle a alors des traits assez pathologiques (non séparabilité en particulier, c'est-à-dire qu'elle n'est pas engendrée par une famille dénombrable de parties).

THÉORÈME. — *Soient $(E, \mathcal{E})$, $(F, \mathcal{F})$ deux espaces mesurables et $\mathcal{C} \subset \mathcal{F}$ une collection de parties mesurables de F telles que $\mathcal{F} = \sigma(\mathcal{C})$.*

Une application $f : E \to F$ est $\mathcal{E}/\mathcal{F}$-mesurable si et seulement si

$$\text{pour tout } C \in \mathcal{C}, \qquad f^{-1}(C) \in \mathcal{E}.$$

Démonstration. — Dans un premier temps, soient E, F deux ensembles, $\mathcal{E}$ une tribu sur E et $f : E \to F$ une application. Notons

$$\tau(f, \mathcal{E}) = \{C \subset F : f^{-1}(C) \in \mathcal{E}\}.$$

Montrons que $\tau(f, \mathcal{E})$ est une tribu sur F :

— On a $f^{-1}(\varnothing) = \varnothing \in \mathcal{E}$, donc $\varnothing \in \tau(f, \mathcal{E})$. (De même, $f^{-1}(F) = E \in \mathcal{E}$, donc $F \in \tau(f, \mathcal{E})$.)

— Si $C \in \tau(f, \mathcal{E})$, $f^{-1}(C) \in \mathcal{E}$ donc $f^{-1}(C^c) = (f^{-1}(C))^c \in \mathcal{E}$ puisque $\mathcal{E}$ est stable par passage au complémentaire, et ainsi $C \in \tau(f, \mathcal{E})$.

— Si $C_1, \ldots, C_n, \ldots$ est une suite à valeurs dans $\tau(f, \mathcal{E})$, c'est-à-dire $f^{-1}(C_n) \in \mathcal{E}$ pour tout n, alors $f^{-1}(\bigcup_n C_n) = \bigcup_n f^{-1}(C_n) \in \mathcal{E}$ puisque $\mathcal{E}$ est stable par réunion au plus dénombrable, c'est-à-dire $\bigcup_n C_n \in \tau(f, \mathcal{E})$.

Dans un deuxième temps, soit $\mathcal{F}$ une tribu sur F. Montrons que $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ est mesurable si et seulement si $\mathcal{F} \subset \tau(f, \mathcal{E})$: on a $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ mesurable $\iff \forall\, C \in \mathcal{F}$, $f^{-1}(C) \in \mathcal{E} \iff \forall\, C \in \mathcal{F}$, $C \in \tau(f, \mathcal{E}) \iff \mathcal{F} \subset \tau(f, E)$.

Dans un troisième temps, nous démontrons le théorème. La condition est évidemment nécessaire. Elle est suffisante : si $f^{-1}(C) \in \mathcal{E}$ pour tout $C \in \mathcal{C}$, c'est-à-dire si $\mathcal{C}$ est inclus dans $\tau(f, \mathcal{E})$, alors la tribu engendrée par $\mathcal{C}$, par hypothèse $\mathcal{F}$, est incluse dans $\tau(f, \mathcal{E})$ puisque cette dernière est une tribu contenant $\mathcal{C}$. L'étape précédente permet alors d'affirmer que f est $\mathcal{E}/\mathcal{F}$-mesurable. □

C. **Classes monotones**

Définition. — Soient E un ensemble et $\mathcal{M}$ une famille de parties de E. La famille $\mathcal{M}$ est une *classe monotone* si et seulement si :

(i) $E \in \mathcal{M}$;

(ii) si A, $B \in \mathcal{M}$ et $A \subset B$, alors $B \backslash A = B \cap A^c \in \mathcal{M}$;

(iii) si $(A_n)_n$ est une suite croissante d'éléments de $\mathcal{M}$ alors $\bigcup_n A_n \in \mathcal{M}$.

Exemples. — *a*) L'ensemble des parties de E est une classe monotone sur E.

b) Toute tribu sur E est une classe monotone sur E.

Théorème et définition. — *Soient E un ensemble et $(B_i)_{i \in I}$ une collection quelconque de parties de E. Il existe une unique classe monotone sur E appelée classe monotone engendrée par $(B_i)_{i \in I}$ et notée $\mathcal{M}((B_i)_{i \in I})$ contenant $(B_i)_{i \in I}$ et telle que toute autre classe monotone contenant elle aussi $(B_i)_{i \in I}$ contient la classe monotone $\mathcal{M}((B_i)_{i \in I})$.*

Lemme. — *Soient E un ensemble et $(\mathcal{M}_j)_{j \in J}$ une collection non vide de classes monotones sur E. Alors*

$$\bigcap_{j \in J} \mathcal{M}_j = \{B \subset E : B \in \mathcal{M}_j,\ \forall\, j \in J\}$$

est une classe monotone sur E.

Démonstration du lemme. — C'est immédiat. □

Démonstration du théorème. — L'ensemble des classes monotones sur E contenant $(B_i)_{i \in I}$ est non vide puisque $\mathcal{P}(E)$ est une telle classe monotone. La classe $\mathcal{M}((B_i)_{i \in I})$ n'est autre que l'intersection de ces classes monotones. □

Théorème. — *Soit $\mathcal{M}$ une classe monotone sur un ensemble E. Si $\mathcal{M}$ est stable par intersections finies, alors $\mathcal{M}$ est une tribu.*

Démonstration. — Avec les axiomes (i) et (ii) d'une classe monotone, on sait que $\mathcal{M}$ contient E et est stable par passage au complémentaire. Il suffit de démontrer qu'une réunion dénombrable de $(A_n)_n \subset \mathcal{M}$ quelconques (non nécessairement croissants) demeure dans $\mathcal{M}$. Par stabilité par passage au complémentaire, l'hypothèse de stabilité par intersections finies équivaut

à la stabilité par réunions finies. En écrivant alors

$$\bigcup_n A_n = \bigcup_n \Big(\bigcup_{k=1}^{n} A_k \Big),$$

(chaque A_k est répété une infinité de fois dans la seconde écriture) on a alors une réunion croissante d'éléments de $\mathcal{M}$ qui est dès lors dans $\mathcal{M}$. □

Théorème des classes monotones. — *Soit $\mathcal{B} \subset \mathcal{P}(E)$ une classe de parties d'un ensemble E. Si $\mathcal{B}$ est stable par intersections finies, alors $\mathcal{M}(\mathcal{B}) = \sigma(\mathcal{B})$.*

Démonstration. — La tribu $\sigma(\mathcal{B})$ est une classe monotone contenant $\mathcal{B}$ donc $\sigma(\mathcal{B}) \supset \mathcal{M}(\mathcal{B})$. Réciproquement, il suffit de montrer que $\mathcal{M}(\mathcal{B})$ est stable par intersections finies. En effet, si c'est le cas, $\mathcal{M}(\mathcal{B})$ est alors une tribu contenant $\mathcal{B}$ et on a nécessairement l'inclusion réciproque $\sigma(\mathcal{B}) \subset \mathcal{M}(\mathcal{B})$.

Posons

$$\mathcal{M}_1 = \{A \in \mathcal{M}(\mathcal{B}) : A \cap B \in \mathcal{M}(\mathcal{B}) \text{ pour tout } B \in \mathcal{B}\}.$$

C'est une classe monotone : (i) $E \in \mathcal{M}_1$ de façon évidente ; (ii) si A_1, $A_2 \in \mathcal{M}_1$, $A_1 \subset A_2$, alors pour $B \in \mathcal{B}$, $(A_1 \backslash A_2) \cap B = (A_1 \cap B) \backslash (A_2 \cap B)$, or $A_1 \cap B$ et $A_2 \cap B$ sont dans $\mathcal{M}(\mathcal{B})$ et $(A_1 \cap B) \supset (A_2 \cap B)$, donc $(A_1 \backslash A_2) \cap B \in \mathcal{M}(\mathcal{B})$, et ainsi $A_1 \backslash A_2 \in \mathcal{M}_1$; (iii) si $(A_n)_n \subset \mathcal{M}_1$ est une suite croissante et $B \in \mathcal{B}$, alors $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B)$ est une réunion croissante d'éléments de $\mathcal{M}(\mathcal{B})$ et est donc dans $\mathcal{M}(\mathcal{B})$, ainsi $\bigcup_n A_n \in \mathcal{M}_1$.

Par hypothèse de stabilité de $\mathcal{B}$ par intersections finies, on constate que $\mathcal{B} \subset \mathcal{M}_1$. Par conséquent $\mathcal{M}_1 \supset \mathcal{M}(\mathcal{B})$ et puisque par définition même de $\mathcal{M}_1$ on avait déjà $\mathcal{M}_1 \subset \mathcal{M}(\mathcal{B})$, on a égalité entre ces deux classes monotones.

Posons

$$\mathcal{M}_2 = \{A \in \mathcal{M}(\mathcal{B}) : A \cap B \in \mathcal{M}(\mathcal{B}) \text{ pour tout } B \in \mathcal{M}(\mathcal{B})\}.$$

C'est une classe monotone : (i) $E \in \mathcal{M}_2$ de façon évidente ; (ii) si A_1, $A_2 \in \mathcal{M}_2$, $A_1 \subset A_2$, alors pour $B \in \mathcal{M}(\mathcal{B})$, $(A_2 \backslash A_1) \cap B = (A_2 \cap B) \backslash (A_1 \cap B)$, or $A_1 \cap B$ et $A_2 \cap B$ sont dans $\mathcal{M}(\mathcal{B})$ et $(A_1 \cap B) \supset (A_2 \cap B)$, donc $(A_2 \backslash A_1) \cap B \in \mathcal{M}(\mathcal{B})$, et ainsi $A_2 \backslash A_1 \in \mathcal{M}_2$; (iii) si $(A_n)_n \subset \mathcal{M}_2$ est une suite croissante et $B \in \mathcal{M}(\mathcal{B})$, alors $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B)$ est une réunion croissante d'éléments de $\mathcal{M}(\mathcal{B})$ et est donc dans $\mathcal{M}(\mathcal{B})$, ainsi $\bigcup_n A_n \in \mathcal{M}_2$.

Il ne reste qu'à constater que $\mathcal{B} \subset \mathcal{M}_2$. Soit $B \in \mathcal{B}$, montrons que pour tout $C \in \mathcal{M}(\mathcal{B})$, $B \cap C \in \mathcal{M}(\mathcal{B})$, ce qui montrera que $B \in \mathcal{M}_2$: c'est immédiat, car $C \in \mathcal{M}(\mathcal{B}) \subset \mathcal{M}_1$, donc $B \cap C \in \mathcal{M}(\mathcal{B})$. Ainsi $\mathcal{M}_2$ est une classe monotone contenant $\mathcal{B}$ et contenue dans $\mathcal{M}(\mathcal{B})$, et donc $\mathcal{M}_2 = \mathcal{M}(\mathcal{B})$. Ceci signifie exactement que la classe monotone $\mathcal{M}(\mathcal{B})$ est stable par intersections finies. C'est donc une tribu. □

D. **Espaces topologiques et mesurabilité**

Définition. — Soit E un espace topologique. Notons $\mathcal{O}(E)$ l'ensemble des parties ouvertes de E et $\mathcal{F}(E)$ l'ensemble de ses parties fermées. La tribu $\sigma(\mathcal{O}(E))$ est égale à la tribu $\sigma(\mathcal{F}(E))$, elle est appelée *tribu borélienne* de E et notée $\mathcal{B}(E)$.

Dans cette définition il est affirmé que la tribu engendrée par les ouverts $\sigma(\mathcal{O}(E))$ est égale à celle qui est engendrée par les fermés. Pour le voir, la tribu engendrée par les ouverts contient les ouverts, donc leurs complémentaires qui sont les fermés. Cette tribu contient donc la plus petite tribu contenant les fermés $\sigma(\mathcal{F}(E))$, et on a donc $\sigma(\mathcal{O}(E)) \supset \sigma(\mathcal{F}(E))$. L'inclusion réciproque vient du fait que les complémentaires des fermés sont les ouverts.

Remarques. — *a*) Soit $E = \mathbb{R}$ avec sa structure d'espace topologique (métrique) usuelle. On peut montrer de manière élémentaire que tout ouvert de $\mathbb{R}$ est une réunion au plus dénombrable d'intervalles (disjoints). Ainsi la tribu borélienne $\mathcal{B}(\mathbb{R})$ est engendrée par les intervalles ouverts $]a, b[$. Sans trop de difficultés, on peut montrer qu'elle est aussi engendrée respectivement par les intervalles de la forme $[a, b]$, $[a, b[$, $]a, b]$, d'extrémités réelles et même simplement rationnelles. On pourra s'attarder sur le fait que $\mathcal{B}(\mathbb{R})$ est engendrée par les intervalles $]-\infty, x]$, $x \in \mathbb{R}$ (ou même $x \in \mathbb{Q}$), à titre d'exercice. Noter que la tribu borélienne de $\mathbb{R}$ est engendrée par une famille dénombrable de parties de $\mathbb{R}$, cette tribu est donc séparable.

b) Soit $E = \mathbb{R}^d$ avec sa structure d'espace topologique (métrique) usuelle. On peut montrer que $\mathcal{B}(\mathbb{R}^d)$ est engendré par les boules ouvertes, boules de la métrique euclidienne ou même boules de la métrique du maximum (les boules sont alors des pavés). Les coordonnées des centres et les rayons peuvent être pris rationnels. Cette tribu est donc séparable.

Définition. — Toute application mesurable entre deux espaces topologiques munis de leur tribus boréliennes est dite *application borélienne*.

Théorème. — *Soient E et F deux espaces topologiques munis de leurs tribus boréliennes. Si $f : E \to F$ est une application continue, alors $f : (E, \mathcal{B}(E)) \to (F, \mathcal{B}(F))$ est mesurable.*

Démonstration. — La continuité de f signifie que pour tout $C \in \mathcal{O}(F)$, $f^{-1}(C) \in \mathcal{O}(E)$. Comme $\mathcal{O}(E) \subset \mathcal{B}(E)$, et que $\sigma(\mathcal{O}(F)) = \mathcal{B}(F)$, le théorème précédent affirme que f est $\mathcal{B}(E)/\mathcal{B}(F)$-mesurable. □

Théorème. — *Soit $X_n : (\Omega, \mathcal{A}) \to (E, \mathcal{B}(E))$, $n \geqslant 1$, une suite d'applications mesurables à valeurs dans un espace topologique E. Supposons que la suite $(X_n)_{n\geqslant 1}$ converge simplement vers une application $X : \Omega \to E$: pour tout $\omega \in \Omega$, la suite $(X_n(\omega))_{n\geqslant 1}$ à valeurs dans E converge, sa limite étant notée $X(\omega)$. Alors, l'application $X : (\Omega, \mathcal{A}) \to (E, \mathcal{B}(E))$ est mesurable.*

Démonstration. — Il suffit de vérifier que pour tout ouvert $B \in \mathcal{O}(E)$ de E, $\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\} = X^{-1}(B) \in \mathcal{A}$. Pour $\omega \in \Omega$ donné, $X(\omega) \in B$ si et seulement si, à partir d'un certain rang $n(\omega)$, $X_k(\omega) \in B$ pour tout $k \geqslant n(\omega)$ (car B est ouvert). Ainsi,

$$\{X \in B\} = \bigcup_{n\geqslant 1} \bigcap_{k\geqslant n} \{X_n \in B\}.$$

Comme chaque $\{X_n \in B\} \in \mathcal{A}$ par mesurabilité de X_n, on a donc $\{X \in B\} \in \mathcal{A}$ par stabilité de $\mathcal{A}$ par intersections et réunions au plus dénombrables. □

Remarques. — *a*) Pour se faire une idée, si on considère les applications boréliennes (mesurables) de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans lui-même, nous savons que parmi elles, il y a les applications continues. Nous venons de voir qu'il y a aussi toutes les limites simples d'applications continues (qui ne sont pas nécessairement continues). L'ensemble des applications boréliennes est donc assez riche et se trouve contenir toutes les fonctions qu'on peut définir « à la main ».

b) Supposons l'espace topologique E muni d'une métrique $d : E \times E \to \mathbb{R}_+$ (métrisable), complet. Une suite $(x_n)_{n\geqslant 1}$ est convergente si et seulement si elle est de Cauchy, condition qui peut s'écrire $\forall\, N \geqslant 1$, $\exists\, n \geqslant 1$, $\forall\, k, \ell \geqslant n$, $d(x_k, d_\ell) \leqslant 1/N$ (N, n, k, ℓ sont des entiers strictement positifs). Si $X_n : (\Omega, \mathcal{A}) \to (E, \mathcal{B}(E))$, $n \geqslant 1$, son ensemble de convergence C_X est l'ensemble des $\omega \in \Omega$ tels que $(X_n(\omega))_{n\geqslant 1}$ est une suite de Cauchy à valeurs dans E, c'est donc

$$C_X = \bigcap_{N\geqslant 1} \bigcup_{n\geqslant 1} \bigcap_{k\geqslant n} \bigcap_{\ell\geqslant n} \{d(X_\ell, X_k) < 1/N\}.$$

L'application $d : E \times E \to \mathbb{R}_+$ étant continue, $d(X_\ell, X_k) : (\Omega, \mathcal{A}) \to \mathbb{R}_+$ est mesurable (*voir* espaces produits et séparabilité), et alors $\{d(X_\ell, X_k) < 1/N\} \in \mathcal{A}$ pour tous $N \geqslant 1$, $k, \ell \geqslant 1$. Il vient que l'ensemble de convergence est un sous-ensemble mesurable de Ω. Lorsque $(\Omega, \mathcal{A})$ est muni d'une mesure de probabilité, une suite de variables aléatoires est dite converger presque sûrement si cet ensemble de convergence C_X est de probabilité 1. Dans ce cas, définissant X sur C_X comme étant la valeur de ces limites et lui donnant sur le complémentaire C_X^c une valeur arbitraire fixe, on peut considérer qu'une suite de variables aléatoires convergeant presque sûrement admet une limite qui est encore une variable aléatoire.

E. **Espaces produits et mesurabilité**

Définition. — Soient $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces mesurables. Considérons les pavés mesurables $B \times C \subset E \times F$, $B \in \mathcal{E}$, $C \in \mathcal{F}$. La tribu qu'ils engendrent est une tribu sur $E \times F$ nommée *tribu produit* et notée $\mathcal{E} \otimes \mathcal{F}$.

Noter que si une intersection de pavés est un pavé, il n'en est pas nécessairement de même d'une réunion de pavés ou du complémentaire de pavés. La tribu $\mathcal{E} \otimes \mathcal{F}$ contient (dès que E et F ont plus de 2 éléments) donc des parties de $E \times F$ ne s'écrivant pas sous la forme d'un produit d'ensembles mesurables. Remarquons que puisque $B \times C = (B \times F) \cap (E \times C)$, la tribu produit $\mathcal{E} \otimes \mathcal{F}$ est aussi engendrée par les pavés très élémentaires $\{B \times F, E \times C : B \in \mathcal{E}, C \in \mathcal{F}\}$.

La construction de la tribu produit d'un produit fini d'espaces mesurables se fait de manière similaire, ou en étendant la cas de 2 facteurs à un nombre plus grand de facteurs en requiérant l'associativité du produit de tribus.

Pour construire la tribu produit d'un produit infini d'espaces mesurables, on procède de même, les pavés élémentaires étant les produits d'ensembles mesurables différant du facteur plein correspondant pour un nombre fini d'indices seulement (voire, pour un seul indice). Notons que pour construire une tribu, sur un ensemble, il n'est pas besoin d'avoir à supposer cet ensemble non vide (il n'y a pas d'axiome du choix ici).

Théorème. — *Soient $((E_i, \mathcal{E}_i))_{i=1}^n$ une collection finie (ou quelconque...) d'espaces mesurables,*

$$(E, \mathcal{E}) = \left(\prod_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{E}_i \right)$$

et $(\Omega, \mathcal{A})$ un espace mesurable. Une application $X : \Omega \to E$ est $\mathcal{A}/\mathcal{E}$-mesurable si et seulement si chacune de ses coordonnées $X_i : \Omega \to E_i$ est $\mathcal{A}/\mathcal{E}_i$-mesurable.

Lemme. — *Les applications coordonnées $f_i : E \to E_i$ sont $\mathcal{E}/\mathcal{E}_i$-mesurables.*

Démonstration du lemme. — Si $B_i \in \mathcal{E}_i$, $f_i^{-1}(B_i) = \prod_{j=1}^{i-1} E_j \times B_i \times \prod_{j=i+1}^n E_j$ qui est un pavé élémentaire et est donc dans $\mathcal{E}$. □

Démonstration du théorème. — Si $X : \Omega \to E$ est $\mathcal{A}/\mathcal{E}$-mesurable, d'après le lemme et par composition, $X_i = f_i \circ X : \Omega \to E_i$ est $\mathcal{A}/\mathcal{E}_i$-mesurable. Réciproquement, supposons les coordonnées X_i $\mathcal{A}/\mathcal{E}_i$-mesurables. La tribu $\tau(X)$ sur E contient tous les $\prod_{j=1}^{i-1} E_j \times B_i \times \prod_{j=i+1}^n E_j$ puisque $X^{-1}(\prod_{j=1}^{i-1} E_j \times B_i \times \prod_{j=i+1}^n E_j) = X_i^{-1}(B_i) \in \mathcal{A}$ par hypothèse. Donc $\mathcal{E} \subset \tau(X)$, ce qui montre que X est $\mathcal{A}/\mathcal{E}$-mesurable. □

Mesurabilité et produit, unicité des mesures, produit de mesures, ...

THÉORÈME. — *Soit $(E, \mathcal{E})$ un espace mesurable. Soit $\mathcal{C} \subset \mathcal{E}$ une classe de parties mesurables stable par intersections finies et engendrant $\mathcal{E}$. Alors deux mesures de probabilité μ_1 et μ_2 sur $(E, \mathcal{E})$ sont égales ($\mu_1(B) = \mu_2(B)$ pour tout $B \in \mathcal{E}$) si et seulement si elles coïncident sur $\mathcal{C}$ ($\mu_1(B) = \mu_2(B)$ pour tout $B \in \mathcal{C}$).*

Remarque. — La condition de stabilité par intersections finies est importante.

F. **Tribus induites**

THÉORÈME. — *Soit $(E, \mathcal{E})$ un espace mesurable et $B \subset E$ un sous-ensemble quelconque de E. Posons $\mathcal{E}_B = \{B \cap C : C \in \mathcal{E}\}$.*

(i) *$\mathcal{E}_B$ est une tribu sur B (elle est nommée tribu trace ou induite de $\mathcal{E}$ sur B).*

(ii) *Soit $(\Omega, \mathcal{A})$ un espace mesurable, $X : \Omega \to B \subset E$ une application. Alors X est $\mathcal{A}/\mathcal{E}$-mesurable si et seulement si X est $\mathcal{A}/\mathcal{E}_B$-mesurable.*

(iii) *Si la tribu $\mathcal{E}$ est engendrée par une famille $(B_i)_{i\in I}$ de E, alors $\mathcal{E}_B$ est engendrée par $(B \cap B_i)_{i\in I}$.*

Démonstration. — (i) On a bien $\mathcal{E}_B \subset \mathcal{P}(B)$, ensuite
— $\varnothing = B \cap \varnothing \in \mathcal{E}_B$ (de même $B = B \cap E \in \mathcal{E}_B$),
— si $B \cap C \in \mathcal{E}_B$, $\complement_B(B \cap C) = B \cap \complement_E C \in \mathcal{E}_B$,
— si $B_1 = B \cap C_1, \ldots, B_n = B \cap C_n, \ldots \in \mathcal{E}_B$, alors $\bigcup_n (B \cap C_n) = B \cap \bigcup_n C_n \in \mathcal{E}_B$ (attention, on se sert ici de l'axiome du choix dénombrable car pour chaque $B_n \in \mathcal{E}_B$, on choisit une écriture de la forme $B \cap C_n$ avec $C_n \in \mathcal{E}$).
Ce qui montre que $\mathcal{E}_B$ est une tribu sur B.

