Théorie radicalement élémentaire des probabilités

Edward Nelson
Département de Mathématique
Université de Princeton

traduit par Thierry Delbecque

French translation posted on the Web April 2003 at
`http:www.math.princeton.edu/~nelson/books.html`

Table des matières

Introduction

Plus que toute autre branche des mathématiques, la théorie des probabilités s'est développée conjointement à leurs applications. Cela fut vrai dès la naissance de cette discipline, au moment où Pascal et Fermat tâchaient de résoudre le problème de partage des mises dans un jeu de hasard,[1] et cela se poursuit à notre époque, où les travaux les plus excitants en probabilités sont menés par des physiciens œuvrant dans le domaine de la mécanique statistique.

Kolmogorov a établi les axiomes de la théorie des probabilités il y a plus de cinquante ans. Je suis convaincu qu'avec moi, de nombreux probabilistes en charge de l'enseignement de débutants ont le sentiment que ces fondations, qui relèvent de la théorie de la mesure, sont là plus pour satisfaire notre bonne conscience de mathématiciens que pour fournir à un praticien désireux d'appliquer la théorie des probabilités un moyen efficace de le faire.

Le présent travail propose de nouvelles fondations pour la théorie des probabilités. Pour cela, il met en œuvre quelques énoncés très simples de l'Analyse Non Standard. Le bagage mathématique requis pour étudier cet ouvrage est à peine supérieur à celui d'un élève d'une classe de Terminale scientifique à la fin de son année scolaire. Aussi mon vœu est-il que par ce travail, des résultats profonds de la théorie moderne des processus stochastiques soient rendus accessibles à quiconque est apte à effectuer une addition ou une multiplication, et à raisonner.

Un tel résultat a été rendu possible parce que l'on a décidé de conserver aux énoncés des résultats leurs formes non standard. Les praticiens de l'Analyse Non Standard adoptent en effet une forme de raisonnement originale qui,

[1] Blaise Pascal (1623, 1662) ; Pierre de Fermat (1601, 1665). Le problème du partage des mises, posé par le chevalier de Méré, est le suivant : soient deux joueurs s'affrontant dans un jeu de hasard, dont la règle prévoit que le premier à avoir gagné N parties emporte la mise S. On suppose que pour une raison quelconque on doive interrompre le jeu. Sachant que chacun des joueurs devrait encore gagner respectivement n et p coups pour remporter la partie, comment répartir équitablement la mise S entre eux deux. (NdT)

si elle n'est pas retraduite en termes conventionnels, paraît remarquablement élémentaire.

Les mathématiciens sont à juste titre plutôt conservateurs et méfiants face aux idées neuves. Ils voudront s'assurer que les résultats développés ici sont aussi féconds que les résultats classiques, et que leur temps ne sera pas perdu s'ils apprennent l'Analyse Non Standard. Ces questions seront abordées dans l'annexe. L'annexe en elle-même exige de la part du lecteur un niveau mathématique nettement supérieur à celui nécessaire pour étudier le cœur de l'ouvrage. En revanche, la lecture du cœur de l'ouvrage se suffit à elle-même.

Remerciements

Je tiens à remercier Eric Carlen, Mai Gehrke, Klaus Kaiser, et Brian White pour leurs précieux commentaires, ainsi que Pierre Cartier pour sa lecture critique du manuscrit, ses corrections des nombreuses erreurs, et ses suggestions très utiles. Ce travail a été en partie soutenu par le N.S.F. (National Science Foundation).

Fine Hall, 1987

Je veux exprimer mes plus vifs remerciements à Thierry Delbecque, qui a conçu et a mené à terme le projet d'une version française. Plus qu'une simple traduction, il s'agit d'une révision critique de l'ouvrage. Il a complété le texte, très concis à la base, en le rendant plus précis et plus facile à suivre. Il a repéré et corrigé quelques erreurs qui s'y trouvaient encore, et a ajouté des notes (NdT) de nature soit plutôt générale, soit plutôt technique.

Le contenu de ce livre m'a servi de support d'un cours que j'ai donné l'année dernière. J'avais alors conclu sur une présentation standard du principe d'invariance (voir le chapitre 18) basée sur le livre de Patrick Billingsley "Convergence of probability measures". C'est un très bel ouvrage, qui regorge de subtilités mathématiques. Mais il y a une chose frappante : dans notre approche radicalement élémentaire, les quatres pages et demi du chapitre 17 font complètement abstraction de ce matériel standard. Je ne comprends pas vraiment comment cela est possible ; il y a là-dessous comme un mystère.

Enfin, je veux faire un commentaire sur l'annexe. Celle-ci ne concerne que les processus voisins, mais un point important de la théorie radicalement élémentaire est qu'elle traite également de processus qui ne sont voisins d'aucun processus standard, par exemple, le processus "désastreux" du chapitre 7.

Fine Hall, 2003

Chapitre 1

Variables aléatoires

Commençons par donner quelques définitions et quelques inégalités à la base de la théorie des probabilités, dans le cadre des espaces probabilisés finis.

Un *espace probabilisé fini* est la donnée d'un ensemble fini Ω et d'une fonction pr telle que :

- pr est une fonction réelle strictement positive définie sur Ω ;
- $\sum pr(\omega) = 1$.

Une *variable aléatoire* x sur Ω est une fonction réelle définie sur Ω. L'*espérance*, ou *moyenne*, d'une variable aléatoire x est définie par la relation :

$$Ex \overset{def}{=} \sum_{\omega \in \Omega} x(\omega) pr(\omega).$$

Un *événement* est un sous-ensemble de Ω. La probabilité d'un événement A est la somme des probabilités des éléments de A :

$$Pr(A) \overset{def}{=} \sum_{\omega \in A} pr(\omega).$$

Etant donné un événement A, la fonction indicatrice de A :

$$\chi_A : \Omega \longrightarrow \{0, 1\}$$

$$(\chi_A(\omega) = 1) \Longleftrightarrow (\omega \in A)$$

est une variable aléatoire sur Ω. On a alors $Pr(A) = E\chi_A$. On désigne par A^c le complémentaire de A dans Ω, c'est-à-dire, l'ensemble des éléments de Ω qui n'appartiennent pas à A.

Si n est le cardinal de Ω, l'ensemble R^Ω des variables aléatoires sur Ω forme un espace vectoriel de dimension n. Sur cet espace vectoriel nous définissons un produit scalaire $\langle . , . \rangle$ par la relation suivante :

$$\forall x \in R^\Omega, \forall y \in R^\Omega, \langle x, y \rangle = Exy.$$

On vérifie aisément la bilinéarité et la symétrie de cette application. Comme $\langle ., . \rangle$ est non-dégénéré ($\langle x, x \rangle$ est nul si et seulement si $x = 0$), l'espace R^Ω muni de ce produit est un espace euclidien, et on a :

$$\forall x \in R^\Omega, \|x\|_2 = \sqrt{Exx}.$$

Comme l'espérance est une forme linéaire de R^Ω dans R, l'ensemble des variables aléatoires de moyenne nulle (le noyau de cette forme) est un hyperplan. La droite orthogonale à cet hyperplan est l'espace des fonctions constantes de R^Ω dans R. Ainsi, l'application $x \mapsto Ex$ est la projection orthogonale de R^Ω sur l'espace des variables aléatoires constantes, que l'on peut identifier à R, et l'application $x \mapsto x - Ex$ est la projection orthogonale de R^Ω sur l'espace des variables aléatoires de moyenne nulle.

Nous définissons également les fonctions suivantes, x et y étant des variables aléatoires :

- variance de x : $var(x) = E((x - Ex)^2)$;
- écart-type de x : $\sigma(x) = \sqrt{var(x)}$;
- covariance de x et de y : $cov(x, y) = E((x - Ex)(y - Ey))$;
- coefficient de corrélation entre x et y :$\rho(x, y) = \frac{cov(x,y)}{\sigma(x)\sigma(y)}$.

Lorsque les variables aléatoires x et y sont de moyenne nulle, l'écart-type de x est égal à sa norme euclidienne, la covariance de x et de y est égale au produit scalaire de x et y, et le coefficient de corrélation entre x et y est égal au cosinus de l'angle que forment entre eux x et y.

D'autres normes définies sur l'espace R^Ω sont souvent utiles. Ainsi, pour tout nombre réel p de l'intervalle $[1, +\infty[$, on construit une norme $\|.\|_p$ en posant :

$$\|x\|_p \stackrel{def}{=} \sqrt[p]{E(|x|^p)}.$$

On crée également une norme $\|.\|_\infty$ grâce à la relation :

$$\|x\|_\infty \stackrel{def}{=} \max_{\omega \in \Omega} |x(\omega)|.$$

On voit que :

$$\forall p \in [1, +\infty[, \|x\|_p \leq \|x\|_\infty. \tag{1.1}$$

De plus, si ω_0 est un point de Ω où x atteint son maximum, alors :

$$\forall p \in [1,+\infty[, \|x\|_p^p \geq (|x(\omega_0)|^p pr(\omega_0)) \tag{1.2}$$

et

$$\lim_{p\to+\infty} \sqrt[p]{(|(x(\omega_0)|^p pr(\omega_0))} = |x(\omega_0)| = \|x\|_\infty.$$

De 1.1 et 1.2, on déduit :

$$\forall x \in R^\Omega, \lim_{p\to+\infty} \|x\|_p = \|x\|_\infty.$$

Pour tout réel p de l'intervalle $]1,+\infty[$, on définit l'exposant conjugué de p comme étant le réel q, contenu dans l'intervalle $]1,+\infty[$, tel que $\frac{1}{p}+\frac{1}{q}=1$. Alors, l'inégalité de Hölder nous dit :

$$\forall x \in R^\Omega, \forall y \in R^\Omega, |Exy| \leq \|x\|_p \|y\|_q. \tag{1.3}$$

En effet, si l'une ou l'autre de x ou de y est uniformément nulle, l'inégalité est triviale. Sinon, on peut supposer que $\|x\|_p = 1$ et $\|y\|_q = 1$, en remplaçant la variable x par la variable $\frac{x}{\|x\|_p}$, et la variable y par la variable $\frac{y}{\|y\|_q}$. Démontrer l'inégalité de Hölder revient donc à démontrer que :

$$\forall x \in R^\Omega, \forall y \in R^\Omega, ((\|x\|_p = 1) \text{ et } (\|y\|_q = 1)) \Rightarrow (|Exy| \leq 1). \tag{1.4}$$

Or, du fait de la concavité du logarithme, et que p et q sont conjugués, on a la relation suivante :

$$\forall \omega \in \Omega, \frac{1}{p}\ln(|x(\omega)|^p) + \frac{1}{q}\ln(|y(\omega)|^q) \leq \ln(\frac{1}{p}|x(\omega)|^p + \frac{1}{q}|y(\omega)|^q)$$

soit :

$$\forall \omega \in \Omega, |x(\omega)y(\omega)| \leq \frac{1}{p}|x(\omega)|^p + \frac{1}{q}|y(\omega)|^q$$

d'où il vient :

$$E|xy| \leq \frac{1}{p}\|x\|_p^p + \frac{1}{q}\|y\|_q^q. \tag{1.5}$$

Comme $\|x\|_p = 1$, $\|y\|_q = 1$, p et q sont conjugués, et $|Exy| \leq E|xy|$, la relation 1.5 permet de déduire la relation 1.4, ce qui conclut la démonstration de l'inégalité de Hölder.

L'inégalité 1.3 est une inégalité stricte, à moins que :

- $|Exy| = E|xy|$, ce qui se produit si et seulement si les fonctions x et y sont systématiquement de même signe ;

et :

- $\forall \omega \in \Omega, |x(\omega)|^p = |y(\omega)|^q$.

Si la variable x vérifie $\|x\|_p = 1$, et si la variable y est définie par la relation

$$\forall \omega \in \Omega, y(\omega) = signe(x(\omega))|x(\omega)|^{\frac{p}{q}},$$

alors $Exy = 1$, et l'égalité dans 1.3 est atteinte. Ainsi, si $\|x\|_p = 1$, on a $\max_{\|y\|_q=1}(Exy) = 1$. D'où il vient :

$$\forall x \in R^\Omega, \|x\|_p = \max_{\|y\|_q=1} |Exy|. \tag{1.6}$$

L'*inégalité de Minkowski* découle immédiatement de la proposition 1.6 :

$$\forall p \in]1, +\infty[, \forall x \in R^\Omega, \forall y \in R^\Omega, \|x+y\|_p \leq \|x\|_p + \|y\|_p. \tag{1.7}$$

Posons que 1 et ∞ sont exposants conjugués l'un de l'autre. Avec cette convention, nous voyons que les propositions 1.3, 1.6, et 1.7 sont valables pour tout exposant p dans l'intervalle fermé $[1, \infty]$.

Soit f une fonction convexe de R dans R. Alors,

$$\forall x \in R^\Omega, f(\sum_{\omega\in\Omega} x(\omega)pr(\omega)) \leq \sum_{\omega\in\Omega} pr(\omega)f(x(\omega))$$

c'est-à-dire :

$$f(Ex) \leq Ef(x). \tag{1.8}$$

La relation 1.8 porte le nom d'*inégalité de Jensen*. Si nous choisissons, pour p dans l'intervalle $[1, +\infty[$, la fonction $f : t \mapsto |t|^p$, l'inégalité de Jensen implique la relation suivante valable pour toute variable aléatoire x :

$$\forall p \in [1, +\infty[, |Ex|^p \leq E|x|^p.$$

Si nous appliquons cette dernière relation à la variable $y = |x|$, on obtient :

$$\forall p \in [1, +\infty[, (E|x|)^p \leq E|x|^p,$$

soit

$$\forall p \in [1, +\infty[, \|x\|_1 \leq \|x\|_p.$$

Et si nous appliquons la même relation à la variable $y = |x|^r$, r étant un réel quelconque de l'intervalle $[1, +\infty[$, on voit de la même manière que :

$$\forall p \in [1, +\infty[, \forall r \in [1, +\infty[, \|x\|_r \leq \|x\|_{rp}$$

d'où l'on déduit que, quelle que soit la variable aléatoire x, la fonction

$$\begin{array}{rcl} [1,+\infty] & \to & R^+ \\ p & \mapsto & \|x\|_p, \end{array}$$

est une fonction croissante de p.

Soit f une fonction positive. La proposition suivante est vraie :

$$\forall\lambda > 0, \forall x \in R^{\Omega},$$
$$Ef(x) \geq \sum_{\{\omega : f(x(\omega)) \geq \lambda\}} f(x(\omega))pr(\omega) \geq \lambda Pr\{\omega : f(x(\omega)) \geq \lambda\}$$

ce qui implique :

$$\forall\lambda > 0, \forall x \in R^{\Omega}, Pr\{f(x) \geq \lambda\} \leq \frac{Ef(x)}{\lambda}. \tag{1.9}$$

Dans cette relation, "$Pr\{f(x) \geq \lambda\}$" est une abréviation pour désigner la valeur $Pr\{\omega : f(x(\omega)) \geq \lambda\}$. Il est de coutume en théorie des probabilités de procéder à de tels raccourcis d'écriture.

Comme les ensembles $\{\omega : |x(\omega)| \geq \lambda\}$ et $\{\omega : |x(\omega)|^p \geq \lambda^p\}$ sont identiques pour tout réel p strictement positif, la proposition 1.9 nous permet d'écrire l'*inégalité de Bienaymé-Tchebychev* :

$$\forall x \in R^{\Omega}, \forall\lambda > 0, \forall p > 0, Pr\{|x| \geq \lambda\} \leq \frac{E(|x|^p)}{\lambda^p}.$$

Chapitre 2

Algèbres de variables aléatoires

L'ensemble R^Ω des variables aléatoires sur Ω est plus qu'un espace vectoriel. Il est en effet muni d'une structure d'algèbre. Nous conviendrons qu'une algèbre de variables aléatoires sur Ω contient toujours les fonctions constantes. Ainsi une sous-algèbre $\mathcal{A}$ de variables aléatoires est un sous-ensemble de R^Ω stable pour l'addition ($x + y$ est dans $\mathcal{A}$ si x et y sont dans $\mathcal{A}$), stable pour la multiplication (xy est dans $\mathcal{A}$ si x et y sont dans $\mathcal{A}$), et qui contient les fonctions constantes.

La structure d'une sous-algèbre $\mathcal{A}$ est très simple. Nous appelons *atome de* $\mathcal{A}$ tout événement (partie de Ω) maximal A, tel que tout élément de $\mathcal{A}$ soit constant sur A. Etant donnée une sous-algèbre $\mathcal{A}$, l'ensemble des atomes de $\mathcal{A}$, noté $at(\mathcal{A})$, forme une partition de Ω :

$$\begin{cases} \bigcup_{A \in at(\mathcal{A})} A = \Omega; \\ \forall A \in at(\mathcal{A}), \forall B \in at(\mathcal{A}), (A \neq B) \Rightarrow (A \cap B = \emptyset). \end{cases}$$

Si A est un atome de $\mathcal{A}$, c'est un événement maximal en ce sens qu'il n'existe pas de partie de Ω contenant strictement A, et sur laquelle chaque variable de $\mathcal{A}$ serait constante. Pour tout élément $\omega \in A^c$, il existe donc une fonction x de $\mathcal{A}$ telle que $x(\omega) \neq x(A)$. Soit la variable aléatoire définie par la relation :

$$x_\omega = \frac{x - x(\omega)}{x(A) - x(\omega)}.$$

Cette fonction satisfait les propriétés suivantes :

$$\begin{cases} x_\omega(A) = 1; \\ x_\omega(\omega) = 0; \end{cases}$$

d'où l'on déduit l'égalité $\chi_A = \prod_{\omega \notin A} x_\omega$. Il s'ensuit que pour tout atome A de $\mathcal{A}$ la fonction indicatrice de A appartient à $\mathcal{A}$, et $\mathcal{A}$ est composée de toutes les fonctions de R^Ω constantes sur chaque élément de $at(\mathcal{A})$.

Inversement, étant donnée une partition P de Ω, l'ensemble de toutes les fonctions de R^Ω constantes sur chaque élément de P constitue une sous-algèbre de R^Ω. Il est par conséquent équivalent de se donner une algèbre de variables aléatoires sur Ω ou une partition de Ω.

Remarquons qu'une sous-algèbre quelconque $\mathcal{A}$ de variables aléatoires contient toute fonction de ses éléments : en effet, quelle que soit la fonction $f : R^n \to R$ et quelles que soient les variables aléatoires $x_1, \cdots, x_n$ de $\mathcal{A}$, la variable $f(x_1, \cdots, x_n)$ est un élément de $\mathcal{A}$.

Illustrons ce que nous venons d'exposer par un modèle du jeu de lancer de dés. Définissons l'ensemble Ω comme étant l'ensemble de tous les couples (i, j) tels que i et j soient des entiers compris dans l'intervalle $[1, 6]$, avec $pr(i, j) = \frac{1}{36}$. Définissons les variables aléatoires x, y, et z par :

- $x(i, j) = i$;
- $y(i, j) = j$;
- $z(i, j) = i + j$.

Soit $\mathcal{A}$ la sous-algèbre de R^Ω engendrée par z (c'est la plus petite sous-algèbre contenant à la fois z et les fonctions constantes). La figure 2.1 représente les atomes de $\mathcal{A}$. L'ensemble $at(\mathcal{A})$ contient 11 atomes, que nous désignons par $A_2, ..., A_{12}$. Dans cette notation, l'indice utilisé correspond à la valeur prise par z. Revenons au cas général. Soit $\mathcal{A}$ une sous-algèbre quelconque de R^Ω. Pour tout atome A de $\mathcal{A}$, construisons la fonction :

$$\begin{cases} pr_A : A & \longrightarrow & [0, 1] \\ \omega & \longmapsto & \frac{pr(\omega)}{Pr(A)}. \end{cases}$$

Muni de cette fonction, A forme un espace probabilisé. Ainsi, la structure d'espace probabilisé de Ω induit sur chaque atome de $\mathcal{A}$ une structure d'espace probabilisé. En utilisant cette propriété, nous serons en mesure de rendre chaque construction et chaque résultat de la théorie des probabilités relatifs à une sous-algèbre $\mathcal{A}$ de variables aléatoires. Lors d'une telle relativisation, les fonctions de $\mathcal{A}$ tiendront lieu de fonctions constantes.

Ainsi pouvons nous définir l'*espérance conditionnelle*, ou *moyenne conditionnelle*, relativement à $\mathcal{A}$, d'une variable aléatoire x de la manière suivante. $E_\mathcal{A}x$ est un élément de $\mathcal{A}$, tel que la valeur prise par $E_\mathcal{A}x$ sur un atome A de $\mathcal{A}$ soit égale à la moyenne de x sur l'espace probabilisé A :

$$\forall x \in R^\Omega, \forall x \in R^\Omega, \forall \omega \in \Omega, \forall A \in at(\mathcal{A}),$$
$$(\omega \in A) \Rightarrow (E_\mathcal{A}x(\omega) = \textstyle\sum_{\xi \in A} x(\xi) pr_A(\xi)).$$

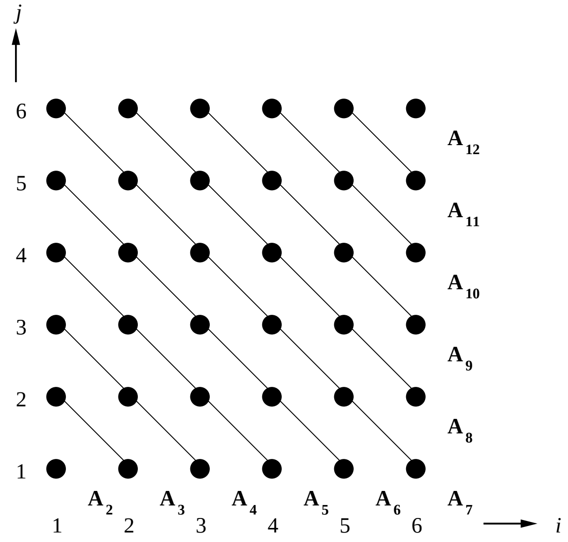

FIG. 2.1 – lancer de dés

De la même manière que l'espérance est une forme linéaire :

$$\forall (x_1, x_2, \lambda_1, \lambda_2) \in R^\Omega \times R^\Omega \times R \times R,$$
$$E(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 E x_1 + \lambda_2 E x_2,$$

l'espérance conditionnelle est une forme $\mathcal{A}$-linéaire :

$$\forall (x_1, x_2, y_1, y_2) \in R^\Omega \times R^\Omega \times \mathcal{A} \times \mathcal{A},$$
$$E_\mathcal{A}(y_1 x_1 + y_2 x_2) = y_1 E_\mathcal{A} x_1 + y_2 E_\mathcal{A} x_2.$$

On dit que l'espérance préserve les constantes pour exprimer le fait que cet opérateur est neutre sur l'espace des variables aléatoires constantes :

$$\forall \lambda \in R, E\lambda = \lambda.$$

De la même manière, on dit que l'espérance conditionnelle relative à une sous-algèbre $\mathcal{A}$ préserve les éléments de $\mathcal{A}$:

$$\forall y \in \mathcal{A}, E_\mathcal{A} y = y.$$

L'espérance est l'opérateur de projection orthogonale de l'espace euclidien R^Ω sur la droite des fonctions constantes. C'est-à-dire que l'on a :

$$\forall x \in R^\Omega, \forall \lambda \in R, E((x - Ex)^2) \leq E((x - \lambda)^2).$$

La relativisation de cette proposition à la sous-algèbre $\mathcal{A}$ s'écrit :

$$\forall x \in R^{\Omega}, \forall y \in \mathcal{A}, E_{\mathcal{A}}((x - E_{\mathcal{A}}x)^2) \leq E_{\mathcal{A}}((x-y)^2).$$

Si $\mathcal{B}$ est une sous-algèbre incluse dans $\mathcal{A}$, alors $E_{\mathcal{B}}E_{\mathcal{A}} = E_{\mathcal{B}}$. En particulier, comme l'ensemble des constantes forme une sous-algèbre (et du fait que nous avons convenu que toute sous-algèbre de variables aléatoires contraindrait les constantes), on a $EE_{\mathcal{A}} = E$. De ce fait, comme par ce qui précède :

$$\forall x \in R^{\Omega}, \forall y \in \mathcal{A}, EE_{\mathcal{A}}((x - E_{\mathcal{A}}x)^2) \leq EE_{\mathcal{A}}((x-y)^2),$$

on en déduit :

$$\forall x \in R^{\Omega}, \forall y \in \mathcal{A}, E((x - E_{\mathcal{A}}x)^2) \leq E((x-y)^2).$$

Autrement dit, l'opérateur $E_{\mathcal{A}}$ est l'opérateur de projection orthogonale de R^{Ω} sur le sous-espace vectoriel $\mathcal{A}$.

On emploie également, pour désigner $E_{\mathcal{A}}x$, la notation $E\{x|\mathcal{A}\}$, et si $\mathcal{A}$ est engendrée par les variables $x_1, \cdots, x_n$, on peut écrire $E\{x|x_1, \cdots, x_n\}$ également.

Revenons à notre exemple du jeu de dés, illustré par la figure 2.1. Pour une raison de symétrie, $E\{x|z\} = E\{y|z\}$, et donc, $E\{x|z\} = \frac{E\{z|z\}}{2} = \frac{z}{2}$.

Relativement à $\mathcal{A}$, pour un événement B quelconque de Ω, nous définissons la *probabilité conditionnelle* de B comme étant l'élément de $\mathcal{A}$ défini par :

$$Pr_{\mathcal{A}}B \stackrel{def}{=} E_{\mathcal{A}}\chi_B.$$

Cette fonction vérifie donc la propriété :

$$\forall \omega \in \Omega, \forall A \in at(\mathcal{A}), (\omega \in A) \Rightarrow \left(Pr_{\mathcal{A}}B(\omega) = \frac{Pr(A \cap B)}{Pr(A)}\right).$$

Voici comment s'expriment quelques inégalités que nous avons rencontrées dans le chapitre précédent, lorsque nous les rendons relatives à la sous-algèbre $\mathcal{A}$:

- inégalité de Hölder, p et q étant conjugués :

$$\forall (x,y) \in R^{\Omega} \times R^{\Omega}, |E_{\mathcal{A}}xy| \leq \sqrt[p]{E_{\mathcal{A}}(|x|^p)}\sqrt[q]{E_{\mathcal{A}}(|y|^q)}.$$

- inégalité de Jensen, f étant une fonction réelle convexe :

$$\forall x \in R^{\Omega}, f(E_{\mathcal{A}}x) \leq E_{\mathcal{A}}f(x).$$

- inégalité de Bienaymé-Tchébychev, f étant une fonction positive :

$$\forall x \in R^{\Omega}, \forall \lambda > 0, Pr_{\mathcal{A}}\{f(x) \geq \lambda\} \leq \frac{E_{\mathcal{A}} f(x)}{\lambda}.$$

L'inégalité de Jensen nous dit que $|E_{\mathcal{A}} x|^p \leq E_{\mathcal{A}}(|x|^p)$, et donc puisque $EE_{\mathcal{A}} = E$, il en découle $\|E_{\mathcal{A}} x\|_p \leq \|x\|_p$, cette identité étant valable pour toute valeur de p prise dans l'intervalle $[1, +\infty]$. On résume cela en disant que l'espérance réduit les normes.

Nous avons vu que la structure d'espace probabilisé sur Ω induit une structure d'espace probabilisé sur chaque élément de $at(\mathcal{A})$, en permettant de définir des fonctions de probabilité locales pr_A. Il est également possible de doter l'ensemble $at(\mathcal{A})$ des atomes de $\mathcal{A}$ d'une structure d'espace probabilisé, $\langle at(\mathcal{A}), pr'_{\mathcal{A}}\rangle$, en définissant $pr'_{\mathcal{A}}$ ainsi :

$$\forall A \in at(\mathcal{A}), pr'_{\mathcal{A}}(A) \stackrel{def}{=} Pr(A).$$

On dit que l'espace $\langle \Omega, pr\rangle$ est un *espace fibré* au-dessus de $\langle at(\mathcal{A}), pr'_{\mathcal{A}}\rangle$ (admettant pour support $\langle at(\mathcal{A}), pr'\mathcal{A}\rangle$), et dont les espaces $\langle A, pr_A\rangle$ sont les *fibres*. Dans l'exemple de la figure 2.1, les fibres apparaissent comme les alignements de points inclinés de 45 degrés dans le sens des aiguilles d'une montre.

L'opérateur espérance au sein de l'espace $\langle at(\mathcal{A}), pr'_{\mathcal{A}}\rangle$ est noté $E'_{\mathcal{A}}$, et la probabilité d'un ensemble d'atomes est notée $Pr'_{\mathcal{A}}$. On vérifie que $E'_{\mathcal{A}} E_{\mathcal{A}} = E$.

Le produit de deux espaces probabilisés $\langle \Omega_1, pr_1\rangle$ et $\langle \Omega_2, pr_2\rangle$ est un cas particulier d'espace fibré. On obtient ce produit en dotant $(\Omega_1 \times \Omega_2)$ de la fonction de probabilité $(pr_1 \times pr_2)$:

$$\forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2, (pr_1 \times pr_2)(\omega_1, \omega_2) = pr_1(\omega_1) pr_2(\omega_2).$$

Soit $\mathcal{A}_1$ la sous-algèbre des variables aléatoires ne dépendant pas de ω_2. Alors l'ensemble $at(\mathcal{A}_1)$ est constitué de tous les ensembles de la forme $\{(\omega_1, \omega_2) : \omega_1 = \eta_1\}$, où η_1 parcourt Ω_1, soit :

$$at(\mathcal{A}_1) = \{\{(\omega_1, \omega_2), \omega_2 \in \Omega_2\}, \omega_1 \in \Omega_1\}.$$

Chapitre 3

Processus stochastiques

Dans le contexte qui nous occupe ici, "stochastique" se référera à la notion d'aléa, et "processus" voudra simplement dire "fonction".[1] Nous disons donc qu'un processus stochastique est une fonction dont les valeurs sont des variables aléatoires. Plus formellement, soit T un ensemble fini et $\langle \Omega, pr \rangle$ un espace probabilisé. Un *processus stochastique indexé par T et défini sur* $\langle \Omega, pr \rangle$ est une fonction $\xi : T \to R^{\Omega}$. Nous conviendrons dans la suite qu'un processus stochastique est toujours indexé par un ensemble fini, et toujours défini sur un espace probabilisé fini.

La notation $\xi(t, \omega)$ désignera la valeur prise par la variable aléatoire $\xi(t)$ au point ω, et la notation $\xi(., \omega)$ désignera, pour chaque ω de Ω, la fonction :

$$\left\{ \begin{array}{rcl} \xi(., \omega) : T & \longrightarrow & R \\ t & \longmapsto & \xi(t, \omega). \end{array} \right.$$

Ainsi, chaque $\xi(t)$ est une variable aléatoire, chaque $\xi(t, \omega)$ est un nombre réel, et chaque $\xi(., \omega)$ est une fonction réelle définie sur T, appelée *trajectoire du processus* ξ.

Soit Λ_ξ l'ensemble des trajectoires du processus ξ. Il constitue un sous-ensemble fini de l'espace vectoriel R^T des fonctions réelles définies sur T. Ce dernier espace est de dimension finie. La fonction de probabilité pr_ξ définie

[1] Etymologiquement, le mot "stochastique" dérive du grec $\sigma\tau o\chi\alpha\sigma\tau\iota\kappa o\varsigma$, "qui vise bien", "habile à conjecturer". De nos jours, "l'adjectif s'applique à ce qui est partiellement produit par le hasard et à ce qui comporte la présence d'une variable aléatoire, ..." (Le Robert, Dictionnaire historique de la langue française).

Le terme "processus", quant à lui, tire ses origines du latin, et a subi jusqu'à nos jours une histoire un peu plus compliquée (voir le même dictionnaire). Cependant, "...il est passé dans l'usage courant en parlant d'un ensemble de phénomènes se déroulant dans le même ordre". (NdT) par :

$$\begin{cases} pr_\xi : \Lambda_\xi & \longrightarrow & R \\ \quad \lambda & \longmapsto & Pr\{\omega : \forall t \in T, \xi(t,\omega) = \lambda(t)\} \end{cases}$$

permet de doter l'ensemble Λ_ξ d'une structure d'espace probabilisé. Muni de cette structure, nous désignerons cet espace par la notation $\langle \Lambda_\xi, pr_\xi \rangle$.

Nous dirons que deux processus stochastiques ξ et ξ', indexés par un même ensemble T mais éventuellement définis sur deux espaces probabilisés distincts, sont *équivalents* si $(\Lambda_\xi = \Lambda_{\xi'})$ et $(pr_\xi = pr_{\xi'})$. Cela revient à dire que deux processus sont équivalents si et seulement si ils ont les mêmes trajectoires, avec les mêmes probabilités d'occurrence. En théorie des probabilités, on ne s'intéresse qu'aux propriétés communes aux processus équivalents, donc aux propriétés des classes d'équivalence des processus si l'on convient de définir ces classes par la relation d'équivalence décrite plus haut. Comme la fonction ψ :

$$\psi : T \longrightarrow R^{\Lambda_\xi}$$

$$\forall t \in T, \forall \omega \in \Omega, \psi(t)(\xi(.,\omega)) = \xi(t,\omega)$$

est un processus stochastique indexé par T, défini sur $\langle \Lambda_\xi, pr_\xi \rangle$, et équivalent à ξ, on peut tout à fait admettre, lorsque l'on étudie un processus stochastique, que celui-ci est défini sur l'ensemble probabilisé de ses trajectoires.

Lorsque T se résume à un seul élément, $T = \{t_0\}$, on peut confondre le processus stochastique ξ avec la variable aléatoire $\xi(t_0)$. On peut également confondre toute variable aléatoire x sur Ω avec un processus stochastique indexé par un singleton :

$$\Lambda_x = \{\lambda \in R : Pr\{x = \lambda\} \neq 0\}$$

$$\forall \lambda \in \Lambda_x, pr_x(\lambda) = Pr\{x = \lambda\}.$$

On voit facilement que :

$$Ex = \sum \lambda \, pr_x(\lambda). \tag{3.1}$$

Le membre de droite de l'égalité 3.1 est la moyenne de la fonction identité, mesurée sur l'espace $\langle \Lambda_x, pr_x \rangle$. Si l'égalité 3.1 n'était pas vérifiée, alors, en vertu du paragraphe précédent, la théorie des probabilité ne se préoccuperait pas de l'espérance des variables aléatoires.

Les valeurs d'un processus stochastique ξ sont qualifiées d'*indépendantes* si la proposition suivante est vérifiée :

$$\forall \lambda \in \Lambda_\xi, pr_\xi(\lambda) = \prod_{t \in T} pr_{\xi(t)}(\lambda(t)).$$

Donnons un exemple. Supposons que $T = \{1, \cdots, \nu\}$, et soit $\langle \Omega, pr \rangle$ un espace probabilisé. Définissons la fonction pr^ν par :

$$\left\{ \begin{array}{rcl} pr^\nu : \Omega^\nu & \longrightarrow & [0,1] \\ (\omega_1, \cdots, \omega_\nu) & \longmapsto & \prod_{i \in \{1, \cdots, \nu\}} pr(\omega_i). \end{array} \right.$$

Alors $\langle \Omega^\nu, pr^\nu \rangle$ est un espace probabilisé fini. Soit x_0 une variable aléatoire sur Ω, on voit que les variables aléatoires x_n, où $1 \leq n \leq \nu$, définies sur Ω^ν par :

$$\forall (\omega_1, \cdots, \omega_\nu) \in \Omega^\nu, x_n(\omega_1, \cdots, \omega_\nu) = x_0(\omega_n)$$

sont indépendantes. La séquence $(x_n)_{n \in \{1, \cdots, \nu\}}$ forme un processus stochastique qui modélise ν observations indépendantes d'une même variable aléatoire x_0.

On dit que ν événements $A_1, \cdots, A_\nu$ sont indépendants si leurs fonctions indicatrices sont indépendantes. Dans notre exemple du jeu de dés (fig. 2.1) désignons par A l'événement "i est pair", par B l'événement "j est pair", et par C l'événement "$i + j$ est pair". A et B sont indépendants, A et C sont indépendants, B et C sont indépendants. Cependant, A, B, et C ne sont pas indépendants. La connaissance de A seul ne nous apporte aucune information sur C, de même que la connaissance de B seul ne nous apporte aucune information sur C. En revanche, la connaissance conjointe de A et B nous renseigne complètement quant à C. Ici réside le secret des bons détectives.

Chapitre 4

Concepts externes

Soit x_0 une variable aléatoire de moyenne nulle et de variance unité. Soit $(x_1, \cdots, x_\nu)$ une séquence de ν réalisations indépendantes de x_0. Considérons la phrase suivante :

> *"Si ν est un grand entier, alors presque sûrement, pour tous les grands entiers n inférieurs ou égaux à ν, le rapport $\frac{\sum_{i=1}^n x_i}{n}$ est presque égal à 0."*

Ce qui précède est une formulation intuitive de la loi forte des grands nombres. Cette formulation n'est pas précise, car nous n'avons pas fixé le sens de "grand entier", "presque sûrement", et "presque égal".

Voici un aperçu de la manière dont on exprime classiquement la loi forte des grands nombres. On remplace la séquence finie d'observations

$$(x_i)_{i \in \{1, \cdots, \nu\}}$$

par une suite infinie d'observations $(x_i)_{i \in N}$. Pour aboutir à cela, il faut construire un produit infini de l'espace probabilisé Ω avec lui-même, soit Ω^N, qui possédera des propriétés originales par rapport à Ω :

- bien que Ω soit un ensemble fini, le résultat du produit infini est un ensemble non dénombrable, dès que Ω contient plus d'un élément ;
- dans ce nouvel espace, la probabilité associée à un point quelconque est nulle ;
- seules certaines parties de Ω^N, forcément mesurables, pourront être qualifiées d' "événement" ;
- la probabilité d'un événement ne se définit plus comme la somme des probabilités de chacun de ses éléments ;
- toutes les fonctions réelles de Ω^N ne pourront pas prétendre au statut de variables aléatoires, celles-ci devant être mesurables. De plus,

pour certaines variables aléatoires, il ne sera pas possible de calculer l'espérance.

Avec tout ceci, la loi forte des grands nombres dit que, en dehors d'un événement de probabilité nulle, on a :

$$\forall \varepsilon \in R^{*+}, \exists m \in N, \forall n \in N, (n > m) \Rightarrow \left(\frac{1}{n}\sum_{i=1}^{n} x_i \leq \varepsilon\right).$$

Nous proposons de procéder différemment, et notre approche présentera l'avantage de rester dans le cadre élémentaire des espaces probabilisés finis. Nous revenons donc à notre séquence $(x_i)_{i\in\{1,\cdots,\nu\}}$, mais nous imposons à ν d'être un entier "non standard". Nous allons préciser ce que recouvre ce terme très bientôt.

Nous dirons aussi qu'un nombre réel est infinitésimal si sa valeur absolue est inférieure à l'inverse d'un entier non standard, et nous dirons que deux nombres sont presque égaux si leur différence est infinitésimale.

Nous dirons qu'une propriété est vérifiée presque sûrement, au sein d'un espace probabilisé, si quelque soit le réel non infinitésimal et non nul ε, il existe un événement N de probabilité inférieure à ε tel que la propriété soit vérifiée pour tous les éléments du complémentaire de N. Ou formellement, en notant $Prop$ la propriété, si l'on a :

$$\forall^{st}\varepsilon > 0, \exists N, (Pr(N) < \varepsilon) \text{ et } (\forall \omega \in N^c, Prop(\omega)).^1$$

Alors, si l'on définit le fait d'être un grand entier comme étant équivalent à être un entier non standard, la formulation de la loi des grands nombres coïncide avec notre formulation intuitive.

