

ENS de CACHAN

COURS ET TRAVAUX DIRIGÉS DE PROBABILITÉS
PRÉPA AGRÉG

CÉDRIC BERNARDIN ET JEAN-MICHEL MOREL

Table des matières

1	La construction des espaces de probabilités	9
1.1	Algèbres, tribus, π -systèmes, λ -systèmes et espaces mesurables	9
1.2	Mesure	11
1.2.1	Complétion d'un espace mesuré	11
1.2.2	Événements	12
1.2.3	Problème posé	12
1.3	Unicité des extensions	12
1.4	Théorème de Carathéodory	12
1.5	Mesures complexes	16
1.6	Mesures de Borel	17
1.7	Absolue continuité	17
1.8	Algèbres, tribus et compagnie	18
2	Variables aléatoires	23
2.1	Définitions	23
2.2	σ -algèbre générée par une collection de variables aléatoires	24
2.3	Distributions et fonctions de répartition	25
2.4	Variables aléatoires à densités et variables aléatoires discrètes	26
2.5	TD Variables aléatoires	26
3	Espérance	29
3.1	Construction de l'intégrale et théorèmes de convergence	29
3.1.1	La machine standard	29
3.1.2	Théorèmes de convergence	31
3.2	Inégalités de Markov	31
3.3	Sommes de variables aléatoires	31
3.4	Inégalité de Jensen	31
3.5	Monotonie des normes \mathcal{L}^p	32
3.6	Hölder et Minkowski	33
3.7	Cas $p = 2$	33
3.8	T.D. Intégration et Lebesgue	34
3.9	Rappel des principaux théorèmes, énoncés dans \mathbb{R}^N	44

4	Produit de mesures	47
4.1	Mesure produit, Théorème de Fubini	48
4.1.1	Théorème de Fubini	48
4.2	Tribus et mesures produits pour une infinité d'espaces	50
4.3	TD Probabilités : Fubini	52
5	Indépendance	55
5.1	Indépendance de familles d'événements et de variables aléatoires	55
5.2	Critères d'indépendance	56
5.3	Indépendance généralisée	58
5.4	Indépendance et événements asymptotiques	59
5.5	Lemmes de Borel-Cantelli, loi du 0-1, mesurabilité relative	60
6	Lois des grands nombres	67
6.1	La preuve la plus rapide de la loi forte des grands nombres	73
7	Les convergences	75
7.1	Convergence des mesures bornées sur \mathbb{R}^d	75
7.1.1	Topologies vague, faible et étroite	75
7.1.2	Tension	78
7.2	Rappel sur les différents types de convergence	79
7.3	Convergence en loi	79
7.4	TD Probabilités : Modes de convergence	81
8	Pile ou face, marches aléatoires	85
8.1	Propriétés élémentaires	86
8.2	Le principe de réflexion et ses conséquences	94
8.3	Lois de l'arc sinus pour les marches aléatoires	97
9	Fonctions génératrices, poules et grands noms	101
9.1	Espérance conditionnelle, poules et oeufs	101
9.2	Fonctions génératrices	104
9.3	Processus de branchement : l'extinction des grands noms	107
10	Fonctions caractéristiques et théorèmes de Lévy	109
10.1	Généralités	109
10.2	Injectivité de la transformée de Fourier	110
10.3	Le Théorème de convergence de Lévy	113
10.4	Fonctions caractéristiques : ce qu'il faut absolument savoir	115
11	Les gaussiennes	117
12	Théorème limite central	121
13	Grandes déviations	127

14	Espérance conditionnelle	131
14.1	Introduction : variables aléatoires discrètes	131
14.2	Théorème d'existence et d'unicité	132
14.3	Variables à densité	133
14.4	Variables discrètes	134
14.5	Noyaux et lois conditionnelles	134
14.6	TD Probabilités : Espérance conditionnelle	136
15	Martingales et inégalités maximales	143
15.1	Williams ou Ouvrard ?	143
15.2	Martingales et temps d'arrêt	144
15.3	Martingales	144
15.3.1	Filtrations et processus adaptés	144
15.3.2	Martingale, Sur-martingale, Sous-martingale	145
15.3.3	Exemples	145
15.3.4	Illustration	146
15.3.5	Temps d'arrêt	147
15.3.6	Sur-martingales arrêtées	147
15.3.7	Théorème d'arrêt de Doob	148
15.4	TD : Jeux, ruine, temps d'arrêt	150
15.5	Le théorème de convergence des martingales	153
16	Martingales bornées dans \mathcal{L}^2.	157
16.1	Le théorème de base de convergence des martingales de \mathcal{L}^2	157
16.1.1	Martingales de \mathcal{L}^2 : orthogonalité des accroissements	157
16.1.2	Décomposition de Doob	158
16.1.3	Le crochet d'une martingale $\langle M \rangle$	159
16.1.4	Application	160
16.2	TD Probabilités : Inégalités maximales	161
17	Martingales uniformément intégrables	167
17.1	L'uniforme intégrabilité	167
17.1.1	Les espérances conditionnelles d'une même variable X sont équiintégrables	168
17.1.2	Convergence en probabilité	169
17.1.3	Théorème de convergence bornée	169
17.1.4	Condition nécessaire et suffisante de convergence L^1	169
17.2	Martingales UI	170
17.2.1	Théorème "montant" de Lévy.	171
17.2.2	La loi du 0-1 par martingale	171
17.2.3	Théorème de Lévy "descendant"	172
17.2.4	Preuve de la LGN par martingale	172
17.3	TD probabilité : Deuxième théorème d'arrêt : [26] p. 391	173
17.4	TD Probabilités : Optimisation d'une vente ([18] p. 511)	174
17.5	TD Probabilités : Théorème de Radon-Nykodim	178

18 Théorie financière	181
18.1 Le problème des options	181
18.2 Le formalisme des modèles discrets	182
18.3 Les stratégies	182
18.4 Marché viable	182
18.5 Marché complet et évaluation des options	183
18.6 Le modèle de Cox, Ross et Rubinstein	184
18.7 Enveloppes de Snell et options américaines	185
19 Chaînes de Markov	191
19.1 Analyse bibliographique brève	191
19.2 Préliminaires	192
19.2.1 Théorème de Kolmogorov	192
19.3 Temps d'arrêt et tribu des événements antérieurs	192
19.4 Chaînes de Markov : la construction	193
19.4.1 Matrices de transition	193
19.4.2 Conséquences de la relation (19.4.1)	195
19.5 Construction de l'espace canonique d'une matrice de transition	195
19.6 La chaîne de Markov canonique associée à une matrice de transition	196
19.6.1 Canonisation d'un processus de Markov	197
19.7 Calculs sur la chaîne canonique et propriété de Markov forte	197
19.8 Première grande application : la probabilité de retour	199
19.9 Opérateurs potentiels	200
19.10 Récurrence et transience	200
19.10.1 Communications entre deux états	201
19.11 Lien entre martingale et chaîne de Markov	203
19.12 Chaînes irréductibles récurrentes	203
19.13 Théorèmes ergodiques	207
19.14 Conséquences du théorème ergodique	208
19.15 Périodicité	209
19.16 Chaînes de Markov à espaces d'états finis ([8] pp. 195-199)	210
19.17 TD Probabilités : Chaînes de Markov I	211
20 Chaînes de Markov et télécommunications	215
20.1 TD Probabilités : Un exemple de télécommunication	215
20.2 TD Probabilités : Système de transmission Aloha	217
20.2.1 Description du modèle	218
20.2.2 Instabilité de ALOHA	218
20.2.3 Quelques outils théoriques : fonctions de Liapounov	219
20.2.4 Stabilisation du protocole Aloha	222
21 Formulaire des chaînes de Markov	225

22	Processus de Poisson	229
22.1	Définitions et propriétés	229
22.2	TD Probabilités : Processus de Poisson	231
22.3	Processus de Poisson (version élégante)	233
22.3.1	Généralités	234
22.4	Superposition, restriction et transformation	234
22.4.1	Théorème d'existence	235
22.4.2	Processus de Poisson marqués	236
23	Limites hydrodynamiques	239
23.1	Marche aléatoires à temps continu	240
23.2	Système de particules libres	241
23.3	Mesures invariantes	241
23.4	Equivalence des ensembles	242
23.5	Equilibre local	244
23.6	Quelques remarques	245
24	Simulations de variables aléatoires et applications	247
24.1	La méthode MCMC	247
24.1.1	Les méthodes classiques	247
24.1.2	Le principe de la méthode MCMC	248
24.1.3	Simulation d'une chaîne de Markov	249
24.1.4	Un modèle d'exclusion	250
24.2	Critique de la méthode MCMC	250
24.3	L'algorithme de Propp-Wilson	251
24.4	Le modèle d'Ising	253
24.4.1	Distribution de Gibbs	253
24.4.2	Distribution de Gibbs en volume infini et transition de phase	254
24.4.3	Simulation d'une distribution de Gibbs par la méthode MCMC	255
24.4.4	Simulation via l'algorithme de Propp-Wilson pour le modèle d'Ising	256
24.4.5	Simulation de la transition de phase	256
24.5	Méthode de gradient stochastique	257
25	Le modèle de Cox-Ross-Rubinstein, Fanny Godet	261
26	Théorie de l'information et de la communication	267
26.1	Introduction	267
26.1.1	L'entropie d'une source	268
26.1.2	Transmission dans un canal bruité	269
26.1.3	Plan	269
26.2	Les étapes du codage : vocabulaire	270
26.3	Entropie, entropie relative, information mutuelle	279

Chapitre 1

La construction des espaces de probabilités

1.1 Algèbres, tribus, π -systèmes, λ -systèmes et espaces mesurables

Soit Ω un ensemble quelconque. On commence par définir toute une typologie de collections d'ensembles que l'on utilisera tout au long du cours.

Définition 1.1.1. (*π, λ -systèmes, algèbres et tribus*)

1. Une classe \mathcal{P} de sous-ensembles de Ω est appelée un π -système ssi :

$$A, B \in \mathcal{P} \Rightarrow A \cap B \in \mathcal{P}$$

2. Une classe \mathcal{L} de sous-ensembles de Ω est appelée un λ -système ssi :

i) $\Omega \in \mathcal{L}$.

ii) Si $A, B \in \mathcal{L}$ et $A \subset B$ alors $B \setminus A \in \mathcal{L}$.

iii) Si $(A_n)_{n \in \mathbb{N}}$ est une suite croissante d'éléments de \mathcal{L} alors $\cup_{n \in \mathbb{N}} A_n \in \mathcal{L}$.

3. Une classe \mathcal{F} de sous-ensembles de Ω est appelée une algèbre ssi :

i) $\Omega \in \mathcal{F}$.

ii) $F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}$.

iii) $F, G \in \mathcal{F} \Rightarrow F \cup G \in \mathcal{F}$.

4. Une classe \mathcal{F} de sous-ensembles de Ω est une tribu (ou σ -algèbre) ssi :

i) \mathcal{F} est une algèbre.

ii) Si $F_n \in \mathcal{F}$ pour tout $n \in \mathbb{N}$ alors $\cup_{n \in \mathbb{N}} F_n \in \mathcal{F}$.

Remarque 1.1.1. Il est facile de voir que si \mathcal{F} est une algèbre alors \mathcal{F} est stable par réunion et intersection finie (mais pas forcément dénombrable) et que la définition d'une algèbre pourrait être donnée de manière équivalente si à la place de la stabilité par réunion finie, on imposait la stabilité par intersection finie.

De même dans la définition d'une tribu, la propriété ii), compte-tenu de i), peut être remplacée par la stabilité par réunion dénombrable croissante (c'est-à-dire en supposant que les F_n forment une suite croissante au sens de l'inclusion) ou bien par réunion disjointe (c'est-à-dire en supposant que les F_n forment une suite d'éléments disjoints de \mathcal{F}).

Proposition 1.1.1. Une collection de sous-ensembles de Ω est une tribu ssi c'est à la fois un λ -système et un π -système.

En général, on ne dispose pas directement d'algèbres, tribus... et on les construit à partir "d'éléments générateurs". Ce qui permet de le faire, c'est la remarque suivante :

“Une intersection quelconque (dénombrable ou pas) de π -systèmes, λ -systèmes, d'algèbres ou de tribus est (respectivement) un π -système, un λ -système, une algèbre ou une tribu.”

En effet, si \mathcal{C} est une collection de sous-ensembles de Ω , puisque $\mathcal{P}(\Omega)$ (ensemble de toutes les parties de Ω) est une tribu (donc un π -système, λ -système et une algèbre) contenant \mathcal{C} , l'intersection de tous les π -systèmes (resp. λ -systèmes, algèbres, tribus) contenant \mathcal{C} est un π -système (resp. λ -système, algèbre, tribu) contenant \mathcal{C} . Et c'est le plus petit π -système (resp. λ -système, algèbre, tribu) contenant \mathcal{C} . On l'appelle alors π -système (resp. λ -système, algèbre, tribu) engendré par \mathcal{C} et on le note $\pi(\mathcal{C})$ (resp. $\lambda(\mathcal{C})$, $\alpha(\mathcal{C})$, $\sigma(\mathcal{C})$).

Le théorème suivant est fondamental dans tout ce qui suit et notamment dans la démonstration du théorème de Carathéodory dont nous parlerons plus loin.

Théorème 1.1.1. (Dynkin)

Si \mathcal{P} est un π -système contenu dans un λ -système alors ce λ -système contient la tribu engendrée par ce π -système.

Démonstration. Voir TD. □

Définition 1.1.2. Une paire (Ω, \mathcal{F}) , où Ω est un ensemble et \mathcal{F} est une tribu sur Ω , est appelée un espace mesurable. Un élément de \mathcal{F} est appelé un ensemble \mathcal{F} -mesurable.

Passons maintenant à un exemple fondamental d'espace mesurable. Soit S un espace topologique muni de sa famille \mathcal{S} d'ouverts. On appelle tribu borélienne sur S la tribu engendrée par \mathcal{S} . On la note $\mathcal{B}(S)$ et ses éléments sont appelés les boréliens de \mathcal{S} . Les éléments de $\mathcal{B}(\mathbb{R}^d)$ par exemple peuvent être très compliqués et il est possible de construire des ensembles non boréliens (voir TD). Néanmoins, il est facile d'exhiber des classes assez simples engendrant la tribu borélienne de \mathbb{R}^d . Cette tribu est en effet engendrée (par exemple) par l'une des classes suivantes (voir TD) :

$$\left\{ \prod_{i=1}^d [a_i; b_i]; a_i, b_i \in \mathbb{R} \right\}, \left\{ \prod_{i=1}^d (-\infty; b_i]; b_i \in \mathbb{R} \right\}, \left\{ \prod_{i=1}^d [a_i; +\infty); a_i \in \mathbb{R} \right\}$$

1.2 Mesure

Définition 1.2.1. Soit Ω un ensemble muni d'une algèbre \mathcal{F} et soit μ une fonction d'ensemble positive pouvant prendre la valeur $+\infty$.

On dira qu'une telle fonction est additive ssi :

1. $\mu(\emptyset) = 0$
2. Si $F, G \in \mathcal{F}$ et si $F \cap G = \emptyset$ alors $\mu(F \cup G) = \mu(F) + \mu(G)$.

On dira qu'elle est σ -additive (ou dénombrablement additive) ssi :

1. $\mu(\emptyset) = 0$.
2. Pour toute suite d'éléments disjoints $F_n \in \mathcal{F}$ tels que $\cup_n F_n \in \mathcal{F}$, $\mu(\cup_n F_n) = \sum_n \mu(F_n)$.

Remarque 1.2.1. Attention à la définition donnée ici. Il est important de noter que μ est définie sur une algèbre et qu'à priori, une réunion dénombrable d'éléments de \mathcal{F} n'est pas obligatoirement dans \mathcal{F} . D'autre part, on a bien sûr que toute fonction d'ensemble positive dénombrablement additive est additive. Enfin, il est aisé de montrer que μ est dénombrablement additive ssi elle est additive et si pour toute suite croissante $(F_n)_n$ d'éléments de \mathcal{F} telle que $F = \cup_n F_n \in \mathcal{F}$, $\mu(F) = \lim_{n \rightarrow \infty} \mu(F_n)$.

Définition 1.2.2. Soit (Ω, \mathcal{F}) un espace mesurable. Une application μ de \mathcal{F} dans $[0, +\infty]$ est appelée mesure sur (Ω, \mathcal{F}) ssi μ est dénombrablement additive. Le triplet $(\Omega, \mathcal{F}, \mu)$ est alors appelé un espace de mesure (ou espace mesuré).

★ On dira que la mesure μ est finie (ou de masse finie) ssi $\mu(\Omega) < +\infty$.

★ On dira qu'elle est σ -finie ssi il existe une suite $(\Omega_n)_n$ d'ensembles mesurables tels que $\mu(\Omega_n) < +\infty$ et $\Omega = \cup_n \Omega_n$.

★ On appellera probabilité toute mesure μ de masse $\mu(\Omega)$ égale à 1. $(\Omega, \mathcal{F}, \mu)$ est alors appelé un espace probabilisé ou espace de probabilité.

1.2.1 Complétion d'un espace mesuré

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré. On définit \mathcal{N} la classe des ensembles négligeables de (Ω, \mathcal{F}) par :

$$N \in \mathcal{N} \text{ ssi } \exists Z \in \mathcal{F}, \mu(Z) = 0 \text{ et } N \subset Z$$

\mathcal{F}^* est la classe des parties F de Ω telles qu'il existe $E, G \in \mathcal{F}$, $E \subset F \subset G$ et $\mu(G \setminus E) = 0$. Il est assez facile de voir que \mathcal{F}^* est une tribu qui plus est engendrée par \mathcal{F} et \mathcal{N} . De plus, les ensembles négligeables de (Ω, \mathcal{F}^*) sont les mêmes que ceux de (Ω, \mathcal{F}) .

Sur \mathcal{F}^* , on définit alors une fonction μ^* par $\mu^*(F) = \mu(E) = \mu(G)$ (remarquer que cette définition a bien un sens). Il n'est alors pas difficile de montrer que $(\Omega, \mathcal{F}^*, \mu^*)$ est un espace mesuré qui "étend" l'espace mesuré de départ. On l'appelle la complétion de $(\Omega, \mathcal{F}, \mu)$. En particulier, la tribu complétée $\mathcal{B}(\mathbb{R})$ pour λ , la mesure de Lebesgue (cf après), est appelée la tribu des Lebesguiens. En ce qui nous concerne, nous n'aurons pratiquement jamais à travailler avec la tribu des Lebesguiens et on se contentera des Boréliens...

1.2.2 Événements

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Les éléments de \mathcal{F} sont appelés les *événements*. Une propriété $\mathcal{P}(\omega)$ dépendant de $\omega \in \Omega$ sera dite vraie presque sûrement (notée p.s.) ssi l'ensemble (non nécessairement mesurable) $\{\omega \in \Omega; \mathcal{P}(\omega) \text{ fausse}\}$ est un ensemble négligeable. On utilisera fréquemment dans la suite les deux événements asymptotiques définis de la manière suivante. Soit E_n une suite d'événements. Les événements $\limsup E_n$ et $\liminf E_n$ sont donnés par :

$$\begin{cases} \limsup E_n = \bigcap_n \bigcup_{k \geq n} E_k = \{E_n \text{ ont lieu infiniment souvent}\} \\ \liminf E_n = \bigcup_n \bigcap_{k \geq n} E_k = \{E_n \text{ ont toujours lieu sauf un nombre fini de fois}\} \end{cases}$$

1.2.3 Problème posé

Le problème posé est le suivant et est bien illustré par le cas de la mesure de Lebesgue λ . On cherche à définir une mesure λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour tous réels $a < b$, $\lambda((a, b)) = b - a$. Première question : existe-t-il une telle mesure ? Si oui, est-elle unique ? La réponse à la première question est apportée par le théorème de Carathéodory, la réponse à la deuxième par le lemme de Dynkin.

1.3 Unicité des extensions

Proposition 1.3.1. (d'unicité des prolongements de mesures)

Soit Ω un ensemble muni d'un π -système \mathcal{P} et soit $\mathcal{F} = \sigma(\mathcal{P})$. Supposons que l'on ait deux mesures μ_1 et μ_2 sur \mathcal{F} de même masse finie, $\mu_1(\Omega) = \mu_2(\Omega)$, qui coïncident sur \mathcal{P} . Alors

$$\mu_1 = \mu_2$$

Démonstration. On définit la classe $\mathcal{D} = \{F \in \mathcal{F}; \mu_1(F) = \mu_2(F)\}$. Il est facile de voir que c'est un λ -système donc par le théorème de Dynkin, \mathcal{D} contient \mathcal{F} . En effet, le fait que les deux mesures aient même masse montre que $\Omega \in \mathcal{D}$ et les masses finies assurent que \mathcal{D} est stable par différence propre. \square

Il s'en suit que deux probabilités qui coïncident sur un π -système coïncident aussi sur la tribu qu'engendre ce π -système.

1.4 Théorème de Carathéodory

Théorème 1.4.1. (Carathéodory)

Soit Ω un ensemble muni d'une algèbre \mathcal{F}_0 et soit \mathcal{F} la tribu engendrée par cette algèbre. Si μ_0 est

une application dénombrablement additive $\mu_0 : \mathcal{F}_0 \rightarrow [0; +\infty]$ alors il existe une mesure μ sur (Ω, \mathcal{F}) telle que

$$\mu = \mu_0 \text{ sur } \mathcal{F}_0$$

Si $\mu_0(\Omega) < +\infty$ alors cette extension est unique.

L'unicité a été traitée dans la section précédente. Avant de passer à la preuve de ce théorème, examinons d'un peu plus près le cas de la mesure de Lebesgue sur $(0, 1]$ muni de la tribu borélienne. On définit la classe \mathcal{F}_0 des sous ensembles F de $(0, 1]$ qui peuvent se mettre sous la forme :

$$F = (a_1, b_1] \cup \dots \cup (a_r, b_r]$$

avec $r \in \mathbb{N}$ et $0 \leq a_1 \leq \dots \leq a_r \leq b_r \leq 1$. On montre facilement que \mathcal{F}_0 est une algèbre et que la tribu qu'elle engendre est la tribu des boréliens sur $(0, 1]$. Sur \mathcal{F}_0 , on définit l'application λ par

$$\lambda(F) = \sum_{k=1}^r (b_k - a_k)$$

si F s'écrit sous la forme donnée ci-dessus. Il faut bien entendu vérifier que cette définition est cohérente (i.e. ne dépend pas de l'écriture de F choisie). On démontre alors, et ce n'est pas totalement trivial, que λ est σ -additive sur \mathcal{F}_0 . Par le théorème d'extension de Carathéodory, λ s'étend en une mesure encore notée λ et appelée *mesure de Lebesgue* sur $((0, 1], \mathcal{B}((0, 1]))$.

Passons maintenant à la preuve du théorème de Carathéodory.

Démonstration.

Nous n'indiquerons que les principales étapes, les détails étant parfaitement expliqués dans [30].

Etape 1 : Soit $\mathcal{G} = \mathcal{P}(\Omega)$ la tribu composée de toutes les parties de Ω . Pour un $G \in \mathcal{G}$, on définit

$$\lambda(G) = \inf \left\{ \sum_n \mu_0(F_n) \right\}$$

où l'infimum est pris sur toutes les suites $(F_n)_n$ d'éléments de \mathcal{F}_0 qui recouvrent G , i.e. $G \subset \cup_n F_n$. λ n'est pas une mesure mais est ce que l'on appelle (improprement) une "*mesure extérieure*", c'est-à-dire qu'elle jouit des propriétés suivantes :

- a) $\lambda(\emptyset) = 0$.
- b) λ est croissante : $G_1 \subset G_2 \Rightarrow \lambda(G_1) \leq \lambda(G_2)$.
- c) λ est dénombrablement sous-additive : si $(G_k)_k$ est une suite d'éléments de \mathcal{G} alors

$$\lambda(\cup_k G_k) \leq \sum_k \lambda(G_k)$$

Seule c) mérite d'être expliquée. Soit $(G_n)_n$ une suite supposée telle que pour chaque n , $\lambda(G_n) < +\infty$ (sinon c'est trivial). On se donne $\varepsilon > 0$ et pour chaque n , on se donne une suite d'éléments $(F_{n,k})_k$ de \mathcal{F}_0 pour lesquels

$$G_n \subset \cup_k F_{n,k} \text{ , } \sum_k \mu_0(F_{n,k}) \leq \lambda(G_n) + \varepsilon 2^{-n}$$

Alors $G \subset \cup_{n,k} F_{n,k}$ et on a :

$$\lambda(G) \leq \sum_{n,k} \mu_0(F_{n,k}) \leq \sum_n \lambda(G_n) + \varepsilon$$

Puisque ε est arbitraire, en le faisant tendre vers 0, on obtient c).

Etape 2 : On introduit alors la notion de λ -ensembles. Un élément L de \mathcal{G} est appelé un λ -ensemble ssi “il décompose tout élément de \mathcal{G} proprement”, c’est-à-dire :

$$\lambda(L \cap G) + \lambda(L^c \cap G) = \lambda(G), \quad \forall G \in \mathcal{G}$$

Lemme 1.4.1. *Les λ -ensembles forment une tribu \mathcal{L} sur laquelle λ est dénombrablement additive si bien que $(\Omega, \mathcal{L}, \lambda)$ est un espace mesuré.*

Démonstration. (du lemme). On montre facilement que \mathcal{L} est une algèbre qui jouit de la propriété suivante :

Si L_1, L_2, \dots, L_n sont des éléments de \mathcal{L} disjoints et $G \in \mathcal{G}$ une partie quelconque de Ω alors

$$\lambda(\cup_{k=1}^n (L_k \cap G)) = \sum_{k=1}^n \lambda(L_k \cap G) \quad (\star)$$

Nous avons donc à démontrer que si $(L_k)_k$ est une suite d’éléments disjoints de \mathcal{L} alors $L = \cup_k L_k \in \mathcal{L}$ (cf. la remarque après la définition de la notion de tribu) et

$$\lambda(L) = \sum_k \lambda(L_k)$$

Par sous-additivité de λ , on a pour tout $G \in \mathcal{G}$:

$$\lambda(G) \leq \lambda(L \cap G) + \lambda(L^c \cap G)$$

Passons maintenant à l’inégalité inverse. Soit $M_n = \cup_{k \leq n} L_k$. Puisque \mathcal{L} est une algèbre (cf ce qui précède), $M_n \in \mathcal{L}$ donc

$$\lambda(G) = \lambda(M_n \cap G) + \lambda(M_n^c \cap G)$$

Or $L^c \subset M_n^c$ si bien que

$$\lambda(G) \geq \lambda(M_n \cap G) + \lambda(L^c \cap G)$$

et par (\star) ,

$$\lambda(G) \geq \sum_{k=1}^n \lambda(L_k \cap G) + \lambda(L^c \cap G)$$

Par passage à la limite quand $n \rightarrow +\infty$, on obtient

$$\lambda(G) \geq \sum_{k=1}^{\infty} \lambda(L_k \cap G) + \lambda(L^c \cap G) \geq \lambda(L \cap G) + \lambda(L^c \cap G)$$

la deuxième inégalité étant une conséquence de la sous-additivité de λ . On en déduit que $L \in \mathcal{L}$. D'autre part, puisque la dernière inégalité est en fait une égalité, on voit qu'il en est de même des autres et que finalement

$$\lambda(L \cap G) = \sum_k \lambda(L_k \cap G)$$

Il suffit maintenant de prendre $G = \Omega$ pour conclure. □

Il nous reste à voir que $\mathcal{F}_0 \subset \mathcal{L}$ et $\lambda = \mu_0$ sur \mathcal{F}_0 pour achever la preuve du théorème.

Etape 3 : Preuve que $\lambda = \mu_0$ sur \mathcal{F}_0 .

Soit $F \in \mathcal{F}_0$. On a bien entendu $\lambda(F) \leq \mu_0(F)$. Supposons maintenant que $F \subset \cup_n F_n$ où $F_n \in \mathcal{F}_0$. On définit une suite $(E_n)_n$ d'éléments disjoints de \mathcal{F}_0 par :

$$E_1 = F_1, \quad E_n = F_n \cap (\cup_{k < n} F_k)^c$$

si bien que $E_n \subset F_n$ et $F \subset \cup_n F_n = \cup_n E_n$. En utilisant la σ -additivité de μ_0 sur \mathcal{F}_0 , on a

$$\mu_0(F) = \mu_0(\cup_n (F \cap E_n)) = \sum_n \mu_0(F \cap E_n)$$

Donc

$$\mu_0(F) \leq \sum_n \mu_0(E_n) \leq \sum_n \mu_0(F_n)$$

si bien que $\lambda(F) \geq \mu_0(F)$ ce qui conclut la troisième étape.

Etape 4 : Preuve de $\mathcal{F}_0 \subset \mathcal{L}$.

Il nous faut voir que si E est un élément de \mathcal{F}_0 alors E décompose proprement les éléments de \mathcal{G} . Soit donc $G \in \mathcal{G}$ et soit pour $\varepsilon > 0$ quelconque une suite $(F_n)_n$ d'éléments de \mathcal{F}_0 tels que $G \subset \cup_n F_n$ avec

$$\sum_n \mu_0(F_n) \leq \lambda(G) + \varepsilon$$

Par définition de λ ,

$$\begin{aligned} \sum_n \mu_0(F_n) &= \sum_n \mu_0(E \cap F_n) + \sum_n \mu_0(E^c \cap F_n) \\ &\geq \lambda(E \cap G) + \lambda(E^c \cap G) \end{aligned}$$

car $E \cap G \subset \cup (E \cap F_n)$ et $E^c \cap G \subset \cup (E^c \cap F_n)$. Puisque ε est aussi petit que l'on veut,

$$\lambda(G) \geq \lambda(E \cap G) + \lambda(E^c \cap G)$$

Puisque λ est sous-additive,

$$\lambda(G) \leq \lambda(E \cap G) + \lambda(E^c \cap G)$$

Donc E est un λ -ensemble. □

1.5 Mesures complexes

Dans cette section, nous donnons un résumé de ce que l'on trouve dans le chapitre 6 de [27] concernant les mesures complexes. (Ω, \mathcal{F}) est toujours un espace mesurable.

Une mesure complexe μ sur \mathcal{F} est une application de \mathcal{F} dans \mathbb{C} telle que $\mu(\emptyset) = 0$ et pour tout $A \in \mathcal{F}$ et toute partition mesurable dénombrable $(A_i)_i$ de A , on ait

$$\mu(A) = \sum_i \mu(A_i)$$

Cette convergence est à entendre au sens de l'absolue convergence. On définit alors la variation totale de μ par

$$|\mu|(A) = \sup \sum_i |\mu(A_i)|$$

où le supremum est pris sur toutes les partitions dénombrables mesurables de A .

Théorème 1.5.1. *La variation totale $|\mu|$ d'une mesure complexe μ sur \mathcal{F} est une mesure positive de masse finie sur \mathcal{F} .*

Il est d'autre part facile de voir que l'ensemble des mesures complexes sur \mathcal{F} forme un espace vectoriel (on définira sans peine ce qu'est la somme de deux mesures complexes et la multiplication par un complexe d'une mesure complexe). La notion de variation totale d'une mesure permet de définir une norme sur cet espace vectoriel : la norme d'une mesure complexe μ est donnée par $\|\mu\| = |\mu|(\Omega) < +\infty$. On obtient donc une structure d'espace vectoriel normé.

Dans le cas où l'on s'intéresse uniquement aux mesures complexes à valeurs dans \mathbb{R} , on parle de *mesures signées*. On définit la variation totale d'une mesure signée comme précédemment puis on définit respectivement les variations positives et négatives de μ par :

$$\mu^+ = \frac{1}{2}(|\mu| + \mu), \quad \mu^- = \frac{1}{2}(|\mu| - \mu)$$

Ce sont des mesures positives de masses finies telles que

$$\mu = \mu^+ - \mu^-$$

On appelle cette décomposition la *décomposition de Jordan* de μ . On a en fait une propriété de minimalité de ces mesures.

Proposition 1.5.1. *Si μ est une mesure signée qui s'écrit $\mu = \lambda_1 - \lambda_2$ où λ_1, λ_2 sont deux mesures positives alors $\lambda_1 \geq \mu^+$ et $\lambda_2 \geq \mu^-$.*

Cette propriété découle du théorème de *décomposition de Hahn* qui dit que μ^+ et μ^- ont leurs masses respectivement concentrées sur $A \in \mathcal{F}$ et $B \in \mathcal{F}$ tels que $A \cap B = \emptyset$ et $A \cup B = \Omega$. Ce théorème est lui-même une conséquence du théorème de Radon-Nykodym (cf. après).

1.6 Mesures de Borel

Pour cette section, on se reportera au chapitre 2 de [27]. On suppose ici que Ω est un espace topologique séparé localement compact et \mathcal{F} est la tribu des boréliens sur Ω .

Définition 1.6.1. Une mesure (complexe ou positive) μ définie sur la tribu des boréliens \mathcal{F} d'un espace Ω séparé localement compact est appelée une mesure borélienne sur Ω .

Une mesure de Borel positive sera dite régulière si elle satisfait :

1. $\forall E \in \mathcal{F}, \mu(E) = \inf \{ \mu(V); E \subset V, V \text{ ouvert} \}$.
2. $\forall E \in \mathcal{F}$ tel que $\mu(E) < +\infty, \mu(E) = \inf \{ \mu(K); K \subset E, K \text{ compact} \}$.

On dira de même d'une mesure complexe qu'elle est régulière si sa variation totale $|\mu|$ qui est une mesure positive est régulière.

On énonce maintenant le théorème de représentation de Riesz tel qu'on le trouve dans [27] dans le chapitre sur les mesures complexes et le chapitre sur les mesures de Borel positive. Il est relativement général. Sa démonstration nécessite des préliminaires topologiques longs. Ce théorème anticipe sur la suite puisqu'il fait appel à la notion d'intégrale relativement à une mesure mais on supposera que le lecteur est familier avec la théorie de l'intégration relativement à une mesure complexe ou tout du moins relativement à une mesure signée.

On note C_c l'espace des fonctions continues sur Ω à support compact et C_0 l'espace des fonctions de Ω dans \mathbb{C} s'annulant à l'infini que l'on munit de la norme uniforme. C_c est dense dans C_0 pour la norme uniforme. Le théorème de représentation de Riesz caractérise toutes les formes linéaires positives (non nécessairement bornées) et les formes linéaires bornées sur C_c muni de la topologie de la convergence uniforme. Bien entendu, toute forme linéaire bornée sur C_c se prolonge de manière unique en une forme linéaire continue sur C_0 .

Théorème 1.6.1. (de représentation de Riesz)

1. Soit Φ une forme linéaire positive sur C_c . Il existe une unique mesure positive borélienne régulière μ sur (Ω, \mathcal{F}) telle que

$$\Phi(f) = \int_{\Omega} f d\mu, \quad (f \in C_c)$$

2. Soit Φ une forme linéaire bornée sur C_0 . Il existe une unique mesure de Borel μ complexe régulière telle que

$$\Phi(f) = \int_{\Omega} f d\mu, \quad (f \in C_0)$$

De plus, $\|\Phi\| = |\mu|(\Omega)$

1.7 Absolue continuité

Là aussi, nous anticipons en supposant connue la notion d'intégrale relativement à une mesure. (Ω, \mathcal{F}) est un espace mesurable quelconque.

Définition 1.7.1. Soit μ une mesure positive sur \mathcal{F} et λ une mesure quelconque sur \mathcal{F} , positive ou complexe. On dit que λ est absolument continue par rapport à μ et on note $\lambda \ll \mu$ si

$$\mu(E) = 0 \implies \lambda(E) = 0$$

A l'opposé de cette notion existe celle de mesures mutuellement singulières.

Définition 1.7.2. Soit λ et μ deux mesures quelconques (positives ou complexes). On dit qu'elles sont mutuellement singulières (et on note $\lambda \perp \mu$) s'il existe deux ensembles mesurables disjoints A et B tels que pour tout $E \in \mathcal{F}$,

$$\lambda(E) = \lambda(E \cap A), \quad \text{et} \quad \mu(E) = \mu(E \cap B)$$

On a le théorème de Radon-Nikodym dont une partie peut être démontrée via les martingales (voir TD).

Théorème 1.7.1. Soit λ et μ deux mesures positives bornées sur la tribu (Ω, \mathcal{F}) .

1. Il existe un couple unique de mesures λ_a et λ_s tel que

$$\lambda = \lambda_a + \lambda_s, \quad \lambda_a \ll \mu, \quad \lambda_s \perp \mu$$

2. Ces mesures sont positives bornées et $\lambda_a \perp \lambda_s$.

3. Il existe un unique $h \in \mathcal{L}^1(\mu)$ tel que

$$d\lambda_a = h d\mu$$

Le couple (λ_a, λ_s) s'appelle la décomposition de Lebesgue de λ relativement à μ .

L'idée de la preuve de ce théorème repose essentiellement sur le théorème de représentation de Riesz des applications linéaires continues dans les espaces de Hilbert. La preuve n'est pas très longue.

Le théorème précédent reste encore vrai dans le cas d'une mesure μ positive σ -finie et d'une mesure complexe (donc bornée). Dans le cas enfin λ, μ positives et σ -finies, la majeure partie du théorème reste valable mais on a plus $h \in \mathcal{L}^1(\mu)$ (on a seulement $\mathcal{L}_{loc}^1(\mu)$). En revanche, s'il n'y a plus d'hypothèse de σ -finitude, on ne peut plus rien dire (cf. [27] pour un contre exemple).

1.8 Algèbres, tribus et compagnie

Exercice 1.8.1. Exercice (Lemme de Dynkin) (Williams, p 193)

Un π -système est un ensemble de parties de Ω stable par intersection finie. Un λ -système est un ensemble \mathcal{L} de parties de Ω tel que :

1. $\Omega \in \mathcal{L}$
2. \mathcal{L} est stable par différence propre (i.e. $A, B \in \mathcal{L}$ et $A \subset B$ entraînent $B \setminus A \in \mathcal{L}$.)
3. \mathcal{L} est stable par réunion dénombrable croissante

1) Démontrer qu'une collection Σ de sous-ensembles de Ω est une tribu (σ -algebra) si et seulement si Σ est à la fois un π -système et un λ -système (en anglais, d -system).

2) On va montrer que si \mathcal{I} est un π -système, alors la tribu $\sigma(\mathcal{I})$ engendrée par \mathcal{I} et le λ -système engendré par \mathcal{I} sont égaux : $\lambda(\mathcal{I}) = \sigma(\mathcal{I})$. La démonstration est détaillée dans les deux questions qui suivent.

2a) Soit $\mathcal{D}_1 = \{B \in \lambda(\mathcal{I}), B \cap C \in \lambda(\mathcal{I}), \forall C \in \mathcal{I}\}$. Montrer que \mathcal{D}_1 hérite de la structure de λ -système de $\lambda(\mathcal{I})$. En déduire que $\mathcal{D}_1 = \lambda(\mathcal{I})$.

2b) Posons $\mathcal{D}_2 = \{A \in \lambda(\mathcal{I}), B \cap A \in \lambda(\mathcal{I}), \forall B \in \lambda(\mathcal{I})\}$. Par la question 2a), \mathcal{D}_2 contient \mathcal{I} . Montrer que \mathcal{D}_2 hérite la structure de λ -système de $\lambda(\mathcal{I})$ et en déduire que $\lambda(\mathcal{I})$ est un π -système. (Se déduit de 2a) en substituant $\lambda(\mathcal{I})$ à \mathcal{I} et en remarquant que $\lambda(\lambda(\mathcal{I})) = \lambda(\mathcal{I})$.

3) Déduire le **théorème de Dynkin** :

Soit \mathcal{P} un π -système et \mathcal{L} un λ -système tel que $\mathcal{P} \subset \mathcal{L}$. Alors la tribu engendrée par \mathcal{P} est contenue dans \mathcal{L} .

Exercice 1.8.2. (Il n'y a pas de tribu dénombrable. Cet exercice est extrait de Le Calcul Integral, H. Buchwalter (ne donne pas la correction).)

Soit \mathcal{F} une tribu sur Ω supposée au plus dénombrable.

a) Démontrer que tout $\omega \in \Omega$ est contenu dans un plus petit élément de \mathcal{F} .

b) En déduire que \mathcal{F} est exactement la tribu engendrée par une partition au plus dénombrable.

c) Montrer alors que \mathcal{F} est nécessairement finie de cardinal une puissance de 2 et que donc il n'existe pas de tribu dénombrable au sens strict.

d) Trouver toutes les tribus de \mathbb{N} .

Réponse : ce sont toutes les tribus engendrées par une partition. En effet, on peut encore montrer la propriété a) : soit n_0 et prenons pour tout n une partie A_n contenant n_0 telle que $n \notin A_n$ (si un tel A_n existe : sinon prendre $A_n = \mathbb{N}$). On pose $A(n_0) = \bigcap_n A_n$. Vérifier que c'est bien le plus petit élément de la tribu contenant n_0 .

Exercice 1.8.3. (Exemple d'ensemble non borélien) (Williams, p 192)

Soit $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ l'espace des réels muni de la tribu borélienne et de la mesure de Lebesgue.

1) On définit la relation binaire suivante sur \mathbb{R}

$$x\mathcal{R}y \text{ ssi } x - y \in \mathbb{Q}$$

Montrer que \mathcal{R} est une relation d'équivalence.

2) On désigne par \mathcal{C} l'ensemble des classes d'équivalence. Montrer que pour toute classe $C \in \mathcal{C}$ il existe $x \in C \cap [0, 1)$. On choisit alors pour chaque classe C un x_C vérifiant la propriété précédente et on note $A = \{x_C / C \in \mathcal{C}\}$.

3) Montrer que

$$[0, 1] \subset \bigcup_{q \in [-1, 1] \cap \mathbb{Q}} q + A \subset [-1, 2)$$

et que les $q + A$ sont deux à deux disjoints.

4) En supposant A borélien, aboutir à une contradiction.

Exercice 1.8.4. (Ensembles de Cantor) (Hauchecorne, "Les contre-exemples en mathématiques", pp. 128-131) On considère pour $0 < \alpha < 1$ l'opération suivante S_α : enlever à un intervalle $[a, b]$ son intervalle central $]\frac{a+b}{2} - \alpha\frac{b-a}{2}, \frac{a+b}{2} + \alpha\frac{b-a}{2}[$. Soit $E \subset \mathbb{R}$ une union finie d'intervalles disjoints. On note $s_\alpha(E)$ le résultat de l'application de s_α à tous les intervalles de E . Quand $\alpha = \frac{1}{3}$, l'ensemble $K_{\frac{1}{3}} = \bigcap_{n \geq 1} s_{\frac{1}{3}}^n([0, 1])$ est appelé ensemble triadique de Cantor. Plus généralement, si $\alpha = (\alpha_1, \dots, \alpha_n, \dots)$ est une suite de coefficients $0 < \alpha_n < 1$, on pose encore

$$K_\alpha^n = s_{\alpha_n} s_{\alpha_{n-1}} \dots s_{\alpha_1}([0, 1]), \quad K_\alpha = \bigcap_{n \geq 1} K_\alpha^n.$$

1) Montrer que pour toute suite α , l'ensemble K_α est un compact d'intérieur vide, équipotent à l'ensemble des suites binaires (gauche, droite) $^{\mathbb{N}}$. (Indication : tout élément de K_α est défini de manière unique par le fait qu'il appartient à l'intervalle de droite ou de gauche de K_α^1 , puis au sous-intervalle de cet intervalle de gauche ou de droite dans K_α^2 , etc.) En déduire que K_α est équipotent à $[0, 1]$.

2) Montrer que K_α est de mesure strictement positive si et seulement si $\sum_n \alpha_n < +\infty$. En déduire qu'il y a des ensembles boréliens équipotents à \mathbb{R} et de mesure nulle et des fermés de mesure positive et d'intérieur vide.

Exercice 1.8.5. (Ouvert dense de mesure petite) (Hauchecorne, "Les contre-exemples en mathématiques", pp. 128-131) Montrer qu'il existe un ouvert de \mathbb{R} contenant \mathbb{Q} et de mesure inférieure à ε . Remarquer que cela prouve qu'il y a des ouverts denses de mesure aussi petite qu'on veut. Indication : énumérer les rationnels q_n , $n \in \mathbb{N}$ et considérer les intervalles $]-\frac{1}{2^n} + q_n, q_n + \frac{1}{2^n}[$.