(ii) Supposons X à valeurs dans B et $\mathcal{A}/\mathcal{E}$-mesurable. Alors pour $C \in \mathcal{E}$, $B \cap C \in \mathcal{E}_B$ et $X^{-1}(B \cap C) = X^{-1}(C) \in \mathcal{A}$. Ce qui montre que X est $\mathcal{A}/\mathcal{E}_B$-mesurable. Réciproquement, X est toujours à valeurs dans B et est supposée $\mathcal{A}/\mathcal{E}_B$-mesurable. Pour $C \in \mathcal{E}$, $X^{-1}(C) = X^{-1}(B \cap C) \in \mathcal{A}$. Ce qui montre que X est $\mathcal{A}/\mathcal{E}$-mesurable.

(iii) Puisque pour tout $i \in I$, $B_i \in \mathcal{E}$, on a $B \cap B_i \in \mathcal{E}_B$, et ainsi la tribu $\mathcal{B} = \sigma\{B \cap B_i : i \in I\}$ est incluse dans $\mathcal{E}_B$.

Réciproquement, considérons $\mathcal{D} = \{C \subset E : B \cap C \in \mathcal{B}\}$. Nous souhaitons montrer que $\mathcal{D}$ contient $\mathcal{E}$ ce qui montrera que $\mathcal{E}_B \subset \mathcal{B}$.

Il est clair que $(B_i)_{i\in I} \subset \mathcal{D}$. Montrons que $\mathcal{D}$ est une tribu sur E :
— il est évident que $\varnothing \in \mathcal{D}$;
— si $C \in \mathcal{D}$, $B \cap C \in \mathcal{B}$ donc $\complement_B(B \cap C) = B \cap \complement_E C \in \mathcal{B} = B \cap C^c$ puisque $\mathcal{B}$ est une tribu, ce qui montre que $C^c \in \mathcal{D}$;
— si $C_1, \ldots, C_n, \ldots \in \mathcal{D}$, pour tout n, $B \cap C_n \in \mathcal{B}$, donc $\bigcup_n (B \cap C_n) = B \cap \bigcup_n C_n \in \mathcal{B}$, et ainsi $\bigcup_n C_n \in \mathcal{D}$.
Puisque $\mathcal{D}$ est une tribu contenant $(B_i)_{i\in I}$ elle contient donc $\mathcal{E} = \sigma((B_i)_{i\in I})$. Ainsi, quelque soit $C \in \mathcal{E}$, $C \in \mathcal{D}$, c'est-à-dire $B \cap C \in \mathcal{B}$, soit $\mathcal{E}_B \subset \mathcal{B}$, ce qui montre l'inclusion réciproque. On a finalement l'égalité $\mathcal{E}_B = \mathcal{B} = \sigma((B \cap B_i)_{i\in I})$. □

Remarque. — Soient E un espace topologique, $\mathcal{E} = \mathcal{B}(E)$ sa tribu borélienne et $B \subset E$ une partie quelconque. La topologie induite ou trace de E sur B a pour ouverts $\{B \cap O : O \text{ ouvert de } E\}$ et pour fermés $\{B \cap F : F \text{ fermé de } E\}$. Le point (iii) nous permet d'affirmer que la tribu induite par la tribu borélienne est la tribu borélienne pour la topologie induite. C'est un fait assez remarquable.

G. **Ensembles négligeables, ensembles presque sûrs**

Définition. — Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On appelle ensemble négligeable (ou presque impossible) toute partie $A \subset \Omega$ telle qu'il existe $A' \in \mathcal{A}$ vérifiant $A \subset A'$ et $\mathbb{P}(A') = 0$.

Théorème. — (i) *Une intersection quelconque d'ensembles négligeables est négligeable.*

(ii) *Toute réunion finie ou dénombrable d'ensembles négligeables est négligeable.*

Démonstration. — (i) Soient $(A_i)_{i\in I}$ une collection quelconque d'ensembles négligeables, $i_0 \in I$ un indice et $A'_{i_0} \in \mathcal{A}$ tel que $A_{i_0} \subset A'_{i_0}$ et $\mathbb{P}(A'_{i_0}) = 0$. On a alors $\bigcap_{i\in I} A_i \subset A_{i_0} \subset A'_{i_0}$, ce qui montre que $\bigcap_{i\in I} A_i$ est négligeable.

(ii) Soit $A_1, \dots, A_n, \dots$ une famille finie ou dénombrable d'ensembles négligables. Pour chaque n, choisissons $A'_n \in\in \mathcal{A}$ tel que $A_n \subset A'_n$ et $\mathbb{P}(A'_n) = 0$. On a alors

$$\bigcup_n A_n \subset \bigcup_n A'_n \qquad \text{et} \qquad \mathbb{P}\Big(\bigcup_n A'_n\Big) \leqslant \sum_n \mathbb{P}(A'_n) = \sum_n 0 = 0$$

par l'inégalité de Boole. Donc $\bigcup_n A_n$ est négligeable. □

Définition. — Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On appelle ensemble presque sûr (ou presque certain) toute partie $A \subset \Omega$ telle qu'il existe $A' \in \mathcal{A}$ vérifiant $A \supset A'$ et $\mathbb{P}(A') = 1$.

Théorème. — *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.*

(i) *Un ensemble $A \subset \Omega$ est presque sûr si et seulement si A^c est négligeable.*

(ii) *Une réunion quelconque d'ensembles presque sûrs est presque sûre, et toute intersection finie ou dénombrable d'ensembles presque sûrs est presque sûre.*

(iii) *Un ensemble $A \subset \Omega$ est presque sûr si et seulement $A = A' \cup N$ avec $A' \in \mathcal{A}$, $\mathbb{P}(A) = 1$ et $N \subset \Omega$ négligeable (pouvant être choisi disjoint de A').*

Démonstration. — (i) On a : A presque sûr $\iff$ il existe $A' \in \mathcal{A}$, $A \supset A'$, $\mathbb{P}(A') = 1$ $\iff$ il existe $(A')^c \in \mathcal{A}$, $A^c \subset (A')^c$, $\mathbb{P}((A')^c) = 0$ $\iff$ A^c négligeable.

(ii) On passe au complémentaire sur les résultats obtenus au théorème précédent.

(iii) Si A est presque sûr, soit $A' \in \mathcal{A}$, $A \supset A'$, $\mathbb{P}(A') = 1$. En posant $N = A\backslash A'$, on constate que $N \subset \Omega\backslash A'$ et que $\mathbb{P}(\Omega\backslash A') = \mathbb{P}(\Omega) - \mathbb{P}(A') = 0$. Donc N est négligeable (ici il a été choisi disjoint de A'. Réciproquement, si $A' \in \mathcal{A}$, $\mathbb{P}(A') = 1$ et N est négligeable, alors $A = A' \cup N$ vérifie $A \supset A'$ avec $\mathbb{P}(A') = 1$, et donc A est presque sûr. □

H. **Complétion d'un espace probabilisé**

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Notons $\mathcal{N} = \mathcal{N}(\Omega, \mathcal{A}, \mathbb{P})$ l'ensemble des parties négligeables de Ω (noter qu'elles dépendent de $\mathcal{A}$ et de $\mathbb{P}$). Soit $\bar{\mathcal{A}} = \mathcal{A} \vee \mathcal{N}$ la tribu engendrée par $\mathcal{A}$ et les négligeables.

Lemme. — *Si $A \in \bar{\mathcal{A}}$, alors $A = A' \cup N$ où $A' \in \mathcal{A}$ et $N \in \mathcal{N}$ (qui peuvent être choisis disjoints).*

Démonstration. — (i) Notons $\hat{\mathcal{A}} = \{A \cup N : A \in \mathcal{A}, N \in \mathcal{N}\}$. Il est clair que $\hat{\mathcal{A}} \subset \bar{\mathcal{A}}$. Montrons que $\bar{\mathcal{A}}$ est une tribu contenant $\mathcal{A}$ et $\mathcal{N}$. ..... □

Chapitre II

PROBABILITÉS CONDITIONNELLES, INDÉPENDANCE

1. **Probabilités conditionnelles**

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé modélisant une expérience aléatoire. Soit $B \in \mathcal{A}$ un événement de probabilité strictement positive : $\mathbb{P}(B) > 0$. On veut déduire du modèle $(\Omega, \mathcal{A}, \mathbb{P})$ un modèle pour lequel B est toujours réalisé.

Exemple. — On lance, les yeux bandés, un dé et une autre personne dit que le résultat est pair. On s'intéresse alors, sachant cette information supplémentaire, qu'elle est la probabilité que le résultat soit égal à un nombre donné.

Définition. — Soit $B \in \mathcal{A}$, $\mathbb{P}(B) > 0$. On appelle *mesure de probabilité conditionnelle* sachant l'événement B l'application :

$$\begin{aligned} \mathbb{P}(\,.\,|\,B) : \mathcal{A} &\longrightarrow [0,1] \\ A &\longmapsto \mathbb{P}(A\,|\,B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\,. \end{aligned}$$

Le nombre $\mathbb{P}(A\,|\,B)$ est appelé probabilité (conditionnelle) de (l'événement) A sachant (l'événement) B.

Premières propriétés. — (i) *Soit $B \in \mathcal{A}$, $\mathbb{P}(B) > 0$. L'application $\mathbb{P}(\,.\,|\,B)$ est une mesure de probabilité sur $(\Omega, \mathcal{A})$ portée par B, i.e., si $A \cap B$, $\mathbb{P}(A\,|\,B) = 0$.*

(ii) *Formule de Bayes : si A, $B \in \mathcal{A}$, $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$, alors*

$$\mathbb{P}(B\,|\,A) = \mathbb{P}(A\,|\,B) \times \frac{\mathbb{P}(B)}{\mathbb{P}(A)}\,.$$

(iii) *Soient $A_1, \ldots, A_n \in \mathcal{A}$ tels que $\mathbb{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$. Alors,*

$$\begin{aligned} \mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_n\,|\,A_1 \cap \ldots \cap A_{n-1}) \times \\ \times\, \mathbb{P}(A_{n-1}\,|\,A_1 \cap \ldots \cap A_{n-2}) \times \cdots \times \mathbb{P}(A_3\,|\,A_1 \cap A_2) \times \mathbb{P}(A_2\,|\,A_1) \times \mathbb{P}(A_1). \end{aligned}$$

Démonstration. — Tout est immédiat, à traiter comme exercice sinon. La seconde identité est une règle de chaînage ; elle est souvent nommée « formule des probabilités composées ». □

Remarques. — *a*) $\mathbb{P}(\,.\,|\,B)$ dépend du choix de B mais aussi de $\mathbb{P}$ comme le rappelle la notation.

b) En troisième lecture : $(B, \mathcal{A}_B, \mathbb{P}(\,.\,|\,B)|_{\mathcal{A}_B})$, où $\mathcal{A}_B = \{A \cap B : A \in \mathcal{A}\}$ et $\mathbb{P}(\,.\,|\,B)|_{\mathcal{A}_B}$ est la restriction de $\mathbb{P}(\,.\,|\,B)$ à $\mathcal{A}_B$ (qu'on pourrait être tenté de noter $\mathbb{P}_B$), est un espace probabilisé.

c) Thomas Bayes (1702–1761) était un révérend anglais qui s'était intéressé à la Logique et au Calcul des Probabilités.

Définition. — (i) On appelle *système complet d'événements* toute partition mesurable, finie ou dénombrable $(B_n)_n \subset \mathcal{A}$ de Ω, c'est-à-dire

$$\text{pour tout } n, \quad B_n \in \mathcal{A}, \qquad \text{si } m \neq n, \quad B_m \cap B_n = \varnothing, \qquad \text{et} \qquad \bigcup_n B_n = \Omega.$$

(ii) On appelle *système complet d'événements généralisé* toute suite finie ou dénombrable $(B_n)_n \subset \mathcal{A}$ telle que

$$\text{si } m \neq n, \qquad \mathbb{P}(B_m \cap B_n) = 0 \qquad \text{et} \qquad \mathbb{P}\left(\bigcup_n B_n\right) = 1.$$

Par la suite nous ne nous servirons que de la notion de système complet d'événements. Les énoncés ultérieurs s'appliquent aussi aux systèmes complets d'événements généralisés en adaptant, facilement, les démonstrations à ce cadre plus général.

Convention. — *On convient que lorsque* $\mathbb{P}(A \cap B) = 0$, *le produit* $\mathbb{P}(A \,|\, B) \times \mathbb{P}(B)$ *a toujours un sens et est nul.*

Cette convention est utile lorsque $\mathbb{P}(B) = 0$. Rappelons que si $\mathbb{P}(B) > 0$, l'expression $\mathbb{P}(A \,|\, B) \times \mathbb{P}(B)$ est bien définie et vaut $\mathbb{P}(A \cap B)$. Cette convention ne fait que prolonger cette identité au cas $\mathbb{P}(B) = 0$.

Théorème. — *Soit* $(B_n)_n \subset \mathcal{A}$ *un système complet d'événements.*

(i) *Pour tout* $A \in \mathcal{A}$,

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A \,|\, B_n) \times \mathbb{P}(B_n) \qquad \textit{(formule des probabilités totales).}$$

(ii) *Pour tout* $A \in \mathcal{A}$ *tel que* $\mathbb{P}(A) > 0$, *et tout indice* N *tel que* $\mathbb{P}(B_N) > 0$,

$$\mathbb{P}(B_N \,|\, A) = \mathbb{P}(A \,|\, B_N) \times \frac{\mathbb{P}(B_N)}{\sum_n \mathbb{P}(A \,|\, B_n) \times \mathbb{P}(B_n)} \qquad \textit{(seconde formule de Bayes).}$$

Démonstration. — (i) On a $A = A \cap \Omega = A \cap \bigcup_n B_n = \bigcup_n (A \cap B_n)$ qui est une écriture de l'événement A en réunion d'événements disjoints. On a ainsi par additivité dénombrable

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A \cap B_n) = \sum_n \mathbb{P}(A \,|\, B_n) \times \mathbb{P}(B_n).$$

(ii) On applique la première formule de Bayes, puis celle des probabilités totales :

$$\mathbb{P}(B_N \,|\, A) = \mathbb{P}(A \,|\, B_N) \times \frac{\mathbb{P}(B_N)}{\mathbb{P}(A)} = \mathbb{P}(A \,|\, B_N) \times \frac{\mathbb{P}(B_N)}{\sum_n \mathbb{P}(A \,|\, B_n) \times \mathbb{P}(B_n)}\,. \qquad \square$$

Application : matrices de transition. — Soient

$$E = (x_1, \ldots, x_m) \qquad \text{et} \qquad F = (y_1, \ldots, y_n)$$

deux ensembles finis, et $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{P}(E))$ et $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (F, \mathcal{P}(F))$ deux variables aléatoires discrètes. On suppose que pour tout $i \in \{1, \ldots, m\}$,

$$\mathbb{P}\{X = x_i\} > 0.$$

Les événements $\{X = x_i\}$ forment un système complet d'événements et pour tout $j = 1, \ldots, n$:

$$\mathbb{P}\{Y = y_j\} = \sum_{i=1}^{m} \mathbb{P}\{X = x_i\} \times \mathbb{P}\{Y = y_j \,|\, X = x_i\},$$

ce qui peut s'écrire sous forme matricielle

$$(\mathbb{P}\{Y = y_1\}, \ldots, \mathbb{P}\{Y = y_n\}) = (\mathbb{P}\{X = x_1\}, \ldots, \mathbb{P}\{X = x_m\}) \times \\ \times \begin{pmatrix} \mathbb{P}\{Y = y_1 \,|\, X = x_1\} & \ldots & \mathbb{P}\{Y = y_n \,|\, X = x_1\} \\ \vdots & \ddots & \vdots \\ \mathbb{P}\{Y = y_1 \,|\, X = x_m\} & \ldots & \mathbb{P}\{Y = y_n \,|\, X = x_m\} \end{pmatrix}.$$

La matrice $(\mathbb{P}\{Y = y_j \mid X = x_i\})_{i=1,\dots,m,\, j=1,\dots,n}$ est une *matrice de transition*. Elle permet par produit matriciel de déduire la loi de Y de celle de X. Nous avons écrit les lois des variables aléatoires comme des vecteurs lignes et non comme des vecteurs colonnes. Ceci est justifié par le fait que sur un espace d'état fini, les fonctions numériques sont identifiées à des vecteurs colonnes indexés par l'espace d'état correspondant ; les lois de probabilités agissant comme des formes linéaires sur les fonctions numériques, elles s'écrivent donc comme des vecteurs lignes.

Dans la pratique, il arrive qu'une telle matrice M soit fixée (ses coefficients sont positifs et leurs sommes par colonne est égale à 1, ce sont ce qu'on appelle des *matrices stochastiques*). Elle est le plus souvent carrée, ce qui correspond à un seul espace d'état $E = F$. On part d'un vecteur ligne $(p_1^0, \dots, p_n^0)$, où $p_i^0 = \mathbb{P}\{X_0 = x_i\}$, qui est l'état initial, et les itérations de la matrice appliquée à ce vecteur $(p_1^0, \dots, p_n^0) \times M^k$ s'interprète comme la loi d'une certaine variable aléatoire X_k. Les variables aléatoires $(X_k)_{k \geqslant 0}$ forment alors une chaîne de Markov...

Définition. — Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ une variable aléatoire.

(i) Soit $B \in \mathcal{A}$, $\mathbb{P}(B) > 0$. La loi de X sachant B est la mesure de probabilité sur $(E, \mathcal{E})$ image par X de la mesure de probabilité $\mathbb{P}(\,.\,\mid B)$, elle est notée $\mathrm{Loi}(X \mid B)$.

(ii) Soit $(B_n)_n \subset \mathcal{A}$ un système complet d'événements. La suite $(\mathrm{Loi}(X \mid B_n))_n$ est appelée loi de X sachant le système complet d'événements $(B_n)_n$.

(iii) Soit $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (F, \mathcal{F})$ une variable aléatoire discrète. La loi de X sachant la variable aléatoire discrète Y est la suite $(\mathrm{Loi}(X \mid Y = y)_{y \in F}$, c'est-à-dire la loi de X sachant le système complet d'événements $(\{Y = y\})_{y \in F}$.

Remarques. — *a*) Les lois conditionnelles d'une variable aléatoire $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ sont des mesures de probabilité sur $(E, \mathcal{E})$. Il n'y a aucune raison que leurs supports soient disjoints.

b) Nous n'avons pas supposé que pour un système complet d'événements $(B_n)_n$, on ait $\mathbb{P}(B_n) > 0$ pour tout n. Ceci a pour conséquence que dans la définition précédente, certaines lois conditionnelles ne soient pas définies. Ceci n'a pratiquement aucune importance puisqu'elles interviennent souvent comme des intermédiaires telles que ci-après.

c) On déduit facilement du théorème précédent la relation entre la loi de X et de telles lois conditionnelles :

$$P_X = \sum_n \mathrm{Loi}(X \mid B_n) \times \mathbb{P}(B_n).$$

L'intérêt des lois conditionnelles est qu'il est parfois plus naturel de les déterminer dans un premier temps par rapport à un système complet d'événement convenable et d'ainsi reconstituer la loi de X.

2. **Indépendance**

Discussion. — Soient A, $B \in \mathcal{A}$, $\mathbb{P}(B) > 0$. L'événement A est indépendant de B si savoir que B est réalisé ne change pas la probabilité que A le soit, c'est-à-dire

$$\mathbb{P}(A \mid B) = \mathbb{P}(A),$$

relation qui s'écrit de manière symétrique $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$. Sous cette forme la relation d'indépendance ne nécessite pas la condition $\mathbb{P}(B) > 0$ et est symétrique en les deux événements. On dira ainsi que A et $B \in \mathcal{A}$ sont des événements indépendants si et seulement si :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).$$

Remarque. — Déduire la notion d'indépendance de deux événements de la définition des probabilités conditionnelles (parfois appelée *axiome de Bayes*) nous paraît très spécieux. Nous nous y sommes conformé sans conviction puisque cette démarche devient rapidement boîteuse en considérant plus de deux événements.

Avec la définition mathématique de l'indépendance de deux événements traduisons quelles sont les propriétés que devraient satisfaire trois événements A, B et $C \in \mathcal{A}$ indépendants.

Les événements A, B et C sont indépendants deux à deux, c'est-à-dire que si on prend deux événements parmi les trois, ils doivent être indépendants :

$$\begin{aligned} &A \text{ et } B \text{ sont indépendants}, && \mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B), \\ &B \text{ et } C \text{ sont indépendants}, && \mathbb{P}(B \cap C) = \mathbb{P}(B) \times \mathbb{P}(C), \\ &A \text{ et } C \text{ sont indépendants}, && \mathbb{P}(A \cap C) = \mathbb{P}(A) \times \mathbb{P}(C). \end{aligned}$$

De plus, tout événement doit demeurer indépendant de la conjonction des deux autres :

$$\begin{aligned} &A \text{ et } B \cap C \text{ sont indépendants}, && \mathbb{P}\big(A \cap (B \cap C)\big) = \mathbb{P}(A) \times \mathbb{P}(B \cap C), \\ &B \text{ et } A \cap C \text{ sont indépendants}, && \mathbb{P}\big(B \cap (A \cap C)\big) = \mathbb{P}(B) \times \mathbb{P}(A \cap C), \\ &C \text{ et } A \cap B \text{ sont indépendants}, && \mathbb{P}\big(C \cap (A \cap B)\big) = \mathbb{P}(C) \times \mathbb{P}(A \cap B). \end{aligned}$$

Si l'indépendance deux à deux est satisfaite, les dernières relations se résument à la seule relation $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C)$. Les événements A, B et $C \in \mathcal{A}$ sont indépendants si et seulement si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B), \quad \mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B), \quad \mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B),$$
$$\text{et} \quad \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C).$$

Après cette discussion, la généralisation à n événements est immédiate : le cas de trois événements indique que l'indépendance consiste à ce que les probabilités des intersections soient égales aux produits des probabilités et ce, pour toutes les intersections qu'on peut former avec les événements considérés. Ainsi, pour $n \geqslant 2$, $A_1, \ldots, A_n \in \mathcal{A}$ sont des événements indépendants si et seulement si pour tout $2 \leqslant k \leqslant n$, $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$,

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \times \cdots \times \mathbb{P}(A_{i_k}).$$

On dénombre aisément $2^n - n - 1$ relations à vérifier dans la définition de l'indépendance de n événements (cardinal de l'ensemble des parties de $\{1, \ldots, n\}$ $[2^n]$ moins le cardinal des parties à 1 élément $[n]$ moins le cardinal des parties à 0 élément $[1]$). La généralisation à une famille quelconque d'événements s'énonce ainsi :

Définition. — Soit $(A_i)_{i \in I} \subset \mathcal{A}$ une collection quelconque d'événements. Les événements $(A_i)_{i \in I} \subset \mathcal{A}$ sont *indépendants* si et seulement si $2 \leqslant k \leqslant \operatorname{Card} I$, $k < \infty$, $\{i_1, \ldots, i_k\} \subset I$,

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \times \cdots \times \mathbb{P}(A_{i_k}).$$

Définition. — Soit $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to (E_i, \mathcal{E}_i)$, $i \in I$ une collection quelconque de variables aléatoires définies sur un même espace probabilisé et à valeurs dans des espaces mesurables quelconques. Les variables aléatoires $(X_i)_{i \in I}$ sont *indépendantes* si et seulement si pour tous $B_i \in \mathcal{E}_i$, $i \in I$, les événements $(\{X_i \in B_i\})_{i \in I}$ sont indépendants.