L'approche conventionnelle que nous avons présentée plus haut procède par idéalisation, car il est impossible de procéder à une suite infinie d'observations. L'autre approche procède également par idéalisation, car vous ne pourrez jamais "voir" un entier non standard. Il est de la nature même des mathématiques de procéder par idéalisations. Différents critères esthétiques entrent en compte dans le choix d'un formalisme plutôt que d'un autre pour exprimer une propriété (par exemple la simplicité, la proximité de l'énoncé avec l'idée fondamentale sous-jacente, la puissance, ...), mais différents formalismes ne s'excluent pas forcément, et il peut être très bénéfique de reconsidérer des résultats classiques à la lumière de concepts nouveaux.

Voyons maintenant d'où provient cette idée d'objets "standard". N est l'ensemble des entiers naturels $\{0, 1, 2, 3, \ldots\}$. La propriété fondamentale de

[1] La notation "$\forall^{st}x$..." s'entend comme "pour tout élément standard x...". (NdT)

cet ensemble nous est offerte par le théorème d'induction, qui affirme que "si une partie S de N contient 0, et si tout élément s de S est tel que $(s+1)$ est dans S, alors $S = N$".

Si $A(n)$ est une proposition qui respecte le langage mathématique conventionnel (par exemple, "n et $(n+2)$ sont premiers", ou "$n > m$"), alors il est licite de considérer l'ensemble S tel que :

$$S = \{n \in N : A(n)\}.$$

Il est important que A soit bien formée, du point de vue des règles de construction des propositions mathématiques. Un ensemble est un objet formel, qui ne peut exister que s'il dispose d'une définition formelle au sein du système axiomatique que l'on a adopté. Ainsi un ensemble défini par la propriété $A(n)$: "n n'est pas très grand selon moi" n'est pas concevable en mathématiques.

Il découle des travaux menés par Gödel dans les années 30 que les objets "naturels" des mathématiques, comme l'ensemble N, ne peuvent être entièrement décrits par un système axiomatique quel qu'il soit. Pour comprendre ce que cela veut dire, adjoignons à notre système d'axiomes initial un nouveau prédicat, "standard", non défini. Ainsi, la proposition "$standard(n)$" (c'est-à-dire, n est standard) n'a aucune signification en mathématiques conventionnelles. Nous dirons d'un énoncé qu'il est *interne* s'il n'utilise pas le prédicat "standard", ni l'un quelconque de ses dérivés (il s'agit donc d'un énoncé des mathématiques classiques), et nous dirons d'un énoncé qu'il est *externe* s'il n'est pas interne.

"$standard(n)$" est l'exemple le plus simple d'une formule externe. "x est infinitésimal" en est un autre, puisque c'est une manière abrégée d'écrire "il existe un entier naturel non standard tel que son inverse soit supérieur à $|x|$". Pour définir un sous-ensemble, il n'est pas permis d'utiliser une formule externe (on ne peut pas parler, par exemple, de l'ensemble des entiers standard[2], ou de l'ensemble des réels infinitésimaux). Lorsque l'on viole cette règle, on dit que l'on procède à une formation illégale d'ensemble.

Nous poserons les axiomes suivants :

(1) 0 est standard ;

(2) pour tout entier naturel n, si n est standard, alors $(n+1)$ est standard ;

Néanmoins, il est impossible de démontrer que tout élément de N est standard. Ceci ne contredit pas le principe d'induction, car "être standard" n'est pas une propriété interne, mais exprime simplement le fait que l'on ne peut

[2]"standard" est un adjectif invariable. (NdT)

pas prouver l'existence d'un sous-ensemble S de N, tel qu'un entier naturel n soit dans S si et seulement si n est standard. Il n'est pas contradictoire d'admettre que :

(3) il existe un entier naturel qui n'est pas standard.

On admettra de plus que, A étant une formule interne ou externe :

(4) si $A(0)$, et si de plus, pour tout entier naturel standard n, $A(n)$ entraîne $A(n+1)$, alors pour tout entier naturel standard n, $A(n)$.

La proposition précédente porte le nom d'*induction externe*. Elle est complémentaire du théorème d'induction ordinaire, dont nous avons vu qu'il peut être mis en défaut lorsqu'on l'applique à des formules externes. Il va de soi que le théorème d'induction ordinaire continue d'être valable pour toute proposition ordinaire, c'est-à-dire pour toute proposition interne. Rien n'est remis en question en ce qui concerne les mathématiques classiques. Nous proposons simplement d'utiliser un langage étendu pour étudier les mêmes objets mathématiques que par le passé.

Le principe d'induction externe nous permet de prouver facilement que tout entier naturel non standard est supérieur à n'importe quel entier naturel standard (pour un entier naturel non standard m quelconque, examiner la proposition $A(n)$: $(n \leq m)$), que la somme de deux entiers naturels standard est standard (pour un entier naturel standard m quelconque, examiner la proposition $A(n)$: $standard(n+m)$), et que le produit de deux entiers naturels standard est standard (pour un entier naturel standard m quelconque, examiner la proposition $A(n)$: $standard(nm)$, en tenant compte de la propriété précédente relative à l'addition).

Nous utiliserons parfois le principe suivant, appelé “principe d'extension généralisé” (voir [5, page 79]). Soit $A(n, x)$ une formule interne ou externe. Supposons que pour tout entier naturel standard n, il existe x tel que $A(n, x)$. Alors, évidemment, il existe x_0 tel que $A(0, x_0)$, il existe x_1 tel que $A(1, x_1)$, il existe x_2 tel que $A(2, x_2)$, et ainsi de suite. Le principe d'extension généralisé pose que :

(*5) si pour tout entier naturel standard n, il existe x tel que $A(n, x)$, alors il existe une suite $(x_n)_{n \in N}$, telle que $\forall^{st} n \in N, A(n, x_n)$.

Ce principe sera mis en œuvre, par exemple, lors de la démonstration du théorème (6.1). Nous signalerons par un astérisque (*) chaque résultat qui utilise le principe d'extension généralisé.

Il découle du postulat (2) qu'il n'existe pas de plus petit entier non standard. Nous pouvons imaginer les entiers naturels comme des points alignés le long d'un ruban (fig 4.1). On peut se contenter de ne considérer que des entiers standard aussi longtemps que n'interviennent que des propositions internes. Mais N contient aussi bien des éléments standard que des éléments non standard. Notons que nous ne sommes pas partis de la portion gauche du ruban, auquel nous aurions adjoint une nouvelle partie droite. Nous avons plutôt considéré le ruban comme achevé, et enrichi notre langage d'un prédicat qui nous permet de distinguer deux portions du ruban.[3] On peut comparer le rôle de notre nouveau prédicat "standard" au rôle de la couleur dans la télévision : l'image reste la même, mais nous voyons apparaître des traits qui n'étaient pas exprimables en noir et blanc.

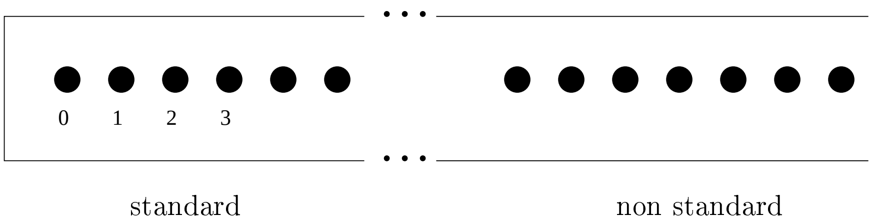

Fig. 4.1 – les entiers naturels

Pendant longtemps, les mathématiciens ont vu l'incomplétude des systèmes axiomatiques comme une malchance. On doit au génie d'Abraham Robinson d'avoir montré, dans les années 60, comment cette incomplétude pouvait être bénéfique en ouvrant de nouvelles voies pour simplifier les raisonnements mathématiques.

[3]Cette notion de "portion de ruban" est un peu particulière. En particulier, on ne peut pas construire de frontière entre les deux portions. (NdT)

Chapitre 5

Infinitésimaux

Nous introduisons dans ce chapitre quelques notions externes utiles dans le cadre du corps R des nombres réels.

On dit qu'un nombre réel x est *infinitésimal* s'il existe un entier naturel non standard ν, tel que $|x| \leq \frac{1}{\nu}$. Comme ν est supérieur à tout entier naturel standard, il s'ensuit que si x est infinitésimal il vérifie, pour tout naturel standard n, $|x| \leq \frac{1}{n}$. La proposition inverse est vraie. En effet, si x est nul, il est infinitésimal par définition. Sinon, soit μ le plus petit entier naturel tel que $|x| \geq \frac{1}{\mu}$. L'hypothèse entraîne que μ est non standard, donc $\nu = (\mu - 1)$ est également non standard. De plus, ν vérifie $|x| \leq \frac{1}{\nu}$, et ainsi x est infinitésimal.

On dit qu'un nombre réel x est *limité* s'il existe un entier naturel standard n tel que $|x| \leq n$. Dans le cas contraire, on dit que x est *illimité*. On rencontre parfois l'usage des termes "fini" et "infini", à la place respectivement de "limité" et "illimité". Cependant, nous préférons éviter cela. En effet, "fini" et "infini" possèdent chacun une signification interne précise. Quel sens devrions-nous accorder, par exemple, à un énoncé tel que "l'intégrale I est finie", si nous adoptions un tel usage ?

Lorsqu'un nombre n'est ni infinitésimal, ni illimité, nous dirons qu'il est *appréciable*.

x et y étant des nombres réels, nous adoptons les notations suivantes :

- $x \approx y$ lorsque $(x - y)$ est infinitésimal. On dit alors que x et y sont *réels équivalents* ;
- Lorsque x et y ne sont pas équivalents, nous dirons qu'ils sont *discernables* ;
- $x \lesssim y$ lorsqu'il existe un nombre infinitésimal α tel que $x \leq y + \alpha$ (c'est-à-dire que soit x est inférieur à y, soit x et y sont équivalents). On dit alors que x est *apparemment inférieur ou égal* à y ;

- $x \gtrsim y$ lorsque $y \lesssim x$. On dit alors que x est *apparemment supérieur ou égal* à y ;
- $x \ll y$ lorsque $x < y$ et que l'on n'a pas $x \approx y$. On dit alors que x est *sensiblement plus petit* que y ;
- $x \gg y$ lorsque $y \ll x$. On dit alors que x est *sensiblement plus grand* que y ;

De plus, lorsque $x \gg 0$, on dit que x est *sensiblement positif.*[1] La droite réelle achevée $\overline{R}$ est obtenue en adjoignant à R deux éléments notés $-\infty$ et ∞, et en étendant la relation d'ordre de R en posant :

$$\forall x \in R, -\infty < x < \infty.$$

Nous écrirons :

- $x \approx \infty$ lorsque x est positif et illimité ;
- $x \approx -\infty$ lorsque $-x \approx \infty$;
- $x \ll \infty$ (ou $\infty \gg x$) lorsque l'on a pas $x \approx \infty$;
- $x \gg -\infty$ (ou $-\infty \ll x$) lorsque l'on a pas $x \approx -\infty$;

Ainsi, il est équivalent d'écrire $|x| \approx \infty$ ou d'écrire que x est illimité. De même, il est équivalent d'écrire $|x| \ll \infty$ ou d'écrire que x est limité.

Traçons un segment de la droite réelle. Alors, quelque soit la précision que nous pourrons pratiquement apporter à nos mesures, deux points équivalents de ce segment nous apparaîtrons toujours comme égaux (les relations $=$ et $\approx$ sont indiscernables), une valeur apparemment inférieure ou égale à une autre nous apparaîtra comme inférieure ou égale à cette même autre (les relations $\leq$ et $\lesssim$ sont indiscernables), et une valeur sensiblement plus petite qu'une autre nous apparaîtra sans hésitation comme une valeur plus petite que cette autre (les relations $\geq$ et $\ll$ sont discernables).

Enonçons quelques propositions qui découlent immédiatement des définitions que nous venons de donner :

1. $(x \approx 0)$ si et seulement si x est infinitésimal ;
2. $(x \approx 0) \Longleftrightarrow (\forall \varepsilon \gg 0, |x| < \varepsilon)$;
3. les nombres infinitésimaux sont limités ;

[1] Dans le texte original, on propose "weakly less" pour $\lesssim$ et "strongly less" pour $\ll$, ce qui aurait pu être traduit respectivement par "faiblement inférieur" et "fortement inférieur". Cependant, la relation $x \lesssim y$ ne garantit pas que x soit vraiment inférieur ou égal à y, mais plutôt "qu'à l'œil nu" on ne peut nier que x soit inférieur ou égal à y. Autrement dit, "selon toute apparence", $x \leq y$, et c'est pourquoi je propose plutôt d'adopter la terminologie "apparemment inférieur ou égal". Le même type d'argument vaut pour la proposition "sensiblement inférieur". (NdT)

4. pour tout nombre réel non nul x, on a $(x \approx 0)$ si et seulement si $(\frac{1}{x}$ est illimité) ;
5. $(|x| \approx \infty) \iff \big((\frac{1}{x}) \approx 0\big)$;
6. si x et y sont limités, il en va de même pour $(x+y)$, ainsi que pour (xy) ;
7. si x et y sont infinitésimaux, il en va de même pour $(x+y)$ ainsi que pour (xy) ;
8. $((x \approx 0)$ et $(|y| \ll \infty)) \Longrightarrow (xy \approx 0)$;
9. $((x \lesssim y)$ et $(y \lesssim x)) \iff (x \approx y)$;
10. $((x \approx y)$ et $(y \approx z)) \Longrightarrow (x \approx z)$;
11. tout entier naturel est standard si et seulement si il est limité ;
12. tout entier naturel est non standard si et seulement si il est illimité.

Théorème 5.1. *Si n est un entier naturel standard, alors :*

$$(\forall i \in \{1, \cdots, n\}, (x_i \approx y_i)) \Rightarrow \left(\sum_{i=1}^{n} x_i \approx \sum_{i=1}^{n} y_i\right).$$

Démonstration. Il faut démontrer que $\sum_{i=1}^{n}(x_i - y_i) \approx 0$. Pour cela, il suffit de mettre en œuvre la proposition (7) ci-dessus, au sein du principe d'induction externe. □

Ce que nous venons de démontrer n'est généralement pas vrai si l'entier n est illimité. En guise de contre-exemple, il suffit de poser, pour n illimité, $x_i = 1/n$ et $y_i = 0$.

Si x et y sont tous deux non nuls, nous dirons que x et y sont *asymptotiques* dès lors que $(x/y \approx 1)$. Nous employons la notation $(x \sim y)$ pour exprimer cette propriété.

Théorème 5.2. *Si x et y sont deux réels appréciables $(0 \ll |x| \ll \infty$ et $0 \ll |y| \ll \infty)$ alors $(x \approx y)$ si et seulement si $(x \sim y)$.*

Démonstration. Soient x et y deux nombres réels appréciables, et supposons que $x \sim y$. Cela implique qu'il existe un réel infinitésimal α, tel que $(x/y = 1 + \alpha)$, donc $(x = y + \alpha y)$. Or, d'après la proposition 8, αy est infinitésimal. Ainsi, $(x \approx y)$. Inversement, supposons que $(x \approx y)$. Il existe donc un réel infinitésimal α, tel que $(x = y + \alpha)$, soit $(x/y = 1 + \alpha/y$. Or, les propositions 5 et 8 nous permettent d'affirmer que α/y est infinitésimal. Ainsi, $(x \sim y)$. □

Théorème 5.3. *Pour tout entier naturel n (standard ou non), et pour toutes familles $(x_i)_{i\in\{1,\cdots,n\}}$ et $(y_i)_{i\in\{1,\cdots,n\}}$ de réels strictement positifs ($x_i > 0$ et $y_i > 0$), on a :*

$$(\forall i \in \{1, \cdots, n\}, (x_i \sim y_i)) \Longrightarrow \left(\sum_{i=1}^{n} x_i \sim \sum_{i=1}^{n} y_i\right).$$

Démonstration. Supposons $\forall i \in \{1, \cdots, n\}, (x_i \sim y_i)$. On observe que :

$$\forall x > 0, \forall y > 0, (x \sim y) \Longleftrightarrow (\forall \varepsilon \gg 0, (1 - \varepsilon)y < x < (1 + \varepsilon)y)\ ,$$

par conséquent :

$$\forall i \in \{1, \cdots, n\}, \forall \varepsilon \gg 0, (1 - \varepsilon)y_i < x_i < (1 + \varepsilon)y_i,$$

et donc

$$(1 - \varepsilon) \sum_{i=1}^{n} y_i < \sum_{i=1}^{n} x_i < (1 + \varepsilon) \sum_{i=1}^{n} y_i,$$

soit

$$\sum_{i=1}^{n} x_i \sim \sum_{i=1}^{n} y_i.$$

□

Le théorème 5.3 ne requiert pas que n soit limité. Ensemble, les théorèmes 5.2 et 5.3 expliquent pourquoi les calculs numériques des intégrales fonctionnent : on ne commet qu'une erreur absolue infinitésimale en effectuant la somme d'un nombre illimité de termes infinitésimaux, si le résultat de la somme est limité, et si sur chaque terme de la somme on ne commet qu'une erreur relative infinitésimale.

J'ai souligné que les nouvelles règles de notre théorie ne nous permettent pas de construire des ensembles définis par des propriétés externes. Dans de nombreuses situations nous pouvons prouver la non-existence de tels ensembles :

Théorème 5.4. *Il n'existe pas d'ensemble A_1, A_2, A_3, A_4, A_5 définis ainsi :*
- *($n \in A_1$) si et seulement si ($n \in N$ et n est standard) ;*
- *($n \in A_2$) si et seulement si ($n \in N$ et n est non standard) ;*
- *($x \in A_3$) si et seulement si ($x \in R$ et x est limité) ;*
- *($x \in A_4$) si et seulement si ($x \in R$ et x est illimité) ;*
- *($x \in A_5$) si et seulement si ($x \in R$ et x est infinitésimal).*

Démonstration. Si A_1 existait, cela contredirait le théorème d'induction ;[2] si A_2 existait, on pourrait construire A_1 en posant $A_1 = N \setminus A_2$; si A_3 existait, on pourrait construire A_2 en posant $A_2 = N \setminus A_3$; si A_4 existait, on pourrait construire A_3 en posant $A_3 = R \setminus A_4$; si A_5 existait, on pourrait construire A_4 en posant $A_4 = \left\{x \in R^* : \frac{1}{x} \in A_5\right\}$. □

Tous ces résultats, qui semblent un peu restrictifs, s'avèrent en fait très utiles. Par exemple, soit une proposition interne P, telle que $P(x)$ soit vraie pour tout réel infinitésimal. Alors, nous sommes sûrs qu'il existe un réel x non infinitésimal tel que $P(x)$ soit vraie, sinon on pourrait poser $A_5 = \{x : P(x)\}$. Nous appellerons le principe sur lequel repose un tel raisonnement le *principe de Cauchy*.[3]

Soit une suite $x_1, x_2, \ldots,$ de nombres réels telle que, pour tout indice i standard, x_i soit infinitésimal ($x_i \approx 0$). On désire démontrer l'existence d'un entier illimité n tel que x_n est infinitésimal. Or, comme "être infinitésimal" n'est pas une formule interne, nous ne pouvons pas appliquer directement le principe de Cauchy, comme nous venons de le faire plus haut, pour prouver l'existence de cet entier. Ce résultat est néanmoins vrai, et il est possible de le démontrer en remplaçant la propriété "$x_n \approx 0$" par une propriété interne plus faible, comme nous allons le voir.

Théorème 5.5 (lemme de Robinson). *Soit $x_1, x_2, \ldots,$ une suite réelle. Si $x_i \approx 0$ pour tout indice i limité, alors il existe un indice illimité ν tel que $x_n \approx 0$ pour tout entier n inférieur ou égal à ν :*

$$\left(\forall^{st} n, x_n \approx 0\right) \Longrightarrow (\exists \nu \approx \infty, \forall n \leq \nu, x_n \approx 0)\,.$$

Démonstration. Soit l'ensemble S :

$$S = \left\{m : \forall n \leq m, |x_n| \leq \frac{1}{n}\right\}.$$

L'ensemble S contient tous les indices standard, donc il contient au moins un indice illimité ν (principe de Cauchy). Soit $n \leq \nu$. Si n est standard, $x_n \approx 0$ par hypothèse. Sinon, $|x_n| \leq \frac{1}{n}$ car ν est dans S, et $x_n \approx 0$ dans ce cas également. Nous avons donc trouvé un entier ν illimité tel que x_n soit infinitésimal pour tout indice n inférieur ou égal à ν, ce qui prouve le lemme. □

[2] On applique le principe classique d'induction à la propriété $P(n) : n \in A_1$, ce qui implique $A_1 = N$, et qui est en contradiction avec notre axiome (3). (NdT)

[3] Sur le principe de Cauchy, voir également [5, page 54]. (NdT)

Nous pourrions être tenté de construire un contre-exemple, en définissant la suite x_n par :

$$x_n = \begin{cases} 0 \text{ si } n \text{ est standard;} \\ 1 \text{ si } n \approx \infty. \end{cases}$$

Mais cette construction est illicite. Une suite est une fonction, et une fonction est un ensemble (un graphe), que nous devons pouvoir définir sans recours à des propriétés externes. Ceci qui n'est pas le cas dans notre tentative de construction du contre-exemple.

Chapitre 6

Analogues externes de notions internes

Soit T un sous-ensemble fini de R. Dans tout le livre, nous adopterons les conventions suivantes :

- Le plus petit élément de T sera noté a, et le plus grand sera noté b ;
- On désignera par T' l'ensemble $T \setminus \{b\}$;
- Pour tout élément t de T', on désignera par $(t + dt)$ son successeur dans T ;
- Pour toute fonction $\xi : T \to R$, lorsque nous écrirons $\xi(t)$, il sera implicite que t appartiendra à T, et on notera $d\xi(t) = \xi(t + dt) - \xi(t)$;
- Si $0 \ll b - a \ll \infty$, et si chaque dt est infinitésimal, alors on dira que T est un *quasi-intervalle*.

La plupart des résultats de l'Analyse sont creux dès lors qu'on les applique à des fonctions dont le domaine de définition est un ensemble fini. Aussi, il n'y aura pas de risque de confusion quand nous utiliserons une terminologie classique au sein d'un discours portant sur des notions externes.[1] Néanmoins, il s'avère parfois avantageux d'adjoindre un qualificatif tel que "quasi", ou "quasiment", à un terme usuel lorsque celui-ci est utilisé dans une analogie externe élémentaire.

Posons $T = \{1, \cdots, \nu\}$, où ν est un entier naturel illimité. L'ensemble T est un analogue externe élémentaire de $N^* = N \setminus \{0\}$. Par analogie avec la notion de convergence des suites réelles, nous dirons qu'une séquence (suite finie) de réels $(x_i)_{i \in \{1, \cdots, \nu\}}$ est *(quasiment) convergente* s'il existe un réel x tel que :

$$\forall n \leq \nu, (n \approx \infty) \Longrightarrow (x_n \approx x) .$$

[1] Où justement on traitera d'ensembles finis. (NdT)

On dit également dans ce cas que la suite converge (quasiment) vers x. On voit de plus que $(x_i)_{i\in\{1,\cdots,\nu\}}$ converge quasiment vers y si et seulement si y est équivalent à x ($y \approx x$).

Sur la figure 6.1, lorsque n est illimité, on ne peut pas discerner à "l'œil nu" x et x_n. Il n'est peut-être pas évident, lorsque l'on regarde cette figure, que la propriété de quasi-convergence corresponde bien à la propriété intuitive de "s'approcher de plus en plus de x". Voyons ce qui peut être dit des valeurs de x_n lorsque n est limité. Soit un réel $\varepsilon \gg 0$, et soit l'entier n_ε défini comme étant le plus petit indice tel que :

$$\forall n \leq \nu, (n \geq n_\varepsilon) \Longrightarrow (|x - x_n| \leq \varepsilon)\,.$$

Alors, n_ε est un entier limité, parce que sinon, on aurait $(n_\varepsilon - 1) \approx \infty$, et donc $|x - x_{n_\varepsilon - 1}| \leq \varepsilon$, ce qui serait contradictoire. La figure 6.2 illustre ce que nous venons de dire.

Il existe plusieurs manières de définir un analogue externe élémentaire pour une notion interne donnée. Soit T un sous-ensemble de R, et soit une fonction $\xi : T \to R$. Nous dirons que ξ *admet k ε-fluctuations* lorsqu'il existe dans T une séquence $t_0 < \cdots < t_k$ de réels telle que :

$$|\xi(t_1) - \xi(t_0)| \geq \varepsilon, |\xi(t_2) - \xi(t_1)| \geq \varepsilon, \cdots, |\xi(t_k) - \xi(t_{k-1})| \geq \varepsilon.$$

Dans ce cas, nous dirons également que les entiers $t_0 < \cdots < t_k$ sont des indices de k ε-fluctuations. Ce sont là des définitions internes. Avec ces définitions, on peut affirmer qu'une suite infinie de réels est convergente si et seulement si, quelque soit le réel $\varepsilon > 0$, il existe un entier k tel que n'existent pas $(k+1)$ ε-fluctuations. Cela nous suggère la définition externe suivante : nous dirons qu'une suite finie ou infinie de réels, ou une fonction réelle de T, est à *fluctuation limitée* si, pour tout réel $\varepsilon \gg 0$, cette suite (ou cette fonction) n'admet pas k ε-fluctuations dés lors que k est un entier illimité. La propriété d'être à fluctuation limitée est, à l'instar de la quasi-convergence, un analogue externe de la notion de convergence. Cependant, elle constitue une propriété plus faible : choisissons en effet un entier illimité $i \leq \nu$, et considérons la suite $(x_i)_{i\in\{1,\cdots,\nu\}}$ définie ainsi :

$$\begin{cases} \forall k \in \{1, \cdots, i\}, x_k = 0; \\ \forall k \in \{i+1, \cdots, \nu\}, x_k = 1. \end{cases}$$

Cette suite est à fluctuation limitée, mais elle n'est pas quasiment convergente. En revanche, il est facile de se convaincre (voir la figure 6.2) que lorsqu'une suite finie est quasiment convergente, elle est à fluctuation limitée.

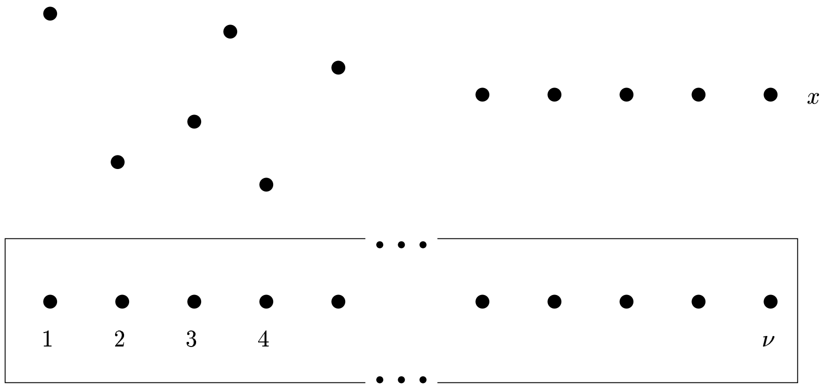

FIG. 6.1 – la quasi-convergence

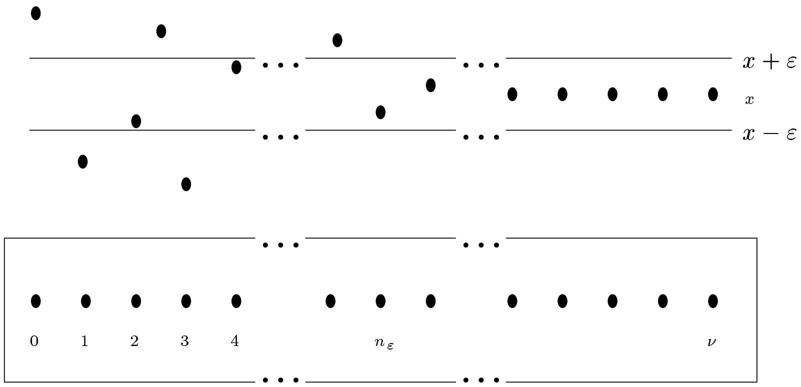

FIG. 6.2 – encore la quasi-convergence

Les notions de quasi-convergence et d' "être à fluctuation limitée" prendront toute leur importance lors de notre étude des fluctuations des processus stochastiques. La quasi-convergence exprime une propriété ordinale de la convergence (à partir d'un certain point, il n'existe plus de ε-fluctuations), tandis que le fait d'être à fluctuation limitée exhibe une propriété cardinale de la convergence (il n'existe qu'un nombre limité de ε-fluctuations).

Soit $A(n)$ une propriété interne ou externe. Le principe externe du plus petit élément stipule que s'il existe un entier naturel standard n, tel que $A(n)$, alors il existe un plus petit entier naturel m, nécessairement standard, tel que $A(m)$. Pour prouver cela on peut avoir recours au principe d'induction externe, de la même manière que l'on peut utiliser l'axiome d'induction classique pour démontrer le principe habituel du plus petit élément.

***Théorème 6.1.** *Soit une séquence de réels $x = (x_i)_{i\in\{1,\cdots,\nu\}}$ avec ν illimité. Si cette suite est à fluctuation limitée, alors il existe un entier illimité $\mu \le \nu$ tel que la séquence $(x_i)_{i\in\{1,\cdots,\mu\}}$ soit convergente.*

Démonstration. Pour tout naturel non nul standard j, l'ensemble des entiers naturels k tels que x n'admette pas k $\left(\frac{1}{j}\right)$-fluctuations contient tous les entiers illimités. Le principe de Cauchy nous permet d'affirmer que cet ensemble contient au moins un élément standard k. En vertu du principe externe du plus petit élément, il existe donc un plus petit entier naturel l inférieur ou égal à k, tel que la suite x n'admette pas $(l+1)$ $\left(\frac{1}{j}\right)$-fluctuations *d'indices illimités.* Si $l = 0$, posons $\mu_j = \nu$. Sinon, soient $n_0, \cdots, n_l$ $(l+1)$ indices illimités de l $\left(\frac{1}{j}\right)$-fluctuations. Posons alors $\mu_j = n_0$. A chaque entier naturel non nul standard j, on associe ainsi un entier μ_j. Le principe d'extension nous permet d'étendre cette association en une suite infinie $j \mapsto \mu_j$ telle que lorsque j est standard, les propriétés de μ_j par rapport à x soient conservées. Transformons maintenant la suite (μ_j) en une suite décroissante $(\tilde{\mu}_j)$ en posant $\tilde{\mu}_j = \inf_{i\le j} \mu_i$. Pour toute valeur standard de j la valeur de $\tilde{\mu}_j$ est illimitée, et en appliquant le lemme de Robinson à la suite $(1/\tilde{\mu}_j)$, on voit qu'il existe un entier k illimité tel que $\tilde{\mu}_k$ soit illimité. Notons $\mu = \tilde{\mu}_k$. Nécessairement, $\forall^{st} j > 0, \mu \le \mu_j$, ce qui implique la propriété suivante :

$$\forall n \le \mu, (n \approx \infty) \Rightarrow \left(\forall^{st} j > 0, |x_n - x_\mu| \le \left|x_n - x_{\mu_j}\right| + \left|x_{\mu_j} - x_\mu\right| \le \frac{2}{j}\right).$$

Autrement dit :

$$\forall n \le \mu, (n \approx \infty) \Longrightarrow (x_n \approx x_\mu),$$

ce qui signifie que la suite $(x_i)_{i\in\{1,\cdots,\mu\}}$ est convergente. □

Nous dirons que $\sum_{i=1}^{\nu} x_i$ est *(quasiment) convergente* (respectivement *à fluctuation limitée*) lorsque la séquence des sommes partielles

$$\left(y_n = \sum_{i=1}^{n} x_i\right)_{n\in\{1,\cdots,\nu\}}$$

est quasiment convergente (respectivement à fluctuation limitée). Nous sommes conscients de procéder ici à un abus d'écriture, dans la mesure où, en toute rigueur, $\sum_{i=1}^{\nu} x_i$ désigne un nombre. On voit que lorsque $\sum_{i=1}^{\nu} x_i$ converge, elle converge vers sa somme $s = \sum_{i=1}^{\nu} x_i$ (et vers toute valeur y équivalente à s). On voit également que $\sum_{i=1}^{\nu} x_i$ converge si et seulement si chaque reste $r_n = \sum_{i=n}^{\nu} x_i$ est infinitésimal lorsque $n \leq \nu$ est illimité.

Les implications suivantes sont vraies, mais leurs inverses sont fausses en général :

$$\begin{array}{ccccc}
 & & \begin{pmatrix} \sum_{i=1}^{\nu} |x_i| \\ \text{converge} \end{pmatrix} & \Rightarrow & \begin{pmatrix} \sum_{i=1}^{\nu} x_i \\ \text{converge} \end{pmatrix} \\
 & & \Downarrow & & \Downarrow \\
\left(\sum_{i=1}^{\nu} |x_i| \ll \infty\right) & \Rightarrow & \begin{pmatrix} \sum_{i=1}^{\nu} |x_i| \\ \text{est à fluctuations} \\ \text{limitées} \end{pmatrix} & \Rightarrow & \begin{pmatrix} \sum_{i=1}^{\nu} x_i \\ \text{est à fluctuations} \\ \text{limitées} \end{pmatrix}
\end{array}$$

Si la séquence $(x_i)_{i\in\{1,\cdots,\nu\}}$ vérifie $\forall^{st} i, |x_i| \ll \infty$, alors elle vérifie également la propriété :

$$\sum_{i=1}^{\nu} |x_i| \text{ converge} \Longrightarrow \sum_{i=1}^{\nu} |x_i| \ll \infty.$$

En effet si $\sum_{i=1}^{\nu} |x_i|$ converge, l'ensemble $\left\{n > 0 : \sum_{i=n+1}^{\nu} |x_i| < 1\right\}$ contient au moins un entier standard k, ce qui nous fournit l'inégalité suivante, dont le terme de gauche est limité :

$$\sum_{i=1}^{\nu} |x_i| < 1 + k \max_{i\leq k} \left(|x_i|\right).$$

Si $\forall i \in \{1, \cdots, \nu\}, |x_i| \ll \infty$, l'équivalence suivante est facilement vérifiée :

$$\sum_{i=1}^{\nu} |x_i| \text{ est à fluctuation limitée } \Leftrightarrow \sum_{i=1}^{\nu} |x_i| \ll \infty.$$

En effet, construisons une séquence d'entier μ_n en posant $\mu_1 = 1$, et pour $n \geq 1$:

$$\mu_{n+1} = \begin{cases} \text{plus petit nombre supérieur à } \mu_n\text{, tel que } \sum_{\mu_n+1}^{\mu_{n+1}} |x_i| > 1. \\ \quad \text{Dans ce cas, } \sum_{\mu_n+1}^{\mu_{n+1}} |x_i| \text{ est limité, du fait que} \\ \quad \forall i \in \{1, \cdots, \nu\}, |x_i| \ll \infty; \\ \nu \text{ si } \sum_{\mu_n+1}^{\nu} |x_i| \leq 1. \end{cases}$$

Si la série est à fluctuation limitée, on ne peut construire qu'un nombre limité k d'entiers μ_n, et on a $\sum_{i=1}^{\nu} |x_i| < k(1 + \max |x_i|) \ll \infty$.

Soit une fonction $\xi : T \to R$. Pour tout élément t de T, nous dirons que ξ est *(quasiment) continue*[2] *en* t lorsque :

$$\forall u \in T, (t \approx u) \Longrightarrow (\xi(t) \approx \xi(u)),$$

et nous dirons que ξ est *(quasiment) continue sur* T si elle est quasiment continue en chaque point de T. Par exemple, si T est un quasi-intervalle tel que $a > 0$, la fonction $\xi : t \mapsto 1/t$ est continue en t si et seulement si $t \gg 0$, donc elle est continue sur T si et seulement si $a \gg 0$.

Soit une fonction ξ quasiment continue en t, et soit $\varepsilon \gg 0$. Définissons l'ensemble Δ_t ainsi :

$$\Delta_t = \{\delta > 0 : \forall s \in T, (|s - t| \leq \delta) \Rightarrow (|\xi(s) - \xi(t)| \leq \varepsilon)\}.$$

Cet ensemble contient tous les réels positifs infinitésimaux, il contient donc au moins un élément non infinitésimal η d'après le principe de Cauchy. Nous avons donc l'équivalence :

$$\begin{gathered} (\xi \text{ est quasiment continue en } t) \\ \Longleftrightarrow \\ (\forall \varepsilon \gg 0, \exists \eta \gg 0, \forall s \in T, (|s - t| \leq \eta) \Rightarrow (|\xi(s) - \xi(t)| \leq \varepsilon)). \end{gathered}$$

Supposons maintenant que ξ soit quasiment continue sur T. Pour chaque élément t de T, définissons le réel δ_t comme étant le plus grand élément de Δ_t dont l'inverse est un entier[3]. Alors $\delta_t \gg 0$, car Δ_t est un intervalle

[2] ou *S-continue*, voir [5, page 93]. (NdT)

[3] Cette construction est justifiée pour bien construire l'ensemble au sein duquel on va calculer $min(\delta_t)$. En effet, on ne peut pas définir simplement δ_t par la relation $\delta_t = \sup \Delta_t$, car on n'aurait pas forcément $\delta_t \in \Delta_t$, et on ne peut pas non plus dire simplement "choisissons un élément appréciable δ_t au sein de chaque Δ_t", car de la sorte on formerait illégalement l'ensemble $\{\delta_t\}$. (NdT)

qui contient des éléments appréciables strictement positifs. Soit η le réel strictement positif défini par :

$$\eta = \min_{t \in T} (\delta_t) .$$

Le réel η est appréciable car, T étant fini, $\eta \in \{\delta_t, t \in T\}$. Nous avons donc :

$$\begin{gathered}(\xi \text{ est quasiment continue sur } T) \\ \Longleftrightarrow \\ \big(\forall \varepsilon \gg 0, \exists \eta \gg 0, \forall (s,t) \in T^2, (|s-t| \leq \eta) \Rightarrow (|\xi(s) - \xi(t)| \leq \varepsilon)\big).\end{gathered}$$

On voit par là que la quasi-continuité en t est un analogue externe élémentaire de la continuité classique en t, et que la quasi-continuité sur T est un analogue externe élémentaire de la continuité uniforme classique sur T.

Présentons maintenant quelques illustrations simples et utiles des concepts précédents. Le principe d'induction externe nous permet d'affirmer que le réel e^n est limité pour tout entier n standard. Il s'ensuit que $\forall t \ll \infty$, $e^t \ll \infty$. Le théorème de la moyenne exprime la propriété que pour toute valeur de h, il existe un réel t' compris entre t et $t+h$, tel que $e^{t+h} = e^t + he^{t'}$. Ainsi, si $t \ll \infty$ et si $h \approx 0$, il vient que $e^{t+h} \approx e^t$, donc que la fonction exponentielle est quasiment continue sur T dès lors que $b \ll \infty$. De la même manière, on a :

$$\log(t+h) = \log(t) + \frac{h}{t'},$$

donc dès que $h \approx 0$ et que $t \gg 0$, $\log(t+h) \approx \log(t)$. La fonction logarithme est donc quasiment continue sur T si $a \gg 0$.