Exercice 1.8.6. (Il existe un ensemble négligeable et non borélien) 1) Soit K l'ensemble triadique de Cantor. Montrer que c'est un compact non vide (donc borélien) négligeable de \mathbb{R} .

2) Montrer par récurrence que l'on peut définir une suite d'applications $\epsilon_n : [0, 1] \rightarrow \{0, 1\}$, $n \geq 1$ telle que

$$\forall x \in [0, 1], \forall n \geq 1, \sum_{i=1}^n \frac{\epsilon_i(x)}{2^i} \leq x < \sum_{i=1}^n \frac{\epsilon_i(x)}{2^i} + \frac{1}{2^n}$$

On définit alors pour x dans $[0, 1]$

$$f(x) = \sum_{i=1}^{\infty} \frac{2\epsilon_i(x)}{3^i}$$

Montrer que f est strictement croissante et en déduire qu'elle est injective et borélienne.

3) Soit A l'ensemble non borélien de l'exercice 7.

1. Montrer que $f([0, 1]) \subset K$
2. Montrer que $f(A)$ est négligeable pour λ .
3. En utilisant l'injectivité de f , établir que $f(A)$ n'est pas borélien.

Exercice 1.8.7. Les seules mesures sur \mathbb{R} finies sur les bornés et invariantes par translation sont des multiples de la mesure de Lebesgue. (Problème 12.1 p. 180 du Billingsley). Indication : appeler a la mesure d'un intervalle de longueur 1, montrer d'abord que les intervalles de longueur $\frac{1}{n}$ ont mesure $\frac{a}{n}$, puis que les intervalles de longueur rationnelle q ont mesure aq , finalement que les intervalles de longueur quelconque b ont mesure ab . Déduire que la mesure considérée coïncide sur un π -système avec la mesure de Lebesgue et conclure.

Exercice 1.8.8. (Théorème de Sard) Soit f une fonction C^1 d'un ouvert de \mathbb{R}^n dans lui-même. On appelle point singulier de f un point $x \in \mathbb{R}^n$ tel que $Df(x)$ est non inversible. On note $|Df(x)|$ le jacobien de f . L'ensemble des points singuliers est donc $S(f) = (|Df|)^{-1}(0)$. Le théorème de Sard dit simplement que l'ensemble $f(S(f))$ est de mesure nulle.

1) Se convaincre que l'on peut se contenter de montrer le théorème de Sard quand f est définie sur un compact. Alors $S(f)$ est compact.

2) On note $B(x, r)$ la boule de centre x et de rayon r . Soit $x_0 \in S(f)$. En écrivant la formule de Taylor, $f(x) = f(x_0) + Df(x_0)(x - x_0) + o(|x - x_0|)$, montrer qu'il existe un sous-espace propre V de \mathbb{R}^N et une constante C tels pour tout $\varepsilon > 0$ il existe $r > 0$ tel que

$$f(B(x_0, r)) \subset f(x_0) + B_V(0, Cr) + C\varepsilon r,$$

où $B_V(0, r)$ désigne la boule de rayon r dans V .

3) Soit K un compact de \mathbb{R}^n . Montrer que pour tout ε il existe r , tel que K soit recouvert par des cubes $(C_i)_{i \in I}$ de diamètre inférieur à r et tels que $mesure((\cup_i C_i) \setminus K) \leq \varepsilon$.

4) En recouvrant $S(f)$ comme indiqué dans la question précédente, montrer le résultat annoncé.

Solution : On écrit $\mathbb{R}^N = \text{Ker}Df(x_0) \oplus (\text{Ker}Df(x_0))^\perp$ et, sur cette somme orthogonale d'espaces, on décompose $x - x_0 = y + z$, de sorte que $\|y\| \leq r$, $\|z\| \leq r$ si $x - x_0 \in B(0, r)$. On note $V = Df(x_0)(\text{Ker}Df(x_0))$. Donc si $\|x - x_0\| < r$ assez petit,

$$f(x) = f(x_0) + Df(x_0)(z) + o(x - x_0) \in f(x_0) + B_V(0, Cr) + B(0, \varepsilon r),$$

avec $C = \|Df(x_0)\|$ et où $B_V(0, Cr)$ est la boule de centre 0 et de rayon Cr dans V . On vérifie aisément que la mesure de $f(B(x_0, r))$ est alors plus petite que $C_1 r^N \varepsilon^d$ où d est la dimension de $\text{Ker}Df(x_0)$ et C_1 une constante adéquate.

Chapitre 2

Variables aléatoires

Dans la suite, $(\Omega, \mathcal{F}, \mathbb{P})$ désigne un espace probabilisé.

2.1 Définitions

Définition 2.1.1. Une variable aléatoire X sur $(\Omega, \mathcal{F}, \mathbb{P})$ est une application mesurable de (Ω, \mathcal{F}) dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, c'est-à-dire une application $X : \Omega \rightarrow \mathbb{R}^d$ telle que

$$\forall A \in \mathcal{B}(\mathbb{R}^d), X^{-1}(A) = \{\omega \in \Omega; X(\omega) \in A\} \in \mathcal{F}$$

Remarque 2.1.1. 1) On réserve parfois le nom de variable aléatoire pour le cas $d = 1$. Pour le cas $d \geq 2$, on utilise le terme plus approprié de vecteur aléatoire.

2) On parle aussi de variables aléatoires pour des applications $X : \Omega \rightarrow \bar{\mathbb{R}} = [-\infty; +\infty]$ pouvant prendre des valeurs infinies. La tribu que l'on met sur $[-\infty; +\infty]$ est la tribu engendrée par les ouverts de $\bar{\mathbb{R}}$, images réciproques des ouverts de $[-1, 1]$ par la bijection bicontinue $f : \mathbb{R} \rightarrow [-1, 1]$ définie par :

$$\forall x \in \mathbb{R}, f(x) = \frac{2}{\pi} \arctan(x); f(\pm\infty) = \pm 1$$

3) On sait que si $X : \Omega \rightarrow \mathbb{R}$ (mesurable ou non mesurable) alors $X^{-1}(\mathcal{B}(\mathbb{R}))$ est une tribu. X est une variable aléatoire ssi $X^{-1}(\mathcal{B}(\mathbb{R}))$ est une sous-tribu de \mathcal{F} . $X^{-1}(\mathcal{B}(\mathbb{R}))$ est donc la plus petite tribu sur Ω rendant X mesurable.

4) $\mathcal{B}(\mathbb{R}^d)$ est engendrée par

$$\left\{ \begin{array}{l} \{\prod_{i=1}^d (-\infty; \alpha_i); \alpha_i \in \mathbb{R}\}, \\ \{\prod_{i=1}^d (-\infty; \alpha_i]; \alpha_i \in \mathbb{R}\}, \\ \{\prod_{i=1}^d (\alpha_i, \beta_i); \alpha_i, \beta_i \in \mathbb{R}\} \dots \end{array} \right.$$

Par conséquent, pour montrer que $X = (X_1, \dots, X_d)$ est une variable aléatoire à valeurs dans \mathbb{R}^d , il suffit de vérifier :

$$\forall \alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d, \{\omega \in \Omega; \forall i \in \mathbb{N}_d, X_i(\omega) < \alpha_i\} \in \mathcal{F}$$

ou

$$\forall \alpha = (\alpha_1, \dots, \alpha_d), \beta = (\beta_1, \dots, \beta_d) \in \mathbb{R}^d, \{\omega \in \Omega; \forall i \in \mathbb{N}_d, \alpha_i < X_i(\omega) < \beta_i\} \in \mathcal{F}$$

En particulier, ceci montre que X est une variable aléatoire à valeurs dans \mathbb{R}^d ssi pour tout $i \in \mathbb{N}_d$, X_i est une variable aléatoire réelle (la raison profonde étant en fait que la tribu borélienne sur \mathbb{R}^d est la tribu produit de d copies de \mathbb{R}).

Notation : On désignera par m l'ensemble des variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et par \bar{m} l'ensemble des variables aléatoires à valeurs dans $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ (qui contient m).

Proposition 2.1.1. *On a les propriétés élémentaires suivantes :*

- 1) $(m, +, \cdot)$ est une algèbre.
- 2) Si $(X_n)_n \in m^{\mathbb{N}}$ alors $\inf_n X_n$, $\sup_n X_n$, $\liminf_n X_n$, $\limsup_n X_n$ sont des éléments de \bar{m} .
- 3) De plus, $\{\omega \in \Omega ; \lim_{n \rightarrow \infty} X_n(\omega) \text{ existe et appartient à } \mathbb{R}\} \in \mathcal{F}$.
- 4) Par suite, si $(X_n)_n$ est une suite d'éléments de \bar{m} et si on pose

$$A = \left\{ \omega \in \Omega ; \lim_{n \rightarrow \infty} X_n(\omega) \text{ existe et appartient à } \mathbb{R} \right\} \in \mathcal{F}$$

alors $(Y_n)_n = (\mathbb{1}_A X_n)_n$ est une suite d'éléments de m .

Ces propriétés, très faciles à démontrer, sont admises. On remarquera qu'elles se généralisent aisément au cas de variables aléatoires multidimensionnelles compte tenu du 4) de la remarque précédente.

A partir de maintenant, on utilisera une forme abrégée pour écrire des événements faisant intervenir des variables aléatoires. Ainsi, pour noter par exemple l'événement $\{\omega \in \Omega ; X(\omega) \leq C\}$, on écrira simplement $\{X \leq C\}$.

2.2 σ -algèbre générée par une collection de variables aléatoires

Définition 2.2.1. *On se donne une famille $(Y_\gamma)_{\gamma \in \mathcal{C}}$ de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ indexée par un ensemble \mathcal{C} quelconque. La tribu engendrée par la famille $(Y_\gamma)_{\gamma \in \mathcal{C}}$, notée $\Gamma = \sigma(Y_\gamma ; \gamma \in \mathcal{C})$, est définie comme la plus petite tribu sur Ω rendant mesurable chaque Y_γ pour $\gamma \in \mathcal{C}$. Autrement dit,*

$$\Gamma = \sigma \left(\{ \omega \in \Omega ; Y_\gamma(\omega) \in B \} ; \gamma \in \mathcal{C} ; B \in \mathcal{B}(\mathbb{R}^d) \right)$$

Remarque 2.2.1. 1) Si X est une variable aléatoire alors $\sigma(X)$ est une sous-tribu de \mathcal{F} .

2) Si X est une variable aléatoire et $f : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ est mesurable alors $f(X)$ est $\sigma(X)$ -mesurable.

Proposition 2.2.1. *Soit X une variable aléatoire. Une fonction $Y : \Omega \rightarrow \mathbb{R}$ est $\sigma(X)$ -mesurable ssi il existe une fonction borélienne f telle que $Y = f(X)$.*

Cette propriété sera démontrée en TD et est une simple conséquence du théorème de la classe monotone. Elle est cependant très importante dans la mesure où elle fait clairement comprendre ce que signifie "être $\sigma(X)$ -mesurable".

2.3 Distributions et fonctions de répartition

Définition 2.3.1. Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On appelle mesure image de \mathbb{P} par X ou encore loi de X la mesure de probabilité \mathbb{P}_X sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ définie par :

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A))$$

On sait que $\left\{ \prod_{i=1}^d (-\infty; c_i] ; c_i \in \mathbb{R} \right\}$ est un π -système qui génère $\mathcal{B}(\mathbb{R}^d)$. Par le théorème d'unicité des probabilités coïncidant sur un π -système, la loi de X est entièrement déterminée par la fonction

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X((-\infty; x])$$

ce qui motive la définition suivante :

Définition 2.3.2. Soit X une variable aléatoire à valeurs dans \mathbb{R}^d . On définit la fonction de répartition de X sur \mathbb{R}^d par

$$\forall x \in \mathbb{R}^d, F_X(x) = \mathbb{P}(X \leq x)$$

et on a par ce qui précède la propriété suivante :

Proposition 2.3.1. Deux variables aléatoires ayant même fonction de répartition ont même loi.

Les fonctions de répartition jouissent de propriétés très caractéristiques, du moins dans le cadre de la dimension 1. Dans le reste de cette section, on se limite donc à la dimension 1.

Proposition 2.3.2. Supposons que F soit la fonction de répartition d'une variable aléatoire réelle X . Alors :

- a) $F : \mathbb{R} \rightarrow [0, 1]$ est croissante.
- b) $\lim_{x \rightarrow +\infty} F(x) = 1$ et $\lim_{x \rightarrow -\infty} F(x) = 0$.
- c) F est continue à droite (avec des limites à gauche puisque croissante).

Démonstration.

- a) trivial.
- b) $\lim_{x \rightarrow +\infty} F(x) = \lim_{x \rightarrow +\infty} \mathbb{P}(\{\omega \in \Omega; X(\omega) \leq x\})$ et cette limite existe car F est croissante. Or $(A_n)_n = (\{X \leq n\})_n$ est une suite croissante d'événements tendant vers Ω . Donc par σ -additivité de \mathbb{P} , $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 1$. Par suite, $\lim_{x \rightarrow +\infty} F(x) = 1$. Un raisonnement analogue montre que $\lim_{x \rightarrow -\infty} F(x) = 0$.
- c) Soit $(x_n)_n$ une suite de points tendant vers x avec pour tout $n \in \mathbb{N}$, $x_n \geq x$. On a $F(x_n) = \mathbb{P}(X \leq x_n)$ et $\{X \leq x_n\} \downarrow \{X \leq x\}$ donc $F(x_n) \downarrow F(x)$ et F est continue à droite. \square

Existence de variables aléatoires ayant une fonction de répartition donnée

Théorème 2.3.1. Soit F une fonction ayant les propriétés a,b,c de la proposition précédente. Il existe alors une variable aléatoire X sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ telle que la fonction de répartition de X soit F .

Démonstration. $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \lambda)$ où λ est la mesure de Lebesgue sur $[0, 1]$. On pose pour $\omega \in \Omega$,

$$X(\omega) = \inf\{z \in \mathbb{R} ; F(z) \geq \omega\} = \sup\{y \in \mathbb{R} ; F(y) < \omega\}$$

Il est facile de voir d'après la définition de X que

$$\omega \leq F(c) \iff X(\omega) \leq c$$

Par conséquent, $\mathbb{P}(X \leq c) = \lambda(\omega \leq F(c)) = F(c)$. \square

2.4 Variables aléatoires à densités et variables aléatoires discrètes

Définition 2.4.1. On dira qu'une variable aléatoire X est à densité (par rapport à la mesure de Lebesgue) ssi la loi de X admet une densité (par rapport à la mesure de Lebesgue), i.e.

$$\begin{aligned} \exists f \in \mathcal{L}_+^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), dx), \forall A \in \mathcal{B}(\mathbb{R}^d), \\ \mathbb{P}_X(A) = \int_A f(x) dx \end{aligned}$$

Exemples fondamentaux :

1) X est une gaussienne de moyenne $m \in \mathbb{R}^d$ et de matrice de covariance $\Gamma \in S_d^+$ (ensemble des matrices symétriques définies positives) ssi X admet la densité sur \mathbb{R}^d donnée par

$$f(x) = \frac{1}{(2\pi)^{d/2} |\det \Gamma|^{1/2}} \exp \left(-\frac{1}{2} \langle \Gamma^{-1}(x - m), (x - m) \rangle \right)$$

où $\langle \cdot, \cdot \rangle$ désigne le produit scalaire euclidien usuel sur \mathbb{R}^d .

2) X est une exponentielle de paramètre λ ssi X admet la densité sur \mathbb{R} donnée par

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

Définition 2.4.2. Une variable aléatoire X est dite discrète s'il existe un ensemble dénombrable \mathcal{E} tel que $\mathbb{P}(X \in \mathcal{E}) = 1$. La loi de X est alors caractérisée par une suite $(p_e)_{e \in \mathcal{E}} \in [0, 1]^{\mathcal{E}}$ telle que $\sum_{e \in \mathcal{E}} p_e = 1$ et $p_e = \mathbb{P}(X = e)$.

Exemples fondamentaux :

1) X suit une loi de Bernoulli de paramètre $p \in [0, 1]$ ssi :

$$\mathbb{P}(X = 0) = 1 - \mathbb{P}(X = 1) = 1 - p$$

2) X suit une loi géométrique de paramètre $p \in [0, 1]$ ssi :

$$\forall n \in \mathbb{N}^*, \mathbb{P}(X = n) = p(1 - p)^{n-1}$$

3) X suit une loi de Poisson de paramètre $\lambda \in \mathbb{R}_+^*$ ssi :

$$\forall k \in \mathbb{N}, \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

2.5 TD Variables aléatoires

Exercice 2.5.1. (Théorème de la classe monotone, [30], p. 205) Soit \mathcal{H} une classe de fonctions bornées de Ω dans \mathbb{R} vérifiant

- (i) \mathcal{H} est un espace vectoriel sur \mathbb{R}
- (ii) la fonction constante 1 est un élément de \mathcal{H}
- (iii) si f_n est une suite croissante de fonctions positives de \mathcal{H} , tendant vers une fonction f bornée sur Ω , alors $f \in \mathcal{H}$.

Alors, si \mathcal{H} contient les fonctions indicatrices de tous les ensembles d'un π -système \mathcal{I} , \mathcal{H} contient aussi toutes les fonctions $\sigma(\mathcal{I})$ -mesurables et bornées sur Ω .

1) Soit \mathcal{D} la classe des ensembles F tels que $\mathbb{1}_F \in \mathcal{H}$. Montrer que \mathcal{D} est un λ -système et en déduire que \mathcal{D} contient $\sigma(\mathcal{I})$.

2) Soit f une fonction $\sigma(\mathcal{I})$ -mesurable telle qu'il existe $K \in \mathbb{N}$, $0 \leq f(s) \leq K$, $\forall s \in \Omega$. Pour $n \in \mathbb{N}$, on considère les fonctions en escalier approximant f ,

$$f_n(s) = \sum_{i=0}^{K2^n} i2^{-n} \mathbb{1}_{A(n,i)}(s),$$

où

$$A(n,i) = \{s, i2^{-n} \leq f(s) \leq (i+1)2^{-n}\}.$$

Montrer que $f_n \in \mathcal{H}$, puis que $f \in \mathcal{H}$. En déduire le théorème.

Chapitre 3

Espérance

Dans ce chapitre $(\Omega, \mathcal{F}, \mathbb{P})$ désigne un espace probabilisé. Nous rappellerons brièvement la construction de l'intégrale relativement à une mesure \mathbb{P} uniquement pour illustrer ce que nous appellerons par la suite "la machine standard". Ce chapitre est donc essentiellement un formulaire et un dictionnaire analyse-probabilités.

L'espérance, notée \mathbb{E} est simplement l'intégrale relativement à \mathbb{P} , c'est-à-dire que si X est une variable aléatoire intégrable relativement à \mathbb{P} ,

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$$

On l'appelle aussi parfois la moyenne de X . En particulier, si $X = \mathbb{1}_A$, $\mathbb{E}(X) = \mathbb{P}(A)$. On note \mathcal{L}^p , pour $p \in [1, +\infty]$, l'espace vectoriel constitué des variables aléatoires X telles que $|X|^p$ est intégrable relativement à \mathbb{P} . \mathbb{E} est une forme linéaire sur ces espaces. On note $\|\cdot\|_p$ les normes associées.

3.1 Construction de l'intégrale et théorèmes de convergence

3.1.1 La machine standard

1) Soit f une fonction simple (ou étagée) positive, c'est-à-dire de la forme

$$f = \sum_{k=1}^N a_k \mathbb{1}_{A_k}$$

où les a_k sont des réels positifs et A_k des éléments de \mathcal{F} . On définit l'intégrale de f relativement à \mathbb{P} (ou espérance) par

$$\mathbb{E}(f) = \sum_{k=1}^N a_k \mathbb{P}(A_k)$$

On vérifie aisément que cette définition est cohérente, c'est-à-dire qu'elle ne dépend pas de l'écriture de f choisie. D'autre part, des propriétés simples comme la linéarité, la positivité (si f, g sont deux

fonctions simples positives et $f \leq g$ alors $\mathbb{E}(f) \leq \mathbb{E}(g)$ se démontrent aisément.

2) Soit maintenant f une fonction positive mesurable. On définit son espérance par

$$\mathbb{E}(f) = \sup\{\mathbb{E}(h); h \text{ fonction simple positive, } h \leq f\}$$

Cette quantité peut être infinie et on dira que f est intégrable si elle est finie. *On vérifie que si h est une fonction simple positive, alors cette nouvelle définition reste cohérente avec l'ancienne.*

On peut construire des modes d'approximation simples pour une fonction mesurable positive. Par exemple si f est positive mesurable,

$$f_k := \sum_{l=0}^{k2^k-1} \frac{l}{2^k} \mathbb{1}_{[\frac{l}{2^k} \leq f < \frac{l+1}{2^k}]} + k \mathbb{1}_{[f \geq k]}$$

converge simplement vers f et uniformément si f est en plus bornée.

Vient alors le théorème de convergence monotone (MON) pour les fonctions mesurables positives :

Si $(f_n)_n$ est une suite de fonctions mesurables positives convergeant en croissant \mathbb{P} p.s. vers une fonction f (mesurable positive) alors

$$\mathbb{E}(f_n) \uparrow \mathbb{E}(f) \tag{3.1.1}$$

La démonstration de la convergence monotone NON TERMINE A REPREDRE

Cette démonstration n'est pas difficile car ce théorème vient essentiellement du fait que si $A_n \uparrow A$, alors $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$. Voici la marche de la démonstration. On va la faire dans le cas $\mathbb{E}f < +\infty$; le cas $\mathbb{E}f = +\infty$ suit exactement la même démarche.

On commence par remarquer qu'il existe une suite croissante de fonctions simples g_n telle que $\mathbb{E}g_n \uparrow \mathbb{E}f$ et $g_n \uparrow f$ p.s.. Pour cela on prend une suite croissante g_n de fonctions simples telles que $\mathbb{E}g_n \uparrow \mathbb{E}f$. La croissance vient du fait qu'on peut toujours remplacer g_n par $\sup_{i=1, \dots, n} g_i$. Supposons par contradiction qu'on n'ait pas $g_n \uparrow f$ p.s.. Alors $\mathbb{P}(\sup g_n < f) > 0$ et par la propriété de continuité de la mesure pour les suites croissantes pour quelque $\eta > 0$, l'ensemble $A := \cap_n \{g_n \leq f - \eta\}$ est de mesure strictement positive. Alors en remplaçant g_n par $h_n := g_n + \eta \mathbb{1}_A$, on arrive à une contradiction car $h_n \leq g$ et

$$\mathbb{E}g \geq \mathbb{E}h_n \geq \mathbb{E}g_n + \eta \mathbb{P}(A) \rightarrow \mathbb{E}g + \eta \mathbb{P}(A).$$

La deuxième étape consiste à fixer g simple telle que $g \leq f$ et $\mathbb{E}g \geq \mathbb{E}f - \varepsilon$. Ensuite on prend par l'étape précédente des fonctions simples g_n telles que $\mathbb{P}(g_n \leq f - \varepsilon) < \frac{1}{2^n}$. Finalement on montre que $\mathbb{E}g_n \rightarrow \mathbb{E}g$. Comme $\mathbb{P}(g - g_n) > \varepsilon \rightarrow 0$, on peut prendre n assez grand pour que cette probabilité soit plus petite que ε . Comme g_n et g sont simples, en nommant A_k la famille finie des intersections d'ensembles où elles sont chacune constantes, on peut écrire $g_n = \sum_k a_k \mathbb{1}_{A_k}$ et $g = \sum_k b_k \mathbb{1}_{A_k}$ et on a alors $\mathbb{P}(\cup_k |b_k - a_k| > \varepsilon) \leq \varepsilon$. Il suffit d'exprimer les espérances de ces deux fonctions simples pour conclure que $\mathbb{E}g - \mathbb{E}g_n < \varepsilon + \varepsilon \sup g$.

Le lemme de Fatou se déduit aisément du théorème de convergence monotone. On démontre aussi à ce stade la linéarité et la positivité de l'espérance.

3) soit f une fonction mesurable de signe quelconque. On décompose f en sa partie positive $f^+ = \sup(f, 0)$ et sa partie négative $f^- = \sup(-f, 0)$ si bien que $f = f^+ - f^-$ et $|f| = f^+ + f^-$. On dira que f est intégrable ssi $|f|$ l'est (ce qui implique que f^+ et f^- le sont). Si f est intégrable, on définit son espérance par

$$\mathbb{E}(f) = \mathbb{E}(f^+) - \mathbb{E}(f^-)$$

L'ensemble des fonctions intégrables forme un espace vectoriel sur lequel l'espérance ainsi définie est une forme linéaire positive. C'est à ce stade que le théorème de convergence dominée s'établit modulo le lemme de Sheffé (voir TD).

On remarquera que la construction de l'intégrale se fait selon trois grandes étapes : on commence par définir l'espérance sur les fonctions simples, puis sur les fonctions mesurables positives grâce à une propriété de monotonie et enfin on passe au cas d'une fonction de signe quelconque grâce à la décomposition de f en sa partie positive et sa partie négative. Cette structure en trois étapes est constamment utilisée dans diverses preuves. Pour être plus explicite, on est souvent amené à montrer une propriété "linéaire" pour les fonctions mesurables. Il est souvent aisé de la montrer pour les fonctions simples puis par le théorème de convergence monotone pour les fonctions mesurables positives (puisque toute fonction mesurable positive est limite simple croissante d'une suite de fonctions simples) et enfin par décomposition en partie positive-négative pour les fonctions mesurables de signe quelconque. Nous appellerons par la suite ce genre de preuve, une "preuve par machine standard".

3.1.2 Théorèmes de convergence

Soit $(X_n)_n$ une suite de variables aléatoires et X une variable aléatoire.

- (MON) : Si $0 \leq X_n \uparrow X$, \mathbb{P} p.s., alors $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$.
- (Lemme de Fatou) : $X_n \geq 0$ alors $\mathbb{E}(\liminf X_n) \leq \liminf \mathbb{E}(X_n)$.
- (Convergence Dominée) : $|X_n| \leq Y$, \mathbb{P} p.s., $\mathbb{E}(Y) < +\infty$ et $(X_n)_n$ tend presque sûrement vers X alors $X_n \rightarrow X$ dans \mathcal{L}^1 .
- (Convergence bornée) : Si $|X_n| \leq C$ et si X_n tend vers X p.s., alors $X_n \rightarrow X$ dans \mathcal{L}^1 .

3.2 Inégalités de Markov

Si $X \in \mathcal{L}^p$, $\mathbb{P}(|X| \geq c) \leq c^{-p} \mathbb{E}(|X|^p)$.

Si X est une variable aléatoire positive, pour tout réel c et tout $\tau > 0$, on a

$$\mathbb{P}(Y \geq c) \leq e^{-\tau c} \mathbb{E}(e^{\tau Y})$$

3.3 Sommes de variables aléatoires

- a) Si X est une variable aléatoire positive et intégrable alors $\mathbb{P}(X < +\infty) = 1$.
- b) Si $(Z_k)_k$ est une suite de variables aléatoires positives alors

$$\mathbb{E}\left(\sum_k Z_k\right) = \sum_k \mathbb{E}(Z_k)$$

- c) Si $(Z_k)_k$ est une suite de variables aléatoires positives telles que $\sum_k \mathbb{E}(Z_k) < +\infty$ alors presque sûrement, on a :

$$\sum_k Z_k < +\infty \text{ et } Z_k \rightarrow 0$$

3.4 Inégalité de Jensen

Théorème 3.4.1. Soit $\phi : G \rightarrow \mathbb{R}$ une fonction convexe et X une variable aléatoire \mathbb{P} p.s. à valeurs dans G où G est un intervalle de \mathbb{R} ouvert.

On suppose que :

$$\mathbb{E}(|X|) < +\infty, \mathbb{E}(|\phi(X)|) < +\infty$$

Alors :

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))$$

Démonstration. (Pour un énoncé et une démonstration en dimension N , voir le TD). Pour $u < v < w$, on pose $\Delta_{u,v} = \frac{\phi(v) - \phi(u)}{v - u}$. Puisque ϕ est convexe,

$$\Delta_{u,v} \leq \Delta_{v,w}$$

$(D_-\phi)(v) = \lim_{u \uparrow v} \Delta_{u,v}$ et $(D_+\phi)(v) = \lim_{w \downarrow v} \Delta_{v,w}$ sont les dérivées à gauche et à droite de ϕ . On a donc

$$(D_-\phi)(v) \leq (D_+\phi)(v)$$

D'autre part, par un argument du même genre en prenant quatre points, $D_-\phi$, $D_+\phi$ sont croissantes.

Si $x < w < v$, $\Delta_{x,w} \leq \Delta_{w,v}$ et par passage à la limite quand w tend en croissant vers v , on obtient $\Delta_{x,v} \leq (D_-\phi)(v)$ car ϕ est continue sur G .

On a donc $\phi(v) - \phi(x) \leq (v - x)(D_-\phi)(v) \leq m(v - x)$ pour $m \in [(D_-\phi)(v); (D_+\phi)(v)]$.

En raisonnant de même si $v < x$, on a pour tout $v, x \in G$, $\phi(x) \geq m(x - v) + \phi(v)$ pour $m \in [(D_-\phi)(v); (D_+\phi)(v)]$.

Donc si $v = \mu = \mathbb{E}(X) \in G$ ($a < X < b \Rightarrow a < \mathbb{E}(X) < b$), et $x = X$, on a :

$$\phi(X) \geq m(X - \mu) + \phi(\mu)$$

et en intégrant on obtient

$$\mathbb{E}(\phi(X)) \geq \phi(\mu) = \phi(\mathbb{E}(X))$$

□

3.5 Monotonie des normes \mathcal{L}^p

On rappelle que \mathcal{L}^p est l'espace vectoriel des variables aléatoires X dont la norme \mathcal{L}^p est finie. La norme \mathcal{L}^p d'une variable aléatoire X est $\|X\|_p = \mathbb{E}(|X|^p)^{1/p} \in [0, +\infty)$. \mathcal{L}^p est un espace vectoriel et $\|\cdot\|_p$ est une semi-norme. Si $X \in \mathcal{L}^p$, on dit que X a un moment d'ordre p . \mathbb{L}^p est le quotient de \mathcal{L}^p pour la relation d'équivalence $X \sim Y \iff X = Y$, \mathbb{P} p.s..

Théorème 3.5.1. \mathbb{L}^p est un espace vectoriel normé complet.

On rappelle que pour démontrer ce théorème, on utilise ce que l'on appelle parfois "la réciproque du théorème de convergence dominée", à savoir :

Proposition 3.5.1. Si $(X_n)_n$ est une suite de Cauchy dans \mathcal{L}^p alors il existe une sous-suite $(Y_n)_n$ de $(X_n)_n$ et une variable aléatoire X ayant un moment d'ordre p telle que $(Y_n)_n$ converge en norme \mathcal{L}^p vers X .

Une propriété particulière aux espaces de probabilités (par rapport aux espaces associés à une mesure quelconque) est la suivante

Proposition 3.5.2. Si $1 \leq p < r \leq +\infty$ alors si Y admet un moment d'ordre r , Y admet un moment d'ordre p et on a

$$\|Y\|_p \leq \|Y\|_r$$

Autrement dit, plus une variable aléatoire a des moments d'ordre élevé, mieux c'est. La démonstration facile repose sur l'inégalité de Jensen appliquée, après troncature, à la fonction $\phi(x) = x^{r/p}$.

3.6 Hölder et Minkowski

On rappelle ces deux inégalités d'un usage constant et qui dérivent de l'inégalité de Jensen

Théorème 3.6.1. (*Inégalité de Hölder*)

Soit $1 < p, q < +\infty$, tels que $p^{-1} + q^{-1} = 1$ et $X \in \mathcal{L}^p, Y \in \mathcal{L}^q$. On a alors l'inégalité de Hölder

$$\mathbb{E}(|XY|) \leq \|X\|_p \|Y\|_q$$

Démonstration. On peut bien sûr se restreindre au cas $X, Y \geq 0$ et $\mathbb{E}(X^p) > 0$. D'autre part, quitte à considérer $X \wedge n$ et $Y \wedge n$, puis à passer à la limite quand n tend vers $+\infty$ par le théorème de convergence monotone, on peut supposer X, Y bornées. On définit une probabilité \mathbb{Q} par

$$d\mathbb{Q} = \frac{X^p d\mathbb{P}}{\mathbb{E}(X^p)}$$

et une variable aléatoire Z par

$$Z = \mathbb{1}_{X>0} \frac{Y}{X^{p-1}}$$

Puisque $\mathbb{E}_{\mathbb{Q}}(Z)^q \leq \mathbb{E}_{\mathbb{Q}}(Z^q)$, on a

$$\mathbb{E}_{\mathbb{P}}(XY) \leq \|X\|_p \|Y \mathbb{1}_{X>0}\|_q \leq \|X\|_p \|Y\|_q$$

□

Théorème 3.6.2. (*Inégalité de Minkowski*)

Soit $X, Y \in \mathcal{L}^p$, on a l'inégalité de Minkowski

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$$

Démonstration. L'inégalité de Hölder donne

$$\begin{aligned} \mathbb{E}(|X + Y|^p) &= \mathbb{E}(|X||X + Y|^{p-1}) + \mathbb{E}(|Y||X + Y|^{p-1}) \\ &\leq A\|X\|_p + A\|Y\|_p \end{aligned}$$

où $A = \| |X + Y|^{p-1} \|_q = \mathbb{E}(|X + Y|)^{1/q}$. Une majoration grossière montre que A est finie et le résultat tombe. □

3.7 Cas $p = 2$

\mathbb{L}^2 jouit d'un place particulière au sein des espaces \mathbb{L}^p puisque c'est un espace de Hilbert pour le produit scalaire $\langle X, Y \rangle = \mathbb{E}(XY)$.

Si X, Y sont deux variables aléatoires ayant des moments d'ordre 2, on appelle covariance de X et Y le réel défini par

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$Cov(\cdot, \cdot)$ est un opérateur bilinéaire symétrique sur \mathbb{L}^2 .

La variance de X est $Var(X) = \sigma_X^2 = Cov(X, X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0$.

En anticipant sur la suite, si X et Y sont indépendantes alors $Cov(X, Y) = 0$ et la réciproque est fautive. Si $Cov(X, Y) = 0$ alors les variances s'ajoutent : $Var(X + Y) = Var(X) + Var(Y)$ (théorème de Pythagore).

3.8 T.D. Intégration et Lebesgue

Certains des exercices qui suivent font appel aux théorèmes de dérivation sous le signe intégral et au théorème de Fubini. Ces théorèmes sont rappelés dans la prochaine section et le théorème de Fubini sera démontré dans le prochain chapitre. On notera λ la mesure de Lebesgue.

Exercice 3.8.1. L'inégalité de Jensen en dimension N . Soit $U \subset \mathbb{R}^n$ un ouvert convexe et $\varphi : U \rightarrow \mathbb{R}^n$ une fonction convexe. On note x, y des éléments de \mathbb{R}^n et $x.y$ leur produit scalaire canonique.

- 1) Rappeler la définition de la convexité, pour un ensemble et pour une fonction.
- 2) Donner la définition de $\limsup_{y \rightarrow x} \varphi(y)$. On rappelle qu'une fonction φ est continue en un point x si et seulement si

$$\liminf_{y \rightarrow x} \varphi(y) \geq \varphi(x) \geq \limsup_{y \rightarrow x} \varphi(y).$$

Montrer que φ est continue. Indication : on vérifiera et utilisera successivement les deux inégalités qui suivent

$$\varphi(y) \leq \frac{1}{2}\varphi(2y - x) + \frac{1}{2}\varphi(x), \quad \varphi(x) - \frac{1}{2}\varphi(x - z) \leq \frac{1}{2}\varphi(x + z).$$

Prendre la limite inférieure ou supérieure des deux membres de ces inégalités quand $y \rightarrow x$ et quand $z \rightarrow 0$.
Solution. Les deux inégalités découlent immédiatement de la convexité de φ , la première parce que y est le milieu de $[2y - x, x]$ et la deuxième parce que x est le milieu de $[x - z, x + z]$. On fait la limite supérieure de la première inégalité quand y tend vers x . Cela donne

$$\limsup_{y \rightarrow x} \varphi(y) \leq \frac{1}{2} \limsup_{y \rightarrow x} \varphi(2y - x) + \frac{1}{2}\varphi(x) \leq \frac{1}{2} \limsup_{y \rightarrow x} \varphi(y) + \frac{1}{2}\varphi(x).$$

En effet si $y \rightarrow x$, $2y - x \rightarrow x$. Donc $\limsup_{y \rightarrow x} \varphi(y) \leq \varphi(x)$. En appliquant maintenant une limite inférieure à la deuxième inégalité,

$$\varphi(x) - \limsup_{z \rightarrow 0} \frac{1}{2}\varphi(x - z) \leq \liminf_{z \rightarrow 0} \frac{1}{2}\varphi(x + z).$$

En utilisant la première inégalité prouvée, $\frac{1}{2}\varphi(x) \geq \frac{1}{2} \limsup_{z \rightarrow 0} \varphi(x - z)$. Donc

$$\frac{1}{2}\varphi(x) \leq \frac{1}{2} \liminf_{z \rightarrow 0} \varphi(x + z).$$

- 3) Indiquer où l'hypothèse que U est ouvert a été utilisée. Donner un exemple de fonction convexe sur un intervalle fermé et qui n'est pas continue.

4) Montrer que l'épigraphe de φ , $\{x, z\} \in \mathbb{R}^n \times \mathbb{R}$, $z > \varphi(x)$ est un ouvert convexe de $\mathbb{R}^n \times \mathbb{R}$.

5) Montrer que tout hyperplan de $\mathbb{R}^n \times \mathbb{R}$ peut s'écrire sous la forme $\{(x, z), a.x + bz = c\}$ où $a \in \mathbb{R}^n$, $b, c \in \mathbb{R}$. On considère le point $(x_0, \varphi(x_0)) \in U$. Montrer qu'il existe des réels $a \in \mathbb{R}^n$, $b, c \in \mathbb{R}$ tels que $\forall x \in U$,

$$a.x + b\varphi(x) \geq c \text{ et } a.x_0 + b\varphi(x_0) = 0.$$

Solution : appliquer Hahn-Banach géométrique à l'épigraphe ouvert de φ et au point $(x_0, \varphi(x_0))$.

- 6) Montrer que $b \neq 0$. En déduire qu'il existe $v \in \mathbb{R}^n$

$$\varphi(x) \geq v.(x - x_0) + \varphi(x_0).$$

- 7) L'ensemble des v vérifiant l'inégalité précédente s'appelle *sous-différentiel* de φ en x_0 et il est noté $\partial\varphi(x_0)$. Montrer que $\partial\varphi(x_0)$ est convexe.

8) On suppose que μ est une mesure de probabilité sur U telle que $\int_U |\varphi(x)| d\mu(x) < \infty$ et $\int_U |x| d\mu(x) < \infty$. Montrer que $\int_U x d\mu(x) = x_0$ appartient à U . Quelle est l'interprétation géométrique de x_0 si μ est la mesure de Lebesgue ?

9) Montrer l'inégalité de Jensen

$$\int \varphi(x) d\mu(x) \geq \varphi\left(\int_U x d\mu(x)\right).$$

Exercice 3.8.2. Tout graphe est de mesure nulle.

1) Démontrer que le graphe d'une application continue f de \mathbb{R} dans \mathbb{R} est un sous-ensemble de mesure nulle dans \mathbb{R}^2 . (On montrera que la portion du graphe comprise entre les abscisses $-k$ et k peut être recouverte par des rectangles dont la somme des mesures est arbitrairement petite).

2) Même question si $f \in \mathbb{L}_{loc}^1(\mathbb{R})$, puis si f est borélienne.

Solution

Comme $[-k, k]$ est un compact, la fonction f considérée est uniformément continue sur $[-k, k]$. Il existe donc $n \in \mathbb{N}$ tel que $|x - y| \leq \frac{1}{n}$, $x, y \in [-k, k] \Rightarrow |f(x) - f(y)| \leq \frac{\varepsilon}{4k}$. On recouvre $[-k, k]$ par n intervalles consécutifs de longueur $\frac{2k}{n}$. On vérifie aisément que les $2n$ rectangles $[l\frac{1}{n}, (l+1)\frac{1}{n}] \times [f((l+\frac{1}{2})\frac{k}{n}) - \frac{\varepsilon}{4k}, f((l+\frac{1}{2})\frac{k}{n}) + \frac{\varepsilon}{4k}]$, $l = -n, -(n-1), \dots, n-1$ recouvrent le graphe de f sur $[-k, k]$. Donc la mesure de ce graphe est plus petite que $2n \frac{2k}{n} \frac{\varepsilon}{4k} = \varepsilon$.

Solution plus élégante : Soit $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $\varphi(x, y) = y - f(x)$. Alors $\varphi^{-1}(0) = \{(x, f(x)), x \in \mathbb{R}\}$ et par le théorème de Fubini-Tonnelli,

$$\lambda(\{(x, f(x)), x \in \mathbb{R}\}) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{y=f(x)}(x, y) dy \right) dx = 0$$

car les singletons sont de mesure de Lebesgue nulle.

Exercice 3.8.3. Soit A un borélien de \mathbb{R}^N et soit h égale à $+\infty$ dans A et à 0 dans son complémentaire. Démontrer que l'on a

$$\int_{\mathbb{R}^N} h(x) dx = 0 \text{ si } \lambda(A) = 0, \quad \int_{\mathbb{R}^N} h(x) dx = +\infty \text{ si } \lambda(A) > 0.$$

Solution

Supposons que $\lambda(A) = 0$. On pose $h_k(x) = \min(h(x), k)$. Comme A est de mesure nulle, on a par définition de $\lambda(A)$, $\int \mathbb{1}_A dx = \lambda(A) = 0$. Donc

$$\int h_k(x) dx = k \int \mathbb{1}_A dx = 0.$$

Par le théorème de la convergence monotone, $\int h(x) dx = \lim_{k \rightarrow \infty} \int h_k(x) dx = 0$. Si maintenant $\lambda(A) > 0$, on a, encore par le théorème de la convergence monotone,

$$\int h(x) dx = \lim_{k \rightarrow \infty} \int h_k(x) dx = \lim_{k \rightarrow \infty} k \lambda(A) = +\infty.$$

On déduit immédiatement que si $f \geq 0$ vérifie $\int f < \infty$, alors l'ensemble $A = \{x, f(x) = +\infty\}$ est de mesure nulle.

Exercice 3.8.4. Soit f et g deux fonctions continues telles que $f(x) = g(x)$ p.p.. Montrer que f et g sont égales partout.

Solution

Comme les boules sont de mesure strictement positive, on déduit que $f(x) = g(x)$ sur un ensemble dense de \mathbb{R}^N , et, par continuité, partout.

Exercice 3.8.5. Soit f une fonction intégrable sur \mathbb{R} . Démontrer que

$$\int_a^b f(x)dx \rightarrow \int_{\mathbb{R}} f(x)dx \text{ quand } a \rightarrow -\infty, b \rightarrow +\infty.$$

Généralisation : soit A_n une suite croissante d'ensembles tels que $\bigcup_n A_n = \mathbb{R}^N$. Démontrer que $\int_{\mathbb{R}^N} f(x)dx = \lim_{n \rightarrow \infty} \int_{A_n} f(x)dx$.

Solution

Appliquer le théorème de convergence dominée de Lebesgue à la suite $(f\mathbb{1}_{A_n})_n$.

Exercice 3.8.6. Soit h une fonction positive intégrable sur \mathbb{R}^N et A_n des boréliens de \mathbb{R}^N tels que $\lambda(A_n) \rightarrow 0$. Démontrer que $\int_{A_n} h(x)dx \rightarrow 0$ quand $n \rightarrow \infty$.