Définition. — Soit $(\mathcal{A}_i)_{i \in I}$ une collection quelconque de sous-tribus de $\mathcal{A}$. Les tribus $(\mathcal{A}_i)_{i \in I}$ sont *indépendantes* si et seulement si, pour tous $A_i \in \mathcal{A}_i$, $i \in I$, les événements $(A_i)_{i \in I}$ sont indépendants.

Remarques. — *a*) L'indépendance est relative à la mesure de probabilité considérée $\mathbb{P}$: deux événements qui sont indépendants pour une mesure de probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{A})$ ne le sont pas nécessairement pour une autre mesure de probabilité $\mathbb{P}'$ sur $(\Omega, \mathcal{A})$.

b) *L'indépendance est une notion distincte de la disjonction.* En effet, si A et B sont disjoints et indépendants on constate immédiatement que $\mathbb{P}(A) = 0$ ou $\mathbb{P}(B) = 0$. Ainsi deux événements de probabilité strictement positive ne peuvent être simultanément indépendants et disjoints.

c) Un événement A est indépendant de lui-même si et seulement si $\mathbb{P}(A) \in \{0, 1\}$. En effet, on a alors

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) \overset{\text{indépendance}}{=} \mathbb{P}(A) \times \mathbb{P}(A) = \mathbb{P}(A)^2,$$

c'est-à-dire $\mathbb{P}(A) \times (\mathbb{P}(A) - 1) = 0$. La réciproque est immédiate.

d) De même, un événement A est indépendant de tout autre événement si et seulement si $\mathbb{P}(A) \in \{0, 1\}$. En effet, si A est indépendant de tout événement, il est indépendant de lui-même et donc $\mathbb{P}(A) \in \{0, 1\}$. Réciproquement, si $\mathbb{P}(A) = 0$, on a pour tout $B \in \mathcal{A}$, $\mathbb{P}(A \cap B) = 0 = \mathbb{P}(A) \times \mathbb{P}(B)$, et si $\mathbb{P}(A) = 1$, on a pour tout $B \in \mathcal{A}$, $\mathbb{P}(A \cap B) = \mathbb{P}(B) = \mathbb{P}(A) \times \mathbb{P}(B)$.

e) La notion d'indépendance de tribus unifie toutes les notions antérieures d'indépendance. Rappelons que si $A \in \mathcal{A}$ est un événement, $\sigma(A) = \{\varnothing, A, A^c, \Omega\}$ est la plus petite tribu contenant A. On montre que l'indépendance d'événements $(A_i)_{i \in I}$ équivaut à celle des tribus $(\sigma(A_i))_{i \in I}$. Enfin il est clair que l'indépendance de variables aléatoires $(X_i)_{i \in I}$ équivaut à celle des tribus $(\sigma(X_i))_{i \in I}$.

Pratique. — 1. L'indépendance avec une variable aléatoire *discrète* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E$ — E est un ensemble au plus dénombrable muni de sa tribu discrète — s'établit en montrant l'indépendance avec les événements de la forme $\{X = x\}$, $x \in E$.

2. L'indépendance avec une variable aléatoire *réelle* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ s'établit en montrant l'indépendance avec les événements de la forme $\{X \in [a, b]\}$, $a \leqslant b \in \mathbb{R}$ (ou $\{X \in [a, b[\}$, $\{X \in]a, b]\}$, $\{X \in]a, b[\}$).

Explications. — 1. Soit $A \in \mathcal{A}$. Supposons que pour tout $B \subset E$, $\mathbb{P}(\{X \in B\} \cap A) = \mathbb{P}\{X \in B\} \cap \mathbb{P}(A)$; alors c'est vrai en particulier pour tout singleton $B = \{x\}$, $x \in E$. Réciproquement, supposons que pour tout $x \in E$, $\mathbb{P}(\{X = x\} \cap A) = \mathbb{P}\{X = x\} \cap \mathbb{P}(A)$; alors pour $B \subset E$ (rappelons que B est au plus dénombrable)

$$\begin{aligned}
\mathbb{P}(\{X \in B\} \cap A) &= \mathbb{P}\left(\left(\bigcup_{x \in B} \{X = x\}\right) \cap A\right) = \mathbb{P}\left(\left(\bigcup_{x \in B} \{X = x\} \cap A\right)\right) \\
&\overset{\text{disjonction}}{=} \sum_{x \in B} \mathbb{P}(\{X = x\} \cap A) \\
&\overset{\text{hypothèse}}{=} \sum_{x \in B} \mathbb{P}\{X = x\} \times \mathbb{P}(A) = \left(\sum_{x \in B} \mathbb{P}\{X = x\}\right) \times \mathbb{P}(A) \\
&\overset{\text{disjonction}}{=} \mathbb{P}\{X \in B\} \times \mathbb{P}(A).
\end{aligned}$$

2. La propriété est admise (la sous-probabilité $B \mapsto \mathbb{P}(\{X \in B\} \cap A)$ sur $\mathbb{R}$ coïncide sur une classe stable par intersections finies engendrant $\mathcal{B}(\mathbb{R})$ avec la sous-probabilité $B \mapsto \mathbb{P}\{X \in B\} \times \mathbb{P}(A)$).

THÉORÈME. — *Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $(\mathcal{A}_i)_{i\in I}$ une famille quelconque de sous-tribus de $\mathcal{A}$, et $(\mathcal{A}'_i)_{i\in I}$ une famille de sous-tribus telle que pour tout $i \in I$, $\mathcal{A}'_i \subset A_i$. Si les tribus $(A_i)_{i\in I}$ sont indépendantes, alors les tribus $(A'_i)_{i\in I}$ sont indépendantes.*

Démonstration. — Ce résultat découle immédiatement de la définition de l'indépendance de tribus. □

THÉORÈME. — *Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to (E_i, \mathcal{E}_i)$, $i \in I$, une famille quelconque de variables aléatoires, et $f_i : (E_i, \mathcal{E}_i) \to (F_i, \mathcal{F}_i)$, $i \in I$, une famille d'applications mesurables. Si les variables aléatoires $(X_i)_{i\in I}$ sont indépendantes, alors les variables aléatoires $(f_i \circ X_i)_{i\in I}$ sont indépendantes.*

Démonstration. — Pour tout $i \in I$, on a l'inclusion de tribus $\sigma(f_i \circ X_i) \subset \sigma(X_i)$. L'énoncé résulte alors de la définition de l'indépendance de variables aléatoires et du théorème précédent. □

Remarque. — On peut aussi énoncer des théorèmes évidents du type : si une collection de tribus, de variables aléatoires, indexée par un ensemble d'indices I est indépendante, alors pour tout sous-ensemble d'indices $J \subset I$, la sous-collection indexée par J est indépendante. Par ailleurs on pourrait aussi définir l'indépendance d'événements et de variables aléatoires, d'événements et de tribus, de variables aléatoires et de tribus, ...

Chapitre III

VARIABLES ALÉATOIRES RÉELLES

1. **Généralités**

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $E \subset \mathbb{R}$. Une application $X : \Omega \to E$ est une variable aléatoire réelle si et seulement si l'une des conditions suivantes est vérifiée :

(i) $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une variable aléatoire ;

(ii) $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{B}(E))$ où $\mathcal{B}(E) = \{B \cap E : B \in \mathcal{B}(\mathbb{R})\}$ est une variable aléatoire.

Remarques. — *a*) Lorsque $E \subset \mathbb{R}$ est au plus dénombrable, alors $\mathcal{B}(E) = \mathcal{P}(E)$ et une variable aléatoire à valeurs dans E est une variable aléatoire discrète.

b) Lorsque E est un intervalle de $\mathbb{R}$, la tribu $\mathcal{B}(E)$ introduite ici $\mathcal{B}(E)$ coïncide avec la tribu $\mathcal{B}(E)$ définie au premier chapitre.

Démonstration. — Soit $X : \Omega \to E$. Il est clair que $\mathcal{B}(E)$ est une tribu sur E.

Supposons $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable. Alors pour tout $B \in \mathcal{B}(E)$, $X^{-1}(B) = X^{-1}(B \cap E) \in \mathcal{A}$, donc $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{B}(E))$ est mesurable.

Réciproquement, si $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{B}(E))$ est mesurable, alors pour tout $B \in \mathcal{B}(\mathbb{R})$, $X^{-1}(B) = X^{-1}(B \cap E) \in \mathcal{A}$, donc $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable. □

Théorème et définition. — *Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire réelle. On appelle* fonction de répartition *de la loi de X l'application*

$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto F_X(x) = \mathbb{P}\{X \leqslant x\} = P_X(]-\infty, x]). \end{aligned}$$

(i) *L'application F_X est croissante, continue à droite, limitée à gauche, et vérifie*

$$\lim_{-\infty} F_X = 0 \qquad \textit{et} \qquad \lim_{+\infty} F_X = 1.$$

(ii) *L'application F_X caractérise la loi de la variable aléatoire X : si Y est une variable aléatoire réelle telle que $F_X = F_Y$ alors $P_X = P_Y$ (identité de mesures de probabilité sur $\mathbb{R}$).*

Explications. — (i) Il est clair que F_X est une application croissante : si $x' \leqslant x''$, alors $]-\infty, x'] \subset]-\infty, x'']$ et donc

$$P_X(]-\infty, x']) = F_X(x') \leqslant F_X(x'') = P_X(]-\infty, x'']).$$

Pour la continuité à droite, soit $(x_n)_n \searrow x \in \mathbb{R}$ une suite décroissante et convergente dans $\mathbb{R}$; alors $]-\infty, x_n] \searrow]-\infty, x]$ et ainsi, par continuité de la mesure de probabilité P_X le long des suites décroissantes d'ensembles mesurables $(F_X(x_n) = P_X(]-\infty, x_n]))_n \searrow P_X(]-\infty, x]) = F_X(x)$ (*voir* chap. I, § 3, théorème (iv-bis)).

Pour l'existence de limites à gauche, soit $(x_n)_n \uparrow x \in \mathbb{R}$ une suite strictement croissante et convergente dans $\mathbb{R}$; alors $]-\infty, x_n] \uparrow]-\infty, x[$ et ainsi, par continuité de la mesure de probabilité P_X le long des suites croissantes d'ensembles mesurables $(F_X(x_n) = P_X(]-\infty, x_n]))_n \uparrow P_X(]-\infty, x[) = F_X(x-)$ (*voir* chap. I, § 3, théorème (iv)).

Le fait que $\lim_{-\infty} F_X = 0$ et $\lim_{+\infty} F_X = 1$ est clair.

(ii) Le résultat est admis. On notera que si $F_X(x) = F_Y(x)$ pour tout $x \in \mathbb{R}$, alors pour tous $a \leqslant b$, $P_X(]a,b]) = F_X(b) - F_X(a) = P_Y(]a,b])$. Donc les mesures de probabilité P_X et P_Y coïncident sur la classe des intervalles $\{]a,b] : a \leqslant b\}$ — classe stable par intersections finies engendrant $\mathcal{B}(\mathbb{R})$ — et sont donc égales.

Affirmation. — *À toute fonction $F : \mathbb{R} \to [0,1]$ vérifiant le point (i) précédent est associée une unique mesure de probabilité μ_F sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

Définition. — La loi d'une variable aléatoire $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est *absolument continue* si et seulement si il existe une application $p_X : \mathbb{R} \to \mathbb{R}_+$ *intégrable*[1] telle que

$$F_X(a) = \int_{-\infty}^{a} p_X(x)\,\mathrm{d}x, \qquad \text{pour tout } a \in \mathbb{R}.$$

Remarque. — Si $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ est de loi absolument continue alors :

1. F_X est continue, i.e., $\mathbb{P}\{X = x\} = F_X(x) - F(x-) = 0$ pour tout $x \in \mathbb{R}$.
2. Si p_X est comme dans la définition précédente, alors

$$\int_{-\infty}^{+\infty} p_X(x)\,\mathrm{d}x = 1 = \lim_{x \to +\infty} F_X(x).$$

Une telle application p_X est appelée *densité* de la loi de X. Il n'y a pas unicité d'une telle application (il suffit par exemple de modifier p_X en un nombre fini ou dénombrable de points).

Remarque. — Une variable aléatoire réelle $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ est discrète si et seulement si F_X est une application en escalier (généralisée) :

$$F_X(x) = \sum_{x' \text{ valeur possible de } X,\ x' \leqslant x} \mathbb{P}\{X = x'\}, \qquad \text{pour tout } x \in \mathbb{R}.$$

Approximations discrètes. — Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire réelle. Soit $\varepsilon > 0$. Définissons $X_\varepsilon : (\Omega, \mathcal{A}, \mathbb{P}) \to \varepsilon\mathbb{Z} \subset \mathbb{R}$ par

$$X_\varepsilon(\omega) = \varepsilon k \qquad \text{si} \qquad X_\varepsilon(\omega) \in [\varepsilon k, \varepsilon(k+1)[,\ k \in \mathbb{Z}.$$

L'application X_ε est une variable aléatoire discrète. En effet ses valeurs possibles sont contenues dans $\varepsilon\mathbb{Z}$ et on a, pour tout $k \in \mathbb{Z}$, $X_\varepsilon^{-1}(\varepsilon k) = X^{-1}([\varepsilon k, \varepsilon(k+1)[) \in \mathcal{A}$ puisque X est une variable aléatoire réelle. De plus on a par construction

$$X_\varepsilon(\omega) \leqslant X(\omega) < X_\varepsilon(\omega) + \varepsilon.$$

La variable aléatoire X_ε approche X par en dessous et uniformément à ε près.

Principe. — *Ce qui marche pour les variables aléatoires réelles discrètes marche pour les variables aléatoires générales par passage à la limite.*

Exercice. — Calculer, lorsque c'est possible, les fonctions de répartition associées aux exemples de lois discrètes et absolument continues donnés au chapitre I.

---

1. Le mot intégrable peut être pris au sens de Riemann-intégrable même si Lebesgue-mesurable suffirait si l'intégrale de Lebesgue était connue.

2. **Intégration de variables aléatoires réelles discrètes**

Préambule. — Les connaissances de base sur les séries numériques sont supposées connues et maîtrisées. La convergence d'une série numérique $\sum_n a_n$ dépend parfois de l'ordre de sommation. Nous nous placerons sous des hypothèses telles que ce ne soit pas le cas :

(i) la série $\sum_n a_n$ est absolument convergente ;

(ii) l'une, au moins, des séries à termes positifs $\sum_n a_n^+$ et $\sum_n a_n^-$ est convergente (on a alors $\sum_n a_n = \sum_n a_n^+ - \sum_n a_n^- \in \bar{\mathbb{R}}$) ;

la condition (i) s'exprime aussi en disant que $\sum_n a_n$ est *sommable dans* $\mathbb{R}$, la condition (ii) en disant que $\sum_n a_n$ est *sommable dans* $\bar{\mathbb{R}}$. La condition (i) est évidemment plus forte que (ii), et si (ii) est satisfaite, la somme de la série ne dépend pas de l'ordre de sommation.

Définition. — Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E \subset \mathbb{R}$ une variable aléatoire réelle discrète.

(i) L'*espérance*, ou la *moyenne*, de la variable aléatoire X est *définie* si et seulement si la série

$$\sum_{x \in E} x \times \mathbb{P}\{X = x\}$$

est sommable dans $\bar{\mathbb{R}}$. Dans ce cas, l'*espérance*, ou la *moyenne*, de X est égale à la somme de cette série (dans $\bar{\mathbb{R}}$) et notée $\mathbb{E}[X]$.

(ii) La variable aléatoire X est *intégrable* si et seulement si la série

$$\sum_{x \in E} x \times \mathbb{P}\{X = x\}$$

est sommable dans $\mathbb{R}$.

Exemples. — a) Si X est une variable aléatoire réelle, discrète et bornée, alors X est intégrable. En effet, si il existe une constante C telle que pour tout $x \in E$, $|x| \leqslant C$, alors

$$\sum_{x \in E} \big| x \times \mathbb{P}\{X = x\} \big| \leqslant \sum_{x \in E} C \times \mathbb{P}\{X = x\} = C \sum_{x \in E} \mathbb{P}\{X = x\} = C \times 1 = C.$$

En particulier, si X n'a qu'un nombre fini de valeurs possibles, alors X est intégrable.

b) Si X est de signe constant, ou plus généralement n'a qu'un nombre fini de valeurs possibles d'un signe donné, alors son espérance est définie.

Lemme. — *Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E \subset \mathbb{R}$ variable aléatoire réelle discrète à valeurs dans un ensemble E au plus dénombrable, telle que $\mathbb{E}[X]$ est définie.*

Soit $(B_n)_n$ un système complet d'événements plus fin que $(\{X = x\})_{x \in E}$ (pour tout n, il existe $x_n \in E$ tel que $B_n \subset \{X = x_n\}$). Alors

$$\mathbb{E}[X] = \sum_n x_n \times \mathbb{P}(B_n)$$

où $x_n = X|_{B_n}$ est l'unique valeur prise par X sur l'événement B_n, et la somme précédente est bien définie.

Démonstration. — Admettons pour l'instant la bonne définition des sommes que nous allons manipuler ce qui revient à considérer que ces sommes seront absolument convergentes. On a

$$\sum_n x_n \times \mathbb{P}(B_n) = \sum_{x \in E} \sum_{\{n \,:\, x_n = x\}} x_n \times \mathbb{P}(B_n) = \sum_{x \in E} x \sum_{\{n \,:\, x_n = x\}} \mathbb{P}(B_n)$$

$$\overset{\text{disjonction}}{=} \sum_{x \in E} x \times \mathbb{P}\left(\bigcup_{\{n \,:\, x_n = x\}} B_n \right) = \sum_{x \in E} x \times \mathbb{P}\{X = x\} = \mathbb{E}[X].$$

L'hypothèse selon laquelle $\mathbb{E}[X]$ est bien définie sert dans la première égalité : l'ordre de sommation n'intervient pas. En prenant la suite d'égalités dans l'autre sens on constate que $\sum_n x_n \times \mathbb{P}(B_n)$ est bien définie. □

Remarques. — *a)* Si $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est un espace probabilisé discret, on a

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega) \times \mathbb{P}\{\omega\}$$

pour toute application $X : \Omega \to \mathbb{R}$ (c'est ici automatiquement une variable aléatoire, discrète) intégrable (appliquer le lemme précédent avec le système complet d'événements $(\{\omega\})_{\omega\in\Omega}$).

b) Si $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est un espace probabilisé classique, on a

$$\mathbb{E}[X] = \frac{1}{\operatorname{Card}\Omega} \sum_{\omega\in\Omega} X(\omega)$$

pour toute application $X : \Omega \to \mathbb{R}$ (qui est ici automatiquement une variable aléatoire, discrète, intégrable).

Théorème de transfert, cas discret. — *Soient $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E$ — E un ensemble au plus dénombrable — une variable aléatoire discrète (non nécessairement réelle) et $f : E \to \mathbb{R}$ une application (quelconque). Alors $f(X) : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ est une variable aléatoire discrète. On a :*

$$\mathbb{E}[|f(X)|] = \sum_{x\in E} |f(x)| \times \mathbb{P}\{X = x\} = \sum_{x\in E} |f(x)| \times P_X\{x\},$$

et si $\mathbb{E}[f(X)]$ est définie, alors

$$\mathbb{E}[f(X)] = \sum_{x\in E} f(x) \times \mathbb{P}\{X = x\} = \sum_{x\in E} f(x) \times P_X\{x\}.$$

Remarque. — La variable aléatoire réelle discrète $f(X)$ est intégrable si et seulement si l'application f est intégrable par rapport à P_X.

Démonstration. — Puisque E est muni de sa tribu discrète $\mathcal{P}(E)$, l'application $f : E \to \mathbb{R}$ est mesurable, et le cardinal de $f(E)$ est inférieur ou égal à celui de E. Ainsi $Y = f(X)$ est une variable aléatoire, réelle, discrète.

Supposons pour simplifier que Y est une variable aléatoire positive. Le système complet d'événements $(\{X = x\})_{x\in E}$ est plus fin que $(\{Y = y\})_{y\in F}$ où $F = f(E)$, et on a pour tout $y \in F$:

$$\{Y = y\} = \bigcup_{\{x\in E\,:\,f(x)=y\}} \{X = x\}.$$

D'après le lemme :

$$\mathbb{E}[Y] = \sum_{x\in E} f(x) \times \mathbb{P}\{X = x\}.$$

Le cas général s'ensuit. □

3. **Intégration de variables aléatoires réelles**

Une variable aléatoire réelle $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ est intégrable si et seulement si pour tout $\varepsilon > 0$ (un seul suffit) son approximation discrète X_ε est intégrable.

Dans ce cas, on montre que $(\mathbb{E}[X_\varepsilon])_{\varepsilon>0}$ admet une limite lorsque $\varepsilon \to 0$. C'est l'espérance de la variable aléatoire X, elle est notée $\mathbb{E}[X]$.

Propriétés fondamentales. — (i) *Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}_+$ une variable aléatoire positive. Alors $\mathbb{E}[X]$ est définie dans $\bar{\mathbb{R}}_+$ et $\mathbb{E}[X] \geqslant 0$ (positivité de l'espérance).*

Si $\mathbb{E}[X] = 0$, alors $X = 0$ presque sûrement, c'est-à-dire $\mathbb{P}\{X = 0\} = 1$.

(ii) *Soient X, $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ deux variables aléatoires réelles intégrables. Alors pour tous α, β et $\gamma \in \mathbb{R}$, $\alpha X + \beta Y + \gamma$ est une variable aléatoire, intégrable, et*

$$\mathbb{E}[\alpha X + \beta Y + \gamma] = \alpha\, \mathbb{E}[X] + \beta\, \mathbb{E}[Y] + \gamma \qquad (\textit{linéarité de l'espérance}).$$

(iii) *Soient X, $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ deux variables aléatoires réelles intégrables. Si pour (presque) tout $\omega \in \Omega$, $X(\omega) \leqslant Y(\omega)$, alors*

$$\mathbb{E}[X] \leqslant \mathbb{E}[Y] \qquad \textit{(monotonie de l'espérance)}.$$

(iv) *Soient X, $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ deux variables aléatoires réelles intégrables. Si X et Y sont indépendantes, alors la variable aléatoire $X \times Y$ est intégrable et on a*

$$\mathbb{E}[X \times Y] = \mathbb{E}[X] \times \mathbb{E}[Y].$$

Remarque. — La seconde partie de la propriété (i) est souvent utilisée pour les variables aléatoires $|X|$ et X^2 pour X une variable aléatoire réelle non nécessairement positive. Le fait que $|X|$ et X^2 soient alors des variables aléatoires provient de la continuité des applications de $\mathbb{R}$ dans $\mathbb{R}_+$ définies par $x \mapsto |x|$ et $x \mapsto x^2$. Cela sera établi dans le chapitre suivant. C'est dans ce même chapitre qu'on trouvera la justification du fait qu'une combinaison affine, ou plus généralement un polynôme, de variables aléatoires réelles est une aléatoire réelle, etc.