Posons $|t| \ll \infty$ et $n \approx \infty$. Il vient alors :

$$\log\left(1 + \frac{t}{n}\right)^n = n \log\left(1 + \frac{t}{n}\right) = n\left(\frac{t}{n} - \frac{1}{2}\left(\frac{t'}{n}\right)^2\right) \approx t,$$

avec t' compris entre 0 et t. Donc, sous ces conditions, $\left(1 + \frac{t}{n}\right)^n \approx e^t$. La séquence $\left(x_n = \left(1 + \frac{t}{n}\right)^n\right)_{n \in \{1, \cdots, \nu\}}$, où $|t| \ll \infty$ et $\nu \approx \infty$, converge donc vers e^t.

La série de Taylor avec reste de Lagrange nous donne :

$$e^t = \sum_{n=0}^{\nu} \frac{t^n}{n!} + e^{t'} \frac{t^{\nu+1}}{(\nu+1)!},$$

où l'on voit facilement que le reste est infinitésimal lorsque t est limité, et que ν est illimité. Ainsi, dans ces conditions, la série $\sum_{n=0}^{\nu} \frac{t^n}{n!}$ converge-t-elle vers e^t.

Pour une fonction $\xi : T \to R$, nous définissons la *variation*, ou *variation totale*, comme étant la valeur $\sum_{t\in T'} |d\xi(t)|$. Lorsque cette valeur est limitée, nous dirons que ξ est à *variation limitée*. Nous avons là un analogue externe élémentaire de la notion de fonction à variation bornée. Nous dirons de plus que ξ est *absolument continue* si pour tout sous-ensemble S de T', on a $\left(\sum_{t\in S} dt \approx 0\right) \Rightarrow \left(\sum_{t\in S} |d\xi(t)| \approx 0\right)$. Il est évident qu'une fonction absolument continue est continue.

Théorème 6.2. *Soit T un quasi-intervalle. Si ξ est absolument continue, alors ξ est à variation limitée.*

Démonstration. Supposons que ξ soit absolument continue, et considérons l'ensemble de tous les réels strictement positifs δ, tels que :

$$\forall S \subseteq T', \left(\sum_{t\in S} dt \leq \delta\right) \Longrightarrow \left(\sum_{t\in S} |d\xi(t)| \leq 1\right).$$

Comme cet ensemble contient tous les réels strictement positifs infinitésimaux, il contient au moins un réel appréciable η. On peut alors couper le quasi-intervalle T en $n = \left[\frac{b-a}{\eta}\right] + 1$ parties S adjacentes (le plus grand élément d'une partie S correspond au plus petit élément de la partie S suivante), telles que $\sum_{t\in S} dt \leq \eta$. Alors, $\sum_{t\in T'} |d\xi(t)| \leq n \ll \infty$. La fonction ξ est donc à variation limitée. □

Chapitre 7

Propriétés vraies presque partout

Les espaces probabilisés finis interviennent habituellement dans les travaux portant sur des problèmes combinatoires. Néanmoins, nous souhaitons en faire le support des discussions touchant aux théorèmes limites classiques de la théorie des probabilités, ainsi qu'aux notions clés de la théorie moderne des processus stochastiques. Le principal concept externe qui nous permettra d'aboutir à ce résultat est le suivant : considérons un espace probabilisé fini $\langle \Omega, pr \rangle$, ainsi qu'une proposition interne ou externe $A(\omega)$. Nous dirons que la propriété A est vraie *presque partout* (pp), ou *presque sûrement* (ps) lorsque, pour tout nombre réel positif appréciable ε, il existe un événement N tel que $Pr(N) \leq \varepsilon$, et tel que la propriété A est vraie pour tout élément du complémentaire de N :

$$(ps : A) \overset{def}{\Longleftrightarrow} (\forall \varepsilon \gg 0, \exists N \subseteq \Omega, (Pr\,(N) \leq \varepsilon) \text{ et } (\forall \omega \in N^c, A\,(\omega)))\,.$$

Lorsque A est une proposition interne, alors nous pouvons construire l'ensemble $\{A\}$ des éléments pour lesquels A est vraie : $\{A\} = \{\omega \in \Omega : A(\omega)\}$. Dans ce cas, A est vraie presque partout si et seulement si la probabilité de $\{A\}$ est équivalente à 1 :

$$(ps : A) \Longleftrightarrow (Pr\,(\{A\}) \approx 1)\,.$$

Cependant, certaines des propositions les plus intéressantes dont nous allons discuter seront externes, aussi devrons-nous avoir recours à la première formulation plutôt qu'à la seconde, afin d'éviter de procéder à des formations illégales d'ensembles.

Mais, que la proposition A soit externe ou interne, l'idée intuitive que nous associons à la propriété d'être "vrai presque partout" est celle de quasi-certitude : quel que soit le seuil petit mais appréciable ε (par exemple, $\varepsilon = 10^{-100}$), il existe un événement N de probabilité inférieure à ε, tel que A soit toujours vraie, à l'exception éventuelle de certains points, qui alors appartiendront nécessairement à N.

Théorème 7.1. *Soit x une variable aléatoire. Alors les propositions suivantes sont équivalentes :*

1. *$x \approx 0$ presque sûrement (ce que l'on note également $x \underset{ps}{\approx} 0$) ;*
2. *$\forall \lambda \gg 0, Pr\{|x| \geq \lambda\} \approx 0$;*
3. *$\exists \lambda \approx 0, Pr\{|x| \geq \lambda\} \approx 0$.*

Démonstration. Supposons 1. Soit $\lambda \gg 0$, et soit ε un réel appréciable. Il existe un évènement N_ε de probabilité inférieure à ε tel que $\forall \omega \in N_\varepsilon^c, x(\omega) \approx 0$. Alors, $\{\omega \in \Omega : |x(\omega)| \geq \lambda\} \subseteq N_\varepsilon$, ce qui implique $Pr\{|x| \geq \lambda\} \leq \varepsilon$. Ceci, étant vrai pour tout ε appréciable indépendamment de λ, prouve que 1⇒2. Supposons maintenant 2. L'ensemble des réels positifs λ tels que

$$Pr\{|x| \geq \lambda\} \leq \lambda$$

contient tous les réels sensiblement positifs (tels que $\lambda \gg 0$). En vertu du principe de Cauchy, il contient également un réel positif infinitésimal μ, qui vérifie donc $Pr\{|x| \geq \mu\} \approx 0$. Ainsi 2⇒3. Enfin, l'implication 3⇒1 est évidente : il suffit de prendre systématiquement $N_\varepsilon = \{\omega \in \Omega : |x(\omega)| \geq \mu\}$. □

Aussi longtemps, que nous n'avons à considérer qu'une seule variable aléatoire x, nous pouvons sereinement admettre, d'un point de vue pratique, que x est uniformément nulle dès lors que x est infinitésimale presque partout. La probabilité dans ce cas d'être sensible à un écart de x par rapport à la valeur nulle est en effet inférieure à 10^{-100}.[1] Cependant, les choses changent du tout au tout dès que l'on a à prendre en compte simultanément un nombre illimité de variables aléatoires $x_1, \cdots, x_\nu$, chacune étant presque sûrement infinitésimale. Par exemple, divisons la durée d'une journée en ν intervalles d'égales largeurs, où ν est un entier illimité. Imaginons une machine dont la probabilité de panne, lorsque l'on est dans l'un de ces intervalles, vaut P_n = (probabilité de panne dans l'intervalle n) $= c/\nu$, avec $0 \ll c \ll \infty$. Si nous notons x_n la variable aléatoire qui est la fonction indicatrice de l'événement "apparition d'une panne dans l'intervalle n", alors nous avons

[1] Ou à tout nombre que l'on pourrait écrire. (NdT)

$\forall n \in \{1, \cdots, \nu\}, x_n \underset{ps}{\approx} 0$, c'est-à-dire que pour chaque indice n, la variable x_n est presque sûrement infinitésimale. En fait, nous avons même plus que cela : pour chaque indice n, la variable x_n est presque sûrement nulle. Maintenant, si l'apparition d'une panne dans la journée engendre une catastrophe, la variable aléatoire que nous préférerions considérer est $x = \max_{i \in \{1, \cdots, \nu\}} (x_i)$, qui est la variable indicatrice d'un désastre provoqué durant la journée par une panne. Sous l'hypothèse d'indépendance des évènements de panne, la probabilité qu'aucune catastrophe ne survienne durant une journée a pour valeur :

$$Pr\left\{\max_{i \in \{1, \cdots, \nu\}} (x_i) = 0\right\} = \left(1 - \frac{c}{\nu}\right)^{\nu} \approx e^{-c} \ll 1.$$

Soit une séquence illimitée $x_1, \cdots, x_\nu$ de variables aléatoires. Nous dirons que $(x_i)_{i \in \{1, \cdots, \nu\}}$ *converge (quasiment et) en probabilité* - ou *(quasiment et) stochastiquement* - vers la variable aléatoire x si, pour tout indice illimité n, on a presque sûrement $x \approx x_n$:

$$\left(x_i \underset{P}{\longrightarrow} x\right) \overset{def}{\Longleftrightarrow} \left(\forall i \in \{1, \cdots, \nu\}, (i \approx \infty) \Rightarrow \left(x_i \underset{ps}{\approx} x\right)\right).$$

Comme notre précédent exemple le montre, la convergence en probabilité n'est pas une propriété très forte. Plus intéressante est la question de savoir si la séquence $(x_i)_{i \in \{1, \cdots, \nu\}}$ converge *(quasiment et) presque sûrement* vers x :

$$\left(x_i \underset{ps}{\longrightarrow} x\right) \overset{def}{\Longleftrightarrow} (ps : \forall i \in \{1, \cdots, \nu\}, (i \approx \infty) \Rightarrow (x_i \approx x)).$$

Pour la convergence en probabilité, l'ensemble N, où devront être confinés les points exceptionnels, peut varier selon l'indice de la variable aléatoire. Pour la convergence presque sûre, en revanche, ceci n'est plus permis.

Théorème 7.2. *Une séquence de variables aléatoires $(x_i)_{i \in \{1, \cdots, \nu\}}$ converge presque sûrement vers 0 si et seulement si :*

$$\forall \lambda \gg 0, \forall n \leq \nu, (n \approx \infty) \Rightarrow \left(Pr\left\{\max_{i \in \{n, \cdots, \nu\}} |x_i| \geq \lambda\right\} \approx 0\right). \tag{7.1}$$

Démonstration. Posons

$$M(n, \lambda) = \left\{\omega \in \Omega : \max_{i \in \{n, \cdots, \nu\}} |x_i(\omega)| \geq \lambda\right\}.$$

Supposons d'abord que la séquence $(x_i)_{i \in \{1, \cdots, \nu\}}$ converge presque sûrement vers 0. Soient $\lambda \gg 0$ et $\varepsilon \gg 0$. Alors, il existe un événement N tel que

$Pr\,(N) \leq \varepsilon$, et tel que, à l'extérieur de N, la séquence converge. Alors, si n est illimité, $M\,(n, \lambda) \subseteq N$, et $Pr\,(M\,(n, \lambda)) \leq \varepsilon$. Cela étant vrai pour tout ε positif appréciable, on en conclut :

$$\forall \lambda \gg 0, \forall n \approx \infty, Pr\,(M\,(n, \lambda)) \approx 0.$$

Inversement, supposons que $\forall \lambda \gg 0, \forall n \approx \infty, Pr\,(M\,(n, \lambda)) \approx 0$. Soit $\varepsilon \gg 0$, et pour tout entier naturel non nul j, définissons l'entier n_j^ε comme étant le plus petit indice vérifiant :

$$Pr\left(M\left(n_j^\varepsilon, \frac{1}{j}\right)\right) \leq \frac{\varepsilon}{2^j}.$$

Soit N_ε l'ensemble défini ainsi :

$$N_\varepsilon = \bigcup_{j=1}^{\infty} M\left(n_j^\varepsilon, \frac{1}{j}\right).$$

De manière évidente, $Pr\,(N_\varepsilon) \leq \varepsilon$. On peut remarquer, bien que cela n'ait pas d'importance dans la suite de la démonstration, que dès que j devient suffisamment grand, $M\left(n_j^\varepsilon, \frac{1}{j}\right)$ est vide. Cela est simplement dû au fait que l'ensemble Ω est fini. Remarquons également, et ceci sera utile dans la suite, que lorsque j est limité, il en va de même pour n_j^ε. En effet, on aurait sinon $n_j^\varepsilon - 1 \approx \infty$ et comme $1/j \gg 0$, $Pr\left(M\left(n_j^\varepsilon - 1, \frac{1}{j}\right)\right)$ serait infinitésimal, par hypothèse, donc inférieur à $\varepsilon/2^j$, ce qui serait en contradiction avec la définition de n_j^ε. Il vient donc :

$$\forall j \ll \infty, \forall \omega \in N_\varepsilon^c, \left(\forall n \leq \nu, (n \approx \infty) \Rightarrow \left(|x_n\,(\omega)| \leq \frac{1}{j}\right)\right),$$

d'où l'on tire :

$$\forall \omega \in N_\varepsilon^c, (\forall n \leq \nu, (n \approx \infty) \Rightarrow (x_n\,(\omega) \approx 0)).$$

Comme $Pr\,(N_\varepsilon) \leq \varepsilon$ pour tout réel ε appréciable, cela démontre que la séquence $(x_i)_{i \in \{1, \cdots, \nu\}}$ converge presque sûrement vers 0. □

Le théorème 7.2 admet le corollaire suivant :

Corollaire. *Soit ξ un processus stochastique indexé par un sous-ensemble fini de R, T. Soit t un point de T. Alors, ξ est presque sûrement continu en t si et seulement si :*

$$\forall \lambda \gg 0, \forall h \approx 0, Pr\left\{\max_{|s-t| \leq h} (|\xi\,(s) - \xi\,(t)|) \geq \lambda\right\} \approx 0. \tag{7.2}$$

Démonstration. Soit $\nu \approx \infty$, construisons la séquence $(x_i)_{i\in\{1,\cdots,\nu\}}$ de variables aléatoires comme suit :

$$\forall n \in \{1, \cdots, \nu\}, x_n = \max_{|s-t|\leq\frac{1}{n}} (|\xi(s) - \xi(t)|).$$

Alors, ξ est continu en t si et seulement si $(x_i)_{i\in\{1,\cdots,\nu\}}$ converge vers 0, et l'expression 7.1 devient équivalente, dans le contexte présent, à l'expression 7.2. □

Théorème 7.3 (Borel-Cantelli, version ordinale). *Soit $(A_1, \cdots, A_\nu)$ une séquence d'événements ($\nu \approx \infty$), et définissons sur Ω une fonction k telle que $k(\omega)$ désigne le plus grand indice k vérifiant $\omega \in A_k$. Lorsque ω n'appartient à aucun sous-ensemble parmi $(A_1, \cdots, A_\nu)$, on conviendra que $k(\omega) = 0$. Plus formellement :*

$$\begin{cases} k : \Omega \longrightarrow \{0, \cdots, \nu\} \\ \forall \omega \in \Omega, \forall n \in \{1, \cdots, \nu\} \begin{cases} (n > k(\omega)) \Rightarrow (\omega \notin A_n) \\ (n = k(\omega)) \Rightarrow (\omega \in A_n) \end{cases} \end{cases}$$

Alors :

1. *si $\sum_{n=1}^{\nu} Pr(A_n)$ converge, alors k est limitée presque partout ;*
2. *lorsque les événements $A_1, \cdots, A_\nu$ sont indépendants, alors $\sum_{n=1}^{\nu} Pr(A_n)$ converge si et seulement si k est limitée presque partout.*

Démonstration. Commençons par la première assertion. Supposons donc que $\sum_{n=1}^{\nu} Pr(A_n)$ converge, et soit $\varepsilon \gg 0$. Définissons j comme étant le plus petit indice tel que $Pr \bigcup_{n=j}^{\nu} A_n \leq \varepsilon$. Alors, j est nécessairement limité, car sinon $j-1$ serait illimité, et on aurait $Pr \bigcup_{n=j-1}^{\nu} A_n \leq \sum_{n=j-1}^{\nu} Pr(A_n) \approx 0$, en contradiction avec la définition de j. Alors, pour tout élément ω de $\left(\bigcup_{n=j}^{\nu} A_n\right)^c$ la valeur $k(\omega)$ est limitée (par construction, elle est strictement inférieure à j). Ceci prouve que $k(\omega)$ est presque partout limitée.

En ce qui concerne la deuxième assertion, supposons l'indépendance des événements $A_1, \cdots, A_\nu$, et que k soit limitée presque partout. La convexité de la fonction $t \mapsto e^{-\lambda t}$, quel que soit λ, nous permet d'écrire

$$\forall \lambda \in R, (1 - \lambda) \leq e^{-\lambda}.$$

Le graphe d'une fonction convexe est en effet partout situé au-dessus de chacune de ses tangentes. On a ainsi, pour tout i :

$$Pr \bigcap_{n=i}^{\nu} A_n^c = \prod_{n=i}^{\nu} (1 - Pr(A_n)) \leq e^{-\sum_{n=i}^{\nu} Pr(A_n)}.$$

Par hypothèse (k est limitée presque partout), lorsque i est illimité,

$$Pr \bigcap_{n=i}^{\nu} A_n^c \approx 1,$$

donc $\sum_{n=i}^{\nu} Pr\left(A_n\right) \approx 0$, ce qui prouve la convergence de la série. Ce résultat, joint à la première assertion, nous permet de conclure. □

Théorème 7.4 (Borel-Cantelli, version cardinale). *Soit* $(A_1, ..., A_\nu)$ *une séquence d'événements* ($\nu \approx \infty$)*, et définissons sur* Ω *une fonction* K *telle que* $K(\omega)$ *désigne le nombre d'indices* n *tels que* $\omega \in A_n$*. Alors :*

1. *si* $\sum_{n=1}^{\nu} Pr\left(A_n\right) \ll \infty$*, alors* K *est limitée presque partout ;*
2. *si les événements* $A_1, \cdots, A_\nu$ *sont indépendants,* $\sum_{n=1}^{\nu} Pr\left(A_n\right) \ll \infty$ *si et seulement si* K *est limitée presque partout. En fait, nous avons* $\sum_{n=1}^{\nu} Pr\left(A_n\right) \approx \infty$ *si et seulement si* K *est illimité presque partout.*

Démonstration. Remarquons avant tout que $EK = \sum_{n=1}^{\nu} Pr\left(A_n\right)$. Pour la première assertion, supposons que $\sum_{n=1}^{\nu} Pr\left(A_n\right) \ll \infty$. L'inégalité de Bienaymé-Tchebychev, appliquée avec l'exposant $p = 1$, nous donne

$$Pr\left\{K \geq \lambda\right\} \leq \frac{EK}{\lambda}.$$

Mais comme par hypothèse $EK \ll \infty$, il est possible, pour tout réel $\varepsilon \gg 0$, de trouver un réel $\lambda \gg 0$ tel que $EK/\lambda < \varepsilon$. Ainsi, $K < \lambda \ll \infty$, sauf pour un événement de probabilité inférieure ou égale à ε. Donc K est presque partout limitée.

Pour prouver la deuxième assertion, il suffit de montrer que lorsque les événements $A_1, \cdots, A_\nu$ sont indépendants, le fait que EK soit illimité entraîne que K soit illimité presque partout. Si nous appliquons l'inégalité de Bienaymé-Tchebychev avec l'exposant $p = 2$, on voit que

$$Pr\left\{|K - EK| \geq \lambda\right\} \leq \frac{var\left(K\right)}{\lambda^2}.$$

Or, du fait de l'hypothèse d'indépendance des événements, nous avons :

$$var\left(K\right) = \sum_{n=1}^{\nu} var\left(\chi_n\right) \leq \sum_{n=1}^{\nu} E\chi_n^2 = EK,$$

donc

$$Pr\left\{|K - EK| \geq \lambda\right\} \leq \frac{EK}{\lambda^2}.$$

Si nous choisissons $\lambda = \frac{EK}{2}$, et que nous supposons que EK est illimité, nous obtenons la relation $Pr\left\{|K - EK| \geq \frac{EK}{2}\right\} \approx 0$, ce qui veut dire qu'à l'exception d'un événement de probabilité infinitésimale, nous aurons $|K - EK| < EK/2$. Ainsi, presque partout, $K \approx \infty$. □

Lorsque les événements $A_1, \cdots, A_\nu$ sont indépendants, alors soit K est limitée presque partout, soit K est illimitée presque partout. Cette propriété n'est pas vraie pour la fonction k de la version ordinale du théorème. Par exemple, posons $A_n = \emptyset$ pour tout indice n, sauf pour un indice i illimité, pour lequel nous construisons un événement A_i de probabilité $1/2$. Alors, ces événements sont indépendants, mais $Pr\{k = 0\} = 1/2$, et $Pr\{k = i\} = 1/2$. Dans cette construction, $\sum Pr(A_n)$ est limitée, mais pas convergente.

Lorsqu'une série de probabilités converge, alors elle converge vers une valeur limitée, du fait que chacun de ses termes est limité. De plus, lorsque k est limité, alors K est limité également. Ceci montre que l'hypothèse et la conclusion de l'assertion 1) sont toutes deux plus fortes dans la version ordinale que dans la version cardinale du théorème de Borel-Cantelli. Le langage des mathématiques conventionnelles ne permet pas de faire la distinction entre une version ordinale et une version cardinale : si nous nous donnons une suite infinie $(A_i)_{i \in N}$ d'événements, alors affirmer qu'il ne survient qu'un nombre fini d'événements est équivalent à affirmer qu'à partir d'un certain indice, plus aucun événement ne survient. Cependant, la distinction est intéressante. L'exposé précédent montre bien que pour estimer K, il suffit d'avoir un majorant de $\sum Pr(A_n)$, alors que pour estimer k, il nous faut procéder en faisant intervenir la petitesse éventuelle des restes de la série $\sum Pr(A_n)$. Dans le cas de notre exemple précédent, qui mettait en scène l'occurrence de pannes catastrophiques, on ne peut pas dire que k soit presque partout limitée (une panne est possible n'importe quand). En revanche, on peut au moins se consoler en constatant que, presque sûrement, le nombre de catastrophes observées lors d'une journée est, quant à lui, limité.

Chapitre 8

Variables aléatoires L^1

Pour une v.a.r. x, et une constante réelle a, nous définissons la variable aléatoire *tronquée* $x^{(a)}$ de la manière suivante :

$$x^{(a)} \stackrel{def}{=} x\chi_{\{|x|\leq a\}}.$$

Ainsi, cette variable est définie par les propriétés :

$$\forall \omega \in \Omega, x^{(a)}(\omega) = \begin{cases} x(\omega) & \text{si } (|x(\omega)| \leq a), \\ 0 & \text{si } (|x(\omega)| > a). \end{cases}$$

Par définition, nous dirons que x est une variable aléatoire L^1 lorsqu'elle vérifie :

$$\forall a > 0, (a \approx \infty) \Longrightarrow \left(E\left|x - x^{(a)}\right| \approx 0\right).$$

Du fait que $E\big|x - x^{(a)}\big| = \sum_{|\lambda|>a} \lambda\, pr_x(\lambda)$, il vient que x est L^1 si et seulement si la séquence $\left(\sum_{|\lambda|\leq n} \lambda\, pr_x(\lambda)\right)_{n\in\{1,\cdots,\nu\}}$, avec $\nu \geq \|x\|_\infty$, converge (ou, de manière plus concise, si et seulement si $\sum \lambda pr\,(\lambda)$ converge). Par conséquent, si x est L^1, alors $E\,|x| \ll \infty$. La proposition inverse n'est pas forcément vérifiée. Ainsi, supposons l'existence d'un point ω_0 tel que $pr\,(\omega_0) \approx 0$, et définissons la variable aléatoire x telle que

$$x(\omega) = \frac{\chi_{\{\omega_0\}}(\omega)}{pr(\omega_0)}.$$

Alors, on a bien $E\,|x| \ll \infty$, mais x n'est pas L^1. En effet, pour $a \approx \infty$, tel que $a < 1/pr\,(\omega_0)$, on a $E\left|x - x^{(a)}\right| = 1$, soit $E\left|x - x^{(a)}\right| \gg 0$.

Théorème 8.1 (Radon-Nikodym). *Une v.a.r. x est L^1 si et seulement si :*

$$\begin{cases} E\,|x| \ll \infty, \\ \textit{et} \\ \textit{pour tout événement } M \textit{ de probabilité infinitésimale } (Pr\,(M) \approx 0)\textit{, on a} \\ E\,(|x|\,\chi_M) \approx 0. \end{cases}$$

Démonstration. Soit x une v.a.r. L^1. Nous savons déjà que $E\,|x| \ll \infty$. Soit M un événement de probabilité infinitésimale, et un réel a illimité et tel que $a \times Pr\,(M) \approx 0$. On peut par exemple choisir $a = 1/\sqrt{Pr\,(M)}$. Alors nous avons :

$$\begin{aligned} E(|x|\chi_M) &\le E\Big(\Big|x^{(a)}\Big|\chi_M\Big) + E\Big(\Big|x - x^{(a)}\Big|\chi_M\Big) \\ &\le aPr(M) + E\Big(\Big|x - x^{(a)}\Big|\Big) \approx 0. \end{aligned}$$

Inversement, supposons que $E|x| \ll \infty$, et que pour tout événement M de probabilité infinitésimale, $E(|x|\chi_M) \approx 0$. Soit $a \approx \infty$ quelconque, et

$$M = \{\omega \in \Omega : |x(\omega)| > a\}.$$

Alors, l'inégalité de Bienaymé-Tchebychev, avec l'exposant $p = 1$, nous dit que $Pr(M) \le E(|x|)/a$, soit $Pr(M) \approx 0$. Donc en vertu de nos hypothèses $E(|x|\chi_M) \approx 0$, ce que l'on peut également écrire $E\big|x - x^{(a)}\big| \approx 0$. Ainsi x est L^1. □

De ce qui précède, nous pouvons déduire les assertions suivantes :
- si x et y sont L^1, la variable $x + y$ est L^1 ;
- si x est L^1, et si y est telle que $|y| \le |x|$, y est L^1 ;
- si x est L^1, et si $\|y\|_\infty \ll \infty$, xy est L^1.

Théorème 8.2 (Lebesgue). *Si x et y sont des variables aléatoires L^1 telles que $x \underset{ps}{\approx} y$, alors $Ex \approx Ey$.*

Démonstration. Soit $z = x - y$. Alors $z \underset{ps}{\approx} 0$, ce qui implique, d'après le théorème 7.1, qu'il existe un réel infinitésimal α tel que $Pr\{|z| \ge \alpha\} \approx 0$. Or $E|z| \le \alpha + E\big(|z|\chi_{\{|z| \ge \alpha\}}\big)$. Comme z est L^1, l'application du théorème 8.1 nous dit que $E\big(|z|\chi_{\{|z| \ge \alpha\}}\big) \approx 0$. Ainsi, $E|z| = E|x - y| \approx 0$ et $Ex \approx Ey$. □

Pour tout réel $p \in]1,\infty[$, nous dirons qu'une v.a.r. x est L^p si la v.a.r. $|x|^p$ est L^1. Nous dirons de plus que x est L^∞ si $\|x\|_\infty \ll \infty$. Si x est L^p et que y est L^q, où p et q sont conjugués $(1/p + 1/q = 1)$ alors l'inégalité suivante déjà mise en œuvre dans le premier chapitre :

$$|xy| \leq \frac{1}{p}|x|^p + \frac{1}{q}|y|^q$$

montre que xy est L^1. De plus, si x est L^p avec $p \gg 1$, alors x est L^1. En effet, soit a un réel illimité. Nous avons :

$$\sum_{|\lambda|>a} |\lambda|\, pr_x(\lambda) \leq \sum_{|\lambda|>a} \frac{|\lambda|^p}{a^{p-1}} pr_x(\lambda) \leq \frac{1}{a^{p-1}} E(|x|^p) \approx 0,$$

d'où le résultat.

Théorème 8.3. *Soit $\mathcal{A}$ une algèbre de variables aléatoires, et soit x une v.a.r. L^p, où $p \in [1,\infty]$. Alors $E_{\mathcal{A}}x$ est une variable L^p.*

Démonstration. Si $p = \infty$, le résultat est évident :

$$\|E_{\mathcal{A}}x\|_\infty = \max |E_{\mathcal{A}}x| \leq \max E_{\mathcal{A}}|x| \leq \max |x| = \|x\|_\infty \ll \infty.$$

Si $1 \leq p < \infty$, l'inégalité de Jensen relativisée donne $|E_{\mathcal{A}}x|^p \leq E_{\mathcal{A}}(|x|^p)$, aussi suffit-il de montrer le théorème pour $p = 1$.

Supposons donc que x soit L^1. Nous avons déjà : $EE_{\mathcal{A}}|x| = E|x| \ll \infty$. Soit maintenant un événement M de probabilité infinitésimale. Posons

$$a = \frac{1}{\sqrt{Pr(M)}}.$$

Nous avons $a \gg 0$, et :

$$\begin{aligned} E(|E_{\mathcal{A}}x|\chi_M) &\leq E\Big(\Big|E_{\mathcal{A}}x^{(a)}\Big|\chi_M\Big) + E\Big(\Big|E_{\mathcal{A}}\Big(x - x^{(a)}\Big)\Big|\chi_M\Big) \\ &\leq a\, Pr(M) + E\Big(\Big|x - x^{(a)}\Big|\Big) \approx 0, \end{aligned}$$

ce qui prouve que $E_{\mathcal{A}}x$ est L^1, d'après le théorème 8.1. □

Théorème 8.4. *Soit $\mathcal{A}$ une algèbre de variables aléatoires, et x une variable aléatoire L^1. Alors x est L^1 sur presque tous les atomes de $\mathcal{A}$.*

Démonstration. Soit un réel $\varepsilon \gg 0$. Pour tout entier naturel non nul n, définissons a_n comme étant le plus petit entier naturel tel que :

$$Pr'_{\mathcal{A}}\left\{E_{\mathcal{A}}\left|x - x^{(a_n)}\right| \geq \frac{1}{n}\right\} \leq \frac{\varepsilon}{2^n}$$

(reportez-vous au chapitre 2 pour un rappel de la définition de $Pr'_{\mathcal{A}}$).

Au préalable notons que si $n \ll \infty$, on a $a_n \ll \infty$. En effet,

$$E_{\mathcal{A}}\left(\left|x - x^{(a)}\right|\right)$$

est une variable aléatoire sur l'espace $at(\mathcal{A})$, dont l'espérance vaut :

$$E'_{\mathcal{A}}E_{\mathcal{A}}\left(\left|x - x^{(a)}\right|\right) = E\left(\left|x - x^{(a)}\right|\right).$$

Donc l'inégalité de Bienaymé-Tchebychev se traduit immédiatement en l'inégalité suivante :

$$Pr'_{\mathcal{A}}\left\{E_{\mathcal{A}}\left(\left|x - x^{(a)}\right|\right) \geq \frac{1}{n}\right\} \leq nE\left(\left|x - x^{(a)}\right|\right)$$

et $E_{\mathcal{A}}\left(\left|x - x^{(a)}\right|\right)$ est infinitésimal (donc inférieur à $\varepsilon/2^n$ si $n \ll \infty$) dès que a est illimité. Le principe de Cauchy permet de conclure cette propriété préliminaire.

Soit M l'ensemble défini ainsi :

$$M = \left\{\alpha \in at(\mathcal{A}) : \exists n \in N, E_{\mathcal{A}}\left(\left|x - x^{(a_n)}\right|\right)(\alpha) \geq \frac{1}{n}\right\}.$$

On a $Pr'_{\mathcal{A}}(M) \leq \varepsilon$. De plus, tous les atomes qui n'appartiennent pas à M vérifient la propriété suivante :

$$\forall a \approx \infty, \forall n \ll \infty, E_{\mathcal{A}}\left(\left|x - x^{(a)}\right|\right) < \frac{1}{n}$$

de par le fait que si a est illimité, $a > a_n$ pour tout n limité. Ceci montre que sur tous les atomes qui ne sont pas dans M x est L^1. Comme $\varepsilon \gg 0$ est arbitraire, nous pouvons conclure la démonstration. □

Inversement, supposons que x soit L^1 sur *tous* les atomes de $\mathcal{A}$. Si a est un réel illimité, alors $E_{\mathcal{A}}\left(\left|x - x^{(a)}\right|\right) \approx 0$ sur tous les atomes de $\mathcal{A}$, donc $E\left(\left|x - x^{(a)}\right|\right) \approx 0$. Ce qui montre que si x est L^1 sur tous les atomes de $\mathcal{A}$, x est L^1. On ne peut pas affirmer plus en général, car il est toujours possible de modifier une variable aléatoire en un point de probabilité infinitésimale pour obtenir une variable qui n'est plus L^1.

Du théorème 8.4, tirons le corollaire suivant :

Corollaire (Fubini). *Si x est L^1 sur $\langle \Omega_1 \times \Omega_2, pr_1 \times pr_2 \rangle$, alors la variable aléatoire x_{ω_1} définie sur $\langle \Omega_2, pr_2 \rangle$ de la manière suivante :*

$$x_{\omega_1}(\omega_2) = x(\omega_1, \omega_2)$$

est L^1 pour presque tous les points ω_1 de Ω_1.

Chapitre 9

Décomposition des processus stochastiques

Nous entrons maintenant dans l'étude des processus stochastiques indexés par un sous-ensemble fini de R, T. Nous ferons usage des notations générales introduites au début du chapitre 6. Les deux situations typiques dans lesquelles nous nous placerons seront celle où $T = \{1, \cdots, \nu\}$, ν étant un entier naturel illimité, et celle où T est un quasi-intervalle de R. Ainsi, bien que nous imposions à T d'être un ensemble fini, nous étudierons aussi bien les suites "infinies" de variables aléatoires que les processus à temps "continu".

Soit $\Phi : t \mapsto \Phi_t$ une fonction croissante de T dans l'ensemble de toutes les algèbres de variables aléatoires sur $\langle \Omega, pr \rangle$. Nous appellerons une telle application une *filtration*. Afin de ne pas alourdir le texte, il sera convenu d'écrire E_t pour désigner E_{Φ_t}. Un Φ-*processus*, ou un *processus adapté à* Φ, est un processus stochastique ξ indexé par T, tel que $\forall t \in T, \xi(t) \in \Phi_t$.[1] Pour un tel processus dès lors que $s \leq t$, $\xi(s) \in \Phi_t$ du fait que $\Phi_s \subseteq \Phi_t$. Tout processus stochastique ξ indexé par T est un Φ-processus si l'on définit Φ_t comme la sous-algèbre engendrée par les variables aléatoires $\{\xi(s) : s \leq t\}$. Cependant, on a intérêt à admettre que Φ_t puisse être une algèbre plus large. L'algèbre Φ_t représente le passé à l'instant t,[2] et si y est une variable aléatoire, $E_t y$ est la meilleure estimation de y qui puisse être proposée, étant donné le passé.

[1] C'est-à-dire $\xi \in \prod_{t \in T} \Phi_t$. (NdT)

[2] C'est un modèle de l'information qui a pu être recueillie lors des instants passés et présent, de provenances endogènes et exogènes. Lorsque les algèbres Φ_t sont simplement les algèbres engendrées par les variables $\{\xi(s) : s \leq t\}$, la seule source d'information est endogène. (NdT)

Un Φ-processus ξ est une *martingale* lorsque :

$$\forall (s,t) \in T^2, (s \leq t) \Longrightarrow (\xi(s) = E_s \xi(t)),$$

ξ est une *sousmartingale* lorsque :

$$\forall (s,t) \in T^2, (s \leq t) \Longrightarrow (\xi(s) \leq E_s \xi(t)),$$

et ξ est une *surmartingale* lorsque :

$$\forall (s,t) \in T^2, (s \leq t) \Longrightarrow (\xi(s) \geq E_s \xi(t)).$$

Ainsi, dans le cas trivial où Ω ne contient qu'un élément, une martingale est simplement une fonction constante de T dans R, une sousmartingale est une fonction croissante, et une surmartingale une fonction décroissante.

Si ξ est une martingale, alors $\forall t \in T, \xi(t) = E_t \xi(b)$. Inversement, si nous nous donnons une filtration Φ et une variable aléatoire quelconque x, le processus ξ défini par $\xi(t) = E_t x$ est une martingale.

Soit ξ un Φ-processus. Nous définissons les processus stochastiques (indexés par T') $D\xi$, $\hat{\xi}$, et σ_ξ^2 de la manière suivante :

$$\begin{array}{lcl} D\xi(t)\,dt & = & E_t\, d\xi(t); \\ d\xi(t) & = & D\xi(t)\,dt + d\hat{\xi}(t); \\ \sigma_\xi^2(t)\,dt & = & E_t\Big(d\hat{\xi}(t)^2\Big). \end{array}$$

$D\xi(t)\,dt$ et $\sigma_\xi^2(t)\,dt$ sont respectivement l'espérance conditionnelle et la variance conditionnelle de l'accroissement $d\xi$ du processus ξ. Les processus $D\xi$ et σ_ξ^2 sont des Φ-processus, en revanche ni $d\xi$ ni $d\hat{\xi}$ n'en sont en général.

On voit que $D\xi = 0$ lorsque ξ est une martingale, que $D\xi \geq 0$ lorsque ξ est une sousmartingale, et que $D\xi \leq 0$ lorsque ξ est une surmartingale. Il va de soi qu'en général $D\xi$ n'est pas forcément de signe constant lorsque t ou ω varient. Nous allons montrer que ces conditions sont tout aussi nécessaires que suffisantes. Nous pouvons écrire :

$$\forall (s,t) \in T^2, (s \leq t) \Rightarrow \left(\xi(t) = \xi(s) + \sum_{s \leq r < t} D\xi(r)\,dr + \sum_{s \leq r < t} d\hat{\xi}(r) \right). \tag{9.1}$$

Maintenant, notons que $\forall r \in T', E_r\, d\hat{\xi}(r) = 0$. Du fait que

$$(s \leq r) \Rightarrow (\Phi_s \subseteq \Phi_r),$$

il vient que $(s \leq r) \Rightarrow (E_s = E_s E_r)$. Cela nous permet d'écrire :

$$\forall (s,t) \in T^2, (s \leq t) \Rightarrow \left(E_s \xi(t) = \xi(s) + E_s \left(\sum_{s \leq r < t} D\xi(r)\, dr \right) \right). \tag{9.2}$$

Nous pouvons donc conclure que :

- ξ est une martingale si et seulement si $D\xi = 0$;
- ξ est une sousmartingale si et seulement si $D\xi \geq 0$;
- ξ est une surmartingale si et seulement si $D\xi \leq 0$.

Le processus $D\xi$ s'appelle la *tendance* de ξ, et lorsque $d\hat{\xi} = 0$ nous dirons que ξ est un processus *prédictible*. Le processus ξ est donc prédictible si et seulement si $\sigma_\xi^2 = 0$, ou de manière équivalente, si et seulement si $d\xi$ est un Φ-processus.

Posons $\tilde{\xi}(t) = \sum_{s<t} D\xi(s)\, ds$, en convenant que $\tilde{\xi}(a) = 0$. Le processus $\tilde{\xi}$ qui est tel que $d\tilde{\xi}(t) = D\xi(t)\, dt$ est adapté à Φ, de sorte que $\tilde{\xi}$ est un processus prédictible. Nous appellerons ce processus le *processus prédictible associé à* ξ. Il est visible que dès que Φ_t est connu, $d\tilde{\xi}(t)$ est parfaitement déterminé : on connaît l'accroissement immédiatement à venir de $\tilde{\xi}$. En revanche, pour connaître l'accroissement suivant, il faudrait se donner $\Phi_{(t+dt)}$, ce qui n'est pas généralement possible lorsque l'on est à l'instant t, la connaissance de Φ_t et de $d\tilde{\xi}(t)$ n'étant pas toujours suffisante.[3]

Posons $\hat{\xi}(t) = \xi(a) + \sum_{s<t} d\hat{\xi}(s)$. Alors, $\hat{\xi}$ est un processus adapté à Φ dont les accroissements successifs, ainsi que les notations le suggèrent, sont les variables $d\hat{\xi}(t)$. Le processus $\hat{\xi}$ est de plus une martingale (car $D\hat{\xi} = 0$), que nous appellerons *martingale associée à* ξ. Si ξ est lui-même une martingale, alors $\hat{\xi} = \xi$.