Solution

Notons $[h \leq M]$ l'ensemble $\{x, h(x) \leq M\}$. On écrit

$$\int_{A_n} h(x) \leq \int_{A_n \cap [h(x) \leq M]} h(x)dx + \int_{A_n \cap [h(x) > M]} h(x)dx \leq M\lambda(A_n) + \int_{[h(x) > M]} h(x)dx. \quad (3.8.1)$$

Par le théorème de la convergence dominée, on a $\int_{[h(x) > M]} h(x)dx \rightarrow 0$. En effet, $h(x)\mathbb{1}_{[h(x) > M]}$ tend vers zéro presque partout car $\lambda([h = +\infty]) = 0$ (Exercice ??) et $|h(x)|$ est chapeau intégrable. On peut donc fixer pour tout $\varepsilon > 0$, M tel que $\int_{[h(x) > M]} h(x)dx < \varepsilon$. On choisit ensuite n suffisamment grand pour que $M\lambda(A_n) < \varepsilon$ et on déduit alors de (3.8.1), $\int_{A_n} h(x)dx < 2\varepsilon$.

Exercice 3.8.7. Calculer $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ à l'aide de l'intégrale $\int_0^1 \frac{dx}{1+x}$. Indication : on posera $f_n(x) = \sum_0^n (-1)^k x^k$ pour $x \in [0, 1]$. On trouvera un chapeau intégrable pour les f_n et on appliquera le théorème de Lebesgue.

Solution

On sait que $f_n(x) \rightarrow \sum_0^{\infty} (-1)^k x^k = \frac{1}{1+x}$ pour tout $0 \leq x < 1$. Par ailleurs, la série étant alternée, $f_0(x) = 1$ est "chapeau intégrable" pour $f_n(x) \geq 0$ sur $[0, 1]$. Par le théorème de Lebesgue, on déduit que $f_n(x) \rightarrow \frac{1}{1+x}$ dans $\mathbb{L}^1(0, 1)$, ce qui implique que $\int_0^1 f_n(x)dx \rightarrow \int_0^1 \frac{1}{1+x}dx = \text{Log}2$. En intégrant terme à terme, on obtient $\sum_1^{\infty} \frac{(-1)^{n+1}}{n} = \text{Log}2$.

Exercice 3.8.8. Soit $f \in \mathbb{L}^1(\mathbb{R})$. Calculer la limite de $n \int_1^{\infty} f(nx)dx$ quand $n \rightarrow +\infty$.

Solution

En posant $y = nx$, on a $n \int_1^{\infty} f(nx)dx = \int_n^{\infty} f(y)dy = \int_{\mathbb{R}} \mathbb{1}_{[n, +\infty[}(y)f(y)dy$. La fonction $f(y)\mathbb{1}_{[n, \infty[}(y)$ tend ponctuellement vers zéro quand $n \rightarrow \infty$ et $|f|$ en est chapeau intégrable. Donc par le théorème de Lebesgue, l'intégrale considérée tend vers zéro.

Exercice 3.8.9. (Dédurre le Lemme de Fatou du théorème de convergence monotone)

On rappelle la définition de la limite inférieure d'une suite a_n de réels :

$$\liminf_{n \rightarrow \infty} a_n = \lim_{N \rightarrow \infty} \inf_{n \geq N} a_n$$

Cette limite est une limite croissante et donc elle existe toujours dans $\mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$.

i) Remarquer que si g_n est une suite de fonctions sommables positives, alors $\int \inf_n g_n \leq \inf_n \int g_n$.

ii) Soit f_n une suite de fonctions à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$. On pose $g_N = \inf_{n \geq N} f_n$. Montrer que $\lim_{N \rightarrow \infty} \int g_N = \int \lim_{N \rightarrow \infty} g_N$.

iii) Dédurre des questions précédentes le lemme de Fatou :

$$\int \liminf f_n \leq \liminf \int f_n.$$

Solution

i) Si g_n est une suite de fonctions sommables positives, on a pour tout n $\int \inf_n g_n \leq \int g_n$. En prenant l'infimum des deux cotés, on obtient $\int \inf_n g_n \leq \inf_n \int g_n$.

ii) La suite $(g_N)_N$ est une suite croissante de fonctions positives ou nulles. Par le théorème de la convergence monotone, on a donc $\lim_{N \rightarrow \infty} \int g_N = \int \lim_{N \rightarrow \infty} g_N$.

iii) En combinant l'égalité du ii) et l'inégalité du i), on obtient directement le lemme de Fatou :

$$\int \liminf f_n \leq \liminf \int f_n.$$

Exercice 3.8.10. *contre-exemples*

Donner des exemples de suites de fonctions $f_n \in \mathbb{L}^1(0,1)$ telles que

- $f_n \rightarrow f$ dans \mathbb{L}^1 et $f_n(x)$ ne tend pas presque partout vers $f(x)$.
- f_n ne tend pas vers f dans \mathbb{L}^1 et $f_n(x)$ tend vers $f(x)$ presque partout.

Solution

- On considère la suite de fonctions f_n définies sur $[0,1]$ par

$$f_1 = 1 \text{ sur } [0,1]$$

$$f_2 = 1 \text{ sur } [0, \frac{1}{2}], 0 \text{ ailleurs}$$

$$f_3 = 1 \text{ sur } [\frac{1}{2}, 1], 0 \text{ ailleurs}$$

$$f_4 = 1 \text{ sur } [0, \frac{1}{4}], 0 \text{ ailleurs}$$

$$f_5 = 1 \text{ sur } [\frac{1}{4}, \frac{1}{2}], 0 \text{ ailleurs}$$

$$f_6 = 1 \text{ sur } [\frac{1}{2}, \frac{3}{4}], 0 \text{ ailleurs...}$$

La formule générale est :

$$\text{Si } 2^{k-1} \leq n < 2^k, \text{ alors } f_n = 1 \text{ sur } [\frac{n-2^{k-1}}{2^{k-1}}, \frac{n+1-2^{k-1}}{2^{k-1}}] = [\frac{n}{2^{k-1}} - 1, \frac{n+1}{2^{k-1}} - 1].$$

Cette suite de fonctions est une "bosse roulante" dont la largeur tend vers zéro, mais qui balaye constamment l'intervalle $[0,1]$. Elle est positive et son intégrale tend vers zéro. Donc $f_n \rightarrow 0$ dans \mathbb{L}^1 . Par contre, $f_n(x)$ ne tend jamais vers 0. En effet, si x est un point de $[0,1]$, on peut poser par division euclidienne $x = \frac{N}{2^{k-1}} + r$, avec $N < 2^{k-1}$ et $r < \frac{1}{2^{k-1}}$. Posons $n = N + 2^{k-1}$, ce qui permet d'écrire x sous la forme $x = \frac{n-2^{k-1}}{2^{k-1}} + r$. On voit que $f_n(x) = 1$, alors que $f_{n+2}(x) = 0$ par exemple. Donc $f_n(x)$ ne converge nulle part !

- On pose $f_n(x) = n$ si $0 \leq x \leq \frac{1}{n}$ et $f_n(x) = 0$ sinon. Alors $f_n(x)$ tend vers 0 pour tout $x \in]0,1]$. Par contre, $\int f_n = 1$ et donc f_n ne tend pas vers 0 dans \mathbb{L}^1 .

Exercice 3.8.11. *Coupe d'un ensemble mesurable*

Soit E un sous-ensemble de mesure bornée de \mathbb{R}^N . Démontrer que pour tout $0 \leq \alpha \leq 1$, E contient un sous-ensemble de mesure $\alpha \cdot \lambda(E)$.

Solution

On pose $x = (x_1, \dots, x_N)$ et

$$m(r) = \lambda(\{x \in E, x_1 \leq r\}) = \int_{\mathbb{R}^N} \mathbb{1}_{\{x, x_1 \leq r\}} dx.$$

Une application immédiate du théorème de Lebesgue montre que m est une fonction continue. Toujours par le même théorème, on montre que $m(r) \rightarrow 0$ quand $r \rightarrow -\infty$ et $m(r) \rightarrow \lambda(E)$ quand $r \rightarrow +\infty$. (Dans tous les cas, on utilise tout bonnement la fonction $\mathbb{1}_E$ comme chapeau intégrable). Par le théorème des valeurs intermédiaires appliqué à la fonction $m(r)$, il existe donc pour tout $0 < a < \lambda(E)$ un réel r tel que $m(r) = a$.

Exercice 3.8.12. *Convergence \mathbb{L}^1 quand il y a "conservation de la masse" (Lemme de Scheffé)*

Montrer que si une suite de fonctions $(w_n)_n \in \mathbb{L}^1$ vérifie $w_n \geq 0$, $w_n(x) \rightarrow w(x)$ presque partout, $w \in \mathbb{L}^1$ et $\int w_n \rightarrow \int w$, alors $w_n \rightarrow w$ dans \mathbb{L}^1 . Indication : écrire $\int w - w_n = \int (w - w_n)^+ - \int (w - w_n)^-$ et $\int |w - w_n| = \int (w - w_n)^+ + \int (w - w_n)^-$ et appliquer le théorème de Lebesgue.

Solution

Comme $w_n \geq 0$, on a $0 \leq (w - w_n)^+ \leq w \in \mathbb{L}^1$ et donc par le théorème de Lebesgue $\int (w - w_n)^+ \rightarrow 0$. On déduit de la relation $\int (w - w_n)^+ - \int (w - w_n)^- = \int w - w_n \rightarrow 0$ que $\int (w - w_n)^-$ tend aussi vers zéro et finalement que la somme des deux mêmes intégrales $\int |w - w_n| = \int (w - w_n)^+ + \int (w - w_n)^-$ tend vers zéro.

Exercice 3.8.13. *Critère d'intégrabilité sur un borné B de \mathbb{R}^N*

On va montrer dans cet exercice que $f \in \mathbb{L}^1(B)$ si et seulement si $\forall \varepsilon \exists \eta, \lambda(K) \leq \eta \Rightarrow \int_K |f| \leq \varepsilon$.

i) Montrer que si le critère d'intégration est vérifié, alors f est intégrable.

ii) Pour la réciproque, on raisonne par contradiction et on suppose qu'il existe $f \in \mathbb{L}^1(B)$ telle que le critère d'intégration ne soit pas vérifié. Montrer qu'il existe alors $\varepsilon > 0$ et une suite K_n de sous-ensembles de B tels que $\int_{K_n} |f| \geq \varepsilon$ et $\lambda(K_n) \leq \frac{1}{2^n}$.

On pose $J_n = \cup_{k=n+1}^{\infty} K_k$. Montrer que $\int_{J_n} |f| \rightarrow 0$ et conclure.

Autre méthode : appliquer directement le résultat de l'exercice 3.8.6!

iii) (facultatif) Adapter le critère d'intégrabilité à \mathbb{R}^N .

Solution

On applique le critère d'intégration avec par exemple $\varepsilon = 1$. Il existe donc η tel que $\lambda(K) \leq \eta \Rightarrow \int_K |f| \leq 1$. Comme B est borné, il peut être recouvert par un nombre fini de boules de mesure inférieure à η , B_1, \dots, B_k . On a donc $\int_B |f| \leq \sum_{i=1}^k \int_{B_i} |f| \leq k$, ce qui prouve que $f \in \mathbb{L}^1(B)$.

Réciproquement, soit $f \in \mathbb{L}^1(B)$ et supposons par contradiction qu'il existe ε et une suite K_n de sous-ensembles de B tels que $\int_{K_n} |f| \geq \varepsilon$ et $\lambda(K_n) \rightarrow 0$. Quitte à extraire une sous-suite, on peut supposer que $\lambda(K_n) \leq \frac{1}{2^n}$. On pose $J_n = \cup_{k=n+1}^{\infty} K_k$. Donc $\text{mes}(J_n) \leq \sum_{k=n+1}^{\infty} \frac{1}{2^k} \leq \frac{1}{2^n}$. La suite g_n des fonctions caractéristiques des J_n est une suite décroissante de fonctions positives dont l'intégrale tend vers zéro. Elle converge donc presque partout vers zéro. Revenant à f , on a $\int_{J_n} |f| = \int_B g_n(x) |f(x)| dx$. Comme $g_n(x) |f(x)|$ tend presque partout vers zéro et que $|f|$ est chapeau intégrable, on conclut que $\int_{J_n} |f| \rightarrow 0$, ce qui contredit le fait que $\int_{J_n} |f| \geq \varepsilon$.

Exercice 3.8.14. *Réciproque du Théorème de Lebesgue*

On va montrer que si f_n est une suite convergente dans \mathbb{L}^1 , alors il existe une sous-suite qui converge presque partout et qui a un chapeau intégrable.

i) Montrer que si la suite f_n converge vers f dans \mathbb{L}^1 , on peut en extraire une sous-suite telle que $\|f_{i_{n+1}} - f_{i_n}\|_1 < 2^{-n}$.

ii) On pose $g_n = f_{i_{n+1}} - f_{i_n}$; montrer en appliquant le théorème de convergence monotone que la série $\sum |g_n(x)|$ appartient à \mathbb{L}^1 et converge donc pour presque tout x . En déduire que f_{i_n} converge presque partout.

iii) Déduire aussi que f_{i_n} a un chapeau intégrable.

iv) Adapter le raisonnement précédent à une suite f_n qui est de Cauchy dans \mathbb{L}^1 pour montrer que \mathbb{L}^1 est un espace complet.

Solution

- i) Si $f_n \rightarrow f$ dans \mathbb{L}^1 , alors $\exists i_n, \|f_{i_n} - f\|_1 < 2^{-n-1}$. Donc $\|f_{i_{n+1}} - f_{i_n}\|_1 < 2^{-n-1} + 2^{-n-2} < 2^{-n}$.
- ii) On a $\sum_n \|g_n\|_1 \leq \sum 2^{-n} = 1$. On pose $h(x) = \sum_n |g_n(x)|$. Cette série est positive et converge partout (éventuellement vers l'infini). Par le théorème de convergence monotone, $\int h(x) dx = \sum_n \int |g_n(x)| dx = \sum_n \|g_n\|_1 \leq 1$. Donc $h(x) \in \mathbb{L}^1$ et on déduit que la série de réels $\sum_n |g_n(x)|$ est convergente pour presque tout x . Il est donc de même pour la série $\sum_n g_n(x)$. Cette dernière série a $h(x)$ pour chapeau intégrable. Par le théorème de Lebesgue, elle est donc convergente dans \mathbb{L}^1 . Comme $f_{i_{n+1}} = f_{i_1} + \sum_1^n g_n$, on en déduit que $(f_{i_n})_n$ converge dans \mathbb{L}^1 .
- iii) $|f_{i_1}| + h(x)$ est un chapeau intégrable pour la suite $(f_{i_n})_n$.
- iv) Adaptation à la démonstration de complétude de \mathbb{L}^1 : Si f_n est une suite de Cauchy dans \mathbb{L}^1 , on peut définir g_n comme précédemment et appliquer exactement le même raisonnement pour prouver que la série $\sum_n g_n(x)$ converge pour presque tout x et a donc une limite ponctuelle $f(x)$. On déduit alors du théorème de Lebesgue que $(f_{i_n})_n$ converge vers f dans \mathbb{L}^1 . Il est alors immédiat (inégalité triangulaire) que toute la suite $(f_n)_n$ converge vers f .

Exercice 3.8.15. Calculer la dérivée à droite en zéro de la fonction

$$t \rightarrow \int_0^1 (g(x) + t^2)^{\frac{1}{2}} dx = \varphi(t),$$

où g vérifie $0 \leq g \leq 1$. Indication : pour évaluer la limite du rapport $\frac{\varphi(t) - \varphi(0)}{t}$, penser à utiliser le théorème de Lebesgue.

Solution

On a

$$\frac{\varphi(t) - \varphi(0)}{t} = \int_0^1 \frac{t}{\sqrt{f(x)^2 + t^2} + \sqrt{f^2(x)}} dx = \int_0^t k_t(x) dx.$$

On voit que $k_t(x) = 1$ si $f(x) = 0$ et $k_t(x) \rightarrow 0$ quand $t \rightarrow 0$ si $f(x) \neq 0$. Donc $k_t(x) \rightarrow \mathbb{1}_{\{f(x)=0\}}$, la fonction caractéristique de l'ensemble des points où $f(x)$ s'annule. Ceci nous donne la limite ponctuelle. Cherchons un chapeau intégrable. On voit immédiatement que $|k_t(x)| \leq 1$. Donc, par le théorème de Lebesgue, $k_t \rightarrow \mathbb{1}_{\{f(x)=0\}}$ quand $t \rightarrow 0$ dans $\mathbb{L}^1(0, 1)$. D'où $\varphi'(0^+) = \int \mathbb{1}_{\{f(x)=0\}} = \lambda(\{x, f(x) = 0\})$.

Exercice 3.8.16. Calculer $\lim_{n \rightarrow +\infty} \Gamma_n$, où $\Gamma_n = \int_0^n (1 - \frac{x}{n})^n e^{\frac{x}{2}} dx$.

Solution

Notons $\mathbb{1}_{[0,n]}$ la fonction caractéristique de $[0, n]$, qui vaut 1 si $x \in [0, n]$ et 0 ailleurs. Alors

$$\Gamma_n = \int_0^{+\infty} (1 - \frac{x}{n})^n e^{\frac{x}{2}} \mathbb{1}_{[0,n]}(x) dx.$$

On calcule la limite ponctuelle de l'intégrand : Comme $(1 - \frac{x}{n})^n = e^{n \log(1 - \frac{x}{n})} \rightarrow e^{-x}$ quand $n \rightarrow +\infty$, on voit que l'intégrand tend vers $e^{-\frac{x}{2}}$. Reste à trouver un chapeau intégrable. Si $x \leq n$, $\log(1 - \frac{x}{n}) \leq -\frac{x}{2n}$. Donc $(1 - \frac{x}{n})^n e^{\frac{x}{2}} \mathbb{1}_{[0,n]}(x) \leq e^{-\frac{x}{2}}$ pour tout $x \geq 0$. On conclut par le théorème de Lebesgue que $\Gamma_n \rightarrow \int_0^{+\infty} e^{-\frac{x}{2}} dx = 2$.

Exercice 3.8.17. On pose, pour $x > 0$, $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$. Démontrer que Γ est de classe C^∞ sur $]0, +\infty[$. Démontrer que $\Gamma(x) \rightarrow +\infty$ quand $x \rightarrow 0$. Pour la deuxième question, penser à utiliser le lemme de convergence monotone !

Solution

On pose $f(t, x) = e^{-t} t^{x-1} = e^{-t+(x-1) \log t}$. Fixons $0 < \alpha < M$. On a, pour $(t, x) \in]0, +\infty[\times]\alpha, M]$, $|\frac{\partial^n f}{\partial x^n}(t, x)| = |e^{-t} (\log t)^n t^{x-1}| \leq h_n(t)$, où $h_n(t) = e^{-t} |(\log t)^n| t^{\alpha-1}$ si $0 < t \leq 1$ et $h_n(t) = e^{-t} |(\log t)^n| t^{M-1}$

si $t > 1$. Les fonctions $h_n(t)$ étant sommables sur $]0, \infty[$, on voit (pour $n = 0$) que Γ est bien définie, puis qu'on peut appliquer successivement à tous ordres le théorème de dérivation sous le signe somme à l'intégrale définissant $\Gamma(x)$. On obtient, pour tout $\alpha < x < M$,

$$\frac{\partial^n \Gamma}{\partial x^n}(t, x) = \int_0^\infty e^{-t} (\log t)^n t^{x-1} dt.$$

Comme α est arbitrairement petit et M arbitrairement grand, on conclut que $\Gamma(x)$ est C^∞ sur $]0, +\infty[$. Quand $x \rightarrow 0$, le théorème de convergence monotone nous dit que

$$\Gamma(x) \geq \int_0^1 e^{-t} t^{x-1} dt \rightarrow \int_0^1 \frac{e^{-t}}{t} dt = +\infty.$$

Remarque : on peut aussi montrer que $\Gamma(z)$ est holomorphe sur le demi-plan $Re(z) > 0$, c'est plus rapide.

Exercice 3.8.18. Soit f une fonction sommable. Démontrer que $\lim_{\alpha \rightarrow +\infty} \alpha \lambda(\{x, |f(x)| \geq \alpha\}) = 0$.

Solution

On a

$$\alpha \int_{\{|f(x)| \geq \alpha\}} dx \leq \int_{\{|f(x)| \geq \alpha\}} |f(x)| dx = \int \chi_{\{|f(x)| \geq \alpha\}}(x) |f(x)| dx.$$

On remarque que $\mathbb{1}_{\{|f(x)| \geq \alpha\}} f(x) \rightarrow 0$ quand $\alpha \rightarrow +\infty$, presque partout, et que $|\mathbb{1}_{\{|f(x)| \geq \alpha\}} f(x)| \leq |f(x)|$ qui est donc chapeau intégrable. Par le théorème de Lebesgue, on a donc

$$\int_{\{|f(x)| \geq \alpha\}} |f(x)| dx \rightarrow 0$$

Exercice 3.8.19. Si $f \in \mathbb{L}^1(\mathbb{R})$, montrer que $\tilde{f}(x) = \sum_{n \in \mathbb{Q}} f(x+n) \in \mathbb{L}^1([0, 1])$ et que $\int_{[0, 1]} \tilde{f}(x) dx = \int_{\mathbb{R}} f(x) dx$.

Solution

Par le théorème de la convergence monotone, on a

$$\int_0^1 \sum_{n \in \mathbb{Z}} |f(x+n)| = \sum_{n \in \mathbb{Z}} \int_0^1 |f(x+n)| dx = \sum_{n \in \mathbb{Z}} \int_n^{n+1} |f(x)| dx = \int_{\mathbb{R}} |f(x)| dx.$$

La série $\sum_{n \in \mathbb{Z}} |f(x+n)|$ converge donc pour presque tout x et il en est de même pour la série $\sum_{n \in \mathbb{Z}} f(x+n)$. Celle-ci a $\sum_{n \in \mathbb{Z}} |f(x+n)|$ comme chapeau intégrable. Par le théorème de Lebesgue, elle est donc convergente dans \mathbb{L}^1 et obtient la formule demandée.

Exercice 3.8.20. Un contre-exemple à méditer au théorème de dérivation sous le signe somme. Soit ϕ une fonction continue sur $[0, 1]$. On considère, dans $[0, 1] \times [0, 1]$ la fonction f définie par $f(x, \lambda) = \phi(x)$ si $x \leq \lambda$ et $f(x, \lambda) = 0$ sinon. On pose $F(\lambda) = \int_0^1 f(x, \lambda) dx$. Pour chaque λ , la dérivée partielle existe sauf en un point, et elle est majorée par une fonction sommable fixe : la fonction 0. Déterminer la dérivée de F .

Solution

On a $F(\lambda) = \int_0^1 f(x, \lambda) dx = \int_0^\lambda \phi(x) dx$. Donc $F'(\lambda) = \phi(\lambda)$. On voit que la conclusion du théorème de dérivation sous le signe somme est fautive. En effet, l'hypothèse b) est invalidée : $\frac{\partial f}{\partial \lambda}(x, \lambda)$ n'existe pas en tout point de $A \times I = [0, 1] \times [0, 1]$, mais seulement en presque tout point.

Exercice 3.8.21. (Inégalité de Hardy, [11])

Soit p un réel strictement supérieur à 1.

1) Soit $f \in \mathbb{L}^p(\mathbb{R}_+)$. On pose $F(x) = \frac{1}{x} \int_0^x f(t)dt$. Le but de la question est de montrer que F est dans $\mathbb{L}^p(\mathbb{R}_+)$ et que :

$$\left| \int_0^\infty |F(x)|^p dx \right| \leq \left(\frac{p}{p-1} \right)^p \int_0^\infty |f^p(x)| dx$$

- a) Etablir le résultat lorsque f est continue positive.
- b) En déduire le résultat lorsque f est continue et de signe quelconque.
- c) Traiter le cas général.

2) Montrer que cette constante est optimale (*ind. : on pourra utiliser des fonctions puissances*).

Solution

1)

a) f étant continue, F est continue (y compris en 0, le vérifier) et dérivable sur $(0, +\infty)$ avec $(xF)' = F + xF' = f$. En écrivant que

$$F^p(x) = F^{p-1}(x)f(x) - xF^{p-1}(x)F'(x)$$

et en effectuant une intégration par parties, on obtient

$$\int_0^T F^p(x)dx = -\frac{T}{p-1}F^p(T) + \left(\frac{p}{p-1} \right) \int_0^T F^{p-1}(x)f(x)dx$$

(on vérifiera qu'il n'y a pas de problème en 0 en prenant l'intégrale entre ε et T et en faisant tendre ε vers 0). Or $-\frac{T}{p-1}F^p(T) \leq 0$ car f est supposée positive. En appliquant l'inégalité de Hölder dans la deuxième intégrale, on a

$$\int_0^T F^p(x)dx \leq \left(\frac{p}{p-1} \right)^p \int_0^T f^p(x)dx$$

b) On généralise l'inégalité précédente au cas où f est de signe quelconque en remarquant qu'alors

$$|F(x)| \leq G(x) = x^{-1} \int_0^x |f(t)|dt$$

et l'inégalité précédente est valable avec le couple $(|f|, G)$. On en déduit facilement l'inégalité pour (f, F) .

c) On vient de montrer que l'application $T : f \in C_0 \rightarrow F \in \mathcal{L}^p$ était linéaire continue avec une norme inférieure à $p/(p-1)$. Puisque C^0 est dense dans \mathcal{L}^p , on a que pour tout $f \in \mathcal{L}^p$, il existe une suite $(f_n)_n$ de fonctions continues tendant en norme \mathcal{L}^p vers f . Soit $F_n = T(f_n) : (F_n)_n$ est une suite de Cauchy dans \mathcal{L}^p donc elle converge vers une certaine fonction G . Par la "réciproque du théorème de Lebesgue", de la suite f_n on peut extraire une sous-suite qui converge simplement (et dans \mathcal{L}^p) vers f . On en déduit facilement que $F(x) = x^{-1} \int_0^x f(t)dt = G(x)$ p.p. Donc F est dans \mathcal{L}^p et vérifie l'inégalité de Hardy.

2) Facile.

Exercice 3.8.22. ([11]) Soit $a > 0$. Calculer

$$\lim_{n \rightarrow \infty} \int_0^{n^{1/a}} \left(1 - \frac{x^a}{n}\right)^n dx$$

(ind. : on pourra utiliser le théorème de convergence monotone)

Solution

Soit $u > 0$. On considère la fonction $g_u(t) = t \log(1 - u/t)$ pour $t > u$. on a $g'_u(t) = \log(1 - u/t) + u/(t - u)$ et $g''_u(t) = -u^2/[t(t - u)^2] \leq 0$ et donc g'_u positive car $g'_u(+\infty) = 0$. Donc $g_u(t)$ est une fonction croissante de $t > u$ et $t \in (u, +\infty) \rightarrow (1 - u/t)^t$ aussi. Il ne reste alors plus qu'à appliquer le théorème de convergence monotone et la limite cherchée est $\int_0^\infty \exp(-x^a) dx = a^{-1} \Gamma(a^{-1})$.

Exercice 3.8.23. ([11]) 1) Soit γ la constante d'Euler. Montrer que

$$\gamma = - \int_1^\infty \left(\frac{1}{x} - \frac{1}{E(x)} \right) dx$$

où $E(x)$ désigne la partie entière de x . En déduire que $\gamma > 0$.

2) Montrer que $\gamma = \int_0^\infty e^{-s} \log(s) ds$.

Solution

1) Trivial.

2) On a

$$\int_0^k (1 - s/k)^k \log(s) ds = \int_0^1 (1 - s/k)^k \log(s) ds + \int_1^k (1 - s/k)^k \log(s) ds$$

Quand k tend vers $+\infty$, la première intégrale tend grâce au théorème de convergence dominée vers $\int_0^1 e^{-s} \log(s) ds$ et la deuxième, grâce au théorème de convergence monotone (voir exercice précédent), vers $\int_1^\infty e^{-s} \log(s) ds$.

On est donc ramené à l'étude de $\int_0^k (1 - s/k)^k \log(s) ds$ qui après un changement de variables donne

$$k \int_0^1 (1 - u)^k \log(u) du - \frac{k}{k+1} \log k = \int_0^1 kv^k \log(1 - v) dv - \frac{k}{k+1} \log k$$

Or $\log(1 - v) = \sum_{j=1}^\infty \frac{v^j}{j}$ d'où par le théorème de convergence monotone

$$\int_0^1 v^k \log(1 - v) dv = \frac{k}{k+1} \left(1 + \frac{1}{2} + \dots + \frac{1}{k+1} \right)$$

On en déduit le résultat.

Exercice 3.8.24. ([13])

Pour $z > 0$, $p, q > 0$, on définit les fonctions

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt, \quad B(p, q) = \int_0^1 t^{p-1} (1 - t)^{q-1} dt$$

Montrer que $B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$

Exercice 3.8.25. Rappels : un ensemble négligeable de \mathbb{R}^N est un ensemble contenu dans un borélien de \mathbb{R}^N de mesure nulle. On ajoute à la tribu des boréliens les ensembles négligeables. On obtient ainsi la tribu de Lebesgue, notée \mathcal{L} .

- 1) Un ensemble est Lebesgue mesurable s'il diffère d'un borélien par un ensemble de mesure nulle.
- 2) On dira qu'une fonction est Lebesgue-mesurable si elle est mesurable de $(\mathbb{R}^N, \mathcal{L})$ dans \mathbb{R}, \mathbb{B} où \mathbb{B} est la tribu de Borel. Une fonction est Lebesgue-mesurable si et seulement si elle est presque partout égale à une fonction borélienne.

Solution 1) Soit \mathcal{L} la tribu de Lebesgue, i.e. la plus petite tribu contenant les boréliens et les ensembles négligeables. Soit \mathcal{E} l'ensemble des ensembles E qui diffèrent d'un borélien par un ensemble négligeable, i.e. $(E = B \cup N_1) \setminus N_2$ où B est un borélien et N_1, N_2 sont des ensembles négligeables. On a $\mathcal{E} \subset \mathcal{L}$ et il suffit donc de vérifier que \mathcal{E} est une tribu, ce qui va tout seul.

2) Si f ne diffère de g borélienne que sur un ensemble négligeable N , les ensembles $[f \leq a]$ et $[g \leq a]$ ne diffèrent que d'un ensemble négligeable et donc $[f \leq a]$ est bien Lebesgue-mesurable. Donc f est Lebesgue-mesurable. Réciproquement, considérons la classe \mathcal{H} des fonctions $\sigma(\mathcal{L})$ -mesurables et la classe \mathcal{J} des fonctions qui ne diffèrent d'une fonction borélienne que sur un ensemble de mesure nulle, et enfin l'ensemble \mathcal{X} des fonctions caractéristiques d'ensembles Lebesgue-mesurables. On a par ce qui précède $\mathcal{X} \subset \mathcal{J} \subset \mathcal{H}$. On voit aisément que \mathcal{J} est un espace vectoriel, qu'il contient 1 et est stable par limite croissante. En appliquant le théorème de la classe monotone, on déduit que \mathcal{J} contient toutes les fonctions $\sigma(\mathcal{L})$ -mesurables et donc \mathcal{H} . Donc $\mathcal{J} = \mathcal{H}$.

2) (solution constructive). Soit f une fonction mesurable au sens de Lebesgue. Soit

$$f_n = \sum_{k \in \mathbb{Z}} (k\varepsilon) \mathbb{1}_{\frac{k}{n} \leq f < \frac{k+1}{n}} = \mathbb{1}_{A_{k,n}}$$

Comme $A_{k,n}$ est Lebesgue-mesurable, il existe $B_{k,n}$ borélien tel que $\mu(A_{k,n} \Delta B_{k,n}) = 0$. Donc, presque partout, $f_n = \tilde{f}_n = \sum_{k \in \mathbb{Z}} (\frac{k}{n}) \mathbb{1}_{B_{k,n}}$. Mais $f_n \rightarrow f$ presque partout. Donc $\tilde{f}_n \rightarrow f$ presque partout. Donc f est égale presque partout à la limite, borélienne, d'une suite de fonctions boréliennes.

3.9 Rappel des principaux théorèmes, énoncés dans \mathbb{R}^N .

Théorème 3.9.1. (Beppo Levi, ou convergence monotone)

Si f_n est une suite croissante de fonctions définies sur \mathbb{R}^N et à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$, on a

$$\int \lim_{n \rightarrow \infty} f_n(x) dx = \lim_{n \rightarrow \infty} \int f_n(x) dx \leq +\infty.$$

Corollaire 3.9.1. Soit $u_n(x)$ une suite de fonctions définies sur \mathbb{R}^N et à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$. On peut alors l'intégrer terme à terme, c'est-à-dire que

$$\int \sum_{n=1}^{\infty} u_n(x) dx = \sum_{n=1}^{\infty} \int u_n(x) dx \leq +\infty.$$

Théorème 3.9.2. (Théorème de Lebesgue)

Soit $f_n(x)$ une suite de fonctions qui converge presque partout vers une fonction f . On suppose qu'il existe une fonction positive sommable fixe h telle que pour tout n , $|f_n(x)| \leq h(x)$ pour presque tout x . Alors

$$\int |f_n(x) - f(x)| \rightarrow 0, \quad \int f_n(x) dx \rightarrow \int f(x) dx.$$

Théorème 3.9.3. (Réciproque du théorème de Lebesgue).

Si $f_n \rightarrow f$ dans \mathbb{L}^1 , alors il existe une sous-suite f_{n_k} de la suite f_n qui converge presque partout vers f et un chapeau intégrable h tel que $\forall k, |f_{n_k}(x)| \leq h(x)$ p.p..

Théorème 3.9.4. (Théorème de dérivation sous le signe somme) Soient I un intervalle de \mathbb{R} et A un sous-ensemble de \mathbb{R}^n . On se donne une fonction f définie sur $A \times I$ vérifiant les trois hypothèses suivantes.

(a) Pour tout $\lambda \in I$, la fonction $x \rightarrow f(x, \lambda)$ est sommable sur A .

(b) La dérivée partielle $\frac{\partial f}{\partial \lambda}(x, \lambda)$ existe en tout point de $A \times I$.

(c) Il existe une fonction h positive et sommable sur A telle que l'on ait $|\frac{\partial f}{\partial \lambda}(x, \lambda)| \leq h(x)$ pour tous x et λ . Alors la fonction F définie par

$$F(\lambda) = \int_A f(x, \lambda) dx$$

est dérivable dans I et on a

$$F'(\lambda) = \int_A \frac{\partial f}{\partial \lambda}(x, \lambda) dx.$$

Théorème 3.9.5. (de changement de variable, [10])

Soient Ω_1 et Ω_2 deux ouverts de \mathbb{R}^n et φ un difféomorphisme entre Ω_1 et Ω_2 . On notera $J_\varphi(x)$ le déterminant de la matrice jacobienne de φ au point x . Soit f une fonction définie sur Ω_2 .

(a) Si la fonction f est à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$, on a l'égalité suivante, où les deux membres ont un sens dans $\mathbb{R}^+ \cup \{+\infty\}$,

$$\int_{\Omega_2} f(y) dy = \int_{\Omega_1} f(\varphi(x)) |J_\varphi(x)| dx.$$

(b) Si f est à valeurs complexes, elle est sommable dans Ω_2 si et seulement si $f(\varphi(x)) J_\varphi(x)$ est sommable dans Ω_1 , et les deux membres de l'égalité précédente sont alors égaux.

Théorème 3.9.6. (Fubini)

Soit $f(x, y)$ une fonction définie dans $\mathbb{R}^p \times \mathbb{R}^q$.

(a) Si f est à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$, on a l'égalité suivante, où les trois membres définissent un élément de $\mathbb{R}^+ \cup \{+\infty\}$,

$$\int_{\mathbb{R}^{p+q}} f(x, y) dx dy = \int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^q} f(x, y) dy \right) = \int_{\mathbb{R}^q} \left(\int_{\mathbb{R}^p} f(x, y) dx \right) dy.$$

Si f est sommable dans \mathbb{R}^{p+q} , les trois membres de l'égalité précédente ont un sens et sont égaux. Plus précisément, dire que le troisième membre a un sens signifie :

- pour presque tout y , la fonction $x \rightarrow f(x, y)$ est sommable dans \mathbb{R}^p ,
- la fonction $\varphi(y) = \int f(x, y) dx$ qui est ainsi définie presque partout est sommable dans \mathbb{R}^q .

Chapitre 4

Produit de mesures

Nous travaillerons avec deux espaces de probabilité mais tout ce que nous dirons s'étend sans changement à un produit fini d'espaces de probabilité. Soient $(\Omega_1, \mathcal{F}_1)$ et $(\Omega_2, \mathcal{F}_2)$ deux espaces munis de leurs tribus respectives. On cherche à définir une notion de tribu produit sur $\Omega_1 \times \Omega_2 = \Omega$ afin de définir la notion de mesure produit sur ce même espace. Soit

$$\rho_1 : \omega = (\omega_1, \omega_2) \in \Omega \rightarrow \omega_1 \in \Omega_1 \text{ et } \rho_2 : \omega = (\omega_1, \omega_2) \in \Omega \rightarrow \omega_2 \in \Omega_2$$

les deux projections canoniques.

Définition 4.0.1. La tribu produit \mathcal{F} sur Ω , notée $\mathcal{F}_1 \otimes \mathcal{F}_2$, est définie comme la tribu engendrée par ρ_1 et ρ_2 ,

$$\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\rho_1, \rho_2) = \sigma(\{B_1 \times \Omega_2, \Omega_1 \times B_2 ; B_1 \in \mathcal{F}_1, B_2 \in \mathcal{F}_2\})$$

De manière plus condensée,

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\rho_1^{-1}(\mathcal{F}_1), \rho_2^{-1}(\mathcal{F}_2)).$$

On remarquera que $\mathcal{I} = \{B_1 \times B_2 ; B_1 \in \mathcal{F}_1, B_2 \in \mathcal{F}_2\}$ est un π -système qui génère (en tant que tribu) $\mathcal{F}_1 \otimes \mathcal{F}_2$. En effet, $(B_1 \times B_2) \cap (C_1 \times C_2) = (B_1 \cap C_1) \times (B_2 \cap C_2)$.

Notation : Si $f : \Omega \rightarrow \mathbb{R}$, on note

$$f_{\omega_1}^1 : \omega_2 \in \Omega_2 \rightarrow f(\omega_1, \omega_2) \in \mathbb{R}, \quad (\omega_1 \in \Omega_1)$$

et

$$f_{\omega_2}^2 : \omega_1 \in \Omega_1 \rightarrow f(\omega_1, \omega_2) \in \mathbb{R}, \quad (\omega_2 \in \Omega_2)$$

Lemme 4.0.1. Soit \mathcal{H} l'ensemble des fonctions $f : \Omega \rightarrow \mathbb{R}$ qui sont \mathcal{F} -mesurables bornées et telles que $f_{\omega_1}^1$ est \mathcal{F}_2 -mesurable pour tout $\omega_1 \in \Omega_1$ et $f_{\omega_2}^2$ est \mathcal{F}_1 -mesurable pour tout $\omega_2 \in \Omega_2$. Alors \mathcal{H} est exactement l'ensemble des fonctions \mathcal{F} -mesurables bornées.

Démonstration. On applique le théorème de la classe monotone. \mathcal{H} est un espace vectoriel constitué de fonctions \mathcal{F} -mesurables bornées et contenant les $\mathbb{1}_A$ pour $A \in \mathcal{I}$, \mathcal{I} étant un π -système générant \mathcal{F} . D'autre part, si $(f_n)_n$ est une suite croissante de fonctions de \mathcal{H} positives uniformément bornées par une constante K , alors la limite simple $f = \lim_{n \rightarrow +\infty} f_n$ est une fonction \mathcal{F} -mesurable bornée. De plus, pour tout $\omega_1, \omega_2 \in \Omega_1$, on a

$$f_{\omega_1}^1(\omega_2) = \lim_{n \rightarrow +\infty} (f_n)_{\omega_1}^1(\omega_2)$$

ce qui montre que $f_{\omega_1}^1$ est \mathcal{F}_2 -mesurable. De même, $f_{\omega_2}^2$ est \mathcal{F}_1 -mesurable. Le théorème de la classe monotone implique que \mathcal{H} contient toutes les fonctions \mathcal{F} -mesurables bornées. \square En appliquant le lemme précédent aux fonctions caractéristiques d'ensembles, on obtient :

Corollaire 4.0.2. Soit $A \subset \Omega_1 \times \Omega_2$ un ensemble \mathcal{F} -mesurable. Alors :

- $\forall \omega_1 \in \Omega_1, \rho_2(A \cap \{\omega_1\} \times \Omega_2)$ est \mathcal{F}_2 -mesurable et
- $\forall \omega_2 \in \Omega_2, \rho_1(A \cap \Omega_1 \times \{\omega_2\})$ est \mathcal{F}_1 -mesurable.

En termes plus imagés, si A est \mathcal{F} -mesurable alors ses coupes verticales et horizontales sont mesurables. Mais attention : la réciproque est fautive ; un ensemble dont les coupes verticales et horizontales sont mesurables n'est pas nécessairement mesurable.

Corollaire 4.0.3. Si $(\Omega_1, \mathcal{F}_1), \dots, (\Omega_n, \mathcal{F}_n)$ sont des espaces munis de tribus, alors pour tout k ,

$$(\otimes_{i=1}^k \mathcal{F}_i) \otimes (\otimes_{i=k+1}^n \mathcal{F}_i) = \otimes_{i=1}^n \mathcal{F}_i.$$

En d'autres termes le produit de tribus est associatif. En particulier :

$$\mathcal{B}(\mathbb{R}^{n+m}) = \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^m).$$

Démonstration. En effet les tribus considérées sont engendrées respectivement par les π -systèmes $(\mathcal{I})^k$, $(\mathcal{I})^{n-k}$ et $(\mathcal{I})^n$. Le dernier est le produit cartésien des deux premiers car le produit cartésien est associatif. \square

4.1 Mesure produit, Théorème de Fubini

4.1.1 Théorème de Fubini

On suppose que $(\Omega_i, \mathcal{F}_i, \mu_i)$, pour $i = 1, 2$, sont des espaces mesurés munis de mesures finies. D'après le lemme précédent, on peut définir pour toute fonction \mathcal{F} -mesurable bornée f :

$$I_1^f(\omega_1) = \int_{\Omega_2} f_{\omega_1}^1(\omega_2) d\mu_2(\omega_2), \quad (\omega_1 \in \Omega_1)$$

et

$$I_2^f(\omega_2) = \int_{\Omega_1} f_{\omega_2}^2(\omega_1) d\mu_1(\omega_1), \quad (\omega_2 \in \Omega_2)$$

Lemme 4.1.1. Soit \mathcal{H} l'espace vectoriel composé des fonctions $f : \Omega \rightarrow \mathbb{R}$ \mathcal{F} -mesurables bornées telles que :

a) I_1^f est \mathcal{F}_1 -mesurable bornée et I_2^f est \mathcal{F}_2 -mesurable bornée.

b) $\int_{\Omega_1} I_1^f(\omega_1) d\mu_1(\omega_1) = \int_{\Omega_2} I_2^f(\omega_2) d\mu_2(\omega_2)$

Alors \mathcal{H} est exactement l'ensemble des fonctions \mathcal{F} -mesurables bornées.

Démonstration. Là aussi, il s'agit d'une simple application du théorème de la classe monotone. En effet \mathcal{H} contient encore les fonctions caractéristiques $\mathbb{1}_{A_1 \times A_2}$ avec A_1 et A_2 \mathcal{F}_1 - et \mathcal{F}_2 -mesurables respectivement. Or l'ensemble $\{A_1 \times A_2\}$ est un π -système engendrant \mathcal{F} . \square

Si $F \in \mathcal{F}$, en posant $f = \mathbb{1}_F$ qui est \mathcal{F} -mesurable bornée, on définit :

$$\mu(F) = \int_{\Omega_1} I_1^f(\omega_1) d\mu_1(\omega_1) = \int_{\Omega_2} I_2^f(\omega_2) d\mu_2(\omega_2)$$

On va voir dans la suite que μ est une mesure sur la tribu produit.