Démonstration (uniquement dans le cas discret). — (i) Si $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E \subset \mathbb{R}_+$ est une variable aléatoire discrète positive, il est clair que $\mathbb{E}[X] = \sum_{x \in E} x \times \mathbb{P}\{X = x\} \geqslant 0$. Si dans ce cas $\mathbb{E}[X] = 0$, alors pour tout $x \in E$, $x > 0$, $\mathbb{P}\{X = x\} = 0$, donc $\mathbb{P}\{X = 0\} = 1$.

(ii) Si $\alpha \in \mathbb{R}$, αX est une variable aléatoire discrète et $\mathbb{E}[\alpha X] = \alpha\, \mathbb{E}[X]$ (théorème de transfert avec $f(x) = \alpha x$).

Si $X = \gamma$ est une variable aléatoire constante (n'ayant qu'une seule valeur possible), il est clair que $\mathbb{E}[X] = \mathbb{E}[\gamma] = \gamma$.

Si $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E$ et $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to F$ sont deux variables aléatoires réelles discrètes, alors $Z = X + Y$ est une variable aléatoire réelle discrète car l'ensemble G de ses valeurs possibles est au plus dénombrable et pour tout $z \in G$, on a :

$$\{Z = z\} = \bigcup_{\{x \in E,\, y \in F\, :\, x+y=z\}} \{X = x,\, Y = y\} = \bigcup_{\{x \in E,\, y \in F\, :\, x+y=z\}} \{X = x\} \cap \{Y = y\}$$

qui est une réunion au plus dénombrable d'éléments de la tribu $\mathcal{A}$ et qui est donc encore dans $\mathcal{A}$. Puisque $(\{X = x,\, Y = y\})_{x \in E,\, y \in F}$ est un système complet d'événements plus fin que

respectivement $(\{X = x\})_{x\in E}$ et $(\{Y = y\})_{y\in F}$, on a d'après le lemme

$$\begin{aligned}
\mathbb{E}[X] + \mathbb{E}[Y] &= \Bigg(\sum_{x\in E,\, y\in F} x\,\mathbb{P}\{X = x,\, Y = y\}\Bigg) + \Bigg(\sum_{x\in E,\, y\in F} y\,\mathbb{P}\{X = x,\, Y = y\}\Bigg) \\
&= \sum_{x\in E,\, y\in F} (x + y)\,\mathbb{P}\{X = x,\, Y = y\} \\
&= \sum_{z\in G} z \sum_{\{x\in E,\, y\in F\,:\, x+y=z\}} \mathbb{P}\{X = x,\, Y = y\} \\
&\overset{\text{disjonction}}{=} \sum_{z\in G} z\,\mathbb{P}\Bigg(\bigcup_{\{x\in E,\, y\in F\,:\, x+y=z\}} \{X = x,\, Y = y\}\Bigg) \\
&= \sum_{z\in G} z\,\mathbb{P}\{Z = z\} = \mathbb{E}[Z],
\end{aligned}$$

car $(\{X = x,\, Y = y\})_{x\in E,\, y\in F}$ est un système complet d'évéments plus fin que $(\{Z = z\})_{z\in G}$.

(iii) Si $X \leqslant Y$ sont deux variables aléatoires réelles discrètes intégrables, alors $Y - X$ est une variable aléatoire discrète positive, donc d'après (ii) puis (i) :

$$\mathbb{E}[Y] - \mathbb{E}[X] \overset{\text{(ii)}}{=} \mathbb{E}[Y - X] \overset{\text{(i)}}{\geqslant} 0.$$

(iv) Si $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E$ et $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to F$ sont deux variables aléatoires réelles discrètes, alors $Z = X \times Y$ est encore une variable aléatoire discrète puisque l'ensemble G de ses valeurs possibles est au plus dénombrable et pour tout $z \in G$, on a :

$$\{Z = z\} = \bigcup_{\{x\in E,\, y\in F\,:\, x\times y=z\}} \{X = x,\, Y = y\} = \bigcup_{\{x\in E,\, y\in F\,:\, x\times y=z\}} \{X = x\} \cap \{Y = y\}$$

qui est une réunion au plus dénombrable d'éléments de la tribu $\mathcal{A}$ et qui est donc encore dans $\mathcal{A}$.

Supposons de plus X et Y intégrables et indépendantes. Ainsi qu'au (ii) :

$$\begin{aligned}
\mathbb{E}[X] \times \mathbb{E}[Y] &\overset{\text{définition}}{=} \Bigg(\sum_{x\in E,\, y\in F} x\,\mathbb{P}\{X = x\}\Bigg) \times \Bigg(\sum_{x\in E,\, y\in F} y\,\mathbb{P}\{Y = y\}\Bigg) \\
&\overset{\substack{\text{produit de séries} \\ \text{absolument convergentes}}}{=} \sum_{x\in E,\, y\in F} (x \times y)\,\mathbb{P}\{X = x\} \times \mathbb{P}\{Y = y\} \\
&\overset{\text{indépendance}}{=} \sum_{x\in E,\, y\in F} (x \times y)\,\mathbb{P}\{X = x,\, Y = y\} \\
&= \sum_{z\in G} z \sum_{\{x\in E,\, y\in F\,:\, x\times y=z\}} \mathbb{P}\{X = x,\, Y = y\} \\
&\overset{\text{disjonction}}{=} \sum_{z\in G} z\,\mathbb{P}\Bigg(\bigcup_{\{x\in E,\, y\in F\,:\, x\times y=z\}} \{X = x,\, Y = y\}\Bigg) \\
&= \sum_{z\in G} z\,\mathbb{P}\{Z = z\} = \mathbb{E}[Z],
\end{aligned}$$

ce qui termine la démonstration. □

Théorème de transfert. — *Soient $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ une variable aléatoire (quelconque) et $f : (E, \mathcal{E}) \to \mathbb{R}$ une application mesurable. Alors $f(X) : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ est une*

variable aléatoire réelle. On a

$$\mathbb{E}[|f(X)|] = \int_E |f(x)|\,\mathbb{P}\{X \in \mathrm{d}x\} = \int_E |f(x)|\,P_X(\mathrm{d}x),$$

et si $\mathbb{E}[X]$ *est définie, alors*

$$\mathbb{E}[f(X)] = \int_E f(x)\,\mathbb{P}\{X \in \mathrm{d}x\} = \int_E f(x)\,P_X(\mathrm{d}x).$$

Cas particuliers. — a) Le cas où X est une variable aléatoire discrète a été vu précédemment.

b) Si $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ est une variable aléatoire réelle de loi absolument continue admettant pour densité $p_X : \mathbb{R} \to \mathbb{R}_+$, alors

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\,p_X(x)\,\mathrm{d}x, \qquad \text{i.e.,} \qquad \mathbb{P}\{X \in \mathrm{d}x\} = P_X(\mathrm{d}x) = p(x)\,\mathrm{d}x.$$

Remarques. — a) On ne calcule que très rarement une espérance par passage à la limite sur une discrétisation d'une variable aléatoire réelle. En général, on identifie la loi P_X de la variable aléatoire réelle X et on calcule :

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\,P_X(\mathrm{d}x) = \begin{cases} \displaystyle\sum_x x\,\mathbb{P}\{X = x\} & \text{si } X \text{ est discrète,} \\ \displaystyle\int_{\mathbb{R}} x\,p_X(x)\,\mathrm{d}x & \text{si } P_X \text{ admet une densité } p_X\,; \end{cases}$$

et plus généralement

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\,P_X(\mathrm{d}x) = \begin{cases} \displaystyle\sum_x f(x)\,\mathbb{P}\{X = x\} & \text{si } X \text{ est discrète,} \\ \displaystyle\int_{\mathbb{R}} f(x)\,p_X(x)\,\mathrm{d}x & \text{si } P_X \text{ admet une densité } p_X. \end{cases}$$

b) Il arrive aussi qu'on fasse de tels calculs en se servant de la linéarité, d'une indépendance éventuelle, des expressions figurant dans l'espérance.

Exemple (variables binomiales). — Soient $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to \{0, 1\}$, $i = 1, \ldots, n$, n variables aléatoires indépendantes et de même loi : $\mathbb{P}\{X_i = 1\} = p = 1 - \mathbb{P}\{X_i = 0\}$ pour tout $i = 1, \ldots, n$ et $p \in [0, 1]$ fixé (la loi de chaque X_i est la loi de Bernoulli de paramètre p, elle est notée $\mathcal{B}(1, p)$).

Si on pose $A_i = \{X_i = 1\}$, les événements $(A_i)_{i=1}^n$ sont indépendants et $\mathbb{P}(A_i) = p$. De plus, les variables aléatoires X_i sont les indicatrices de ces événements.

Alors, $Y = \sum_{i=1}^n X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to \{0, 1, \ldots, n\}$ est une variable aléatoire discrète dont la loi est donnée par

$$P_Y\{k\} = \mathbb{P}\{Y = k\} = \begin{cases} C_n^k\,p^k(1-p)^{n-k} & \text{pour } k \in \{0, 1, \ldots, n\}, \\ 0 & \text{sinon,} \end{cases}$$

qui est la loi binomiale de paramètres n et p et qui est notée $\mathcal{B}(n, p)$.

Démonstration. — La preuve de ceci peut être conduite par récurrence sur n ou directement tant elle est simple. □

Remarque. — *a*) La formule du binôme de Newton :

$$(x+y)^n = \sum_{k=0}^{n} C_n^k x^k y^{n-k},$$

valable dans tout anneau commutatif, est à connaître. On déduit de celle-ci que la somme des probabilités décrites plus haut est 1 :

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

b) On a $X_i = \mathbb{1}_{A_i}$ et ainsi $\mathbb{E}[X_i] = \mathbb{E}[\mathbb{1}_{A_i}] = \mathbb{P}(A_i) = p$. Ainsi,

$$\mathbb{E}[Y] = \mathbb{E}\bigg[\sum_{i=1}^{n} X_i\bigg] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n \times p.$$

c) On pourrait vérifier que $\mathbb{E}[Y] = np$ en établissant directement que

$$np = \sum_{k=0}^{n} k \times C_n^k\, p^k (1-p)^{n-k} = \sum_{k=0}^{n} k \times \mathbb{P}\{Y=k\}.$$

4. **Un peu d'analyse**

Proposition et définition. — *Une variable aléatoire réelle* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ *est de puissance* p-ième intégrable, $p \geqslant 1$, *si et seulement si*

$$\mathbb{E}[|X|^p] < \infty.$$

Si une variable aléatoire réelle X *est de puissance* p*-ième intégrable,* $p \geqslant 1$, *alors* X *est de puissance* q*-ième intégrable pour tout* $1 \leqslant q \leqslant p$.

Nous notons $\mathcal{L}(\Omega, \mathcal{A}, \mathbb{P}\,;\, \mathbb{R})$, *ou s'il n'y a pas d'ambigüité* $\mathcal{L}^p$, *l'ensemble des variables aléatoires de puissance* p*-ième intégrable, et* $\|X_p\| = \mathbb{E}[|X|^p]^{1/p}$ *pour toute variable aléatoire* $X \in \mathcal{L}^p$.

Démonstration. — Soit $q \in [1, p]$. Pour tout $x \in \mathbb{R}$ on a $|x|^q \leqslant 1 + |x|^p$, par conséquent on a l'inégalité entre variables aléatoires $|X|^q \leqslant 1 + |X|^p$. En intégrant cette inégalité, on a par monotonie,

$$\mathbb{E}[|X|^q] \leqslant \mathbb{E}[1 + |X|^p] = 1 + \mathbb{E}[|X|^p]$$

d'où la conclusion. □

Remarque. — Pour $p = 2$, on utilise l'expression *variable aléatoire de carré intégrable*. Pour $p = 1$, c'est simplement la notion de variable aléatoire intégrable.

Inégalité de Jensen. — *Soient* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ *une variable aléatoire réelle intégrable et* $\phi : \mathbb{R} \to \mathbb{R}$ *une fonction convexe. Alors*

$$\phi(\mathbb{E}[X]) \leqslant \mathbb{E}[\phi(X)] \in \mathbb{R} \cup \{+\infty\}.$$

Démonstration. — Une fonction $\phi : \mathbb{R} \to \mathbb{R}$ convexe étant continue, elle est mesurable lorsque $\mathbb{R}$ est muni de sa tribu borélienne, ainsi $\phi(X)$ est encore une variable aléatoire réelle.

Supposons que $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E \subset \mathbb{R}$ est une variable aléatoire discrète. L'inégalité affirmée précédemment s'écrit alors

$$\phi\bigg(\sum_{x \in E} x \times \mathbb{P}\{X=x\}\bigg) \leqslant \sum_{x \in E} \phi(x) \times \mathbb{P}\{X=x\}$$

qui est une conséquence usuelle de la convexité de l'application ϕ. Remarquer que nous savons que le terme de gauche de l'inégalité est fini, le terme de droite est alors bien défini dans $\mathbb{R} \cup \{+\infty\}$.

Pour le cas d'une variable aléatoire réelle générale, il suffit de remarquer que la fonction convexe ϕ est l'enveloppe supérieure des fonctions affines qu'elle domine. Ainsi, si $(a, b) \in \mathbb{R}^2$ sont tels que, pour tout $x \in \mathbb{R}$, $ax + b \leqslant \phi(x)$ on a, d'une part par linéarité, d'autre part par monotonie,

$$a\,\mathbb{E}[X] + b = \mathbb{E}[aX + b] \leqslant \mathbb{E}[\phi(X)] \in \mathbb{R} \cup \{+\infty\}.$$

En passant au supremum sur le membre de gauche parmi de tels couples (a, b), on obtient la valeur de ϕ évaluée en $\mathbb{E}[X]$, et obtient donc l'inégalité voulue.

Formulons une nouvelle explication. Soit $x = \mathbb{E}[X]$. La fonction ϕ étant convexe, elle est au dessus de sa tangente en x lorsqu'elle est définie, ou du moins au dessus d'une droite passant par $(x, \phi(x))$ lorsque ϕ n'est pas dérivable en x. Soient (a, b) les coefficients d'une telle droite. Nous avons $ax + b = \phi(x)$, soit $b = \phi(x) - ax$. Puisque ϕ est au dessus de cette droite, $ax' + b \leqslant \phi(x')$ pour tout $x' \in \mathbb{R}$. Ainsi, $aX + b \leqslant \phi(X)$. En intégrant, $\phi(\mathbb{E}[X]) = \phi(x) = ax + b = a\,\mathbb{E}[X] + b = \mathbb{E}[aX + b] \leqslant \mathbb{E}[\phi(X)]$, soit $\phi(\mathbb{E}[X]) \leqslant \mathbb{E}[\phi(X)]$, ce qu'il fallait démontrer. □

Remarque. — Dans la plupart des textes académiques, une prudence excessive est faite autour de cette inégalité. Il est notamment très fréquent qu'on ajoute l'hypothèse selon laquelle $\phi(X)$ doit être intégrable. Nous constatons avec cette démonstration qu'il suffit que X soit intégrable, $\mathbb{E}[\phi(X)]$ ayant un sens dans ce cas. Si $\mathbb{E}[X]$ a un sens mais est infinie, l'usage de cette inégalité ne semble pas avoir beaucoup d'intérêt, elle en demeure néanmoins vraie.

Lemme. — *Soient $p, q > 1$ deux réels tels que $1/p + 1/q = 1$ (exposants conjugués). Pour tout $x, y \geqslant 0$, on a $xy \leqslant x^p/p + y^q/q$, avec égalité si et seulement si $x^p = y^q$.*

Démonstration. — Si $x = 0$ ou $y = 0$, l'inégalité est évidente. Soit $y > 0$ fixé. Considérons la fonction $\phi : x \in \mathbb{R}_+^* \mapsto xy - x^p/p$ (noter qu'elle s'impose naturellement compte tenu de l'inégalité à démontrer). On a $\phi'(x) = y - x^{p-1}$, ce qui montre que ϕ croît strictement sur $]0, y^{1/(p-1)}[$, et décroît strictement sur $]y^{1/(p-1)}, +\infty[$. Notons que $1/(p-1) = q/p$ et que $\phi(y^{1/(p-1)}) = \phi(y^{q/p}) = y^q - y^q/p = y^q/q$. De l'inégalité $\phi(x) \geqslant \phi(y^{1/(p-1)}) = y^q/q$, on a immédiatement l'inégalité du lemme, et on a égalité si et seulement si $x = y^{1/(p-1)} = y^{q/p}$, autrement dit si $x^p = y^q$. □

Inégalité de Hölder. — *Soient $p, q > 1$ deux réels tels que $1/p + 1/q = 1$ (exposants conjugués). Si X est une variable aléatoire de puissance p-ième intégrable, Y de puissance q-ième intégrable, alors XY est une variable aléatoire intégrable et*

$$\|XY\|_1 \leqslant \|X\|_p \times \|Y\|_q$$

avec égalité si et seulement si $|X|^p$ et $|Y|^q$ sont proportionnelles presque sûrement.

Démonstration. — Si $\|X\|_p = 0$ ou $\|Y\|_q$, alors $X = 0$ ou $Y = 0$ presque sûrement auquel cas $XY = 0$ presque sûrement et l'inégalité cherchée se résume à $0 \leqslant 0$. Nous pouvons donc supposer que $\|X\|_p > 0$ et $\|Y\|_q > 0$ et sont finis. En appliquant l'inégalité du lemme à $x = |X|/\|X\|_p$ et $y = |Y|/\|Y\|_p$, il vient

$$\frac{|XY|}{\|X\|_p \times \|Y\|_q} \leqslant \frac{1}{p}\left(\frac{|X|}{\|X\|_p}\right)^p + \frac{1}{q}\left(\frac{|Y|}{\|Y\|_q}\right)^q = \frac{1}{p}\frac{|X|^p}{\|X\|_p^p} + \frac{1}{q}\frac{|Y|^q}{\|Y\|_q^q},$$

et en intégrant cette inégalité

$$\frac{\|XY\|_1}{\|X\|_p \times \|Y\|_q} = \mathbb{E}\left[\frac{|XY|}{\|X\|_p \times \|Y\|_q}\right] \leqslant \mathbb{E}\left[\frac{1}{p}\frac{|X|^p}{\|X\|_p^p} + \frac{1}{q}\frac{|Y|^q}{\|Y\|_q^q}\right]$$
$$= \frac{1}{p}\frac{\mathbb{E}\big[|X|^p\big]}{\|X\|_p^p} + \frac{1}{q}\frac{\mathbb{E}\big[|Y|^q\big]}{\|Y\|_q^q} = \frac{1}{p}\frac{\|X\|_p^p}{\|X\|_p^p} + \frac{1}{q}\frac{\|Y\|_q^q\big]}{\|Y\|_q^q} = \frac{1}{p} + \frac{1}{q} = 1,$$

d'où on déduit immédiatement l'inégalité cherchée. Pour qu'il y ait égalité, on doit avoir

$$\frac{|XY|}{\|X\|_p \times \|Y\|_q} = \frac{1}{p}\frac{|X|^p}{\|X\|_p^p} + \frac{1}{q}\frac{|Y|^q}{\|Y\|_q^q} \qquad \text{presque sûrement,}$$

et, donc, toujours d'après le lemme $|X|^p/\|X\|_p^p = |Y|^q/\|Y\|_q^q$ presque sûrement, ce qui achève la démonstration. □

Inégalité de Minkowski. — *Soit $p \geqslant 1$. Si X et Y sont deux variables aléatoires de puissance p-ième intégrable, alors $X + Y$ est une variable aléatoire de puissance p-ième intégrable et*

$$\|X + Y\|_p \leqslant \|X\|_p + \|Y\|_q$$

avec égalité si et seulement si X et Y sont combinaisons linéaires positives l'une de l'autre.

Démonstration. — Puisque l'application $x \in \mathbb{R} \mapsto |x|^p$ est convexe, on a

$$|X/2 + Y/2|^p \leqslant |X|^p/2 + |Y|^p/2 \qquad \text{soit} \qquad |X + Y|^p \leqslant 2^{p-1}(|X|^p + |Y|^p),$$

ce qui montre que si X et Y sont de puissance p-ième intégrable, alors c'est aussi le cas de $X + Y$.

Supposons dans un premier temps X et Y positives. Puisque $(p-1)q = p$, la variable aléatoire $(X + Y)^{p-1}$ est de puissance q-ième intégrable, on peut alors appliquer l'inégalité de Hölder à chaque terme du membre de droite de l'identité

$$(X + Y)^p = X(X + Y)^{p-1} + Y(X + Y)^{p-1}$$

pour obtenir

$$\begin{aligned}\mathbb{E}\big[(X + Y)^p\big] &= \mathbb{E}\big[X(X + Y)^{p-1} + Y(X + Y)^{p-1}\big] \\ &= \mathbb{E}\big[X(X + Y)^{p-1}\big] + \mathbb{E}\big[Y(X + Y)^{p-1}\big] \\ &\leqslant \|X\|_p \times \|(X + Y)^{p-1}\|_q + \|Y\|_p \times \|(X + Y)^{p-1}\|_q \\ &= (\|X\|_p + \|Y\|_p) \times \|(X + Y)^{p-1}\|_q \\ &= (\|X\|_p + \|Y\|_p) \times \mathbb{E}\big[(X + Y)^p\big]^{1/q}.\end{aligned}$$

Si le facteur de droite du membre de droite est nul, l'inégalité à démontrer se résume à $0 \leqslant \|X\|_p + \|Y\|_p$. Nous pouvons le supposer strictement positif, et en divisant par celui-ci obtenir

$$\mathbb{E}\big[(X + Y)^p\big]^{1-1/q} \leqslant \|X\|_p + \|Y\|_p$$

et puisque $1 - 1/q = 1/p$, $\|X + Y\|_p \leqslant \|X\|_p + \|Y\|_p$ qui est l'inégalité de Hölder. Lorsque X et Y ne sont pas nécessairement de signe positif, avec l'inégalité intermédiaire $|X + Y| \leqslant |X| + |Y|$, on obtient la même majoration.