Nous pouvons décomposer n'importe quel Φ-processus sous la forme d'un processus prédictible plus une martingale de la manière suivante :

$$\xi = \tilde{\xi} + \hat{\xi},$$

et cette décomposition est unique si l'on impose $\tilde{\xi}(a) = 0$ (contrainte de normalisation).

Si les variables $d\xi(t)$ sont indépendantes entre elles,[4] et si la filtration Φ est engendrée par le processus ξ (chaque Φ_t est engendrée par la famille des variables $\xi(s), s \leq t$), alors $D\xi(t)dt = E_t\, d\xi(t) = E\, d\xi(t)$. Par conséquent, lorsque les variables $d\xi(t)$ sont indépendantes entre elles et de moyenne

[3] On retrouve là l'idée d'information exogène. (NdT)

[4] Et indépendantes de la variable $\xi(a)$. (NdT)

nulle, le processus ξ défini par la relation $\xi(t) = \sum_{s<t} d\xi(s)$ est une martingale. Pour définir un tel processus (à l'équivalence près), il suffit de préciser les lois de probabilité de chaque incrément $d\xi(t)$. Voici deux exemples.

Le premier exemple fournit un type de processus que l'on appelle une *marche aléatoire de Wiener* :

$$d\xi(t) = \begin{cases} \sqrt{dt} & \text{avec une probabilité de } \frac{1}{2}; \\ -\sqrt{dt} & \text{avec une probabilité de } \frac{1}{2}. \end{cases}$$

Le second exemple fournit un type de processus que l'on appelle *une marche aléatoire de Poisson* : en supposant que pour tout t de T, $dt \leq 1$, on pose

$$d\xi(t) = \begin{cases} 1 & \text{avec une probabilité de } \frac{dt}{2}; \\ 0 & \text{avec une probabilité de } 1 - dt; \\ -1 & \text{avec une probabilité de } \frac{dt}{2}. \end{cases}$$

Soit une martingale ξ. Si $r_1 < r_2$, alors $d\xi(r_1) \in \Phi_{r_2}$. Comme E_{r_2} est un opérateur Φ_{r_2} -linéaire, $E_{r_2}\, d\xi(r_1)\, d\xi(r_2) = d\xi(r_1) E_{r_2}\, d\xi(r_2) = 0$. Mais puisque pour tout indice $s \leq r_2$ on a $E_s = E_s E_{r_2}$, il vient :

$$((r_1 \neq r_2) \text{ et } (s \leq \max\{r_1, r_2\})) \Rightarrow (E_s\, d\xi(r_1) d\xi(r_2) = 0).$$

De plus, lorsque $s \leq t$ on a $\xi(t) - \xi(s) = \sum_{s \leq r < t} d\xi(r)$. On obtient donc l'égalité $E_s\Big((\xi(t) - \xi(s))^2\Big) = E_s \sum_{s \leq r < t} \sigma_\xi^2(r)\, dr$.

En passant à l'espérance absolue dans les égalités précédentes, on voit que l'ensemble des accroissements successifs de ξ, $(d\xi(t))_{t \in T'}$, forme une famille orthogonale, et que :

$$(s \leq t) \Rightarrow \left(\|\xi(t) - \xi(s)\|_2^2 = \sum_{s \leq r < t} \|d\xi(r)\|_2^2 \right).$$

Le symbole ξ désignant toujours une martingale, soit η un Φ-processus quelconque, et construisons le Φ-processus ζ ainsi :

$$\forall t \in T, \zeta(t) = \sum_{s<t} \eta(s)\, d\xi(s)$$

si bien que $\zeta(a) = 0$ et que $d\zeta = \eta\, d\xi$. Ainsi $D\zeta = \eta D\xi = 0$, ce qui montre que ζ est également une martingale, et $\sigma_\zeta^2 = \eta^2 \sigma_\xi^2$. Le processus ξ peut représenter par exemple un jeu équilibré, et η peut représenter la stratégie adoptée par le joueur. A chaque instant t, le joueur s'appuie sur sa connaissance du passé pour décider du facteur $\eta(t)$ à appliquer à son pari. Dans ce contexte, la valeur $\zeta(t)$ mesure le gain (ou la perte) effectué par le joueur à l'instant t depuis le début du jeu. Alors, peu importe la subtilité que le joueur aura pu apporter au choix de sa stratégie : à tout moment, le gain potentiel sera contrebalancé par une perte potentielle d'un même montant, puisque $E\zeta(t) = 0$.

Une variable aléatoire x sera qualifiée de *normalisée* si $Ex = 0$, et si de plus $var(x) = 1$. Pour une variable x qui n'est pas constante, on peut construire la variable normalisée

$$\check{x} = \frac{(x - \mu)}{\sigma}$$

où μ et σ sont respectivement la moyenne et l'écart-type de x, et nous pouvons représenter x sous la forme :

$$x = \mu + \sigma\check{x}.$$

Dans le cas où x est constante, une telle représentation est encore possible en choisissant pour $\check{x}$ une variable normalisée arbitraire. Le seul cas où on ne peut aboutir à cette forme d'écriture est celui où Ω ne contient qu'un seul élément. La solution passe alors par l'extension de Ω en un espace plus large.

Ces rappels élémentaires nous inspirent de nouvelles notions à propos des processus stochastiques. Nous qualifierons un processus stochastique ξ de *normalisé* lorsqu'à la fois $D\xi = 0$ et $\sigma_\xi^2 = 1$. On voit en particulier qu'un processus normalisé est toujours une martingale. La marche aléatoire de Wiener et celle de Poisson, sont toutes les deux des processus normalisés. Un processus stochastique ξ sera dit *non-dégénéré* dès lors que $\forall(t,\omega) \in T' \times \Omega, \sigma_\xi^2(t,\omega) > 0$. Pour un tel processus, définissons le processus $\check{\xi}(t) = \sum_{s<t} \sigma_\xi^{-1}(s)\, d\hat{\xi}(s)$, où $\sigma_\xi^{-1} = 1/\sqrt{\sigma_\xi^2}$, et où $\hat{\xi}$ est la martingale associée à ξ. Il est clair que $\check{\xi}$ est normalisé, et que ξ admet la représentation suivante :

$$\xi(t) = \xi(a) + \sum_{s<t} D\xi(s)\, ds + \sum_{s<t} \sigma_\xi\, d\check{\xi}(s). \tag{9.3}$$

Cette écriture reste possible lorsque ξ est dégénéré. Il faut alors étendre l'espace probabilisé. Soit η un processus normalisé indexé par T, et défini sur un espace probabilisé fini $\langle \Omega', pr' \rangle$. Définissons le processus σ_ξ^{-1} de la manière suivante :

$$\sigma_\xi^{-1}(t) = \begin{cases} \frac{1}{\sqrt{\sigma_\xi^2(t)}} & \text{lorsque } \sigma_\xi^2(t) > 0\,; \\ 0 & \text{lorsque } \sigma_\xi^2(t) = 0\,; \end{cases}$$

et construisons sur l'espace $\langle \Omega \times \Omega', pr \times pr' \rangle$ le processus $\check{\xi}$ ainsi :

$$\begin{array}{rcl} \check{\xi}(a) & = & 0; \\ d\check{\xi}(s) & = & \sigma_\xi^{-1}(s)\, d\hat{\xi}(s) + \chi_{\left\{\sigma_\xi^2(s)=0\right\}}\, d\eta(s). \end{array}$$

Le processus $\check{\xi}$ est normalisé, et nous pouvons l'utiliser pour obtenir la représentation 9.3.

Chapitre 10

Variation totale d'un processus

Le propos du chapitre précédent était exclusivement interne. Maintenant, nous allons nous intéresser essentiellement aux processus pour lesquels $D\xi$ et σ_ξ^2 sont de norme limitée. Dans ce cas, les accroissements $d\tilde{\xi}(t) = D\xi(t)\,dt$ du processus prédictible associé seront de l'ordre de dt, tandis que les accroissements $d\hat{\xi}(t)$ de la martingale associée, seront de l'ordre de $\sqrt{dt}$. Les fluctuations intéressantes d'un processus stochastique sont celles qui viennent de sa martingale associée. Cependant, dans le présent chapitre nous étudierons les fluctuations d'un processus par l'intermédiaire de l'estimation de la variable $\sum |d\xi(t)|$, approche trop grossière en soit, cependant, pour être appliquée de manière fructueuse à l'étude des martingales. Nous allons voir que lorsqu'un processus stochastique ξ est tel que son taux d'accroissement $d\xi/dt$ est L^1 à la fois en ω et en t (en un sens que nous allons préciser bientôt), alors ξ est presque partout absolument continu (c'est-à-dire que, presque sûrement, les trajectoires du processus sont absolument continues), et presque sûrement la martingale associée reste en chaque instant infiniment proche de sa valeur initiale. Cette dernière propriété est vraie également dans le cas d'un processus croissant dont les incréments sont partout infinitésimaux.

Commençons, à l'aide de la définition $pr'(t) \overset{def}{=} \frac{dt}{b-a}$, par doter l'ensemble T' d'une structure d'espace probabilisé fini $\langle T', pr' \rangle$. Nous noterons la moyenne dans cet espace par E_T. Si ξ est une fonction réelle définie sur T', la variation totale de ξ a pour valeur $(b-a)\,E_T|d\xi/dt|$. Sous l'hypothèse que $(b-a)$ est appréciable, ξ est à variation limitée si et seulement si la grandeur $E_T|d\xi/dt|$ est limitée. Ceci en plus du théorème 8.1 (Radon-Nikodym), montre que ξ est absolument continue si et seulement si $d\xi/dt$ est L^1 sur $\langle T', pr' \rangle$.

Soit maintenant un processus stochastique ξ indexé par T. Nous décidons de désigner par la notation ξ_n la valeur $\xi(n)$ lorsque T est un sous-ensemble $\{1,\cdots,\nu\}$ de N.

Théorème 10.1. *On a les cas suivants :*

1. *si la somme $\sum_{t\in T'}\|d\xi(t)\|_1$ est limitée, alors la v.a.r. $\sum_{t\in T'}|d\xi(t)|$ est presque sûrement limitée (autrement dit, ξ est presque sûrement à variation limitée) ;*
2. *lorsque $T=\{1,\cdots,\nu\}$, si la série $\sum_{n\in T'}\|d\xi_n\|_1$ converge, alors le processus ξ converge presque sûrement ;*
3. *si $d\xi/dt$ est L^1 sur l'espace $\langle T'\times\Omega, pr'\times pr\rangle$, alors presque sûrement les trajectoires de ξ sont absolument continues.*

Démonstration. Pour l'assertion 1, posons $c=\sum_{t\in T'}\|d\xi(t)\|_1$, et

$$x=\sum_{t\in T'}|d\xi(t)|,$$

si bien que $c=Ex$. Alors, l'assertion 1 dit simplement que lorsque c est limité, la variable x l'est presque sûrement, ce que montre immédiatement l'inégalité de Bienaymé-Tchebychev.

Dans le cas 2, supposons que les conditions de l'assertion soient vérifiées. Du fait du théorème 7.2, nous pouvons tâcher de montrer alors que, quel que soit le réel $\lambda\gg 0$, et quel que soit l'indice n inférieur ou égal à ν et infiniment grand, nous avons $Pr\left(\sum_{i=n}^{\nu-1}|d\xi_i|\geq\lambda\right)\approx 0$. Or on a :

$$Pr\left(\sum_{i=n}^{\nu-1}|d\xi_i|\geq\lambda\right)\leq\sum_{i=n}^{\nu-1}\|d\xi_i\|_1/\lambda$$

d'après l'inégalité de Bienaymé-Tchebychev, d'où le résultat.

L'assertion 3 est une conséquence triviale du théorème de Fubini (8.4). Ce qui clôt la démonstration. □

L'assertion 3 est exprimée pour un processus stochastique général. Mais le théorème 10.4 nous montrera que les conditions de cette assertion sont très restrictives, le processus ξ étant alors pratiquement égal à un processus prédictible en ce sens que presque partout sa martingale associée (quelle que soit la filtration à laquelle ξ est adaptée) vérifie la proposition $\forall t\in T, \hat{\xi}(t)\approx\hat{\xi}(a)$.

Théorème 10.2. *Soit ξ un processus adapté à une filtration Φ, tel que $d\xi/dt$ soit L^1 sur $\langle T'\times\Omega, pr'\times pr\rangle$. Alors $D\xi$ est L^1 sur $T'\times\Omega$, de même que l'est le processus $\frac{d\hat{\xi}}{dt}=\frac{d\xi}{dt}-D\xi$.*

Démonstration. Définissons sur $T' \times \Omega$ l'algèbre de variables aléatoires Φ' comme étant l'algèbre engendrée par les variables y de l'espace $T' \times \Omega$ telles qu'en chaque instant t de T', la variable $y(t,.)$ soit un élément de Φ_t. Alors, $D\xi = E_{\Phi'} \frac{d\xi}{dt}$, et le théorème 8.3 nous permet de conclure. □

Le résultat qui suit est un lemme de troncature. Il affirme que sous certaines conditions, il est possible de modifier très légèrement un processus (c'est-à-dire de manière telle que presque partout le processus modifié reste en chaque instant infiniment proche du processus initial), afin que les accroissements élémentaires du nouveau processus soient en tout point infinitésimaux.

Théorème 10.3. *Soit une martingale ξ indexée par un quasi-intervalle T et telle que $d\xi/dt$ est L^1 sur $T' \times \Omega$. Pour tout réel $\alpha > 0$ nous définissons la martingale $\xi_{(\alpha)}$ ainsi :*

$$\begin{array}{rcl} \xi_{(\alpha)}(a) & = & \xi(a) \\ d\xi_{(\alpha)}(t) & = & d\xi(t)^{(\alpha)} - E_t\, d\xi(t)^{(\alpha)} \end{array}$$

de manière que $\forall (t,\omega) \in T' \times \Omega, \left|d\xi_{(\alpha)}(t,\omega)\right| \leq 2\alpha$. Alors, il existe un réel infinitésimal α tel que presque partout la propriété suivante soit vérifiée :

$$\forall t \in T', \xi_{(\alpha)}(t) \approx \xi(t).$$

Démonstration. Sous les conditions du théorème, on peut appliquer le théorème 10.1 (assertion 3), pour voir que les trajectoires de ξ sont presque sûrement absolument continues, donc presque sûrement continues sur T. L'ensemble T étant un quasi-intervalle, on en conclut que presque sûrement $\max_{t \in T'} |d\xi(t)| \approx 0$. Alors, le théorème 7.1 nous assure l'existence d'un réel α infinitésimal vérifiant $Pr(\max_{t \in T'} |d\xi(t)| \geq \alpha) \approx 0$. Pour ce réel, la propriété suivante est vérifiée presque partout :

$$\forall t \in T, \xi(t) = \xi(a) + \sum_{s<t} d\xi^{(\alpha)}(s).$$

Ainsi, en se souvenant de la définition de $\xi_{(\alpha)}$ on voit que pour montrer que $\forall t \in T', \xi_{(\alpha)}(t) \approx \xi(t)$ presque partout il suffit d'établir en plus la relation $\sum_{t \in T'} \left| E_t\, d\xi(t)^{(\alpha)} \right| \approx 0$ presque partout. Pour cela, il suffit de montrer qu'il est possible de choisir α pour avoir en plus $E\Big(\sum_{t \in T'} \left| E_t\, d\xi(t)^{(\alpha)} \right|\Big) \approx 0$. Or, ξ étant une martingale, $\forall t \in T', E_t\, d\xi(t) = 0$, et :

$$E\left(\sum_{t \in T'} \left| E_t\, d\xi(t)^{(\alpha)} \right|\right) = E\left(\sum_{t \in T'} \left| E_t\Big(d\xi(t) - d\xi(t)^{(\alpha)}\Big) \right|\right)$$

d'où

$$E\left(\sum_{t\in T'}\left|E_t\,d\xi(t)^{(\alpha)}\right|\right) \leq E\left(\sum_{t\in T'}E_t\left|d\xi(t)-d\xi(t)^{(\alpha)}\right|\right)$$
$$= E\left(\sum_{t\in T'}\left|\frac{d\xi}{dt}-\left(\frac{d\xi}{dt}\right)^{\left(\frac{\alpha}{dt}\right)}\right|dt\right)$$

ce qui donne, si l'on convient que $c = \frac{\alpha}{\max_{t\in T'}(dt)}$:

$$E\left(\sum_{t\in T'}\left|E_t\,d\xi(t)^{(\alpha)}\right|\right) \leq E\left(\sum_{t\in T'}\left|\frac{d\xi}{dt}-\left(\frac{d\xi}{dt}\right)^{(c)}\right|dt\right).$$

Si α était sensiblement positif ($\alpha \gg 0$), il s'ensuit que c serait infiniment grand ($c \approx \infty$), et comme $d\xi/dt$ est L^1 sur $T' \times \Omega$, on aurait alors

$$E\left(\sum_{t\in T'}\left|\frac{d\xi}{dt}-\left(\frac{d\xi}{dt}\right)^{(c)}\right|dt\right) \approx 0.$$

Ainsi l'ensemble des réels x tels que $E\left(\sum_{t\in T'}\left|E_t\,d\xi(t)^{(x)}\right|\right) \leq x$ contient tous les réels sensiblement positifs. En vertu du principe de Cauchy, il est donc possible de choisir α infinitésimal suffisamment grand pour qu'en plus de ses propriétés déjà acquises, α satisfasse $E\left(\sum_{t\in T'}\left|E_t\,d\xi(t)^{(\alpha)}\right|\right) \approx 0$. Ceci conclut la démonstration. □

Lors des preuves des deux théorèmes qui suivent, nous aurons besoin d'un résultat (théorème 11.1) dont la démonstration sera exposée ultérieurement.

Théorème 10.4. *Soit ξ un processus stochastique indexé par un quasi-intervalle T, tel que $d\xi/dt$ soit L^1 sur $\langle T' \times \Omega, pr' \times pr\rangle$, et soit $\hat{\xi}$ sa martingale associée (pour une filtration quelconque). Alors ξ vérifie la proposition suivante :*

$$ps : \forall t \in T, \hat{\xi}(t) \approx \hat{\xi}(a).$$

Démonstration. On ne restreint pas le résultat en supposant que $\xi(a) = 0$. De surcroît, les théorèmes 10.2 et 10.3 nous permettent de supposer que $\xi = \hat{\xi}$ (c'est-à-dire que ξ est une martingale)[1] et qu'il existe un réel infinitésimal α tel que en tout point de $T' \times \Omega$, $|d\xi(t)| \leq \alpha$. Alors on voit que :

$$\forall t \in T', \|d\xi(t)\|_2^2 = E\left(d\xi(t)^2\right) \leq \alpha\, E|d\xi(t)|$$

[1]La propriété qui dit que $d\xi/dt$ est L^1 sur $T' \times \Omega$ se transmet à $\hat{\xi}$, et le résultat ne concerne que la martingale, on peut donc concentrer son attention sur celle-ci. (NdT)

et du fait que $d\xi/dt$ est L^1 sur $T' \times \Omega$, on a :

$$\sum_{t\in T'} E|d\xi(t)| = E\left(\sum_{t\in T'} \left|\frac{d\xi(t)}{dt}\right| dt\right) \ll \infty.$$

Ainsi :

$$\|\xi(b)\|_1^2 \leq \|\xi(b)\|_2^2 = \sum_{t\in T'} \|d\xi(t)\|_2^2 \leq \alpha \sum_{t\in T'} E|d\xi(t)| \approx 0$$

et les théorèmes 7.1 et 11.1 nous enseignent que dans ce cas on a presque partout $\max_{t\in T} |\xi(t)| \approx 0$, ce qui achève la preuve. □

Théorème 10.5. *Soit ξ un processus stochastique croissant, d'accroissement infinitésimal en tout point de $T' \times \Omega$, et soit $\hat{\xi}$ sa martingale associée. Alors ξ vérifie la proposition suivante :*

$$(E(\xi(b) - \xi(a)) \ll \infty) \Rightarrow \Big(ps : \forall t \in T, \hat{\xi}(t) \approx \hat{\xi}(a)\Big).$$

Démonstration. Il n'y a pas d'inconvénient à supposer que $\xi(a)=0$. Soit $\alpha = \max_{(t,\omega)\in T'\times\Omega} |d\xi(t,\omega)|$. Par hypothèse, α est infinitésimal. Alors :

$$\begin{aligned}\left\|\hat{\xi}(b)\right\|_1^2 \leq \left\|\hat{\xi}(b)\right\|_2^2 &= \sum_{t\in T'} \left\|d\hat{\xi}(t)\right\|_2^2 \leq \sum_{t\in T'} \|d\xi(t)\|_2^2 = \sum_{t\in T'} E\Big(d\xi(t)^2\Big) \\ &\leq \alpha\, E\left(\sum_{t\in T'} d\xi(t)\right) = \alpha\, E(\xi(b) - \xi(a)) \approx 0.\end{aligned}$$

Ici encore, les théorèmes 7.1 et 11.1 nous permettent de conclure. □

Chapitre 11

Convergence des martingales

Dans ce chapitre, nous chercherons à estimer la valeur maximale que peut atteindre une surmartingale ou une sousmartingale, et à utiliser ce résultat pour étudier des propriétés de convergence et de continuité en un instant donné d'un processus.

Imaginons que le prix sur un marché public d'une action de la société des "Transports des Encombrants Industriels" (T.E.I.) soit bien modélisé par le processus stochastique ξ. Une stratégie d'investissement est dite haussière (bullish dans la terminologie anglo-saxonne) si elle consiste à acheter une action à l'instant a, et à la conserver tant que le prix de l'action n'a pas augmenté d'au moins λ Euros par rapport au prix d'achat. Dès que cet événement survient, l'investisseur revend son action et empoche son gain. Dans ce scénario, le gain de l'investisseur à l'instant t vaut :

$$\zeta_1(t) = \sum_{s<t} \eta_1(s)\, d\xi(s)$$

où $\eta_1(s)$ vaut 1 si à l'instant s l'investisseur est encore en possession de l'action, et 0 dans le cas contraire. La valeur de $\eta_1(s)$ est entièrement déterminée par les valeurs prises par le processus ξ en chaque instant $r \leq s$, aussi η_1 est-il un processus adapté. On peut se reporter au chapitre 9 pour une discussion générale sur les sommes de cette forme ("intégrales stochastiques"). Soit Λ l'événement qui correspond au succès de la stratégie que nous venons de décrire, c'est-à-dire :

$$\Lambda = \left\{\omega \in \Omega : \max_{t\in T} \left(\xi(t,\omega) - \xi(a,\omega)\right) \geq \lambda\right\}.$$

Dans l'hypothèse favorable où l'investisseur a rencontré le succès, son gain à l'instant b sera au moins égal à λ. Dans le cas contraire, il sera exactement égal à $(\xi(b) - \xi(a))$. Ainsi :

$$\zeta_1(b) \geq \lambda\chi_\Lambda + (\xi(b) - \xi(a))\chi_{\Lambda^c}. \tag{11.1}$$

Imaginons maintenant que, pour le plus grand malheur de notre investisseur, le processus ξ soit une surmartingale, c'est-à-dire que le cours de l'action T.E.I. soit à la baisse. Du fait que η_1 est un processus adapté positif, il vient que $D\zeta_1(t) = \eta_1(t)\, D\xi(t) \leq 0$, où l'on voit que ζ_1 est aussi une surmartingale. Comme $\zeta_1(a) = 0$, $E\zeta_1(b) \leq 0$. Alors, en introduisant cette inégalité dans la relation 11.1, on voit que :

$$Pr\left\{\max_{t \in T}\,(\xi(t) - \xi(a)) \geq \lambda\right\} \leq \frac{\|\xi(b) - \xi(a)\|_1}{\lambda}. \tag{11.2}$$

Le scénario que nous venons d'imaginer ne fait pas apparaître grand chose si ξ est une sousmartingale. Mais supposons maintenant que l'investisseur adopte une stratégie baissière (bearish). Cette stratégie consiste à attendre que le prix de l'action augmente de λ Euros par rapport à son prix à l'instant a, puis à partir de ce moment, à conserver l'action. Dans cette situation, le gain à l'instant t se calcule ainsi :

$$\zeta_2(t) = \sum_{s<t} \eta_2(s)\, d\xi(s)$$

où $\eta_2 = 1 - \eta_1$, si bien que η_2 est un processus positif adapté. Le gain final (à l'instant b) est nul, sauf dans l'éventualité où le cours de l'action depuis l'instant a a rencontré à un certain moment les conditions d'achat. Si cela se produit, le gain $\zeta_2(b)$ vaut au plus $\xi(b) - \xi(a) - \lambda$. Donc :

$$\zeta_2(b) \leq (\xi(b) - \xi(a) - \lambda)\chi_\Lambda. \tag{11.3}$$

En remarquant que $\zeta_1 + \zeta_2 = \xi - \xi(a)$, nous voyons que l'inégalité 11.3 n'est rien d'autre qu'une reformulation de l'inégalité 11.1.

Lorsque ξ est une sousmartingale (action à la hausse), il en est de même de ζ_2 du fait que $D\zeta_2(t) = \eta_2(t)\, D\xi(t) \geq 0$. Comme $\zeta_2(a) = 0$, $E\zeta_2(b) \geq 0$, et en introduisant ceci dans la relation 11.3, nous voyons que la propriété 11.2 reste vraie pour une sousmartingale.

La propriété 11.2 est donc vraie aussi bien pour les surmartingales que pour les sousmartingales. Elle est par conséquent également vérifiée pour une martingale, qui est en même temps une surmartingale et une sousmartingale. Comme $-\xi$ est une surmartingale si et seulement si ξ est une sousmartingale, et vice versa, l'inégalité suivante est également valide à la fois pour les

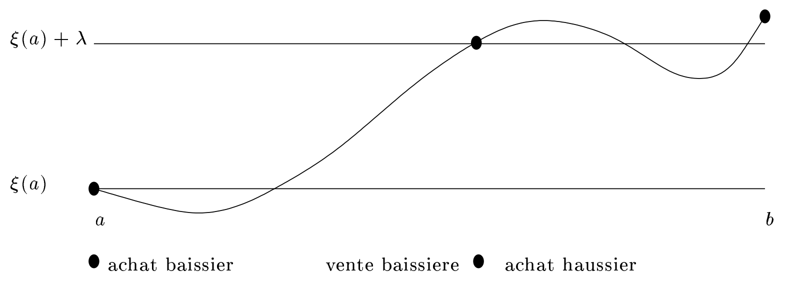

FIG. 11.1 – deux stratégies d'investissement

surmartingales et les sousmartingales :

$$Pr\left\{\max_{t\in T}(\xi(a)-\xi(t))\geq\lambda\right\}\leq\frac{\|\xi(b)-\xi(a)\|_1}{\lambda}.$$

Nous venons ainsi de démontrer le théorème suivant :

Théorème 11.1. *Soit ξ une surmartingale ou une sousmartingale. Alors :*

$$\forall\lambda>0, Pr\left\{\max_{t\in T}(\xi(t)-\xi(a))\geq\lambda\right\}\leq\frac{1}{\lambda}\|\xi(b)-\xi(a)\|_1,$$
$$\forall\lambda>0, Pr\left\{\max_{t\in T}|\xi(t)-\xi(a)|\geq\lambda\right\}\leq\frac{2}{\lambda}\|\xi(b)-\xi(a)\|_1. \tag{11.4}$$

Dans le cas où ξ est une martingale, le théorème suivant nous permet de rendre la relation 11.4 deux fois plus précise.

Théorème 11.2. *Nous avons les implications suivantes :*

1. *si ξ est une martingale, et si f est une fonction réelle convexe, alors $f\circ\xi$ est une sousmartingale ;*
2. *si ξ est une sousmartingale, est si f est une fonction réelle convexe croissante, alors $f\circ\xi$ est une sousmartingale.*

Démonstration. Lorsque f est une fonction convexe, nous pouvons appliquer la relativisation de l'inégalité de Jensen : $f(E_s\xi(t))\leq E_sf(\xi(t))$. Or si ξ est une martingale, pour $s\leq t$ on a $E_s\xi(t)=\xi(s)$. On voit donc que dans ce cas $f\circ\xi$ est une sousmartingale. Si maintenant ξ est une sousmartingale, pour $s\leq t$ on a $E_s\xi(t)\geq\xi(s)$. Alors, si f est à la fois convexe et croissante, $E_sf(\xi(t))\geq f(E_s\xi(t))\geq f(\xi(s))$, ce qui montre ici encore que $f\circ\xi$ est une sousmartingale. □

Si ξ est une martingale, alors $(\xi - \xi(a))$ est une martingale,[1] et le théorème précédent nous dit que $|\xi - \xi(a)|$ est une sousmartingale. La première relation du théorème (11.1) nous permet dans ce cas de remplacer la relation 11.4 par la relation :

$$\forall \lambda > 0, Pr\left\{\max_{t \in T} |\xi(t) - \xi(a)| \geq \lambda\right\} \leq \frac{1}{\lambda}\|\xi(b) - \xi(a)\|_1.$$

Chacune des normes $\|.\|_p, p \in [1, \infty]$, dote l'ensemble R^{Ω} des v.a.r. de $\langle \Omega, pr \rangle$ d'une structure d'espace métrique que nous désignerons par $\mathbf{L}^p$. Pour des valeurs de p différentes, nous conservons le même ensemble d'éléments R^{Ω}, et nous faisons simplement varier la structure métrique de cet ensemble. Nous dirons que la séquence de variables aléatoires $x_1, \cdots, x_\nu$ *converge (quasiment) vers y dans $\mathbf{L}^p$* lorsque nous aurons :

$$\forall n \leq \nu, (n \approx \infty) \Longrightarrow \Big(\|x_n - y\|_p \approx 0\Big).$$

Si ξ est un processus stochastique indexé par un sous-ensemble T de R, pour tout point t de T nous dirons que ξ est *(quasiment) continu en t dans $\mathbf{L}^p$* si :

$$\forall s \in T, (s \approx t) \Longrightarrow \Big(\|\xi(t) - \xi(s)\|_p \approx 0\Big)$$

et dans le cas contraire, nous dirons que ξ est *(fortement) discontinu en t dans $\mathbf{L}^p$*.

La continuité en un point d'un processus dans $\mathbf{L}^p$ est une propriété analytique qui est facilement vérifiable ou réfutable. En revanche, la continuité presque sûre en un point des trajectoires d'un processus est une propriété probabiliste intéressante et subtile. Un grand nombre de théorèmes portant sur les processus stochastiques établissent la vérité presque sûre d'une proposition donnée relative aux trajectoires, conditionnellement à des hypothèses analytiques adéquates.

Théorème 11.3. *Nous avons :*

1. *soit la séquence de variables aléatoires $x_1, \cdots, x_\nu$ formant une surmartingale ou une sousmartingale convergente dans $\mathbf{L}^1$. Alors cette séquence converge presque sûrement ;*
2. *soit le processus stochastique ξ formant une surmartingale ou une sousmartingale, et continu en t dans $\mathbf{L}^1$. Alors ξ est presque sûrement continu en t.*

[1] En effet $\xi(a)$ appartient à chaque algèbre Φ_t, et est donc invariant pour chacun des opérateurs d'espérance conditionnelle. (NdT)

La démonstration découle immédiatement du théorème 11.1 ainsi que du théorème 7.2 et de son corollaire.

Le théorème suivant établit dans certains cas les propositions inverses.

Théorème 11.4. *Nous avons :*

1. *soit une séquence de variables aléatoires $x_1, \cdots, x_\nu$ formant une martingale telle que x_ν soit L^1. Alors cette séquence converge dans $\mathbf{L}^1$ si et seulement si elle converge presque sûrement ;*
2. *soit un processus stochastique ξ formant une martingale, et tel que $\xi(b)$ soit L^1. Alors, pour tout t de T, ξ et continu en t dans $\mathbf{L}^1$ si et seulement si ξ est presque sûrement continu en t.*

Démonstration. Etablissons uniquement la deuxième assertion, la première assertion se démontrant suivant le même principe. La variable $\xi(b)$ étant L^1, en tout point t de T la variable $\xi(t) = E_t\xi(b)$ est L^1 (théorème 8.3). Le théorème de Lebesgue (8.2) nous montre alors que si ξ est presque sûrement continu en t, ξ est continu en t dans $\mathbf{L}^1$. Le sens inverse est simplement le théorème 11.3. □

Pour montrer que la continuité (ou la convergence) presque sûre entraîne la continuité (ou la convergence) dans $\mathbf{L}^1$, l'hypothèse que ξ est une martingale ne nous a servi qu'à propager le caractère L^1 de $\xi(b)$ (ou de x_ν) à tous les $\xi(t)$ (à tous les x_n). Aussi ce résultat reste-t-il vrai en supposant simplement que chaque $\xi(t)$ (ou chaque x_n) est L^1.

Nous dirons que t est une *discontinuité (forte) fixe* de ξ lorsque ξ n'est pas presque sûrement continu en t. La deuxième assertion du théorème 11.4 peut être reformulée en disant que, lorsque ξ est une martingale telle que $\xi(b)$ soit L^1, pour tout t de T, t est une discontinuité fixe de ξ si et seulement si t est une discontinuité de ξ dans $\mathbf{L}^1$. Le contre-exemple suivant illustre un cas où $\xi(b)$ n'est pas L^1.

Soit Ω un ensemble ordonné de 2^ν points, avec $\nu \approx \infty$. A chaque point de Ω nous attribuons la probabilité $2^{-\nu}$. Pour tout indice n de $\{1, \cdots, \nu\}$ définissons la variable aléatoire x_n sur Ω dont la valeur est :

$$\begin{cases} -2^{(n-1)} & \text{sur les } 2^{(\nu-n)} \text{ premiers points de } \Omega; \\ +2^{(n-1)} & \text{sur les } 2^{(\nu-n)} \text{ derniers points de } \Omega; \\ \hfill 0 & \text{sur tous les points intermédiaires.} \end{cases}$$

Les valeurs non nulles du processus $(x_n)_{n\in\{1,\cdots,\nu\}}$ sont illustrées par la figure 11.2, où pour épargner la forêt nous nous sommes contentés de choisir $\nu = 4$ plutôt que $\nu \approx \infty$.

FIG. 11.2 – une mauvaise martingale

Pour la filtration Φ engendrée par $(x_n)_{n\in\{1,\cdots,\nu\}}$ (telle que chaque algèbre Φ_i est engendrée par la famille $\{x_j : j \leq i\}$) le processus $(x_n)_{n\in\{1,\cdots,\nu\}}$ est une martingale qui vérifie les propriétés suivantes :

$$\begin{array}{ll} \forall n \in \{1,\cdots,\nu\}, & (Ex_n = 0) \text{ et } (\|x_n\|_1 = 1), \\ \forall (n,m) \in \{1,\cdots,\nu\}^2, & (n \neq m) \Rightarrow (\|x_n - x_m\|_1 \geq 1). \end{array}$$

Soit alors un entier k tel que k et $\nu - k$ soient tous les deux illimités, et soit T un quasi-intervalle contenant $\nu - k$ points. Définissons le processus stochastique ξ tel que $\xi(t) = x_{(n+k)}$ si t est le n-ième élément de T. Alors, en tout point de T, ξ est discontinu dans $\mathbf{L}^1$. Cependant, comme presque partout $\forall t \in T, \xi(t) = 0$, ξ est presque partout continu en tout point de T.

Nous nous attachons dans le reste de ce chapitre à montrer que si ξ est une martingale telle que $\xi(b)$ est L^1, alors ξ présente très peu de discontinuités fixes.

Soit un réel $\varepsilon > 0$, et un espace métrique $\langle M, \rho\rangle$ (par exemple $\langle R, |.|\rangle$, ou $\langle \mathbf{L}^1, \|.\|_1\rangle$). Nous dirons qu'une fonction $\xi : T \to M$ est *(quasiment) ε-continue* au point t si elle vérifie la propriété suivante :

$$\forall (t_1,t_2) \in T^2, ((t_1 \approx t) \text{ et } (t_2 \approx t)) \Rightarrow (\rho(\xi(t_1),\xi(t_2)) \leq \varepsilon),$$

et si cette propriété n'est pas vraie en t, nous qualifierons t de *ε-discontinuité (forte)* de ξ.

Remarquons que si $s \approx t$, alors s est une ε-continuité de ξ si et seulement si t en est une.

Si t est une ε-discontinuité de ξ, alors il existe deux éléments de T infiniment proches de t, t_1 et t_2, tels que $\rho(\xi(t_1), \xi(t_2)) > \varepsilon$. En utilisant l'inégalité triangulaire, il apparaît qu'il existe t' (qui est soit t_1 soit t_2) dans T, tel que $\rho(\xi(t), \xi(t')) > \varepsilon/2$. Remarquons de plus que ξ n'est pas quasiment continue en t si et seulement si il existe un réel $\varepsilon \gg 0$ tel que t soit une ε-discontinuité de ξ.

Théorème 11.5. *Soit ξ une martingale indexée par un quasi-intervalle T, telle que $\xi(b)$ est L^1, et soit un réel $\varepsilon \gg 0$. Alors, il existe un nombre limité de points de T, $t_1, \cdots, t_n$, deux à deux discernables (non équivallents), tels que, pour tout élément t de T, t est une ε-discontinuité si et seulement si t est infiniment proche d'un élément de l'ensemble $\{t_1, \cdots, t_n\}$.*

Démonstration. L'ensemble $\left\{x > 0 : \left\|\xi(b) - \xi(b)^{(x)}\right\|_1 \leq \varepsilon/4\right\}$ contient tous les réels positifs infiniment grands. En vertu du principe de Cauchy, il contient également un élément positif appréciable c. Posons $\xi_c(t) = E_t\xi(b)^{(c)}$. Ainsi que nous l'avons vu au chapitre 9, le processus ainsi défini est une martingale. Du fait que l'espérance conditionnelle réduit les normes dans $\mathbf{L}^p$ (voir le chapitre 2), il vient que l'on a :

$$\forall t \in T, \|\xi_c(t) - \xi(t)\|_1 \leq \frac{\varepsilon}{4}.$$

Supposons maintenant l'existence de n points $t_1, \cdots, t_n$ deux à deux discernables, avec pour chaque indice i de $\{1, \cdots, n\}$ l'existence de deux points s_i et r_i dans T, tels que $s_i \approx r_i \approx t_i$, et $\|\xi(s_i) - \xi(r_i)\|_1 > \varepsilon$. Convenons que $s_i > r_i$. Les intervalles $[r_i, s_i]$ sont alors deux à deux disjoints. Par l'inégalité triangulaire, on voit que $\|\xi_c(s_i) - \xi_c(r_i)\|_1 > \varepsilon/2$ et donc que

$$\|\xi_c(s_i) - \xi_c(r_i)\|_2 > \frac{\varepsilon}{2}.$$

Du fait que les accroissements d'une martingale forment un système orthogonal, nous pouvons écrire les relations suivantes :

$$\begin{aligned} n\frac{\varepsilon^2}{4} &\leq \sum_{i=1}^{n} \|\xi_c(s_i) - \xi_c(r_i)\|_2^2 \leq \|\xi_c(b) - \xi_c(a)\|_2^2 \\ &= \|\xi_c(b) - E_a\xi_c(b)\|_2^2 \leq \|\xi_c(b)\|_2^2 \leq c^2. \end{aligned}$$

Ainsi $n \leq \frac{4c^2}{\varepsilon^2} \ll \infty$. Nous avons donc montré l'existence d'un entier limité qui possède la propriété d'être supérieur où égal à la taille de tout ensemble $\{t_1, \cdots, t_n\}$ constitué de ε-discontinuités de ξ deux à deux discernables (il suffit de considérer la partie entière de $\frac{4c^2}{\varepsilon^2}$). En vertu du principe externe du plus petit élément, il existe un entier n_0 qui possède cette même propriété, et un ensemble $T = \{t_1, \cdots, t_{n_0}\}$ de ε-discontinuité deux à deux discernables de ξ. L'ensemble T et n_0 possèdent bien la propriété cherchée : s'il existait une ε-discontinuité de ξ, s, qui ne soit pas infiniment proche d'un élément de T, la construction de l'ensemble $T \bigcup \{s\}$ conduirait à une contradiction. □

***Théorème 11.6.** *Soit ξ une martingale indexée par un quasi-intervalle T, telle que $\xi(b)$ soit L^1. Alors :*

- *soit il y a un nombre limité de points de T, $t_1, \cdots, t_n$, deux à deux discernables et tels que pour tout élément t de T, t est une discontinuité fixe de ξ si et seulement si t est équivalent à l'un des points de l'ensemble $\{t_1, \cdots, t_n\}$;*
- *soit il existe une suite réelle $(t_i)_{i \in N}$, dont les éléments d'indice limité sont deux à deux discernables, et telle que pour tout t de T, t est une discontinuité fixe de ξ si et seulement si il existe un indice limité tel que $t \approx t_i$.*

Démonstration. D'après la deuxième assertion du théorème 11.4, t est une discontinuité fixe de ξ si et seulement si il existe un naturel non nul standard k tel que t est une $\frac{1}{k}$-discontinuité.