Nous allons construire l'intégrale par machine standard. Soit f une fonction positive \mathcal{F} -mesurable. Pour tout entier n , $f \wedge n$ est \mathcal{F} -mesurable positive bornée. Si $\omega_1 \in \Omega_1$, $\{(f \wedge n)_{\omega_1}^1\}_n$ est une suite de fonctions \mathcal{F}_2 -mesurables tendant en croissant vers $f_{\omega_1}^1$ ce qui montre que $f_{\omega_1}^1$ est \mathcal{F}_2 -mesurable. De même, pour tout $\omega_2 \in \Omega_2$, $f_{\omega_2}^2$ est \mathcal{F}_1 -mesurable. On peut donc définir I_1^f et I_2^f (en admettant la valeur $+\infty$).

$I_1^{(f \wedge n)}$ et $I_2^{(f \wedge n)}$ sont bien définies, respectivement \mathcal{F}_1 -mesurable et \mathcal{F}_2 -mesurable. Et si $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2$, on a

$$I_1^{(f \wedge n)}(\omega_1) \uparrow I_1^f(\omega_1) \quad \text{et} \quad I_2^{(f \wedge n)}(\omega_2) \uparrow I_2^f(\omega_2)$$

si bien que I_1^f et I_2^f sont encore des fonctions \mathcal{F}_1 -mesurables et \mathcal{F}_2 -mesurables. Qui plus est, toujours par le théorème de convergence monotone,

$$\int_{\Omega_1} I_1^f d\mu_1 = \int_{\Omega_2} I_2^f d\mu_2.$$

Donc le résultat du lemme 4.1.1 s'étant aux fonctions positives.

Théorème 4.1.1. (Théorème de Fubini)

1) La fonction μ définie sur (Ω, \mathcal{F}) est une mesure (finie) appelée mesure produit de μ_1 et μ_2 et notée $\mu_1 \otimes \mu_2$. De plus, μ est l'unique mesure sur (Ω, \mathcal{F}) telle que :

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2), \quad A_1 \in \mathcal{F}_1, \quad A_2 \in \mathcal{F}_2$$

2) Si f est une fonction \mathcal{F} -mesurable positive alors :

$$\int f d\mu = \int_{\Omega_1} I_f^1(\omega_1) d\mu_1(\omega_1) = \int_{\Omega_2} I_f^2(\omega_2) d\mu_2(\omega_2) \quad (\star)$$

3) Si f est \mathcal{F} -mesurable et si $\mu(|f|) < +\infty$ alors (\star) est encore valable.

Démonstration. 1) Le fait que μ soit une mesure provient de la linéarité de l'intégrale et du théorème de convergence monotone. En fait, la question délicate dans ce théorème concerne les problèmes de mesurabilité, travail préliminaire qui a déjà été fait via le théorème de la classe monotone. L'unicité est une simple conséquence du théorème de Dynkin.

2) On a déjà prouvé la seconde égalité dans (\star) . Quant à la première égalité, elle provient là encore du théorème de la classe monotone (on sait en effet qu'elle est vraie pour les indicatrices d'éléments de \mathcal{F}). On a donc le résultat pour les fonctions \mathcal{F} -mesurables bornées. Si f est une fonction \mathcal{F} -mesurable positive, en considérant la suite de fonctions bornées $(f \wedge n)_n$ et en appliquant le théorème de convergence monotone, on obtient 2).

3) Il suffit comme d'habitude de décomposer la fonction f en la différence de sa partie positive f^+ et de sa partie négative f^- .
□

Exercice 4.1.1. Un contreexemple à Fubini si f n'est pas sommable : ([14], p 472) :

Soit $f(x, y) = e^{-xy} - 2e^{-2xy}$. Vérifier que :

$$\int_0^1 \int_1^\infty f(x, y) dy dx > 0$$

$$\int_1^\infty \int_0^1 f(x, y) dx dy < 0$$

Remarque 4.1.1. 1) Tout ce qui précède s'étend sans peine si au lieu de considérer des mesures finies, on prend des mesures σ -finies mais pas au-delà (contre-exemple dans [30]). La conclusion 2) du théorème de Fubini s'obtient par simple convergence monotone et la conclusion 3) par le théorème de Lebesgue.

2) D'autre part, au lieu de considérer deux espaces, on peut très bien en considérer un nombre fini arbitraire $n : (\Omega_i, \mathcal{F}_i, \mu_i)_{i=1 \dots n}$. La tribu produit $\otimes_{i=1}^n \mathcal{F}_i$ est définie comme la tribu engendrée par les (ρ_1, \dots, ρ_n) , les projections canoniques de Ω sur $\Omega_1, \dots, \Omega_n$. Cette tribu produit est générée par le π -système $\{A_1 \times \dots \times A_n ; A_i \in \mathcal{F}_i\}$.

3) On remarquera que pour aboutir au théorème de Fubini, on a utilisé de manière fréquente le théorème de la classe monotone. Il est en fait très difficile de définir les mesures produits sans y avoir recours d'une manière ou d'une autre. En particulier, la "machine standard" ne fonctionne pas alors que la plupart du temps où l'on utilise le théorème de la classe monotone, un argument "machine standard" peut être utilisé.

Exemple : Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}), dx)$ une variable aléatoire. Soit $\Omega \times \mathbb{R}$ muni de $d\mathbb{P} \otimes dx$ et $A = \{(\omega, x) / 0 \leq x < X(\omega)\}$. A est $\mathcal{F} \otimes \mathcal{B}(\mathbb{R})$ -mesurable et l'on a par Fubini :

$$\mu(A) = \int_{[0;+\infty)} \mathbb{P}(X > x) dx = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \mathbb{E}(X)$$

et donc

$$\mathbb{E}(X) = \int_{\mathbb{R}_+} (1 - F(x)) dx$$

Un autre conséquence importante du théorème de la classe monotone est la suivante :

Proposition 4.1.1. f est $\otimes_{i=1}^n \mathcal{F}_i$ -mesurable ssi $f = G(\rho_1, \dots, \rho_n)$ où G est une fonction borélienne.

4.2 Tribus et mesures produits pour une infinité d'espaces

Cette section s'inspire de [14].

Théorème 4.2.1. (de Kolmogorov)

On suppose donnée une suite de mesures de probabilités $(\mu_n)_n$ sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ consistantes, c'est-à-dire :

$$\mu_{n+1}((a_1, b_1] \times \dots \times (a_n, b_n] \times \mathbb{R}) = \mu_n((a_1, b_1] \times \dots \times (a_n, b_n])$$

Alors il existe une unique probabilité \mathbb{P} sur $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ telle que :

$$(*) \mathbb{P}(\omega ; \omega_i \in (a_i, b_i], 1 \leq i \leq n) = \mu_n((a_1, b_1] \times \dots \times (a_n, b_n])$$

Démonstration. Précisons d'abord ce que nous entendons par $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$. Cette tribu est la tribu borélienne sur $\mathbb{R}^{\mathbb{N}}$ muni de la topologie de la convergence simple. C'est aussi, comme dans le cas fini-dimensionnel, la tribu produit $\otimes_{n=0}^{\infty} \mathcal{B}(\mathbb{R})$ engendrée par les projections canoniques. Ces deux tribus sont en effet engendrées par les rectangles fini-dimensionnels $\{\omega \in \mathbb{R}^{\mathbb{N}} ; \omega_i \in (a_i, b_i], i = 1, \dots, n\}$ où $-\infty \leq a_i < b_i \leq +\infty$ (rectangles non nécessairement bornés).

Mais il faut faire attention, si C est un ensemble non dénombrable, $\mathcal{B}(\mathbb{R}^C)$ (\mathbb{R}^C est muni de la topologie la moins fine rendant continues les projections canoniques) est plus grand que $\otimes_{\gamma \in C} \mathcal{B}(\mathbb{R})$ (que l'on peut encore définir que C soit dénombrable ou pas). Bref, dans le cas dénombrable, tout se passe bien mais dans le cas non dénombrable, quelque petites pathologies peuvent survenir.

L'ensemble des réunions finies de rectangles fini-dimensionnels est une algèbre qui génère $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ notée \mathcal{A} . On définit alors \mathbb{P} sur \mathcal{A} par (*). La condition de consistance garantit la cohérence de la définition de $\mathbb{P}(A)$ (i.e. ne dépend pas de l'écriture de A).

Par définition même de \mathbb{P} , \mathbb{P} est additive sur \mathcal{A} . Il reste à montrer, afin d'utiliser le théorème de Carathéodory, qu'elle est σ -additive sur \mathcal{A} .

Lemme 4.2.1. Si $B_n \in \mathcal{A}$, $B_n \downarrow \emptyset$ alors $\mathbb{P}(B_n) \downarrow 0$.

Démonstration. Supposons que $\mathbb{P}(B_n) \downarrow \delta > 0$. Pour chaque n , on écrit B_n sous la forme d'une réunion disjointe de rectangles fini-dimensionnels

$$B_n = \cup_{k=1}^{K_n} \{\omega ; \omega_i \in (a_i^k, b_i^k], 1 \leq i \leq n\}$$

En fait, on devrait écrire

$$B_n = \cup_{k=1}^{K_n} \{\omega ; \omega_i \in (a_i^k, b_i^k], 1 \leq i \leq \phi(n)\}$$

où $(\phi(n))_n$ est une suite strictement croissante. Afin d'alléger les écritures, on a posé $\phi(n) = n$ (mais la preuve reste exactement la même sinon).

L'idée de la preuve consiste à approximer B_n par des compacts D_n qui ont presque la même probabilité que B_n (donc non vides car de probabilité strictement positive) et d'utiliser un argument d'extraction diagonal pour montrer que $\cap_n D_n = \emptyset$ (ce qui est exclu pour une intersection de compacts non vides).

Pour chaque n , il existe des suites de réels $(\bar{a}_i^k)_{i=1\dots n, k=1\dots K_n}, (\bar{b}_i^k)_{i=1\dots n, k=1\dots K_n}$ telles que

$$\mu_n \left(\prod_{i=1}^n (a_i^k, b_i^k] - \prod_{i=1}^n [\bar{a}_i^k, \bar{b}_i^k] \right) \leq \frac{\delta}{2^{n+1} K_n}$$

donc en prenant $C_n = \cup_{k=1}^{K_n} \{\omega ; \omega_i \in [\bar{a}_i^k, \bar{b}_i^k], 1 \leq i \leq n\}$, on a

$$\mathbb{P}(B_n \setminus C_n) \leq \frac{\delta}{2^{n+1}}$$

On pose $D_n = \cap_{m=1}^n C_m$. On a

$$\mathbb{P}(B_n \setminus D_n) \leq \sum_{m=1}^n \mathbb{P}(B_m \setminus C_m) \leq \delta/2$$

Par conséquent, $\mathbb{P}(D_n) \downarrow \varepsilon \geq \delta/2$. En particulier, $(D_n)_n$ est non vide pour tout n .

On introduit les sous-ensembles de \mathbb{R}^n , C_n^*, D_n^* tels que

$$C_n = C_n^* \times \mathbb{R} \times \dots \times \mathbb{R} \times \dots$$

$$D_n = D_n^* \times \mathbb{R} \times \dots \times \mathbb{R} \times \dots$$

C_n^* est un compact non vide pour tout n et on a :

$$D_n^* = C_n^* \cap \cap_{m=1}^{n-1} (C_m^* \times \mathbb{R}^{n-m})$$

qui est donc un compact non vide de \mathbb{R}^n (non vide car D_n est non vide).

On utilise alors un argument d'extraction diagonale. Pour tout $m \in \mathbb{N}^*$, soit $\omega^m \in D_m \subset \mathbb{R}^{\mathbb{N}}$. $\omega^m \in D_m \subset D_1$ donc $\omega_1^m \in D_1^*$ qui est compact. On peut extraire une sous-suite $(\omega^{\phi_1(m)})_m$ de $(\omega^m)_m$ telle que $(\omega_1^{\phi_1(m)})_m$ converge vers un certain réel $\theta^1 \in D_1^*$. Or $\omega^{\phi_1(m)} \in D_{\phi_1(m)} \subset D_2$ donc on peut extraire une sous-suite $(\omega^{\phi_1 \circ \phi_2(m)})_m$ de $(\omega^{\phi_1(m)})_m$ telle que $((\omega_1^{\phi_1 \circ \phi_2(m)}, \omega_2^{\phi_1 \circ \phi_2(m)})_m)$ converge vers un certain $\theta^2 \in D_2^* \subset \mathbb{R}^2$ et bien entendu $\theta_1^2 = \theta^1 \in D_1^* \dots$ Bref, on construit $\theta \in \mathbb{R}^{\mathbb{N}^*}$ tel que pour tout $i \in \mathbb{N}^*, \theta_i \in D_i^*$ donc tel que $\theta \in D_k$ pour tout k et donc

$$\emptyset \neq \cap_k D_k \subset \cap_k B_k = \emptyset$$

On aboutit à une contradiction et le lemme est ainsi prouvé. □

Ceci conclut la preuve du théorème de Kolmogorov (l'unicité découle comme d'habitude du théorème de Dynkin). □

Remarque 4.2.1. 1. Ce théorème est encore valable quand \mathbb{R} est remplacé par un espace de Lusin E (i.e. E homéomorphe à un Borélien d'un espace métrisable compact). On pourra trouver des compléments dans *Diffusions, Markov processes and Martingales, Rogers and Williams, Vol.1*.

2. Le théorème de Kolmogorov montre que si l'on se donne une loi μ sur \mathbb{R} , on peut trouver un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et des variables aléatoires sur (Ω, \mathcal{F}) de loi μ et indépendantes entre elles (prendre $\mu_n = \otimes^n \mu$). En TD (modèle de pile ou face), on verra une construction à la main dans ce cas.

4.3 TD Probabilités : Fubini

Exercice 4.3.1. Calcul de l'intégrale de Gauss

Démontrer que $\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}$. On prendra le carré de cette intégrale, le considérera par Fubini comme une intégrale sur \mathbb{R}^2 que l'on calculera en coordonnées polaires. Penser à justifier rigoureusement le changement de coordonnées effectué dans l'intégrale à l'aide du théorème de changement de variables.

Exercice 4.3.2. Propriétés de la convolution

1) Si f et g appartiennent à $\mathbb{L}^1(\mathbb{R}^n)$, alors $f * g$ définie par

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y)g(x-y)dy$$

aussi (vérifier que la définition a un sens!), et $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$. (Application directe du théorème de Fubini.) Montrer que $f * g = g * f$ et que $(f * g) * h = f * (g * h)$.

2) On appelle $\mathbb{L}_{loc}^1(\mathbb{R}^n)$ l'ensemble des fonctions \mathbb{L}^1 sur tout borné de \mathbb{R}^n . Montrer que si $f \in \mathbb{L}_{loc}^1(\mathbb{R}^n)$, et si $g \in \mathbb{L}^1(\mathbb{R}^n)$ est nulle en dehors d'un borné, alors on peut définir $f * g$ et $f * g \in \mathbb{L}_{loc}^1$.

3) On suppose que $1 \leq p \leq \infty$, $f \in \mathcal{L}^1(\mathbb{R}^n)$ et $g \in \mathcal{L}^p(\mathbb{R}^n)$. Montrer que $f * g$ définie par

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)dy$$

existe pour presque tout x , appartient à $\mathcal{L}^p(\mathbb{R}^n)$ et vérifie $\|f * g\|_p \leq \|f\|_1 \|g\|_p$.

4) On suppose maintenant que $f \in \mathcal{L}^p(\mathbb{R}^n)$, $g \in \mathcal{L}^q(\mathbb{R}^n)$ avec p et q des exposants conjugués. Montrer que $f * g$ est uniformément continue et que si $1 < p < \infty$ alors $h = f * g$ est continue, nulle à l'infinie. Donner un contre-exemple dans le cas $p = 1$.

5) Si $f \in \mathbb{L}_{loc}^1(\mathbb{R}^n)$, $\varphi \in C_0^m(\mathbb{R}^n)$, alors $f * \varphi \in C^m(\mathbb{R}^n)$ et,

$$\forall |\alpha| \leq m, \quad \partial^\alpha (f * \varphi) = f * \partial^\alpha \varphi.$$

Solution

Applications directes du théorème de Fubini et du théorème de dérivation sous le signe somme. Le seul point un peu délicat est le 3). On mène le calcul en supposant f et g positives puisqu'on peut toujours remplacer f par $|f|$ et g par $|g|$. On a

$$\int |f * g|^p dx = \int \left(\int f(y)g(x-y)dy \right)^p dx.$$

Or par l'inégalité de Hölder,

$$\int f(y)g(x-y) = \int f(y)^{\frac{1}{p}} g(x-y) f^{\frac{p-1}{p}}(y) \leq \left(\int f(y)g(x-y)^p \right)^{\frac{1}{p}} \left(\int f(y) \right)^{\frac{p-1}{p}}.$$

Donc par le théorème de Fubini,

$$\begin{aligned} \int \left(\int f(y)g(x-y) \right)^p &\leq \int \left(\int f(y)g(x-y)^p dy \right) \left(\int f(y)dy \right)^{p-1} dx \\ &\leq \int f(y)dy \int g(x)^p dx \left(\int f(y) \right)^{p-1} \\ &\leq \left(\int f(y)dy \right)^p \left(\int g(x)^p dx \right). \end{aligned}$$

Cela donne $\|f * g\|_{L^p} \leq \|f\|_{L^1} \|g\|_{L^p}$.

Exercice 4.3.3. Contrexemples pour le théorème de Fubini, [14], pp. 471-473 1) On considère la mesure de comptage sur $\mathbb{N} \times \mathbb{N}$. On pose pour $m \geq 1$, $f(m, m) = 1$ et $f(m + 1, m) = -1$, on pose sinon $f(m, n) = 0$. Vérifier que

$$\sum_m \sum_n f(m, n) = 1, \text{ mais que } \sum_n \sum_m f(m, n) = 0.$$

Quelle est l'hypothèse du théorème de Fubini qui n'est pas vérifiée? 2) On considère sur $[0, 1] \times [1, +\infty[$ muni de la mesure de Lebesgue, $f(x, y) = e^{-xy} - 2e^{-2xy}$. Montrer que

$$\int_0^1 \int_1^\infty f(x, y) dy dx = \int_0^1 x^{-1}(e^{-x} - e^{-2x}) dx > 0$$

et

$$\int_1^\infty \int_0^1 f(x, y) dx dy = \int_1^\infty y^{-1}(e^{-2y} - e^{-y}) dy < 0.$$

Quelle est l'hypothèse du théorème de Fubini qui n'est pas vérifiée?

3) Soit $X = [0, 1]$ muni de la mesure de Lebesgue λ et $Y = [0, 1]$ muni de la tribu des parties et de la mesure de comptage, μ . On pose $f(x, y) = 1$ si $x = y$ et $f(x, y) = 0$ sinon. Montrer que

$$\int_X \int_Y f(x, y) d\mu(y) d\lambda(x) = 1 \text{ et } \int_Y \int_X f(x, y) d\lambda(x) d\mu(y) = 0.$$

Vérifier que f est bien mesurable pour la mesure produit. Quelle est l'hypothèse du théorème de Fubini qui n'est pas vérifiée?

Solution

Dans la première question, c'est l'hypothèse que f soit sommable qui n'est pas vérifiée. Dans la seconde, aussi, car si f était sommable, l'ordre d'intégration serait indifférent. On peut d'ailleurs vérifier directement que $\int |f(x, y)| dx dy = \int_0^{+\infty} (|e^{-y} - 2e^{-2y}| \int_0^{\min(1, y)} \frac{1}{x} dx) dy$. Dans la troisième, la fonction f est bien bornée, mesurable et sommable, mais la seconde mesure n'est pas σ -finie.

Exercice 4.3.4. On admet que l'on peut ordonner les réels de $[0, 1]$ par une relation d'ordre total $<'$ de telle manière que $\forall x \in [0, 1], \{y, y <' x\}$ soit un ensemble dénombrable.

(Cette propriété s'appelle l'hypothèse du continu. Il a été démontré par Kurt Gödel qu'elle ne contredisait pas les axiomes classiques de la théorie des ensembles et par Paul Cohen que son contraire ne les contredisait pas non plus.)

On pose $X = Y = [0, 1]$, munis de la tribu borélienne et de la mesure de Lebesgue. On pose $f(x, y) = 1$ si $x <' y$ et $f(x, y) = 0$ sinon. Montrer que

$$\forall x \int_X f(x, y) dx = 0, \text{ et } \forall y \int_Y f(x, y) dy = 1.$$

Quelle est l'hypothèse du théorème de Fubini qui n'est pas vérifiée?

Solution

Toutes les hypothèses du théorème de Fubini sont vérifiées, ... sauf une, la mesurabilité de f .

Exercice 4.3.5. Application directe du théorème de Fubini, [14] p. 473 Soit $X = \mathbb{N}$ muni de la mesure de comptage et (Y, \mathcal{T}, μ) un espace mesurable de mesure σ -finie. Si les fonctions f_n sont \mathcal{T} -mesurables et si $\sum_n \int |f_n| d\mu < \infty$, alors

$$\sum_n \int f_n d\mu = \int \sum_n f_n d\mu.$$

Exercice 4.3.6. Application directe du théorème de Fubini, [14] p. 473.

Soit $g \geq 0$ mesurable sur (X, \mathcal{T}, μ) et λ la mesure de Lebesgue sur \mathbb{R} . Montrer que

$$\int_X g d\mu = (\mu \otimes \lambda)(\{(x, y), 0 \leq y < g(x)\}) = \int_0^\infty \mu(\{x, g(x) > y\}) dy.$$

(Vérifier que $\{(x, y), 0 \leq y < g(x)\}$ est mesurable en remarquant que $(x, y) \rightarrow (g(x), y)$ mesurable et que $(a, b) \rightarrow b - a$ aussi.)

Exercice 4.3.7. Application 3 du théorème de Fubini, [14] p. 473.

Soit μ une mesure finie sur \mathbb{R} et $F(x) = \mu(]-\infty, x])$. Montrer que

$$\int (F(x+c) - F(x)) dx = c\mu(\mathbb{R}).$$

Exercice 4.3.8. Exercice 8.1 page 29 de [26] (calculs de densités de v.a.)

Chapitre 5

Indépendance

Dans tout ce chapitre, $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé.

5.1 Indépendance de familles d'événements et de variables aléatoires

Définition 5.1.1. a) Les événements A et B sont dits indépendants si et seulement si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
b) Deux familles d'événements \mathcal{A}_1 et \mathcal{A}_2 sont dites indépendantes si et seulement si tout élément de \mathcal{A}_1 est indépendant de tout élément de \mathcal{A}_2 .

Théorème 5.1.1. Soient \mathcal{C}_1 et \mathcal{C}_2 deux π -systèmes contenus dans \mathcal{F} . On note \mathcal{F}_i la tribu engendrée par \mathcal{C}_i . Alors \mathcal{F}_1 est indépendant de \mathcal{F}_2 si et seulement si \mathcal{C}_1 est indépendant de \mathcal{C}_2 .

Démonstration. La condition nécessaire étant triviale, passons à la condition suffisante. C'est une conséquence de la proposition 1.3.1, qui est une conséquence directe du théorème de Dynkin. En effet, pour un $C_1 \in \mathcal{C}_1$ fixé, les deux applications définies sur \mathcal{F} par :

$$I \in \mathcal{F} \rightarrow \mathbb{P}(I)\mathbb{P}(C_1) \quad , \quad I \in \mathcal{F} \rightarrow \mathbb{P}(I \cap C_1)$$

Ce sont deux mesures finies de même masse et coïncidant sur un π -système, à savoir \mathcal{C}_2 . Elles coïncident donc sur la tribu engendrée par \mathcal{C}_2 :

$$\forall C_1 \in \mathcal{C}_1, \forall F_2 \in \mathcal{F}_2, \mathbb{P}(F_2)\mathbb{P}(C_1) = \mathbb{P}(F_2 \cap C_1).$$

Soit maintenant $F_2 \in \mathcal{F}_2$ fixé. Considérons les deux applications

$$I \in \mathcal{F} \rightarrow \mathbb{P}(I)\mathbb{P}(F_2) \quad , \quad I \in \mathcal{F} \rightarrow \mathbb{P}(I \cap F_2)$$

sont deux mesures de masses égales finies coïncidant sur le π -système \mathcal{C}_1 : elles coïncident sur la tribu engendrée \mathcal{F}_1 , c.q.f.d. \square

On verra dans la section 3 une généralisation de ce théorème (théorème des coalitions).

Définition 5.1.2. Soient X_1 et X_2 deux variables aléatoires. On dit que X_1 et X_2 sont indépendantes si et seulement si les tribus engendrées $\sigma(X_1)$ et $\sigma(X_2)$ le sont.

Remarque 5.1.1. Si X_1 et X_2 sont deux variables aléatoires indépendantes et si f_1 et f_2 sont des fonctions boréliennes alors $f_1(X_1)$ et $f_2(X_2)$ sont indépendantes (car $\sigma(f_i(X_i)) \subset \sigma(X_i)$). Par exemple, si X_1 et X_2 sont deux vecteurs aléatoires alors toute composante (ou marginale) de X_1 est indépendante de toute composante (marginale) de X_2 .

Lois jointes

Définition 5.1.3. Soient X_1 et X_2 deux variables aléatoires à valeurs respectivement dans \mathbb{R}^n et \mathbb{R}^m . La loi jointe de X_1 et X_2 est la loi du couple $(X_1, X_2) \in \mathbb{R}^{n+m}$. C'est donc une probabilité sur $(\mathbb{R}^{n+m}, \mathcal{B}(\mathbb{R}^{n+m})) = (\mathbb{R}^n \times \mathbb{R}^m, \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^m))$.

On n'a en général bien-sûr pas $\mathbb{P}_{(X_1, X_2)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$. C'est cependant vrai dans le cas de deux variables aléatoires indépendantes.

Proposition 5.1.1. X_1 et X_2 sont indépendantes si et seulement si la loi du couple (X_1, X_2) est le produit tensoriel des lois de X_1 et de X_2 , $\mathbb{P}_{(X_1, X_2)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$.

Démonstration. Si X_1 et X_2 sont indépendantes, pour tout borélien A de \mathbb{R}^n et tout borélien B de \mathbb{R}^m ,

$$\mathbb{P}_{(X_1, X_2)}(A \times B) = \mathbb{P}(X_1 \in A, X_2 \in B) = \mathbb{P}(X_1 \in A)\mathbb{P}(X_2 \in B) = \mathbb{P}_{X_1}(A)\mathbb{P}_{X_2}(B)$$

et $\mathbb{P}_{(X_1, X_2)}$ est une probabilité sur $(\mathbb{R}^{n+m}, \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^m))$. Par l'unicité dans le théorème de Fubini,

$$\mathbb{P}_{(X_1, X_2)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$$

La réciproque se traite de même en utilisant ici la propriété caractéristique des mesures produits dans le théorème de Fubini. \square

Proposition 5.1.2. Soient X_1 et X_2 deux variables aléatoires à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Les affirmations suivantes sont équivalentes :

i) X_1 est indépendante de X_2 .

ii) Pour toutes fonctions boréliennes positives f_1 et f_2 , on a

$$\mathbb{E}[f_1(X_1)f_2(X_2)] = \mathbb{E}[f_1(X_1)]\mathbb{E}[f_2(X_2)]$$

iii) Pour toutes fonctions boréliennes bornées f_1 et f_2 , on a

$$\mathbb{E}[f_1(X_1)f_2(X_2)] = \mathbb{E}[f_1(X_1)]\mathbb{E}[f_2(X_2)]$$

Démonstration. La démonstration peut se faire très rapidement en invoquant la proposition précédente et le théorème de Fubini. Nous choisirons ici de faire la preuve avec la "machine standard".

i) \Rightarrow ii) : ii) est vraie pour les fonctions indicatrices de boréliens par définition de l'indépendance. Par linéarité de l'espérance, c'est aussi vrai en développant pour les fonctions étagées de la forme $\sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$ où les α_i sont des scalaires et les A_i des boréliens. Or toute fonction borélienne positive est limite croissante de fonctions de cette forme. Par application du théorème de Beppo-Levi, on a ii).

ii) \Rightarrow iii) : Il suffit d'écrire f_i comme la différence de sa partie positive et de sa partie négative $f_i = f_i^+ - f_i^-$ et de développer à nouveau.

iii) \Rightarrow i) : Prendre pour $A_1, A_2 \in \mathcal{B}(\mathbb{R}^d)$, $f_1 := \mathbb{1}_{A_1}$, $f_2 := \mathbb{1}_{A_2}$. Alors $\mathcal{E}(f_1(X_1)f_2(X_2)) = \mathbb{P}(X_1 \in A_1 \text{ et } X_2 \in A_2)$ et $\mathcal{E}(f_i(X_i)) = \mathbb{P}(X_i \in A_i)$, $i = 1, 2$. \square

5.2 Critères d'indépendance

1) En termes de fonction de répartition

On rappelle que si X est une variable aléatoire à valeurs dans \mathbb{R}^d , la fonction de répartition de X est donnée par

$$\forall x \in \mathbb{R}^d, F_X(x) = \mathbb{P}[X \leq x]$$

où l'on a muni \mathbb{R}^d de l'ordre partiel usuel. On rappelle d'autre part que la loi de X est la mesure image de \mathbb{P} par X , c'est-à-dire la probabilité sur \mathbb{R}^d , \mathbb{P}_X définie par

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X(A) = \mathbb{P}(X \in A)$$

On rappelle enfin que la fonction de répartition caractérise entièrement la loi de X (conséquence du théorème de Dynkin).

X_1 est indépendante de X_2 si et seulement si $F_{(X_1, X_2)} = F_{X_1} \otimes F_{X_2}$ où $F_{X_1} \otimes F_{X_2}$ est définie par

$$(F_{X_1} \otimes F_{X_2})(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$$

$F_{(X_1, X_2)} = F_{X_1} \otimes F_{X_2}$ est en effet équivalent à $\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = \mathbb{P}(X_1 \leq x_1)\mathbb{P}(X_2 \leq x_2)$. Or $\mathcal{C}_i = \{\{X_i \leq x_i\}; x_i \in \mathbb{R}^{d_i}\}$ est un π -système engendrant $\sigma(X_i)$. La condition précédente nous dit que \mathcal{C}_1 est indépendant de \mathcal{C}_2 donc par le théorème 5.1.1, $\sigma(X_1)$ est indépendant de $\sigma(X_2)$, c.q.f.d.

2) *En termes de densités*

a) Si les variables aléatoires X_1 et X_2 sont indépendantes et admettent des densités f_1 et f_2 alors la variable aléatoire (X_1, X_2) admet une densité $f_1 \otimes f_2$, produit direct de f_1 et f_2 :

$$\forall x_1 \in \mathbb{R}^{d_1}, \forall x_2 \in \mathbb{R}^{d_2}, (f_1 \otimes f_2)(x_1, x_2) = f_1(x_1)f_2(x_2)$$

b) Inversement, si la variable aléatoire (X_1, X_2) admet une densité produit direct de deux fonctions intégrables positives f_1 et f_2 alors à des constantes multiplicatives près, X_1 admet une densité f_1 et X_2 admet une densité f_2 et X_1 et X_2 sont indépendantes.

a) En utilisant le théorème de Fubini, on voit aisément que $\mathbb{P}_{(X_1, X_2)}$ et la probabilité définie par $f_1 \otimes f_2$ coïncident sur le π -système $\mathcal{C} = \{A_1 \times A_2; A_1 \in \mathcal{B}(\mathbb{R}^{d_1}), A_2 \in \mathcal{B}(\mathbb{R}^{d_2})\}$ donc sont égales par l'unicité du prolongement des mesures (Dynkin : deux mesures finies qui coïncident sur un π -système \mathcal{P} coïncident sur $\sigma(\mathcal{P})$).

b) Si l'on se donne des boréliens A_1 et A_2 , on a

$$\mathbb{P}((X_1, X_2) \in A_1 \times A_2) = \left(\int_{A_1} \frac{f_1(x_1)}{\int_{\mathbb{R}^{d_1}} f_1} dx_1 \right) \left(\int_{A_2} \frac{f_2(x_2)}{\int_{\mathbb{R}^{d_2}} f_2} dx_2 \right)$$

En prenant successivement $A_1 = \mathbb{R}^{d_1}$ puis $A_2 = \mathbb{R}^{d_2}$, on obtient que la loi de X_1 est donnée par la densité $\frac{f_1}{\int f_1}$ et celle de X_2 par la densité $\frac{f_2}{\int f_2}$. On déduit de l'égalité précédente l'indépendance de X_1 et X_2 .

3) *Pour des variables aléatoires discrètes*

Si les variables aléatoires X_1 et X_2 sont discrètes, il en est de même pour (X_1, X_2) . Pour que X_1 et X_2 soient indépendantes, il faut et il suffit que

$$\forall x_1 \in X_1(\Omega), \forall x_2 \in X_2(\Omega), \mathbb{P}(X_1 = x_1, X_2 = x_2) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)$$

La démonstration de cette propriété est laissée en exercice. Le point important est de remarquer que pour une variable discrète X à valeurs dans un ensemble discret E les ensembles atomiques $\{X = e\}$, $e \in E$ forment un π -système qui engendre $\sigma(X)$.

Nous passons maintenant à une propriété vérifiée par les variables aléatoires indépendantes. Inutile de dire que la réciproque de cette propriété est en général fausse.

Proposition 5.2.1. *Soit X_1 et X_2 deux variables aléatoires réelles indépendantes.*

1) *Si X_1 et X_2 sont dans \mathcal{L}^1 alors $X_1 X_2$ est dans \mathcal{L}^1 et $\mathbb{E}(X_1 X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2)$.*

2) *Si X_1 et X_2 sont dans \mathcal{L}^2 alors $\text{cov}(X_1, X_2) = 0$ et $\sigma_{X_1+X_2}^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2$.*

Rappel : La covariance de deux variables aléatoires réelles X et Y est définie par $\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ et $\sigma_X^2 = \text{Cov}(X, X)$.

Démonstration. 1) Pour montrer que $X_1 X_2$ est intégrable, on applique le ii) de la proposition 5.1.2 avec $f_1 = f_2 = |\cdot|$. Pour établir l'égalité entre les espérances, on écrit X_1 et X_2 comme différence de leurs parties positives et négatives respectives. Par linéarité de l'espérance, on se ramène ainsi au cas de variables aléatoires positives. On applique alors à nouveau la proposition 5.1.2, ii) avec $f_1 = f_2 = id$.

2) Evident.

□

5.3 Indépendance généralisée

Soit I un ensemble quelconque, dénombrable ou pas.

Définition 5.3.1. (Indépendance de familles d'événements et de variables aléatoires)

– Soit $(A_i)_{i \in I}$ une famille d'événements. Les événements $(A_i)_{i \in I}$ sont dits indépendants ssi :

$$\forall J \subset \subset I, \mathbb{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j)$$

- Des familles d'événements $(\mathcal{A}_i)_{i \in I}$ sont indépendantes ssi pour tout choix d'un A_i dans \mathcal{A}_i , la famille d'événements $(A_i)_{i \in I}$ est constituée d'événements indépendants.
- On dira qu'une famille de variables aléatoires $(X_i)_{i \in I}$ est une famille de variables aléatoires indépendantes ssi les $(\sigma(X_i))_{i \in I}$ le sont.

Le théorème suivant qui est une généralisation du théorème vu dans la section 1 signifie simplement que l'indépendance se conserve par "regroupements disjoints".

Théorème 5.3.1. ("des coalitions")

Soit une famille $(\mathcal{C}_i)_{i \in I}$ de π -systèmes indépendants contenus dans une tribu \mathcal{F} . On note $\mathcal{F}_i = \sigma(\mathcal{C}_i)$ la tribu engendrée par \mathcal{C}_i . Soit $(I_j)_{j \in J}$ une partition quelconque de I et pour tout $j \in J$, $\mathcal{G}_j = \sigma(\cup_{i \in I_j} \mathcal{C}_i)$. Alors les tribus $(\mathcal{G}_j)_{j \in J}$ sont indépendantes. En particulier, les tribus $(\mathcal{F}_i)_{i \in I}$ sont indépendantes.

Démonstration. La démonstration dans le cas fini est faite dans le Durett et dans l'Ouvrard. Durett donne pratiquement tous les éléments pour traiter le cas infini. Dans le cas infini, on pourra trouver une démonstration bien faite dans le Buchwalter.

On commence par montrer le théorème pour $J = I$ et $I_j = \{j\}$. Dans ce cas, il est clair que l'on peut se ramener au cas I fini. On suppose pour simplifier les notations que $I = \{1, \dots, n\}$ et on rappelle que le cas $n = 2$ a été vu dans la section 1.

On fixe $A_2 \in \mathcal{C}_2, \dots, A_n \in \mathcal{C}_n$ et on considère les deux applications suivantes sur \mathcal{F}_1

$$B_1 \in \mathcal{F}_1 \rightarrow \mathbb{P}(B_1 \cap A_2 \cap \dots \cap A_n)$$

$$B_1 \in \mathcal{F}_1 \rightarrow \mathbb{P}(B_1) \mathbb{P}(A_2) \dots \mathbb{P}(A_n)$$

Ce sont deux mesures de même masse finie (faire $B_1 = \Omega_1$) qui coïncident sur le π -système \mathcal{C}_1 donc sur \mathcal{F}_1 . On fixe ensuite $B_1 \in \mathcal{F}_1, A_3 \in \mathcal{C}_3, \dots, A_n \in \mathcal{C}_n$ et on considère les deux applications

$$B_2 \in \mathcal{F}_2 \rightarrow \mathbb{P}(B_1 \cap B_2 \cap A_3 \dots \cap A_n)$$

$$B_2 \in \mathcal{F}_2 \rightarrow \mathbb{P}(B_1) \mathbb{P}(B_2) \mathbb{P}(A_3) \dots \mathbb{P}(A_n)$$

qui sont des mesures de même masse finie (appliquer ce qui précède en faisant $B_2 = \Omega_2$) coïncidant sur le π -système \mathcal{C}_2 donc sur \mathcal{F}_2 . Et on continue ainsi pour obtenir le résultat annoncé.

On a donc obtenu le théorème pour le cas de la partition triviale. On définit

$$\mathcal{L}_j = \left\{ \bigcap_{i \in I_j} A_i; A_i \in \mathcal{C}_i \text{ pour un nombre fini d'indices et } A_i = \Omega \text{ autrement} \right\}$$

C'est un π -système contenant tous les \mathcal{C}_i pour $i \in I_j$ donc $\sigma(\mathcal{L}_j)$ contient \mathcal{G}_j . Or les \mathcal{C}_i sont indépendants donc les \mathcal{L}_j le sont : il suffit de l'écrire. D'après ce qui précède, les $\sigma(\mathcal{L}_j)_j$ sont des tribus indépendantes et il en est donc de même des $(\mathcal{G}_j)_j$ puisque $\mathcal{G}_j \subset \sigma(\mathcal{L}_j)$. \square

Remarque 5.3.1. *Voici une application discrète du précédent théorème. Soient A_1, \dots, A_n des événements indépendants. On considère les événements $\tilde{A}_1, \dots, \tilde{A}_n$ où $A_i = A_i$ ou A_i^c . Montrer que les \tilde{A}_i sont indépendants. Cela peut aussi se vérifier à la main, mais ce n'est pas si facile (faire une récurrence sur le nombre de complémentaires A_i^c dans la séquence \tilde{A}_i).*

5.4 Indépendance et événements asymptotiques

Dans cette section, nous énonçons trois théorèmes concernant la tribu asymptotique : loi du 0-1, lemme de Borel-Cantelli 1 et lemme de Borel-Cantelli 2. Le lemme de Borel-Cantelli 1 n'a rien à voir avec l'indépendance mais son énoncé étant une sorte de réciproque du 2, il est d'usage de le placer ici.

Définition 5.4.1. *Soit $(\mathcal{F}_n)_n$ une suite de sous-tribus de \mathcal{F} . On appelle tribu asymptotique des $(\mathcal{F}_n)_n$ la tribu*

$$\mathcal{F}_\infty = \bigcap_{n \in \mathbb{N}} \sigma \left(\bigcup_{k \geq n} \mathcal{F}_k \right)$$

Les éléments de \mathcal{F}_∞ sont appelés les événements asymptotiques et une variable aléatoire \mathcal{F}_∞ -mesurable est dite asymptotique.

Si $(X_n)_n$ est une suite de variables aléatoires, la tribu asymptotique associée aux X_n est la tribu asymptotique associée aux $\mathcal{F}_n = \sigma(X_n)$.

Par exemple, l'événement $\{\omega \in \Omega; X_n(\omega) \text{ converge dans } \mathbb{R}\}$ est un événement asymptotique (par rapport aux X_n) et $\limsup X_n, \liminf X_n$ sont des variables aléatoires asymptotiques. Traitons un exemple :

$$\liminf_n X_n \geq a \Leftrightarrow (\forall l > 0, \exists n, \forall k \geq n, X_n \geq a - \frac{1}{l}) \Leftrightarrow \omega \in \bigcap_l \bigcup_n \bigcap_{k \geq n} [X_k \geq a].$$

Théorème 5.4.1. *(Loi du "Tout ou Rien", ou du "0-1")
Si la suite $(\mathcal{F}_n)_n$ est constituée de tribus indépendantes alors*

$$\forall A \in \mathcal{F}_\infty, \mathbb{P}(A) = 0 \text{ ou } \mathbb{P}(A) = 1$$

Démonstration. On démontre que \mathcal{F}_∞ est indépendante d'elle-même ce qui implique trivialement le résultat. D'après le théorème des coalitions, $\sigma(\bigcup_{k \geq n} \mathcal{F}_k)$ est indépendant de \mathcal{F}_p pour $p < n$. Or \mathcal{F}_∞ est l'intersection des $\sigma(\bigcup_{k \geq n} \mathcal{F}_k)$ donc \mathcal{F}_∞ est indépendant de \mathcal{F}_p pour tout p . Il en résulte que \mathcal{F}_∞ est indépendant de $\sigma(\bigcup_{p \geq 1} \mathcal{F}_p)$ et $\mathcal{F}_\infty \subset \sigma(\bigcup_{p \geq 1} \mathcal{F}_p)$ donc \mathcal{F}_∞ est indépendante d'elle-même. \square

On rappelle que si l'on a une suite d'événements $(A_n)_n$, on définit

$$\limsup A_n = \bigcap_n \bigcup_{p \geq n} A_p = \{\omega; \omega \text{ appartient à une infinité de } A_n\}$$

(en anglais *infinitely often, i.o.*)

$$\liminf A_n = \bigcup_n \bigcap_{p \geq n} A_p = \{\omega; \omega \text{ appartient à tous les } A_n \text{ sauf à un nombre fini}\}$$

(en anglais, *eventually, ev.*).

Lemme 5.4.1. *(Borel-Cantelli 1)*

Si $\sum_n \mathbb{P}(A_n) < +\infty$ alors $\mathbb{P}(\limsup A_n) = 0$. (C'est à dire : A_n ne se produit qu'un nombre fini de fois).

Démonstration. Pour tout n , $\limsup A_k \subset \cup_{k \geq n} A_k$ donc

$$\mathbb{P}(\limsup A_k) \leq \mathbb{P}(\cup_{k \geq n} A_k) \leq \sum_{k \geq n} \mathbb{P}(A_k)$$

et le dernier terme de droite tend vers 0 quand n tend vers l'infini par hypothèse d'où le résultat. \square

Lemme 5.4.2. (Borel-Cantelli 2)

Si les $(A_n)_n$ sont des événements indépendants et si $\sum_n \mathbb{P}(A_n) = +\infty$ alors $\mathbb{P}(\limsup A_n) = 1$.

Démonstration. Il est aisé de voir que

$$\{\limsup A_k\}^c = \liminf A_k^c = \cup_n \cap_{k \geq n} A_k^c$$

Posons $p_k = \mathbb{P}(A_k)$. On a $\mathbb{P}(\cap_{k \geq n} A_k^c) = \prod_{k \geq n} (1 - p_k)$ car par indépendance,

$$\mathbb{P}(\cap_{m \geq k \geq n} A_k^c) = \prod_{m \geq k \geq n} (1 - p_k)$$

et on peut faire tendre m vers l'infini grâce au théorème de convergence monotone.