Pour qu'il y ait égalité, on doit avoir $|X + Y| \leqslant |X| + |Y|$ presque sûrement et en plus être dans le cas de l'égalité pour l'inégalité de Hölder appliquée à $|X|$ et $|Y|$. Cette dernière condition implique que $|X|^p$ et $(|X + Y|^{p-1})^q = |X + Y|^p$, ainsi que $|Y|^p$ et $(|X + Y|^{p-1})^q = |X + Y|^p$, sont proportionnelles presque sûrement, et donc $|X|$ et $|Y|$ sont proportionnelles

presque sûrement, alors la première condition oblige à ce que X et Y soit positivement proportionnelles presque sûrement. □

Corollaire. — *Pour $p \geqslant 1$, l'espace $L^p(\Omega, \mathcal{A}, \mathbb{P}\,;\,\mathbb{R})$ des variables aléatoires de puissance p-ième intégrable modulo l'égalité presque sûre est un espace vectoriel normé muni de la norme $\|\,.\,\|_p$.*

Démonstration. — Nous savons déjà que l'ensemble des variables aléatoires réelles définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace vectoriel. Le sous-ensemble des variables aléatoires de puissance p-ième intégrable est stable par produit par un facteur scalaire (évident) et par addition compte tenu de l'inégalité de Minkowski. C'est donc un sous-espace vectoriel. Modulo l'égalité presque sûre, la distance déduite de $\|\,.\,\|_p$ vérifie l'axiome de séparabilité. L'homogénéité est évidente, l'inégalité triangulaire est donnée par l'inégalité de Minkowski. □

Inégalité de Cauchy–Schwarz. — *Si X et Y sont deux variables aléatoires réelles de carré intégrable, alors $X \times Y$ est une variable aléatoire réelle intégrable et*

$$\big|\mathbb{E}[X \times Y]\big| \leqslant \sqrt{\mathbb{E}[X^2]} \times \sqrt{\mathbb{E}[Y^2]}.$$

Par ailleurs, il y a égalité dans l'inégalité précédente si et seulement si il existe $(\alpha, \beta) \in \mathbb{R}^2 \backslash \{0, 0\}$ tel que $\alpha X + \beta Y = 0$ presque sûrement, c'est-à-dire que l'une des variables aléatoires est (presque sûrement) proportionnelle à l'autre.

Démonstration. — La preuve pourrait être presque identique à celle qu'on a dans le cadre des espaces préhilbertiens. L'inégalité de Hölder pour $p = q = 2$ donne le résultat. □

Remarque. — L'inégalité de Cauchy–Schwarz peut être appliquée aux variables aléatoires $|X|$ et $|Y|$ et en combinaison avec l'inégalité de Jensen (la fonction valeur absolue est convexe) on obtient :

$$\big|\mathbb{E}[X \times Y]\big| \leqslant \mathbb{E}[|X \times Y|] \leqslant \sqrt{\mathbb{E}[X^2]} \times \sqrt{\mathbb{E}[Y^2]}.$$

Définition. — Soient X et Y deux variables aléatoires réelles de carré intégrable.

(i) Le réel positif $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \overset{(\dots)}{=} \mathbb{E}[X^2] - \mathbb{E}[X]^2$ est appelé *variance* de la variable aléatoire X. Le réel positif $\sqrt{\mathrm{Var}(X)}$ est appelé *écart-type* de X, c'est l'écart quadratique moyen de la variable aléatoire X à sa moyenne $\mathbb{E}[X]$.

(ii) Le réel $\mathrm{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X]) \times (Y - \mathbb{E}[Y])] \overset{(\dots)}{=} \mathbb{E}[X \times Y] - \mathbb{E}[X] \times \mathbb{E}[Y]$ est appelé *covariance* des variables aléatoires X et Y.

(iii) Si $\mathrm{cov}(X, Y) = 0$, on dit que les variables aléatoires X et Y sont *non corrélées* (et corrélées sinon).

Remarque. — Les identités $\overset{(\dots)}{=}$ exigent un peu de calcul. Elles sont pourtant assez élémentaires et reçoivent dans la littérature le nom de formule de Huyghens.

Propriétés. — *L'application* cov *est une forme bilinéaire symétrique et* Var *est sa forme quadratique associée ; elles sont définies sur l'ensemble des variables aléatoires réelles de carré intégrable : soient X et Y deux variables aléatoires réelles de carré intégrable, alors*

(i) *pour tous α, $\beta \in \mathbb{R}$,* $\mathrm{Var}(\alpha X) = \alpha^2 \,\mathrm{Var}(X)$, $\mathrm{cov}(\alpha X, \beta Y) = \alpha\beta \,\mathrm{cov}(X, Y)$;

(ii) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{cov}(X, Y)$;

(iii) *De plus, si X et Y sont indépendantes alors* $\mathrm{cov}(X, Y) = 0$.

Démonstration. — Tout ceci est évident. □

Exemple (variables binomiales, suite). — Si Y est une variable aléatoire de loi binomiale de paramètres n et p telle que décrite précédemment, on a

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \overset{\text{indépendance}}{=} \sum_{i=1}^{n} \mathrm{Var}(X_i),$$

comme on le constate aisément en utilisant les propriétés précédentes. Or, $\mathrm{Var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = \mathbb{E}[X_i] - \mathbb{E}[X_i]^2 = p - p^2 = p(1-p)$ car $X_i^2 = X_i$ puisque X_i est à valeurs dans $\{0,1\}$. D'où

$$\mathrm{Var}(Y) = np(1-p).$$

Théorème. — (i) *Soit $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}_+$ une variable aléatoire positive. Alors pour tout $a > 0$,*

$$\mathbb{P}\{Y > a\} \leqslant \frac{1}{a}\,\mathbb{E}[Y] \qquad (\textit{inégalité de Markov}).$$

(ii) *Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire intégrable, $m = \mathbb{E}[X]$. Alors pour tout $\varepsilon > 0$, on a*

$$\mathbb{P}\{|X - m| > \varepsilon\} \leqslant \frac{1}{\varepsilon^2}\,\mathrm{Var}(X) \qquad (\textit{inégalité de Bienaymé–Tchebychev}).$$

Démonstration. — (i) On a $\mathbb{1}_{\{Y>a\}} \leqslant \mathbb{1}_{\{Y>a\}} \times (Y/a) \leqslant Y/a$. Ainsi

$$\mathbb{P}\{Y > a\} = \mathbb{E}[\mathbb{1}_{\{Y>a\}}] \leqslant \mathbb{E}\left[\frac{Y}{a}\right] = \frac{1}{a}\,\mathbb{E}[Y].$$

(ii) Le résultat découle immédiatement de l'application de l'inégalité de Markov à $Y = (X-m)^2$ et $a = \varepsilon^2$. □

Remarques. — *a)* L'inégalité de Markov peut s'écrire aussi

$$\mathbb{P}\{Y \geqslant a\} \leqslant \frac{1}{a}\,\mathbb{E}[Y]$$

pour Y une variable aléatoire positive et $a > 0$.

L'inégalité de Bienaymé–Tchebychev peut s'écrire aussi

$$\mathbb{P}\{|X - m| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2}\,\mathrm{Var}(X)$$

pour X une variable aléatoire intégrable, $m = \mathbb{E}[X]$ et $\varepsilon > 0$.

b) Il est à noter que pour l'inégalité de Markov, on ne suppose pas X intégrable, c'est-à-dire dans ce cas $\mathbb{E}[X] < \infty$. Le cas $\mathbb{E}[X] = \infty$ est admissible pour cette inégalité mais n'apporte rien en pratique.

De même, pour l'inégalité de Bienaymé–Tchebychev, on ne suppose pas que X soit de carré intégrable mais seulement que X soit intégrable — ce qui est le minimum pour pouvoir considérer $m = \mathbb{E}[X]$. Le cas $\mathbb{E}[X^2] = \infty$, qui implique $\mathrm{Var}(X) = \infty$, est admissible pour cette inégalité mais n'apporte rien non plus en pratique.

Application possible. — Présenter et démontrer le théorème de Weierstrass.

5. **Fonctions caractérisant la loi d'une variable aléatoire réelle**

Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire réelle. Nous avons vu au cours de la première section que sa fonction de répartition F_X définie par

$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto F_X(x) = P_X(]-\infty, x]) = \mathbb{P}\{X \leqslant x\}, \end{aligned}$$

ou plus exactement la fonction de répartition de la loi P_X de la variable aléatoire X, caractérise la loi de X, c'est-à-dire que deux variables aléatoires réelles ayant même fonction de répartition ont même loi (ou encore que deux mesures de probabilité sur $\mathbb{R}$ ayant même fonction de répartition sont égales).

D'autres fonctions caractérisant la loi d'une variable aléatoire peuvent être définies.

Théorème et définition. — *Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{N}$ une variable aléatoire entière. La* fonction génératrice *de la loi de X est la fonction G_X définie comme la somme de la série entière en la variable s*

$$\sum_{n \geqslant 0} s^n \, \mathbb{P}\{X = n\}$$

dont le rayon de convergence R_X est supérieur ou égal à 1. Pour $s \in \mathbb{C}$, $|s| < R_X$, on a

$$G_X(s) = \sum_{n \geqslant 0} s^n \, \mathbb{P}\{X = n\} = \mathbb{E}\big[s^X\big].$$

Elle caractérise la loi de X : si $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{N}$ est une variable aléatoire entière telle que $G_X(s) = G_Y(s)$ au moins dans un voisinage de 0, alors $P_X = P_Y$, i.e., *les lois de X et de Y sont les mêmes.*

Démonstration. — Tout ceci est évident. □

Théorème. — *La fonction génératrice d'une somme de variables aléatoires entières indépendantes est égale, sur le disque unité ouvert, au produit des fonctions génératrices.*

Démonstration. — Tout ceci est évident puisque l'espérance d'un produit de variables aléatoires indépendantes intégrables est égal au produit de leurs espérances. Ainsi, si $X_1, \ldots, X_n$ sont des variables aléatoires entières indépendantes, $X_1 + \cdots + X_n$ est une variable aléatoire entière, et, pour $|s| < 1$,

$$\begin{aligned} G_{X_1+\ldots+X_n}(s) = \mathbb{E}\big[s^{X_1+\cdots+X_n}\big] &= \mathbb{E}\big[s^{X_1} \times \cdots \times s^{X_n}\big] \\ &= \mathbb{E}\big[s^{X_1}\big] \times \cdots \times \mathbb{E}\big[s^{X_n}\big] = G_{X_1}(s) \times \cdots \times G_{X_n}(s). \end{aligned}$$

Prendre garde néanmoins aux différents rayons de convergence. □

Exemples. — *a*) Lois de Bernoulli.

b) Lois binomiales.

c) Lois géométriques.

d) Lois de Poisson.

Plus généralement,

Théorème et définition. — *Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire réelle. Pour tout $\theta \in \mathbb{R}$, l'application $\mathrm{e}^{i\theta X}$ de Ω dans $\mathbb{S}^1$ le cercle unité de $\mathbb{C}$ est une variable aléatoire bornée, et donc intégrable. La fonction*

$$\begin{aligned} \varphi_X : \mathbb{R} &\longrightarrow \mathbb{C} \\ \theta &\longmapsto \varphi_X(\theta) = \mathbb{E}\big[\mathrm{e}^{i\theta X}\big] \end{aligned}$$

est appelée fonction caractéristique *(de la loi) de X : si $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ est une variable aléatoire telle que, pour tout $\theta \in \mathbb{R}$, $\varphi_X(\theta) = \varphi_Y(\theta)$, alors $P_X = P_Y$,* i.e.*, les lois de X et de Y sont les mêmes.*

Remarques. — *a*) L'espérance $\mathbb{E}[\,.\,]$ est étendue aux variables aléatoires à valeurs dans $\mathbb{C}$ en les écrivant comme combinaisons de leurs parties réelle et imaginaire. De même, elle s'étend aux variables aléatoires à valeurs dans $\mathbb{R}^d$ et $\mathbb{C}^n$. On verra au chapitre suivant qu'une application de Ω à valeurs dans l'un de ces derniers ensembles ou espaces est une variable aléatoire si et seulement si chacune de ses composantes est une variable aléatoire réelle.

b) À quelques facteurs multiplicatifs près, la fonction caractéristique de (la loi de) X est la transformée de Fourier de la loi de X. (Le fait qu'elle caractérise la loi peut être vu comme une conséquence de l'injectivité de la transformée de Fourier des distributions tempérées [*voir* Laurent Schwartz, *Théorie des distributions*].)

c) On utilise les fonctions caractéristiques surtout lorsqu'on a à faire avec des combinaisons linéaires de variables aléatoires réelles indépendantes. En effet, si X, $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ sont deux variables aléatoires réelles *indépendantes*, alors pour tous α, β, γ et $\theta \in \mathbb{R}$, les variables aléatoires $\mathrm{e}^{\mathrm{i}\theta\alpha X}$ et $\mathrm{e}^{\mathrm{i}\theta\beta Y}$ sont indépendantes et

$$\begin{aligned} \varphi_{\alpha X+\beta Y+\gamma}(\theta) &= \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta(\alpha X+\beta Y+\gamma)}\big] \\ &= \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta\alpha X+\mathrm{i}\theta\beta Y+\mathrm{i}\theta\gamma}\big] \\ &= \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta\alpha X}\big] \times \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta\beta Y}\big] \times \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta\gamma}\big] \\ &= \varphi_X(\alpha\theta) \times \varphi_Y(\beta\theta) \times \mathrm{e}^{\mathrm{i}\theta\gamma}. \end{aligned}$$

Des contre-exemples montrent que la réciproque n'a pas lieu, c'est-à-dire qu'il est possible que la fonction caractéristique d'une somme de variables aléatoires soit le produit de leurs fonctions caractéristiques sans que les variables aléatoires soient indépendantes.

d) Il est possible de montrer que si φ_X est la fonction caractéristique d'une variable aléatoire X, φ_X est une fonction continue de $\mathbb{R}$ dans $\mathbb{C}$, bornée en module par 1 et $\varphi_X(0) = 1$ (cette dernière propriété peut être utile pour vérifier si un calcul de fonction caractéristique est juste ou non). Si X est intégrable, on montre que φ est dérivable en 0 et alors $\varphi'_X(0) = \mathrm{i}\ \mathbb{E}[X]$ et si X est de carré intégrable alors φ est deux fois dérivable en 0 et $\varphi''_X(0) = -\,\mathbb{E}[X^2]$, etc. (Dériver l'exponentielle en θ sous l'espérance.)

Exemples. — *a*) Lois de Bernoulli.

b) Lois binomiales.

c) Lois géométriques.

d) Lois de Poisson.

e) Loi uniforme sur $[0, 1]$, sur $[a, b]$.

f) Lois exponentielles

g) Lois normales.

Chapitre IV

COMPLÉMENTS DE PROBABILITÉS, VECTEURS ALÉATOIRES

1. **Génération de tribus**

Théorème et définition. — *Soient E un ensemble et $(B_i)_{i\in I}$ une collection quelconque de parties de E. Il existe une unique tribu sur E appelée tribu engendrée par $(B_i)_{i\in I}$ et notée $\sigma((B_i)_{i\in I})$ contenant $(B_i)_{i\in I}$ et telle que toute autre tribu sur E contenant elle aussi $(B_i)_{i\in I}$ contient la tribu $\sigma((B_i)_{i\in I})$.*

Lemme. — *Soient E un ensemble et $(\mathcal{B}_j)_{j\in J}$ une collection non vide de tribus sur E. Alors*

$$\bigcap_{j\in J} \mathcal{B}_j = \{B \subset E : B \in \mathcal{B}_j,\ \forall\, j \in J\}$$

est une tribu sur E.

Démonstration du lemme. — C'est immédiat : pour tout $j \in J$, $E \in \mathcal{B}_j$, donc $E \in \bigcap_{j\in J} \mathcal{B}_j$; si $B \in \bigcap_{j\in J} \mathcal{B}_j$, alors $B^c \in \mathcal{B}_j$ pour tout $j \in J$, donc $B^c \in \bigcap_{j\in J} \mathcal{B}_j$; si $(B_n)_n \subset \bigcap_{j\in J} \mathcal{B}_j$, alors pour tout $j \in J$, $(B_n)_n \subset \mathcal{B}_j$ et $\bigcap_n B_n \in \mathcal{B}_j$, donc $\bigcap_n B_n \in \bigcap_{j\in J} \mathcal{B}_j$. □

Démonstration du théorème. — Considérons l'ensemble des tribus sur E contenant $(B_i)_{i\in I}$, notons cet ensemble $(\mathcal{B}_j)_{j\in J}$. Cet ensemble est non vide puisqu'il contient $\mathcal{P}(E)$. Posons $\mathcal{B} = \bigcap_{j\in J} \mathcal{B}_j$. C'est une tribu sur E qui contient $(B_i)_{i\in I}$. Si $\mathcal{B}'$ est une tribu sur E contenant $(B_i)_{i\in I}$, alors $\mathcal{B}' \in (\mathcal{B}_j)_{j\in J}$ et donc $\mathcal{B} \subset \mathcal{B}'$ (minimalité). L'unicité est immédiate. □

Exemples. — *a)* Si $X : \Omega \to E$ est une application quelconque et l'ensemble E est muni d'une structure mesurable $(E, \mathcal{E})$, $\sigma(X) = \{X^{-1}(B) : B \in \mathcal{E}\} = (X^{-1}(B)_{B\in\mathcal{E}}$ est une tribu. Ainsi $\sigma(\sigma(X)) = \sigma(X)$. C'est la plus petite tribu sur Ω pour laquelle X est mesurable.

b) Soit E un espace topologique. La tribu borélienne de E est la tribu engendrée par l'ensemble des ouverts de E. Elle est égale, par passage au complémentaire, à la tribu engendrée par l'ensemble des fermés de E. La tribu borélienne de E est notée $\mathcal{B}(E)$.

Exercice. — Considérons l'espace topologique $\mathbb{R}$ (topologie usuelle). Montrer que $\mathcal{B}(\mathbb{R})$ est engendrée par, respectivement,

(i) $]a, b[$, $a < b \in \mathbb{R}$ (ou $\mathbb{Q}$) ;
(ii) $[a, b]$, $a < b \in \mathbb{R}$ (ou $\mathbb{Q}$) ;
(iii) $[a, b[$, $a < b \in \mathbb{R}$ (ou $\mathbb{Q}$) ;
(iv) $]a, b]$, $a < b \in \mathbb{R}$ (ou $\mathbb{Q}$).

Indication. — *Tout ouvert de $\mathbb{R}$ s'écrit comme une réunion au plus dénombrable d'intervalles ouverts disjoints.*

Lemme. — *Soit $(E, \mathcal{E})$ un espace mesurable, F un ensemble et $f : E \to F$ une application. Alors $\tau(f) = \{C \subset F : f^{-1}(C) \in \mathcal{E}\}$ est une tribu sur $\mathcal{F}$.*

Démonstration. — Puisque $E = f^{-1}(F)$, on a $F \in \tau(f)$. Si $C \in \tau(f)$, puisque $f^{-1}(C^c) = f^{-1}(C)^c$, alors $C^c \in \tau(f)$. Enfin si $(C_n)_n \subset \tau(f)$, comme $f^{-1}(\bigcap_n C_n) = \bigcap_n f^{-1}(C_n)$, on a $\bigcap_n C_n \in \tau(f)$. □

Remarque. — La tribu $\tau(f)$ est la plus grande tribu sur F telle que $f : E \to F$ soit mesurable. La notation $\tau(f)$ n'est pas standard et cette notion ne servira que comme intermédiaire de démonstration.

THÉORÈME. — *Soient $(\Omega, \mathcal{A})$ et $(E, \mathcal{E})$ deux espaces mesurables, $(B_i)_{i\in I} \subset \mathcal{E}$ telle que $\sigma((B_i)_{i\in I}) = \mathcal{E}$, et $X : \Omega \to E$ une application. Alors l'application X est $\mathcal{A}/\mathcal{E}$-mesurable si et seulement si pour tout $i \in I$, $X^{-1}(B_i) \in \mathcal{A}$.*

Démonstration. — La condition est évidemment nécessaire. Elle est suffisante car par hypothèse la tribu $\tau(X)$ contient $(B_i)_{i\in I}$ et donc aussi $\sigma((B_i)_{i\in I}) = \mathcal{E}$. □

THÉORÈME. — *Soient E, F deux espaces topologiques et $f : E \to F$ une application continue. Alors f est $\mathcal{B}(E)/\mathcal{B}(F)$-mesurable (on dit dans ce cas que f est* borélienne*).*

Démonstration. — Puisque f est continue, les images réciproques des ouverts de F par f sont des ouverts de E et sont donc dans $\mathcal{B}(E)$. La tribu $\tau(f)$ sur F contient donc tous les ouverts de F et donc $\mathcal{B}(F)$, ce qui montre que f est borélienne. □

PROPOSITION 1. — *Soient E un espace topologique et $\mathcal{B}(E)$ sa tribu borélienne. Alors pour toute partie F de E, la tribu induite*

$$\mathcal{E}_F = \{B \cap F : B \in \mathcal{B}(E)\}$$

de $\mathcal{B}(E)$ sur F est égale à la tribu borélienne $\mathcal{B}(F)$ de F muni de la topologie induite de E.

Démonstration. — Notons $\mathcal{O}_E$ l'ensemble des ouverts de E. Par définition la topologie induite sur $F \subset E$ est $\mathcal{O}_F = \{O \cap F : O \in \mathcal{O}_E\}$. Soit $\mathcal{B}(F) = \sigma(\mathcal{O}_F)$ la tribu borélienne de F.

Comme pour tout $O \in \mathcal{O}_E$, $O \cap F \in \mathcal{E}_F$, alors $\mathcal{O}_F \subset \mathcal{E}_F$, et donc $\mathcal{B}(F) = \sigma(\mathcal{O}_F) \subset \mathcal{E}_F$.

Réciproquement, considérons l'application d'inclusion $f : x \in F \mapsto x \in E$. Cette application est continue car pour tout $O \in \mathcal{O}_E$, $f^{-1}(O) = O \cap F \in \mathcal{O}_E$. D'après la proposition précédente, $f : (F, \mathcal{B}(F)) \to (E, \mathcal{B}(E))$ est mesurable. Or $\mathcal{E}_F = f^{-1}(\mathcal{E})$, donc $\mathcal{E}_F \subset \mathcal{B}(F)$. On a donc l'égalité des deux tribus. □

Remarque. — Ceci est un cas particulier d'un résultat proposé en exercice au premier chapitre.

2. **Ensembles négligeables**

DÉFINITION. — Soit $(E, \mathcal{E}, \mu)$ un espace probabilisé. Un sous ensemble $B \subset E$ est (extérieurement) *négligeable* si et seulement si il existe $B' \in \mathcal{E}$, $B \subset B'$ et $\mu(B') = 0$. L'ensemble des négligeables de $(E, \mathcal{E}, \mu)$ est noté $\mathcal{N}(\mathcal{E}, \mu)$.

Lorsque l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ modélise une expérience aléatoire, ses ensembles négligeables sont dits ensembles *presque impossibles*. Un ensemble qui est de complémentaire presque impossible est dit *presque certain* et une identité qui a lieu sur un ensemble presque certain est dite *presque sûre*.

FAITS IMMÉDIATS. — (i) *Une réunion au plus dénombrable d'ensembles négligeables est négligeable.*

(ii) *Une intersection quelconque d'ensembles négligeables est négligeable.*

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Nous notons $\bar{\mathcal{A}} = \sigma(\mathcal{A}, \mathcal{N}(\mathcal{A}, \mathbb{P}))$ la tribu engendrée par $\mathcal{A}$ et $\mathcal{N}(\mathcal{A}, \mathbb{P})$.