Grâce au théorème 11.5, on montre qu'à tout naturel non nul standard k on peut associer un ensemble $E_k = \{t_{k1}, \cdots, t_{kn_k}\}$, avec $n_k \ll \infty$, tel que les éléments de E_k soient deux à deux discernables, et tel qu'être infiniment proche d'un élément (et d'un seul) de E_k soit équivalent à être une $\frac{1}{k}$-discontinuité de ξ sans être une $\frac{1}{m}$-discontinuité pour $m < k$.

Le principe d'extension nous garantit l'existence d'une suite $k \mapsto E_k$, telle que pour tout indice standard la propriété précédente soit vérifiée.

S'il existe un entier standard j tel que $\forall k > j, E_k = \emptyset$, construisons l'ensemble $\{t_1, \cdots, t_n\} = \bigcup_{k \leq j} E_k$. Cet ensemble vérifie les propriétés de la première assertion du théorème 11.6 que nous sommes en train de démontrer. Si un tel entier n'existe pas, soit j un entier non standard quelconque, et construisons la séquence $\{t_1, \cdots, t_\nu\} = \bigcup_{k \leq j} E_k$. Complétons cette séquence en une suite réelle en posant par exemple $t_i = 0$ pour tout $i > \nu$. La suite ainsi obtenue vérifie les propriétés de la deuxième assertion du théorème. □

On peut s'assurer que la deuxième partie du théorème 11.6 exprime bien le fait qu'il y a peu de discontinuités fixes. En effet :

Théorème 11.7. *Soit T un quasi-intervalle. Il n'existe pas de suite réelle telle que chaque élément de T soit équivalent à un élément d'indice standard de cette suite.*

Démonstration. Procédons par l'absurde en posant l'existence d'une telle suite $(t_i)_{i \in N}$. On peut supposer, sans restreindre la généralité du propos, que $a = 0$ et $b = 1$. Soit x un élément de $[0, 1]$, et soit t le plus grand élément de T' tel que $t \leq x$. Alors nous avons $t \leq x \leq t + dt$, si bien que x est infiniment proche de t et donc par hypothèse, d'une valeur t_i où i est standard. Or, le procédé de la diagonale de Cantor nous permet de construire un réel x appartenant au segment $[0, 1]$ tel que pour tout entier i non nul, la i-ième décimale de x soit différente de la i-ième décimale de t_i d'au moins 2. Un tel entier vérifie donc la proposition suivante :

$$\forall^{st} i \in N^*, |x - t_i| > 10^{-1} \gg 0$$

ce qui constitue une contradiction. □

Nous avons vu qu'en terme de "cardinalité" il y avait peu de discontinuités fixes discernables. Maintenant, montrons que les discontinuités fixes sont rares également en terme de "mesure".

***Théorème 11.8.** *Soit ξ une martingale indexée par un quasi-intervalle T, telle que $\xi(b)$ soit L^1, et soit une fonction $f : R^+ \to R^+$ telle que $f(h) \approx 0$ pour tout réel positif $h \approx 0$. Pour tout réel $\varepsilon \gg 0$, il existe un entier naturel ν et des intervalles $[r_i, s_i], i = 1, \cdots, \nu$, tels que $\sum_{i=1}^{\nu} f(s_i - r_i) \leq \varepsilon$, et tels que chaque discontinuité fixe de ξ réside dans l'un de ces intervalles.*

Démonstration. Soit un réel infinitésimal $\alpha > 0$. Pour tout entier naturel i, définissons le réel positif h_i comme étant le plus grand multiple entier de α inférieur à 1 et tel que $f(h_i) \leq \varepsilon/2^i$, ou 0 si un tel nombre n'existe pas. Alors, pour toute valeur standard de i, $h_i \gg 0$ et $f(h_i) \leq \varepsilon/2^i$. Soit (t_i) une séquence ou une suite infinie de réels, dont le théorème 11.6 a démontré l'existence. Si nous sommes dans le premier cas de ce théorème (il existe un nombre limité n de t_i), alors posons $\nu = n$. Si nous sommes dans le second cas, choisissons arbitrairement $\nu \approx \infty$. Maintenant, pour chaque i de $\{1, \cdots, \nu\}$, définissons les réels r_i et s_i par les relations $r_i = t_i - h_i/2$ et $s_i = t_i + h_i/2$. Alors, nous avons $\sum_{i=1}^{\nu} f(s_i - r_i) \leq \varepsilon$, et chaque discontinuité fixe de ξ est dans l'un des intervalles $[r_i, s_i]$. □

En particulier, soit ξ une martingale indexée par un quasi-intervalle T, telle que $\xi(b)$ soit L^1, et choisissons la fonction $f(h) = h$. Nous voyons alors que, pour la probabilité pr_T ("mesure de Lebesgue normalisée sur le segment $[a,b]$"), presque tout point t de T n'est pas une discontinuité fixe de ξ.

Pour clore ce chapitre, notons que si ξ est une martingale telle que $\xi(b)$ est L^1, alors $\xi(a) = E_a\xi(b)$ est également L^1, ainsi que $\xi(b) - \xi(a)$. Tous les théorèmes 11.4 jusqu'à 11.8 restent vrais sous l'hypothèse que $(\xi(b) - \xi(a))$ est L^1, car ξ ne se différencie que trivialement de $\xi - \xi(a)$. Dans le prochain chapitre nous étudierons les martingales en nous contentant de l'hypothèse plus faible selon laquelle $\|\xi(b) - \xi(a)\|_1 \ll \infty$.

Chapitre 12

Fluctuations des martingales

Lorsque ξ est une marche aléatoire de Wiener ou de Poisson,

$$E\, d\xi(r)^2 = dr$$

et de ce fait $\|\xi(t) - \xi(s)\|_2^2 = |t - s|$. Ainsi, dans ces deux cas, le processus est en tout point continu dans $\mathbf{L}^2$, et par conséquent dans $\mathbf{L}^1$. Nous pouvons en déduire, grâce au théorème 11.3 (deuxième assertion), que, pour tout point t de T, le processus est presque sûrement continu en t. C'est à dire qu'un processus appartenant à l'une de ces deux catégories n'admet pas de discontinuité fixe. Cela ne veut cependant pas dire que presque sûrement, les trajectoires de tels processus sont continues sur T, car il peut exister un ensemble de points exceptionnels différent à chaque instant t. Considérons une marche de Poisson ξ indexée par un quasi-intervalle. Comme ce type de processus ne prend que des valeurs entières, la seule possibilité pour une trajectoire d'être continue est de rester uniformément égale à sa valeur initiale $\xi(a)$. Or, la probabilité qu'un saut survienne à l'instant t est égale à dt. Comme les accroissements successifs du processus sont indépendants, on en déduit que la probabilité pour qu'une trajectoire ne présente pas de sauts est égale à $\prod_{t \in T'} (1 - dt)$. Or :

$$\log\left(\prod_{t \in T'} (1 - dt)\right) = \sum_{t \in T'} \log(1 - dt) \sim -\sum_{t \in T'} dt = -(b - a)$$

(on met en œuvre le théorème 5.3, joint au fait que $\log(1 - dt) \sim -dt$). Ainsi la probabilité pour qu'aucun saut ne survienne entre les instants a et b est équivalente à $e^{-(b-a)}$, qui est une valeur sensiblement inférieure à 1. Dans le prochain chapitre nous serons en mesure de montrer qu'une marche de Wiener, en revanche, est presque sûrement continue en tout point de T. La figure 12.1 montre un exemple typique de trajectoire pour chacun de ces deux types de processus.

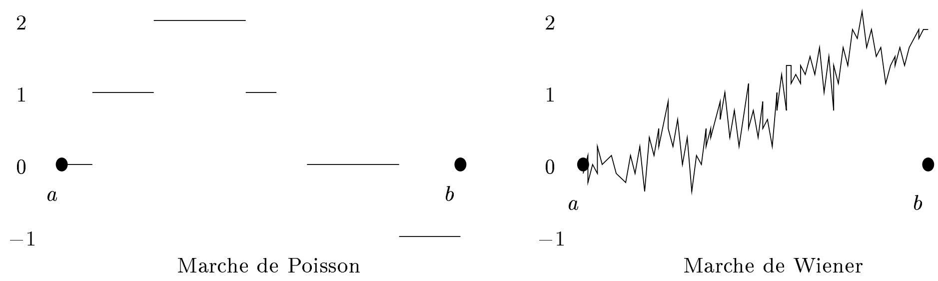

FIG. 12.1 – deux trajectoires

Ceci soulève la question de la qualification des discontinuités des trajectoires d'une martingale. Pour une fonction bornée, on distingue habituellement deux types de discontinuités : les discontinuités de type "saut", et les discontinuités de type "oscillations". La figure 12.2 illustre ces deux cas (il faut imaginer, en regardant la figure 12.2 (b), la présence d'une infinité d'oscillations au voisinage de l'origine, comme pour la fonction $x \mapsto \sin(1/x)$). La propriété d'être à fluctuation limitée que nous avons définie au chapitre 6 est un analogue externe, pour une fonction dont le domaine de définition est borné, de la propriété interne de ne présenter que des discontinuités de type "saut". Si on se reporte à la figure 12.2, (a) illustre une fonction à fluctuation limitée, et (b) une fonction à fluctuation illimitée.

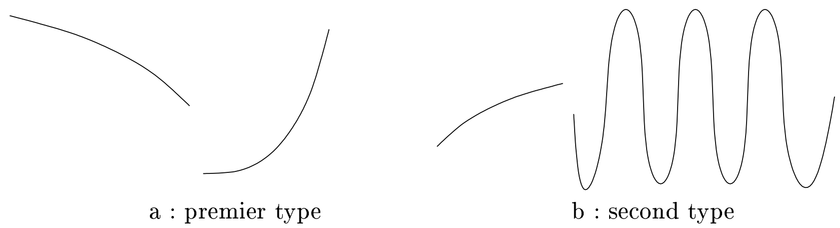

FIG. 12.2 – discontinuités

Le résultat principal de ce chapitre dit que, moyennant des conditions très peu restrictives, les trajectoires d'une surmartingale ou d'une sousmartingale sont presque sûrement à fluctuation limitée. Nous allons commencer par établir un lemme technique, puis nous mettrons en scène une stratégie d'investissement qui nous permettra d'estimer le nombre de fois qu'un même intervalle peut être franchi sur la durée du processus.

Théorème 12.1. *Soit ξ un processus stochastique indexé par T, un sous ensemble fini de R. ξ est presque sûrement à fluctuation limitée si et seulement si :*

$$\forall \varepsilon \gg 0, \forall k \approx \infty, Pr\{\xi \text{ admet } k\ \varepsilon\text{-fluctuations}\} \approx 0.$$

Démonstration. Il est évident que la condition est nécessaire. Inversement, supposons que la condition soit vérifiée. Définissons l'événement $M(\varepsilon, k)$ correspondant au fait que ξ admette k ε-fluctuations, et soit $\delta \gg 0$. Pour tout naturel n supérieur à 0, définissons également l'entier k_n comme étant le plus petit entier tel que

$$Pr\left\{M\left(\frac{1}{n}, k_n\right)\right\} \leq \frac{\delta}{2^n},$$

et soit l'ensemble N :

$$N = \bigcup_{n=1}^{\infty} M\left(\frac{1}{n}, k_n\right).$$

On a $PrN \leq \delta$. Par hypothèse lorsque n est limité k_n l'est également, et donc lorsque ω est dans N^c, $\xi(\omega)$ est à fluctuation limitée. Comme $\delta \gg 0$ est arbitraire, ceci conclut la démonstration. □

Un spéculateur espère réaliser de grosses plus-values en pariant sur une valeur très dynamique, dont il modélise le cours par le processus ξ. Il adopte le principe "acheter bas, vendre élevé". Pour cela, il attend le premier instant s_1 où le prix de l'action tombe en dessous d'un seuil λ_1, pour acquérir une action. Puis il attend le prochain instant $t_1 > s_1$ auquel la valeur de l'action dépasse un seuil λ_2 pour la revendre. Il répète ce procédé autant de fois que possible. A l'instant t, le gain du spéculateur vaut :

$$\zeta_3(t) = \sum_{s<t} \eta_3(s)\, d\xi(s)$$

où η_3 vaut 1 lorsque la personne est encore en possession de l'action (c'est à dire aux instants t tels qu'il existe un indice vérifiant $s_k \leq t$, mais pas $t_k \leq t$) et 0 aux autres instants. Le processus η_3 est un processus positif adapté. Le gain intermédiaire (avant l'instant b) du spéculateur vaut au moins $\beta\,(\lambda_2 - \lambda_1)$, où β désigne le nombre de franchissements dans le sens croissant de l'intervalle $[\lambda_1, \lambda_2]$ effectués jusque-là par le cours (β est la plus grande valeur de k pour laquelle t_k est défini). Le spéculateur peut enregistrer une perte en fin de partie, si le dernier instant s_k n'est pas suivi pas une dernière opportunité de vente t_k. Mais dans ce cas, la perte vaut au plus $(\lambda_1 - \xi(b))^+$, où nous désignons par x^+ la valeur $\max(0, x)$. Ainsi :

$$\zeta_3(b) \geq \beta\,(\lambda_2 - \lambda_1) - (\lambda_1 - \xi(b))^+. \tag{12.1}$$

Supposons maintenant que ξ soit une surmartingale, ce qui implique que $0 \geq E\zeta_3(b)$. Alors 12.1 nous permet d'écrire les inégalités suivantes :

$$(\lambda_2 - \lambda_1)E\beta \leq E(\lambda_1 - \xi(b))^+ \leq \|\xi(b)\|_1 + |\lambda_1|.$$

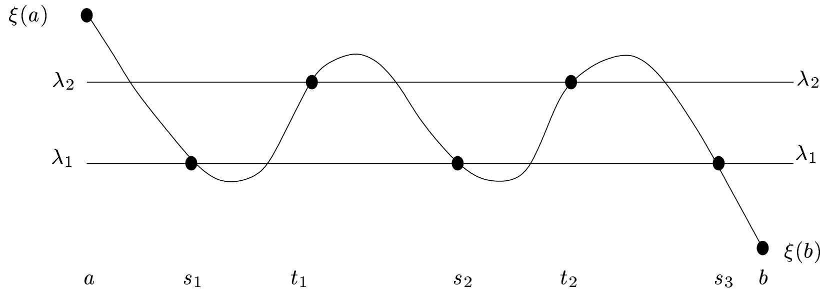

FIG. 12.3 – franchissements de $[\lambda_1, \lambda_2]$

Le scénario précédent nous a permit de démontrer le théorème suivant :

Théorème 12.2. *Soit ξ une surmartingale, soient $\lambda_1 < \lambda_2$, et soit β le nombre de franchissements croissants de l'intervalle $[\lambda_1, \lambda_2]$ effectués par ξ. Alors :*

$$E\beta \leq \frac{\|\xi(b)\|_1 + |\lambda_1|}{\lambda_2 - \lambda_1}.$$

Bien que nous ne ferons pas usage de cette constatation dans la suite de l'exposé, nous pouvons remarquer que le théorème 12.2 reste valable si ξ est une sousmartingale. Pour s'en convaincre, imaginons le cas d'un investisseur qui veut se constituer un portefeuille d'actions solides, et qui pour cela suit une stratégie rigoureusement opposée à celle du spéculateur de notre précédent scénario. Ainsi, il achète une action (dont le cours est ici encore modélisé par le processus ξ) à l'instant a, à moins que la valeur de cette action soit alors inférieure à λ_1 (il achète si $\xi(a) > \lambda_1$). A l'instant s_1, il vend son action en acceptant sa perte pour des raisons fiscales. Puis il attend l'instant t_1 pour acheter à nouveau, et ainsi de suite. A l'instant t son gain est donné par la relation suivante :

$$\zeta_4(t) = \sum_{s<t} \eta_4(s)\, d\xi(s)$$

avec $\eta_4 = 1 - \eta_3$. Le processus positif η_4 est adapté également. Alors l'inégalité suivante est vérifiée :

$$\zeta_4(b) \leq (\xi(b) - \lambda_1)^+ - \beta(\lambda_2 - \lambda_1). \tag{12.2}$$

Pour montrer cette inégalité, supposons d'abord que $\xi(a) \geq \lambda_1$. Il commence par acheter au prix $\xi(a)$, et perd $(\lambda_2 - \lambda_1)$ à chaque croisement de l'intervalle $[\lambda_2 - \lambda_1]$ dans le sens croissant. Donc :

$$\begin{aligned} \zeta_4(b) &\leq \xi(b) - \xi(a) - \beta(\lambda_2 - \lambda_1) \\ &\leq (\xi(b) - \lambda_1)^+ - \beta(\lambda_2 - \lambda_1). \end{aligned}$$

Si maintenant $\xi(a) < \lambda_1$ et $\beta = 0$, il n'effectue aucun achat, et $\zeta_4(b) = 0$ ce qui rend 12.2 trivialement vraie. Enfin, si $\xi(a) < \lambda_1$ et $\beta > 0$, il effectue son premier achat à un prix supérieur à λ_2 juste après le premier franchissement de l'intervalle $[\lambda_2 - \lambda_1]$ dans le sens croissant, si bien que :

$$\begin{aligned} \zeta_4(b) &\leq (\xi(b) - \lambda_2) - (\beta - 1)(\lambda_2 - \lambda_1) \\ &\leq (\xi(b) - \lambda_1)^+ - \beta(\lambda_2 - \lambda_1). \end{aligned}$$

Ainsi 12.2 est vérifiée dans tous les cas, et le théorème (12.2) vaut également pour les sousmartingales (du fait que $0 \leq E\zeta_4(b)$).

Théorème 12.3. *Soit ξ une surmartingale ou une sousmartingale telle que $\|\xi(b) - \xi(a)\|_1 \ll \infty$. Alors, ξ est presque partout à fluctuation limitée.*

Démonstration. Du fait que ξ est une surmartingale si et seulement si $(-\xi)$ est une sousmartingale, il nous suffit de démontrer le théorème dans le seul cas des surmartingales. On ne perd pas en généralité en supposant de plus que $\xi(a) = 0$. Soit un réel $\delta \gg 0$, et soit $\lambda = \frac{2\|\xi(b)\|_1}{\delta} \ll \infty$. Le théorème 11.1 nous permet d'écrire :

$$Pr\left\{\max_{t \in T} |\xi(t)| \geq \lambda\right\} \leq \frac{2\|\xi(b)\|_1}{\lambda} \leq \delta. \tag{12.3}$$

Soit un réel $\varepsilon \gg 0$ tel que $n = \frac{\lambda}{\varepsilon}$ soit un entier (forcément standard). Effectuons la partition de l'intervalle $[-\lambda, \lambda]$ en $2n$ intervalles de largeurs égales à ε.

Supposons que $\max_{t \in T} |\xi(t)| < \lambda$, et que le processus ξ admette $(2n + 2k)$ 2ε-fluctuations, où k est un multiple illimité de $2n$. Chaque 2ε-fluctuation donne lieu au franchissement de l'un des $2n$ sous-intervalles de $[-\lambda, \lambda]$, si bien qu'il existe un sous-intervalle pour lequel le nombre de franchissements est égal à $\left(1+\frac{k}{n}\right)$. Soit β le nombre de franchissements de cet intervalle dans le sens croissant. Comme les nombres de franchissements dans le sens croissant et dans le sens décroissant diffèrent au plus de 1, il vient que $\beta \geq \frac{k}{2n}$. Or, d'après l'inégalité de Bienaymé-Tchebychev et le théorème 12.2, on a les inégalités suivantes :

$$Pr\left\{\beta \geq \frac{k}{2n}\right\} \leq \frac{2n}{k}E\beta \leq \frac{n}{k\varepsilon}(\|\xi(b)\|_1 + \lambda).$$

Comme nous avons créé $2n$ sous-intervalles, la probabilité pour que

$$\max_{t\in T}|\xi(t)| < \lambda$$

et que ξ admette $(2n+2k)$ 2ε-fluctuations est majorée par

$$\frac{2n^2}{k\varepsilon}(\|\xi(b)\|_1 + \lambda),$$

qui est infinitésimal. En rappelant l'inégalité 12.3, on obtient donc la relation suivante :

$$Pr\{\xi \text{ admet } (2n+2k) \ 2\varepsilon\text{-fluctuations}\} \lesssim \delta.$$

Comme le réel $\delta \gg 0$ est arbitraire, le théorème (12.1) permet de conclure la démonstration. $\square$

Chapitre 13

Discontinuités des martingales

Nous avons vu précédemment que, moyennant des hypothèses assez peu contraignantes, presque toutes les trajectoires d'une surmartingale ou d'une sousmartingale sont à fluctuation limitée. Ceci nous renseigne quelque peu sur la nature des discontinuités que l'on peut observer sur les trajectoires de ces processus, mais nous serions heureux d'en savoir davantage. Ainsi, soit un réel $\varepsilon \gg 0$; si les bonnes hypothèses sont vérifiées ($\|\xi(b) - \xi(a)\|_1$ est limité), pour presque toutes les trajectoires le nombre de points tels que $d\xi(t) \geq \varepsilon$ est limité. Est-il possible d'observer, sur une même trajectoire, deux points de ce type infiniment proches l'un de l'autre ? Peut-on trouver, pendant un laps de temps infinitésimal, une séquence illimitée d'accroissements infinitésimaux telle que l'effet cumulé de chacun d'eux produise un accroissement en valeur absolue supérieur à ε, si bien que l'on aurait une discontinuité alors même que chaque accroissement est infinitésimal ? Une trajectoire peut-elle être discontinue à la fois à droite et à gauche en un point t ? En bref, est-il possible de trouver 2 ε-fluctuations au sein d'un même intervalle de largeur infinitésimale ? De tels phénomènes sont généralement possibles pour une fonction à fluctuation limitée. Cependant, nous allons voir que si ξ est une surmartingale ou une sousmartingale indexée par T, avec $(b - a)$ appréciable, et telle que σ_ξ^2 est un processus L^1 sur l'espace $T' \times \Omega$, alors pour presque toutes les trajectoires de ξ, de tels phénomènes ne se produisent pas. Souvenons nous (théorème 8.4 et son inverse) que pour que σ_ξ^2 soit un processus L^1 sur $T' \times \Omega$, il est suffisant que chaque variable $\sigma_\xi^2(t)$ soit L^1 sur Ω.

Soit ξ un processus stochastique indexé par T, un sous-ensemble fini de R. Le *temps propre* de ξ est le processus τ_ξ défini ainsi :

$$\forall t \in T, \tau_\xi(t) \stackrel{def}{=} \sum_{s<t} \sigma_\xi^2(s)\, ds.$$

Le temps propre est un processus croissant, mais sa manière de croître varie d'une trajectoire à l'autre. Pour $s \leq t$, nous appelons la quantité positive $(\tau_\xi(t) - \tau_\xi(s))$ la *durée propre* de l'intervalle $[s,t]$.

Supposons que nous modélisions le prix d'une action d'une société par une martingale ξ telle que $\|\xi(b) - \xi(a)\|_1 \ll \infty$. Un spéculateur veut éviter d'immobiliser son capital pendant une période trop longue (ce qui est une notion subjective et en rapport avec le temps propre du processus, car le temps d'immobilisation lui semblera d'autant plus long que le marché subira des fluctuations rapides). Il a de plus pour règle d'acheter lorsque les valeurs montrent des signes nets d'activité.

Pour agir, il choisit un réel $\varepsilon \gg 0$ et un réel $\alpha \approx 0$ positif.[1] Posons $s_0 = a$, et soit s_1 le premier instant postérieur à s_0 tel que $|\xi(s_1) - \xi(s_0)| \geq \varepsilon/4$, s_2 le premier instant postérieur à s_1 tel que $|\xi(s_2) - \xi(s_1)| \geq \varepsilon/4$ et ainsi de suite.[2] Notre investisseur choisit *a priori* un entier j, et achète une action à l'instant s_j. Il revend cette action soit au premier instant s postérieur à s_j tel que $\tau_\xi(s + ds) - \tau_\xi(s_j) > \alpha$, soit à l'instant s_{j+1}, si $\tau_\xi(s_{j+1}) - \tau_\xi(s_j) \leq \alpha$. Ensuite, aucun achat n'est plus effectué. Alors, soit $\zeta_j(t)$ le gain apporté à l'instant t par cette stratégie :

$$\zeta_j(t) = \sum_{s<t} \eta_j(s)\, d\xi(s).$$

Comme $\tau_\xi(s + ds) = \tau_\xi(s) + \sigma_\xi^2(s)\, ds$, il vient que $\tau_\xi(s + ds) \in \Phi_s$, ce qui fait que la variable $\eta_j(s)$ appartient à Φ_s. Le processus η est donc un Φ-processus. Nous avons :

$$\begin{aligned}
\|\zeta_j(b)\|_2^2 &= E\left(\sum_{t\in T'} \eta_j(t)\, d\xi(t)\right)^2 \\
&= E\left(\sum_{t\in T'} \eta_j(t)^2 \sigma_\xi^2(t)\, dt\right) \\
&\leq \alpha \approx 0
\end{aligned}$$

et, suivant l'inégalité de Bienaymé-Tchebychev, on peut donc affirmer que $\zeta_j(b) \underset{ps}{\approx} 0$.

[1] Il ne s'agit là que d'une expérience de pensée pour établir un résultat général sur les processus stochastiques, et en aucun cas d'un exposé sur une pratique financière. Il est impossible pratiquement de choisir une valeur infinitésimale. (NdT)

[2] L'ensemble des instants s_i dépend donc de la trajectoire qu'observe l'investisseur. (NdT)

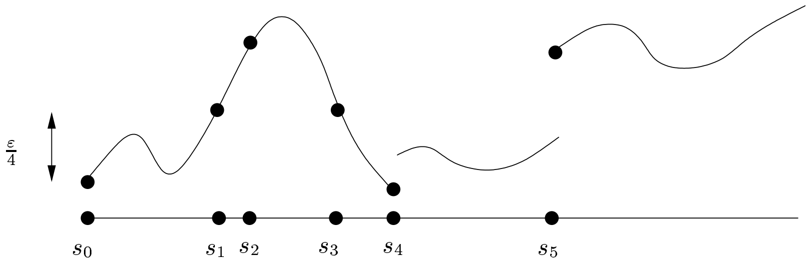

FIG. 13.1 – une autre stratégie d'investissement

Les hypothèses que nous avons posées sur ξ et le théorème 12.3 nous disent que, presque sûrement, le cour de l'action que constate le spéculateur est à fluctuation limitée, c'est à dire que presque sûrement l'ensemble des instants s_i est de cardinal limité. Nous pouvons donc également affirmer que :

$$\max_j |\zeta_j(b)| \underset{ps}{\approx} 0. \tag{13.1}$$

Soit $M(\varepsilon, \alpha)$ l'événement qui marque la présence d'au moins 2 ε-fluctuations au sein d'un même intervalle de durée propre inférieure ou égale à α. Notons que si $t_1 \leq t_2 \leq t_3 \leq t_4$ vérifient $|\xi(t_2) - \xi(t_1)| \geq \varepsilon$ et $|\xi(t_4) - \xi(t_3)| \geq \varepsilon$, alors il y a au moins deux instants s_j distincts dans l'intervalle $[t_1, t_4]$ car on doit en trouver au moins un dans $[t_1, t_2]$, et au moins un dans $[t_3, t_4]$. Par conséquent, si l'événement $M(\varepsilon, \alpha)$ se réalise il existe au moins un indice j vérifiant l'inégalité $\tau_\xi(s_{j+1}) - \tau_\xi(s_j) \leq \alpha$, et de ce fait $\max_j |\zeta_j(b)| \geq \varepsilon/4$. Avec la relation 13.1, ceci nous montre que si $\varepsilon \gg 0$ et si $\alpha \approx 0$, alors $PrM(\varepsilon, \alpha) \approx 0$.

Soit un réel $\delta \gg 0$. Pour tout naturel non nul n, définissons le réel α_n comme étant le plus grand élément de l'ensemble $\{|\tau_\xi(t) - \tau_\xi(s)| : s \in T, t \in T\}$ tel que :

$$PrM\left(\frac{1}{n}, \alpha_n\right) \leq \frac{\delta}{2^n}.$$

Alors, si n est limité, α_n est sensiblement supérieur à 0 ($\alpha_n \gg 0$).[3]

Posons $N = \bigcup_{n \in N^*} M\left(\frac{1}{n}, \alpha_n\right)$. Par construction, $PrN \leq \delta$. Soit $\omega \in N^c$. Pour tout naturel n limité, la trajectoire $\xi(., \omega)$ ne présente pas deux $\frac{1}{n}$-fluctuations au sein d'un même intervalle de durée propre infinitésimale. Étant δ un réel sensiblement positif quelconque, nous venons de démontrer le théorème suivant :

[3]Ou bien $\alpha_n = \mu = \max\{|\tau_\xi(t) - \tau_\xi(s)| : s \in T, t \in T\}$, dans le cas où μ est infinitésimal. Ceci ne change pas la suite de la démonstration. (NdT)

Théorème 13.1. *Soit ξ une martingale telle que $\|\xi(b)-\xi(a)\|_1 \ll \infty$. Alors, presque sûrement, quel que soit le réel $\varepsilon \gg 0$, on ne peut pas trouver deux ε-fluctuations dans un même intervalle de durée propre infinitésimale.*

On constate que si ξ est une martingale normalisée, son temps propre est trivialement équivalent à son temps réel :

$$\forall t \in T, \tau_\xi(t) = t - a,$$

et l'on a de plus :

$$\|\xi(b) - \xi(a)\|_1 \leq \|\xi(b) - \xi(a)\|_2 = \sqrt{b-a}$$

ce qui fait que pour avoir $\|\xi(b) - \xi(a)\|_1 \ll \infty$, il suffit que $|b-a| \ll \infty$. Aussi, sous ces conditions (ξ est une martingale normalisée, et $|b-a| \ll \infty$), les conclusions du théorème (13.1) sont vérifiées. Si de plus en chaque point de $T' \times \Omega$ le processus $d\xi$ est infinitésimal, presque toutes les trajectoires de ξ sont continues sur toutes leurs durées. Ceci vient du fait que si t_1 et t_2 sont deux instants tels que $|\xi(t_2) - \xi(t_1)| \gg 0$, il existe nécessairement un instant t dans $]t_1, t_2[$ tel que $|\xi(t_1) - \xi(t)| \gg 0$ et $|\xi(t_2) - \xi(t)| \gg 0$.[4] Le théorème (13.1) admet donc le corollaire suivant :

Corollaire. *Presque toutes les trajectoires d'une marche aléatoire de Wiener sont continues sur toute leur durée.*

Nous voulons maintenant traduire la donnée relative aux durées propres, telle qu'on la voit apparaître dans le théorème (13.1), en une donnée relative aux durées ordinaires.

Théorème 13.2. *Soit ξ un processus stochastique tel que $(b-a)$ soit appréciable, et tel que σ_ξ^2 soit L^1 sur $T' \times \Omega$. Alors, presque sûrement, le temps propre τ_ξ est absolument continu. En particulier, presque sûrement, tout intervalle de largeur infinitésimale est de durée propre infinitésimale.*

Démonstration. Il vient du théorème de Fubini que presque toutes les trajectoires $t \mapsto \sigma_\xi^2(t)$ sont L^1 sur T'. Par le théorème de Radon-Nykodim, ceci implique que presque sûrement τ_ξ est absolument continu (en effet, $\frac{d\tau_\xi}{dt} = \sigma_\xi^2$ et $(b-a) \gg 0$). La conclusion du théorème vient simplement du fait qu'une fonction absolument continue est partout continue. □

[4] Et si $t_1 \approx t_2$ on observerait deux ε-fluctuations ($\varepsilon \gg 0$) au sein d'un intervalle de durée propre infinitésimale. On sait que les trajectoires qui ont cette particularité sont rares. (NdT)

Soit T un sous-ensemble fini de R, et une fonction $\xi : T \to R$. Pour tout point s de T', nous dirons que s est un *ε-saut* de ξ lorsque $|d\xi(s)| \geq \varepsilon$, et nous dirons que s est un *saut* de ξ lorsqu'il existe un réel $\varepsilon \gg 0$ tel que s soit un ε-saut de ξ. Si s est un saut de ξ, avec $ds \approx 0$, et si $t \approx s$, alors t est une discontinuité de ξ. Nous dirons que t est une *discontinuités de type "saut"* de ξ s'il existe un unique saut de ξ, s, tel que $t \approx s$.[5]

Théorème 13.3. *Soit T un sous-ensemble fini de R, et soit une fonction $\xi : T \to R$ telle que, quel que soit le réel $\varepsilon \gg 0$, il n'existe pas deux ε-fluctuations au sein d'un même intervalle de largeur infinitésimale. Alors :*
- *si t est une discontinuité de ξ, t est une discontinuité de type saut ;*
- *si s est le saut de ξ tel que $s \approx t$, alors ξ est continue a gauche en t si et seulement si $t \leq s$, et ξ est continue à droite en t si et seulement si $s < t$;*
- *si tous les accroissements de ξ sont infinitésimaux, ξ est partout continue ;*
- *on ne peut pas trouver deux sauts de ξ infiniment proches l'un de l'autre ;*
- *si $(b - a) \ll \infty$, alors quel que soit le réel $\varepsilon \gg 0$, il n'y a qu'un nombre limité de ε-sauts ;*

Démonstration. Soit t une discontinuité de ξ. C'est à dire que l'on peut trouver t_1 et t_2 équivalents à t, avec $t_1 < t_2$, tels que $|\xi(t_2) - \xi(t_1)| \geq \delta$ pour un certain réel $\delta \gg 0$. Soit s le premier instant postérieur ou égal à t_1, tel que $|\xi(s + ds) - \xi(t_1)| \geq \delta/2$. Les hypothèses que nous avons émises sur ξ imposent que $|\xi(s + ds) - \xi(t_1)| \gtrsim \delta$, et que $|\xi(s) - \xi(t_1)| \approx 0$. De ce fait, $|\xi(s + ds) - \xi(s)| \gtrsim \delta$, et nous voyons que s est un $\frac{\delta}{2}$-saut. Donc ξ admet un saut s équivalent à t, et tel que $ds \leq (t_2 - t_1) \approx 0$, ce qui fait de t une discontinuité de type saut. Les assertions suivantes coulent de source. $\square$

Les théorèmes (13.1) à (13.3) nous fournissent le résultat suivant :

Théorème 13.4. *Soit ξ une martingale telle que $(b - a)$ soit appréciable, et telle que σ_ξ^2 soit L^1 sur $T' \times \Omega$. Les conclusions du théorème (13.3) sont valables pour presque toutes les trajectoires de ξ.* [6]

En mettant en œuvre le principe d'extension généralisé, on montrerait également sous ces hypothèses que presque sûrement il existe une séquence de points (s_i) de T, telle que s est un saut de ξ si est seulement si $s = s_i$ pour un indice i standard.

[5] Et ds infinitésimal. (NdT)

[6] Si σ_ξ^2 est L^1 sur $T' \times \Omega$, alors $\|\xi(b) - \xi(a)\|_1 \ll \infty$. (NdT)

Chapitre 14

La condition de Lindeberg

Nous dirons qu'un processus stochastique ξ remplit la *condition de Lindeberg (approchée)* s'il vérifie la propriété suivante :

$$\forall \varepsilon \gg 0, E \sum_{t \in T'} d\xi(t)^2 \approx E \sum_{t \in T'} d\xi(t)^{(\varepsilon)^2}. \tag{14.1}$$

Dans le cas d'une martingale normalisée, la partie gauche de cette relation est égale à $(b-a)$. Une marche aléatoire de Wiener remplit toujours la condition de Lindeberg, car les deux membres de l'égalité sont rigoureusement égaux. En revanche, une marche aléatoire de Poisson ne la remplit pas, car dès que $\varepsilon < 1$, le membre de droite est nul.

Il est possible de construire une martingale normalisée de la manière suivante. Considérons une séquence $x_1, \cdots, x_\nu$ de v.a.r. indépendantes ($\nu \approx \infty$), de moyennes nulles, et vérifiant $\forall k \in [1, \nu], 0 < Ex_k^2 \ll \infty$. Posons :

$$\forall n \in [1, \nu], s_n = \sqrt{\sum_{k=1}^{n} Ex_k^2} \ ,$$

et admettons que $s_\nu \approx \infty$. Soit T l'ensemble suivant :

$$T = \left\{ \frac{s_n^2}{s_\nu^2} : n \in [1, \nu] \right\}.$$

L'ensemble T est clairement un quasi-intervalle, avec $a = 0$, et $b = 1$. Définissons maintenant le processus ξ indexé par T ainsi :

$$\forall t \in T, \left(t = \frac{s_n^2}{s_\nu^2} \right) \Longrightarrow \left(\xi(t) = \frac{1}{s_\nu} \sum_{k=1}^{n} x_k \right).$$

Alors, ξ est une martingale[1] normalisée, que nous appellerons la martingale normalisée *associée* à la séquence $x_1, \cdots, x_\nu$. Cette martingale satisfait la condition de Lindeberg approchée si et seulement si :

$$\forall \varepsilon \gg 0, \frac{1}{s_\nu^2} \sum_{k=1}^{\nu} E x_k^{(\varepsilon s_\nu)^2} \approx 1. \tag{14.2}$$

La propriété 14.2 est très similaire à la forme habituelle de la condition de Lindeberg,[2] mais l'expression 14.1 portant sur la martingale associée est plus parlante.