Pour tout $x \geq 0$, $1 - x \leq \exp(-x)$ donc

$$\prod_{k \geq n} (1 - p_k) \leq \exp\left(-\sum_{k \geq n} p_k\right) = 0$$

Donc $\mathbb{P}((\limsup A_k)^c) = 0$. \square

5.5 Lemmes de Borel-Cantelli, loi du 0-1, mesurabilité relative

Exercice 5.5.1. Application du théorème de Fubini, [14] p. 473

Rappel : Soit F croissante sur \mathbb{R} , continue à droite. Alors il y a une unique mesure $\mu = dF$ sur \mathbb{R} muni de la tribu borélienne telle que $\mu([a, b]) = F(b) - F(a)$. On dit que F est une fonction de Stieltjes et μ une mesure de Stieltjes. La preuve est essentiellement la même que celle qui donne l'existence de la mesure de Lebesgue.

1) Rappeler pourquoi l'ensemble des sauts d'une fonction croissante est dénombrable. Peut-on énumérer ces sauts x_n de telle sorte que $x_{n+1} \leq x_n$, $n \in \mathbb{N}$?

2) Soit $(x_i)_{i \in I}$ l'ensemble des sauts de F et α_i la valeur du saut en x_i . Montrer que

$$dF = \sum_i \alpha_i \delta_{x_i} + d\tilde{F},$$

où \tilde{F} est une fonction croissante continue.

3) Soient F , μ et G , ν deux couples de Stieltjes. On note $\mu = dF$, $\nu = dG$. Montrer que $(\mu \times \nu)(\{a < x = y \leq b\}) = \sum_{x \in]a, b]} \mu(\{x\})\nu(\{x\})$. (On appliquera le théorème de Fubini à l'intégrale $\int \int \mathbb{1}_{y=x} d\mu \otimes d\nu$).

4) Montrer que pour $-\infty < a \leq b < +\infty$,

$$(i) \int_{]a, b]} (F(y) - F(a)) dG(y) = (\mu \times \nu)(\{(x, y), a < x \leq y \leq b\});$$

5) Dédire des deux questions précédentes que

$$(ii) \int_{]a, b]} F(y) dG(y) + \int_{]a, b]} G(y) dF(y) = F(b)G(b) - F(a)G(a) + \sum_{x \in]a, b]} \mu(\{x\})\nu(\{x\}).$$

(Indication : Dans $(\mu \otimes \nu)(\{(x, y), ; a < x \leq y \leq b\}) + \mu \otimes \nu(\{(x, y), a < y \leq x \leq b\})$, la diagonale apparaît deux fois!).

6) Vérifier sur un exemple comment apparaît le second terme en posant $F(x) = G(x) = \mathbb{1}_{[0, +\infty[}(x)$ en prenant $a < 0 < b$.

$$(iii) \int_{]a, b]} 2F(y)dF(y) = F^2(b) - F^2(a) \text{ si } F = G \text{ est continue.}$$

Solution

1) et 2) : classique, poser $d\tilde{F} = dF - \sum_i \alpha_i \delta_{x_i}$ et vérifier que \tilde{F} n'a plus de sauts.

3) En appliquant le théorème de Fubini on a

$$\int \int \mathbb{1}_{y=x} d\mu \otimes \delta\nu = \int \left(\int \mathbb{1}_{y=x}(y) d\nu(y) \right) d\mu(x) = \int \nu(\{x\}) d\mu(x) = \sum_x \nu(\{x\}) \mu(\{x\}).$$

En effet, la deuxième intégrale est en fait une somme dénombrable puisque $d\mu(\{x\})$ est nulle sauf pour x appartenant à l'ensemble dénombrable des sauts de F . On intègre donc une fonction nulle sauf en un ensemble dénombrable de points par rapport à la mesure $d\nu$.

4) Preuve de (i) :

$$\begin{aligned} \int_{]a, b]} (F(y) - F(a)) dG(y) &= \int_{]a, b]} \left(\int_{]a, y]} dF(x) \right) dG(y) = \int_{]a, b]} \left(\int_{]a, b]} \mathbb{1}_{x \leq y}(x, y) dF(x) \right) dG(y) = \\ &= \int \int_{]a, b]^2} \mathbb{1}_{x \leq y}(x, y) (d\mu \otimes d\nu)(x, y) = (\mu \otimes \nu)(\{(x, y), a < x \leq y \leq b\}), \end{aligned}$$

où l'avant dernière égalité s'obtient par le théorème de Fubini.

5) Preuve de (ii). En utilisant le (i),

$$\begin{aligned} \int_{]a, b]} F(y) dG(y) + \int_{]a, b]} G(y) dF(y) &= F(a) \int_{]a, b]} dG(y) + G(a) \int_{]a, b]} dF(x) + \\ &= (\mu \otimes \nu)(\{a < x \leq y \leq b\}) + (\mu \otimes \nu)(\{a < y \leq x \leq b\}) = \\ &= (\mu \otimes \nu)(\{]a, b]^2\}) + F(a)(G(b) - G(a)) + G(a)(F(b) - F(a)) + (\mu \otimes \nu)(\{a < x = y \leq b\}) = \\ &= F(b)G(b) - F(a)G(a) + (\mu \otimes \nu)(\{a < x = y \leq b\}). \end{aligned}$$

Exercice 5.5.2. Un modèle pour le jeu de pile ou face ([26], 9.3., p. 53-59)

1) Soit $x \in [0, 1]$. On définit par récurrence les suites de terme général

$$R_0(x) = x$$

et pour tout $n > 0$ $D_n(x) = [2R_{n-1}(x)]$ et $R_n(x) = 2R_{n-1}(x) - D_n(x)$ Vérifier que l'on a pour tout $n > 0$

$$x = \sum_{j=1}^n \frac{D_j(x)}{2^j} + \frac{1}{2^n} R_n(x).$$

Montrer que $x \in [0, 1]$ s'écrit d'une manière unique sous la forme précédente, $x = 0, x_1 x_2 \dots x_k \dots$ avec $x_i \in \{0, 1\}$ si et seulement si x n'est pas un rationnel dyadique, c'est-à-dire un rationnel de la forme $\frac{p}{2^q}$ avec $p, q \in \mathbb{N}$ et $p \leq 2^q$. Préciser la forme des $D_n(x)$ quand x est un rationnel dyadique. Le but de l'exercice est de montrer la proposition suivante :

Proposition 5.5.1. Sur $[0, 1)$ muni de la tribu des Boréliens et la mesure de Lebesgue, la suite $(D_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre $\frac{1}{2}$. La variable aléatoire R_n est de loi uniforme sur $[0, 1)$ et la variable R_n et les variables (D_1, \dots, D_n) sont indépendantes.

2) On va montrer que $(D_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre $1/2$. On note $\varepsilon^n = (\varepsilon_1, \dots, \varepsilon_n)$ un élément de $\{0, 1\}^n$ et $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n, \dots)$ un élément de $\{0, 1\}^{\mathbb{N}}$.

2a) Montrer que $\cap_{j=1}^n (D_j = \varepsilon_j)$ est un intervalle de $[0, 1]$ que l'on précisera et calculer sa probabilité. (On trouve 2^{-n}).

2b) En déduire que pour toute partie non vide J de $\{1, 2, \dots, n\}$ on a

$$\mathbb{P}(\cap_{j \in J} (D_j = \varepsilon_j)) = \frac{1}{2^{|J|}}.$$

Conclure que les D_n sont bien indépendantes.

Solution

On a

$$\begin{aligned} \mathbb{P}(\cap_{j \in J} (D_j = \varepsilon_j)) &= \sum_{\varepsilon^{n-|J|} \in \{0, 1\}^{J^c}} \mathbb{P}(\cap_{j \in J} (D_j = \varepsilon_j) \cap (\cap_{k \in J^c} (D_k = \varepsilon^{n-|J|}(k)))) = \\ &= \sum_{\varepsilon \in \{0, 1\}^{J^c}} \frac{1}{2^n} 2^{n-|J|} = \frac{1}{2^{|J|}}. \end{aligned}$$

3) On va montrer que, pour tout $n \in \mathbb{N}^*$, la variable aléatoire R_n est de loi uniforme sur $[0, 1)$ et que les variables aléatoires R_n et (D_1, D_2, \dots, D_n) sont indépendantes.

3a) Vérifier que $R_n(x) = 2^n x - \sum_{j=1}^n 2^{n-j} D_j(x)$. Remarquer qu'en développement dyadique, $R_n(x) = 0, D_{n+1}(x)D_{n+2}(x)\dots$ et il est donc immédiat que $R_n \in \sigma(D_{n+1}, D_{n+2}, \dots)$ est indépendante de D_1, \dots, D_n .

3b) Montrer que la loi de R_n est uniforme, i.e.

$$\mathbb{E}(f(R_n)) = \int_{\mathbb{R}} \mathbb{1}_{[0,1[}(y) f(y) dy.$$

3c) Déduire des deux questions précédentes que pour toute fonction borélienne positive $f : \mathbb{R} \rightarrow \mathbb{R}$ et toute partie J de $\{1, 2, \dots, n\}$,

$$\mathbb{E} [f(R_n) \prod_{i \in J} \mathbb{1}_{D_i = \varepsilon_i}] = \mathbb{P}(\cap_{i \in J} (D_i = \varepsilon_i)) \left[\int_{\mathbb{R}} \mathbb{1}_{[0,1[}(y) f(y) dy \right]. \quad (5.5.1)$$

Remarquer que le résultat du 3b) est (5.5.1) dans le cas $J = \{1, 2, \dots, n\}$. En déduire le cas général (5.5.1).

4) En vérifiant et utilisant la relation $D_n = -R_n + 2R_{n-1}$, montrer que la suite de v.a. R_n n'est pas une suite de variables aléatoires indépendantes.

5) Soit $(\mu_j)_{j>0}$ une suite de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On se propose de construire explicitement des variables aléatoires $(X_j)_{j>0}$ indépendantes de lois respectives $(\mu_j)_{j>0}$.

5a) Soit $(N_j)_{j>0}$ une suite de sous-ensembles infinis de \mathbb{N}^* formant une partition. Pour tout $j > 0$, soit $\phi_j \in \mathbb{N}^*$ une énumération des N_j . Pour tout $j > 0$, on pose

$$Y_j = \sum_{k=1}^{\infty} \frac{1}{2^k} D_{\phi_j(k)}$$

Montrer que les variables $(Y_j)_{j>0}$ sont indépendantes. On utilisera le théorème des coalitions ([26] p. 47) :

Théorème 5.5.1. Soit une famille $(C_i)_{i \in I}$ de π -systèmes contenus dans \mathcal{A} et indépendants; on note \mathcal{F}_i la tribu engendrée par C_i . Soit $(I_j)_{j \in J}$ une partition de I et \mathcal{A}_j la tribu engendrée par $\cup_{i \in I_j} C_i$. Alors les tribus $(\mathcal{A}_j)_{j \in J}$ sont indépendantes.

5b) Montrer que la loi de $Y_{j,n} = \sum_{k=1}^n \frac{1}{2^k} D_{\varphi_j(k)}$ est la même que celle de $Z_n = \sum_{k=1}^n \frac{1}{2^k} D_k$ et en déduire que la loi de Y_j est uniforme sur $[0, 1]$.

5c) (Voir exercice précédent 4.3.8.) Soit F_j la fonction de répartition de μ_j . On définit la pseudo-inverse de F_j par

$$G_j(t) = \inf_{x \in \mathbb{R}} \{F_j(x) \geq t\}$$

On pose $X_j = G_j(Y_j)$. Montrer que la loi de X_j est μ_j et conclure.

6) Lecture : [26] pp. 59-60, modélisation du jeu de pile ou face.

Quelques notations ([30], pp. 26-27). Soit A_n une suite d'événements d'un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$. On écrit " $A_n, i.o.$ " (*infinitely often*), pour l'événement

$$\{A_n \text{ i.o.}\} = \limsup_n A_n = \bigcap_n \bigcup_{k \geq n} A_k.$$

On dit que A_n se produit "infiniment souvent" si $\mathbb{P}(A_n, i.o.) = 1$. On définit de même " $A_n, ev.$ " (*eventually*), par

$$\{A_n \text{ ev.}\} = \liminf_n A_n = \bigcup_n \bigcap_{k \geq n} A_k.$$

Si $\mathbb{P}(A_n, ev.) = 1$, on dit que A_n "finit presque sûrement par se produire". Remarquer que $\{A_n \text{ ev.}\} = \{A_n^c \text{ i.o.}\}^c$. Avec ces notations, on rappelle les deux lemmes de Borel-Cantelli ([30] pp. 27 et 40)

(BC1) : Si E_n est une suite d'événements telle que $\sum_n \mathbb{P}(E_n) < \infty$, alors $\mathbb{P}(E_n, i.o.) = 0$.

(BC2) : Si E_n est une suite d'événements indépendants, alors

$$\sum_n \mathbb{P}(E_n) = +\infty \Rightarrow \mathbb{P}(E_n, i.o.) = 1.$$

Le but des deux exercices suivants est de donner deux contre-exemples à la loi des grands nombres (chapitre suivant). Le premier exercice prouve que celle-ci peut être fautive si les v.a. X_n ne sont pas équidistribuées. Le deuxième prouve que la loi des grands nombres peut être fautive si $E(|X_1|) = +\infty$.

Exercice 5.5.3. Un jeu qui n'est équitable qu'en apparence ([30] p. 228) Soient X_1, \dots, X_n des v.a.i. telles que $X_n = n^2 - 1$ avec probabilité n^{-2} et $X_n = -1$ avec probabilité $1 - n^{-2}$. Démontrer que $\mathbb{E}(X_n) = 0$, mais que si on pose $S_n = X_1 + X_2 + \dots + X_n$, alors

$$\frac{S_n}{n} \rightarrow -1.$$

Solution

On a $\mathbb{P}(X_n = n^2 - 1) = n^{-2}$, et donc $\sum_n \mathbb{P}(X_n = n^2 - 1) < \infty$. Par (BC1), $\mathbb{P}(X_n = n^2 - 1, i.o.) = 0$ et donc $\mathbb{P}(X_n = -1, \text{ à partir d'un certain rang.}) = 1$, ce qui prouve le résultat.

Exercice 5.5.4. Réciproque de la loi forte des grands nombres ([30], E4.6 p. 227)

1) Soit $Z \geq 0$ une v.a. et Y sa partie entière. Montrer que

$$Y = \sum_{n \in \mathbb{N}^*} \mathbb{1}_{Z \geq n}.$$

2) Dédurre que

$$\sum_{n \in \mathbb{N}^*} \mathbb{P}(Z \geq n) \leq \mathbb{E}(Z) \leq \sum_{n \in \mathbb{N}^*} \mathbb{P}(Z \geq n) + 1.$$

3) Soient X_n , v.a.i.i.d. telles que $\mathbb{E}(|X_n|) = +\infty$. Démontrer que

$$\sum_n \mathbb{P}(|X_n| \geq kn) = \infty \quad (k \in \mathbb{N}) \text{ et que } \limsup_n \frac{|X_n|}{n} = \infty, \text{ p.s.}$$

4) En déduire que si $S_n = X_1 + \dots + X_n$, alors $\limsup \frac{|S_n|}{n} = \infty$ p.s.

Solution

Le 1) est immédiat ; le 2) se déduit en utilisant les inégalités $Y \leq Z \leq Y + 1$ et en prenant l'espérance. Pour le 3), posons $Z = X_0$. Alors on sait par le 2) que $\sum_n \mathbb{P}(|Z| \geq n) = +\infty$. Comme $\mathbb{P}(|X_n| \geq n) = \mathbb{P}(|Z| \geq n)$, on en déduit que $\sum_n \mathbb{P}(|X_n| \geq n) = +\infty$. Les événements $\{|X_n| \geq n\}$ étant indépendants, on a par (BC2) : $|X_n| \geq kn$ i.o. et donc $\mathbb{P}(\limsup_n \frac{|X_n|}{n} \geq k) = 1$. Ceci étant vrai pour tout k , on conclut. 4) Utiliser : $\frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1} = \frac{X_n}{n}$ et minorer $\frac{|S_n|}{n}$.

Exercice 5.5.5. Limite supérieure d'une suite de variables normales indépendantes, [30], E4.5 p. 227

1) Soit G une v.a. normale $\mathcal{N}(0, 1)$. On pose $\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$. Démontrer que pour $x > 0$,

$$(x + x^{-1})^{-1} \varphi(x) \leq \mathbb{P}(G \geq x) \leq \frac{1}{x} \varphi(x).$$

2) Soit $(X_n)_n$ une suite de v.a.i. normales $\mathcal{N}(0, 1)$. Montrer qu'avec probabilité 1, $L \leq 1$, où L est définie par

$$L := \limsup_n \frac{X_n}{\sqrt{2 \log n}}.$$

3) Montrer que $\mathbb{P}(L = 1) = 1$.

4) On pose $S_n = X_1 + \dots + X_n$ et on rappelle que $\frac{S_n}{\sqrt{n}}$ est encore normale $\mathcal{N}(0, 1)$. Montrer que $\mathbb{P}(|S_n| < 2\sqrt{n \log n}, \text{ ev.}) = 1$. (voir la section 14.7 de [30] pour un résultat plus précis : *la loi du Logarithme Itéré*). Remarquer que cela implique la loi forte des grands nombres dans ce cas particulier : $\mathbb{P}(\frac{S_n}{n} \rightarrow 0) = 1$.

Solution

1) Pour l'inégalité de droite, utiliser $\varphi'(y) = -y\varphi(y)$ et pour l'inégalité de gauche, $(\frac{\varphi(y)}{y})' = -(1 + \frac{1}{y^2})\varphi(y)$.

2) Par l'inégalité de droite de la question 1), on a

$$\mathbb{P}(|X_n| \geq L\sqrt{2 \log n}) \leq \frac{1}{L\sqrt{4\pi \log n}} e^{-L^2 \log n} = \frac{1}{L\sqrt{4\pi \log n}} n^{-L^2}.$$

Si $L > 1$, alors $\sum_n \mathbb{P}(|X_n| \geq L\sqrt{2 \log n}) < \infty$ et, par (BC1), $\mathbb{P}(|X_n| \geq L\sqrt{2 \log n} \text{ i.o.}) = 0$, et donc $\mathbb{P}(\limsup_n \frac{|X_n|}{\sqrt{2 \log n}} \leq L) = \mathbb{P}(|X_n| \leq L\sqrt{2 \log n} \text{ ev.}) = 1$.

3) On par l'inégalité de gauche du 1),

$$\mathbb{P}(X_n \geq \sqrt{2 \log n}) \geq (L\sqrt{2 \log n} + \frac{1}{L\sqrt{2 \log n}}) n^{-L^2}.$$

Donc, si $L \leq 1$, on obtient $\sum_n \mathbb{P}(|X_n| \geq L\sqrt{2 \log n}) = +\infty$, et donc, par (BC2), $\mathbb{P}(|X_n| \geq L\sqrt{2 \log n}, i.o.) = 1$. Utilisant ceci, et le résultat du 2), on conclut que $L = 1$, p.s.

4) Les S_n ne sont pas indépendantes, mais on peut toujours invoquer le résultat du 2), qui n'utilise pas l'indépendance des X_n . On a donc

$$\limsup_n \left(\frac{S_n}{\sqrt{2n \log n}} \right) \leq 1, \text{ p.s.}$$

Cela implique que pour tout $C > \sqrt{2}$, $\mathbb{P}(\frac{S_n}{\sqrt{2n \log n}} \leq C, \text{ ev.}) = 1$ et on obtient le résultat annoncé en prenant par exemple $C = 2 > \sqrt{2}$.

Exercices 9.10, 9.11, 9.12 de [26], p. 84-85.

Exercice 5.5.6. mesurabilité relative, très important! ([30], p. 206) Soient X et Y deux variables aléatoires réelles $\Omega \rightarrow \mathbb{R}$. On note $\sigma(Y)$ la tribu engendrée par les ensembles $Y^{-1}(E)$, où E est un borélien quelconque de \mathbb{R} . On va montrer le

Théorème : X est $\sigma(Y)$ -mesurable si et seulement si elle est une fonction $X = f(Y)$ de Y , où $f : \mathbb{R} \rightarrow \mathbb{R}$ est borélienne.

Remarque 5.5.1. *Cet énoncé n'a rien à voir avec le fait que X et Y soient réelles et que l'on considère la tribu de Borel. On l'énonce avec des variables réelles pour le rendre plus intuitif. Se reporter à [30] (loc. cit.) pour la formulation générale.*

1) Expliquer pourquoi il suffit de prouver que X est bornée et $\sigma(Y)$ mesurable si et seulement s'il existe f borélienne bornée telle que $X = f(Y)$.

2) Soit \mathcal{H} la classe de toutes les fonctions bornées X sur Ω telles que $X = f(Y)$ pour une fonction f borélienne bornée. On va appliquer à \mathcal{H} le théorème de la classe monotone.

2a) Montrer que \mathcal{H} contient les fonctions indicatrices d'éléments de $\sigma(Y)$.

2b) Montrer que \mathcal{H} est un espace vectoriel contenant les constantes et que si $X_n \geq 0$, $X_n \in \mathcal{H}$, tend en croissant vers X bornée, alors $X \in \mathcal{H}$.

3) Conclusion. Après avoir terminé l'exercice, lire attentivement les pages 206 et 207 de [30].

Solution

1) Remplacer X par $\arctan(X)$ par exemple.

2a) Si $F \in \sigma(Y)$, alors $F = Y^{-1}(B)$ où B est borélien. Donc $\mathbb{1}_F = \mathbb{1}_B(Y)$.

2b) On a $X_n = f_n(Y)$ et f_n boréliennes. En appliquant \limsup_n à cette relation, on obtient $X = \limsup_n f_n(Y) = (\limsup_n f_n)(Y)$. Donc $X = f(Y)$ avec $f =: \limsup_n f_n$, qui est encore borélienne. (On pose $f = 0$, par exemple, sur le complémentaire de l'image de Y .)

3) Par le théorème de la classe monotone, \mathcal{H} contient toutes les fonctions bornées et $\sigma(Y)$ -mesurables. En appliquant le 1), on obtient le théorème.

Exercice 5.5.7. Deux exemples parlants de variables aléatoires deux à deux indépendantes et d'évènements deux à deux indépendants.

1) On considère $\Omega_1 = [0, 1]^3$ muni de la mesure de Lebesgue. Montrer que les coordonnées X, Y et Z sont des variables aléatoires indépendantes.

2) On considère l'espace de probabilité $\Omega_2 = \Omega_1 \cup (1, 1, 0) + \Omega_1 \cup (1, 0, 1) + \Omega_2 \cup (0, 1, 1) + \Omega_1$ muni de la mesure de Lebesgue divisée par 4. Dessiner Ω_2 . Montrer que l'on a $\mathbb{P}(X \in [a, b]) = b - a$.

3) En déduire que X, Y, Z sont encore des variables aléatoires indépendantes deux à deux. En considérant l'évènement $[X \geq 1, Y \geq 1, Z \geq 1]$ montrer que X, Y et Z ne sont pas indépendantes.

4) Version discrète du même exemple : $\Omega = \{0, 1\}^3$ d'éléments (x, y, z) avec $x, y, z \in \{0, 1\}$, muni de la probabilité vérifiant $\mathbb{P}((0, 0, 0)) = \mathbb{P}((1, 1, 0)) = \mathbb{P}((0, 1, 1)) = \mathbb{P}((1, 0, 1)) = \frac{1}{4}$ et tous les autres atomes de probabilité nulle. Montrer que les évènements $X = 1, Y = 1$ et $Z = 1$ sont indépendants deux à deux mais pas dans leur ensemble.



FIG. 5.1 – L'ensemble Ω_2 . Les coordonnées ne sont pas indépendantes sur Ω_2 mais le sont deux à deux.

Chapitre 6

Lois des grands nombres

Dans ce chapitre, on donne plusieurs variantes de la loi des grands nombres parmi les plus classiques et les plus simples. On joint une photocopie de [5], pp. 66-67, avec une preuve synthétique en deux pages de la loi forte des grands nombres.

Loi faible des grands nombres

Le premier théorème important et très simple à montrer est la loi faible des grands nombres. On rappelle d'abord l'inégalité de Markov

$$\mathbb{P}(|f(X)| \geq a) \leq \frac{\mathbb{E}|f(X)|}{a}$$

dont on déduit l'inégalité de Tchebychev, pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\sigma^2(X)}{\varepsilon^2}.$$

La preuve consiste à appliquer l'inégalité de Markov à $f(X) = |X - \mathbb{E}X|^2$.

Théorème 6.0.2. *Soit une suite de variables aléatoires indépendantes X_i d'espérance $\mathbb{E}X$, $i = 1, \dots, n, \dots$ et $S_n := \sum_{i=1}^n X_i$. Alors $\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}X| \geq \varepsilon) \rightarrow 0$ quand $n \rightarrow \infty$.*

En effet l'espérance de S_n est $n\mathbb{E}X$ et la variance de S_n est $n\sigma^2(X)$. Donc la moyenne de $\frac{S_n}{n}$ est $\mathbb{E}X$ et sa variance est $\sigma^2 = \frac{\sigma^2(X)}{n}$. Appliquons l'inégalité de Tchebychev. On obtient

$$\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}X| \geq \varepsilon) \leq \frac{\sigma^2(X)}{n\varepsilon^2} \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

La loi forte

Le second théorème à retenir est le suivant :

Théorème 6.0.3. *Soit X_1, X_2, \dots des variables aléatoires indépendantes identiquement distribuées et intégrables d'espérance μ . Soit $(S_n)_n$ la suite des sommes partielles des $(X_i)_i$. Alors*

$$n^{-1}S_n \longrightarrow \mu \quad \mathbb{P} \text{ p.s. et dans } \mathbb{L}^1$$

Exercice 6.0.8. Lemme de Kronecker ([30] p. 117)

Soit $(b_n)_n$ une suite de réels strictement positifs croissante et tendant vers l'infini. Soit $(x_n)_n$ une suite de réels et définissons $s_n = x_1 + \dots + x_n$. On a alors :

$$\sum_n \frac{x_n}{b_n} \text{ converge} \Rightarrow \frac{s_n}{b_n} \rightarrow 0.$$

1) Montrer d'abord le lemme de Cesàro : si b_n est comme ci-dessus et $v_n \rightarrow v_\infty$, alors (en posant $b_0 = 0$) :

$$\frac{1}{b_n} \sum_1^n (b_k - b_{k-1})v_k \rightarrow v_\infty.$$

(Marche à suivre : fixer $\varepsilon > 0$ et N tel que $v_k \geq v_\infty - \varepsilon$ pour $k \geq N$. Montrer en coupant la somme en deux que $\liminf_n \frac{1}{b_n} \sum_1^n (b_k - b_{k-1})v_k \geq 0 + v_\infty - \varepsilon$.)

2) Démonstration du Lemme de Kronecker : soit $u_n := \sum_{k \leq n} (\frac{x_k}{b_k})$. Par hypothèse, $u_\infty = \lim u_n$ existe et on a $u_n - u_{n-1} = \frac{x_n}{b_n}$. Ecrire $s_n = \sum_1^n b_k(u_k - u_{k-1})$ et déduire du lemme de Césaro que $\frac{s_n}{b_n} \rightarrow 0$.

Exercice 6.0.9. Très facile!! Loi faible des grands nombres avec moment d'ordre 4 borné, ([30] p. 72)
Soit X_1, X_2, \dots des variables aléatoires indépendantes telles qu'il existe une constante $K > 0$ vérifiant :

$$\mathbb{E}(X_k) = 0, \quad \mathbb{E}(X_k^4) \leq K$$

Soit $S_n = X_1 + \dots + X_n$. Alors

$$\mathbb{P}(n^{-1}S_n \rightarrow 0) = 1, \quad \lim_{n \rightarrow \infty} \frac{S_n}{n} = 0 \text{ dans } \mathbb{L}^4$$

1) Vérifier que $\mathbb{E}X_i^2$ et $\mathbb{E}X_i^3$ ont un sens. Montrer que

$$E(S_n^4) = E \left(\sum_k X_k^4 + 6 \sum_{i < j} X_i^2 X_j^2 \right)$$

En utilisant $(E(X_i^2))^2 \leq EX_i^4$ (le montrer), en déduire que

$$E(S_n^4) \leq nK + 3n(n-1)K \leq 3Kn^2.$$

2) Montrer que $E(\sum_n (\frac{S_n}{n})^4) \leq 3K \sum n^{-2}$ et conclure.

Exercice 6.0.10. Très général et pratique : la loi forte des grands nombres avec v.a. non-corrélées, pas nécessairement indépendantes, ([6], p. 228).

Soit X_1, X_2, \dots des variables aléatoires identiquement distribuées. Supposons que leur moyenne $\mu = \mathbb{E}(X_1)$ soit définie et qu'elles aient une variance finie σ^2 . Supposons de plus qu'elles soient non corrélées, c'est-à-dire :

$$i \neq j \implies \mathbb{E}[(X_i - \mu)(X_j - \mu)] = 0$$

Soit $S_n = X_1 + \dots + X_n$. Alors

$$\mathbb{P}(n^{-1}S_n \rightarrow \mu) = 1$$

1) Montrer que l'on peut se ramener au cas $\mu = 0$.

2) On pose $Z_m = \sup_{1 \leq k \leq 2m+1} |X_{m^2+1} + \dots + X_{m^2+k}|$. Montrer qu'il suffit de montrer

$$\mathbb{P} \text{ p.s.}, \quad \lim_{m \rightarrow +\infty} \frac{S_{m^2}}{m^2} = 0, \tag{6.0.1}$$

$$\mathbb{P} \text{ p.s.}, \quad \lim_{m \rightarrow +\infty} \frac{Z_m}{m^2} = 0 \tag{6.0.2}$$

On considérera l'entier $m(n)$ tel que $m(n)^2 \leq n \leq (m(n) + 1)^2$ et on remarquera que

$$\left| \frac{S_n}{n} \right| \leq \left| \frac{S_{m(n)^2}}{m(n)^2} \right| + \left| \frac{Z_{m(n)}}{m(n)^2} \right|.$$

3) Montrer que $\mathbb{P}(|S_{m^2}/m^2| \geq \epsilon) \leq \frac{\sigma^2}{m^2\epsilon^2}$. En déduire, en utilisant le lemme de Borel-Cantelli, l'assertion (6.0.1).

4) On définit $\zeta_k^{(m)} = X_{m^2+1} + \dots + X_{m^2+k}$. Montrer que

$$\mathbb{P}\left(\frac{Z_m}{m^2} \geq \epsilon\right) \leq \sum_{k=1}^{2m+1} \mathbb{P}(|\zeta_k^{(m)}| \geq m^2\epsilon).$$

Montrer que $\mathbb{P}(|\zeta_k^{(m)}| \geq m^2\epsilon) \leq \frac{(2m+1)\sigma^2}{m^4\epsilon^2}$. En utilisant le lemme de Borel-Cantelli, en déduire l'assertion (6.0.2) et conclure.

Remarque 6.0.2. Dans [14], p. 36, on trouve un théorème de lgn pour des variables non corrélées qui ressemble à celui-ci. Il n'y est cependant prouvé que la convergence en probabilité.

Exercice 6.0.11. Inégalité de Kolmogorov, ([16], tome 1, p 234-235)

On reverra cette inégalité quand on traitera les martingales. Soit X_1, \dots, X_n des variables aléatoires indépendantes centrées. On pose $S_n = X_1 + \dots + X_n$ et s_k^2 la variance de S_k . On a l'inégalité suivante :

$$\mathbb{P}(\exists k \in \{1 \dots n\}, |S_k| \geq ts_n) \leq t^{-2}.$$

On pose pour $\nu \in \{1 \dots n\}$, $Y_\nu = 1_{|S_\nu| \geq ts_n} 1_{\{|S_k| < ts_n \text{ pour } k=1, \dots, \nu-1\}}$.

1) Montrer que $Y_1 + \dots + Y_n = 1_{\{\exists k \in \{1 \dots n\}, |S_k| \geq ts_n\}}$.

2) Montrer que $\sum_{k=1}^n \mathbb{E}(Y_k S_n^2) \leq s_n^2$.

3) On pose $U_k = S_n - S_k$. En remarquant que U_k est indépendant de $Y_k S_k$, montrer que

$$\mathbb{E}(Y_k S_n^2) \geq \mathbb{E}(Y_k S_k^2)$$

4) Montrer que $Y_k S_k^2 \geq t^2 s_n^2 Y_k$ et déduire de 2) et 3) le résultat annoncé.

Solution

1) En effet, les évènements $Y_\nu = 1$ sont disjoints et leur union est bien l'évènement qui nous intéresse.

2) En effet, $\sum_{k=1}^n Y_k \leq 1$. Il suffit de multiplier par S_n^2 et de prendre l'espérance.

3) $\mathbb{E}(Y_k S_n^2) = \mathbb{E}(Y_k (S_k + U_k)^2) = \mathbb{E}(Y_k S_k^2) + 2\mathbb{E}(Y_k S_k U_k) + \mathbb{E}(Y_k U_k^2)$. Le deuxième terme est nul par l'indépendance mentionnée et le troisième terme est positif.

4) L'inégalité demandée est une conséquence immédiate de la définition de Y_k . En utilisant successivement cette dernière inégalité, puis celle de 3) et finalement celle de 2), on a

$$\begin{aligned} \mathbb{P}(\exists k \in \{1, \dots, n\} |S_k| \geq ts_n) &= \mathbb{E}(Y_1 + \dots + Y_n) \leq \frac{1}{t^2 s_n^2} \mathbb{E}\left(\sum_k Y_k S_k^2\right) \\ &\leq \frac{1}{t^2 s_n^2} \mathbb{E}\left(\sum_{k=1}^n Y_k S_n^2\right) \leq t^{-2}. \end{aligned}$$

Exercice 6.0.12. Loi forte des grands nombres avec une contrainte sur la variance ([16], tome 2).

Soit $(W_n)_{n \in \mathbb{N}}$ une suite de variables indépendantes centrées ayant des variances $\sigma_n^2 = \text{Var}(W_n)$ telles que :

$$\sum_n \frac{\sigma_n^2}{n^2} < +\infty$$

Alors $\mathbb{P} p.s. \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n W_k = 0$

1) Démontrer le théorème de Kronecker.

2) On pose $X_n = W_n/n$ qui sont centrées et telles que :

$$\sum_n \text{Var}(X_n) < +\infty$$

et $S_n = \sum_{k=1}^n X_k$. On va montrer qu'avec probabilité 1, $\sum_n X_n$ converge.

2a) Montrer qu'il suffit de prouver que pour tout $\epsilon > 0$,

$$\lim_{n_0 \rightarrow +\infty} \mathbb{P}(\exists m, n \geq n_0, |S_m - S_n| > \epsilon) = 0$$

2b) Montrer que

$$\lim_{m \rightarrow +\infty} \mathbb{P}(\exists k \in \{n_0 \dots m\}, |S_{n_0} - S_k| > \epsilon/2) \geq \mathbb{P}(\exists m, n \geq n_0, |S_m - S_n| > \epsilon)$$

3) Utiliser l'inégalité de Kolmogorov et conclure.

Solution

2a) Supposons que l'on ait démontré que pour tout $\epsilon > 0$

$$\lim_{n_0 \rightarrow +\infty} \mathbb{P}(\exists m, n \geq n_0, |S_m - S_n| > \epsilon) = 0$$

On a :

$$\mathbb{P}(\{\exists \epsilon > 0, \forall n_0, \exists m, n \geq n_0, |S_m - S_n| > \epsilon\}) \leq \sum_{\epsilon \in \mathbb{Q}^+} \mathbb{P}(\{\forall n_0, \exists m, n \geq n_0, |S_m - S_n| > \epsilon\})$$

et il suffit donc de montrer que pour $\epsilon > 0$ fixé,

$$\mathbb{P}(\{\forall n_0, \exists m, n \geq n_0, |S_m - S_n| > \epsilon\}) = 0$$

Ce dernier terme s'écrit $\mathbb{P}(\bigcap_{n_0} A_{n_0})$ où les $A_{n_0} = \{\exists m, n \geq n_0, |S_m - S_n| > \epsilon\}$ sont décroissants. D'où :

$$\mathbb{P}(\{\forall n_0, \exists m, n \geq n_0, |S_m - S_n| > \epsilon\}) = \lim_{n_0 \rightarrow +\infty} \mathbb{P}(A_{n_0}) = 0 \text{ par hypothèse.}$$

2b) On remarque aisément en utilisant l'inégalité triangulaire que :

$$A_{n_0} \subset \bigcup_{m \geq n_0} \{|S_m - S_{n_0}| > \epsilon/2\}$$

Or

$$\bigcup_{m \geq n_0} \{|S_m - S_{n_0}| > \epsilon/2\} = \bigcup_{m \geq n_0} \bigcup_{m \geq k \geq n_0} \{|S_k - S_{n_0}| > \epsilon/2\}$$

Et les événements $B_m = \bigcup_{m \geq k \geq n_0} \{|S_k - S_{n_0}| > \epsilon/2\}$ sont croissants. D'où 2b).

3) En appliquant soigneusement l'inégalité de Kolmogorov, on a :

$$\mathbb{P}(\exists k \in \{n_0 \dots m\}, |S_{n_0} - S_k| > \epsilon/2) \leq \frac{4\text{Var}(X_{n_0+1} + \dots + X_m)}{\epsilon^2}$$

Le terme de droite s'écrit en raison de l'indépendance des X_k ,

$$\frac{4}{\epsilon^2} \sum_{k=n_0+1}^m \text{Var}(X_k)$$

Prenant la limite quand m tend vers l'infini puis la limite quand n_0 tend vers l'infini, le résultat tombe.

Exercice 6.0.13. Lemme de troncature de Kolmogorov ([30], 12.9 p. 118)

Soit X_1, X_2, \dots des variables aléatoires indépendantes identiquement distribuées ayant la même loi que X et intégrables. Soit $\mu = \mathbb{E}(X)$. On définit :

$$Y_n = X_n \mathbb{1}_{|X_n| \leq n}$$

Alors :

i) $\mathbb{E}(Y_n) \rightarrow \mu$

ii) $\mathbb{P}(X_n = Y_n \text{ sauf pour un nombre fini de } n) = 1$

iii) $\sum_n (\text{Var}(Y_n)/n^2) < +\infty$

On pose $Z_n = X \mathbb{1}_{|X| \leq n}$. Montrer que la loi de Z_n est la même que celle de Y_n .

1) Démontrer i).

2) Montrer que $\sum_n \mathbb{P}(X_n \neq Y_n) < +\infty$ et en déduire ii).

3) Montrer que $\sum_n (\text{Var}(Y_n)/n^2) \leq \mathbb{E}(|X|^2 f(|X|))$ où $f(z) = \sum_{n \geq \sup(1, z)} 1/n^2$. Prouver que $f(z) \leq 2/\sup(1, z)$ et en déduire iii).

Solution

1) On pose $Z_n = \mathbb{1}_{\{|X| \leq n\}} X$ qui est majorée en valeur absolue par $|X|$ intégrable et qui tend \mathbb{P} presque sûrement vers X . Le théorème de convergence dominée donne le résultat.

2) $\sum_n \mathbb{P}(Y_n \neq X_n) = \sum_n \mathbb{P}(|X_n| > n) = \sum_n \mathbb{P}(|X| > n)$. Or on a :

$$\sum_n \mathbb{P}(|X| > n) = \mathbb{E} \left(\sum_n \mathbb{1}_{\{|X| > n\}} \right) = \mathbb{E} \left(\sum_{n < |X|} 1 \right) \leq \mathbb{E}(|X|) < +\infty$$

Par le lemme de Borel-Cantelli 1, on obtient ii).

3) Il suffit d'écrire que

$$\sum_n \frac{\text{Var}(Y_n)}{n^2} \leq \sum_n \frac{\mathbb{E}(|X|^2; |X| < n)}{n^2} = \mathbb{E}(|X|^2 f(|X|))$$

L'inégalité concernant f se montre facilement et on obtient donc

$$\sum_n \frac{\text{Var}(Y_n)}{n^2} \leq 2\mathbb{E}(|X|)$$

Exercice 6.0.14. Loi forte des grands nombres : démonstration via les troncatures ([30])

Soit X_1, X_2, \dots des variables aléatoires indépendantes identiquement distribuées et intégrables d'espérance μ . Alors

$$\mathbb{P}(n^{-1}S_n \rightarrow \mu) = 1$$

Cette démonstration se fait au moyen de deux ingrédients : le lemme de troncature de Kolmogorov et une loi forte sous contrainte de variance. Démontrer le théorème.

Solution

On pose $Y_n = 1_{\{|X_n| \leq n\}} X_n$. D'après le ii) du lemme de troncature, $Y_n = X_n$ sauf pour un nombre fini de n d'où, \mathbb{P} presque sûrement,

$$\lim_{n \rightarrow \infty} \left\{ \frac{1}{n} \sum_{k=1}^n Y_k - \frac{1}{n} \sum_{k=1}^n X_k \right\} = 0$$

Il nous reste donc à montrer que \mathbb{P} presque sûrement

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Y_k = \mu$$

Or

$$\frac{1}{n} \sum_{k=1}^n Y_k = \frac{1}{n} \sum_{k=1}^n \mathbb{E}(Y_k) + \frac{1}{n} \sum_{k=1}^n W_k$$

où $W_k = Y_k - \mathbb{E}(Y_k)$.

Mais par le i) du lemme de troncature de Kolmogorov,

$$\mathbb{E}(Y_k) \rightarrow \mu$$

et par le théorème de Césaro,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E}(Y_k) = \mu$$

Quant au deuxième terme, $\frac{1}{n} \sum_{k=1}^n W_k$, il se traite grâce à la loi forte des grand nombres avec une contrainte sur la variance. En effet, les W_k sont i.i.d. centrées et leurs variances sont données par $Var(W_k) = Var(Y_k)$. Par iii) du lemme de troncature de Kolmogorov et la loi des grands nombres avec contrainte sur la variance, on conclut.

Exercice 6.0.15. La démonstration d'Ettemadi (1981). Nous donnons cette démonstration sous la forme élégante proposée par [5], p. 66-67).

1) Lire attentivement les sept inégalités et égalités de la page 66 et les justifier.

2) Lire attentivement. Remarquer que le haut de la page 67 est le lemme de troncature de Kolmogorov.

Exercice 6.0.16. On rappelle le théorème de Sheffé : *Si une suite de variables aléatoires $(X_n)_n \in \mathbb{L}^1$ vérifie $X_n \geq 0$, $X_n \rightarrow X$ presque sûrement, $X \in \mathbb{L}^1$ et $\mathbb{E}(X_n) \rightarrow \mathbb{E}(X)$, alors $X_n \rightarrow X$ dans \mathbb{L}^1 .*

Soit $(X_n)_n$ une suite i.i.d. de variables aléatoires intégrables. Montrer que $(S_n)_n = (n^{-1} \sum_{k=1}^n X_k)_n$ converge dans \mathbb{L}^1 vers $\mathbb{E}(X_1)$.

Solution

Si Z est une variable aléatoire, on note $Z^+ = \sup(Z, 0)$ sa partie positive et $Z^- = \sup(-Z, 0)$ sa partie négative si bien que $Z = Z^+ - Z^-$. On considère les variables aléatoires i.i.d. intégrables $(X_n^+)_n$ d'espérance $\mathbb{E}(X_1^+)$ et on forme la suite $(S_{n,+})_n$ définie par

$$S_{n,+} = n^{-1} (X_1^+ + \dots + X_n^+)$$

On sait que \mathbb{P} p.s., la suite $(S_{n,+})_n$ converge vers $\mathbb{E}(X_1^+)$ et d'autre part, on a $\mathbb{E}(S_{n,+}) = \mathbb{E}(X_1^+)$ donc par le théorème de Sheffé, on a que la convergence de $(S_{n,+})_n$ vers la constante $\mathbb{E}(X_1^+)$ a lieu dans \mathbb{L}^1 . De même, en considérant $(X_n^-)_n$, et $S_{n,-} = n^{-1} (X_1^- + \dots + X_n^-)$, on a la convergence de $(S_{n,-})_n$ vers $\mathbb{E}(X_1^-)$. Vu que $S_n = S_{n,+} - S_{n,-}$ pour tout n , on en déduit immédiatement le résultat.