PROPOSITION. — *Un sous-ensemble A de Ω appartient à $\bar{\mathcal{A}}$ si et seulement si il existe A', $A'' \in \mathcal{A}$ tels que $A' \subset A \subset A''$ et tels que $\mathbb{P}(A') = \mathbb{P}(A'')$ (ou encore $\mathbb{P}(A''\backslash A') = 0$).*

Démonstration (en seconde lecture). — Notons $\hat{\mathcal{A}}$ la collection des parties de Ω vérifiant la condition précédente : $A \in \hat{\mathcal{A}}$ si et seulement si il existe A', $A'' \in \mathcal{A}$ tels que $A' \subset A \subset A''$ et $\mathbb{P}(A') = \mathbb{P}(A'')$ (le fait que cette dernière condition soit équivalente à $\mathbb{P}(A''\backslash A') = 0$ est évident). Soit $A \in \hat{\mathcal{A}}$ et A', A'' comme dans la condition précédente. On a alors $A = A' \cup (A\backslash A')$ qui est la réunion d'un élément de $\mathcal{A}$ et d'un négligeable : $(A\backslash A') \subset (A''\backslash A')$ et $\mathbb{P}(A''\backslash A') = 0$. Réciproquement, supposons que $A \subset \Omega$ s'écrive $A' \cup N$ où $A' \in \mathcal{A}$ et $N \in \mathcal{N}(\mathcal{A}, \mathbb{P})$. Soit $N' \in \mathcal{A}$ tel que $N \subset N'$ et $\mathbb{P}(N') = 0$. Alors posons $A'' = A' \cup N' \in \mathcal{A}$. On a $A' \subset A \subset A''$ et $\mathbb{P}(A') \leqslant \mathbb{P}(A'') \leqslant \mathbb{P}(A') + \mathbb{P}(N') = \mathbb{P}(A')$, donc $\mathbb{P}(A') = \mathbb{P}(A'')$, et on constate que $A \in \hat{\mathcal{A}}$. Ainsi $A \subset \Omega$ est dans $\hat{\mathcal{A}}$ si et seulement si $A = A' \cup N$ avec N négligeable et $A' \in \mathcal{A}$.

On constate immédiatement que si $A \in \hat{\mathcal{A}}$, alors $A \in \bar{\mathcal{A}}$ et donc $\hat{\mathcal{A}} \subset \bar{\mathcal{A}}$. Pour montrer la réciproque, il suffit de vérifier que $\hat{\mathcal{A}}$ est une tribu puisque $\mathcal{A} \subset \hat{\mathcal{A}}$ et $\mathcal{N}(\mathcal{A}, \mathbb{P}) \subset \hat{\mathcal{A}}$:

(i) $E, \varnothing \in \mathcal{A} \subset \hat{\mathcal{A}}$;

(ii) si $A \in \hat{\mathcal{A}}$, $A' \subset A \subset A''$, $\mathbb{P}(A''\backslash A') = 0$, alors $A''^c \subset A^c \subset A'^c$ et on a de plus $A'^c\backslash A''^c = A''\backslash A'$, ce qui montre que $A^c \in \hat{\mathcal{A}}$;

(iii) si $(A_n)_n \subset \hat{\mathcal{A}}$, $A_n = A'_n \cup N_n$, $\bigcup_n A_n = (\bigcup_n A'_n) \cup (\bigcup_n N_n)$ réunion d'un élément de $\mathcal{A}$ et d'un négligeable qui est donc dans $\hat{\mathcal{A}}$.

Nous en déduisons donc l'inclusion réciproque $\bar{\mathcal{A}} \subset \hat{\mathcal{A}}$. □

PROPOSITION. — *Il existe une unique mesure de probabilité $\bar{\mathbb{P}}$ sur $(\Omega, \bar{\mathcal{A}})$ prolongeant $\mathbb{P}$.*

Démonstration (en seconde lecture). — Si $A \in \bar{\mathcal{A}}$, $A' \subset A \subset A''$, avec A', $A'' \in \mathcal{A}$, $\mathbb{P}(A') = \mathbb{P}(A'')$, on pose $\bar{\mathbb{P}}(A) = \mathbb{P}(A') = \mathbb{P}(A'')$. L'application $\bar{\mathbb{P}} : \bar{\mathcal{A}} \to [0,1]$ est bien définie : soit $A^* \subset A \subset A^{**}$ avec A^* et A^{**} vérifiant les mêmes conditions que A' et A'', et posons $A^\dagger = A' \cup A^* \in \mathcal{A}$. On a $A' \subset A^\dagger \subset A \subset A''$ et $A^* \subset A^\dagger \subset A \subset A^{**}$. Alors $\mathbb{P}(A') \leqslant \mathbb{P}(A^\dagger) \leqslant \mathbb{P}(A'')$, ce qui montre l'égalité $\mathbb{P}(A') = \mathbb{P}(A^\dagger) = \mathbb{P}(A'')$. Mais on a aussi $\mathbb{P}(A^*) \leqslant \mathbb{P}(A^\dagger) \leqslant \mathbb{P}(A^{**})$, ce qui montre l'égalité $\mathbb{P}(A^*) = \mathbb{P}(A^\dagger) = \mathbb{P}(A^{**})$. Donc toutes ces probabilités sont égales et $\bar{\mathbb{P}}(A)$ ne dépend pas du choix de A' et A''.

Dans un deuxième temps, il faut vérifier que $\bar{\mathbb{P}}$ est une mesure de probabilité. Évidemment, $\bar{\mathbb{P}}(\Omega) = \mathbb{P}(\Omega) = 1$. Soit $(A_n)_n \subset \bar{\mathcal{A}}$ disjoints. Pour chaque n, $A_n = A'_n \cup N_n$ où $A'_n \in \mathcal{A}$ et $N_n \in \mathcal{N}(\mathcal{A}, \mathbb{P})$. Les $(A'_n)_n$ sont disjoints, et $\bigcup_n N_n$ est extérieurement négligeable donc inclus dans un certain $N \in \mathcal{A}$ de probabilité nulle : $\mathbb{P}(N) = 0$. On a $\bigcup_n A'_n \subset \bigcup_n A_n = (\bigcup_n A'_n) \cup (\bigcup_n N_n) \subset (\bigcup_n A'_n) \cup N$, et aussi $\sum_n \mathbb{P}(A'_n) = \mathbb{P}(\bigcup_n A'_n) \leqslant \mathbb{P}((\bigcup_n A'_n) \cup N) \leqslant \mathbb{P}(\bigcup_n A'_n) + \mathbb{P}(N) = \mathbb{P}(\bigcup_n A'_n) = \sum_n \mathbb{P}(A'_n)$. Comme $\bar{\mathbb{P}}(A_n) = \mathbb{P}(A'_n)$ pour tout n, on en déduit que $\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n)$. Ainsi $\bar{\mathbb{P}}$ est bien une mesure de probabilité sur $(\Omega, \bar{\mathcal{A}})$.

Enfin, il nous reste à vérifier l'unicité de ce prolongement : si $\hat{\mathbb{P}}$ est une autre mesure de probabilité prolongeant $\mathbb{P}$ et si $A \in \bar{\mathcal{A}}$, alors pour tous A', $A'' \in \mathcal{A}$, $A' \subset A \subset A''$ et $\mathbb{P}(A') = \mathbb{P}(A'')$; on a $\mathbb{P}(A') = \hat{\mathbb{P}}(A') \leqslant \hat{\mathbb{P}}(A) \leqslant \hat{\mathbb{P}}(A'') = \mathbb{P}(A'')$, il y a donc égalité ; ainsi, par définition de $\bar{\mathbb{P}}$, $\hat{\mathbb{P}}(A) = \bar{\mathbb{P}}(A)$. □

Exemples. — *a)* Si E est au plus dénombrable, $\mathcal{E} = \mathbb{P}(E)$ et μ est une mesure de probabilité sur $(E, \mathcal{P}(E))$, les ensembles négligeables sont les sous-ensembles de $\{x \in E : \mu\{x\} = 0\}$.

b) Soit $\mathbb{R}$ muni de sa tribu borélienne et d'une mesure de probabilité absolument continue. Alors tout ensemble au plus dénombrable est négligeable. D'autres types d'ensembles négligeables existent (par exemple l'ensemble triadique de Cantor).

3. **Tribus et espaces produit**

Dans cette section nous nous limiterons à des collections finies d'ensembles mesurables et de variables aléatoires. Pourtant, quitte à modifier quelques énoncés en préservant une place à la finitude, on peut prolonger cette discussion à des collections quelconques...

Définition. — Soit $((E_i, \mathcal{E}_i))_{i=1}^n$ une collection finie d'espaces mesurables. La *tribu produit* sur $\prod_{i=1}^n E_i$ des tribus $(\mathcal{E}_i)_{i=1}^n$ est la tribu engendrée par l'ensemble des $\prod_{i=1}^n B_i$ où $B_i \in \mathcal{E}_i$ (pavés) pour tout $i = 1, \dots, n$. Cette tribu est notée $\bigotimes_{i=1}^n \mathcal{E}_i$.

Remarque. — L'intersection de pavés étant encore un pavé, on constate que la tribu $\bigotimes_{i=1}^n \mathcal{E}_i$ est engendrée par les $\prod_{j=1}^{i-1} E_j \times B_i \times \prod_{j=i+1}^n E_j$ (pavés élémentaires) pour $B_i \in \mathrm{e}_i$, $i = 1, \dots, n$. On peut même remplacer $B_i \in E_i$ par B_i appartenant à une collection de parties de E_i engendrant $\mathcal{E}_i$.

Théorème. — *Soient $((E_i, \mathcal{E}_i))_{i=1}^n$ une collection finie d'espaces mesurables,*

$$(E, \mathcal{E}) = \left(\prod_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{E}_i \right)$$

et $(\Omega, \mathcal{A})$ un espace mesurable. Une application $X : \Omega \to E$ est $\mathcal{A}/\mathcal{E}$-mesurable si et seulement si chacune de ses coordonnées $X_i : \Omega \to E_i$ est $\mathcal{A}/\mathcal{E}_i$-mesurable.

Lemme. — *Les applications coordonnées $f_i : E \to E_i$ sont $\mathcal{E}/\mathcal{E}_i$-mesurables.*

Démonstration du lemme. — Si $B_i \in \mathcal{E}_i$, $f_i^{-1}(B_i) = \prod_{j=1}^{i-1} E_j \times B_i \times \prod_{j=i+1}^n E_j$ qui est un pavé élémentaire et est donc dans $\mathcal{E}$. □

Démonstration du théorème. — Si $X : \Omega \to E$ est $\mathcal{A}/\mathcal{E}$-mesurable, d'après le lemme et par composition, $X_i = f_i \circ X : \Omega \to E_i$ est $\mathcal{A}/\mathcal{E}_i$-mesurable. Réciproquement, supposons les coordonnées X_i $\mathcal{A}/\mathcal{E}_i$-mesurables. La tribu $\tau(X)$ sur E contient tous les $\prod_{j=1}^{i-1} E_j \times B_i \times \prod_{j=i+1}^n E_j$ puisque $X^{-1}(\prod_{j=1}^{i-1} E_j \times B_i \times \prod_{j=i+1}^n E_j) = X_i^{-1}(B_i) \in \mathcal{A}$ par hypothèse. Donc $\mathcal{E} \subset \tau(X)$, ce qui montre que X est $\mathcal{A}/\mathcal{E}$-mesurable. □

Par la suite nous serons amené à admettre certains énoncés.

Théorème. — *Soit $(E, \mathcal{E})$ un espace mesurable. Soit $\mathcal{C} \subset \mathcal{E}$ une classe de parties mesurables stable par intersections finies et engendrant $\mathcal{E}$. Alors deux mesures de probabilité μ_1 et μ_2 sur $(E, \mathcal{E})$ sont égales ($\mu_1(B) = \mu_2(B)$ pour tout $B \in \mathcal{E}$) si et seulement si elles coïncident sur $\mathcal{C}$ ($\mu_1(B) = \mu_2(B)$ pour tout $B \in \mathcal{C}$).*

Remarque. — La condition de stabilité par intersections finies est importante.

Théorème. — *Soient $(E_i, \mathcal{E}_i, \mu_i)$, $i = 1, \dots, n$, des espaces probabilisés, alors il existe une et une seule mesure de probabilité μ sur $(\prod_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{E}_i)$ telle que*

$$\mu(B_1 \times \cdots \times B_n) = \mu_1(B_1) \times \cdots \times \mu_n(B_n) \qquad \text{pour tous } B_i \in \mathcal{B}_i,\ i = 1, \dots, n.$$

Cette mesure de probabilité est appelée mesure de probabilité produit *et est notée $\mu_1 \otimes \cdots \otimes \mu_n$.*

Remarque. — On peut remplacer $B_i \in \mathcal{B}_i$ dans la condition ci-dessus par $B_i \in \mathcal{C}_i$, pour $\mathcal{C}_i \subset \mathcal{E}_i$ stable par intersections finies et telle que $\sigma(\mathcal{C}_i) = \mathcal{B}_i$.

Exemples. — *a)* Soient $(E_i, \mathcal{E}_i, \mu_i)$, $i = 1, \dots, n$, des espaces probabilisés discrets : pour tout $i = 1, \dots, n$, E_i est un ensemble au plus dénombrable et $\mathcal{E}_i = \mathcal{P}(E_i)$. Alors $\mathcal{E}_i \otimes \cdots \otimes \mathcal{E}_n = \mathcal{P}(E_1 \times \cdots \times E_n)$ et $\mu_1 \otimes \cdots \otimes \mu_n$ est caractérisée par

$$\mu\{(x_1, \dots, x_n)\} = \mu_1\{x_1\} \times \cdots \times \mu_n\{x_n\} \qquad \text{pour tout } (x_1, \dots, x_n) \in E_1 \times \cdots \times E_n.$$

b) La tribu borélienne de $\mathbb{R}^d$ (de dimension finie) est égale à la tribu produit : $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes n}$ (on constate que les boules ouvertes et les pavé ouverts engendrent la même tribu). Si $\mu_i(\mathrm{d}x) = p_i(x)\,\mathrm{d}x$, $i = 1, \dots, n$, sont des mesures de probabilité absolument continues sur $\mathbb{R}$ (p_i application mesurable positive d'intégrale égale à 1), alors la mesure produit $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ est absolument continue et on a

$$\mu(\mathrm{d}x_1 \dots \mathrm{d}x_n) = p_1(x_1) \times \cdots \times p_n(x_n)\,\mathrm{d}x_1 \dots \mathrm{d}x_n.$$

4. **Vecteurs aléatoires**

Définition. — Un *vecteur aléatoire* X est un n-uplet $(X_1, \dots, X_n)$ de variables aléatoires $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to (E_i, \mathcal{E}_i)$ (avec $(E_i, \mathcal{E}_i) \subset (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ le plus souvent). La loi de $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\prod_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{E}_i)$ est appelée *loi conjointe* de $(X_1, \dots, X_n)$. Les lois de chaque X_i sont appelées *lois marginales* de X.

Théorème. — *Des variables aléatoires $X_1, \dots, X_n$ définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ sont indépendantes si et seulement si la loi de $X = (X_1, \dots, X_n)$ est la loi produit de ses lois marginales.*

Démonstration. — C'est immédiat. □

Exemples. — *a*) Soient $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E$ et $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to F$ deux variables aléatoires discrètes (les ensembles E et F sont au plus dénombrables et munis de leurs tribus discrètes). Soit $Z = (X, Y) : (\Omega, \mathcal{A}, \mathbb{P}) \to E \times F$ qui est une variable aléatoire discrète. La loi de Z est la donnée des nombres $P_Z\{(x, y)\} = \mathbb{P}\{Z = (x, y)\} = \mathbb{P}\{X = x,\, Y = y\}$, $x \in E$, $y \in F$, positifs et de somme égale à 1.

La loi de X s'obtient en sommant en y :

$$P_X\{x\} = \mathbb{P}\{X = x\} = \sum_{y \in F} \mathbb{P}\{X = x, Y = y\}.$$

La loi de Y s'obtient en sommant en x :

$$P_Y\{y\} = \mathbb{P}\{Y = y\} = \sum_{x \in E} \mathbb{P}\{X = x, Y = y\}.$$

Les variables aléatoires X et Y sont indépendantes si et seulement si pour tous $x \in E$, $y \in F$, $P_Z\{(x, y)\} = P_X\{x\} \times P_Y\{y\}$.

Dans le cas où les ensembles E et F sont finis, la loi conjointe de X peut se voir comme un tableau :

	x_1	…	x_i	…	x_m	total
y_1	$P_Z\{(x_1, y_1)\}$	…	$P_Z\{(x_i, y_1)\}$	…	$P_Z\{(x_m, y_1)\}$	$P_Y\{y_1\}$
⋮	⋮		⋮		⋮	⋮
y_j	$P_Z\{(x_1, y_j)\}$	…	$P_Z\{(x_i, y_j)\}$	…	$P_Z\{(x_m, y_j)\}$	$P_Y\{y_j\}$
⋮	⋮		⋮		⋮	⋮
y_n	$P_Z\{(x_1, y_n)\}$	…	$P_Z\{(x_i, y_n)\}$	…	$P_Z\{(x_m, y_n)\}$	$P_Y\{y_n\}$
total	$P_X\{x_1\}$	…	$P_X\{x_i\}$	…	$P_X\{x_m\}$	1

et tout tableau de réels positifs dont la somme totale est 1 peut s'interpréter comme la loi conjointe de deux variables aléatoires discrètes.

b) Soient X, $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ des variables aléatoires réelles et $Z = (X, Y)$, variable aléatoire à valeurs dans $\mathbb{R}^2$. Supposons que la loi de Z admette une densité $p_Z : \mathbb{R}^2 \to \mathbb{R}_+$ intégrable d'intégrale 1. On démontre (théorème de Fubini) qu'alors les lois de X et de Y admettent des densités données respectivement par

$$p_X(x) = \int_{\mathbb{R}} p_Z(x, y)\,\mathrm{d}y \qquad \text{et} \qquad p_Y(y) = \int_{\mathbb{R}} p_Z(x, y)\,\mathrm{d}x.$$

Si on suppose que les lois de X et de Y admettent des densités, c'est-à-dire sont absolument continues, il n'est pas nécessaire que ce soit le cas de la loi de Z.

En effet, soient $X : (\Omega, \mathcal{A}, \mathbb{P}) \to [0, 1]$ une variable aléatoire, $f : [0, 1] \to [0, 1]$ une fonction de classe C^1 et $Y = f(X) : (\Omega, \mathcal{A}, \mathbb{P}) \to [0, 1]$. Alors la loi de $Z = (X, Y)$ est portée par le graphe de f qui est d'aire nulle ; elle ne peut donc pas être absolument continue par rapport à la mesure d'aire $\mathrm{d}x\,\mathrm{d}y$ sur le carré $[0, 1]^2$. Pourtant, si f est assez régulière (C^1-difféomorphisme par exemple), la loi de Y ainsi définie admet — puisque c'est le cas de la loi de X — une densité.

Rappelons que nous avons déjà vu que si les variables aléatoires X et Y sont indépendantes et si leurs lois admettent une densité, la loi de $Z = (X, Y)$ admet elle aussi une densité.

Définition. — Soit $X = (X_1, \ldots, X_n) : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^n$ un vecteur aléatoire. Si chacune des coordonnées de X est intégrable, alors X est intégrable et

$$\mathbb{E}[X] = \big(\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n]\big) \in \mathbb{R}^n.$$

Lorsque chacune des coordonnées de X est de carré intégrable, X est dit de carré intégrable et on définit la matrice de covariance de X comme la matrice de covariance de ses coordonnées :

$$\begin{aligned}\mathrm{Cov}(X, X) &= \begin{pmatrix} \mathrm{cov}(X_1, X_1) & \ldots & \mathrm{cov}(X_1, X_j) & \ldots & \mathrm{cov}(X_1, X_n) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X_i, X_1) & \ldots & \mathrm{cov}(X_i, X_j) & \ldots & \mathrm{cov}(X_i, X_n) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X_n, X_1) & \ldots & \mathrm{cov}(X_n, X_j) & \ldots & \mathrm{cov}(X_n, X_n) \end{pmatrix} \\ &= \begin{pmatrix} \mathrm{Var}(X_1) & \ldots & \mathrm{cov}(X_1, X_j) & \ldots & \mathrm{cov}(X_1, X_n) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X_i, X_1) & \ldots & \mathrm{cov}(X_i, X_j) & \ldots & \mathrm{cov}(X_i, X_n) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X_n, X_1) & \ldots & \mathrm{cov}(X_n, X_j) & \ldots & \mathrm{Var}(X_n) \end{pmatrix}\end{aligned}$$

qui est une matrice symétrique de dimensions $n \times n$, et plus généralement, si $Y : (\Omega, \mathcal{A}, \mathbb{P} \to \mathbb{R}^m$ est un autre vecteur aléatoire de carré intégrable,

$$\mathrm{Cov}(X, Y) = \begin{pmatrix} \mathrm{cov}(X_1, Y_1) & \ldots & \mathrm{cov}(X_1, Y_j) & \ldots & \mathrm{cov}(X_1, Y_m) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X_i, Y_1) & \ldots & \mathrm{cov}(X_i, Y_j) & \ldots & \mathrm{cov}(X_i, Y_n) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X_n, Y_1) & \ldots & \mathrm{cov}(X_n, Y_j) & \ldots & \mathrm{cov}(X_n, Y_m) \end{pmatrix}$$

qui est une matrice de dimensions $n \times m$ et qui n'est pas nécessairement symétrique. En

notations matricielles, considérons X et Y comme des vecteurs colonnes

$$X \times {}^tY = \begin{pmatrix} X_1Y_1 & \dots & X_1Y_j & \dots & X_1Y_m \\ \vdots & & \vdots & & \vdots \\ X_iY_1 & \dots & X_iY_j & \dots & X_iY_m \\ \vdots & & \vdots & & \vdots \\ X_nY_1 & \dots & X_nY_j & \dots & X_nY_m \end{pmatrix}$$

et ainsi

$$\mathrm{Cov}(X,Y) = \mathbb{E}\big[\big(X - \mathbb{E}[X]\big) \times {}^t\big(Y - \mathbb{E}[Y]\big)\big]$$

en étendant l'espérance coordonnées par coordonnées aux matrices. Les notations matricielles sont souvent à privilégier lorsqu'on considère des covariances de vecteurs aléatoires.

Chapitre V

CONVERGENCES

Ce chapitre est assez superficiel. Son intérêt réside d'une part dans les calculs, et, d'autre part, dans la préparation à des concepts qui — dans le cadre du programme du Capes — interviennent en Statistique.

1. **Définitions, exemples**

Définition. — Soient $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$, $n \geqslant 1$, une suite de variables aléatoires réelles, $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire réelle et P une mesure de probabilité sur $\mathbb{R}$.