Théorème 14.1. *Soit ξ un processus stochastique satisfaisant la condition de Lindeberg. Alors, sur presque toutes les trajectoires, chaque accroissement $d\xi(t)$ est infinitésimal.*

Démonstration. Soit $\varepsilon \gg 0$. On a :

$$\begin{aligned} 0 \approx E \sum_{t \in T'} \left(d\xi(t)^2 - d\xi(t)^{(\varepsilon)^2} \right) &= \sum_{t \in T'} \sum_{|\lambda| > \varepsilon} \lambda^2 pr_{d\xi(t)}(\lambda) \\ \geq \varepsilon^2 \sum_{t \in T'} Pr\{|d\xi(t)| > \varepsilon\} &\geq \varepsilon^2 Pr\left\{ \max_{t \in T'} |d\xi(t)| > \varepsilon \right\}. \end{aligned}$$

Comme $\varepsilon^2 \gg 0$, il s'ensuit que $Pr\{\max_{t \in T'} |d\xi(t)| > \varepsilon\} \approx 0$, et ε étant arbitraire, le théorème 7.1 nous permet de conclure que $\max_{t \in T'} |d\xi(t)| \underset{ps}{\approx} 0$. $\square$

Théorème 14.2. *Soit ξ une martingale indexée par T tel que $(b-a)$ soit appréciable, telle que σ_ξ^2 soit L^1 sur $T' \times \Omega$, et vérifiant de plus la condition de Lindeberg. Alors presque toutes les trajectoires de ξ sont continues.*

---

[1] Il faut se souvenir (chapitre 9) qu'une somme d'accroissements indépendants et de moyennes nulles forme une martingale. (NdT)

[2] La version classique de la condition de Lindeberg est la suivante : soit un n-uplet $(x_1, \cdots, x_n)$ de v.a.r. de moyennes nulles et de variances σ_i^2 finies. Soit $s_n = \sum_{i=1}^n \sigma_i^2$ la variance de la variable $S_n = \sum_{i=1}^n x_i$. Posons :

$$\Lambda_n(\varepsilon) = \sum_{k=1}^{n} E\left(\left(\frac{x_k}{s_n} \right)^2 \chi_{\{|x_k/s_n| > \varepsilon\}} \right).$$

La condition de Lindeberg exige :

$$\forall \varepsilon > 0, \lim_{n \to \infty} \Lambda_n(\varepsilon) = 0.$$

(NdT)

Démonstration. Ce résultat découle immédiatement des théorèmes 14.1 et 13.4. □

Nous définissons la *variation quadratique* q_ξ d'une fonction $\xi : T \to R$ de la manière suivante :

$$q_\xi(t) \stackrel{def}{=} \sum_{s<t} d\xi(s)^2.$$

Lorsque T est un quasi-intervalle et que ξ est une fonction suffisamment régulière, la variation quadratique de ξ est infinitésimale. En revanche, si ξ est une martingale, on a $Eq_\xi(t) = \|\xi(t) - \xi(a)\|_2^2$, si bien qu'il faut s'attendre dans ce cas à voir des trajectoires de variation quadratique appréciable.

Souvenons nous des exemples de trajectoires typiques des marches aléatoires de Poisson et de Wiener, illustrés par la figures 12.1. Nous avons représenté sur la figure 14.1 les variations quadratiques correspondantes. On remarque que dans le cas de la marche aléatoire de Wiener, mais pas dans le cas de celle de Poisson, la variation quadratique est indépendante de la trajectoire. Afin de comprendre la connexion entre ce phénomène et la condition de Lindeberg nous allons commencer par établir un lemme de troncature similaire au théorème (10.3).

Théorème 14.3. *Soit ξ une martingale qui remplit la condition de Lindeberg. Pour $\alpha > 0$, on définit la martingale ξ_α ainsi :*

$$\begin{array}{rcl} \xi_\alpha(a) & = & \xi(a), \\ \forall t \in T', d\xi_\alpha(t) & = & d\xi(t)^{(\alpha)} - E_t\Big(d\xi(t)^{(\alpha)}\Big), \end{array}$$

de manière que $\forall (t,\omega) \in T' \times \Omega, |d\xi_\alpha(t,\omega)| \leq 2\alpha$. Alors, il existe un réel infinitésimal α tel que presque toutes les trajectoires possèdent les propriétés suivantes :

$$\forall t \in T, \left\{ \begin{array}{rcl} \xi_\alpha(t) & \approx & \xi(t); \\ q_{\xi_\alpha}(t) & \approx & q_\xi(t); \\ \tau_{\xi_\alpha}(t) & \approx & \tau_\xi(t). \end{array} \right.$$

Démonstration. Nous savons que $\max_{t\in T'} |d\xi(t)| \underset{ps}{\approx} 0$, grâce au théorème (14.1). Le théorème 7.1 nous garantit que, pour un réel infinitésimal α suffisamment grand, $Pr\{\max_{t\in T'} |d\xi(t)| \geq \alpha\} \approx 0$. De ce fait nous pouvons écrire :

$$ps : \forall t \in T, \xi(t) = \xi(a) + \sum_{s<t} d\xi(s)^{(\alpha)}.$$

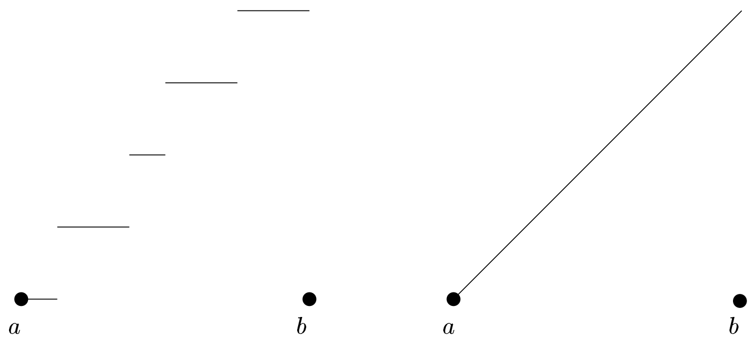

FIG. 14.1 – variation quadratique

Il s'ensuit que pour montrer que $\forall t \in T, \xi_\alpha(t) \approx \xi(t)$ pour presque toutes les trajectoires, il suffit de montrer que $\sum_{t\in T'} \left| E_t\, d\xi(t)^{(\alpha)} \right| \underset{ps}{\approx} 0$, ce qui revient à démontrer que $E \sum_{t\in T'} \left| E_t\, d\xi(t)^{(\alpha)} \right| \approx 0$.

Nous avons, du fait que $E_t\, d\xi(t) = 0$:

$$\begin{aligned} E \sum_{t\in T'} \left| E_t\, d\xi(t)^{(\alpha)} \right| &= E \sum_{t\in T'} \left| E_t \Big(d\xi(t) - d\xi(t)^{(\alpha)} \Big) \right| \\ &\leq E \sum_{t\in T'} E_t \left| d\xi(t) - d\xi(t)^{(\alpha)} \right| \\ &\leq \frac{1}{\alpha} E \sum_{t\in T'} \Big(d\xi(t)^2 - d\xi(t)^{(\alpha)^2} \Big) \end{aligned}$$

En effet, soit $|d\xi(t)| \leq \alpha$ et dans ce cas :

$$0 = \left| d\xi(t) - d\xi(t)^{(\alpha)} \right| = \frac{1}{\alpha} \Big(d\xi(t)^2 - d\xi(t)^{(\alpha)^2} \Big),$$

soit $|d\xi(t)| > \alpha$ et dans ce cas :

$$\left| d\xi(t) - d\xi(t)^{(\alpha)} \right| = |d\xi(t)| \leq \frac{1}{\alpha} |d\xi(t)|^2 = \frac{1}{\alpha} \Big(d\xi(t)^2 - d\xi(t)^{(\alpha)^2} \Big).$$

Si α est appréciable ($\alpha \gg 0$), $\frac{1}{\alpha} E \sum_{t\in T'} \Big(d\xi(t)^2 - d\xi(t)^{(\alpha)^2} \Big)$ est infinitésimal du fait de la condition de Lindeberg. Donc l'ensemble des réels :

$$\left\{ x : E \sum_{t\in T'} \left| E_t\, d\xi(t)^{(x)} \right| \leq x \right\}$$

contient tous les réels positifs appréciables. Il contient donc un réel infinitésimal α que nous pouvons choisir suffisamment grand (principe de Cauchy).

On voit également que $\sum_{t\in T'}\left(d\xi(t)^2 - d\xi(t)^{(\alpha)^2}\right) \underset{ps}{\approx} 0.$[3] Or $d\xi_\alpha(t)^2$ diffère de $d\xi(t)^{(\alpha)2}$ par la quantité $\left(E_t d\xi(t)^{(\alpha)}\right)^2 - 2d\xi(t)^{(\alpha)}E_t\, d\xi(t)^{(\alpha)}$, qui est en valeur absolue inférieure à $3\alpha\left|E_t\, d\xi(t)^{(\alpha)}\right|$. Puisque nous avons montré que $\sum_{t\in T'}\left|E_t\, d\xi(t)^{(\alpha)}\right| \approx 0$, il vient que $ps : \forall t \in T, q_\xi(t) \approx q_{\xi_\alpha}(t)$.

Enfin, en suivant le même raisonnement après avoir appliqué l'opérateur E_t à chaque membre, on voit facilement que $ps : \forall t \in T, \tau_\xi(t) \approx \tau_{\xi_\alpha}(t)$. □

Théorème 14.4. *Soit ξ une martingale respectant la condition de Lindeberg, telle que $\|\xi(b) - \xi(a)\|_2 \ll \infty$. Alors, presque sûrement :*

$$\forall t \in T, \tau_\xi(t) \approx q_\xi(t).$$

Démonstration. Le théorème (14.3) nous montre que l'on peut supposer l'existence d'un nombre réel infinitésimal α tel que :

$$\forall (t,\omega) \in T' \times \Omega, |d\xi(t,\omega)| \le \alpha.$$

De plus, on remarque que $(q_\xi - \tau_\xi)$ est une martingale dont la valeur en a est nulle. Ainsi nous pouvons écrire :

$$\begin{aligned}\|q_\xi(b) - \tau_\xi(b)\|_1^2 \le \|q_\xi(b) - \tau_\xi(b)\|_2^2 &= \sum_{t\in T'} \|dq_\xi(t) - d\tau_\xi(t)\|_2^2 \\ &= \sum_{t\in T'} \left\|d\xi(t)^2 - E_t\, d\xi(t)^2\right\|_2^2 \le \sum_{t\in T'} \left\|d\xi(t)^2\right\|_2^2 = \sum_{t\in T'} E\, d\xi(t)^4 \\ &\le \alpha^2 \sum_{t\in T'} \|d\xi(t)\|_2^2 = \alpha^2 \|\xi(b) - \xi(a)\|_2^2 \approx 0.\end{aligned}$$

Alors le théorème (11.1) nous permet d'écrire :

$$\forall \lambda \gg 0, Pr\left\{\max_{t\in T} |q_\xi(t) - \tau_\xi(t)| \ge \lambda\right\} \approx 0,$$

et le théorème (7.1) nous permet de conclure :

$$\max_{t\in T} |q_\xi(t) - \tau_\xi(t)| \underset{ps}{\approx} 0.$$

□

[3] $E\left|\sum_{t\in T'}\left(d\xi(t)^2 - d\xi(t)^{(\alpha)^2}\right)\right| = E\sum_{t\in T'}\left(d\xi(t)^2 - d\xi(t)^{(\alpha)^2}\right) \approx 0$ par la condition de Lindeberg. (NdT)

Chapitre 15

Maximum des martingales

Commençons par énoncer un théorème.

Théorème 15.1. *Soit ξ une sousmartingale. Alors, ξ vérifie les relations suivantes :*

$$\lambda Pr\{\max \xi(t) \geq \lambda\} \leq E\xi(b)\chi_{\{\max \xi(t) \geq \lambda\}} \leq \|\xi(b)\|_1. \tag{15.1}$$

Démonstration. Souvenons nous de la démonstration du théorème (11.1). Si ξ est une sousmartingale, alors la relation 11.3 nous donne :

$$0 \leq E(\xi(b) - \xi(a) - \lambda)\chi_{\{\max (\xi(t) - \xi(a)) \geq \lambda\}}.$$

Nous pouvons relativiser cette relation à l'algèbre Φ_a en remplaçant E par E_a, et λ par n'importe quel élément de Φ_a. En particulier, si nous remplaçons λ par $(\lambda - \xi(a))$, nous obtenons :

$$0 \leq E_a(\xi(b) - \lambda)\chi_{\{\max \xi(t) \geq \lambda\}}$$

ce qui donne, en calculant l'espérance absolue :

$$\lambda Pr\{\max \xi(t) \geq \lambda\} \leq E\xi(b)\chi_{\{\max \xi(t) \geq \lambda\}} \leq \|\xi(b)\|_1.$$

□

On peut suggérer une deuxième démonstration pour montrer ce théorème. Posons :

$$\begin{aligned} \Lambda &= \{\omega \in \Omega : \max_{t \in T} \xi(t, \omega) \geq \lambda\}; \\ \Lambda_t &= \{\omega \in \Omega : (\xi(t, \omega) \geq \lambda) \text{ et } (\forall s < t, \xi(s, \omega) < \lambda)\}. \end{aligned}$$

Alors, la famille $\{\Lambda_t : t \in T\}$ forme une partition de Λ, et $\chi_{\Lambda_t} \in \Phi_t$. Nous pouvons écrire :

$$\begin{aligned}\|\xi(b)\|_1 &\geq E\xi(b)\chi_\Lambda = E\sum_{t\in T}\xi(b)\chi_{\Lambda_t} = E\sum_{t\in T}E_t(\xi(b)\chi_{\Lambda_t})\\ &= E\sum_{t\in T}\chi_{\Lambda_t}E_t\xi(b) \geq E\sum_{t\in T}\chi_{\Lambda_t}\xi(t) \geq E\sum_{t\in T}\lambda\chi_{\Lambda_t} = \lambda Pr\Lambda,\end{aligned}$$

ce qui prouve 15.1.

Le théorème (11.2) nous apprenait que lorsque ξ est une martingale, alors, pour tout réel $p \geq 1, |\xi|^p$ est une sousmartingale. Du fait que :

$$\forall\lambda > 0, \{\max|\xi(t)| \geq \lambda\} = \{\max|\xi(t)|^p \geq \lambda^p\}$$

et du théorème (15.1), nous avons :

$$Pr\{\max|\xi(t)| \geq \lambda\} \leq \frac{1}{\lambda^p}E|\xi(b)|^p. \tag{15.2}$$

Si nous avions $\|\max|\xi(t)|\|_p \leq \|\xi(b)\|_p$, l'inégalité 15.2 découlerait immédiatement du théorème de Bienaymé-Tchebychev. Cela n'est cependant pas vrai en général, mais nous avons quand même l'inégalité remarquable suivante :

$$\|\max|\xi(t)|\|_p \leq p'\|\xi(b)\|_p$$

dans laquelle $p \in]1,\infty]$, et p' est l'exposant conjugué de p. Ce résultat est une conséquence du théorème (15.1) et du théorème (15.2) suivant :

Théorème 15.2. *Soient x et y deux v.a.r. positives telles que :*

$$\forall\lambda > 0, Pr\{y \geq \lambda\} \leq \frac{1}{\lambda}Ex\chi_{\{y\geq\lambda\}}.$$

Alors, pour toute valeur $p \in]1,\infty]$, d'exposant conjugué p', nous avons :

$$\|y\|_p \leq p'\|x\|_p.$$

Démonstration. Lorsque $p \in]1,\infty[$, nous avons les relations suivantes :

$$\begin{aligned}\|y\|_p^p = Ey^p &= \sum_{\lambda\in\{y(\omega):\omega\in\Omega\}}\lambda^p pr_y(\lambda) = -\int_0^{+\infty}\lambda^p\,dPr\{y \geq \lambda\}\\ &= \int_0^{+\infty}p\lambda^{p-1}Pr\{y\geq\lambda\}\,d\lambda \leq \int_0^{+\infty}p\lambda^{p-2}Ex\chi_{\{y\geq\lambda\}}\,d\lambda\\ &= E\left(x\int_0^y p\lambda^{p-2}\,d\lambda\right) = \frac{p}{p-1}Exy^{p-1} \leq p'\|x\|_p\left\|y^{p-1}\right\|_{p'}\\ &= p'\|x\|_p\left\|y^{p/p'}\right\|_{p'} = p'\|x\|_p\,\|y\|_p^{p-1}.\end{aligned}$$

Le cas $p = \infty$ s'en déduit en faisant tendre p vers ∞.[1] $\square$

[1] Il est possible de vérifier directement que $\max x(t) \geq \max y(t)$. (NdT)

Chapitre 16

La loi des grands nombres

Pour une raison ou pour une autre, les probabilistes préfèrent dire "loi des grands nombres" plutôt que "loi des moyennes" qui est pourtant plus expressif. Car c'est pourtant bien sur les propriétés des moyennes

$$y_n = \frac{x_1 + \cdots + x_n}{n}$$

d'une suite de v.a.r. que cette loi porte.

Soit $x_1, \cdots, x_\nu$ une séquence de v.a.r. indépendantes de moyennes nulles et de variances égales à 1. Soit Φ_n l'algèbre de v.a.r. engendrée par les variables $x_1, \cdots, x_n$, et posons $y_n = \dfrac{x_1 + \cdots + x_n}{n}$. Nous voyons que :

$$dy_n = \frac{x_1 + \cdots + x_{n+1}}{n+1} - \frac{x_1 + \cdots + x_n}{n} = -\frac{y_n}{n+1} + \frac{x_{n+1}}{n+1}. \tag{16.1}$$

Comme les variables x_i sont indépendantes entre elles, $E_n x_{n+1} = E x_{n+1} = 0$, et de ce fait :

$$\begin{aligned} Dy_n &= -\frac{y_n}{n+1}; \\ d\hat{y}_n &= \frac{x_{n+1}}{n+1}. \end{aligned}$$

Lorsque $n \approx \infty$, $\|y_n\|_2^2 = \frac{1}{n} \approx 0$, si bien que conformément à l'inégalité de Bienaymé-Tchebychev on a $y_n \underset{ps}{\approx} 0$. C'est à dire que la séquence $(y_i)_{i \in \{1, \cdots, \nu\}}$ converge en probabilité vers 0. Ceci constitue la *loi faible des grands nombres*. Nous allons également prouver *la loi forte des grands nombres*, qui dit que la séquence $(y_i)_{i \in \{1, \cdots, \nu\}}$ converge presque sûrement vers 0.

Afin d'effectuer cette démonstration, remarquons d'abord que :

$$\|Dy_n\|_1 \leq \|Dy_n\|_2 = \frac{1}{(n+1)\sqrt{n}},$$

ce qui indique que la série $\sum_{n=1}^{\nu-1} \|Dy_n\|_1$ converge. La deuxième assertion du théorème (10.1) nous permet d'en conclure que le processus prédictible associé à $(y_i)_{i\in\{1,\cdots,\nu\}}$ converge presque sûrement. De plus, la série

$$\sum_{n=1}^{\nu-1} \|d\hat{y}_n\|_2^2 = \sum_{n=1}^{\nu-1} \frac{1}{(n+1)^2}$$

converge également, d'où l'on déduit conformément à la première assertion du théorème (11.3) que la martingale associée à $(y_i)_{i\in\{1,\cdots,\nu\}}$ converge presque sûrement. En faisant intervenir ici la loi faible des grands nombres, on voit que $(y_i)_{i\in\{1,\cdots,\nu\}}$ converge presque sûrement vers 0.

Nous venons d'établir la loi forte des grands nombres dans le cas que nous évoquions au commencement du chapitre 4. Nous sommes cependant en mesure de montrer des résultats plus forts. Il apparaît que dans la démonstration que nous venons de dérouler nous n'avons pas réellement fait usage de l'hypothèse d'indépendance des v.a.r. x_i, mais nous avons plutôt utilisé le fait que ces v.a.r. étaient les accroissements d'une martingale.

Théorème 16.1. *Soient les v.a.r. $x_1, \cdots, x_\nu$ constituant les accroissements d'une martingale. Alors :*

(i) *si* $\displaystyle\sum_{i=1}^{\nu} \frac{Ex_i^2}{i^2}$ *converge, alors* $\displaystyle\frac{x_1+\cdots+x_n}{n}$ *converge presque sûrement vers* 0 *;*

(ii) *si* $\displaystyle\sum_{i=1}^{\nu} \frac{Ex_i^2}{i^2} \ll \infty$*, alors* $\displaystyle\frac{x_1+\cdots+x_n}{n}$ *est presque sûrement à fluctuation limitée.*

Démonstration. Posons :

$$\begin{aligned} y_n &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ z_n &= \sum_{i=1}^{n} \frac{x_i}{i} \end{aligned}$$

si bien que

$$y_n = -\frac{1}{n}\sum_{j=1}^{n} z_j + \frac{n+1}{n} z_n \tag{16.2}$$

(ce que l'on voit facilement en développant chaque terme en x_i). On convient que $z_0 = 0$. La séquence $(z_i)_{i\in\{0,\cdots,\nu\}}$ forme une martingale et :

$$\|z_\nu\|_2^2 = \sum_{i=1}^{\nu} \frac{Ex_i^2}{i^2} \ll \infty.$$

D'après le théorème (11.1), le long de presque toutes les trajectoires chaque z_n est limité. Dans le cas (i) que nous voulons démontrer, la première assertion du théorème (11.3) implique que $(z_n)_{n\in\{0,\cdots,\nu\}}$ converge presque sûrement. Dans le cas (ii), le théorème (12.3) implique que $(z_n)_{n\in\{0,\cdots,\nu\}}$ est presque sûrement à fluctuation limitée. Il s'ensuit que nous aurons achevé la démonstration si nous pouvons prouver les deux énoncés suivants :

(I) si la trajectoire z_n est limitée et convergente, alors la trajectoire y_n converge vers 0 ;

(II) si la trajectoire z_n est limitée et à fluctuation limitée, alors la trajectoire y_n est à fluctuation limitée.

Commençons par démontrer (I). Soit $n \approx \infty$. Comme z_n est limité,

$$\frac{n+1}{n} z_n \approx z_n \approx z_\nu.$$

Soit $\varepsilon \gg 0$. Définissons l'entier k comme étant le plus grand entier inférieur à ν tel que $|z_k - z_\nu| > \varepsilon$ (dans le cas où un tel indice n'existe pas, nous conviendrons que $k = 0$). Alors $k \ll \infty$. Pour tout indice j, $|z_j| \ll \infty$ et comme n est illimité, on a :

$$\frac{1}{n}\sum_{j=1}^{k} z_j \approx 0.$$

De ce fait :

$$\left| \frac{1}{n}\sum_{j=1}^{n} z_j - z_\nu \right| \lesssim \varepsilon.$$

Comme le réel $\varepsilon \gg 0$ est arbitraire, la partie droite de l'égalité 16.2 est presque égale à $-z_\nu + z_\nu = 0$. Ceci conclut la preuve de (I).

Démontrons maintenant (II). On remarque que dès que $n \approx \infty$, $z_n/n \approx 0$. Ainsi z_n/n converge vers 0, et est donc à fluctuation limitée. On peut montrer grâce à l'inégalité triangulaire que la somme de deux suites à fluctuation limitée est elle-même à fluctuation limitée. Il s'ensuit que $\frac{n+1}{n} z_n = z_n + \frac{1}{n} z_n$ est à fluctuation limitée.

Posons $w_n = \frac{1}{n}\sum_{j=1}^{n} z_j$. Il nous suffit de montrer que w_n est à fluctuation limitée. Pour cela, reprenons les notations du chapitre 12. Suivant la démonstration du théorème (12.3), on sait que si w_n n'est pas à fluctuation limitée, il existe deux réels $\lambda_1 \ll \lambda_2$ tels que w_n franchit l'intervalle $[\lambda_1, \lambda_2]$ un nombre illimité de fois. Comme dans le chapitre 12, définissons pour la séquence w_n une suite d'indices s_1, t_1, s_2, t_2, etc. (voir la figure 12.3). Si la suite w_n franchit dans le sens croissant l'intervalle $[\lambda_1, \lambda_2]$, il existe nécessairement un indice $j_1 \in [s_1, t_1]$ tel que $z_{j_1} \geq \lambda_2$. De la même manière, si un nouveau franchissement en sens croissant se produit, on doit avoir un indice $k_1 \in [t_1, s_2]$ tel que $z_{k_1} \leq \lambda_1$. On voit par là que si w_n présente β franchissements dans le sens croissant de $[\lambda_1, \lambda_2]$, z_n doit en présenter au moins $\beta - 1$. Ceci interdit à β d'être illimité, car z_n est à fluctuation limitée, et conclut la démonstration de (II). □

Voici un exemple où nous sommes dans le cas (ii), mais pas dans le cas (i). Définissons une séquence $(x_i)_{i\in\{1,\cdots,\nu\}}$ de v.a.r. de la manière suivante :

- $x_n = 0$ si $n < \nu$;
- $x_\nu = +\nu$ avec une probabilité de 1/2, et $-\nu$ avec une probabilité de 1/2.

Ainsi, $\sum_{i=1}^{\nu} \frac{Ex_i^2}{i^2} = 1$. La moyenne de y_n est nulle pour tout $n \leq \nu$, mais y_ν vaut $+1$ ou -1 avec une probabilité de 1/2.

Dans le cas (ii), le fait que n soit illimité n'implique donc pas nécessairement que y_n soit infinitésimal. En revanche, lorsque n est limité, les y_n doivent avoir tendance à devenir et à rester en valeur absolue inférieurs à n'importe quel réel appréciable $\varepsilon \gg 0$. En effet, d'après le théorème (6.1), si $\sum_{i=1}^{\nu} \frac{Ex_i^2}{i^2} \ll \infty$, il existe un indice illimité $\mu \leq \nu$ tel que $\sum_{i=1}^{\mu} \frac{Ex_i^2}{i^2}$ converge. Nous nous retrouvons ainsi dans le cas (i) appliqué à la séquence $(x_i)_{i\in\{1,\cdots,\mu\}}$, et nous savons alors que $(y_i)_{i\in\{1,\cdots,\mu\}}$ converge vers 0.

Il est possible de généraliser le théorème (16.1) à des exposants p différents de 2, mais la démonstration repose sur une méthode de troncature qui permet de se ramener au cas $p = 2$.

Théorème 16.2. *Soient les v.a.r. $x_1, \cdots, x_\nu$ constituant les accroissements d'une martingale, et soit un réel $0 \ll p \leq 2$. Alors :*

(i) *si $\displaystyle\sum_{i=1}^{\nu} \frac{E|x_i|^p}{i^p}$ converge, alors $\displaystyle\frac{x_1 + \cdots + x_n}{n}$ converge presque sûrement vers 0 ;*

(ii) *si $\displaystyle\sum_{i=1}^{\nu} \frac{E|x_i|^p}{i^p} \ll \infty$, alors $\displaystyle\frac{x_1 + \cdots + x_n}{n}$ est presque sûrement à fluctua-*

tion limitée.

Démonstration. Montrons d'abord le résultat préliminaire suivant. Dans les deux cas, nous sommes assurés que :

$$ps : \exists c \ll \infty, \forall n \in [1, \nu], |x_n| \leq cn.$$

En effet, soit $\varepsilon \gg 0$, posons $a = \sum_{i=1}^{\nu} \frac{E|x_i|^p}{i^p}$ ainsi que $c = \left(\frac{a}{\varepsilon}\right)^{\frac{1}{p}}$. Comme $a \ll \infty$ et $p \gg 0$, il s'ensuit que $c \ll \infty$. Appliquons l'inégalité de Bienaymé-Tchebychev :

$$\begin{aligned} Pr\{\exists n \in [1, \nu], |x_n| > cn\} &\leq \sum_{n=1}^{\nu} Pr\{|x_n| > cn\} \\ &\leq \sum_{n=1}^{\nu} \frac{E|x_n|^p}{c^p n^p} = \frac{a}{c^p} = \varepsilon. \end{aligned}$$

Comme $\varepsilon \gg 0$ est arbitraire, cela prouve ce résultat préliminaire.

Convenons de dire, à partir de maintenant, qu'une séquence de v.a.r. *fonctionne* lorsqu'elle converge presque sûrement vers 0 dans le cas (i), ou lorsqu'elle est presque sûrement à fluctuation limitée dans le cas (ii). D'après le résultat préliminaire que nous venons de prouver, il suffit pour montrer le théorème (16.2) de prouver que pour tout $c \ll \infty$ la séquence $\frac{x_1^{(c)} + x_2^{(2c)} + \cdots + x_n^{(nc)}}{n}$ fonctionne. Les variables $\left(x_n^{(nc)} - E_{n-1} x_n^{(nc)}\right)$ sont forcément les accroissements d'une martingale, bien que les variables $x_n^{(nc)}$ ne constituent pas forcément de tels accroissements.[1]

Nous avons les relations suivantes :

$$\begin{aligned} \sum_{n=1}^{\nu} \left\| E_{n-1} \frac{x_n^{(nc)}}{n} \right\|_1 &= \sum_{n=1}^{\nu} \frac{1}{n} \left\| E_{n-1}\left(x_n - x_n^{(nc)}\right) \right\|_1 \\ &= \sum_{n=1}^{\nu} \frac{1}{n} \sum_{|\lambda| > nc} |\lambda| pr_{x_n}(\lambda) \\ &\leq \sum_{n=1}^{\nu} \sum_{|\lambda| > nc} \frac{|\lambda|^p}{n^p c^{p-1}} pr_{x_n}(\lambda) \\ &\leq \frac{1}{c^{p-1}} \sum_{n=1}^{\nu} \frac{E|x_n|^p}{n^p}. \end{aligned}$$

[1] En convenant que $E_0 = E$. (NdT)

Donc, d'après les théorèmes 11.3 et 12.3, la série $\sum_{n=1}^{\nu} \frac{E_{n-1}x_n^{(nc)}}{n}$ converge presque sûrement dans le cas (i), et est presque sûrement à fluctuation limitée dans le cas (ii). Dans les deux cas, on a :

$$ps : \forall n \in [1, \nu]\,, \left| \frac{E_{n-1}x_n^{(nc)}}{n} \right| \ll \infty.$$

En réutilisant les résultats (I) et (II) que nous avons montrés lors de la démonstration de (16.1), nous voyons que la séquence

$$\frac{E_0 x_1^{(c)} + E_1 x_2^{(2c)} + \cdots + E_{n-1} x_n^{(nc)}}{n}$$

fonctionne. Il nous reste donc à prouver que la séquence des moyennes successives de $\left(x_n^{(nc)} - E_{n-1}x_n^{(nc)}\right)$ fonctionne.[2] Vérifions que cette séquence vérifie les hypothèses du théorème (16.1) :

$$\begin{aligned}\sum_{n=1}^{\nu} \frac{E\left(x_n^{(nc)} - E_{n-1}x_n^{(nc)}\right)^2}{n^2} &\leq \sum_{n=1}^{\nu} \frac{E x_n^{(nc)2}}{n^2} \\ &= \sum_{n=1}^{\nu} \frac{1}{n^2} \sum_{|\lambda| \leq nc} |\lambda|^2 pr_{x_n}(\lambda)\end{aligned}$$

et du fait que $p \leq 2$:

$$\begin{aligned}\sum_{n=1}^{\nu} \frac{1}{n^2} \sum_{|\lambda| \leq nc} |\lambda|^2 pr_{x_n}(\lambda) &\leq \sum_{n=1}^{\nu} \frac{1}{n^2} \sum_{|\lambda| \leq nc} \frac{|\lambda|^p}{(nc)^{p-2}} pr_{x_n}(\lambda) \\ &= \frac{1}{c^{p-2}} \sum_{n=1}^{\nu} \frac{1}{n^p} \sum_{|\lambda| \leq nc} |\lambda|^p pr_{x_n}(\lambda) \\ &\leq \frac{1}{c^{p-2}} \sum_{n=1}^{\nu} \frac{E|x_n|^p}{n^p}.\end{aligned}$$

Ce qui conclut la démonstration. □

Corollaire. *Soient $x_1, \cdots, x_\nu$ les accroissements d'une martingale, et soit $p \gg 1$. Si chaque $\|x_n\|_p$ est limité, alors la séquence*

$$\frac{x_1 + \cdots + x_n}{n}$$

converge presque sûrement vers 0.

[2] C'est à dire la séquence $\frac{\left(x_1^{(c)} - E_0 x_1^{(c)}\right) + \cdots + \left(x_n^{(nc)} - E_{n-1}x_n^{(nc)}\right)}{n}$. (NdT)

Démonstration. On peut admettre sans perte de généralité que $p \leq 2$.[3] Posons $M = \max_{n \in [1,\nu]} \|x_n\|_p$. Le réel M est limité, du fait qu'il existe un indice i tel que $M = \|x_i\|_p$. Ceci nous montre aussitôt que les hypothèses du théorème (16.2) sont vérifiées. □

Notons que pour $p = 1$, le théorème (16.2) est trivial. Ce résultat particulier découle du théorème (10.1) ainsi que des résultats (I) et (II) montrés précédemment, et ne nécessite même pas que les x_n soient les accroissements d'une martingale. En revanche, le corollaire est faux pour $p = 1$, comme peut nous en convaincre le contre-exemple suivant.

Pour $n = 3, \cdots, \nu$, considérons une séquence de v.a.r. x_n indépendantes et vérifiant les propriétés suivantes :

$$\begin{aligned} Pr\{x_n = n\} &= \tfrac{1}{2n \log n}; \\ Pr\{x_n = 0\} &= \tfrac{1}{2} - \tfrac{1}{2n \log n}; \\ Pr\Big\{x_n = -\tfrac{1}{\log n}\Big\} &= \tfrac{1}{2}. \end{aligned}$$

Alors, pour tout n, $Ex_n = 0$, et $\|x_n\|_1 = \frac{1}{\log n}$. On voit que chaque variable x_n est L^1, et de plus que $(n \approx \infty) \Rightarrow (\|x_n\|_1 \approx 0)$. La version cardinale du théorème de Borel-Cantelli (deuxième assertion du théorème (7.4)) nous dit que, presque partout, il existe un nombre illimité d'indices n tels que $x_n = n$. Donc, presque partout, il existe un nombre illimité d'indices n tels que :

$$y_n = \frac{x_3 + \cdots + x_n}{n-2} \geq 1 - \frac{1}{\log 3}.^{4}$$

Ainsi, presque partout, y_n ne converge pas vers 0. On peut voir, en utilisant la relation (16.1), que presque partout y_n est à fluctuation illimitée.

Le théorème suivant montre qu'il y a un cas important, celui d'une suite d'observations indépendantes d'une même v.a.r., pour lequel la loi forte des grands nombres s'applique moyennant des hypothèses de type L^1.

Théorème 16.3. *Soit x une v.a.r., et soient $x_1, \cdots, x_\nu$ des v.a.r. indépendantes suivant la même loi de probabilité que x.*

(i) *Si x est L^1, alors $\dfrac{x_1 + \cdots + x_n}{n}$ converge presque sûrement vers Ex.*

(ii) *Si $\|x\|_1 \ll \infty$, alors $\dfrac{x_1 + \cdots + x_n}{n}$ est presque sûrement à fluctuation limitée.*

---

[3] Car si $p > 2$, $\|x_n\|_2 \leq \|x_n\|_p$. (NdT)

[4] Minorer brutalement, en supposant que $x_n = n$, et $x_i = -\frac{1}{\log n}$ pour $i < n$. (NdT)

Démonstration. Nous allons procéder de la même manière que pour le théorème (16.2), mais des différences apparaîtront par endroit. Nous admettrons sans perte de généralité que $Ex = 0$.

Commençons par établir le même résultat préliminaire que pour le théorème (16.2) :

$$ps : \exists c \ll \infty, \forall n \in [1, \nu], |x_n| \leq cn.$$

En effet, soit $\varepsilon \gg 0$, et posons :

$$\begin{aligned} c &= \frac{E|x|}{\varepsilon}, \\ E_k &= \{\omega \in \Omega : ck < |x(\omega)| \leq \ c(k+1)\}. \end{aligned}$$

Nous avons les relations suivantes :

$$\begin{aligned} Pr\{\exists n \in [1, \nu], |x_n| > nc\} &\leq \sum_{n=1}^{\nu} Pr\{|x_n| > nc\} \\ &= \sum_{n=1}^{\nu} Pr\{|x| > nc\} = \sum_{n=1}^{\nu} \sum_{k=n}^{\infty} PrE_k \\ &= \sum_{k=0}^{\infty} \sum_{n=1}^{\min(k,\nu)} PrE_k \leq \sum_{k=0}^{\infty} kPrE_k \\ &\leq \frac{1}{c} \sum_{\lambda} |\lambda| pr_x(\lambda) = \frac{1}{c} E|x| = \varepsilon, \end{aligned}$$

ce qui démontre le résultat intermédiaire.

Il découle de ce résultat qu'il nous suffit de montrer que pour tout $c \ll \infty$ la séquence $\frac{x_1^{(c)} + x_2^{(2c)} + \cdot + x_n^{(nc)}}{n}$ fonctionne pour prouver le théorème. Nous avons :

$$Ex_n^{(nc)} = Ex^{(nc)} = E\Big(x^{(nc)} - x\Big),$$

ce qui nous montre que :

$$\Big|Ex_n^{(nc)}\Big| \leq \sum_{|\lambda| > nc} |\lambda| pr_x(\lambda).$$

De plus, pour $n < m$, nous avons :

$$\Big|Ex_n^{(nc)} - Ex_m^{(mc)}\Big| = \Big|Ex^{(nc)} - Ex^{(mc)}\Big| \leq \sum_{nc < \lambda \leq mc} |\lambda| pr_x(\lambda)$$

ce qui montre que la séquence de variables $Ex_n^{(nc)}$ est toujours limitée, et qu'elle fonctionne. Il s'ensuit que la séquence $\frac{Ex_1^{(c)}+Ex_2^{(2c)}+\cdots+Ex_n^{(nc)}}{n}$ fonctionne, ce dont on peut se convaincre en suivant les mêmes raisonnements que pour montrer (I) et (II).

Il nous suffit donc de montrer que $\left(x_n^{(nc)} - Ex_n^{(nc)}\right)$ vérifie les hypothèses du théorème (16.1) (c'est ici que la condition sur l'indépendance des variables x_n intervient, elle permet d'utiliser les espérances absolues plutôt que les espérances conditionnelles pour voir que les $\left(x_n^{(nc)} - Ex_n^{(nc)}\right)$ sont les incréments d'une martingale). Or, soit m tel que $\frac{\sqrt{m}}{c} \geq 2$. Nous avons :

$$\begin{aligned}
\sum_{n=m}^{\nu} \frac{E\left(x_n^{(nc)} - Ex_n^{(nc)}\right)^2}{n^2} &\leq \sum_{n=m}^{\nu} \frac{Ex^{(nc)2}}{n^2} \\
&= \sum_{n=m}^{\nu} \frac{1}{n^2} \sum_{|\lambda|\leq nc} \lambda^2 pr_x(\lambda) \\
&= \sum_{\lambda\in x(\Omega)} \lambda^2 pr_x(\lambda) \sum_{n\geq \max\left(m,\frac{|\lambda|}{c}\right)} \frac{1}{n^2} \\
&\leq \sum_{|\lambda|\leq\sqrt{m}} \lambda^2 pr_x(\lambda) \sum_{n=m}^{\nu} \frac{1}{n^2} + \sum_{|\lambda|>\sqrt{m}} \lambda^2 pr_x(\lambda) \sum_{n=\frac{|\lambda|}{c}}^{\nu} \frac{1}{n^2} \\
&\leq \sum_{|\lambda|\leq\sqrt{m}} \lambda^2 pr_x(\lambda) \frac{1}{m-1} + \sum_{|\lambda|>\sqrt{m}} \lambda^2 pr_x(\lambda) \frac{1}{\frac{|\lambda|}{c} - 1} \\
&\leq \frac{\sqrt{m}}{m-1} \sum_{\lambda\in x(\Omega)} |\lambda| pr_x(\lambda) + \sum_{|\lambda|>\sqrt{m}} 2c|\lambda| pr_x(\lambda)
\end{aligned}$$

ce qui nous donne dans le cas (i) une valeur infinitésimale si $m \approx \infty$ (montrant par là la convergence de la série), et dans le cas (ii) une valeur limitée. D'où le résultat. □

Soient $x_1, \cdots, x_\nu$ et x des v.a.r. que l'on ne suppose pas forcément définies sur le même espace probabilisé. On dira que la famille $\{x_1, ..., x_\nu\}$ est *dominée en distribution* par x lorsque :

$$\exists a \ll \infty, \forall n \in [1,\nu], \forall k \in N, Pr\{k < |x_n| \leq k+1\} \leq a Pr\{k < |x| \leq k+1\}.$$

De manière évidente, on est dans ce cas si chaque variable $x_1, \cdots, x_\nu$ possède la même distribution de probabilité que x. En revanche, dans le contre-exemple que nous avons donné précédemment dans ce chapitre, il n'existe aucune variable aléatoire dominant en distribution la famille $\{x_3, \cdots, x_\nu\}$.