Remarque 6.0.3. 1. Nous n'avons pas énoncé le théorème des trois séries de Kolmogorov, que l'on pourra trouver dans [30] par exemple. Il est d'usage d'en parler lorsqu'on traite la loi des grands nombres mais il en est indépendant (même si les techniques sont les mêmes).

2. On pourra consulter [16], tome 1, pp. 243-263, et surtout les exercices pour des généralisations de la loi des grands nombres. Il sera aussi bon de s'intéresser au théorème ergodique qui est une généralisation de la loi des grands nombres pour les processus stationnaires ([18]).

3. La loi des grands nombres est énormément utilisée en statistique (pour construire des estimateurs par exemple), d'un point de vue numérique pour estimer des intégrales (méthode de Monte Carlo)...

6.1 La preuve la plus rapide de la loi forte des grands nombres

(V. S. Borkar, Probability Theory, Springer, 1995, pages 66-67).

La loi des grands nombres est justifiée par Vivek Borkar de la manière suivante : "Observez que des variables aléatoires de carré intégrable et de moyenne zéro sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ forment un espace vectoriel avec produit scalaire $\langle X, Y \rangle := \mathbb{E}[XY]$. Mais si on ajoute des variables aléatoires indépendantes et identiquement distribuées dans cet espace, c'est la même chose que d'ajouter n vecteurs orthogonaux de même longueur. Donc la moyenne arithmétique de ces vecteurs doit être d'ordre $n^{-\frac{1}{2}}$." C'est exactement la ligne de raisonnement que suit la loi faible des grands nombres, qui suppose les variables aléatoires de variance bornée. La loi forte précise la loi faible. Sa démonstration très détaillée suivant la ligne de Kolmogorov consiste à tronquer les variables X_n pour que leurs variances soient contrôlées et que l'on puisse appliquer un argument hilbertien du même type que pour la loi faible. Ensuite, on montre que la différence entre X_n et sa troncature Y_n est asymptotiquement négligeable par le lemme de Borel-Cantelli. La preuve d'Etemadi suit toujours cette démarche, mais de manière particulièrement heureuse, en peu de lignes.

Théorème 6.1.1. Soient X_n , $n \geq 1$, intégrables, indépendantes deux à deux et identiquement distribuées et $S_n := \sum_{i=1}^n X_i$, $n \geq 1$. Alors

$$\frac{S_n}{n} \rightarrow EX_1 \text{ p.s..}$$

Démonstration. Cette preuve est due à Etemadi. Comme X_n^+ et X_n^- vérifient les mêmes hypothèses du théorème on montre tout avec $X_i \geq 0$. Soit $Y_i := X_i \mathbb{1}_{X_i \leq i}$ et $S_n^* := \sum_{i=1}^n Y_i$, $n \geq 1$. Soient $\varepsilon > 0$, $\alpha > 1$ et $k_n := [\alpha^n]$ la partie entière de α^n . Dans tout ce qui suit C désigne une constante positive variant d'étape en étape. Soit $\mu := \mathbb{P}_{X_i}$ la loi de X_i . On note $\sigma^2(X)$ la variance de X . On a alors

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P} \left(\left| \frac{S_{k_n}^* - \mathbb{E}[S_{k_n}^*]}{k_n} \right| \right) &\leq C \sum_{n=1}^{\infty} \frac{\sigma^2(S_{k_n}^*)}{k_n^2} \\ &= C \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{i=1}^{k_n} \sigma^2(Y_i) \\ &\leq C \sum_{i=1}^{\infty} \frac{\mathbb{E}[Y_i^2]}{i^2} \\ &= C \sum_{i=1}^{\infty} \frac{1}{i^2} \int_0^i x^2 \mu(dx) \\ &= C \sum_{i=1}^{\infty} \frac{1}{i^2} \sum_{k=0}^{i-1} \int_k^{k+1} x^2 \mu(dx) \\ &\leq C \sum_{k=0}^{\infty} \frac{1}{k+1} \int_k^{k+1} x^2 \mu(dx) \\ &\leq C \sum_{k=0}^{\infty} \int_k^{k+1} x \mu(dx) = CEX_1 < \infty. \end{aligned}$$

On a aussi

$$\mathbb{E}X_1 = \lim_{n \rightarrow \infty} \int_0^n x \mu(dx) = \lim_{n \rightarrow \infty} \mathbb{E}Y_n = \lim_{n \rightarrow \infty} \frac{\mathbb{E}S_{k_n}^*}{k_n}.$$

Donc par le lemme de Borel-Cantelli,

$$\frac{S_{k_n}^*}{k_n} \rightarrow EX_i \text{ p.s.}$$

Mais

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(Y_n \neq X_n) &= \sum_{n=1}^{\infty} \mathbb{P}(X_n > n) \\ &= \sum_{n=1}^{\infty} \int_n^{\infty} \mu(dx) \\ &= \sum_{n=1}^{\infty} \sum_{i=n}^{\infty} \int_i^{i+1} \mu(dx) \\ &= \sum_{i=1}^{\infty} i \int_i^{i+1} \mu(dx) \\ &\leq EX_1 < \infty. \end{aligned}$$

Par Borel-Cantelli, $\mathbb{P}(X_n \neq Y_n \text{ i.o.}) = 0$. Donc

$$\lim_{n \rightarrow \infty} \frac{S_{k_n}}{k_n} = EX_1 \text{ p.s.}$$

Pour $n \geq 1$ soit $m(n) \geq 0$ tel que $k_{m(n)} \leq n \leq k_{m(n)+1}$. Comme $n \rightarrow S_n$ est croissante,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{S_n}{n} &\geq \liminf_{n \rightarrow \infty} \frac{S_{k_{m(n)}}}{k_{m(n)}} \frac{k_{m(n)}}{k_{m(n)} + 1} \\ &\geq \frac{1}{\alpha} \lim_{n \rightarrow \infty} \frac{S_{k_{m(n)}}}{k_{m(n)}} = \frac{1}{\alpha} EX_1 < \infty. \end{aligned}$$

De même,

$$\limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq \alpha EX_1 \text{ p.s..}$$

Comme $\alpha > 1$ est arbitraire on conclut en faisant $\alpha \rightarrow 1$. □

Chapitre 7

Les convergences

Dans ce chapitre, les mesures considérées sont des mesures sur \mathbb{R}^d mais tout ceci reste valable si on remplace \mathbb{R}^d par un espace métrique localement compact dénombrable à l'infini.

7.1 Convergence des mesures bornées sur \mathbb{R}^d

On commence par donner quelques notations utilisées tout au long du chapitre.

Ensembles de mesures (sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$) :

\mathcal{M} : ensemble des mesures positives de masse finie.

\mathcal{M}_b : ensemble des mesures de masse positive inférieure ou égale à b .

\mathcal{M}^1 : ensemble des probabilités.

Ensembles de fonctions (sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$) :

C_K : ensemble des fonctions continues à support compact.

C_0 : ensemble des fonctions continues tendant vers 0 à l'infini.

C_b : ensemble des fonctions continues bornées.

Remarque 7.1.1. C_0, C_K munis de la norme uniforme sont séparables mais C_b ne l'est pas. $(C_b, \|\cdot\|_\infty)$ est un espace de Banach dont C_0 est un sev fermé (mais pas C_K) et C_K est dense dans C_0 (mais pas dans C_b).

On rappelle que l'on dit qu'un sous-ensemble H d'un espace vectoriel E est *total* ssi $Vect(H)$ est dense dans E .

7.1.1 Topologies vague, faible et étroite

Les topologies vague, faible et étroite sont respectivement les topologies les moins fines sur \mathcal{M} (celles qui contiennent le moins d'ouverts) rendant continues les applications $\mu \rightarrow \int f d\mu$ (avec $f \in C_K, C_0$ et C_b respectivement). En particulier, $(\mu_n)_n$ converge vers μ :

- vaguement si $\forall f \in C_K, \lim_{n \rightarrow \infty} \mu_n(f) = \mu(f)$
- faiblement si $\forall f \in C_0, \lim_{n \rightarrow \infty} \mu_n(f) = \mu(f)$
- étroitement si $\forall f \in C_b, \lim_{n \rightarrow \infty} \mu_n(f) = \mu(f)$

On rappelle qu'une base de voisinages de μ pour ces topologies est composée des ensembles

$$V_{\varepsilon, f_1 \dots f_n}^\mu = \{\nu \in \mathcal{M}, \forall i \in \mathbb{N}_n, |\mu(f_i) - \nu(f_i)| \leq \varepsilon\} = \bigcap_{i=1}^n V_{\varepsilon, f_i}^\mu$$

où $\varepsilon > 0$ et f_1, \dots, f_n appartiennent à l'espace considéré.

La topologie la plus fine est l'étroite, puis vient la faible et enfin la vague. Ainsi, une suite convergeant étroitement converge faiblement et une suite convergeant faiblement converge vaguement.

Proposition 7.1.1. *On a les propriétés suivantes :*

- 1) Sur \mathcal{M}_b , les topologies vagues et faibles coïncident avec la topologie la moins fine rendant continues les applications $\mu \rightarrow \int f d\mu$ lorsque f parcourt un ensemble total H de C_K (resp. C_0).
- 2) Sur \mathcal{M}^1 , les trois topologies coïncident.
- 3) \mathcal{M}_b est métrisable compact pour la topologie faible.

Remarque 7.1.2. 1) n'a d'intérêt que si H n'est pas trop gros, typiquement dénombrable, ce qui permet de rendre \mathcal{M}_b métrisable. En fait, un résultat analogue à celui-ci pour la convergence étroite n'aurait pas d'intérêt car un ensemble H total dans C_b n'est pas dénombrable (et même très gros).

Démonstration. 1) On traite uniquement le cas de la convergence vague, le cas de la convergence faible étant semblable. Soit H un ensemble total de C_K . On peut toujours supposer que H est un espace vectoriel quitte à considérer $Vect(H)$. Soit alors $f \in C_K$ et $g \in H$ telle que $\|f - g\|_\infty \leq \varepsilon$. Si $\nu, \mu \in \mathcal{M}_b$, on a

$$\begin{aligned} |\mu(f) - \nu(f)| &\leq |\mu(f) - \mu(g)| + |\mu(g) - \nu(g)| + |\nu(g) - \nu(f)| \\ &\leq 2b\varepsilon + |\nu(g) - \mu(g)| \end{aligned}$$

et par suite $V_{\varepsilon, g}^\mu \subset V_{(2b+1)\varepsilon, f}^\mu$ donc puisque la topologie vague est celle qui contient le moins d'ouverts, les deux topologies coïncident.

2) Il nous suffit donc de montrer maintenant que sur \mathcal{M}^1 , les topologies vagues et étroites coïncident. Il suffit pour cela de montrer que $\forall \mathbb{P} \in \mathcal{M}^1, \forall \varepsilon > 0, \forall f \in C_b, V_{\varepsilon, f}^\mathbb{P}$ contient un voisinage pour la topologie vague. Soit $(h_p)_p$ une suite de fonctions positives de C_K qui convergent vers 1 simplement et en croissant. Pour toute probabilité \mathbb{Q} , on a

$$\begin{aligned} |\mathbb{P}(f) - \mathbb{Q}(f)| &\leq |\mathbb{P}(f - fh_p)| + |\mathbb{P}(fh_p) - \mathbb{Q}(fh_p)| + |\mathbb{Q}(fh_p) - \mathbb{Q}(f)| \\ &\leq \|f\|_\infty \{\mathbb{Q}(1 - h_p) + \mathbb{P}(1 - h_p)\} + |\mathbb{P}(fh_p) - \mathbb{Q}(fh_p)| \\ &= \|f\|_\infty \{2 - \mathbb{Q}(h_p) - \mathbb{P}(h_p)\} + |\mathbb{P}(fh_p) - \mathbb{Q}(fh_p)| \\ &\leq \|f\|_\infty \{2\mathbb{P}(1 - h_p) + \mathbb{P}(h_p) - \mathbb{Q}(h_p)\} + |\mathbb{P}(fh_p) - \mathbb{Q}(fh_p)| \end{aligned}$$

Remarquons maintenant que par le théorème de la convergence monotone, $\lim_{p \rightarrow \infty} \mathbb{P}(1 - h_p) = 0$. On fixe alors p tel que $2\mathbb{P}(1 - h_p) \leq \varepsilon$. Il est aisé de voir que l'inégalité précédente ne dit rien d'autre que

$$\mathcal{M}^1 \cap V_{\varepsilon, h_p}^\mathbb{P} \cap V_{\varepsilon, fh_p}^\mathbb{P} \subset \mathcal{M}^1 \cap V_{(2\varepsilon\|f\|_\infty + \varepsilon), f}^\mathbb{P}$$

autrement dit, que tout voisinage de la topologie étroite sur \mathcal{M}^1 contient un voisinage ouvert de la topologie vague sur \mathcal{M}^1 , c.q.f.d.

3) On prend $(f_n)_n$ une suite d'éléments de C_K dense dans C_0 . On définit une distance d sur \mathcal{M}_b par

$$d(\mu, \nu) = \sum_{n \geq 1} \frac{1}{2^n \|f\|_\infty} |\mu(f_n) - \nu(f_n)|$$

Il est aisé de montrer que c'est bien une distance. D'après 1), la topologie induite par d coïncide avec la topologie faible.

Montrons maintenant que \mathcal{M}_b est faiblement compact. On peut se restreindre à montrer que \mathcal{M}_b est séquentiellement compact (car métrisable). Soit donc $(\mu_p)_p \in \mathcal{M}_b^{\mathbb{N}}$. On veut en extraire une sous-suite faiblement convergente. Or

$$\forall n \in \mathbb{N}, |\mu_p(f_n)| \leq b \|f_n\|_{\infty}$$

Par un procédé d'extraction diagonale, on peut extraire une sous-suite $(\mu_{p_k})_k$ telle que

$$\lim_{k \rightarrow \infty} \mu_{p_k}(f_n) = \Psi(f_n), \quad n \in \mathbb{N}$$

Par densité, il existe pour tout f de C_K , un réel $\Psi(f)$ tel que

$$\mu_{p_k}(f) \xrightarrow{k \rightarrow \infty} \Psi(f)$$

Le théorème de Riesz assure l'existence d'une unique mesure μ telle que $\mu(f) = \Psi(f)$ pour tout $f \in C_K$. (Attention, [26] ne détaille pas ce passage : il passe à C_0 directement). Il reste à montrer que $\mu \in \mathcal{M}_b$. On prend alors une suite $h_p \in C_K$, $0 \leq h_p \leq 1$ tendant simplement en croissant vers 1. On a pour tout k , $\mu_{p_k}(h_p) \leq b$ donc $\mu(h_p) \leq b$ donc $\mu(\mathbb{R}^d) \leq b$. Ceci conclut la preuve. \square

Notation : On notera le fait que $(\mu_n)_n$ converge étroitement vers μ par $\mu_n \Rightarrow \mu$.

Remarque 7.1.3. Si $(\mu_n)_n$ est une suite de probabilités, pour que $(\mu_n)_n$ converge étroitement vers une mesure $\mu \in \mathcal{M}$, il suffit que $(\mu_n(f))_n$ converge vers $\mu(f)$ pour f appartenant à un ensemble total H de C_K . Mais si $(\mu_n)_n$ est une suite de probabilités telle que pour tout $f \in C_K$, $\mu_n(f) \rightarrow_{n \rightarrow \infty} \Psi(f)$ alors $(\mu_n)_n$ converge faiblement vers une mesure μ de masse inférieure ou égale à 1 mais ce n'est pas forcément une probabilité ("perte de masse"). La propriété suivante précise un peu ce phénomène.

Proposition 7.1.2. Soit $(\mu_n)_n \in \mathcal{M}_b^{\mathbb{N}}$ et $\mu \in \mathcal{M}_b$ telles que $(\mu_n)_n$ converge vers μ vaguement (ou faiblement puisque les deux notions coïncident sur \mathcal{M}_b). Alors $(\mu_n)_n$ converge étroitement vers μ ssi $\lim_{n \rightarrow \infty} \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d)$.

La preuve de cette propriété est facile et donc admise.

Remarque 7.1.4. Sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\delta_{1/n} \Rightarrow \delta_0$ mais $\delta_{1/n}(\{0\}) = 0$ ne converge pas vers $\delta_0(\{0\}) = 1$. Aussi, $\mu_n \Rightarrow \mu$ n'implique pas que $(\mu_n(A))_n$ converge vers $\mu(A)$ pour tout borélien A .

Définition 7.1.1. Soit μ une mesure sur \mathbb{R}^d . On dira qu'un borélien A est de μ -continuité ssi $\mu(\partial A) = 0$ (∂A désigne la frontière de A).

Proposition 7.1.3. Soit $(\mu_n)_n \in \mathcal{M}_b^{\mathbb{N}}$ et $\mu \in \mathcal{M}_b$. Il y a équivalence entre :

- a) $\mu_n \Rightarrow \mu$
- b) Pour tout fermé F , $\limsup \mu_n(F) \leq \mu(F)$ et $\lim_{n \rightarrow \infty} \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d)$.
- c) Pour tout ouvert O , $\liminf \mu_n(O) \geq \mu(O)$ et $\lim_{n \rightarrow \infty} \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d)$.
- d) Pour tout borélien A de μ -continuité, $\mu_n(A) \rightarrow_{n \rightarrow \infty} \mu(A)$.

Démonstration. Nous ne donnons ici que la démonstration de quelques implications. Pour plus de détails, se reporter à [26].

a) \Rightarrow b) : La deuxième partie de b) découle de la proposition précédente. Passons à la première. Soit $\phi_j : \mathbb{R} \rightarrow [0, 1]$ continue définie par $\phi_j(x) = 1$ si $x \leq 0$, $\phi_j(x) = 0$ si $x \geq 1/j$ et ϕ_j affine sur $[0, 1/j]$. On définit alors $f_j \in C_b$ par $f_j(x) = \phi_j(d(x, F))$ qui tend en décroissant vers la fonction caractéristique 1_F de F (car F fermé). D'après le théorème de convergence monotone décroissante,

$$\mu(F) = \lim_{j \rightarrow \infty} \mu(f_j)$$

donc pour tout $\varepsilon > 0$, il existe un entier j_0 tel que $\mu(F) \leq \mu(f_{j_0}) \leq \mu(F) + \varepsilon$.

Or pour tout $n \in \mathbb{N}$, $\mu_n(F) \leq \mu_n(f_{j_0})$ donc

$$\limsup \mu_n(F) \leq \limsup \mu_n(f_{j_0}) = \mu(f_{j_0}) \leq \mu(F) + \varepsilon$$

b) \iff c) : Passer au complémentaire.

c) \implies d) : On a $A^\circ \subset A \subset \bar{A}$ et on sait b) et c) car $b \iff c$. Donc on a :

$$\mu(A^\circ) \leq \liminf \mu_n(A^\circ) \leq \liminf \mu_n(A) \leq \limsup \mu_n(A) \leq \limsup \mu_n(\bar{A}) \leq \mu(\bar{A})$$

or $\mu(\partial A) = 0$ et par conséquent, toutes les inégalités sont en fait des égalités.

c) \implies a) On rappelle qu'une application classique du théorème de Fubini permet d'écrire que pour toute fonction $f \in C_b$ positive et pour toute mesure $\mu \in \mathcal{M}_b$,

$$\int_{\mathbb{R}^d} f(x) d\mu(x) = \int_0^{\|f\|_\infty} \mu(f \geq u) du$$

et

$$\int_{\mathbb{R}^d} f(x) d\mu(x) = \int_0^{\|f\|_\infty} \mu(f > u) du \quad (*)$$

Soit alors $f \in C_b$ positive. Par le théorème de Fatou, on a

$$\begin{aligned} \limsup_n \int_{\mathbb{R}^d} f(x) d\mu_n(x) &= \limsup_n \int_0^{\|f\|_\infty} \mu_n(f \geq u) du \\ &\leq \int_0^{\|f\|_\infty} \limsup_n \mu_n(f \geq u) du \end{aligned}$$

Or $\{f \geq u\}$ est fermé et on a supposé b) vraie (car $b \iff c$) donc on a

$$\limsup_n \int_{\mathbb{R}^d} f(x) d\mu_n(x) \leq \int_0^{\|f\|_\infty} \mu(f \geq u) du = \int_{\mathbb{R}^d} f d\mu$$

On a de même, en utilisant (*) et c), que

$$\liminf_n \int_{\mathbb{R}^d} f(x) d\mu_n(x) \geq \int_{\mathbb{R}^d} f d\mu$$

On a donc montré que $\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f(x) d\mu_n(x) = \int_{\mathbb{R}^d} f(x) d\mu(x)$ pour toute fonction continue bornée et positive. Par linéarité, on en déduit le résultat pour toute fonction f continue bornée de signe quelconque. \square

7.1.2 Tension

La propriété de tension est une propriété qui permet de limiter les phénomènes de perte de masse.

Définition 7.1.2. Une suite $(\mu_n)_n$ de mesures de \mathcal{M}_b est tendue ssi pour tout $\varepsilon > 0$, il existe un compact K de \mathbb{R}^d tel que

$$\sup_n \mu_n(K^c) \leq \varepsilon$$

Proposition 7.1.4. 1) Si $(\mu_n)_n \in \mathcal{M}_b$ converge étroitement vers $\mu \in \mathcal{M}_b$ alors $(\mu_n)_n$ est tendue.

2) Si $(\mu_n)_n$ est tendue et si $(\mu_n)_n \in (\mathcal{M}^1)^\mathbb{N}$ alors il existe une sous-suite $(\mu_{n_k})_k$ de $(\mu_n)_n$ et une probabilité $\mu \in \mathcal{M}^1$ telle que $\mu_{n_k} \Rightarrow \mu$.

Démonstration. 1) Soit $\varepsilon > 0$ fixé. Il existe une boule ouverte O telle que $\mu(O) \geq \mu(\mathbb{R}^d) - \varepsilon$. D'autre part, par la propriété précédente, $\liminf \mu_n(O) \geq \mu(O)$. Donc il existe un entier N_2 tel que pour $n \geq N_2$,

$$\mu_n(O) \geq \mu(\mathbb{R}^d) - \varepsilon$$

D'autre part, toujours par la propriété précédente, $\lim_{n \rightarrow \infty} \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d)$ donc il existe un rang $N_1 \geq N_2$ tel que pour $n \geq N_1$, $\mu_n(\mathbb{R}^d) \leq \mu(\mathbb{R}^d) + \varepsilon$.

On en déduit que $\forall n \geq N_1, \mu_n(\bar{O}^c) \leq \mu_n(O^c) \leq 2\epsilon$.

N_1 est maintenant fixé. Pour chaque $n \leq N_1$, il existe un compact K_n tel que $\mu_n(K_n^c) \leq \epsilon$. Donc si on pose $K = \cup_{n \leq N_1} K_n \cup \bar{O}$, K est un compact et pour tout n ,

$$\mu_n(K^c) \leq 2\epsilon$$

ce qui prouve la tension de la suite $(\mu_n)_n$.

2) Réciproquement, si $(\mu_n)_n$ est une suite de mesures tendue, on sait déjà qu'il existe une mesure μ de masse inférieure ou égale à 1 et une sous-suite $(\nu_n)_n$ de $(\mu_n)_n$ telle que $(\nu_n)_n$ converge faiblement vers μ . Pour montrer que $\nu_n \Rightarrow \mu$ et que $\mu \in \mathcal{M}_1$, il suffit de montrer que $\mu(\mathbb{R}^d) = 1$.

Or pour tout $\epsilon > 0$, il existe un compact K_ϵ tel que $\mu_n(K_\epsilon) \geq 1 - \epsilon$ pour tout n . Soit alors $f_\epsilon \in C_0$ une fonction telle que :

$$1 \geq f_\epsilon \geq 1_{K_\epsilon} \geq 0$$

On a $\mu_n(f_\epsilon) \geq \mu_n(K_\epsilon) \geq 1 - \epsilon$ et quand $n \rightarrow \infty$, on obtient donc $\mu(f_\epsilon) \geq 1 - \epsilon$. Puisque $\mu(\mathbb{R}^d) \geq \mu(f_\epsilon)$, en faisant maintenant tendre ϵ vers 0, il en résulte que $\mu(\mathbb{R}^d) \geq 1$ et puisque l'on sait déjà que l'on a l'inégalité inverse, μ est une probabilité. \square

7.2 Rappel sur les différents types de convergence

On rappelle les principales notions de convergence. Ici $(X_n)_n$ est une suite de variables aléatoires à valeurs dans \mathbb{R}^d et X une variable aléatoire à valeur dans \mathbb{R}^d toutes définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

- Convergence presque sûre : $X_n \rightarrow X$ \mathbb{P} p.s. ssi $\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1$.
- Convergence en probabilité : $X_n \rightarrow X$ en probabilité ssi $\forall \epsilon > 0, \mathbb{P}(|X_n - X| > \epsilon) \rightarrow_{n \rightarrow \infty} 0$.
- Convergence \mathcal{L}^p : $X_n \rightarrow X$ dans \mathcal{L}^p ssi $\mathbb{E}(|X_n - X|^p) \rightarrow_{n \rightarrow \infty} 0$.
- Convergence en loi : $X_n \rightarrow X$ en loi ssi $\mathbb{E}(f(X_n)) \rightarrow_{n \rightarrow \infty} \mathbb{E}(f(X))$ pour tout $f \in C_b$.

La convergence en loi est juste la transcription de la convergence faible en termes de variables aléatoires. Les liens qui existent entre ces différentes notions de convergence sera vu en TD.

7.3 Convergence en loi

Définition 7.3.1. Soit $X_n : (\Omega_n, \mathcal{F}_n, \mathbb{P}^n) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ une suite de variables aléatoires non nécessairement définies sur le même espace de probabilité. On dira que les X_n convergent en loi vers la variable aléatoire X (définie sur un certain espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$) ssi

$$\mathbb{P}_{X_n}^n \Rightarrow \mathbb{P}_X$$

c'est-à-dire

$$\forall f \in C_b, \lim_{n \rightarrow \infty} \mathbb{E}^n [f(X_n)] = \mathbb{E} [f(X)]$$

Remarque 7.3.1. 1) Puisque les topologies faibles, vagues et étroites coïncident sur \mathcal{M}^1 , dans la définition ci-dessus, on peut en fait remplacer C_b par C_K ou C_0 .

2) On dira aussi parfois que X_n converge en loi vers une certaine probabilité μ et non pas vers une variable aléatoire. Ceci est à comprendre dans le sens $(\mu_n)_n$ converge faiblement vers μ (où μ_n est la loi de X_n).

Notation : On utilise indifféremment la notation $X_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} X$ ou bien $X_n \Rightarrow X$.

Exemple : Si X_n est une variable aléatoire suivant la loi binomiale $\mathcal{B}(n, \lambda/n)$ où $\lambda > 0$ est un paramètre alors $(X_n)_n$ converge en loi vers une loi de Poisson de paramètre λ (Cf. [26]).

Ce qui suit sera démontré en TD.

Proposition 7.3.1. Si $(X_n)_n$ converge en loi vers X et si $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ est continue alors $(f(X_n))_n$ converge en loi vers $f(X)$.

Exemple : Si (X_n, Y_n) converge en loi vers (X, Y) alors $(X_n + Y_n)_n$ converge en loi vers $X + Y$. Mais il n'est pas vrai que si $(X_n)_n$ (resp. $(Y_n)_n$) converge en loi vers X (resp. Y) alors $(X_n + Y_n)_n$ converge en loi vers $X + Y$. C'est cependant vrai si X_n est indépendante de Y_n pour tout n et X indépendante de Y .

Proposition 7.3.2. La convergence en probabilités implique la convergence en loi. La réciproque est fautive mais vraie si la limite est une constante.

Pour les variables aléatoires discrètes, on a un critère très simple de convergence en loi.

Proposition 7.3.3. Soit $(X_n)_n, X$ des variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans \mathbb{Z} . Alors

$$X_n \Rightarrow X \text{ ssi } \forall r \in \mathbb{Z}, \lim_{n \rightarrow +\infty} \mathbb{P}(X_n = r) = \mathbb{P}(X = r)$$

Convergence en loi en termes de fonctions de répartition (dimension $d = 1$)

Théorème 7.3.1. $(X_n)_n$ converge en loi vers X ssi $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ en tout point de continuité x de F_X .

Démonstration.

\Rightarrow : Si x est un point de continuité de F_X , alors $(-\infty; x]$ est un ensemble de continuité de \mathbb{P}_X donc

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \lim_{n \rightarrow \infty} \mathbb{P}_{X_n}((-\infty; x]) = \mathbb{P}_X((-\infty; x]) = F_X(x)$$

\Leftarrow : On rappelle qu'il n'y a au plus qu'un nombre dénombrable de points de discontinuité de F_X (car il ne peut y avoir pour chaque $\varepsilon > 0$ qu'un nombre fini de sauts de taille plus grande que ε puisque F_X est bornée et croissante). Soit donc $f \in C_0(\mathbb{R})$. Pour chaque $\varepsilon > 0$, on peut trouver (voir TD pour plus de détails) une fonction g de la forme

$$g = \sum_{j=1}^k \alpha_j 1_{(a_j; b_j]}, \quad a_j < b_j \leq a_{j+1} < b_{j+1}$$

avec a_j et b_j qui ne sont pas des points de discontinuité de F_X et

$$\|f - g\|_\infty \leq \varepsilon$$

On a alors

$$\int g d\mathbb{P}_{X_n} \xrightarrow{n \rightarrow \infty} \sum_{j=1}^k \alpha_j (F_X(b_j) - F_X(a_j)) = \int g d\mathbb{P}_X$$

Il suffit maintenant d'appliquer l'inégalité triangulaire pour obtenir

$$\left| \int f d\mathbb{P}_{X_n} - \int f d\mathbb{P}_X \right| \leq 2\varepsilon + \left| \int g d\mathbb{P}_{X_n} - \int g d\mathbb{P}_X \right| \leq 3\varepsilon$$

dès que n est assez grand. □

Nous donnons maintenant le théorème d'Helly-Bray dont on pourra trouver une preuve élémentaire dans [30] mais, qui vu le travail précédent, apparaît comme une simple conséquence du théorème 7.3.1 et de la proposition 7.1.4.

Théorème 7.3.2. (Helly-Bray)

Soit $(F_n)_n$ une suite de fonctions de répartition tendue (c'est-à-dire dont la suite de mesures de probabilité associées est tendue). Il existe alors une fonction de répartition F et une sous-suite $(F_{n_k})_k$ de $(F_n)_n$ telle que

$$\lim_{k \rightarrow \infty} F_{n_k}(x) = F(x)$$

en tout point de continuité x de F .

7.4 TD Probabilités : Modes de convergence

Exercice 7.4.1. lien entre la convergence en probabilité et la convergence en norme \mathbb{L}^p , [6], p. 176 et [7], p. 227

Montrer à l'aide de l'inégalité de Markov que la convergence \mathbb{L}^p implique la convergence en probabilité.

Exercice 7.4.2. Lien entre la convergence en probabilité et la convergence presque sûre, [7], pp. 175-176.

1) Montrer que la convergence presque sûre implique la convergence en probabilité. Indication : si $X_n \rightarrow X$ p.s., poser $B_k = \limsup_n \{|X_n - X| \geq \frac{1}{k}\}$. Montrer que si $\omega \notin \bigcup_k B_k$, on a $\lim X_n(\omega) = X(\omega)$. On pourra aussi appliquer le théorème de convergence dominée.

2) Montrer que la réciproque est fautive. On pourra considérer une suite $(X_n)_n$ de Bernoulli indépendantes de paramètres respectifs $p_n = 1 - q_n$ avec $\lim_{n \rightarrow \infty} p_n = 0$ mais tels que la série de terme général p_n soit divergente. (*ind.* : Considérer $N = \{X_n = 1 \text{ i.o.}\}$ et montrer que $\mathbb{P}(N) = 1$)

3) Montrer que de toute suite de variables aléatoires convergeant en probabilité, on peut extraire une sous-suite convergeant presque sûrement. (*ind.* : Si $(X_n)_n$ converge en probabilité vers X , montrer que l'on peut construire une sous-suite $(X_{n_k})_k$ de $(X_n)_n$ telle que $\mathbb{P}(\|X_{n_k} - X\| \geq 1/k) \leq 1/2^k$ et conclure par le lemme de Borel-Cantelli).

Exercice 7.4.3. Lien entre la convergence en probabilité et la convergence en loi, [26], p. 315.

Rappel : $X_n \rightarrow X$ en loi signifie $\mathbb{P}_{X_n} \rightarrow \mathbb{P}_X$ étroitement, i.e.

$$\forall f \in C_b(\mathbb{R}), \int_{\mathbb{R}} f d\mathbb{P}_{X_n} \rightarrow \int_{\mathbb{R}} f d\mathbb{P}_X, \quad (\text{soit } \mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)])$$

Soit $(X_n)_n$ une suite de variables aléatoires et X une variable aléatoire définies sur le même espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$.

1) ([26], proposition 10.5) On suppose que les $(X_n)_n$ convergent en probabilité vers X . Soit f une fonction continue de \mathbb{R}^d dans \mathbb{R} .

a) Montrer que pour tout $\epsilon, a > 0$, il existe $\eta > 0$ tel que

$$\{|f(X_n) - f(X)| > \epsilon\} \subset \{|X| > a\} \cup \{|X_n - X| > \inf(\eta, a)\}$$

b) En choisissant convenablement a , en déduire que $(f(X_n))_n$ converge en probabilité vers $f(X)$.

2) On suppose que les $(X_n)_n$ convergent en probabilité vers X . Montrer que les $(X_n)_n$ convergent en loi vers X .

Indication : on montrera que

$$\forall \epsilon > 0, \forall f \in C_b(\mathbb{R}), \left| \int_{\mathbb{R}} f d\mathbb{P}_{X_n} - \int_{\mathbb{R}} f d\mathbb{P}_X \right| \leq \epsilon + 2\|f\|_{\infty} \mathbb{P}(|f(X_n) - f(X)| > \epsilon)$$

3) Montrer que la réciproque est fautive en considérant le contre-exemple suivant : pour tout n , $X_n = X$ et X est une Bernoulli de paramètre $1/2$. Montrer que $(X_n)_n$ converge en loi vers $(1 - X)$ mais que $(X_n)_n$ ne peut pas converger en probabilité vers $(1 - X)$.

4) Montrer néanmoins que si les X_n convergent en loi vers une constante a alors X_n convergent en probabilité vers a . (*ind.* : remarquer que pour $\epsilon > 0$, si δ_a désigne la masse de Dirac en a et $B(a, \epsilon)$ la boule fermée de centre a et de rayon ϵ , $\delta_a(\partial B(a, \epsilon)) = 0$. En déduire $\lim_{n \rightarrow +\infty} \mathbb{P}(\|X_n - a\| \leq \epsilon) = \delta_a(B(a, \epsilon)) = 1$.)

Conclusion : On vient de démontrer que la convergence en norme \mathbb{L}^p et la convergence presque sûre impliquent la convergence en probabilités qui implique la convergence en loi. Les réciproques sont fausses et sont partiellement vraies si on rajoute certaines hypothèses. Regarder en détail le tableau de la hiérarchie des convergences dans [6] p. 179 et [7], p.238.

Exercice 7.4.4. [26], Exercice 14.7

Soient sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ deux suites $(X_n)_n$ et $(Y_n)_n$ de variables aléatoires qui convergent en loi respectivement vers X et Y .

1) Montrer que si pour tout n , X_n et Y_n sont indépendantes et si X et Y sont indépendantes alors $(X_n, Y_n)_n$ converge en loi vers (X, Y) . On utilisera les fonctions caractéristiques et le théorème de Lévy ou bien les fonctions de répartition.

2) On étudie un contre-exemple dans le cas où l'on supprime l'hypothèse "pour tout n , X_n et Y_n sont indépendantes". Soit donc X et Y deux variables aléatoires réelles indépendantes de même loi de Bernoulli de paramètre $1/2$. On pose pour tout n .

$$X_n = X + 1/n, \quad Y_n = (1 - X) - 1/n$$

Etudier la convergence en loi des trois suites $(X_n)_n$, $(Y_n)_n$ et $(X_n + Y_n)_n$ et conclure.

Le lemme de Slutsky donne une alternative à l'indépendance pour assurer la propriété de convergence en loi étudiée à la première question de l'exercice précédent.

Exercice 7.4.5. Exercice 14.8, [26], lemme de Slutsky

Soient sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ deux suites $(X_n)_n$ et $(Y_n)_n$ de variables aléatoires qui convergent en loi respectivement vers X et une constante y_0 .

1) a) En utilisant le théorème de Stone-Weierstrass, montrer que l'ensemble

$$\mathcal{H} = \{(x, y) \rightarrow f(x)g(y) \mid f, g \in C_0(\mathbb{R})\}$$

est total dans $C_0(\mathbb{R}^2)$ ($C_0(\mathbb{R}^k)$ est l'espace des fonctions continues sur \mathbb{R}^k tendant vers 0 en l'infini).

b) On rappelle que $(Y_n)_n$ converge en loi vers une constante donc converge en probabilité vers cette même constante (exercice 7.4.3). Soit f et g deux éléments de $C_0(\mathbb{R})$. Montrer que

$$\begin{aligned} |\mathbb{E}(f(X_n)g(Y_n)) - \mathbb{E}(f(X)g(y_0))| &\leq \|f\|_\infty [\epsilon + 2\|g\|_\infty \mathbb{P}(|g(Y_n) - g(y_0)| > \epsilon)] \\ &+ \|g\|_\infty |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \end{aligned}$$

c) Conclure.

Exercice 7.4.6. Stabilité du caractère gaussien

Soit $(X_n)_n$ une suite de variables aléatoires gaussiennes qui convergent en moyenne quadratique vers une variable aléatoire X . Montrer que X est une variable gaussienne. (*ind.* : on remarquera que la convergence en moyenne quadratique implique la convergence en loi, on passera aux fonctions caractéristiques et utilisera le théorème de convergence de Lévy.)

Exercice 7.4.7. Convergence en loi pour les variables aléatoires discrètes, [26], page 317.

Soit $(X_n)_n$ une suite de variables aléatoires et X une variable aléatoire définies sur le même espace probabilisé et toutes à valeurs dans \mathbb{Z} . Montrer que $(X_n)_n$ converge en loi vers X si et seulement si pour tout $r \in \mathbb{Z}$, $\mathbb{P}(X_n = r) \xrightarrow{n \rightarrow \infty} \mathbb{P}(X = r)$. (*ind.* : pour la condition nécessaire, approcher la fonction indicatrice de r par un élément de $C_K(\mathbb{R})$ à support dans $[r - \frac{1}{2}, r + \frac{1}{2}]$, et pour la condition suffisante, remarquer que $\mathbb{E}(f(X_n)) = \sum_{r \in K} f(r)\mathbb{P}(X_n = r)$ pour toute fonction continue à support dans le compact K .)

Solution

Soit $f \in C_K(\mathbb{R})$, $\text{supp}(f) \subset [-r + \frac{1}{2}, r + \frac{1}{2}]$, $f(r) \neq 0$. On a $\int_{\mathbb{R}} f d\mathbb{P}_{X_n} \rightarrow \int_{\mathbb{R}} f d\mathbb{P}_X$ et donc $f(r)\mathbb{P}(X_n = r) \rightarrow f(r)\mathbb{P}(X = r)$. Réciproquement, par la relation donnée, on a $\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X))$, $\forall f \in C_k(\mathbb{R})$. Donc $\mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X) \forall f \in C_K(\mathbb{R})$. Comme sur \mathcal{M}' les topologies faibles, vagues et étroites coïncident, on conclut.

Exercice 7.4.8. Convergence en loi et fonction de répartition, [26], pp. 318-319

Soit $(X_n)_n$ une suite de variables aléatoires et X une variable aléatoire définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On note F_n la fonction de répartition de X_n et F celle de X . On a le critère suivant :

$$X_n \Longrightarrow X \quad \text{ssi} \quad \lim_{n \rightarrow +\infty} F_n(x) = F(x)$$

en tout point de continuité x de F .

1) Pour la condition nécessaire, remarquer que si x est un point de continuité de F alors $(-\infty; x]$ est un ensemble de \mathbb{P}_X -continuité. (C'est-à-dire que $\mathbb{P}_X(\{x\}) = 0$ et donc $P_X(\cdot - \infty, y] \rightarrow P_X(\cdot - \infty, x]$ quand $y \rightarrow x$.) Rappel : si μ_n converge étroitement vers μ , alors $\lim_n \mu(A_n) = \mu(A)$ pour tout borélien de μ -continuité.

2) Pour la condition suffisante, on se donne un $f \in C_0(\mathbb{R})$ et $\epsilon > 0$.

a) Se rappeler pourquoi l'ensemble des points de discontinuité de F est au plus dénombrable.

b) Justifier l'existence d'une fonction $g = \sum_{j=1}^k \alpha_j 1_{]a_j; b_j]}$ où $a_j < b_j \leq a_{j+1} < b_{j+1}$ et a_j et b_j sont des points de continuité de F_X et telle que $\|f - g\|_{\infty} \leq \epsilon$. (*ind.* : on pourra se servir du fait que f est uniformément continue).

c) Montrer que $\lim_{n \rightarrow +\infty} \mathbb{E}(g(X_n)) = \mathbb{E}(g(X))$ et conclure. *Indication* : $\mathbb{E}g(X_n) = \int_{\mathbb{R}} g d\mathbb{P}_{X_n} = \sum_j \alpha_j (F_n(b_j) - F_n(a_j)) \rightarrow \sum_j \alpha_j (F(b_j) - F(a_j)) = \mathbb{E}g(X)$.

Exercice 7.4.9. Représentation de Skorokhod d'une fonction de répartition, [30], p. 34.

On va montrer que $F : \mathbb{R} \rightarrow [0, 1]$ est la fonction de répartition d'une variable aléatoire réelle si et seulement si F est croissante, continue à droite et si $\lim_{z \rightarrow -\infty} F(z) = 0$, $\lim_{z \rightarrow +\infty} F(z) = 1$.

1) Dessiner X^+ et X^- (cf. question 3 pour la définition) pour les fonctions F suivantes : $F = \mathbb{1}_{\mathbb{R}^+}$, et $F(z) = 0$ pour $z \leq 0$, $F(z) = z/2$ pour $0 \leq z \leq 1$, $F(z) = 1/2$ pour $1 \leq z < 2$, $F(z) = 1$ pour $z \geq 2$.

2) Montrer que ces conditions sont nécessaires.

3) Supposons que F ait les quatre propriétés précédentes. On pose pour $\omega \in [0, 1]$ muni de la tribu des boréliens et de la mesure de Lebesgue,

$$X^+(\omega) =: \inf\{z : F(z) > \omega\} = \sup\{y : F(y) \leq \omega\},$$

$$X^-(\omega) =: \inf\{z : F(z) \geq \omega\} = \sup\{y : F(y) < \omega\}.$$

Remarquer que $(z > X^-(\omega)) \Rightarrow (F(z) \geq \omega)$. En utilisant la continuité à droite de F montrer que

$$F(X^-(\omega)) \geq \omega \text{ et que } (X^-(\omega) \leq c) \Rightarrow (\omega \leq F(X^-(\omega)) \leq F(c)).$$

Déduire que

$$\mathbb{P}(X^- \leq c) = F(c).$$

4) On va montrer que X^+ a aussi pour fonction de répartition F et que $\mathbb{P}(X^+ = X^-) = 1$. Expliquer successivement pourquoi $(\omega < F(c)) \Rightarrow (X^+(\omega) \leq c)$, $F(c) \leq \mathbb{P}(X^+ \leq c)$, $X^- \leq X^+$. En remarquant que

$$\mathbb{P}(X^- \neq X^+) = \bigcup_{c \in \mathbb{Q}} \{X^- \leq c < X^+\},$$

déduire en utilisant le 2) que $X^+ = X^-$ presque partout. Conclure.