(i) la suite $(X_n)_n$ *converge presque sûrement* vers X si et seulement si il existe $A \in \mathcal{A}$, $\mathbb{P}(A) = 1$, tel que pour tout $\Omega \in A$, $X_n(\omega) \to X(\omega)$ quand $n \to \infty$;

(ii) la suite $(X_n)_n$ *converge en probabilité* vers X si et seulement si, pour tout $\varepsilon > 0$, $\mathbb{P}\{|X_n - X| > \varepsilon\} \to 0$ quand $n \to \infty$;

(iii) la suite $(X_n)_n$ *converge en loi* vers P si et seulement si les lois P_{X_n} convergent vers P quand $n \to \infty$, c'est-à-dire si l'une des conditions équivalentes suivantes est satisfaite :

— pour toute application $f : \mathbb{R} \to \mathbb{R}$ continue bornée,

$$\mathbb{E}[f(X_n)] = \int_{\mathbb{R}} f(x)\, P_{X_n}(\mathrm{d}x) \longrightarrow \int_{\mathbb{R}} f(x)\, P(\mathrm{d}x) \qquad \text{quand } n \to \infty\,;$$

— pour tout $\theta \in \mathbb{R}$,

$$\varphi_{X_n}(\theta) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\theta X_n}] = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\theta x}\, P_{X_n}(\mathrm{d}x) \longrightarrow \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\theta x}\, P(\mathrm{d}x) \qquad \text{quand } n \to \infty\,;$$

— pour tout $x_c \in \mathbb{R}$ où $F : x \mapsto \int_{-\infty}^{x} P(\mathrm{d}x) = P(]-\infty, x])$ est continue, on a

$$F_{X_n}(x_c) \longrightarrow F(x_c) \qquad \text{quand } n \to \infty.$$

Pour le point (iii), la première condition est une condition de convergence étroite de mesures de probabilité sur $\mathbb{R}$ (qui équivaut dans ce cadre à la convergence vague), la seconde est la convergence simple des fonctions caractéristiques vers la fonction caractéristique de la mesure de probabilité P, la troisième est la convergence simple des fonctions de répartition en tout point de continuité de la fonction de répartition de la mesure de probabilité P. L'équivalence annoncée est un théorème difficile qui sera ici admis.

Proposition. — *La convergence presque sûre implique la convergence en probabilité, et cette dernière implique la convergence en loi.*

Démonstration. — Admis. □

Exemples de convergences presque sûres. — *a)* Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire réelle. Nous avons déjà vu qu'il est possible d'approcher uniformément X par des variables aléatoires discrètes : pour $\varepsilon > 0$, on définit $X_\varepsilon : (\Omega, \mathcal{A}, \mathbb{P}) \to \varepsilon\mathbb{Z} \subset \mathbb{R}$ par $X_\varepsilon = \varepsilon \times$ partie entière(X/ε) ; on a alors $X_\varepsilon \leqslant X < X_\varepsilon + \varepsilon$, ce qui montre que X_ε approche X

uniformément sur Ω à ε près. Lorsque ε tend vers 0, la suite $(X_\varepsilon)_{\varepsilon>0}$ converge uniformément vers X et donc aussi simplement sur Ω. Cette convergence est donc presque sûre puisque sûre.

Il existe des procédés d'approximation de variables aléatoires réelles par des variables aléatoires discrètes plus naturels pour le probabiliste. Ils consistent souvent à considérer la suite des projections d'une variable sur une suite croissante de sous-tribus au plus dénombrables de $\mathcal{A}$ (*voir* les notions d'espérances conditionnelles, filtrations et de martingales). Dans ce cadre, on ne peut guère qu'espérer une convergence presque sûre puisque ces projections ne sont elles même définies que presque sûrement.

b) Étant donnée une variable aléatoire réelle $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$, il est fréquent de considérer des suites de troncations de X, par exemple en posant pour tout $n \geqslant 1$,

$$\omega \in \Omega, \qquad X_n(\omega) = \begin{cases} n & \text{si } X(\omega) \geqslant n, \\ X(\omega) & \text{si } X(\omega) \in [-n, n], \\ -n & \text{si } X(\omega) \leqslant -n. \end{cases}$$

Il est clair qu'une telle suite converge sûrement vers X lorsque n tend vers l'infini. L'intérêt, essentiellement d'ordre théorique, de la troncation d'une variable aléatoire réelle est d'obtenir des ordres d'intégrabilité plus grands.

c) Il est fréquent de voir apparaître des suites de variables aléatoires comme des sommes partielles de séries. Un exemple usuel est de la forme suivante : soit un nombre $b > 1$ et $\xi_k : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$, $k \geqslant 1$, une suite de variables aléatoires telles que pour tout $k \geqslant 1$, $\mathbb{P}\{\xi_k \in [0, b[\} = 1$. On a alors

$$\begin{aligned} \mathbb{P}\{\xi_k \in [0, b[, \text{ pour tout } k \geqslant 1\} &= \mathbb{P}\left(\bigcap_{k\geqslant 1} \{\xi_k \in [0, b[\} \right) \\ &= 1 - \mathbb{P}\left(\bigcup_{k\geqslant 1} \{\xi_k \notin [0, b[\} \right) \geqslant 1 - \sum_{k\geqslant 1} \mathbb{P}\{\xi_k \notin [0, b[\} = 1, \end{aligned}$$

et donc $\mathbb{P}\{\xi_k \in [0, b[, \text{ pour tout } k \geqslant 1\} = 1$. En posant pour tout $n \geqslant 1$,

$$X_n = \sum_{k=1}^{n} \frac{\xi_k}{b^k}$$

on définit une suite de variables aléatoires qui sont les sommes partielles de la série aléatoire

$$X = \sum_{k=1}^{\infty} \frac{\xi_k}{b^k}$$

qui est presque sûrement convergente puisque presque sûrement à termes positifs majorés par $(1/b^{k-1})_{k\geqslant 1}$ dont la série associée est géométrique et convergente

$$\sum_{k=1}^{\infty} \frac{1}{b^{k-1}} = \sum_{\ell=0}^{\infty} \frac{1}{b^\ell} = \frac{1}{1 - 1/b} = \frac{b}{b-1} < \infty.$$

Notons que la variable X n'est *a priori* définie que sur un événement de probabilité 1 contenant $\{\xi_k \in [0, b[, \text{ pour tout } k \geqslant 1\}$ (on peut montrer de plus que l'ensemble de éléments de Ω où la série converge est un événement). On pourra la prolonger à Ω en la posant égale à 0, par exemple, en dehors de cet événement.

Supposons de plus que b est un entier supérieur ou égal à 2, $(\xi_k)_{k\geqslant 1}$ est une suite de variables aléatoires indépendantes uniformément distribuées sur $\{0, 1, \ldots, b-1\}$. On peut montrer que pour tout $n \geqslant 1$, X_n est uniformément distribuée sur l'ensemble $\{0, 1/b^n, \ldots, (b^n -$

$1)/b^n\}$ qui comporte b^n éléments également espacés dans $[0,1]$, et finalement, que la variable limite X est uniformément distribuée sur $[0,1]$ (ceci se fait particulièrement facilement en considérant les fonctions caractéristiques).

Ce qui précède est lié à l'écriture en base b des nombres de l'intervalle $[0,1]$ muni de sa tribu borélienne et de sa mesure uniforme. Lorsque $b=2$, les sommes partielles correspondent aux *nombres dyadiques*, lorsque $b=10$, aux *nombres décimaux*.

d) *Voir* la loi forte des grands nombres évoquée à la section suivante.

Exemples de convergences en probabilité. — a) Toute suite de variables aléatoires qui converge presque sûrement converge en probabilité (admis).

b) Nous nous plaçons sur $\Omega = \mathbb{R} \bmod 1 \sim [0,1[$ le cercle de circonférence 1 muni de sa tribu borélienne et de sa mesure de probabilité uniforme. Pour $n \geqslant 1$, définissons

$$X_n(\omega) = \begin{cases} 1 & \text{si } \omega \in \left[\displaystyle\sum_{k=1}^{n-1} \frac{1}{k} \bmod 1,\ \sum_{k=1}^{n} \frac{1}{k} \bmod 1 \right] \\ 0 & \text{sinon.} \end{cases}$$

Nous avons pour tout $\varepsilon > 0$,

$$\mathbb{P}\{|X_n - 0| > \varepsilon\} = \mathbb{P}\{X_n = 1\} = \frac{1}{n} \longrightarrow 0 \qquad \text{quand } n \text{ tend vers l'infini,}$$

ce qui montre que la suite de variables aléatoires $(X_n)_{n\geqslant 1}$ converge vers 0 en probabilité. Cependant cette suite ne converge nulle part au sens de la convergence simple.

En effet, par construction X_n est une « fonction palier » au dessus du cercle dont le palier est de longueur $1/n$, X_{n+1} est alors la « fonction palier » dont le palier jouxte le précédent et est de longueur $1/(n+1)$, et ainsi de suite, de sorte qu'entre le rang n et $n+m$, tous les points situés entre (en suivant l'orientation du cercle et en faisant éventuellement plusieurs tours) $x_{n-1} = \sum_{k=1}^{n-1} 1/k \bmod 1$ et $x_{n+m} = \sum_{k=1}^{n+m} 1/k \bmod 1$ sont couverts par un palier au moins. Par divergence de la série $\sum_{k\geqslant n} 1/k$, ce sont donc tous les points du cercle qui sont recouverts pour m assez grand. Comme cela est vrai pour tout $n \geqslant 1$, on en déduit que chaque point du cercle est recouvert une infinité de fois par un palier. Aussi, pour tout $\omega \in \Omega$, la suite numérique $(X_n(\omega))_{n\geqslant 1}$ prend une infinité de fois la valeur 1 ; elle prend aussi une infinité de fois la valeur 0 car pour tout $n \geqslant 2$ tel que $X_n(\omega) = 1$, on a $X_{n+1}(\omega) = 0$ (en fait elle prend bien plus souvent la valeur 0 que la valeur 1). Ainsi il n'y a nulle part convergence simple.

Cette exemple classique est parfois appelé exemple de la « bosse flottante ». Il montre que la convergence en probabilité est strictement plus faible que la convergence presque sûre.

c) *Voir* la loi faible des grands nombres à la section suivante.

Venons en maintenant à deux exemples importants de convergences de lois de probabilité. Ils seront complétés à la dernière section par le théorème central limite.

Cas particulier. — *Soient $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to E \subset \mathbb{R}$, $n \geqslant 1$, une suite de variables aléatoires discrètes, avec E un sous-ensemble discret (au sens topologique, c'est-à-dire dont tous les points sont isolés) de $\mathbb{R}$ (par exemple un sous-ensemble fini, $\mathbb{N}$, etc., mais pas $\mathbb{Q}$). Alors $(X_n)_n$ converge en loi vers la mesure de probabilité P si et seulement si pour tout $x \in E$*

$$P_{X_n}\{x\} = \mathbb{P}\{X_n = x\} \longrightarrow P\{x\} \qquad \textit{quand } n \to \infty.$$

Nous utiliserons dans la suite les résultats suivants :

Résultats asymptotiques. — (i) *Pour tout $x \in \mathbb{R}$ et toute suite $(x_n)_n$ convergeant vers x, on a*

$$\Big(1 + \frac{x_n}{n}\Big)^n \longrightarrow \mathrm{e}^x \qquad \textit{quand } n \textit{ tend vers l'infini.}$$

(ii) *Soient $\alpha \in \mathbb{R}_+^*$ et $k \in \mathbb{N}$. Alors*

$$\frac{(\alpha n)\,!}{(\alpha n - k)\,!} \sim \alpha^k n^k \qquad \textit{quand } n \textit{ tend vers l'infini.}$$

Démonstration. — (i) Soit $x \in \mathbb{R}$ et $N \in \mathbb{N}^*$ tel que $x_n/n \in \,]-1, 1[$ pour tout $n \geqslant N$. Pour $n \geqslant N$, on a

$$(1 + x_n/n)^n = \exp\big(n \ln(1 + x_n/n)\big) = \exp\big(n(x/n + o(1/n))\big) = \exp\big(x + o(1)\big),$$

d'où le résultat (bien connu) quand n tend vers l'infini.

(ii) Soit $\alpha \in \mathbb{R}_+^*$ et $k \in \mathbb{N}$ sont fixés. La factorielle d'un nombre réel positif x est définie par $x\,! = \Gamma(x+1)$ où Γ est la fonction Gamma d'Euler

$$\Gamma(x) = \int_0^\infty t^{x-1}\,\mathrm{e}^{-t}\,\mathrm{d}t \qquad \text{pour tout } x \in [1, \infty[.$$

La propriété $\Gamma(x+1) = x\Gamma(x)$ s'itère pour obtenir que, pour tout $x \geqslant k$,

$$x\,! = x \times (x-1) \times \cdots \times (x-k+1) \times \Gamma(x-k+1) = x \times (x-1) \times \cdots \times (x-k+1) \times (x-k)\,!$$

Ainsi

$$\frac{(\alpha n)\,!}{(\alpha n - k)\,!} = (\alpha n) \times (\alpha n - 1) \times \cdots \times (\alpha n - k + 1) = \alpha^k n^k + o\big(n^k\big)$$

quand n tend vers l'infini comme on le voit en développant ce produit d'un nombre fixe égal à k de termes. L'équivalence en découle. □

Rappelons que pour n, $N \in \mathbb{N}^*$ et $p \in [0,1]$ tels que $Np \in \mathbb{N}$ et $n \leqslant N$, la *loi hypergéométrique* de paramètres (N, n, p) est la loi de probabilité discrète π de support inclus dans $\mathbb{N}$ vérifiant

$$\pi\{k\} = \begin{cases} \dfrac{C_{Np}^k C_{N(1-p)}^{n-k}}{C_N^n} & \text{pour } \max\big(0, n - N(1-p)\big) \leqslant k \leqslant \min(n, Np), \\ 0 & \text{sinon.} \end{cases}$$

On note généralement cette loi de probabilité $\mathcal{H}(N, n, p)$.

Théorème. — *Pour tous $n \in \mathbb{N}^*$ et $p \in [0,1]$, les lois hypergéométriques $(\mathcal{H}(N, n, p))_N$ convergent vers la loi binomiale $\mathcal{B}(n, p)$ quand $N \to \infty$.*

Interprétation. — Une loi hypergéométrique $\mathcal{H}(N, n, p)$ apparaît typiquement dans les situations qui se ramènent à celle-ci : considérons une urne comportant N boules dont Np boules blanches et $N(1-p)$ boules noires ; considérant un tirage sans remise de n boules dans l'urne, la loi du nombre de boules blanches tirées est la loi hypergéométrique $\mathcal{H}(N, n, p)$. Ce que dit ce théorème est que si N est très supérieur à n, le tirage peut être considéré comme étant presque avec remise. En effet, dans ces conditions les tirages successifs n'affectent quasiment pas les conditions du tirage, à savoir les proportions successives de boules blanches et noires restantes dans l'urne.

Démonstration. — Nous allons montrer que pour tout $k \in \{0, 1, \ldots, n\}$,

$$\frac{C_{Np}^k \, C_{N(1-p)}^{n-k}}{C_N^n} \longrightarrow C_n^k \, p^k (1-p)^{n-k} \qquad \text{quand } N \to \infty.$$

Supposons $p \in]0, 1[$. Nous avons, pour N assez grand,

$$\begin{aligned}\frac{C_{Np}^k \, C_{N(1-p)}^{n-k}}{C_N^n} &= \frac{n\,!}{k\,!\,(n-k)\,!} \times \frac{(Np)\,!\,(N(1-p))\,!\,(N-n)\,!}{(Np-k)\,!\,(N(1-p)-n+k)\,!\,N\,!} \\ &= C_n^k \times \frac{(Np)\,!}{(Np-k)\,!} \times \frac{(N(1-p))\,!}{(N(1-p)-n+k)\,!} \times \frac{(N-n)\,!}{N\,!}\end{aligned}$$

En appliquant le second résultat asymptotique avec $\alpha = p$, $\alpha = 1-p$ et $\alpha = 1$, lorsque N tend vers l'infini, on a

$$\frac{C_{Np}^k \, C_{N(1-p)}^{n-k}}{C_N^n} \sim C_n^k \times p^k N^k \times (1-p)^{n-k} N^{n-k} \times \frac{1}{1^n N^n} = C_n^k \, p^k (1-p)^{n-k},$$

d'où la conclusion pour $p \in]0, 1[$.

Pour $p \in \{0, 1\}$, nous avons identiquement $\mathcal{H}(N, n, p) = \mathcal{B}(n, p)$ pour tout $N \geqslant n$. En effet, il n'y a alors dans l'urne alors que des boules d'une seule couleur. □

Théorème. — *Soit $(p_n)_n \subset [0, 1]$ une suite telle que $(np_n)_n$ converge vers un certain $\lambda \in \mathbb{R}_+$. Alors, les lois binomiales $(\mathcal{B}(n, p_n))_n$ convergent vers la loi de Poisson $\mathcal{P}(\lambda)$ quand $n \to \infty$.*

Démonstration. — Précisons que $p_n \to 0$ quand $n \to \infty$. Comme λ peut être nul, nous nous garderons d'utiliser une équivalence du type $p_n \sim \lambda/n$ mais préférerons écrire le développement limité $p_n = \lambda/n + o(1/n)$ quand n tend vers l'infini.

Pour $k \in \mathbb{N}$ fixé et $n \geqslant k$ tendant vers l'infini, on a

$$\begin{aligned}C_n^k \, p_n^k (1-p_n)^{n-k} &= C_n^k \, p_n^k \, \mathrm{e}^{(n-k)\ln(1-p_n)} = C_n^k \, p_n^k \, \mathrm{e}^{(n-k)\ln(1-\lambda/n+o(1/n))} \\ &= C_n^k p_n^k \, \mathrm{e}^{(n-k)(-\lambda/n+o(1/n))} = \frac{n\,!}{k\,!\,(n-k)\,!} \, p_n^k \, \mathrm{e}^{-\lambda+o(1)} \\ &= \frac{n^k + o(n^k)}{k\,!} \, p_n^k \, \mathrm{e}^{-\lambda} = \frac{(np_n)^k + p_n^k o(n^k)}{k\,!} \, \mathrm{e}^{-\lambda} \longrightarrow \frac{\lambda^k}{k\,!} \, \mathrm{e}^{-\lambda},\end{aligned}$$

où nous avons appliqué le second résultat asymptotique (avec $\alpha = 1$) sous la forme d'un développement limité pour éviter d'écrire des équivalences entre termes éventuellement nuls (si $p_n = 0$ pour certains n). La convergence de $(np_n)^k$ vers λ^k est acquise du fait de la continuité de la fonction puissance k-ième, celle du terme $p_n^k o(n^k)$ vers 0 est immédiate. D'où la conclusion. □

L'exemple suivant n'est plus du même type que les deux précédents. Le support des lois auxquelles on s'intéresse n'est plus fixe et celui de la loi limite n'est pas discret puisque c'est $\mathbb{R}_+$. Il

Théorème. — *Soient $(p_n)_n \subset [0, 1]$ une suite telle que $(np_n)_n$ converge vers un certain $\lambda \in \mathbb{R}_+$ et $(X_n)_n$ une suite quelconque de variables aléatoires telle que pour chaque n, X_n est de loi gémétrique $\mathcal{G}(p_n)$. Alors, la suite des lois des variables aléatoires $(X_n/n)_n$ converge vers la loi exponentielle $\mathcal{E}(\lambda)$ quand $n \to \infty$.*

Démonstration. — Nous allons établir la convergence des lois par la convergence simple des fonctions de répartition. Soit $x \in \mathbb{R}_+$. Pour tout $n \geqslant 1$, on a

$$\mathbb{P}\left\{\frac{X_n}{n} > x\right\} = \mathbb{P}\{X_n > nx\} = \mathbb{P}\{X_n > \lfloor nx \rfloor\} = (1-p_n)^{\lfloor nx \rfloor}$$

où $\lfloor nx \rfloor$ désigne la partie entière de nx. On a l'encadrement

$$(1-p_n)^{nx} \leqslant (1-p_n)^{\lfloor nx \rfloor} \leqslant (1-p_n)^{nx-1}$$

c'est-à-dire

$$\big((1-p_n)^n\big)^x \leqslant (1-p_n)^{\lfloor nx \rfloor} \leqslant \big((1-p_n)^n\big)^x (1-p_n)^{-1}$$

puisque np_n converge vers λ, on a d'après le premier résultat asymptotique

$$(1-p_n)^n = \Big(1 - \frac{np_n}{n}\Big)^n \longrightarrow \mathrm{e}^{-\lambda} \qquad \text{quand } n \text{ tend vers l'infini.}$$

Puisque $t \mapsto t^x$ est continue sur $\mathbb{R}_+$ et que $(p_n)_n$ tend vers 0, les termes extrêmes de l'encadrement tendent tous deux vers $(\mathrm{e}^{-\lambda})^x = \mathrm{e}^{-\lambda x}$. Ainsi on a pour tout $x \geqslant 0$

$$\mathbb{P}\left\{\frac{X_n}{n} > x\right\} \longrightarrow \mathrm{e}^{-\lambda x} \qquad \text{quand } n \text{ tend vers l'infini.}$$

Par ailleurs la convergence est évidente pour $x \leqslant 0$, car les variables étant positives, ces probabilités sont toutes égales à 1. Nous avons donc montré que 1 moins les fonctions de répartition convergent simplement vers 1 moins la fonction de répartition de la loi exponentielle de paramètre λ, d'où la convergence simple annoncée. □

Remarques. — *a)* Il est facile, mais peu éclairant, de démontrer le résultat de convergence vers une loi de Poisson à l'aide des fonctions caractéristiques. En effet, la fonction caractéristique de la loi $\mathcal{B}(n, p_n)$ est $\theta \mapsto (1-p_n+p_n\,\mathrm{e}^{\mathrm{i}\theta})^n = (1+p_n(\mathrm{e}^{\mathrm{i}\theta}-1))^n$. Par un calcul d'équivalents rapide, si $np_n \to \lambda$, on constate que ces fonctions convergent simplement vers $\theta \mapsto \exp(\lambda(\mathrm{e}^{\mathrm{i}\theta}-1))$ qui est la fonction caractéristique de la loi de Poisson de paramètre λ.

Puisque les lois sont des lois de variables aléatoires entières, il est aussi possible de considérer les fonctions génératrices : la fonction génératrice de la loi $\mathcal{B}(n, p_n)$ est $s \mapsto (1-p_n+p_n s)^n = (1+p_n(s-1))^n$; si $np_n \to \lambda$, la suite de ces fonctions converge simplement vers $\theta \mapsto \exp(\lambda(s-1))$ qui est la fonction génératrice de la loi de Poisson de paramètre λ.

b) De même, pour la convergence vers une loi exponentielle, nous autrons pu avoir recours aux fonctions caractéristiques. En effet la fonction caractéristique de la loi exponentielle de paramètre λ est, pour $\theta \neq 0$,

$$\varphi(\theta) = \int_0^\infty \lambda \exp\big((\mathrm{i}\theta - \lambda)x\big)\,\mathrm{d}x = \frac{\lambda}{\mathrm{i}\theta - \lambda}\big[\exp\big((\mathrm{i}\theta-\lambda)x\big)\big]_0^\infty = \frac{\lambda}{\lambda - \mathrm{i}\theta}\,;$$

et celle de la loi de X_n/n, où X_n est de loi géométrique de paramètre p_n, est, pour $\theta \neq 0$,

$$\begin{aligned}
\varphi_{X_n/n}(\theta) = \varphi_{X_n}(\theta/n) &= \frac{p_n\,\mathrm{e}^{\mathrm{i}\theta/n}}{1-(1-p_n)\,\mathrm{e}^{\mathrm{i}\theta/n}} \\
&= \frac{(\lambda/n + o(1/n)) \times (1+o(1))}{1-(1-\lambda/n+o(1/n)) \times (1+\mathrm{i}\theta/n+o(1/n))} \\
&= \frac{(\lambda+o(1))\,\mathrm{e}^{\mathrm{i}\theta/n}}{n-(n-\lambda+o(1)) \times (1+\mathrm{i}\theta/n+o(1/n))} \\
&= \frac{\lambda+o(1)}{n-n+\lambda-\mathrm{i}\theta+o(1)} = \frac{\lambda}{\lambda-\mathrm{i}\theta} + o(1),
\end{aligned}$$

d'où la conclusion.