Soient $x_1, \cdots, x_\nu$ des v.a.r. indépendantes, de moyennes nulles, et dominées en distribution par une variable x. Si on relit la démonstration du théorème (16.3), on constate que chaque argument reste valable dans le cas plus général présent, à l'exclusion de l'estimation de $\left|Ex_n^{(nc)} - Ex_m^{(mc)}\right|$ utilisée dans le cas (ii). Nous avons donc le théorème suivant :

Théorème 16.4. *Soit x une v.a.r. et soient $x_1, \cdots, x_\nu$ des v.a.r. indépendantes, de moyennes nulles, dominées en distribution par x. Alors, si x est L^1, la séquence $\frac{x_1+\cdots+x_n}{n}$ converge presque sûrement vers* 0.

Donnons un contre-exemple afin de montrer que l'assertion (ii) peut être fausse dans les conditions du théorème (16.4). Soit $\nu \approx \infty$, et soit μ tel que $\mu/\nu \approx \infty$. Soit x une v.a.r. vérifiant les propriétés suivantes :

$$\begin{aligned} Pr\{x = 1\} &= 1 - \tfrac{1}{\mu}, \\ Pr\{x = 1 - \mu\} &= \tfrac{1}{\mu}. \end{aligned}$$

La variable x n'est pas L^1, cependant $\|x\|_1 = 2 - \frac{2}{\mu} \ll \infty$ et $Ex = 0$. Soient $x'_1, \cdots, x'_\nu$ des observations de x indépendantes. Nous avons :

$$ps : \forall n \in [1, \nu], x'_n = 1.$$ [5]

Soient $x_1, \cdots, x_\nu$ des v.a.r. définies comme étant égales à $x'_1, \cdots, x'_\nu$, sauf que l'on remplace x'_1 par 0, on laisse les 2! variables x'_i suivantes inchangées, on remplace les 3! suivantes par 0, on laisse les 4! suivantes inchangées, et ainsi de suite. Alors, les variables x_i sont indépendantes, de moyennes nulles, et dominées en distribution par la variable x. La séquence $\frac{x_1+\cdots+x_n}{n}$ est presque partout à fluctuation illimitée bien que $\|x\|_1 \ll \infty$.

La loi forte des grands nombres met en exergue le fait que les moyennes d'un nombre croissant d'observations indépendantes d'une v.a.r. x, $\frac{x_1+\cdots+x_n}{n}$, finissent par se fixer dans un voisinage de Ex. Cela est vrai lorsque x est L^1, conformément à l'assertion (i) du théorème (16.3). Cependant si nous pouvons seulement affirmer que $\|x\|_1 \ll \infty$, la deuxième assertion du théorème (16.3) ne nous apprend pas grand chose. Elle nous dit uniquement que les valeurs des moyennes successives ne fluctueront de manière sensible qu'un nombre limité de fois. Néanmoins, si $\|x\|_1 \ll \infty$, le théorème (6.1) appliqué

[5] Il suffit de vérifier que le produit des probabilités individuelles est équivalent à 1. (NdT)

à la séquence $\left(E\left|x^{(n)}\right|\right)_{n\in\{1,\cdots,\nu\}}$ nous assure l'existence d'un réel $a \gg 0$ tel que $x^{(a)}$ est L^1. Nous pouvons de plus choisir un entier illimité $\mu \leq \nu$ tel que :

$$ps : \forall n \leq \mu, x_n = x_n^{(a)}.^6$$

L'assertion (i) du théorème (16.3) nous dit alors que la séquence $\frac{x_1+\cdots+x_n}{n}$, avec $n \leq \mu$, converge presque sûrement vers $Ex^{(a)}$.

Nous dirons qu'une v.a.r. x est *presque* L^1 s'il existe une v.a.r. L^1 y telle que $x \underset{ps}{\approx} y$, et nous dirons de plus dans ce cas que Ey est une *espérance réduite* de x. Soit z une v.a.r. L^1 telle que $z \underset{ps}{\approx} x$. Nous savons d'après le théorème de Lebesgue que dans ce cas $Ey \approx Ez$. Ainsi, l'espérance réduite de x est unique à un infinitésimal près.

***Théorème 16.5.** *Soit x une v.a.r. Les propriétés suivantes sont équivalentes :*

(i) *x est presque L^1.*

(ii) *Il existe un réel $a \approx \infty$ tel que $Pr\{|x| > a\} \approx 0$ et $E\left|x^{(a)}\right| \ll \infty$.*

(iii) *Il existe une réel $a \approx \infty$ tel que $Pr\{|x| > a\} \approx 0$ et $x^{(a)}$ est L^1.*

Démonstration. Comme nous l'avons vu précédemment, le théorème (6.1) nous montre que (ii)⇒(iii).[7] L'implication (iii)⇒(i) est immédiate. Enfin, supposons (i). Alors évidemment $\forall a \approx \infty, Pr\{|x| > a\} \approx 0$. Soit y une v.a.r. L^1 telle que $x \underset{ps}{\approx} y$, et soit K un réel tel que $E|y| \ll K \ll \infty$. Supposons que $\forall a \approx \infty, E\left|x^{(a)}\right| \approx \infty$. Alors, l'ensemble $\left\{u > 0 : E\left|x^{(u)}\right| \geq K\right\}$ contient tous les réels positifs illimités, donc il contient au moins un réel positif limité a d'après le principe de Cauchy. Or, $\left|x^{(a)}\right| \leq |x| \underset{ps}{\approx} |y|$. Comme $x^{(a)}$ est L^1, ainsi que y, il s'ensuit grâce au théorème de Lebesgue que $E\left|x^{(a)}\right| \lesssim E|y| \ll K$, ce qui est contradictoire. Ceci conclut la démonstration. □

[6] En effet $Pr\{|x| \geq a\} \approx 0$, donc $Pr\left\{x \neq x^{(a)}\right\} \approx 0$. De plus :

$$\forall^{st} n, Pr\left\{\exists k \leq n, x_n \neq x_n^{(a)}\right\} \leq nPr\left\{x \neq x^{(a)}\right\} \leq \frac{1}{n}.$$

Donc l'ensemble

$$\left\{n : Pr\left\{\exists k \leq n, x_n \neq x_n^{(a)}\right\} \leq \frac{1}{n}\right\}$$

contient tous les naturels positifs standards. Il contient donc au moins un entier illimité μ en vertu du principe de Cauchy. (NdT)

[7] Egalement grâce à l'inégalité de Bienaymé-Tchebychev, qui nous assure que pour tout $b \approx \infty$ inférieur à a, $E|x^{(b)}| \ll \infty$ (du fait que $E|x^{(b)} \leq |x^{(a)}| \ll \infty$). (NdT)

Nous avons les implications suivantes :

$$\big(x \text{ est } L^1\big) \Rightarrow (\|x\|_1 \ll \infty) \Rightarrow \big(x \text{ est presque } L^1\big),$$

mais aucune des implications inverses n'est vraie en général.

Le théorème suivant est un corollaire de l'assertion (i) du théorème (16.3) et du théorème (16.5) :

***Théorème 16.6.** *Soit x une v.a.r. presque L^1, et soient $x_1, \cdots, x_\nu$, avec $\nu \approx \infty$, des v.a.r. indépendantes de même distribution de probabilité que x. Alors il existe un entier illimité $\mu \leq \nu$ tel que la séquence $\frac{x_1+\cdots+x_n}{n}$, avec $1 \leq n \leq \mu$, converge presque sûrement vers l'espérance réduite de x.*

Chapitre 17

Processus quasiment équivalents

Au chapitre 3 nous avons défini la notion interne d'équivalence entre processus stochastiques. Nous allons définir dans le présent chapitre la notion de quasi-équivalence, qui est, elle, une notion externe. L'idée intuitive que recouvre cette nouvelle notion est celle de processus que l'on ne pourrait distinguer que par des moyens d'observation de résolutions infinitésimales.

Souvenons-nous que les trajectoires d'un processus stochastique ξ indexé par un ensemble fini T et défini sur un ensemble probabilisé fini $\langle\Omega, pr\rangle$ forment un sous-ensemble fini de R^T, que nous notons Λ_ξ. Si Λ est un sous-ensemble fini de R^T, et si F est une fonction réelle de Λ, $F : \Lambda \to R$, nous dirons que F est une *fonctionnelle* définie sur Λ. Si F est une fonctionnelle définie sur Λ, et si $\Lambda_\xi \subseteq \Lambda$, $F(\xi) = F \circ \xi$ est une variable aléatoire définie sur $\langle\Omega, pr\rangle$. L'ensemble Λ est un espace métrique pour la métrique ρ :

$$\rho(\lambda, \mu) = \max_{t \in T} |\lambda(t) - \mu(t)|$$

et la fonctionnelle F est *(quasiment) continue* si et seulement si :

$$\forall(\lambda, \mu) \in \Lambda^2, (\rho(\lambda, \mu) \approx 0) \Rightarrow (F(\lambda) \approx F(\mu)),$$

ou, de manière équivalente, si et seulement si :

$$\forall(\lambda, \mu) \in \Lambda^2, (\forall t \in T, \lambda(t) \approx \mu(t)) \Rightarrow (F(\lambda) \approx F(\mu)).$$

Nous dirons que F est *limitée* si $\forall\lambda \in \Lambda, |F(\lambda)| \ll \infty$. Comme $|F|$ atteint son maximum (Λ est fini), cela est équivalent à écrire $\max_{\lambda\in\Lambda} |F(\lambda)| \ll \infty$.

Soient ξ et η deux processus stochastiques indexés par le même ensemble fini T, mais éventuellement définis sur deux espaces probabilisés distincts. On peut remarquer que ξ et η sont équivalents si et seulement si on a $EF(\xi) = EF(\eta)$ quelle que soit la fonctionnelle F (définie au moins sur $\Lambda_\xi \bigcup \Lambda_\eta$, mais si les processus sont équivalents, on a $\Lambda_\xi = \Lambda_\eta$). Nous dirons que ξ et η sont *quasiment équivalents* dès que $EF(\xi) \approx EF(\eta)$ quelle que soit la fonctionnelle F limitée et continue définie sur $\Lambda_\xi \bigcup \Lambda_\eta$.

Théorème 17.1. *Soit ξ un processus stochastique défini sur $\langle \Omega \times pr \rangle$. Supposons de plus que Ω soit muni d'une seconde structure d'espace probabilisé par la donnée d'une seconde densité de probabilité pr' telle que*

$$\sum_{\omega \in \Omega} \left| pr(\omega) - pr'(\omega) \right| \approx 0.$$

Notons par ξ' le processus ξ considéré comme un processus stochastique sur l'espace $\langle \Omega \times pr' \rangle$. Alors ξ et ξ' sont quasiment équivalents.

Démonstration. Soit F une fonctionnelle limitée définie sur Λ_ξ. Nous avons :

$$\begin{aligned} \left| EF(\xi) - E'F\big(\xi'\big) \right| &\leq \sum_{\omega \in \Omega} |F(\xi(\omega))| \left| pr(\omega) - pr'(\omega) \right| \\ &\leq \max |F| \sum_{\omega \in \Omega} \left| pr(\omega) - pr'(\omega) \right| \end{aligned}$$

d'où le résultat. □

Théorème 17.2. *Soient ξ et η deux processus stochastiques indexés par T et définis sur le même espace probabilisé fini. Si l'on a :*

$$ps : \forall t \in T, \xi(t) \approx \eta(t)$$

alors ξ et η sont quasiment équivalents.

Démonstration. Si F est une fonctionnelle limitée, $F(\xi)$ et $F(\eta)$ sont L^1, et si F est continue, $F(\xi) \underset{ps}{\approx} F(\eta)$. Par le théorème de Lebesgue, nous pouvons donc écrire $EF(\xi) \approx EF(\eta)$, ce qui conclut la démonstration. □

Dans le théorème suivant, A désigne indifféremment une formule interne ou externe. Ce théorème va nous rassurer sur le fait que la définition de la quasi-équivalence que nous avons adoptée correspond bien au concept intuitif que nous voulions obtenir.

Théorème 17.3. *Supposons que :*

$$\forall(\lambda,\mu) \in \Lambda^2, (\rho(\lambda,\mu) \approx 0) \Rightarrow (A(\lambda) \Leftrightarrow A(\mu)).$$

Soient ξ et η deux processus stochastiques quasiment équivalents, tels que $\Lambda_\xi \bigcup \Lambda_\eta \subseteq \Lambda$. Alors :

$$(ps : A(\xi)) \Longleftrightarrow (ps : A(\eta)).^1$$

Démonstration. Sans sacrifier la généralité du propos, on peut admettre que ξ et η sont définis respectivement sur Λ_ξ et Λ_η. Supposons $A(\xi)$ presque partout. Il nous faut montrer que dans ces conditions on a $A(\eta)$ presque partout.

Soit $\varepsilon \gg 0$. Il existe un sous-ensemble Φ de Λ_ξ tel que :

$$\begin{aligned} &Pr_\xi \Phi \geq 1 - \varepsilon, \\ &\forall\lambda \in \Phi, A(\lambda). \end{aligned}$$

Pour tout réel $\beta > 0$, définissons sur Λ la fonctionnelle F_β de la manière suivante :

$$\forall\lambda \in \Lambda, F_\beta(\lambda) = \left(1 - \frac{\rho(\lambda,\Phi)}{\beta}\right)^+.$$

Chaque F_β est limitée. De plus, dès que $\beta \gg 0$, F_β est continue sur Λ. Posons maintenant :

$$\Phi_\beta = \{\lambda \in \Lambda : \rho(\lambda,\Phi) \leq \beta\}.$$

On a évidemment :

$$\forall\beta > 0, \chi_{\Phi_\beta} \geq F_\beta \geq \chi_\Phi,$$

d'où l'on déduit :

$$\begin{aligned} &Pr_\eta \Phi_\beta \geq EF_\beta(\eta), \\ &EF_\beta(\xi) \geq Pr_\xi \Phi \geq 1 - \varepsilon. \end{aligned}$$

Si $\beta \gg 0$, $EF_\beta(\eta) \approx EF_\beta(\xi)$, et il en découle $Pr_\eta \Phi_\beta \gtrsim 1 - \varepsilon$. Nous avons donc :

$$\forall\beta \gg 0, Pr_\eta \Phi_\beta \geq 1 - 2\varepsilon.$$

[1] A peut être vu comme un procédé de qualification stochastique de processus : ξ possède une qualité si et seulement si on a presque partout $A(\xi)$. Dans le cas présent, A n'est pas un moyen de discrimination des trajectoires de résolution infinitésimale. Etant donnés deux processus stochastiques quasiment équivalents, si A qualifie l'un des processus, il qualifie également le second. Deux processus stochastiques quasiment équivalents sortent donc de la limite de résolution de A, ce qui est bien le concept recherché. (NdT)

Ceci étant vrai pour tout réel $\beta \gg 0$, il existe, en vertu du principe de Cauchy, une valeur infinitésimale de β pour laquelle l'inégalité est encore vérifiée. Or, nous avons, d'après les hypothèses :

$$\forall \beta > 0, (\beta \approx 0) \Longrightarrow (\forall \lambda \in \Phi_\beta, A(\lambda)).$$

Du fait que $\varepsilon \gg 0$ est arbitraire, on en déduit que, presque partout, $A(\eta)$. Ceci conclut la démonstration. $\square$

Ce résultat appelle deux corollaires :

Corollaire 1. *Soient ξ et η deux processus stochastiques quasiment équivalents, et soit F une fonctionnelle. Si F est continue en presque toutes les trajectoires de ξ, F est continue en presque toutes les trajectoires de η.*

Corollaire 2. *Soient ξ et η deux processus stochastiques quasiment équivalents indexés par un sous-ensemble fini de R, T. Si presque toutes les trajectoires de ξ sont continues, alors presque toutes les trajectoires de η sont continues.*

Théorème 17.4. *Soient ξ et η deux processus stochastiques quasiment équivalents, et soit F une fonctionnelle limitée, continue en presque toutes les trajectoires de ξ. Alors, $EF(\xi) \approx EF(\eta)$.*

Démonstration. On peut supposer, sans perte de généralité, que ξ et η sont définis respectivement sur Λ_ξ et Λ_η, et que $1 \leq F \leq c \ll \infty$.[2] Soit $\varepsilon \gg 0$. Le corollaire 1 nous assure l'existence d'un sous-ensemble Φ de $\Lambda_\xi \bigcup \Lambda_\eta$ tel que :

$$\begin{array}{rcl} Pr(\Phi \bigcap \Lambda_\xi) & \geq & 1 - \varepsilon, \\ Pr(\Phi \bigcap \Lambda_\eta) & \geq & 1 - \varepsilon, \end{array}$$

et tel que F soit continue partout sur Φ. Soit G la fonctionnelle définie de la manière suivante :

$$G(\lambda) = \begin{cases} F(\lambda) & \text{si } \lambda \in \Phi\,; \\ \dfrac{\min_{\mu \in \Phi} (\rho(\lambda, \mu) F(\mu))}{\rho(\lambda, \Phi)} & \text{si } \lambda \notin \Phi. \end{cases}$$

A tout élément λ de $\Lambda = \Lambda_\xi \bigcup \Lambda_\eta$ associons deux éléments de Φ, λ^* et λ', tels que :

$$\begin{array}{rcl} \rho(\lambda, \lambda^*) & = & \rho(\lambda, \Phi), \\ \rho(\lambda, \lambda') F(\lambda') & = & \min_{\mu \in \Phi} (\rho(\lambda, \mu) F(\mu)). \end{array}$$

[2] Il suffit éventuellement d'ajouter à F une constante adéquate (par exemple $1 + |\min_{\lambda \in \Lambda} F(\lambda)|$) (NdT).

Alors on peut voir que $1 \leq G \leq c$. En effet, si $\lambda \in \Phi$, l'inégalité est trivialement vérifiée. Si $\lambda \notin \Phi$, nous avons :

$$1 \leq \frac{\rho(\lambda,\lambda')F(\lambda')}{\rho(\lambda,\Phi)} = G(\lambda) \leq \frac{\rho(\lambda,\lambda^*)F(\lambda^*)}{\rho(\lambda,\Phi)} = F(\lambda^*) \leq c. \tag{17.1}$$

Montrons que, de plus, G est partout continue sur Λ. Soient λ_1 et λ_2 deux éléments de Λ tels que $\rho(\lambda_1,\lambda_2) \approx 0$. Si $(\lambda_1,\lambda_2) \in \Phi^2$, alors

$$G(\lambda_1) - G(\lambda_2) = F(\lambda_1) - F(\lambda_2) \approx 0.$$

Imaginons que $\lambda_1 \notin \Phi$, mais que $\rho(\lambda_1,\Phi) \approx 0$. Alors, par définition de λ_1', et du fait que $F \geq 1$, on a $\rho(\lambda_1,\lambda_1') \approx 0$ Il s'ensuit que $\rho(\lambda_1^*,\lambda_1') \approx 0$, donc $F(\lambda_1') \approx F(\lambda_1^*)$. En utilisant la relation 17.1, il vient donc que :

$$1 \leq \frac{\rho(\lambda_1,\lambda_1')}{\rho(\lambda_1,\Phi)} \approx 1.$$

Ainsi, $G(\lambda_1) \approx F(\lambda_1^*)$. De la même manière, $G(\lambda_2) \approx F(\lambda_2^*)$ (à moins que $\lambda_2 \in \Phi$, au quel cas $G(\lambda_2) = F(\lambda_2)$). On en déduit que $G(\lambda_1) \approx G(\lambda_2)$. Enfin, imaginons pour finir que $\rho(\lambda_1,\Phi) \gg 0$. La définition de G montre clairement que dans ce cas $G(\lambda_1) \approx G(\lambda_2)$.

Nous avons ainsi prouvé la continuité de G. Nous pouvons maintenant affirmer que $EG(\xi) \approx EG(\eta)$, les deux processus étant quasiment équivalents. Or, $E|F(\xi) - G(\xi)| \leq c\varepsilon$, de même que $E|F(\eta) - G(\eta)| \leq c\varepsilon$. Comme le réel $\varepsilon \gg 0$ est arbitraire, on en déduit $EF(\xi) \approx EF(\eta)$, d'où le théorème. $\square$

Chapitre 18

Théorème de de Moivre-Laplace-Lindeberg-Feller-Wiener-Lévy-Doob-Erdős-Kac-Donsker-Prokhorov

Soit ξ un processus stochastique indexé par un quasi-intervalle T (par exemple ξ peut être la martingale normalisée associée à une séquence de variables aléatoires, telle que nous l'avons définie au chapitre 14). Nous dirons que ξ est un *processus de Wiener (approché)* lorsque ξ est quasiment équivalent à la marche aléatoire de Wiener définie sur T. Ce qui suit est une version du théorème central limite de de Moivre-Laplace, qui contient :

- le fait, prouvé par Lindeberg, que la condition du théorème de Lindeberg est suffisante ;
- le fait, prouvé par Feller, que cette condition est également nécessaire ;
- le théorème de Wiener sur la continuité des processus de Wiener ;
- le théorème de Lévy-Doob qui caractérise ces processus comme étant les seules martingales normalisées dont les trajectoires sont continues ;
- le principe d'invariance de Erdős-Kac, étendu par Donsker et Prokhorov.

Théorème 18.1. *Soit ξ une martingale normalisée, indexée par un quasi-intervalle, telle que $\xi(a) = 0$. Les propriétés suivantes sont équivalentes :*

(i) *ξ est un processus de Wiener ;*

(ii) *presque toutes les trajectoires de ξ sont continues, et $\xi(b)$ est L^2 ;*

(iii) *ξ remplit la condition de Lindeberg.*

Démonstration. Nous allons montrer les implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i).

Supposons (i). D'après le corollaire du théorème (13.1), presque toutes les trajectoires de la marche aléatoire de Wiener quasiment équivalente à ξ sont continues, d'où l'on déduit, en accord avec le deuxième corollaire du théorème (17.3), que presque toutes les trajectoires de ξ sont continues sur toute leur durée. Montrons maintenant que $\xi(b)$ est L^2.

Définissons la fonction réelle f_c ainsi

$$f_c(\lambda) = \begin{cases} \lambda^2 & \text{si } |\lambda| \leq c\,; \\ c^2 & \text{si } |\lambda| > c. \end{cases}$$

On peut voir qu'une v.a.r. y est L^2 si et seulement si :

$$\forall \varepsilon \gg 0, \exists c \ll \infty, Ef_c(y) \geq Ey^2 - \varepsilon.[1]$$

Si $c \ll \infty$, f_c est limitée et continue, et si nous posons $F_c(\xi) = f_c(\xi(b))$, nous obtenons une fonctionnelle limitée et continue.

Soit w la marche aléatoire de Wiener quasiment équivalente à ξ. On se rend compte facilement que $w(b)$ est L^2. Ceci ressort par exemple des

[1]En effet, supposons que y soit L^2. Soit $\varepsilon \gg 0$. Nous avons :

$$\forall a \approx \infty, E\Big|y^2 - y^{2(a)}\Big| \leq \varepsilon,$$

donc :

$$\exists c \ll \infty, E\Big|y^2 - y^{2(c)}\Big| \leq \varepsilon.$$

Or :

$$E\big|y^2 - f_c(y)\big| \leq E\Big|y^2 - y^{2(c)}\Big|.$$

Inversement, si par hypothèse

$$\forall \varepsilon \gg 0, \exists c \ll \infty, E\big(y^2\big) \leq \varepsilon + Ef_c(y),$$

il est possible d'écrire (en choisissant $\varepsilon = 1$) $\exists c \ll \infty, E\big(y^2\big) \leq 1 + c^2$. De plus, soit un événement M de probabilité infinitésimale. Nous avons :

$$\begin{aligned} E\big(y^2\chi_M\big) &= E\big(y^2 - f_c(y)\big)\chi_M + Ef_c(y)\chi_M \\ &\leq \varepsilon PrM + c^2 PrM \approx 0, \end{aligned}$$

et le théorème de Radon-Nikodym (8.1) nous permet de conclure. (NdT) relations suivantes[2] :

$$\begin{aligned}E\Big(w(b)^4\Big) &= E\left(\sum_{t\in T'} dw(t)\right)^4 = E\sum_{(t,u,r,s)\in T'^4} dw(t)\,dw(u)\,dw(r)\,dw(s)\\ &= 3\sum_{\substack{(t,s)\in T'^2\\ t\neq s}} E\Big(dw(t)^2 dw(s)^2\Big) + \sum_{t\in T'} E\,dw(t)^4\\ &= 3\sum_{\substack{(t,s)\in T'^2\\ t\neq s}} dt\,ds + \sum_{t\in T'} dt^2 \approx 3(b-a)^2\\ &\ll \infty.\end{aligned}$$

Il vient donc que :

$$\forall\varepsilon \gg 0, \exists c \ll \infty, Ef_c(w(b)) \geq b-a-\varepsilon,$$

ainsi :

$$\forall\varepsilon \gg 0, \exists c \ll \infty, Ef_c(\xi(b)) \gtrsim b-a-\varepsilon = E\xi(b)^2 - \varepsilon,$$

ce qui montre que $\xi(b)$ est L^2. Nous venons de prouver (i) $\Rightarrow$ (ii).

Supposons (ii). ξ vérifie l'identité suivante :

$$\xi(b)^2 = 2\sum_{t\in T'} \xi(t)\,d\xi(t) + \sum_{t\in T'} d\xi(t)^2.$$

Or, $\xi(b)^2$ est L^1 par hypothèse. De même, le terme $2\sum_{t\in T'}\xi(t)\,d\xi(t)$ est L^1, comme le montre la majoration suivante de l'espérance de son carré par un réel appréciable :

$$\begin{aligned}E\left(2\sum_{t\in T'}\xi(t)\,d\xi(t)\right)^2 &= 4E\sum_{t\in T'} E_t\Big(\xi(t)^2\,d\xi(t)^2\Big)\\ &= 4E\sum_{t\in T'} \xi(t)^2 E_t d\xi(t)^2\\ &= 4E\sum_{t\in T'} \xi(t)^2\,dt = 4\sum_{t\in T'}(t-a)\,dt\\ &\approx 2(b-a)^2 \ll \infty.\end{aligned}$$

[2]Si x est une v.a.r. telle que $Ex^4 \ll \infty$, alors x est L^2. En effet :

$$\forall a \approx \infty, a\sum_{\lambda^2>a} \lambda^2 pr_x(\lambda) \leq \sum_{\lambda^2>a}\lambda^4 pr_x(\lambda) \ll \infty$$

ce qui montre que $\forall a \approx \infty, \sum_{\lambda^2>a}\lambda^2 pr_x(\lambda) \approx 0$. (NdT)

Ainsi, le terme $\sum_{t\in T'} d\xi(t)^2$, différence des deux termes précédents, est L^1, de même que la variable $\sum_{t\in T'} d\xi(t)^{(\varepsilon)2}$ pour toute valeur du réel ε.

L'hypothèse (ii) (continuité de presque toutes les trajectoires) entraîne que lorsque $\varepsilon \gg 0$, presque toutes les trajectoires vérifient l'égalité suivante :

$$\sum_{t\in T'} d\xi(t)^2 = \sum_{t\in T'} d\xi(t)^{(\varepsilon)2}.$$

Le théorème de Lebesgue nous permet donc d'écrire, lorsque $\varepsilon \gg 0$:

$$E\sum_{t\in T'} d\xi(t)^2 \approx E\sum_{t\in T'} d\xi(t)^{(\varepsilon)2},$$

ce qui est précisément la condition de Lindeberg. D'où nous concluons (ii)$\Rightarrow$(iii).

Supposons maintenant (iii). L'idée sur laquelle repose la démonstration que (iii)$\Rightarrow$(i) est simple. Observons sur la figure 18.1 une réalisation du processus ξ aux instants t_n, où ε est un réel infinitésimal "très grand". Cette trajectoire va varier entre deux instants t_n de $\sqrt{\varepsilon}$ ou de $-\sqrt{\varepsilon}$ environ, et du fait des propriétés des martingales, accroissements et décroissements surviendront à parts approximativement égales. Les instants t_n constituent des variables aléatoires, mais comme la variation quadratique de la trajectoire (qui vaut presque $n\varepsilon$) est à peu près égale au temps écoulé $(t-a)$, deux instants t_n consécutifs se comportent comme s'ils étaient espacés en moyenne de ε l'un de l'autre (ceci n'apparaît pas sur la figure 18.1), ainsi la trajectoire semble être issue d'un processus de Wiener.

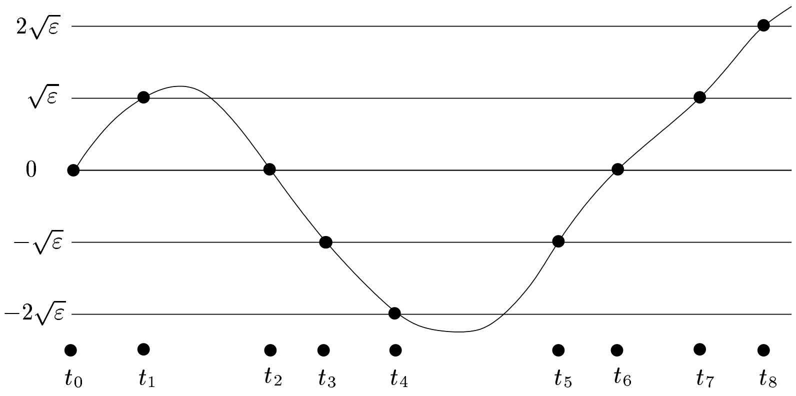

FIG. 18.1 – un processus de Wiener

Nous allons montrer que (iii)$\Rightarrow$(i) sous une l'hypothèse un peu plus faible qui suppose que ξ est *quasiment normalisé*, ce par quoi on entend que ξ vérifie la proposition suivante :

$$ps : \forall t \in T, \tau_\xi(t) \approx t - a.$$

Grâce aux théorèmes (14.3) et (17.2), on ne perd pas en généralité si l'on suppose :

$$\exists \alpha \approx 0, \forall (t,\omega) \in T' \times \Omega, |d\xi(\omega,t)| \leq \alpha.$$

Soit $\bar{b}$ tel que $\bar{b} - b \approx \infty$, et définissons l'ensemble $\overline{T}$ comme l'union de T et de l'ensemble de tous les réels de la forme :

$$b + k\frac{\bar{b}-b}{m}, k = 1, \cdots, m, \text{ où } \frac{\bar{b}-b}{m} \approx 0.$$

Nous généralisons ξ à tout l'ensemble $\overline{T}$ en posant $\xi(t) = \xi(b) + \bar{w}(t)$, $\bar{w}$ étant la marche aléatoire de Wiener définie sur $\overline{T} - T$. Le processus ainsi étendu conserve les propriétés que nous avons attribuées au processus original. L'objet de cette opération est d'éviter de se préoccuper du fait que les instants t_n peuvent cesser d'être définis pour le processus original.

Désignons par $t - d_*t$ le prédécesseur de t dans $\overline{T}$. Nous dirons que ξ *croise* λ à l'instant t dès lors que :

- soit $\xi(t - d_*t) < \lambda$ et $\xi(t) \geq \lambda$;
- soit $\xi(t - d_*t) > \lambda$ et $\xi(t) \leq \lambda$.

Soit $\varepsilon > 0$ tel que $\sqrt{\varepsilon} > 2\alpha$, et soit $\nu = \left[\frac{b-a}{\varepsilon}\right]$. Posons $t_0 = a$, et définissons récursivement t_n, pour $n = 1, \cdots, \nu$, comme étant le premier instant postérieur à t_{n-1} au quel le processus ξ croise un niveau $k_n\sqrt{\varepsilon}$ où k_n et un entier de même parité que n. Si un tel instant n'existe pas, on convient de poser $t_n = \bar{b}$. La figure 18.1 illustre cette définition.

Soit $\overline{\Phi}_n$ l'algèbre de v.a.r. engendrée par la famille $\{\xi(t_1), \cdots, \xi(t_n)\}$. Le processus $\psi_n = \xi(t_n)$ est une martingale pour cette filtration. Pour voir cela, choisissons un événement A tel que $\chi_A \in \overline{\Phi}_n$. Alors :

$$\chi_A(\psi_{n+1} - \psi_n) = \sum_{s \in T'} \eta(s)\, d\xi(s),$$

où η est un Φ-processus tel que $\eta(s) = 1$ si l'événement A est survenu et si $t_n \leq s < t_{n+1}$, et $\eta(s) = 0$ sinon. Ainsi, $E\chi_A(\psi_{n+1} - \psi_n) = 0$, et comme cette relation est vraie quel que soit le choix de A dans $\overline{\Phi}_n$, il vient que :

$$E_{\overline{\Phi}_n}(\psi_{n+1} - \psi_n) = 0,$$

ce qui établit la propriété des martingales.

Supposons que $0 \ll \varepsilon \ll \infty$. Alors, presque sûrement, l'une des deux inégalités suivantes est vérifiée pour tout $n < \nu$:

$$\left| \begin{array}{l} |\xi(t_{n+1}) - \xi(t_n) - \sqrt{\varepsilon}| \leq 2\alpha, \\ \text{ou} \\ |\xi(t_{n+1}) - \xi(t_n) + \sqrt{\varepsilon}| \leq \ 2\alpha. \end{array} \right.$$

En effet, la seule possibilité pour qu'aucune de ces inégalités ne soit vérifiée est celle où $t_{n+1} = \bar{b}$, mais comme $\big(\bar{b} - b\big) \approx \infty$, et que $\nu \ll \infty$ si $\varepsilon \gg 0$, on voit facilement que presque sûrement cette éventualité ne se réalise pas.

Remarquons de plus que l'on a toujours :

$$\forall n < \nu, |\xi(t_{n+1}) - \xi(t_n)| \leq \sqrt{\varepsilon} + 2\alpha \ll \infty.$$

Du fait que la moyenne de $(\xi(t_{n+1}) - \xi(t_n))$ est nulle, il nous est possible d'écrire :

$$\left\{ \begin{array}{l} Pr\Big\{\Big|\xi(t_{n+1}) - \xi(t_n) - \sqrt{\varepsilon}\Big| \leq 2\alpha\Big\} \approx 1/2, \\ Pr\Big\{\Big|\xi(t_{n+1}) - \xi(t_n) + \sqrt{\varepsilon}\Big| \leq 2\alpha\Big\} \approx 1/2. \end{array} \right.$$

Ces relations sont également valables pour les probabilités conditionnelles relativement à $\overline{\Phi}_n$. Comme $\varepsilon \gg 0$ implique que $\nu \ll \infty$, cela signifie que :

$$\sum_{\pi} \left| Pr\Big\{\forall n < \nu, \Big|\xi(t_{n+1}) - \xi(t_n) - (-1)^{\pi(n)}\sqrt{\varepsilon}\Big| \leq 2\alpha\Big\} - \frac{1}{2^\nu} \right| \approx 0, \quad (18.1)$$

où la somme est effectuée sur l'ensemble des applications de $\{0, \cdots, \nu - 1\}$ dans $\{0, 1\}$.

Ici le principe de Cauchy intervient. Comme l'ensemble :

$$\left\{ \varepsilon > 0 : \sum_{\pi} \left| Pr\Big\{\forall n < \nu, \Big|\xi(t_{n+1}) - \xi(t_n) - (-1)^{\pi(n)}\sqrt{\varepsilon}\Big| \leq 2\alpha\Big\} - \frac{1}{2^\nu} \right| \leq \varepsilon \right\}$$

contient tous les réels appréciables, il contient tout infinitésimal suffisamment grand. La relation 18.1 vaut donc encore pour des infinitésimaux assez grands. Maintenant choisissons ε (notre infinitésimal "très grand") tel que $\varepsilon \approx 0$, $\alpha/\varepsilon \approx 0$, et que 18.1 soit vérifié.

Soit la variable aléatoire $\nu(t)$ définie comme étant le plus grand entier $n \leq \nu$ tel que $t_n \leq t$. Définissons le processus stochastique ζ indexé par T de la manière suivante :

$$\zeta(t) = \sum_{n=1}^{\nu(t)} (\xi(t_n) - \xi(t_{n-1}))^2 \ + \big(\xi(t) - \xi\big(t_{\nu(t)}\big)\big)^2. \quad (18.2)$$

Alors :

$$d\zeta(t) = d\xi(t)^2 + 2\big(\xi(t) - \xi\big(t_{\nu(t)}\big)\big)\, d\xi(t),$$

si bien que, en posant $\gamma = \alpha + 2(\sqrt{\varepsilon} + \alpha) \approx 0$, nous avons la proposition suivante :

$$\forall(\omega, t) \in \Omega \times T', |d\zeta(\omega, t)| \leq \gamma |d\xi(t)|.$$

Par conséquent :

$$\left\| \sum_{t \in T'} d\zeta(t) - \sum_{t \in T'} E_t\, d\zeta(t) \right\|_2^2 \leq \sum_{t \in T'} \|d\zeta(t)\|_2^2 \leq \gamma^2 \sum_{t \in T'} \|d\xi(t)\|_2^2 \approx 0.$$

En vertu des théorèmes 11.1 et 7.1, nous avons :

$$ps : \forall t \in T, \zeta(t) \approx \sum_{s<t} E_s\, d\zeta(s) = \sum_{s<t} E_s d\xi(s)^2 \approx \tau_\xi(t) \approx t - a.$$

Comme $\alpha/\varepsilon \approx 0$, il vient que, presque sûrement, chaque terme de la somme figurant au sein de la relation 18.2 est asymptotique à ε. Le dernier terme de 18.2 étant équivalent à 0, nous avons :

$$ps : \forall t \in T, \zeta(t) \approx \nu(t)\varepsilon,$$

d'où :

$$ps : \forall t \in T, \nu(t)\varepsilon \approx t - a. \tag{18.3}$$

Soit Π l'espace probabilisé fini de toutes les applications π de l'ensemble $\{1, \cdots, \nu\}$ dans l'ensemble $\{0, 1\}$, la fonction de probabilité pr étant définie par :

$$\forall \pi \in \Pi, pr(\pi) = \frac{1}{2^\nu}.$$

Soit w_ε le processus stochastique défini sur $\langle \Pi, pr \rangle$ et indexé par T tel que :

$$w_\varepsilon(t) = \sum_{n \leq \frac{t-a}{\varepsilon}} (-1)^{\pi(n)} \sqrt{\varepsilon}.$$

Du fait des relations 18.1 et 18.3, et des théorèmes (17.1) et (17.2), ξ est quasiment équivalent à w_ε. Si ξ' est un autre processus stochastique satisfaisant nos hypothèses initiales (en particulier si ξ' est la marche aléatoire de Wiener indexée par T), alors pour un réel positif infinitésimal ε suffisamment grand ξ et ξ' sont tous les deux quasiment équivalents à w_ε. Les processus ξ et ξ' sont donc quasiment équivalents l'un à l'autre. ξ est ainsi un processus de Wiener, ce qui établit que (iii)$\Rightarrow$(i). □

Posons ici, arbitrairement, un point d'arrêt. Bien plus peut être écrit, et je souhaite qu'il se trouvera quelqu'un pour réaliser un ouvrage élémentaire sur les processus stochastiques dans la lignée de ce livre, comprenant également des exercices et des exemples d'applications.