Solution

3) $z > X(\omega) \Rightarrow F(z) \geq \omega$ par définition de $X^-(\omega)$. On a donc $\forall z > X^-(\omega), F(z) \geq \omega$ et comme F est continue à droite, $F(X^-(\omega)) \geq \omega$. Si $X^-(\omega) \leq c$, on a $F(X^-(\omega)) \leq F(c)$ et donc $\omega \leq F(c)$. En résumé, $X^-(\omega) \leq c \Rightarrow \omega \leq F(c)$ et réciproquement $\omega \leq F(c) \Rightarrow c \geq X^-(\omega)$ par définition de $X^-(\omega)$. Donc $X^-(\omega) \leq c \Leftrightarrow \omega \leq F(c)$. En conséquence, $\mathbb{P}(X^- \leq c) = \mathbb{P}(\omega \leq F(c)) = F(c) - 0 = F(c)$.

4) $(\omega < F(c)) \Rightarrow (X^+(\omega) \leq c)$ par définition de X^+ . $F(c) \leq \mathbb{P}(X^+ \leq c)$ s'en déduit immédiatement puisque $\mathbb{P}(\omega < F(c)) = F(c)$. $X^- \leq X^+$ est évident. On a $\mathbb{P}(X^- \leq c \leq X^+) = \mathbb{P}(\{X^- \leq c\} \setminus \{X^+ \leq c\}) = \mathbb{P}(X^- \leq c) - \mathbb{P}(X^+ \leq c) \leq F(c) - F(c) = 0$. Donc $\mathbb{P}(X^- \neq X^+) = 0$. On obtient $X^+ = X^-$ p.s., ce qui implique que X^+ et X^- ont tous deux F comme fonction de répartition.

Remarque 7.4.1. *Etant donné une fonction de répartition F , ce théorème montre comment simuler une variable aléatoire X^- de fonction de répartition F . Bien entendu, cette méthode suppose que l'on sache calculer X^- à partir de F ...*

Exercice 7.4.10. La représentation de Skorokhod, [30], p.182.

1) Dédurre des exercices précédents que si X_n , de fonction de répartition F_n , tend presque sûrement vers X , de fonction de répartition F , alors $F_n(x) \rightarrow F(x)$ en tout point x de continuité de F . Notre but est de montrer une sorte de réciproque :

Théorème 7.4.1. *Si F_n et F sont des fonctions de répartition et si $F_n \rightarrow F$ en tout point de continuité de F , alors il existe des v.a. X_n et X sur $([0, 1], \mathcal{B}, \mathbb{P})$ (où \mathbb{P} est la mesure de Lebesgue) telles que $F = F_X$, $F_n = F_{X_n}$ et $X_n \rightarrow X$ p.s.*

2) On pose $X^+ =: \inf\{z : F(z) > \omega\}$, $X^- =: \inf\{z : F(z) \geq \omega\}$ et on définit de même X_n^+ et X_n^- . On rappelle (exercice 7.4.9) que X^+ et X^- ont pour fonction de répartition F et que $\mathbb{P}(X^+ = X^-) = 1$. On fixe $\omega \in [0, 1]$. Soit z un point non atomique pour F vérifiant $z > X^+(\omega)$. Montrer que $\limsup_n X_n^+(\omega) \leq z$ et en déduire que $\limsup_n X_n^+(\omega) \leq X^+(\omega)$.

3) Adapter l'argument précédent pour montrer que $\liminf_n X_n^-(\omega) \geq X^-(\omega)$ et conclure.

Solution

2) Par définition de X^+ , $F(z) > \omega$. Donc pour n grand, $F_n(z) > \omega$. Par la définition de X_n^+ , on a alors $X_n^+ \leq z$ et donc $\limsup_n X_n^+(\omega) \leq z$. Comme ceci est vrai $\forall z > X^+(\omega)$, on obtient $\limsup_n X_n^+(\omega) \leq X^+(\omega)$.

Exercice 7.4.11. (Théorème de Helly-Bray), [30] pp. 183-184.

Soit $(F_n)_n$ une suite de fonctions de répartition sur \mathbb{R} . Le but de l'exercice est de montrer qu'il existe une fonction croissante continue à droite F à valeurs dans $[0, 1]$ et une sous-suite $(F_{n_k})_k$ telle que

$$\lim_k F_{n_k}(x) = F(x) \text{ en tout point de continuité } x \text{ de } F$$

1) En utilisant le principe d'extraction diagonale, montrer qu'il existe une sous-suite $(F_{n_k})_k$ telle que pour tout $q \in \mathbb{Q}$, il existe un réel $H(q) \in [0, 1]$ telle que

$$\lim_{k \rightarrow +\infty} F_{n_k}(q) = H(q)$$

2) Pour $x \in \mathbb{R}$, on définit $F(x)$ par

$$F(x) = \lim_{\substack{q \in \mathbb{Q}, q > x \\ q \rightarrow x}} H(q)$$

Montrer que F est continue à droite.

3) Conclure.

Chapitre 8

Pile ou face, marches aléatoires

Ces travaux dirigés suivent [18], section 3.9 pages 71 et suivantes. Nous suivrons la présentation élémentaire (sans le formalisme martingales et chaînes de Markov). On montre tous les résultats par des raisonnements élémentaires, le principe de réflexion et un usage habile de la fonction génératrice. Cette présentation élémentaire est à connaître. L'option probabilité verra des résultats complémentaires, et retrouvera les mêmes, grâce aux martingales et aux chaînes de Markov. Le nombre des questions qu'on peut se poser sur le jeu de pile ou face est immense. En fait, la liste des questions et réponses présentée aussi dépend aussi de l'accessibilité technique du résultat. Un ouvrage entier classique, en deux tomes, le Feller, est entièrement consacré au jeu de pile ou face. Voici une liste succincte des questions que l'on se pose avec une indication des réponses démontrées dans ce chapitre.

Soit X_n une suite de variables de Bernoulli i.i.d. centrées de paramètre p , c'est-à-dire $\mathbb{P}(X_n = 1) = p$ et $\mathbb{P}(X_n = -1) = 1 - p = q$. Soit S_0 une v.a. indépendante des X_i telle que $\mathbb{P}(S_0 = k) > 0$ pour tout $k \in \mathbb{Z}$. On pose $S_n = S_0 + \sum_{i=1}^n X_i$. On peut interpréter S_n comme la fortune d'un joueur de roulette ou de pile ou face. La probabilité de gagner pour le joueur est p et sa mise est supposée unitaire. X_n désigne le n -ième tirage. Par exemple si le joueur joue à rouge ou noir, sa probabilité est $\frac{18}{37}$, car il y a 36 chiffres colorés dont 18 noirs et 18 rouges, plus le zéro. La banque ramasse tout quand le zéro sort. (Lire *Le Joueur* de Dostoïevsky pour tous les détails, probablement la plus belle analyse littéraire et psychologique du jeu de roulette mais avec une suite d'observations concrètes). La suite S_n s'appelle une *marche aléatoire simple*. Elle correspond aussi au mouvement aléatoire d'une particule sur \mathbb{N} , progressant de $+1$ avec probabilité p et reculant de -1 avec probabilité q . La marche est appelée symétrique si $p = q = \frac{1}{2}$.

- S_n est-elle une chaîne de Markov (le futur ne dépend que du présent ?). Oui, calculs élémentaires de probabilité conditionnelle.
- Ruine du joueur : le joueur commence avec une fortune de k , la banque a une fortune de N : le jeu s'arrête-t-il en temps fini ? (oui, technique de temps d'arrêt élémentaire). Permet de montrer "a la main" le théorème d'arrêt de Doob.
- Ruine du joueur : quelle est l'espérance du temps de jeu ? $D_k = k(N - k)$. L'interprétation de ce résultat est passionnante. Technique d'espérance conditionnelle (discrète).
- Ruine du joueur : Probabilité de ruine du joueur : $p_k = 1 - \frac{k}{N}$. Là encore l'interprétation doit faire l'objet d'une discussion (la meilleure illustration est le film de Luigi Comencini, *L'argent de la vieille*).
- Le retour à l'origine partant de l'origine. Calcul de la probabilité de retour : $1 - |p - q|$. Technique : usage intensif des fonctions génératrices de distributions discrètes.
- Egalement par la même technique : calcul de la probabilité de retour en exactement k pas.
- Egalement par la même technique : calcul des probabilités d'atteindre $1, 2, r$ partant de 0 . On utilise aussi le *principe de réflexion*.
- Probabilité d'aller de $a > 0$ en $b > 0$ sans passer par 0 (c'est-à-dire sans se ruiner). Ce calcul utilise l'astuce remarquable du **principe de réflexion**.
- Une suite de résultats sur la probabilité des longueurs d'excursions sans ruine permet d'établir le

théorème de ballottage ou de scrutin, qui donne la probabilité (étonnamment élevée) que lors d'un dépouillement de votes le résultat du dépouillement partiel contredise le résultat final. La loi de l'arcsinus montre évalue la probabilité que dans un jeu un des joueurs soit gagnant ($S_n > 0$) sur une très longue période. L'interprétation de ces résultats permet de comprendre mieux le phénomène du jeu de hasard, l'existence des casinos, les suspenses paradoxaux provoqués par le jeu. **Lors d'une leçon sur pile ou face, il est essentiel d'interpréter chaque résultat donné.**

8.1 Propriétés élémentaires

Exercice 8.1.1. 1) Représentation graphique standard : tirer à pile ou face 10 fois. Noter les résultats. Représenter les résultats avec en abscisse n et en ordonnée la valeur de S_n , avec $S_0 = 0$. Joindre les points par des segments de droite.

2) Homogénéité spatiale. Montrer que

$$\mathbb{P}(S_n = j | S_0 = a) = \mathbb{P}(S_n = j + b | S_0 = a + b). \quad (8.1.1)$$

Indication : Montrer que les deux cotés sont égaux à $\mathbb{P}(\sum_1^n X_i = j - a)$.

3) Homogénéité temporelle Montrer que

$$\mathbb{P}(S_n = j | S_0 = a) = \mathbb{P}(S_{m+n} = j | S_m = a). \quad (8.1.2)$$

Indication : Le membre de gauche est $\mathbb{P}(\sum_1^n X_i = j - a)$ et le membre de droite est $\mathbb{P}(\sum_{m+1}^{m+n} X_i = j - a)$.

4) Les marches aléatoires simples ont la propriété de Markov

$$\mathbb{P}(S_{m+n} = j | S_0, S_1, \dots, S_m) = \mathbb{P}(S_{m+n} = j | S_m), \quad n \geq 0. \quad (8.1.3)$$

D'abord, comprenons bien ce que veut dire cette relation, à savoir que pour toutes les valeurs possibles de a_0, a_1, \dots, a_m , on a

$$\mathbb{P}(S_{m+n} = j | S_0 = a_0, S_1 = a_1, \dots, S_m = a_m) = \mathbb{P}(S_{m+n} = j | S_m = a_m). \quad (8.1.4)$$

La "démonstration" de [18] est la suivante "If one knows the value of S_m , then the distribution of S_{m+n} depends only on the jumps X_{m+1}, \dots, X_{m+n} and cannot depend on further information concerning the values of S_0, S_1, \dots, S_{m-1} . Vous paraît-elle suffisante ?

5) Montrer (8.1.3) en menant les calculs grâce à (8.1.4) et en utilisant la définition de la probabilité conditionnelle.

Solution

(C'est juste une indication, je vous conseille de détailler un peu plus...)

$$\begin{aligned} \mathbb{P}(S_{m+n} = j | (S_0 \dots S_m) = (a_0 \dots a_m)) &= \mathbb{P}\left(\sum_{m+1}^n X_i = j - a_m | (S_0 \dots S_m) = (a_0 \dots a_m)\right) = \\ &= \frac{\mathbb{P}\left(\sum_{m+1}^n X_i = j - a_m\right) \mathbb{P}\left((S_0 \dots S_m) = (a_0 \dots a_m)\right)}{\mathbb{P}\left((S_0 \dots S_m) = (a_0 \dots a_m)\right)} = \mathbb{P}\left(\sum_{m+1}^n X_i = j - a_m\right) = \\ &= \mathbb{P}(S_{m+n} = j | S_m = a_m). \end{aligned}$$

Exercice 8.1.2. Pile ou face-La ruine du joueur, [18] 1.7 p. 17 et 3.9 p. 74. Un joueur a une fortune de k et il désire atteindre en jouant à pile ou face une fortune de N , $0 < k < N$. Il parie toujours sur pile, qui a probabilité p . La question est de savoir avec quelle probabilité il se ruinera avant d'atteindre son but. On désigne par A_k l'événement "ruine du joueur, partant d'une fortune $S_0 = k$ " et par B l'événement que le résultat du premier tirage est pile. On munit l'espace Ω des suites infinies de tirages de la probabilité \mathbb{P} construite par le théorème de Kolmogorov. On note $\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots)$ un élément de Ω . On note également $\omega = (\omega_1, \omega^1)$ avec $\omega^1 = (\omega_2, \dots, \omega_n, \dots)$.

1) Donner une définition formelle de l'événement "ruine du joueur". On utilisera le temps d'arrêt

$$t_k(\omega) = \inf\{n, k + \omega_1 + \dots + \omega_n = 0 \text{ ou } N\}.$$

Montrer que t_k est mesurable. Montrer que pour $1 < k < N - 1$,

$$\mathbb{P}(A_k) = p\mathbb{P}(A_{k+1}) + q\mathbb{P}(A_{k-1}).$$

On commencera par interpréter cette formule. Vérifier que cette formule est encore valable pour $k = 1$ ou $N - 1$ en remarquant que $\mathbb{P}(A_0) = 1$ et $\mathbb{P}(A_N) = 0$

Indication : montrer que pour $\omega \in \{\omega_1 = 1\}$, on a $t_k(\omega) = t_{k+1}(\omega^1) + 1$. Attention : ni [18] ni [16] ne jugent bon de donner cette preuve. Le formalisme de [18] est discutable : ils notent $\mathbb{P}_k(A)$ ce que nous notons $\mathbb{P}(A_k)$, l'événement A étant alors un événement absolu appelé "ruine du joueur". L'ennui, c'est que cet événement dépend du point de départ k et donc, A a de toutes manières une signification différente quand k change. Il vaut mieux le noter A_k et garder le même espace de probabilité Ω pour toutes les marches aléatoires. Pour le reste, nous suivons [18].

2) On pose $p_k = \mathbb{P}(A_k)$. Si $p = \frac{1}{2}$, déduire que $p_k = \frac{1}{2}(p_{k+1} + p_{k-1})$, $0 < k < N$.

3) En utilisant les valeurs de p_0 et p_N , montrer que

$$\mathbb{P}(A_k) = 1 - \frac{k}{N}.$$

4) Interpréter cette relation. On pourra s'inspirer comme suggéré par Grimmett-Stirzaker de "Millionaires should always gamble, poor men never" (J.M. Keynes). Et voir *L'argent de la vieille* de Luigi Comencini.

5) Plus généralement, on a l'équation $p_k = p.p_{k+1} + q.p_{k-1}$, $1 \leq k \leq N - 1$, avec conditions au bord $p_0 = 1$, $p_N = 0$. Montrer que

$$p_k = \frac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}.$$

6) En déduire q_k , la probabilité que le joueur fasse fortune.

7) Vérifier que $p_k + q_k = 1$. Interpréter.

8) Montrer que le temps d'arrêt est presque sûrement fini.

9) On pose $D_k = \mathbb{E}t_k$. Grimmett-Stirzaker n'écrit pas cette définition mais décrit D_k comme "the mean number of steps before the particle hits one of the absorbing barriers, starting from k ". Justifier la relation proposée par ces auteurs

$$D_k = p(1 + D_{k+1}) + q(1 + D_{k-1}), \quad 1 \leq k \leq N - 1, \quad D_0 = D_N = 0.$$

Pour cela, on appliquera au temps d'arrêt les formules de l'espérance conditionnelle pour des variables aléatoires discrètes et des événements B que l'on rappelle :

$$\mathbb{E}Y = \sum_x \mathbb{E}(Y|X = x)\mathbb{P}(X = x), \quad (8.1.5)$$

$$\mathbb{E}(Y|B) = \sum_l l\mathbb{P}(Y = l|B). \quad (8.1.6)$$

10) En déduire que

$$D_k = \frac{1}{q-p} \left[k - N \left(\frac{1 - (\frac{q}{p})^k}{1 - (\frac{q}{p})^N} \right) \right] \quad \text{si } p \neq \frac{1}{2},$$

$$D_k = k(N - k) \quad \text{si } p = \frac{1}{2}.$$

11) Interpréter ces résultats.

Solution

1) On a

$$A_k = \{t_k(\omega) = \inf\{n, k + \omega_1 + \dots + \omega_n = 0 \text{ ou } N\} < \infty \text{ et } k + \omega_1 + \dots + \omega_{t_k(\omega)} = 0\}.$$

On écrit $A_k = (A_k \cap \{\omega_1 = 1\}) \cup (A_k \cap \{\omega_1 = -1\})$. Mais pour $\omega \in \{\omega_1 = 1\}$, on a

$$t_k(\omega) = \inf\{n, k + 1 + \omega_2 + \dots + \omega_n = 0 \text{ ou } N\} < \infty = t_{k+1}(\omega^1) + 1.$$

On en déduit que

$$A_k \cap \{\omega_1 = 1\} = \{t_{k+1}(\omega^1) < \infty \text{ et } k + 1 + \omega_2 + \dots + \omega_{t_{k+1}(\omega^1)} = 0\} \cap \{\omega_1 = 1\}.$$

Comme les X_i sont i.i.d., on obtient

$$\mathbb{P}(A_k \cap \{\omega_1 = 1\}) = p\mathbb{P}(A_{k+1}).$$

On conclut aisément.

2), 3) 4) Immédiat, simple calcul. Le résultat nous dit que plus on est pauvre, moins on a de chances de devenir riche. L'interprétation est toutefois facilitée si on considère le jeu de manière symétrique : le premier joueur possède k et le second possède $N - k$. On peut alors reformuler la règle du jeu ainsi : le jeu s'arrête quand l'un des deux joueurs est ruiné. La probabilité de ruine du premier joueur est $\frac{N-k}{N}$ et celle du second $\frac{k}{N}$. L'une tend vers 1 et l'autre vers 0 quand $\frac{N}{k}$ tend vers l'infini. Les riches gagnent donc presque toujours. Mais, quand ils perdent, il perdent gros, comme on peut le vérifier en remarquant que l'espérance des gains des deux joueurs au moment de l'arrêt du jeu sont les mêmes et égales à zéro. En effet, l'espérance du gain du premier joueur est $-k\mathbb{P}(A_k) + (N - k)(1 - \mathbb{P}(A_k)) = 0$ et on vérifie immédiatement que celle du second est aussi zéro. Ce résultat est connu sous le nom de théorème d'arrêt de Doob ([30] page 100), mais pour appliquer directement ce théorème il faudrait d'abord démontrer que le temps d'arrêt $t_k(\omega)$ est presque sûrement fini. Une autre manière de calculer $\mathbb{P}(A_k)$ est donc, pour ceux qui connaissent les martingales, de :

- montrer que le temps d'arrêt du jeu est presque sûrement fini.
- utiliser le théorème d'arrêt qui dit que sous cette condition, $\mathbb{E}S_{t_k} = \mathbb{E}S_0 = k$, ce qui veut exactement dire que l'espérance de gain est nulle.

Il est quand même préférable de savoir que l'on peut tout faire avec des moyens élémentaires !

5) Calculs classiques avec la formule de récurrence.

6) La valeur de q_k s'interprète comme une probabilité de ruine en échangeant les rôles de p et q et de k et $N - k$. On obtient donc $q_k = \frac{\left(\frac{p}{q}\right)^{N-k} - \left(\frac{p}{q}\right)^N}{1 - \left(\frac{p}{q}\right)^N}$.

7) 8) La relation est vite vérifiée. On en déduit que presque sûrement, le joueur se ruine ou fait fortune en temps fini, ce qui veut impliquer que t_k est presque sûrement fini.

9) On a en utilisant (8.1.6),

$$\begin{aligned} \mathbb{E}(t_k | X_1 = -1) &= \sum_l l \frac{\mathbb{P}(t_k = l; X_1 = -1)}{\mathbb{P}(X_1 = -1)} = \\ &= \sum_l l \frac{\mathbb{P}(\inf\{n, k-1 + X_2 + \dots + X_n = 0 \text{ ou } N\} = l)q}{q} = \\ &= \sum_l l \mathbb{P}(\inf\{n, k-1 + X_1 + \dots + X_{n-1} = 0 \text{ ou } N\} = l) = \\ &= \sum_l l \mathbb{P}(\inf\{n-1, k-1 + X_1 + \dots + X_{n-1} = 0 \text{ ou } N\} = l-1) = \\ &= 1 + \sum_l (l-1) \mathbb{P}(t_{k-1} = l-1) = 1 + \mathbb{E}t_{k-1} = 1 + D_{k-1} \end{aligned}$$

On a utilisé le fait que $X_2 + \dots + X_n$ et $X_1 + \dots + X_{n-1}$ ont la même loi, l'indépendance de X_1 et des X_i pour $i \geq 2$ et le fait que t_k est presque sûrement fini. On a de même

$$\mathbb{E}(t_k | X_i = +1) = 1 + D_{k+1}.$$

En utilisant la formule (8.1.5), appliquée à $Y = t_k$ et $X = X_1$, on obtient comme demandé

$$D_k = p(1 + D_{k+1}) + q(1 + D_{k-1}).$$

10)-11) Simple calcul. L'interprétation n'est pas très excitante si p est très différent de q , car alors l'un des joueurs a beaucoup plus de chances que l'autre. On va donc interpréter le cas $p = \frac{1}{2}$ et remarquer qu'aux jeux de casino, p est assez proche de $\frac{1}{2}$ et la première formule donnée pour D_k peut alors s'interpréter dans la même ligne que la seconde. Le point important qui a été démontré est que

le temps de jeu $D_k = k(N - k)$ est proportionnel au produit des fortunes des deux joueurs.

Ceci explique pourquoi les casinos existent. On doit d'une part avoir $N - k$ très grand pour que le casino ne perde jamais. Cela garantit de plus un temps moyen de jeu élevé pour tous les joueurs et donc permet de les "accrocher". La fortune du casino étant grande et fixée, on remarque que la

même formule donne un temps de jeu à peu près proportionnel à la fortune du joueur. Pour s’amuser longtemps dans un casino, il vaut mieux faire de petites mises, et, à fortune égale, le temps de jeu est grosso modo double si les mises sont deux fois plus petites. On voit surgir d’intéressantes questions de stratégie pour le casino : il faut imposer des mises pas trop petites, car sinon le client risque de ressortir par simple ennui, avant d’avoir tout perdu. Inversement, une mise minimale trop grosse va chasser trop vite les petits clients qui auront manqué les plaisirs des “haut et bas” qui font le charme du jeu. Il y a donc un problème d’optimisation de la mise minimale imposée en fonction de la fortune moyenne des clients et de l’estimation du temps de jeu optimal (une fin de soirée). L’idéal est que les clients aient le temps de s’amuser avant de sortir les poches vides.

Remarque : ce TD était complété pour les rapides par les exercices pages 65-68 et 85-86 de [26]

Exercice 8.1.3. Le retour à l'origine

On considère une marche aléatoire sur \mathbb{Z} ,

$$S_n = X_1 + \dots + X_n, \quad S_0 = 0$$

où les X_i sont des Bernoulli indépendantes prenant la valeur 1 avec probabilité p et la valeur -1 avec la probabilité $1 - p = q$.

On pose $p_0(n) = \mathbb{P}(S_n = 0)$ et pour $n \geq 1$, $f_0(n) = \mathbb{P}(S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = 0)$ la probabilité que le premier retour en 0 se fasse après n pas. On note les fonctions génératrices associées

$$P_0(s) = \sum_0^{\infty} p_0(n)s^n, \quad F_0(s) = \sum_{n=1}^{\infty} f_0(n)s^n.$$

On va montrer le théorème suivant

Théorème 8.1.1. *On a*

(a) $P_0(s) = 1 + P_0(s)F_0(s)$,

(b) $P_0(s) = (1 - 4pqs^2)^{-\frac{1}{2}}$,

(c) $F_0(s) = 1 - (1 - 4pqs^2)^{\frac{1}{2}}$.

1) Soit $A = \{S_n = 0\}$ et B_k l'évènement que le premier retour à l'origine se fait à k pas. Remarquer que

$$\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A|B_k)\mathbb{P}(B_k).$$

2) Montrer que $\mathbb{P}(A|B_k) = p_0(n - k)$.

3) Dédurre que

$$p_0(n) = \sum_{k=1}^n p_0(n - k)f_0(k) \tag{8.1.7}$$

si $n \geq 1$ et obtenir (a).

4) Montrer par un simple argument combinatoire que $p_0(n) = C_n^{\frac{n}{2}}(pq)^{\frac{n}{2}}$ si n est pair, $= 0$ sinon et déduire (b) et (c).

5) Dédurre des résultats précédents que :

- la probabilité de retour à l'origine est $1 - |p - q|$. (Commencer par définir cette probabilité)
- cette probabilité n'est égale à 1 que si $p = \frac{1}{2}$ et alors le temps moyen de retour est infini. (Commencer par définir le temps moyen de retour).

Solution

1) C'est la règle des probabilités disjointes car les B_k sont disjoints.

2) On a

$$\begin{aligned} \mathbb{P}(A \cap B_k) &= \mathbb{P}(S_n = 0, S_k = 0, S_{k-1} \neq 0, \dots, S_1 \neq 0) \\ &= \mathbb{P}(X_{k+1} + \dots + X_n = 0, S_k = 0, S_{k-1} \neq 0, \dots, S_1 \neq 0) \\ &= \mathbb{P}(X_{k+1} + \dots + X_n = 0)\mathbb{P}(B_k) = \mathbb{P}(S_{n-k} = 0)\mathbb{P}(B_k) \end{aligned}$$

car les X_i sont i.i.d. et les lois de $X_1 + \dots + X_{n-k}$ et de $X_{k+1} + \dots + X_n$ sont donc égales. (On peut aussi utiliser l'homogénéité temporelle)

3) La première relation découle de 1) et 2). La seconde découle du fait que le calcul de la fonction génératrice transforme une convolution de suites en produit. Plus précisément, en utilisant la relation du 3) :

$$P_0(s)F_0(s) = \sum_1^\infty s^n \left[\sum_{k=1}^n f_0(k)p_0(n-k) \right] = \sum_1^\infty s^n p_0(n) = P_0(s) - 1.$$

4) $S_n = 0$ si et seulement si la particule fait un nombre égal à $\frac{n}{2}$ de pas vers la droite et un nombre égal vers la gauche. Ceci peut se faire de $C_n^{\frac{n}{2}}$ manières et chacune de ces marches a probabilité $(pq)^{\frac{n}{2}}$. Si n est impair il est impossible que $S_n = 0$ et donc $p_0(n) = 0$. Pour (b) : à vos développements en série ! La relation (c) découle de (a) et (b).

5) La probabilité de retour à l'origine est $\sum_{n=1}^\infty f_0(n) = F_0(1) = 1 - |p - q|$. Le retour à l'origine n'est certain que si $F_0(1) = 1$ et donc $p = \frac{1}{2}$. Le temps moyen de retour se définit comme l'espérance du temps de premier retour, soit $\mathbb{E}T_0 = \sum_{n=1}^\infty n f_0(n) = F_0'(1) = +\infty$. (La seconde égalité vient du théorème de convergence monotone).

Exercice 8.1.4. Le temps de première visite en 1

Soit T_1 le premier temps (éventuellement infini) pour lequel la marche atteint 1. Soit F_1 la fonction génératrice de $1_{\{T_1 < +\infty\}} T_1$ et pour tout entier n positif ou nul $f_1(n) = \mathbb{P}(T_1 = n)$.

1a) Montrer que $f_1(1) = p$ et que pour tout entier $n \geq 2$, on a la relation de récurrence $f_1(n) = q \sum_{\nu=2}^{n-1} f_1(\nu - 1) f_1(n - \nu)$

1b) En déduire que l'on a $F_1(s) = qsF_1(s)^2 + ps$, puis que F_1 satisfait $F_1(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2qs}$

2) En déduire la distribution de T_1 :

$$\mathbb{P}(T_1 = 2k) = 0, \quad \mathbb{P}(T_1 = 2k - 1) = \frac{(-1)^{k-1}}{2q} C_{1/2}^k (4pq)^k.$$

Solution

1a) Le résultat pour $n = 1$ étant trivial, soit donc $n \geq 2$. On a :

$$\begin{aligned} \mathbb{P}(T_1 = n) &= \mathbb{P}(S_n = 1, S_{n-1} \leq 0, \dots, S_1 \leq 0) = \\ &= \sum_{\nu=2}^{n-1} \mathbb{P}(X_1 = -1, \inf\{j \geq 2, -1 + X_2 + \dots + X_j = 0\} = \nu, \inf\{k \geq \nu + 1, X_{\nu+1} + \dots + X_k = 1\} = n) = \\ &= q \sum_{\nu=2}^{n-1} \mathbb{P}(\inf\{j \geq 2; -1 + X_2 + \dots + X_j = 0\} = \nu) \mathbb{P}(\inf\{k \geq \nu + 1, X_{\nu+1} + \dots + X_k = 1\} = n) = \\ &= q \sum_{\nu=2}^{n-1} \mathbb{P}(\inf\{j \geq 1; X_1 + \dots + X_j = 1\} = \nu - 1) \mathbb{P}(\inf\{k \geq 1, X_1 + \dots + X_k = 1\} = n - \nu) = \\ &= q \sum_{\nu=2}^{n-1} f_1(\nu - 1) f_1(n - \nu). \end{aligned}$$

La première égalité résulte du fait que l'on est en dimension 1 et que la marche est plus proche voisin. Puisque $n \geq 2$, c'est que la marche est partie à gauche (sinon $T_1 = 1$) et puisque la marche est plus proche voisin, la marche ne peut redevenir strictement positive sans passer par 1.

La deuxième égalité repose sur l'observation suivante : puisque $n \geq 2$, la marche part à gauche, reste strictement négative jusqu'au temps $\nu - 1$, redevient nulle à l'instant ν (pour la première fois depuis l'instant 0) et au temps n devient strictement positive égale à 1 (et pour la première fois!).

La troisième et quatrième égalité sont de simples conséquences de l'indépendance et du caractère équidistribué des X_k .

1b) On a :

$$\begin{aligned} \sum_{n=1}^{\infty} f_1(n)s^n &= qs \sum_{n=2}^{\infty} \left(\sum_{j=1}^{n-1} f_1(j)f_1(n-1-j) \right) s^{n-1} + ps \\ &= qs \sum_{n=1}^{\infty} \left(\sum_{j=0}^n f_1(j)f_1(n-j) \right) s^n + ps \end{aligned}$$

soit

$$F_1(s) = qsF_1(s)^2 + ps$$

On obtient donc une équation du second degré pour $F_1(s)$ et on a :

$$F_1(s) = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2qs}$$

et puisque $F_1(s)$ tend vers 0 quand $s \rightarrow 0$, on a donc :

$$F_1(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2qs}.$$

2) En utilisant le développement en série entière de $\sqrt{1-x}$, on obtient le résultat annoncé dans la question 2. Remarquer qu'il était évident que $\mathbb{P}(T_1 = 2k) = 0$ (on doit faire un nombre impair de pas pour de 0 aller en 1).

Exercice 8.1.5. Le temps de première visite en r

Soit T_r le temps de première visite au point r . On pose $f_r(n) = \mathbb{P}(S_1 \neq r, \dots, S_{n-1} \neq r, S_n = r)$, c'est-à-dire la probabilité que la première visite se produise au n -ième pas. On note sa fonction génératrice $F_r(s) = \sum_{n=1}^{\infty} f_r(n)s^n$.

1) Montrer que $f_r(n) = \sum_{k=1}^{n-1} f_{r-1}(n-k)f_1(k)$ si $r > 1$.

2) En déduire que pour $r \geq 1$,

$$F_r(s) = [F_1(s)]^r.$$

3) Montrer que la probabilité que la marche aléatoire entre dans $\mathbb{N}^* = \{1, 2, \dots, n, \dots\}$ est

$$F_1(1) = \frac{1 - |p - q|}{2q} = \min\left(1, \frac{p}{q}\right).$$

(Cette probabilité est la même que celle de passer par 1, qui est le vestibule de \mathbb{N}^* .)

4) Démontrer, en considérant les probabilités conditionnelles par rapport aux valeurs possibles de X_1 , que $f_0(n) = qf_1(n-1) + pf_{-1}(n-1)$ et que donc $F_0(s) = qsf_1(s) + psf_{-1}(s)$.

5) Montrer que $F_{-1}(s) = \frac{1-(1-4pqs^2)^{\frac{1}{2}}}{2ps}$. Pour cela, on appliquera une *méthode de réflexion* : Soit ω une marche quelconque arrivant en -1 et ω^* la marche symétrique par rapport à 0 , qui arrive donc en 1 . Soit $\mathbb{P}_{p,q}(\pi)$ la probabilité de π . Alors $\mathbb{P}_{p,q}(\pi) = \mathbb{P}_{q,p}(\pi^*)$.

6) En déduire à nouveau $F_0(s)$: par le principe de réflexion on peut donc déduire directement F_0 de F_1 .

Solution

1) On reprendra point par point l'argument qui donnait la relation (8.1.7).

2) On multiplie par s^n la relation du 1) et on somme sur n pour obtenir $F_r(s) = F_{r-1}(s)F_1(s) = F_1(s)^r$.

4) L'argument suggéré conduit à échanger les rôles de p et q dans la formule donnant F_1 pour obtenir F_{-1} .

5) Simple calcul.

8.2 Le principe de réflexion et ses conséquences

[18], 3.10, pages 75 à 82.

Dans la suite, on pose $S_0 = a$, $S_n = a + \sum_{i=1}^n X_i$ où X_i sont des variables de Bernoulli i.i.d. de paramètre p : $\mathbb{P}(X_1 = 1) = p$, $\mathbb{P}(X_1 = -1) = 1 - p = q$. On représente une marche aléatoire par la suite des couples temps-position, $\{(n, S_n), n \geq 0\}$. On note $N_n(a, b)$ le nombre de chemins distincts allant de $(0, a)$ à (n, b) et $N_n^0(a, b)$ le nombre de chemins allant de $(0, a)$ à (n, b) et passant au moins une fois par 0 : de tels chemins contiennent $(k, 0)$ pour un k tel que $0 \leq k \leq n$.

Exercice 8.2.1. 1) Démontrer que

$$\mathbb{P}(S_n = b) = \begin{cases} C_n^{\frac{1}{2}(n+b-a)} p^{\frac{1}{2}(n+b-a)} q^{\frac{1}{2}(n-b+a)} & \text{si } \frac{1}{2}(n+b-a) \in \{0, 1, \dots, n\} \\ 0 & \text{sinon} \end{cases}$$

2) *Le principe de réflexion.* On suppose que $S_0 = a$ et $S_n = b$. Montrer que si $a, b > 0$, alors $N_n^0(a, b) = N_n(-a, b)$. On appliquera le principe de réflexion, qui consiste à associer à tout chemin de $(0, -a)$ à (n, b) passant pour la première fois en 0 à l'instant k le chemin de $(0, a)$ à (n, b) obtenu en changeant en leur opposé tous les points avant k . (Faire un dessin où l'axe des x représente le temps et l'axe des y la position S_n . La réflexion se fait par rapport à l'axe Ox .)

3) Montrer que pour tous $n \in \mathbb{N}$, $a < b$, tels que $0 \leq \frac{1}{2}(n+b-a) \leq n$, on a $N_n(a, b) = C_n^{\frac{1}{2}(n+b-a)}$, et $N_n(a, b) = 0$ si cette condition sur n, a, b n'est pas satisfaite.

4) Montrer le *théorème de scrutin* : Si $b > 0$, alors le nombre de chemins de $(0, 0)$ à (n, b) qui ne repassent pas par 0 est égal à $\frac{b}{n}N_n(0, b)$.

5) On prend $a = 0$. Montrer que

$$\mathbb{P}(S_1 S_2 \dots S_n \neq 0, S_n = b) = \frac{|b|}{n} \mathbb{P}(S_n = b)$$

6) En déduire que

$$\mathbb{P}(S_1 S_2 \dots S_n \neq 0) = \frac{1}{n} \mathbb{E}|S_n|.$$

7) Interpréter ces résultats.

Solution

1) Si $S_n = b$, et s'il y a eu r pas vers la droite et l pas vers la gauche, on a $a + r - l = b$ et $r + l = n$. On trouve $r = \frac{1}{2}(n + b - a)$ et $l = \frac{1}{2}(n - b + a)$. Le nombre de chemins distincts est le nombre de manières de placer les r pas vers la droite parmi les n pas, soit $C_n^r = C_n^{\frac{1}{2}(n+b-a)}$. Tous les chemins ont la même probabilité $p^r q^l$, d'où le résultat.

2) La réflexion définit une bijection entre les chemins de a à b passant par 0 et les chemins de $-a$ à b . Leurs nombres sont donc égaux.

3) Ce raisonnement a déjà été fait au 1).

4) Comme tous les chemins ne repassant pas par 0 vont d'abord en $(1, 1)$, leur nombre est $N_{n-1}(1, b) - N_{n-1}^0(1, b) = N_{n-1}(1, b) - N_{n-1}(-1, b)$, par le 2). Un calcul facile utilisant 3) donne le résultat.

5) Prenons d'abord $b > 0$. Le nombre de chemins dans l'événement considéré est par le théorème de scrutin $\frac{b}{n}N_n(0, b)$. et chaque chemin a $\frac{1}{2}(n + b)$ pas vers la droite et $\frac{1}{2}(n - b)$ pas vers la gauche. Donc

$$\mathbb{P}(S_1 S_2 \dots S_n \neq 0, S_n = b) = \frac{b}{n} N_n(0, b) p^{\frac{1}{2}(n+b)} q^{\frac{1}{2}(n-b)} = \frac{b}{n} \mathbb{P}(S_n = b).$$

Un calcul du même type traite le cas $b < 0$.

6) Il suffit de sommer la relation du 5) sur toutes les valeurs de b .

7) Les formules du 5) et du 1) sont intéressantes pour le problème de la fortune du joueur : elle permettent de calculer la probabilité de fortune ($S_n = b$) sous la condition que le joueur ne se ruine pas ($S_0 \dots S_n \neq 0$).

Exercice 8.2.2. Interprétation du théorème de scrutin Ce théorème ("Ballot theorem") est parfois traduit improprement par "théorème de ballotage". En fait, il s'agit plutôt du théorème de scrutin et il concerne le dépouillement des votes. La question est de savoir : supposons que le candidat A a plus de votes que le candidat B : quelle est la probabilité que cette supériorité soit vérifiée tout au long du dépouillement ?

1) Supposons que le candidat A ait α votes et que le candidat B ait β votes, avec $\alpha > \beta$. Montrer en utilisant le théorème de scrutin que la probabilité que le candidat A ait tout au long du dépouillement un score supérieur à celui de B est égale à $\frac{\alpha - \beta}{\alpha + \beta}$.

2) Interpréter.

Solution

1) Le dépouillement se fait bulletin par bulletin. On pose pour $i = 1, \dots, \alpha + \beta$ $X_i = 1$ si le bulletin est pour A et -1 s'il est pour B . On suppose les X_i i.i.d. On a $S_0 = 0$ et S_n représente l'avantage de A sur B . Les chemins tels que S_k soit toujours différent de 0 pour $k \geq 1$ sont en nombre $N_n^0(0, b) = \frac{b}{n} N_n(0, b)$, et comme tous les chemins arrivant en $\alpha + \beta$ sont équiprobables, la probabilité recherchée est la proportion de chemins ne repassant pas par zéro, soit $\frac{b}{n} = \frac{\alpha - \beta}{\alpha + \beta}$.

2) Remarquons d'abord que le résultat ne dépend pas de la connaissance de p , la probabilité de voter A . Dans les démocraties, α est en général proche de β (1% de différence) et un ordre de grandeur de la probabilité précédente est donc $p = \frac{1}{100}$. Il est donc très probable que le dépouillement soit incertain. Toutefois, on peut se demander si cette incertitude ne se produit pas, en fait, au début du dépouillement, et s'il n'y a pas stabilisation du résultat bien avant la fin du dépouillement. Peut-on utiliser directement la loi des grands nombres, le théorème central limite ou des inégalités de grande déviation pour démontrer une stabilisation des proportions ? La réponse est "non", car nous avons une information a priori supplémentaire, à savoir $S_{\alpha+\beta} = \alpha - \beta$ et par contre nous ignorons la probabilité p de vote pour A .

Exercice 8.2.3. La méthode d'inversion du temps

On va montrer :

Théorème 8.2.1. *La probabilité $f_b(n)$ qu'une marche aléatoire S_n partant de 0 arrive en $b > 0$ pour la première fois au temps n vérifie*

$$f_b(n) = \frac{|b|}{n} \mathbb{P}(S_n = b), \quad n \geq 1.$$

Cette conclusion ressemble beaucoup à celle du théorème de scrutin et s'en déduit grâce à la méthode d'inversion du temps. On associe à la marche aléatoire

$$\{0, S_1, S_2, \dots, S_n\} = \{0, X_1, X_1 + X_2, \dots, \sum_1^n X_i\}$$

la marche aléatoire à rebours

$$\{0, T_1, T_2, \dots, T_n\} = \{0, X_n, X_n + X_{n-1}, \dots, \sum_1^n X_i\}.$$

1) Montrer que les deux marches ont la même loi (quelle que soit la valeur de p).

2) Montrer que la marche directe ne repasse pas par 0 et vérifie $S_n = b$ si et seulement si la marche inverse vérifie $T_n = b$ et que la première visite de T_n en b se produit au temps n (faire un dessin). En déduire le théorème.

Solution

Le fait que les marches aient même loi est immédiat, puisque les X_i sont i.i.d. On $S_n = b$ et $S_1 S_2 \dots S_n \neq 0$ si et seulement si $T_n = b$ et $T_n - T_{n-i} = X_1 + \dots + X_i > 0$ pour tout $i \geq 1$, ce qui veut bien dire que la première visite de T_n en b se fait au temps n .

Exercice 8.2.4. Faire une synthèse des questions posées et des résultats obtenus sur les marches aléatoires : que sait-on ?

8.3 Lois de l'arc sinus pour les marches aléatoires

[18], 3.10, pages 80-82.

Pour comprendre vraiment ce qui est démontré dans cette section, il faut absolument que vous fassiez quelques simulations longues de marches aléatoires et que vous visualisiez la fonction $n \rightarrow S_n$. On observe notamment que la marche aléatoire a tendance à s'éloigner de zéro pour de longues périodes, et qu'il est en particulier assez probable qu'une marche partant de zéro ne revienne en zéro qu'après un temps assez long. Ceci ajoute une explication de plus au mystère de l'existence des casinos, qui peut se formuler ainsi : comment peut-on aller dans un endroit où l'on est statistiquement sûrs de perdre son argent ? La réponse est là encore dans la durée du plaisir obtenu à peu de frais : une mise minimale donne une chance importante de jouer longtemps.

Exercice 8.3.1. Soit une marche aléatoire $S_n = X_1 + \dots + X_n$ où les X_i sont i.i.d. et Bernoulli. On pose, pour indiquer le point de départ, $S_0 = 0$. On suppose que $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$. On va montrer le théorème suivant.

Théorème 8.3.1. La probabilité que la dernière visite à 0 avant (et y compris) le temps $2n$ se produise au temps $2k$ est $\mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2n-2k} = 0)$.

Tout d'abord, interprétons cette relation.

1) On appelle $\alpha_{2n}(2k)$ la probabilité précédente. On a donc $\alpha_{2n-2k} = u_{2k}u_{2n-2k}$, en posant $u_{2k} = \mathbb{P}(S_{2k} = 0)$. On rappelle que $\mathbb{P}(S_{2k} = 0) = C_{2k}^k 2^{-2k}$. Montrer en utilisant la formule de Stirling $n! \approx n^n e^{-n} \sqrt{2\pi n}$ que

$$\alpha_{2n}(2k) \approx \frac{1}{\pi \sqrt{k(n-k)}} = \frac{1}{\pi \sqrt{k(n-k)}} \left(1 + \varepsilon\left(\frac{1}{k}, \frac{1}{n-k}\right)\right),$$

cette équivalence étant valide quand $n \rightarrow \infty$, $k \rightarrow \infty$, et $n-k \rightarrow \infty$ simultanément.