Applications. — L'approximation de lois hypergéométriques par des lois binomiales est fréquente dans la pratique, en particulier en Statistique appliquée : on considère une population, par exemple les habitants d'un pays, sur laquelle les individus possèdent ou non un certain

caractère ; on prélève dans cette population un échantillon souvent sans remise et on approche la loi du nombre d'individus possédant ce caractère dans un tel échantillon — qui est hypergéométrique — par la loi binomiale correspondante.

L'approximation de lois binomiales par des lois de Poisson est appelée *loi des petits nombres* : lorsque le paramètre $p \in [0,1]$ est proche de 0, la loi $\mathcal{B}(n,p)$ (loi du « nombre de succès ») est proche de la loi de Poisson $\mathcal{P}(np)$. À l'inverse, lorsque p est proche de 1, $1-p$ est proche de 0, et on approche $\mathcal{B}(n,1-p)$ (qui est la loi du « nombre d'échecs ») par la loi de Poisson $\mathcal{P}(n(1-p))$. Des conditions d'approximation typiquement utilisées sont les suivantes :

— $n \geqslant 50$, $np \leqslant 10$, alors $\mathcal{B}(n,p)$ est proche de $\mathcal{P}(np)$;

— $n \geqslant 50$, $n(1-p) \leqslant 10$, alors $\mathcal{B}(n,1-p)$ est proche de $\mathcal{P}(n(1-p))$;

en rappelant que si une variable aléatoire X a pour loi $\mathcal{B}(n,p)$, la variable aléatoire $Y = n-X$ a pour loi $\mathcal{B}(n,1-p)$.

2. **Loi faible des grands nombres**

Définition. — Une suite de variables aléatoires $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$, $n \geqslant 1$, intégrables de même espérance $\mathbb{E}[X_n] = m$, satisfait *la loi faible (resp. forte) des grands nombres*, si et seulement si la suite de variables aléatoires

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \qquad n \geqslant 1,$$

converge en probabilité (resp. presque sûrement) vers m.

Rappelons l'inégalité de Bienaymé–Tchebychev : soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire intégrable, $m = \mathbb{E}[X]$; alors pour tout $\varepsilon > 0$, on a

$$\mathbb{P}\{|X-m| > \varepsilon\} \leqslant \frac{1}{\varepsilon^2}\operatorname{Var}(X).$$

Théorème (loi faible des grands nombres). — *Soit $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$, $n \geqslant 1$, une suite de variable aléatoires indépendantes, identiquement distribuées, de carré intégrable, $m = \mathbb{E}[X_n]$. Alors*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{} m \qquad \textit{en probabilité,}$$

c'est-à-dire que la suite $(X_n)_{n\geqslant 1}$ satisfait la loi faible des grands nombres.

Démonstration. — Posons $\sigma^2 = \operatorname{Var}(X_n)$ qui ne dépend pas de $n \geqslant 1$. Soit $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Alors $\mathbb{E}[\bar{X}_n] = m$ et $\operatorname{Var}(\bar{X}_n) = \operatorname{Var}(X_1+\cdots+X_n)/n^2 = (\operatorname{Var} X_1 + \cdots + \operatorname{Var} X_n)/n^2 = \sigma^2/n$ par non corrélation de $X_1, \ldots, X_n$. Pour $\varepsilon > 0$, par l'inégalité de Tchebychev, on a

$$\mathbb{P}\big\{\big|\bar{X}_n - m\big| \geqslant \varepsilon\big\} \leqslant \frac{\sigma^2}{\varepsilon^2 n} \longrightarrow 0 \qquad \text{quand } n \to \infty,$$

ce qui montre la convergence en probabilité. □

Remarques. — *a*) Plus haut, on a seulement besoin que $(X_n)_{n\geqslant 1}$ soient non corrélées, $\mathbb{E}[X_n] = m$, et $\operatorname{Var}(X_n) \leqslant \sigma^2$ pour que la suite $(X_n)_{n\geqslant 1}$ satisfasse la loi faible des grands nombres.

b) Il existe un résultat plus pointu : si $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes, intégrables, $\mathbb{E}[X_n] = m$, et si l'une des deux conditions suivantes est satisfaite :

$$(X_n)_{n\geqslant 1} \text{ est identiquement distribuée,} \qquad \text{ou} \qquad \sum_{n=1}^{\infty} \operatorname{Var}(X_n)/n^2 < \infty\,;$$

alors $(X_n)_{n\geqslant 1}$ satisfait la loi forte des grands nombres, et donc aussi la loi faible. Ce résultat est appelé *la loi forte des grands nombres de Kolmogorov*.

c) Dans le language statistique, n variables aléatoires $X_1, \ldots, X_n$ définies sur un même espace probabilisé, indépendantes, de même loi qu'une variable aléatoire X (définie éventuellement sur un autre espace probabilisé) est appelé échantillon de taille n de la variable X. Si X est intégrable et $m = \mathbb{E}[X]$, la convergence en probabilité de

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad \text{vers} \qquad m \qquad \text{lorsque } n \text{ tend vers l'infini}$$

s'énonce en disant que $\bar{X}_n$ est un *estimateur consistant* du paramètre m. Comme on a de plus $\mathbb{E}[\bar{X}_n] = m$, $\bar{X}_n$ est dit être un *estimateur sans biais* de m. Finalement quand la loi faible des grands nombres est satisfaite, $\bar{X}_n$ est un estimateur consistant et sans biais de m.

Exemple. — Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes identiquement distribuées de loi $\mathcal{B}(1, p)$, alors $\bar{X}_n \to p$ quand $n \to \infty$ (presque sûrement et en probabilité).

On a comme application — assez anecdotique — l'estimation expérimentale d'un tel paramètre p pour un lancer de pile ou face.

De manière plus profonde, les lois des grands nombres montrent que la probabilité d'un événement est la limite des fréquences de réalisation de cet événement au cours d'épreuves répétées indépendantes. Nous retrouvons la définition de la probabilité d'un événement correspondant à la notion d'expérience aléatoire (*voir* chapitre premier).

3. **Théorème central limite**

Cette section porte sur une propriété universelle des lois normales qui est appelée *théorème central limite*. Ce théorème est asymptotique, d'où le mot « limite », et il joue un rôle « central » en particulier en Statistique.

Lemme. — *Soient $m \in \mathbb{R}$ et $\sigma > 0$. La fonction caractéristique de la loi normale de moyenne m et d'écart-type σ est*

$$\theta \in \mathbb{R} \longmapsto \exp\bigl(\mathrm{i}m\theta - \sigma^2\theta^2/2\bigr).$$

Démonstration. — Soit X une variable aléatoire réelle de loi $\mathcal{N}(m, \sigma^2)$. Nous savons alors que $Z = (X - m)/\sigma$ est une variable aléatoire réelle de loi $\mathcal{N}(0, 1)$. On a, pour tout $\theta \in \mathbb{R}$,

$$\begin{aligned}\varphi_X(\theta) = \mathbb{E}\bigl[\mathrm{e}^{\mathrm{i}\theta X}\bigr] = \mathbb{E}\bigl[\mathrm{e}^{\mathrm{i}\theta(X-m)} \times \mathrm{e}^{\mathrm{i}\theta m}\bigr] &= \mathbb{E}\bigl[\mathrm{e}^{\mathrm{i}\sigma\theta(X-m)/\sigma} \times \mathrm{e}^{\mathrm{i}\theta m}\bigr] \\ &= \mathbb{E}\bigl[\mathrm{e}^{\mathrm{i}\sigma\theta(X-m)/\sigma}\bigr] \times \mathrm{e}^{\mathrm{i}\theta m} = \mathbb{E}\bigl[\mathrm{e}^{\mathrm{i}\sigma\theta Z}\bigr] \times \mathrm{e}^{\mathrm{i}\theta m} = \varphi_Z(\sigma\theta) \times \mathrm{e}^{\mathrm{i}\theta m}.\end{aligned}$$

Il nous suffit donc de calculer la fonction caractéristique φ_Z de la loi normale $\mathcal{N}(0, 1)$. Nous avons

$$\varphi_Z(\theta) = \int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}\theta z}\, \mathrm{e}^{-z^2/2}\, \frac{\mathrm{d}z}{\sqrt{2\pi}}\,.$$

Démontrer que φ_Z est de classe C^1 et que sa fonction dérivée s'obtient par dérivation sous le signe somme est classique, nous nous en dispensons. Nous avons donc

$$\varphi'_Z(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}\theta z}\,\mathrm{e}^{-z^2/2}\,\frac{\mathrm{d}z}{\sqrt{2\pi}}\right) = \int_{-\infty}^{+\infty} \frac{\partial}{\partial\theta}\big(\mathrm{e}^{\mathrm{i}\theta z}\,\mathrm{e}^{-z^2/2}\big)\,\frac{\mathrm{d}z}{\sqrt{2\pi}}$$
$$= \int_{-\infty}^{+\infty} \frac{\partial}{\partial\theta}\big(\mathrm{e}^{\mathrm{i}\theta z}\big)\,\mathrm{e}^{-z^2/2}\,\frac{\mathrm{d}z}{\sqrt{2\pi}} = \int_{-\infty}^{+\infty} \mathrm{i}z\,\mathrm{e}^{\mathrm{i}\theta z}\,\mathrm{e}^{-z^2/2}\,\frac{\mathrm{d}z}{\sqrt{2\pi}}\,.$$

Puis nous intégrons par parties en la variable z en notant que $(\mathrm{e}^{-z^2/2})' = -z\,\mathrm{e}^{-z^2/2}$

$$\varphi'_Z(\theta) = \int_{-\infty}^{+\infty} -\mathrm{i}\,\mathrm{e}^{\mathrm{i}\theta z}\,\frac{\mathrm{d}}{\mathrm{d}z}(\mathrm{e}^{-z^2/2})\,\frac{\mathrm{d}z}{\sqrt{2\pi}} = \left[-\mathrm{i}\,\mathrm{e}^{\mathrm{i}\theta z}\,\mathrm{e}^{-z^2/2}\frac{1}{\sqrt{2\pi}}\right]_{-\infty}^{+\infty}$$
$$+ \int_{-\infty}^{+\infty} \mathrm{i}^2\theta\,\mathrm{e}^{\mathrm{i}\theta z}\,\mathrm{e}^{-z^2/2}\,\frac{\mathrm{d}z}{\sqrt{2\pi}} = 0 - \theta\int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}\theta z}\,\mathrm{e}^{-z^2/2}\,\frac{\mathrm{d}z}{\sqrt{2\pi}} = -\theta\times\varphi_Z(\theta).$$

Ainsi la fonction φ_Z vérifie l'équation différentielle $\varphi' = f(\theta,\varphi) = -\theta\times\varphi$ avec f continue, localement lipschitzienne en la deuxième variable (elle est en fait C^1 sur $\mathbb{R}\times\mathbb{R}$) et la condition initiale $\varphi_Z(0) = 1$. D'après le théorème de Cauchy–Lipschitz, il y a unicité de la solution maximale associée à ce problème de Cauchy. Comme la fonction $\theta\mapsto \mathrm{e}^{-\theta^2/2}$ vérifie aussi ces conditions, par unicité, on a $\varphi_Z(\theta) = \mathrm{e}^{-\theta^2/2}$ pour tout $\theta\in\mathbb{R}$. La formule annoncée dans le lemme s'en déduit immédiatement : pour $\theta\in\mathbb{R}$,

$$\varphi_X(\theta) = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta X}\big] = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}m\theta}\times\mathrm{e}^{\mathrm{i}\theta(X-m)}\big] = \mathrm{e}^{\mathrm{i}m\theta}\times\mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta(X-m)}\big]$$
$$= \mathrm{e}^{\mathrm{i}m\theta}\times\mathbb{E}\big[\mathrm{e}^{\mathrm{i}(\sigma\theta)(X-m)/\sigma}\big] = \mathrm{e}^{\mathrm{i}m\theta}\times\mathbb{E}\big[\mathrm{e}^{\mathrm{i}(\sigma\theta)Z}\big] = \mathrm{e}^{\mathrm{i}m\theta}\times\varphi_Z(\sigma\theta) = \exp\big(\mathrm{i}m\theta - \sigma^2\theta^2/2\big),$$

ce qui était annoncé. □

Remarque. — On notera que le résultat est encore vrai lorsque $\sigma = 0$: on obtient la fonction caractéristique de la loi d'une variable aléatoire constante égale à m.

Théorème. — *Soient $X_i : (\Omega,\mathcal{A},\mathbb{P})\to\mathbb{R}$, $i = 1,\dots,n$, des variables aléatoires indépendantes de lois respectives $\mathcal{N}(m_i,\sigma_i^2)$, $m_i\in\mathbb{R}$, $\sigma_i\geqslant 0$, $i = 1,\dots,n$. Alors la variable aléatoire $X = X_1+\cdots+X_n$ a pour loi la loi normale $\mathcal{N}(m_1+\cdots+m_n,\sigma_1^2+\cdots+\sigma_n^2)$.*

Démonstration. — On calcule la fonction caractéristique de la loi de X : pour tout $\theta\in\mathbb{R}$, par indépendance de $X_1,\dots,X_n$,

$$\varphi_X(\theta) = \varphi_{X_1+\cdots+X_n}(\theta) = \prod_{i=1}^{n}\varphi_{X_i}(\theta) = \prod_{i=1}^{n}\mathrm{e}^{\mathrm{i}m_i\theta-\sigma_i^2\theta^2/2} = \mathrm{e}^{\mathrm{i}(m_1+\cdots+m_n)\theta-(\sigma_1^2+\cdots+\sigma_n^2)\theta^2/2}.$$

C'est la fonction caractéristique de la loi normale de moyenne $m_1+\cdots+m_n$ et de variance $\sigma_1^2+\cdots+\sigma_n^2$ qui est donc la loi de X. □

Comme nous savons déjà que toute transformation affine d'une variable aléatoire de loi normale est encore de loi normale, nous pouvons en déduire que toute combinaison affine de variables aléatoires indépendantes de lois normales est de loi normale. Les paramètres se retrouvent aisément en évaluant l'espérance et la variance d'une telle combinaison. Le théorème suivant montre qu'outre cette propriété de stabilité des lois normales, elles ont une propriété d'« attracteurs ».

Théorème central limite. — *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles, indépendantes, identiquement distribuées, de carré intégrable, $m = \mathbb{E}[X_n]$ et $\sigma^2 = \mathrm{Var}(X_n)$.*

Alors

$$\mathrm{Loi}\left(\frac{\left(\sum_{k=1}^n X_k\right) - nm}{\sqrt{n\sigma^2}}\right) \longrightarrow \mathcal{N}(0,1) \qquad \textit{quand } n \to \infty.$$

Démonstration (plus que facultative). — Soit Z une variable aléatoire réelle de carré intégrable, alors la fonction caractéristique φ_Z est de classe C^2 et on a

$$\begin{aligned}\varphi_Z(t) &= \varphi_Z(0) + t\varphi_Z'(0) + \frac{t^2}{2}\,\varphi_Z''(0) + o(t^2)\\ &= \mathbb{E}\big[\mathrm{e}^{\mathrm{i}0Z}\big] + t\,\mathbb{E}\big[\mathrm{i}Z\,\mathrm{e}^{\mathrm{i}0Z}\big] + \frac{t^2}{2}\,\mathbb{E}\big[-Z^2\,\mathrm{e}^{\mathrm{i}0Z}\big] + o(t^2)\\ &= 1 + \mathrm{i}t\,\mathbb{E}[Z] - \frac{t^2}{2}\,\mathbb{E}\big[Z^2\big] + o(t^2).\end{aligned}$$

On pose

$$Y_n = \frac{\left(\sum_{k=1}^n X_k\right) - nm}{\sqrt{n\sigma^2}} = \sum_{k=1}^n \frac{X_k - m}{\sqrt{n\sigma^2}}$$

qui est une somme de variables aléatoires identiquement distribuées et indépendantes. On a

$$\begin{aligned}\varphi_{\frac{X_k-m}{\sqrt{n\sigma^2}}}(\theta) = \varphi_{X_k-m}\left(\frac{\theta}{\sqrt{n\sigma^2}}\right) &= 1 + \mathrm{i}\,\frac{\theta}{\sqrt{n\sigma^2}}\,\mathbb{E}[X_k - m] - \frac{\theta^2}{2n\sigma^2}\,\mathbb{E}\big[(X_k-m)^2\big] + o\Big(\frac{1}{n}\Big)\\ &= 1 + \mathrm{i}\,\frac{\theta}{\sqrt{n\sigma^2}} \times 0 - \frac{\theta^2}{2n\sigma^2}\sigma^2 + o\Big(\frac{1}{n}\Big) = 1 - \frac{\theta^2}{2n} + o\Big(\frac{1}{n}\Big)\end{aligned}$$

Ainsi

$$\begin{aligned}\varphi_{Y_n}(\theta) = \prod_{k=1}^n \varphi_{\frac{X_k-m}{\sqrt{n\sigma^2}}}(\theta) = \left(1 - \frac{\theta^2}{2n} + o\Big(\frac{1}{n}\Big)\right)^n &= \exp\left(n\ln\left(1 - \frac{\theta^2}{2n} + o\Big(\frac{1}{n}\Big)\right)\right)\\ &= \exp\left(n\left(-\frac{\theta^2}{2n} + o\Big(\frac{1}{n}\Big)\right)\right) = \exp\left(-\frac{\theta^2}{2} + o(1)\right) \longrightarrow \exp\left(-\frac{\theta^2}{2}\right)\end{aligned}$$

quand $n \to \infty$ (attention : il y a des difficultés cachées avec l'exponentielle ou le logarithme complexe). Les fonctions caractéristiques convergent simplement vers la fonction caractéristique de $\mathcal{N}(0,1)$, d'où la convergence en loi. □

Remarques. — *a*) On notera que l'énoncé du théorème porte sur la loi de $X_1 + \cdots + X_n$ diminuée de sa moyenne nm divisée par son écart-type $\sqrt{n}\,\sigma$, qui est dès lors de moyenne 0 et d'écart-type 1 pour tout n. Le résultat porte donc sur une somme de variables aléatoires indépendantes identiquement distribuées qui a été centrée et réduite. Par une transformation affine, nous obtiendrions bien évidemment des résultats similaires avec pour loi limite la même transformée affine de la loi normale $\mathcal{N}(0,1)$.

b) Il est fondamental de comparer le résultat affirmé par le théorème central limite avec la loi faible/forte des grands nombres. Cette dernière dit que, sous de bonnes hypothèses, on a $(X_1 + \cdots + X_n)/n \sim m$ lorsque n tend vers l'infini. Le théorème central limite quantifie l'écart asymptotique entre cette moyenne de Césaro et sa limite.

Application : approximation de $\mathcal{B}(n,p)$. — Si $(X_k)_{k=1}^n$ sont des variables aléatoires indépendantes identiquement distribuées de loi $\mathcal{B}(1,p)$, alors $Z_n = \sum_{k=1}^n X_k$ est de loi $\mathcal{B}(n,p)$ et

$$\mathrm{Loi}\left(\frac{Z_n - np}{\sqrt{np(1-p)}}\right) \longrightarrow \mathcal{N}(0,1) \qquad \text{quand } n \to \infty.$$

Pratique. — Si $n \geqslant 50$, $np \geqslant 10$ et $n(1-p) \geqslant 10$, on applique l'approximation correspondante. Par exemple, si X est de loi $\mathcal{B}(n,p)$, avec $n \geqslant 50$, $np \geqslant 10$ et $n(1-p) \geqslant 10$, alors pour $0 \leqslant k \leqslant l \leqslant n$

$$\begin{aligned}
\mathbb{P}\big\{X \in \{k,\dots,l\}\big\} &= \mathbb{P}\{k \leqslant X \leqslant l\} = \mathbb{P}\{k-0{,}5 \leqslant X \leqslant l+0{,}5\} \\
&= \mathbb{P}\left\{\frac{k-0{,}5-np}{\sqrt{np(1-p)}} \leqslant \frac{X-np}{\sqrt{np(1-p)}} \leqslant \frac{l+0{,}5-np}{\sqrt{np(1-p)}}\right\} \\
&\approx \int_{\frac{k-0{,}5-np}{\sqrt{np(1-p)}}}^{\frac{l+0{,}5-np}{\sqrt{np(1-p)}}} \mathrm{e}^{-x^2/2}\,\frac{\mathrm{d}x}{\sqrt{2\pi}} \\
&= \Phi\left(\frac{l+0{,}5-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-0{,}5-np}{\sqrt{np(1-p)}}\right).
\end{aligned}$$

où Φ est la fonction de répartition de la loi $\mathcal{N}(0,1)$. Le $\pm 0{,}5$ est souvent appelé « correction de continuité », et est souvent, assez légitimement, oublié.

TABLE DES MATIÈRES

Introduction . . . 1
A. Quelques repères . . . 1
B. Exemples d'expériences aléatoires . . . 1
C. Expériences aléatoires, tentative de définition . . . 2
D. Modélisation mathématique . . . 3
E. Bibliographie sommaire . . . 4

Chapitre premier. Formalisme général . . . 5
1. Définitions générales . . . 5
2. Cas usuels . . . 6
2.1. *Espaces probabilisés discrets* . . . 6
2.2. *Variables aléatoires discrètes* . . . 7
2.3. *Mesures de probabilité absolument continues sur* $\mathbb{R}$ . . . 8
3. Propriétés essentielles . . . 10

Compléments au chapitre I . . . 13
A. Séries à termes positifs . . . 13
B. Génération de tribus . . . 13
C. Classes monotones . . . 14
D. Espaces topologiques et mesurabilité . . . 15
E. Espaces produits et mesurabilité . . . 17
F. Tribus induites . . . 18
G. Ensembles négligeables, ensembles presque sûrs . . . 19
H. Complétion d'un espace probabilisé . . . 19

Chapitre II. Probabilités conditionnelles, Indépendance . . . 20
1. Probabilités conditionnelles . . . 20
2. Indépendance . . . 22

Chapitre III. Variables aléatoires réelles . . . 26
1. Généralités . . . 26
2. Intégration de variables aléatoires réelles discrètes . . . 28
3. Intégration de variables aléatoires réelles . . . 29
4. Un peu d'analyse . . . 33
5. Fonctions caractérisant la loi d'une variable aléatoire réelle . . . 38

Chapitre IV. Compléments de Probabilités, vecteurs aléatoires 40

1. Génération de tribus 40

2. Ensembles négligeables 41

3. Tribus et espaces produit 43

4. Vecteurs aléatoires 44

Chapitre V. Convergences 47

1. Définitions, exemples 47

2. Loi faible des grands nombres 53

3. Théorème central limite 54

Table des matières 58