Annexe

Introduction

Dans cette annexe, nous voulons montrer que les théorèmes de la théorie conventionnelle des probabilités peuvent être déduits de leurs analogues élémentaires grâce à des arguments que l'on qualifie habituellement de non-sens généralisé ; aucun raisonnement probabiliste n'apparaîtra ici. On illustre ainsi le fait que la théorie non standard élémentaire des processus stochastiques peut être utilisée pour obtenir des résultats conventionnels. Inversement, cela indique que la machinerie sophistiquée de la théorie classique et les stratagèmes de l'Analyse Non Standard, nécessaires pour prouver l'équivalence entre les résultats élémentaires et leurs équivalents classiques, n'apportent rien d'important aux probabilités : la théorie élémentaire a le même contenu scientifique que la théorie classique. Cette annexe est en quelque sorte une conclusion autodestructrice.

Nous supposons que le lecteur connaît la théorie classique de la mesure, ainsi que l'Analyse Non Standard dans la formulation de la Théorie des Ensembles Internes (T.E.I.) au sujet de laquelle il peut se reporter à [1] ainsi qu'à [5].[3]

Les axiomes élémentaires du chapitre 4 sont des résultats de la T.E.I. Les axiomes (1) et (2) découlent du principe de transfert, l'axiome (3) vient du principe d'idéalisation, et le principe d'induction externe (4) vient du principe de standardisation. Le principe d'extension généralisé (*5) est un cas particulier du principe de saturation ; on peut se reporter à [3]. Par conséquent, chaque théorème de ce texte est un théorème de la T.E.I.

[3] La bibliographie figure après cette annexe.

Processus élémentaires voisins

Tous les théorèmes des mathématiques classiques sont internes. Ainsi, si nous voulons démontrer un résultat classique portant sur un processus stochastique, par transfert nous pouvons supposer que le processus est standard. Ceci implique que l'ensemble des indices et l'espace probabilisé sur lesquels et défini le processus sont standard.

Soit ξ_0 un processus stochastique standard indexé par T_0, et défini sur $\langle \Omega_0, S_0, Pr_0 \rangle$. Nous définissons un *processus élémentaire voisin* comme un processus stochastique ξ indexé par T, sous-ensemble fini de T_0 contenant tous les éléments standard de T_0, défini sur $\langle \Omega_0, S_0, Pr_0 \rangle$ mais ne prenant qu'un nombre fini de valeurs distinctes, tel que, à l'exception d'un ensemble de mesure infinitésimale on observe :

$$\sum_{t \in T} |\xi(t) - \xi_0(t)| \approx 0, \tag{1}$$

et tel que si chaque $\xi_0(t)$ est L^p, (avec $p \geq 1$ standard, ou $p = \infty$), alors :

$$\sum_{t \in T} \|\xi(t) - \xi_0(t)\|_p \approx 0. \tag{2}$$

On peut remarquer que s'il existe une valeur p standard telle que chaque $\xi_0(t)$ soit L^p, alors (1) se déduit de (2) en vertu du théorème de Bienaymé-Tchebychev. Soit S la σ-algèbre engendrée par la famille $\{\xi(t) : t \in T\}$. L'ensemble S est une sous-algèbre booléenne de S_0. Soit Ω l'ensemble de tous les atomes de S de mesure strictement positive. Définissons sur Ω la fonction pr ainsi :

$$pr(\omega) \stackrel{def}{=} Pr_0(\omega).$$

L'espace $\langle \Omega, pr \rangle$ est un espace probabilisé fini. Nous pouvons considérer le processus ξ comme étant défini aussi bien sur $\langle \Omega, pr \rangle$ comme nous l'avons fait dans cet ouvrage, que sur $\langle \Omega_0, S_0, Pr_0 \rangle$ afin de pouvoir le comparer à ξ_0. Nous ne ferons pas de distinction entre ces deux définitions dans la suite.

Pour obtenir un processus élémentaire voisin, il nous suffit de ne conserver qu'un nombre illimité suffisamment grand de termes dans le développement décimal de $\xi_0(t)$ avec $t \in T$. Pour tout réel positif x, et tout entier naturel n, définissons le réel $x_{[n]}$ comme étant le plus grand réel inférieur ou égal à x, et tel que :

$$x_{[n]} = \sum_{k=-n}^{n} a_k 2^{-k},$$
$$\forall k \in \{-n, \cdots, n\}, a_k \in \{0, 1\}.$$

Pour $x < 0$, nous posons $x_{[n]} = -(-x)_{[n]}$.

Théorème A.1. *Soit ξ_0 un processus stochastique standard. Il existe un processus élémentaire voisin de ξ_0.*

Démonstration. L'existence de T découle du principe d'idéalisation. Soit $\varepsilon > 0$ un réel infinitésimal. Soit P l'ensemble des valeurs p de l'intervalle $[1, \infty]$ telles que chaque $\xi_0(t)$ soit L^p. Si P n'est pas vide, alors, par idéalisation, il contient un élément p_0 supérieur ou égal à tout élément standard de P. Du fait du théorème de la convergence dominée de Lebesgue, pour n suffisamment grand nous avons :

$$\sum_{t \in T} \left\| \xi_0(t)_{[n]} - \xi_0(t) \right\|_{p_0} \leq \varepsilon.$$

Posons $\xi(t) = \xi_0(t)_{[n]}, \forall t \in T$. Alors, la relation (2) est vérifiée pour toute valeur standard de p dans P. Ainsi que nous l'avons noté, ceci entraîne que la relation (1) est elle-même vérifiée.

Dans le cas où P est vide, posons :

$$N_n = \left\{ \omega \in \Omega_0, \max_{t \in T} |\xi_0(\omega, t)| \geq 2^n \right\}.$$

Choisissons un entier n suffisamment grand pour que :

- $Pr_0 N_n \leq \varepsilon$ (ce qui est possible du fait que la suite N_n décroît vers l'ensemble vide) ;
- n est supérieur au cardinal de T ;
- $n\, 2^{-n} \leq \varepsilon$.

Posons ici encore $\xi(t) = \xi_0(t)_{[n]}, \forall t \in T$. Alors, la relation (1) est vérifiée, car sauf sur N_n nous avons $|\xi(t) - \xi_0(t)| \leq 2^{-n}$. □

Il est visible que, par construction, les variables $\xi(t)$ sont indépendantes si les variables $\xi_0(t)$ le sont. Si ξ_0 est une martingale, il nous faut modifier un peu la méthode de construction de ξ afin d'obtenir une martingale élémentaire.

Théorème A.2. *Soit ξ_0 un processus stochastique standard indexé par un sous-ensemble de R, T_0, dont chaque composante est L^1, est qui est soit une surmartingale, soit une sousmartingale, soit une martingale. Alors, il existe un processus élémentaire voisin disposant des mêmes propriétés.*

Démonstration. Choisissons un sous-ensemble fini T de T_0, contenant tous les éléments standard de T_0. Nous allons mettre en œuvre les notations

$(T', dt, a, \ldots)$ que nous avons utilisées précédemment dans l'ouvrage. Supposons que ξ_0 soit adapté à une filtration Φ_0. Pour le choix d'un entier n, nous construisons une filtration Φ telle que Φ_t soit l'algèbre de Boole finie engendrée par $\left\{\xi_0(s)_{[n]} : s \in T \text{ et } s \leq t\right\}$ si bien que $\Phi_t \subseteq \Phi_{0t}$. Définissons le processus ξ ainsi :

$$\begin{aligned} \forall t \in T', d\xi(t) &= d\Big(\xi_0(t)_{[n]}\Big) + E\left\{d\xi_0(t) - d\Big(\xi_0(t)_{[n]}\Big) \mid \Phi_t\right\}, \\ \forall t \in T, \xi(t) &= \xi_0(a)_{[n]} + \textstyle\sum_{s<t} d\xi(s). \end{aligned}$$

Alors :

$$E\{d\xi(t) \mid \Phi_t\} = E\{d\xi_0(t) \mid \Phi_t\}.$$

Du fait que $\Phi_t \subseteq \Phi_{0t}$, cette espérance a le signe attendu ; c'est-à-dire que si ξ_0 est une surmartingale, une sousmartingale, ou une martingale, il en va de même pour ξ. Soit $\varepsilon > 0$ un réel infinitésimal, et soit p_0 défini de la même manière que dans la démonstration précédente. Si n est suffisamment grand, d'après le théorème de la convergence dominée de Lebesgue, on peut écrire :

$$\sum_{t \in T} \|\xi(t) - \xi_0(t)\|_{p_0} \leq \varepsilon.$$

$\square$

Ces preuves montrent que la valeur infinitésimale implicite dans la définition des processus élémentaires voisins peut être choisie aussi petite que voulue.

Equivalence des propriétés analytiques.

Un théorème typique de la théorie des processus stochastiques affirme que, moyennant certaines hypothèses analytiques, certaines conclusions probabilistes sont vérifiées par presque toutes les trajectoires. Pour obtenir un théorème classique à partir de son analogue élémentaire, il nous faut montrer que les hypothèses analytiques internes entraînent les hypothèses analytiques élémentaires analogues, et que les conclusions probabilistes élémentaires qui en découlent entraînent les conclusions probabilistes internes correspondantes. Mais pour nous convaincre du fait que les deux formulations expriment la même chose dans les deux langages, il nous faut montrer l'équivalence entre les deux formes de l'hypothèse, et entre les deux formes de la conclusion.

Dans cette section, nous allons vérifier que certaines propriétés analytiques internes d'un processus stochastique standard sont équivalentes aux propriétés analytiques externes correspondantes portant sur un processus élémentaire voisin. Les conventions d'écriture suivantes sont en usage : ξ_0 est un processus stochastique standard, indexé par T_0, et défini sur $\langle \Omega_0, S_0, Pr_0 \rangle$; la fonction ξ est un processus élémentaire voisin, indexé par T, et défini sur $\langle \Omega, pr \rangle$ comme nous l'avons déjà vu (et ξ est une surmartingale, ou autre, si ξ_0 lui même en est une) ; a et b sont respectivement les premier et dernier éléments de T, $T' = T - \{b\}$, pour tout t de T', $t + dt$ désigne le successeur de t dans T, et pour toute fonction f définie sur T, on pose :

$$\forall t \in T', df(t) = f(t + dt) - f(t).$$

Lorsque $T_0 = N^*$, nous choisirons T tel que $T = \{1, \cdots, \nu\}$ où ν désigne un entier naturel illimité. Remarquons que si T_0 est un intervalle fermé $[a, b]$, alors T est un quasi-intervalle dont le premier élément est a, et le dernier élément est b.

Théorème A.3. *Soit ξ_0 dont chaque composante est L^1. Si $T_0 = N^*$, alors :*

1) $\sum_{n=1}^{\infty} \|\xi_0(n)\|_1$ *converge si et seulement si* $\sum_{n=1}^{\nu} \|\xi(n)\|_1$ *converge (quasiment) ;*

2) $\xi_0(n)$ *converge dans L^1 si et seulement si $\xi(n)$ converge (quasiment) dans* $\mathbf{L}^1$.

Si $T_0 \subseteq R$, et si t_0 est un élément standard de T_0, alors :

3) ξ_0 *est continu en t_0 dans L^1 si et seulement si ξ est quasiment continu en t_0 dans* $\mathbf{L}^1$.

Si $T_0 \subseteq R$, alors :

4) ξ_0 *est à variation bornée dans L^1 si et seulement si* $\sum_{t \in T'} \|d\xi(t)\|_1 \ll \infty$.

Si T_0 est un intervalle fermé $[a, b]$, alors :

5) ξ_0 *appartient à $L^1(T_0 \times \Omega_0)$ si et seulement si ξ est L^1 sur $T' \times \Omega$;*

6) ξ_0 *est absolument continu dans L^1 si et seulement si il existe un processus élémentaire voisin tel que $\frac{d\xi}{dt}$ soit L^1 sur $T' \times \Omega$.*

Démonstration. Comme ξ est un processus élémentaire voisin, la conclusion de 1) est équivalente à la proposition :

$$\sum_{n=1}^{\nu} \|\xi_0(n)\|_1 \text{ converge (quasiment).} \tag{3}$$

De ce fait, 1) exprime le fait qu'une série standard converge si et seulement si sa somme partielle jusqu'à un indice illimité ν converge (quasiment), ce que l'on montre facilement. On démontre 2) et 3) de manière similaire.

Supposons que l'hypothèse de 4) soit vérifiée. Par transfert, cela implique l'existence d'un nombre standard K tel que pour tout sous-ensemble fini de T_0, et pour T en particulier, on a $\sum_{t\in T'} \|d\xi_0(t)\|_1 \le K$. La conclusion est alors vérifiée. Inversement, supposons que la conclusion de 4) soit vérifiée. Comme T contient tous les points standard de T_0, il vient qu'il existe une valeur standard fixe qui borne la variation de ξ_0 quel que soit le sous-ensemble standard fini de T_0. Par transfert, on obtient l'hypothèse de 4).

L'hypothèse de 5) est équivalente à la proposition suivante :

$$\lim_{\substack{n\to\infty \\ m\to\infty}} \left\|\xi_0^{(n)} - \xi_0^{(m)}\right\|_1 = 0. \tag{4}$$

Comme ξ_0 est standard, (4) est équivalent à :

$$\forall n \approx \infty, \forall m \approx \infty, \left\|\xi_0^{(n)} - \xi_0^{(m)}\right\|_1 \approx 0,$$

ce qui est en soit équivalent à la conclusion de 5) car ξ est un processus élémentaire voisin.

Pour toute collection finie I de sous-intervalles de $[a,b]$ ne se recouvrant pas, désignons par $|I|$ sa longueur totale, et par $var_0(I)$ la quantité

$$\sum_i \|\xi_0(b_i) - \xi_0(a_i)\|_1$$

où les $[a_i, b_i]$ sont les éléments de I. Avec ces notations, l'hypothèse de 6) devient :

$$\forall \varepsilon > 0, \exists \delta > 0, \forall I, (|I| \le \delta) \Rightarrow (var_0(I) \le \varepsilon), \tag{5}$$

mais ceci est équivalent à :

$$\forall I, (|I| \approx 0) \Rightarrow (var_0(I) \approx 0). \tag{6}$$

Pour s'en convaincre, appliquons l'algorithme de réduction à (6) de la manière suivante. Ecrivons (6) sous cette forme :

$$\forall I, \big(\forall^{st}\delta > 0, |I| \le \delta\big) \Rightarrow \big(\forall^{st}\varepsilon > 0, var_0(I) \le \varepsilon\big),$$

puis sous cette forme :

$$\forall^{st}\varepsilon, \forall I, \exists^{st}\delta, (|I| \le \delta) \Rightarrow (var_0(I) \le \varepsilon).$$

Utilisons le principe d'idéalisation pour dériver la forme équivalente suivante :

$$\forall^{st}\varepsilon, \exists^{st,fini}\delta', \forall I, \exists \delta \in \delta', (|I| \leq \delta) \Rightarrow (var_0(I) \leq \varepsilon).$$

En choisissant le plus petit élément δ dans δ', et en utilisant le principe de transfert, on obtient (5).

Disons que I est *satisfaisant* lorsque les extrémités de chacun de ses éléments sont dans T.

Lemme. (6) *est équivalent à :*

$$\forall^{satisfaisant} I, (|I| \approx 0) \Rightarrow (var_0(I) \approx 0). \tag{7}$$

Il est clair que (6) implique (7). En appliquant l'algorithme de réduction à (7), nous voyons que (7) est équivalent à :

$$\forall^{st}\varepsilon > 0, \exists^{st}\delta > 0, \forall^{satisfaisant} I, (|I| \leq \delta) \Rightarrow (var_0(I) \leq \varepsilon).$$

Comme T contient tous les éléments standard de $[a,b]$, cela entraîne :

$$\forall^{st}\varepsilon > 0, \exists^{st}\delta > 0, \forall^{st} I, (|I| \leq \delta) \Rightarrow (var_0(I) \leq \varepsilon),$$

ce qui est, par transfert, équivalent à (5), donc à (6). Le lemme est ainsi prouvé.

(5) et (7) sont donc équivalents. Pour un ensemble I satisfaisant, définissons la quantité $var(I)$ comme la quantité $\sum_i \|\xi(b_i) - \xi(a_i)\|_1$. Comme ξ est un processus voisin, (7) est équivalent à :

$$\forall^{satisfaisant} I, (|I| \approx 0) \Rightarrow (var(I) \approx 0).$$

Mais cette proposition exprime le fait que pour des ensembles de la forme $M = I \times \Omega$, si la probabilité de M dans $T' \times \Omega$ est infinitésimale alors :

$$E\left|\frac{d\xi}{dt}\right|\chi_M \approx 0,$$

ce qui est vérifié dès que $\frac{d\xi}{dt}$ est L^1 dans $T' \times \Omega$ d'après le théorème 8.1. Inversement, on remarque que l'hypothèse de 6) est équivalente à l'existence d'une fonction standard η_0 dans $L^1(T_0 \times \Omega_0)$ telle que presque sûrement :

$$\forall t \in T_0, \xi_0(t) = \xi_0(a) + \int_a^t \eta_0(s)\, ds.$$

Soit η un processus élémentaire voisin de η_0. η est L^1 dans $T' \times \Omega$ d'après 5). Choisissons $\xi(a)$ comme étant une variable aléatoire élémentaire voisine de $\xi_0(a)$ (cela revient à faire le choix d'un processus élémentaire voisin du processus stochastique standard indexé par le singleton $\{a\}$ et dont la valeur en a est $\xi_0(a)$). Construisons le processus ξ ainsi :

$$\xi(t) = \xi(a) + \sum_{a \leq s < t} \eta(s)\, ds.$$

On peut facilement se rendre compte que ξ est un processus voisin de ξ_0, et que $\eta = \frac{d\xi}{dt}$ est L^1 dans $T' \times \Omega$. □

Théorème A.4. *Soit ξ_0 un processus dont chaque composante est L^1, et supposons que $T_0 \subseteq R$. Si ξ_0 est une martingale ou une sousmartingale positive, alors $\|\xi(a)\|_1 \ll \infty$, et de plus $\|\xi(b)\|_1 \ll \infty$ si et seulement si la fonction $t \mapsto \|\xi_0(t)\|_1$ est bornée. Si ξ_0 est une surmartingale positive alors on a $\|\xi(b)\|_1 \ll \infty$, et de plus $\|\xi(a)\|_1 \ll \infty$ si et seulement si la fonction $t \mapsto \|\xi_0(t)\|_1$ est bornée.*

Démonstration. Dans la première hypothèse, $\|\xi_0(t)\|_1$ et $\|\xi(t)\|_1$ sont des fonctions croissantes de t, donc $\|\xi(a)\|_1 \ll \infty$. Si $\|\xi(b)\|_1 \ll \infty$, il existe un réel standard K tel que :

$$\forall t \in T, \|\xi(t)\|_1 \leq K,$$

si bien que l'on a :

$$\forall^{st} t \in T_0, \|\xi_0(t)\|_1 \leq K + 1,$$

donc par transfert $t \mapsto \|\xi_0(t)\|_1$ est bornée. Inversement, si $t \mapsto \|\xi_0(t)\|_1$ est bornée, elle a une borne standard K, si bien que $\|\xi_0(b)\|_1 \leq K$, et que $\|\xi(b)\|_1 \ll \infty$. La démonstration dans le cadre de la deuxième hypothèse (les normes sont alors décroissantes), se déroule de manière strictement similaire. □

Mesures de probabilité régulières

Soit η_0 un processus stochastique indexé par l'ensemble T_0. Lorsque T_0 n'est pas dénombrable, des complications apparaissent au sein de la théorie de la mesure. Afin de remédier à la plupart de ces complications, nous considérerons toujours la version *canonique* ξ_0 du processus (voir [4]), qui est un processus indexé par T_0, et équivalent à η_0. Il est défini sur l'espace des trajectoires :

$$\Omega_0 = \prod_{t \in T_0} \dot{R},$$

où $\dot{R}$ est le compactifié de R obtenu en lui ajoutant un point à l'infini. Alors, Ω_0 est un espace compact pour la topologie produit. Soit S_0 la σ-algèbre des ensembles boréliens de Ω_0, et soit Pr_0 l'unique mesure de probabilité régulière telle que le processus stochastique ξ_0 défini sur $\langle \Omega_0, S_0, Pr_0 \rangle$ par la relation :

$$\forall t \in T_0, \forall \omega_0 \in \Omega_0, \xi_0(t, \omega_0) = \omega_0(t),$$

a les mêmes distributions discrètes jointes que η_0. Si η_0 est standard, il en va de même pour ξ_0.

L'un des avantages de la version canonique est que la plupart des sous-ensembles intéressants de l'espace des trajectoires sont des boréliens, et qu'ils sont de ce fait mesurables.

Nous aurons besoin de deux résultats internes concernant les mesures de probabilité régulières. Par définition, un ensemble D est dit *filtrant supérieurement* s'il est muni d'une relation binaire transitive $\prec$ telle que tout sous-ensemble fini de D possède une borne supérieure. Un *réseau* est une fonction $F \mapsto \Theta_F$ définie sur D. Lorsque les Θ_F sont des ensembles, on dira que le réseau est *croissant* si :

$$(F_1 \prec F_2) \Longrightarrow (\Theta_{F_1} \subseteq \Theta_{F_2}),$$

et on dira qu'il est *décroissant* si :

$$(F_1 \prec F_2) \Longrightarrow (\Theta_{F_2} \subseteq \Theta_{F_1}).$$

L'interêt du théorème suivant est qu'il s'applique à des familles non dénombrables d'ensembles.

Théorème A.5. *Soit Pr_0 une mesure de probabilité régulière définie sur un espace compact Ω_0. Soit $F \mapsto \Theta_F$ un réseau décroissant de sous-ensembles fermés de Ω_0, et $G \mapsto \Gamma_G$ un réseau croissant de sous-ensembles ouverts de Ω_0. Alors :*

$$Pr_0 \bigcap_{F \in D} \Theta_F = \inf_{F \in D} Pr_0 \Theta_F, \tag{8}$$

$$Pr_0 \bigcup_{G \in D} \Gamma_G = \sup_{G \in D} Pr_0 \Gamma_G. \tag{9}$$

Démonstration. Soit $\varepsilon > 0$. Par définition de la régularité de la mesure, il existe un sous-ensemble compact $\Theta \subseteq \bigcup_{G \in D} \Gamma_G$ tel que :

$$Pr_0 \left(\bigcup_{G \in D} \Gamma_G \setminus \Theta \right) \leq \varepsilon.$$

Du fait de la compacité de Θ, il existe une famille finie $\{G_1, \cdots, G_n\}$ telle que :

$$\Theta \subseteq \bigcup_{i=1}^{n} \Gamma_{G_i}.$$

Soit G_0 une borne supérieure de cette famille. Il vient que $\Theta \subseteq \Gamma_{G_0}$, si bien que :

$$Pr_0\left(\bigcup_{G\in D} \Gamma_G \ \setminus \Gamma_{G_0}\right) \leq \varepsilon.$$

Comme ε est arbitraire, cela prouve (9). (8) s'en déduit par dualité. □

Théorème A.6. *Soit Pr_0 une mesure de probabilité régulière sur un espace compact Ω_0. Soient Θ_{jkF} des sous-ensembles fermés de Ω_0, et soient Γ_{kjG} des sous-ensembles ouverts de Ω_0, où j et k parcourent N, et F et G parcourent un ensemble filtrant supérieurement D. Supposons que Θ_{jkF} soit une fonction décroissante en j et en F, et que Γ_{kjG} soit une fonction croissante en k et en G. Alors :*

$$Pr_0 \bigcap_j \bigcup_k \bigcap_F \Theta_{jkF} = \sup_{\tilde{k}} \inf_{j,F} Pr_0 \Theta_{j\tilde{k}(j)F}, \tag{10}$$

où $\tilde{k}$ parcourt l'ensemble des fonctions de N dans N, et :

$$Pr_0 \bigcup_k \bigcap_j \bigcup_G \Gamma_{kjG} = \sup_{k,\tilde{G}} \inf_{j} Pr_0 \Gamma_{kj\tilde{G}(j)}, \tag{11}$$

où $\tilde{G}$ parcourt l'ensemble des fonctions de N dans D.

Démonstration. Posons :

$$p = Pr_0 \bigcap_j \bigcup_k \bigcap_F \Theta_{jkF}.$$

Soit $\varepsilon > 0$. Alors :

$$\forall j \in N, \exists k \in N, Pr_0 \bigcap_{F\in D} \Theta_{jkF} \geq p - \varepsilon,$$

d'où l'on déduit qu'il existe une fonction $\tilde{k}$ telle que :

$$\forall j \in N, Pr_0 \bigcap_{F\in D} \Theta_{j\tilde{k}(j)F} \geq p - \varepsilon.$$

En vertu du théorème précédent, nous avons :

$$\inf_{j,F} Pr_0 \Theta_{j\tilde{k}(j)F} \geq p - \varepsilon.$$

Comme ε est arbitraire, cela prouve l'inégalité $\leq$ dans la relation (10). Mais du fait que :

$$\bigcap_j \bigcup_k \bigcap_F \Theta_{jkF} = \bigcup_{\tilde{k}} \bigcap_{j,F} \Theta_{j\tilde{k}(j)F},$$

l'inégalité $\geq$ est triviale.

Maintenant, posons :

$$p = Pr_0 \bigcup_k \bigcap_j \bigcup_G \Gamma_{kjG},$$

et soit $\varepsilon > 0$. Alors :

$$\exists k \in N, \forall j \in N, \exists G \in D, Pr_0 \Gamma_{kjG} \geq p - \varepsilon,$$

en vertu du théorème précédent. Donc il existe k et $\tilde{G}$ tels que :

$$\forall j \in N, Pr_0 \Gamma_{kj\tilde{G}(j)} \geq p - \varepsilon.$$

Comme ε est arbitraire, cela prouve l'inégalité $\leq$ dans (11), et ici encore, l'inégalité $\geq$ est triviale. $\square$

Equivalence entre propriétés probabilistes

Considérons une fonction standard $\xi_0 : T_0 \to R$. La notion de fonction élémentaire voisine ξ est claire : c'est le cas particulier d'un processus stochastique élémentaire voisin dans le cas où l'espace probabilisé sous-jacent est réduit à un singleton. Nous rencontrons souvent des couples de propriétés (A, B), A étant interne et B externe, tels que A est vérifiée pour ξ_0 si et seulement si B l'est pour ξ. Par exemple, A peut désigner la propriété d'être continue en un point standard t_0, et B peut designer la propriété d'être quasiment continue en ce même point t_0. Cependant, si ξ_0 n'est pas standard, l'équivalence de A et B n'est plus forcément vérifiée. Imaginons maintenant que ξ_0 soit un processus stochastique standard dont ξ est un processus élémentaire voisin, et demandons-nous si le fait que A soit vraie presque partout pour ξ_0 équivaille au fait que B soit vraie presque partout pour ξ. Nous devons d'abord nous assurer que l'ensemble sur lequel A est vraie est mesurable, sinon la question que nous nous posons n'a pas de sens.

Même lorsque c'est le cas cependant, un raisonnement basé sur un examen individuel des trajectoires ne fonctionnerait pas. En effet, il existe un ensemble fini contenant toutes les trajectoires standard de ξ_0, et pour la majorité des processus stochastiques intéressants tout sous-ensemble fini de trajectoires est négligeable (c'est-à-dire que la réalisation d'un processus standard est généralement non-standard !). On en déduit donc que l'équivalence entre A pour une fonction standard et B pour une fonction élémentaire voisine ne nous aide pas à répondre à notre question.

Théorème A.7. *Soit ξ_0 un processus stochastique standard dans sa version canonique. Si $T_0 = N^*$, alors :*

1) $(ps : \sum_{n=1}^{\infty} |\xi_0(n)| < \infty) \Leftrightarrow (ps : \sum_{n=1}^{\nu} |\xi(n)| \ll \infty)$ *;*

2) $(ps : \sum_{n=1}^{\infty} |\xi_0(n)| < \infty) \Leftrightarrow (ps : \sum_{n=1}^{\nu} |\xi(n)|$ *converge quasimement*$)$.

Si t_0 est un point standard de T_0, et si $T_0 \subseteq R$, alors :

3) *(ps : ξ_0 est continue en t_0)$\Leftrightarrow$(ps : ξ est quasiment continue en t_0).*

Si T_0 est un sous-ensemble compact de R, alors :

4) *(ps : ξ_0 est continue)$\Leftrightarrow$(ps : ξ est quasiment continue),*

5) *(ps : ξ_0 n'a pas de discontinuités du type "oscillations")$\Leftrightarrow$(ps : ξ est à fluctuation limitée).*

Si $T_0 \subseteq R$, alors :

6) *(ps : ξ_0 est à variation bornée) $\Leftrightarrow$(ps : ξ est à variation limitée).*

Quel que soit T_0 :

7) $(ps : \xi_0 = 0) \Leftrightarrow (ps : \xi \approx 0)$.

Démonstration. Pour $i = 1, \cdots, 7$, l'équivalence i) est de la forme $A_i \Leftrightarrow B_i$ où A_i est une formule interne et B_i est une formule externe.

Par définition, il existe un sous-ensemble N de l'espace des trajectoires Ω_0 tel que $Pr_0 N \approx 0$, et tel que sur N^c on a :

$$\sum_{t \in T} |\xi_0(t) - \xi(t)| \approx 0.$$

Supposons qu'il soit faux que A_1 soit vérifiée presque partout. Posons :

$$\begin{aligned} \Theta^1_{jk} &= \left\{ \sum_{n=1}^{k} |\xi_0(n)| \geq j \right\}, \\ \Theta^1 &= \bigcap_j \bigcup_k \Theta^1_{jk}. \end{aligned}$$

Alors $Pr_0\Theta^1 > 0$ et par transfert on sait qu'il existe un réel standard $\varepsilon > 0$ tel que $Pr_0\Theta^1 > \varepsilon$. Grâce au théorème (A.6) et au principe de transfert, il existe une fonction standard $\tilde{k}$ telle que $Pr_0\Theta^1_{\tilde{k}} > \varepsilon$, avec :

$$\Theta^1_{\tilde{k}} = \bigcap_j \Theta^1_{j\tilde{k}(j)}.$$

Si j est standard $\tilde{k}(j)$ l'est également si bien que $\tilde{k}(j) < \nu$. D'où il vient que sur $\Theta^1_{\tilde{k}}$ on a :

$$\forall^{st} j, \sum_{n=1}^{\nu} |\xi_0(n)| > j,$$

et donc :

$$\sum_{n=1}^{\nu} |\xi_0(n)| \approx \infty.$$

Ainsi, sur $N^c \bigcap \Theta^1_{\tilde{k}}$ on a :

$$\sum_{n=1}^{\nu} |\xi(n)| \approx \infty.$$

Or, $Pr_0\left(N^c \bigcap \Theta^1_{\tilde{k}}\right) \gtrsim \varepsilon$, et comme cela est vrai pour un certain réel standard $\varepsilon > 0$, il est faux de dire que B_1 est vérifiée presque partout. Ainsi, si B_1 est vérifiée presque partout, A_1 est vérifiée presque partout.

A fortiori, B_2 presque partout implique A_2 (qui est identique à A_1) presque partout.

Supposons qu'il soit faux que A_3 soit vérifiée presque partout. Pour tout sous-ensemble fini G de T_0, posons :

$$\begin{aligned} \Gamma^3_{kjG} &= \bigcup_{\substack{s \in G \\ |s-t_0| \le \frac{1}{j}}} \left\{ |\xi_0(s) - \xi_0(t_0)| > \frac{1}{k} \right\}, \\ \Gamma^3 &= \bigcup_k \bigcap_j \bigcup_G \Gamma^3_{kjG}. \end{aligned}$$

Γ^3 est l'ensemble des trajectoires discontinues en t_0. Remarquons que c'est un borélien, car l'union non dénombrable sur G est une union d'ouverts. Alors, il existe un réel standard $\varepsilon > 0$ tel que $Pr_0\Gamma^3 > \varepsilon$. Par le théorème (A.6) et le principe de transfert, on peut trouver k standard, ainsi que $\tilde{G}$ standard, tels que $Pr_0\Gamma^3_{k\tilde{G}} > \varepsilon$, où :

$$\Gamma^3_{k\tilde{G}} = \bigcap_j \Gamma^3_{kj\tilde{G}(j)}.$$

Lorsque j est standard, $\tilde{G}(j)$ est un ensemble fini standard, donc chacun de ses éléments est standard. On a donc dans ce cas $\tilde{G}(j) \subseteq T$. Donc, sur $\Gamma^3_{k\tilde{G}}$, pour tout entier standard j il existe un point s dans T tel que $|s - t_0| \leq \frac{1}{j}$ et $|\xi_0(s) - \xi_0(t_0)| > \frac{1}{k}$, si bien que, par le principe de Cauchy, il existe un point s dans T tel que $s \approx t_0$ et $|\xi_0(s) - \xi_0(t_0)| > \frac{1}{k}$. Par conséquent, sur $N^c \bigcap \Gamma^3_{k\tilde{G}}$, le processus ξ n'est pas quasiment continu en t_0, et il est faux de dire que B_3 est vrai presque partout. Ainsi, B_3 presque partout implique A_3 presque partout.

On procède de manière similaire, en se contentant de remplacer la valeur fixe t_0 par une valeur variable t, pour montrer que B_4 presque partout implique A_4 presque partout.

La preuve que B_5 presque partout implique A_5 presque partout est similaire. Pour tout sous-ensemble G de T_0, constitué des j points $t_1 < \cdots < t_j$, posons :

$$\begin{aligned} \Gamma^5_{kjG} &= \bigcap_{i=1}^{j-1} \left\{ |\xi_0(t_i) - \xi_0(t_{i+1})| > \frac{1}{k} \right\}, \\ \Gamma^5 &= \bigcup_k \bigcap_j \bigcup_G \Gamma^5_{kjG}, \end{aligned}$$

et mettons en œuvre les mêmes arguments que précédemment. Notons que le borélien Γ^5 est l'ensemble des trajectoires possédant une discontinuité du type "oscillations".

La preuve que B_6 presque partout implique A_6 presque partout est également similaire. Pour tout sous-ensemble fini G de T_0, désignons par G' l'ensemble G privé de son plus grand élément, et pour tout point t dans G', désignons par $t + dt$ le successeur de t dans G. Posons :

$$\begin{aligned} \Gamma^6_{jG} &= \left\{ \sum_{t \in G'} |\xi_0(t + dt) - \xi_0(t)| > j \right\}, \\ \Gamma^6 &= \bigcap_j \bigcup_G \Gamma^6_{jG}, \end{aligned}$$

et procédons comme précédemment.

Imaginons qu'il soit faux que A_7 presque partout. Alors, pour tout sous-ensemble fini G de T_0, posons :

$$\begin{aligned} \Gamma^7_{kG} &= \left\{ \max_{t \in G} |\xi_0(t)| > \frac{1}{k} \right\}, \\ \Gamma^7 &= \bigcup_k \bigcup_G \Gamma^7_{kG}. \end{aligned}$$

Alors, il existe un réel standard $\varepsilon > 0$ tel que $Pr_0\Gamma^7 > \varepsilon$, et donc un entier standard k ainsi qu'un sous-ensemble fini standard G tels que $Pr_0\Gamma^7_{kG} > \varepsilon$. On a $G \subseteq T$, et sur l'ensemble Γ^7_{kG} le processus ξ n'est pas infinitésimal le long de T. On n'a donc pas B_7 presque partout. On en déduit que B_7 presque partout implique A_7 presque partout.

Considérons maintenant le sens inverse. Supposons A_2 presque partout, et posons :

$$\begin{aligned} \Theta^2_{jk} &= \left\{ \sum_{n=k}^{\infty} |\xi_0(n)| \leq \frac{1}{j} \right\}, \\ \Theta^2 &= \bigcap_j \bigcup_k \Theta^2_{jk}. \end{aligned}$$

Alors $Pr_0\Theta^2 = 1$. Soit $\varepsilon > 0$ un réel standard. En vertu du théorème (A.6), et du principe de transfert, il existe une fonction standard $\tilde{k}$ qui vérifie $Pr_0\Theta^2_{\tilde{k}} \geq 1 - \varepsilon$, en notant :

$$\Theta^2_{\tilde{k}} = \bigcap_j \Theta^2_{j\tilde{k}(j)}.$$

Au sein de $\Theta^2_{\tilde{k}}$ on a, pour tout j standard :

$$\sum_{n=\tilde{k}(j)}^{\infty} |\xi_0(n)| \leq \frac{1}{j},$$

et donc, au sein de $N^c \bigcap \Theta^2_{\tilde{k}}$:

$$\sum_{n=\tilde{k}(j)}^{\nu} |\xi(n)| \lesssim \frac{1}{j}.$$

Du fait que $\tilde{k}(j)$ est standard dès que j est standard, on a pour tout entier illimité $\mu \leq \nu$:

$$\sum_{n=\mu}^{\nu} |\xi(n)| \approx 0,$$

ce qui signifie que $\sum_{n=1}^{\nu} |\xi(n)|$ converge quasiment. On a donc B_2 presque partout. Ainsi, A_2 presque partout implique B_2 presque partout.

A fortiori, A_1 presque partout implique B_1 presque partout.

Supposons que A_3 soit vraie presque partout, et posons :

$$\begin{aligned}\Theta^3_{jk} &= \bigcap_{|s-t_0|\leq \frac{1}{k}} \left\{ |\xi_0(s) - \xi_0(t_0)| \leq \frac{1}{j} \right\},\\ \Theta^3 &= \bigcap_j \bigcup_k \Theta^3_{jk}.\end{aligned}$$

Alors, $Pr_0\Theta^3 = 1$. Soit $\varepsilon > 0$ un réel standard. Alors, il existe une fonction standard $\tilde{k}$ telle que $Pr_0\Theta^3_{\tilde{k}} \geq 1 - \varepsilon$, en notant :

$$\Theta^3_{\tilde{k}} = \bigcap_j \Theta^3_{j\tilde{k}(j)}.$$

Au sein de $N^c \bigcap \Theta^3_{\tilde{k}}$ le processus ξ est quasiment continu en t_0, et donc A_3 presque partout implique B_3 presque partout.

Les autres cas se démontrent de manière similaire. $\square$

Bibliographie

[1] Edward Nelson, *Internal set theory : A new approach to nonstandard analysis*, Bull. Amer. Math. Soc. 83 (1977), 1165-1198.

[2] –, *Théorie des ensembles internes : une nouvelle approche de l'analyse non standard.* Traduction de [1] extraite de [6].

[3] –, *The syntax of nonstandard analysis*, Ann. of Pure and Appl. Logic.

[4] –, *Regular probability measures on function space*, Ann. of Math. 69 (1959), 630-643.

[5] Francine Diener, Georges Reeb, *Analyse Non Standard*, Hermann (1989)

[6] Hervé Barreau, Jacques Harthong, *La mathématique non standard*, Editions du CNRS (1989)

[7] Claude Lobry, *et pourtant... ils ne remplissent pas N !* ALEAS Editeur (1989)

[8] Jean-Michel Salanskis, *Le constructivisme non standard*, Presses universitaires du Septentrion (1999)

[9] Patrick Billingsley *Convergence of probability measures.* Second edition, Wiley & Sons, Inc., New York (1999)

Index

Table des figures