2) On note T_{2n} le temps de la dernière visite en 0 avant le temps $2n$. Montrer que

$$\mathbb{P}(\varepsilon n \leq T_{2n} \leq 2xn) \approx \sum_{\varepsilon n \leq k \leq xn} \frac{1}{\pi \sqrt{k(n-k)}} \approx \int_{\varepsilon n}^{xn} \frac{1}{\pi \sqrt{u(n-u)}} du = \frac{2}{\pi} \arcsin \sqrt{x} - \frac{2}{\pi} \arcsin \sqrt{\varepsilon}.$$

En déduire que

$$\liminf_{n \rightarrow +\infty} \mathbb{P}(T_{2n} \leq 2xn) \geq \frac{2}{\pi} \arcsin \sqrt{x}.$$

3) En déduire qu'il y a, quand n devient grand, plus d'une chance sur deux que la marche aléatoire ne repasse pas par 0 entre n et $2n$ et plus d'une chance sur 5 qu'elle ne repasse pas par 0 entre $\frac{n}{5}$ et $2n$.

4) Montrer la loi de l'arcsinus :

$$\lim_n \mathbb{P}(T_{2n} \leq 2xn) = \frac{2}{\pi} \arcsin \sqrt{x}.$$

Solution

1), 2) et 3) viennent tous seuls. Montrons juste le 4). On a

$$\lim_n \mathbb{P}(2\varepsilon n \leq T_{2n} \leq 2xn) = \frac{2}{\pi}(\arcsin \sqrt{x} - \arcsin \sqrt{\varepsilon}) \rightarrow 1,$$

quand $x \rightarrow 1$ et $\varepsilon \rightarrow 0$. Donc, pour tout ε' il existe η tel que pour $x \geq 1 - \eta$ et $\varepsilon \leq \eta$ on ait

$$\lim_n (\mathbb{P}(T_{2n} \leq 2\varepsilon n) + \mathbb{P}(T_{2n} \geq 2xn)) \leq \varepsilon' \text{ et donc}$$

$$\limsup_n \mathbb{P}(T_{2n} \leq 2\varepsilon n) \leq \varepsilon'.$$

On en déduit que

$$\begin{aligned} & \limsup_n |\mathbb{P}(T_{2n} \leq 2xn) - \frac{2}{\pi} \arcsin \sqrt{x}| \leq \\ & \limsup_n |\mathbb{P}(2\varepsilon n \leq T_{2n} \leq 2xn) - \frac{2}{\pi} \arcsin \sqrt{x} + \frac{2}{\pi} \arcsin \sqrt{\varepsilon} - \frac{2}{\pi} \arcsin \sqrt{\varepsilon} + \mathbb{P}(T_{2n} \leq 2\varepsilon n)| \\ & \leq \limsup_n |\mathbb{P}(2\varepsilon n \leq T_{2n} \leq 2xn) - \frac{2}{\pi} \arcsin \sqrt{x} + \frac{2}{\pi} \arcsin \sqrt{\varepsilon}| + \frac{2}{\pi} \arcsin \sqrt{\varepsilon} + \varepsilon' \\ & = \frac{2}{\pi} \arcsin \sqrt{\varepsilon} + \varepsilon'. \end{aligned}$$

Cette dernière quantité étant arbitrairement petite, on obtient le résultat.

Exercice 8.3.2. Preuve de la loi en arc sinus pour les retours à l'origine.

1) Montrer que $\alpha_{2n}(2k) = \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_1 S_2 \dots S_{2n-2k} \neq 0)$. On pose dans la suite $m = 2n - 2k$.

2) On rappelle que

$$\mathbb{P}(S_1 S_2 \dots S_n \neq 0, S_n = b) = \frac{|b|}{n} \mathbb{P}(S_n = b)$$

et que $\mathbb{P}(S_{2m} = 2j) = C_{2m}^{m+j} \left(\frac{1}{2}\right)^{2m}$. En déduire que

$$\mathbb{P}(S_1 S_2 \dots S_{2m} \neq 0) = \mathbb{P}(S_{2m} = 0).$$

Indication : vérifier et utiliser $\frac{2k}{2m} C_{2m}^{m+k} = C_{2m-1}^{m+k-1} - C_{2m-1}^{m+k}$.

Solution

1)

$$\alpha_{2n}(2k) = \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2k+1} S_{2k+2} \dots S_{2n} \neq 0 | S_{2k} = 0) \tag{8.3.1}$$

$$= \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_1 S_2 \dots S_{2n-2k} \neq 0). \tag{8.3.2}$$

2) Par la loi des probabilités disjointes,

$$\mathbb{P}(S_1 S_2 \dots S_{2m} \neq 0) = 2 \sum_{k=1}^m \frac{2k}{2m} \mathbb{P}(S_{2m} = 2k) = 2 \sum_{k=1}^m \frac{2k}{2m} C_{2m}^{m+k} \left(\frac{1}{2}\right)^{2m} \quad (8.3.3)$$

$$= 2 \left(\frac{1}{2}\right)^{2m} \sum_{k=1}^m [C_{2m-1}^{m+k-1} - C_{2m-1}^{m+k}] \quad (8.3.4)$$

$$= 2 \left(\frac{1}{2}\right)^{2m} C_{2m-1}^m \quad (8.3.5)$$

$$= C_{2m}^m \left(\frac{1}{2}\right)^{2m} = \mathbb{P}(S_{2m} = 0). \quad (8.3.6)$$

Exercice 8.3.3. Loi de l'arc sinus pour les périodes de gain On dit que la marche aléatoire passe l'intervalle $(k, k+1)$ à droite de l'origine si $S_k > 0$ ou $S_{k+1} > 0$, ou les deux. On va montrer le théorème suivant :

Théorème 8.3.2. *La probabilité qu'une marche aléatoire symétrique partant de 0 passe exactement $2k$ intervalles de temps à droite de l'origine jusqu'au temps $2n$ est égale à $\mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2n-2k} = 0)$.*

- 1) Interpréter le résultat dans les mêmes lignes que la loi arc sinus pour les retours à l'origine.
- 2) Lire la solution page 81-82 du Grimmett. Attention, cette solution donne les idées mais a quelques petites erreurs à corriger.

Solution

Par les mêmes équivalents que pour la loi de l'arc sinus pour les retours à l'origine, on obtient que la probabilité que la marche passe $2xn$ unités de temps ou moins à droite de l'origine est de l'ordre de $\frac{2}{\pi} \arcsin \sqrt{x}$. Si on prend par exemple $x = \frac{1}{5}$, cette probabilité est de l'ordre de $\frac{1}{5}$. Du point de vue du jeu de pile ou face, il y a donc une chance sur cinq pour que l'un des joueurs soit gagnant pendant plus des quatre cinquièmes du temps.

Chapitre 9

Fonctions génératrices, poules et grands noms

9.1 Espérance conditionnelle, poules et oeufs

Source : [18], pp. 67-68

Exercice 9.1.1. Définition de l'espérance conditionnelle

Définition 9.1.1. Soient X et Y deux variables discrètes sur $(\Omega, \mathcal{F}, \mathbb{P})$. Pour tout x tel que $\mathbb{P}(X = x) > 0$, la distribution conditionnelle de Y sachant $X = x$, notée $F_{Y|X}(\cdot|x)$, est définie par

$$F_{Y|X}(y|x) = \mathbb{P}(Y \leq y | X = x)$$

et on pose $F_{Y|X}(y|x) = 0$ si $\mathbb{P}(X = x) = 0$. La loi conditionnelle de Y sachant $X = x$, notée $f_{Y|X}(\cdot|x)$ est définie par

$$f_{Y|X}(y|x) = \mathbb{P}(Y = y | X = x).$$

1) On note $f_X(x) = \mathbb{P}(X = x)$ et de même $f_{X,Y}(x, y) = \mathbb{P}(X = x; Y = y)$ la densité jointe de X et Y . Vérifier que $f_{Y|X} = \frac{f_{X,Y}}{f_X}$ et que X et Y sont indépendantes si et seulement si $f_{Y|X} = f_Y$.

Définition 9.1.2. Soit Y une variable aléatoire discrète sommable. On appelle espérance conditionnelle de Y sachant $X = x$ la fonction $\psi(x)$ notée et définie par :

$$\psi(x) = \mathbb{E}(Y|X = x) = \sum_y y f_{Y|X}(y|x) = \sum_y \mathbb{P}(Y = y | X = x).$$

L'espérance conditionnelle de Y sachant X est $\psi(X)$.

2) (L'espérance de l'espérance conditionnelle est encore l'espérance). Montrer que $\psi(X) = \mathbb{E}(Y|X)$ est une variable aléatoire et plus précisément qu'elle est X -mesurable. Démontrer que

$$\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}(Y). \tag{9.1.1}$$

On peut réécrire cette relation

$$\mathbb{E}(Y) = \sum_x \mathbb{E}(Y|X = x) \mathbb{P}(X = x).$$

Solution

2) La fonction $x \rightarrow \mathbb{P}(Y = y | X = x)$ est constante sur chaque ensemble $\{X = x\}$. Donc elle est X -mesurable. La fonction $\psi(x)$ est une limite presque sûre de fonctions X -mesurables et est donc X -mesurable. Comme Y est sommable on a

$$\mathbb{E}Y = \sum_y y \mathbb{P}(Y = y) = \sum_{x,y} y \mathbb{P}(Y = y, X = x) = \sum_x \sum_y \mathbb{P}(Y = y | X = x) \mathbb{P}(X = x) = \mathbb{E}[\mathbb{E}(Y | X)].$$

Attention : justifier soigneusement les différentes étapes avec Fubini !

Exercice 9.1.2. La limite de Poisson. On considère N variables de Bernoulli indépendantes X_n d'ordre 1 et de paramètre $\mathbb{P}(X_n = 1) = p$ très petit, $\mathbb{P}(X_n = 0) = 1 - p$.

1) Donner la loi (dite binômiale) de $S_N = X_1 + \dots + X_N$, c'est-à-dire calculer $\mathbb{P}(S_N = k)$. Montrer que quand $N \rightarrow \infty$, $p \rightarrow 0$ avec $Np \rightarrow \lambda$, on a $\mathbb{P}(S_N = k) \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$. Cette distribution sur \mathbb{N} s'appelle loi de Poisson. Vérifier que c'est une distribution de probabilité. Montrer que la loi binômiale converge étroitement vers la loi de Poisson.

2) Donner une interprétation et discuter soigneusement des exemples. (Nombre de clients arrivant à une file d'attente dans un petit intervalle de temps donné, désintégration atomique, arrivée à maturité d'oeufs de poisson,...)

Solution

1) La loi binômiale est

$$C_N^k (1-p)^{N-k} p^k = C_N^k \left(1 - \frac{\lambda}{N}\right)^{N-k} \frac{\lambda^k}{k!} \frac{N!}{N^k (N-k)!} \sim e^{-\lambda} \frac{\lambda^k}{k!}$$

quand $N \rightarrow \infty$. Il est immédiat de vérifier que la somme de ces derniers coefficients est 1. Ceci prouve qu'il y a "conservation de la masse". Donc la loi binômiale, dont nous venons de voir qu'elle converge vaguement vers la loi de Poisson, converge aussi étroitement (Proposition 7.1.2).

2) Si par exemple un très grand nombre N de voyageurs a une faible probabilité p de se diriger dans un petit intervalle de temps donné vers un certain guichet, le nombre de voyageurs arrivant au guichet suit une loi de Poisson. L'exemple de la désintégration atomique est l'exemple idéal, les atomes ayant tous la même probabilité de se désintégrer. Le nombre de descendants arrivant à maturité d'un poisson donné aussi (un très grand nombre d'oeufs, une très faible probabilité de survie). L'exemple de la poule ci-dessous est plus discutable : peut-être vaut-il mieux la remplacer par un saumon sauvage...

Exercice 9.1.3. Application : une poule pond N oeufs, N suivant une loi de Poisson de paramètre λ , c'est-à-dire $f_N(n) = \mathbb{P}(N = n) = e^{-\lambda} \frac{\lambda^n}{n!}$. Chaque oeuf éclôt avec probabilité p , indépendamment des autres oeufs. On note $q = 1 - p$. Soit K le nombre de poussins. Notre but est de calculer, puis d'interpréter les valeurs de $\mathbb{E}(K|N)$, $\mathbb{E}K$ et $\mathbb{E}(N|K)$.

1) Montrer que $f_{K|N}(k|n) = C_n^k p^k (1-p)^{n-k}$ et que $\mathbb{E}(K|N) = pN$. En déduire que $\mathbb{E}(K) = p\lambda$. Interpréter.

2) Si vous ne l'avez jamais vue, montrer la règle de Bayes (disposer d'un moyen mnémotechnique pour s'en souvenir).

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

3) En utilisant la règle de Bayes, montrer que si $n \geq k$,

$$f_{N|K}(n|k) = \mathbb{P}(N = n|K = k) = \frac{(q\lambda)^{n-k}}{(n-k)!} e^{-q\lambda}.$$

Conclure que $\mathbb{E}(N|K) = K + q\lambda$. Interpréter.

Solution

1) Il faut bien sûr supposer que le nombre d'oeufs N et la variable de Bernoulli d'éclosion de chacun des oeufs sont des v.a. indépendantes! Grâce à cette indépendance :

$$\mathbb{P}(K = k | N = n) = \frac{\mathbb{P}(K = k, N = n)}{\mathbb{P}(N = n)} = \frac{\mathbb{P}(\sum_{i=1}^n X_i = k, N = n)}{\mathbb{P}(N = n)} = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k}.$$

On a

$$\begin{aligned} \mathbb{E}(K | N = n) &= \sum_k k \frac{\mathbb{P}(K = k, N = n)}{\mathbb{P}(N = n)} = \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} \\ &= pn \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} = pn(p + (1-p))^{n-1} = pn. \end{aligned}$$

2) On a

$$\begin{aligned} \mathbb{P}(N = n | K = k) &= \frac{\mathbb{P}(K = k | N = n) \mathbb{P}(N = n)}{\mathbb{P}(K = k)} = \\ &= \frac{C_n^k p^k (1-p)^{n-k} \left(\frac{\lambda^n}{n!}\right) e^{-\lambda}}{\sum_{m \geq k} C_m^k p^k (1-p)^{m-k} \left(\frac{\lambda^m}{m!}\right) e^{-\lambda}} = \frac{(q\lambda)^{n-k}}{(n-k)!} e^{-q\lambda}, \end{aligned}$$

car

$$\sum_{m \geq k} \frac{m!}{k!(m-k)!} p^k (1-p)^{m-k} \frac{\lambda^m}{m!} = \lambda^k \frac{p^k}{k!} \sum_l \frac{(\lambda(1-p))^l}{l!} = \frac{\lambda^k p^k}{k!} e^{q\lambda}.$$

Donc

$$\mathbb{E}(N|K = k) = \sum_{n \geq k} n \frac{(q\lambda)^{n-k}}{(n-k)!} e^{-q\lambda} = \sum_{m=0}^{\infty} (k+m) \frac{(q\lambda)^m}{m!} e^{-q\lambda} = k + q\lambda.$$

Donc $\mathbb{E}(N|K) = K + q\lambda$. Ce n'est pas difficile à interpréter. On a $\mathbb{E}N = \lambda$, $\mathbb{E}K = p\lambda$. A titre de vérification des résultats, remarquer que

$$\lambda = \mathbb{E}N = \mathbb{E}(\mathbb{E}(N|K)) = \mathbb{E}(K + q\lambda) = p\lambda + q\lambda = \lambda.$$

L'espérance du nombre d'oeufs sachant le nombre d'oeufs éclos est égale à la somme du nombre, connu, d'oeufs éclos, K , plus l'espérance du nombre d'oeufs non éclos, soit $q\lambda$. Formellement, on remplace $p\lambda$ par K dans la formule $\lambda = p\lambda + q\lambda$. Voir aussi le théorème 9.2.1 plus loin qui prouve le résultat plus précis : K suit une loi de Poisson de paramètre λp .

Exercice 9.1.4. Formule générale caractérisant l'espérance conditionnelle

Montrer que $Z = \mathbb{E}(Y|X)$ vérifie

$$\mathbb{E}(Zg(X)) = \mathbb{E}(Yg(X)) \quad (9.1.2)$$

pour toute fonction g telle que les deux espérances existent. Montrer qu'il y a une unique fonction $Z = \psi(X)$ vérifiant (9.1.2) et que c'est donc $Z = \mathbb{E}(Y|X)$.

Solution

Par hypothèse $\sum_x |\psi(x)g(x)|f_X(x) < \infty$ et $\sum_{x,y} |yg(x)|f_{X,Y}(x,y) < \infty$. Cette dernière hypothèse permet d'appliquer Fubini et de justifier tous les calculs qui suivent :

On a $\mathbb{E}(Y|X) = \psi(X)$ avec $\psi(x) = \sum_y yf_{Y|X}(y|x)$. Donc

$$\begin{aligned} \mathbb{E}(\mathbb{E}(Y|X)g(X)) &= \sum_x \psi(x)g(x)f_X(x) = \\ &= \sum_{x,y} yf_{Y|X}(y|x)g(x)f_X(x) = \sum_{x,y} yg(x)f_{X,Y}(x,y) = \mathbb{E}(Yg(X)). \end{aligned}$$

Unicité : soit $Z = \psi(X)$ vérifiant (9.1.2). On choisit $g(X) = \mathbb{1}_x(X)$. Cela donne

$$\begin{aligned} \psi(x)\mathbb{P}(X = x) &= \sum_{y,z} y\mathbb{1}_x(z)\mathbb{P}(X = z; Y = y) \\ &= \sum_y y\mathbb{P}(X = x; Y = y) \\ &= \sum_y y\mathbb{P}(Y = y|X = x)\mathbb{P}(X = x) \end{aligned}$$

Donc $\psi(x) = \sum_y y\mathbb{P}(Y = y|X = x)$ si $\mathbb{P}(X = x) \neq 0$. Donc Z et $\psi(X)$ sont égales \mathbb{P}_X -p.s.

9.2 Fonctions génératrices

Source : [18], pp. 148-154

Si X est une variable aléatoire à valeurs dans \mathbb{N} , la fonction génératrice de X est définie comme

$$G_X(s) = \mathbb{E}(s^X) = \sum_{k=0}^{\infty} p_k s^k$$

où $p_k = \mathbb{P}(X = k)$ et où on adopte la convention $0^0 = 1$.

Exercice 9.2.1. 0) Rayon de convergence de la série ?

1) Montrer que la fonction génératrice de X caractérise complètement sa loi.

2) Montrer que G_X est C^∞ sur $(-1; 1)$ et que si X est intégrable alors

$$\lim_{s \rightarrow 1^-} G'_X(s) = \mathbb{E}(X)$$

Donner un résultat analogue pour la variance. Montrer de même que $\mathbb{E}(X(X-1)\dots(X-k+1)) = G^{(k)}(1)$. On utilisera et montrera le théorème d'Abel : Soit $a_n \geq 0$ et $G(s) = \sum_{n \in \mathbb{N}} a_n s^n$. Si $G(s)$ est fini pour $s < 1$, alors $\lim_{s \rightarrow 1} G(s) = \sum_{n \in \mathbb{N}} a_n$.

3) Montrer que si X et Y sont deux variables aléatoires à valeurs entières indépendantes alors la fonction génératrice de $X + Y$ est le produit de G_X et de G_Y . Réciproque ?

4) Calculer les fonctions génératrices des distributions suivantes : v.a. constante $X = c$, variable de Bernoulli de paramètre p , distribution géométrique ($\mathbb{P}(X = k) = p(1-p)^{k-1}$ pour $k \geq 1$), distribution de Poisson, distribution binomiale $\mathcal{B}(n, p)$. Vérifier sans calcul que la somme de deux variables indépendantes $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$ est $\mathcal{B}(n+m, p)$.

5) On appelle fonction génératrice jointe de deux v.a. X_1 et X_2 à valeurs dans \mathbb{N} la fonction

$$G_{X_1, X_2}(s_1, s_2) = \mathbb{E}(s_1^{X_1} s_2^{X_2}).$$

Montrer que X_1 et X_2 sont indépendantes si et seulement si

$$G_{X_1, X_2}(s_1, s_2) = G_{X_1}(s_1)G_{X_2}(s_2)$$

pour tous s_1 et s_2 .

Solution

0) Le rayon de convergence de la série est supérieur ou égal à 1.

1) Remarquer que l'on a $p_k = \frac{1}{k!} G_X^{(k)}(0)$.

2) G_X est une série entière et donc en tant que telle C^∞ sur son disque de convergence. D'autre part, pour tout $s \in (-1, 1)$, on a (en utilisant le théorème d'Abel)

$$G'_X(s) = \sum_{k=1}^{\infty} k p_k s^{k-1} \underset{s \rightarrow 1}{\uparrow} \sum_{k=1}^{\infty} k p_k = \mathbb{E}(X)$$

De même, on montre que si X a un moment d'ordre 2

$$Var(X) = \lim_{s \rightarrow 1} \left[G_X^{(2)}(s) + G_X^{(1)}(s) - [G_X^{(1)}(s)]^2 \right]$$

3) $G_{X+Y}(s) = \mathbb{E}(s^X s^Y) = \mathbb{E}(s^X) \mathbb{E}(s^Y) = G_X(s) G_Y(s)$ par indépendance de X et Y . La réciproque est fautive (comme elle l'est pour les fonctions caractéristiques). Ce qui est en revanche vrai, c'est que si la fonction génératrice du couple (X, Y) (définie par $G_{(X, Y)}(s_1, s_2) = \mathbb{E}(s_1^X s_2^Y)$) s'écrit comme le produit direct de G_X et G_Y (i.e. $G_{(X, Y)}(s_1, s_2) = G_X(s_1) G_Y(s_2)$) alors les deux variables X et Y sont indépendantes (voir la question 5)).

4) On obtient successivement $G(s) = s^c$ (variable constante), $G(s) = (1-p) + ps$ (variable de Bernoulli), $G(s) = \sum_{k=1}^{\infty} s^k p(1-p)^{k-1} = \frac{ps}{1-s(1-p)}$ (variable géométrique de paramètre p), $G(s) =$

$\sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda(s-1)}$ (variable de Poisson), $G(s) = (q + ps)^n$ (variable binomiale $\mathcal{B}(n, p)$, $q = 1 - p$.) On peut voir que la somme d'une $\mathcal{B}(n, p)$ et d'une $\mathcal{B}(m, p)$ indépendantes est une $\mathcal{B}(n+m, p)$ de deux manières : d'une part on peut interpréter la somme comme la somme de $n+m$ variables de Bernoulli indépendantes. Mais on peut simplement multiplier les fonctions génératrices en utilisant le 3)

5) Si X_1 et X_2 sont indépendantes, $s_1^{X_1}$ et $s_2^{X_2}$ aussi et on a donc $\mathbb{E}(s_1^{X_1} s_2^{X_2}) = \mathbb{E}(s_1^{X_1}) \mathbb{E}(s_2^{X_2})$. Réciproquement, si cette relation est vraie, on a

$$\sum_{m,n} \mathbb{P}(X_1 = m; X_2 = n) s_1^m s_2^n = \left(\sum_m \mathbb{P}(X_1 = m) s_1^m \right) \left(\sum_n \mathbb{P}(X_2 = n) s_2^n \right).$$

Il suffit de développer le deuxième membre et d'identifier terme à terme pour déduire que $\mathbb{P}(X_1 = m; X_2 = n) = \mathbb{P}(X_1 = m) \mathbb{P}(X_2 = n)$.

Exercice 9.2.2. Poules et oeufs 2

On va montrer le théorème suivant :

Théorème 9.2.1. *Si X_1, X_2, \dots est une suite de v.a.i.i.d. de fonction génératrice commune G_X et si N est une variable aléatoire à valeurs dans \mathbb{N} indépendante des X_i et de fonction génératrice G_N , alors $S_N = X_1 + X_2 + \dots + X_N$ a pour fonction génératrice*

$$G_{S_N}(s) = G_N(G_X(s)).$$

1) Moyen mnémotechnique : pour se souvenir de l'ordre dans lequel les fonctions sont composées, remarquer que si N est déterministe ($\mathbb{P}(N = n) = 1$) alors on doit retrouver la formule donnant la fonction génératrice de n v.a.i.i.d. soit $G_{S_N}(s) = G_X(s)^n$, et l'on a $G_N(s) = s^n$.

2) Démonstration : appliquer la formule de l'espérance conditionnelle (9.1.1) $\mathbb{E}(s^{S_N}) = \mathbb{E}(\mathbb{E}(s^{S_N} | N))$.

3) Une poule pond N oeufs, et N est une distribution de Poisson de paramètre λ . Chaque oeuf éclôt avec probabilité p et les éclosions sont des événements indépendants. Soit K le nombre de poussins. Calculer la fonction génératrice de K et en déduire que K suit une loi de Poisson de paramètre λp .

Solution

2) Par la formule de l'espérance conditionnelle, on a

$$\begin{aligned} G_{S_N}(s) &= \sum_n \mathbb{E}(s^{S_N} | N = n) \mathbb{P}(N = n) = \sum_n \mathbb{E}(s^{X_1 + \dots + X_n}) \mathbb{P}(N = n) = \\ &= \sum_n \mathbb{E}(s^{X_1}) \dots \mathbb{E}(s^{X_n}) \mathbb{P}(N = n) = \\ &= \sum_n G_X(s)^n \mathbb{P}(N = n) = G_N(G_X(s)). \end{aligned}$$

L'avant dernière ligne s'obtient par indépendance des X_i .

3) On a $G_X(s) = q + ps$, $G_N(s) = \sum_{n=0}^{\infty} s^n \frac{\lambda^n}{n!} e^{-\lambda} = e^{\lambda(s-1)}$. Donc $G_K(s) = G_N(G_X(s)) = e^{\lambda(q+ps-1)} = e^{\lambda p(s-1)}$. K suit donc une loi de Poisson de paramètre λp .

9.3 Processus de branchement : l'extinction des grands noms

Source : [18], 5.4 pp. 171-174

Au dix-neuvième siècle, en 1873 exactement, H.W. Watson répond à une question posée par F. Galton la même année. Il s'agit d'expliquer la tragique et progressive disparition des grands noms de l'aristocratie anglaise. Des noms des grands personnages de Shakespeare, bien peu ont laissé descendance : où sont maintenant les Gloucester, et les Buckingham ? Les processus de Galton-Watson, appelés aussi processus de branchement, modélisent la reproduction et permettent de calculer la probabilité d'extinction. On suppose qu'une population évolue par générations et que Z_n désigne le nombre d'individus de la n -ième génération. Chaque membre de la n -ième génération donne le jour à une famille de membres de la $n + 1$ -ième génération et la taille de cette famille est une variable aléatoire. On fait les hypothèses suivantes :

- (a) Les tailles X_i des familles de tous les individus sont des variables indépendantes
- (b) Toutes ces tailles X_i ont même loi f et pour fonction génératrice G .

On supposera dans la suite que $Z_0 = 1$. On appelle $G_n(s) = \mathbb{E}(s^{Z_n})$ la fonction génératrice de Z_n . Notre but est de montrer le théorème suivant et ses conséquences.

Théorème 9.3.1. On a $G_n(s) = G(G(\dots(G(s))\dots))$, l'itérée n fois de G .

Exercice 9.3.1. 1) Démontrer le théorème 9.3.1.

2) On pose $\mu_n = EZ_1$ et $\sigma_n^2 = \text{var}(Z_1)$. En dérivant une fois, puis deux fois en $s = 1$ la relation $G_n(s) = G(G_{n-1}(s))$, montrer que $\mathbb{E}(Z_n) = \mu^n$, $\text{var}(Z_n) = n\sigma^2$ si $\mu = 1$,

$$\text{var}(Z_n) = \frac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1} \quad \text{si } \mu \neq 1.$$

3) On suppose que les tailles de familles ont la distribution géométrique $\mathbb{P}(X = k) = qp^k$. Avez-vous une interprétation pour cette hypothèse ? Vérifier que $G(s) = q(1 - ps)^{-1}$ et, par récurrence, que

$$G_n(s) = \frac{n - (n-1)s}{n + 1 - ns} \quad \text{si } p = q = \frac{1}{2},$$

$$G_n(s) = \frac{q(p^n - q^n - ps(p^{n-1} - q^{n-1}))}{p^{n+1} - q^{n+1} - ps(p^n - q^n)} \quad \text{si } p \neq q.$$

4) En déduire que

$$\mathbb{P}(Z_n = 0) = \frac{n}{n+1} \quad \text{si } p = q = \frac{1}{2},$$

$$\mathbb{P}(Z_n = 0) = \frac{q(p^n - q^n)}{p^{n+1} - q^{n+1}} \quad \text{si } p \neq q.$$

5) En conclure que, sous les hypothèses précédentes, la probabilité d'extinction de la population des descendants vérifie

$$\mathbb{P}(\text{extinction}) = 1 \quad \text{si } p \leq q, \quad \mathbb{P}(\text{extinction}) = \frac{q}{p} \quad \text{si } p > q.$$

Interpréter ce résultat !

6) Le résultat précédent est valide en général comme le montre le théorème suivant

Théorème 9.3.2. *Quand $n \rightarrow \infty$, $\mathbb{P}(Z_n = 0) \rightarrow \eta = \mathbb{P}(\text{extinction})$ et η peut être calculée comme la plus petite racine positive de l'équation $s = G(s)$. De plus, $\eta = 1$ si $\mu < 1$ et $\eta < 1$ si $\mu > 1$. Si $\mu = 1$, alors $\eta = 1$ du moment que la variance de taille de famille est positive.*

Interpréter ce résultat ! La démonstration n'est pas difficile ([18], p.173). La lire attentivement.

Solution

1) C'est très simple, mais la démonstration du Grimmett-Stirzaker est pourtant un peu confuse. On a $Z_n = X_1 + \dots + X_{Z_{n-1}}$, où X_i est la taille de la famille du i -ème individu de la génération $n-1$. Il s'agit d'une somme d'un nombre aléatoire de v.a.i.i.d. de fonctions génératrices G , ce nombre aléatoire Z_{n-1} ayant une loi G_n . Donc par le théorème 9.2.1, on a $G_n = G_{n-1} \circ G$ et on conclut parce que $G_1 = G$.

2) On rappelle que si X est une v.a. de fonction génératrice G , alors $\mathbb{E}X = G'(1)$ et $\text{var}(X) = G''(1) + G'(1) - G'(1)^2$. En dérivant comme indiqué, on a $G'_n(1) = G'(1)G'_{n-1}(1)$ et donc $\mu_n = \mu\mu_{n-1}$, soit $\mu_n = \mu^n$. En dérivant deux fois, on obtient $G''_n(s) = G''(G_{n-1}(s))G'_{n-1}(s)^2 + G'(G_{n-1}(s))G''_{n-1}(s)$, soit $G''_n(1) = G''(1)\mu^{2(n-1)} + \mu G''_{n-1}(1)$. On en tire la relation de récurrence $\sigma_n^2 = \sigma^2 \mu^{2(n-1)} + \mu \sigma_{n-1}^2$ qui donne bien les résultats annoncés.

3) Simple calcul.

4) On a $\mathbb{P}(Z_n = 0) = G_n(0)$, ce qui donne tout de suite le résultat annoncé en utilisant la question précédente. On définit l'évènement "extinction" comme l'union (croissante) des évènements "extinction au temps n ". Par la propriété fondamentale des mesures positives, on a donc $\mathbb{P}(\text{extinction}) = \lim \mathbb{P}(Z_n = 0)$.

Passons à l'interprétation, du point de vue de Galton et Watson : si $p \leq q$, on a $p \leq \frac{1}{2}$ et donc la distribution des tailles décroît très vite. De plus on vérifie facilement que l'espérance du nombre de descendants par individu est $G'(1) = \frac{p}{1-p}$. Donc si $p < \frac{1}{2}$, cette espérance est inférieure à 1. Dans ce cas, il y a extinction ! Si $p \geq \frac{1}{2}$, l'espérance du nombre de descendants est supérieure à 1 et tend vers l'infini quand $p \rightarrow 1$. Dans ce cas, il y a une probabilité de "survie du nom". Mais la probabilité d'extinction reste importante puisque elle est égale à $\frac{q}{p}$. Il peut y avoir extinction même dans une famille prolifique et les grands noms, donc, peuvent disparaître.

Chapitre 10

Fonctions caractéristiques et théorèmes de Lévy

Bien que l'on présente souvent la transformée de Fourier de manière autonome en probabilités, l'essentiel de la théorie, et en particulier les théorèmes d'inversion, sont plus complètement décrits grâce à la théorie des distributions. Voir par exemple [4]. Dans ce T.D. nous suivrons essentiellement [26] (chapitre 12) car la fonction caractéristique y est traitée pour des v.a. à valeurs dans \mathbb{R}^d . Pour un traitement très élégant et rapide dans le cas de v.a. réelles, voir le chapitre 16 de [30]. On a besoin de v.a. vectorielles à cause, notamment, du traitement des vecteurs gaussiens.

10.1 Généralités

Définition 10.1.1. Soit μ une mesure bornée sur \mathbb{R}^d et X une variable aléatoire de loi $\mathbb{P}_X = \mu$. On appelle transformée de Fourier de μ l'application de \mathbb{R}^d dans \mathbb{C} définie par

$$(\mathcal{F}\mu)(t) = \hat{\mu}(t) = \mathbb{E}(e^{iX \cdot t}) = \int_{\mathbb{R}^d} e^{ix \cdot t} d\mu(x).$$

(Ici, $x \cdot t$ est le produit scalaire usuel de x et $t \in \mathbb{R}^d$)

On appelle fonction caractéristique d'une variable aléatoire X la transformée de Fourier de sa loi $\mu = \mathbb{P}_X$. Elle est notée φ_X .

Exercice 10.1.1. Propriétés élémentaires

Montrer que

- 1) $\hat{\mu}(0) = \mu(\mathbb{R}^d)$, $\varphi_X(0) = 1$.
- 2) $\forall t \in \mathbb{R}^d$, $|\hat{\mu}(t)| \leq \mu(\mathbb{R}^d)$, $|\varphi_X(t)| \leq 1$.
- 3) $\hat{\mu}(-t) = \overline{\hat{\mu}(t)}$, $\varphi_X(-t) = \overline{\varphi_X(t)}$.
- 4) Soient $A \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^k)$ et $b \in \mathbb{R}^k$, alors

$$\varphi_{AX+b}(t) = \varphi_X(A^*t)e^{ib \cdot t},$$

où A^* désigne l'adjoint de A .

5) Les fonctions $\hat{\mu}$ et φ_X sont uniformément continues sur \mathbb{R}^d . *Indication : Fixer $\varepsilon > 0$ et choisir n tel que $\mu(B(0, n)^c) \leq \varepsilon$, montrer et utiliser l'inégalité $|e^{ix \cdot u} - e^{ix \cdot t}| \leq \|u - t\| \|x\|$*

Exercice 10.1.2. Exemple très important : la gaussienne

On pose

$$g_\sigma(x) = \frac{1}{(\sigma\sqrt{2\pi})^d} e^{-\frac{\|x\|^2}{2\sigma^2}}.$$

Vous devez savoir démontrer sans aucune hésitation les relations suivantes :

(a) La fonction g_d est une densité de probabilité sur \mathbb{R}^d . Se ramener au cas de la dimension 1 et se souvenir comment l'on montre que $\int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi}$.

(c) Pour tout $f \in C_b(\mathbb{R}^d)$, $\lim_{\sigma \rightarrow 0} (f * g_\sigma)(x) = f(x)$. Pour cela, utiliser le théorème de Lebesgue après changement de variable $y = \sigma z$.

(d) $\hat{g}_1(t) = e^{-\frac{\|t\|^2}{2}} = (\sqrt{2\pi})^d g_1(t)$

Pour montrer cette dernière formule riche en argument techniques, il est recommandé de la montrer par TROIS arguments différents :

- Commencer par se ramener au calcul en dimension 1 par le théorème de Fubini.
- Première méthode très simple que l'on peut trouver dans le chapitre 9 de [4] sur la transformée de Fourier. Poser $g(s) = \int e^{isx} e^{-x^2} dx$ et montrer en utilisant le théorème de dérivation sous le signe somme que g est C^1 et déduire de l'expression de g' en intégrant par parties que g est solution de l'équation différentielle $2g'(s) + sg(s) = 0$.
- Une autre méthode utilise la formule de Cauchy appliquée à la fonction holomorphe $g(z) = e^{-\frac{z^2}{2}}$ sur le circuit composé des segments $[(-r, 0), (r, 0)]$, $[(r, 0), (r, t)]$, $[(r, t), (-r, t)]$, $[(-r, t), (-r, 0)]$.
- Une troisième méthode élégante consiste à remarquer que

$$\hat{g}_1(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(x-it)^2}{2}} e^{-\frac{t^2}{2}}.$$

On pose alors

$$f(z) := \int_{\mathbb{R}} e^{-\frac{(x-z)^2}{2}} dx.$$

On démontre très soigneusement en utilisant le théorème de dérivation sous le signe somme que f est holomorphe. Comme f est constante et égale à $\sqrt{2\pi}$ sur \mathbb{R} (pourquoi ??) on conclut.

10.2 Injectivité de la transformée de Fourier

Définition 10.2.1. (Convolution d'une mesure bornée et d'une fonction). Soit μ une mesure bornée sur \mathbb{R}^d et f une fonction borélienne telle que, pour tout x , la fonction $y \rightarrow f(y-x)$ soit μ -intégrable. On pose

$$(f * \mu)(x) = \int_{\mathbb{R}^d} f(x-y) d\mu(y).$$

Exercice 10.2.1. Théorème d'injectivité de la transformée de Fourier sur les mesures bornées

Il s'agit de montrer que *deux mesures bornées sur \mathbb{R}^d qui ont même transformée de Fourier sont égales*. Cette démonstration suit [26], pp. 202-204.

1) Soit μ une mesure bornée sur \mathbb{R}^d . Montrer en appliquant le théorème de Fubini et la relation (d) de l'exercice précédent que

$$(g_\sigma * \mu)(y) = \frac{1}{(\sqrt{2\pi})^d} \int_{\mathbb{R}^d} \hat{\mu}(v) g_1(\sigma v) e^{-iy \cdot v} dv.$$

2) Expliquer pourquoi une mesure bornée sur \mathbb{R}^d est déterminée par la donnée pour tout $f \in C_K(\mathbb{R}^d)$ des intégrales $\int_{\mathbb{R}^d} f d\mu$. *Indication : il suffit de montrer que ces mesures coïncident sur les ouverts car ils forment un π -système des boréliens. On montrera brièvement que toute fonction caractéristique d'ouvert est une limite croissante de fonctions continues à support compact (En effet, soit $g_r(x) = \max(0, 1 - r||x||)$. Quand $r \rightarrow 0$, cette fonction tend en croissant vers 1. Posant $h_r(x) = \inf(1, \frac{1}{r}d(x, \Omega^c))$, puis $k_r(x) = g_r(x)h_r(x)$, on vérifie que cette suite de fonctions convient, quand $r \rightarrow 0$).*

En déduire qu'il suffit de donner une expression de $\int f d\mu$ en fonction de $\hat{\mu}$ pour montrer l'unicité.

3) Montrer en appliquant le théorème de Lebesgue que

$$\int_{\mathbb{R}^d} f d\mu = \lim_{\sigma \rightarrow 0} \int_{\mathbb{R}^d} (f * g_\sigma)(x) d\mu(x).$$

4) En appliquant le théorème de Fubini, en déduire que

$$\int_{\mathbb{R}^d} f d\mu = \lim_{\sigma \rightarrow 0} \int_{\mathbb{R}^d} f(y) (g_\sigma * \mu)(y) dy.$$

5) En utilisant le 1), déduire que

$$\int_{\mathbb{R}^d} f d\mu = \frac{1}{(\sqrt{2\pi})^d} \lim_{\sigma \rightarrow 0} \int f(y) \left(\int \hat{\mu}(v) g_1(\sigma v) e^{-iy \cdot v} dv \right) dy.$$

6) Enoncer le théorème d'injectivité de la transformée de Fourier.

7) Application : soit μ une mesure bornée sur \mathbb{R}^d telle que $\hat{\mu}$ soit Lebesgue-intégrable. Alors μ est absolument continue par rapport à la mesure de Lebesgue et sa densité est donnée par

$$h(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{\mu}(t) e^{-ix \cdot t} dt.$$

Indication : Commencer par montrer en appliquant le théorème de Lebesgue que

$$\int_{\mathbb{R}^d} \hat{\mu}(v) g_1(\sigma v) e^{-iy \cdot v} dv \rightarrow \frac{1}{(\sqrt{2\pi})^d} \int_{\mathbb{R}^d} \hat{\mu} e^{-iy \cdot v} dv$$

Puis, appliquer à nouveau le théorème de Lebesgue dans la relation 5) avec $f \in C_K(\mathbb{R}^d)$.

Exercice 10.2.2. Transformée de Fourier et indépendance ([26], pp. 204-208).

On considère des mesures bornées μ_1 et μ_2 sur \mathbb{R}^{d_1} et \mathbb{R}^{d_2} et la mesure produit $\mu_1 \otimes \mu_2$.

1) Montrer en appliquant le théorème de Fubini (on rappelle que \mathcal{F} désigne la transformation de Fourier) que

$$[\mathcal{F}(\mu_1 \otimes \mu_2)](t_1, t_2) = [\mathcal{F}\mu_1](t_1)[\mathcal{F}\mu_2](t_2).$$

2) Soit $X = (X_1, X_2)$ une v.a. à valeurs dans $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Dédurre du 1) que X_1 et X_2 sont indépendantes si et seulement si

$$\varphi_{(X_1, X_2)}(t_1, t_2) = \varphi_{X_1}(t_1)\varphi_{X_2}(t_2).$$

On utilisera l'injectivité de la transformée de Fourier.

3) Remarquer que $\varphi_{X_1}(t_1) = \varphi_{X_1, X_2}(t_1, 0)$.

4) Si μ_1 et μ_2 sont des mesures bornées sur \mathbb{R}^d , on a $\mathcal{F}(\mu_1 * \mu_2) = \mathcal{F}\mu_1 \mathcal{F}\mu_2$.

Rappel ([26] p. 63) : la convolée de deux mesures μ_1 et μ_2 , $\mu_1 * \mu_2$ est la mesure image de $\mu_1 \otimes \mu_2$ par l'application somme. Cela signifie que pour toute fonction mesurable positive,

$$\int_{\mathbb{R}^d} f d(\mu_1 * \mu_2) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x_1 + x_2) d(\mu_1 \otimes \mu_2)(x_1, x_2).$$

(Appliquer cette formule avec $f(x) = e^{ix \cdot t}$ et utiliser le théorème de Fubini).

Exercice 10.2.3. Un contreexemple ([7], p. 134, exercice 13).

On considère le sous-ensemble de \mathbb{R}^2 ,

$$C = \{(x, y) \in [0, 1]^2, x - \frac{1}{2} \leq y \leq x \text{ ou } y \geq x + \frac{1}{2}\}$$

1) Dessiner C .

2) Soit X et Y deux variables aléatoires de densité de probabilité $f_{X,Y}$ constante sur C , nulle ailleurs. Calculer la valeur de la constante et les densités de X et Y . En déduire que X et Y ne sont pas indépendantes.

3) Calculer les fonctions caractéristiques de X , Y et $X + Y$ pour montrer que

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

Solution

Un calcul élémentaire montre que la constante vaut 2. On désigne par f_X la densité de X et par f_Y celle de Y . On a

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy = \mathbb{1}_{[0,1]}(x)$$

et

$$f_Y(y) = \mathbb{1}_{[0,1]}(y).$$

X et Y ne sont pas indépendantes puisque

$$f_X(x)f_Y(y) \neq f_{X,Y}(x,y).$$

Or on a

$$\varphi_{X+Y}(u) = \left(\frac{e^{iu} - 1}{iu} \right)^2, \quad \varphi_X(u) = \varphi_Y(u) = \left(\frac{e^{iu} - 1}{iu} \right).$$

Exercice 10.2.4. ([30] p. 238) Calculer la fonction caractéristique de la loi uniforme sur $[-1, 1]$. Montrer qu'il n'existe pas de variables aléatoires i.i.d. X et Y telles que $X - Y$ ait cette loi uniforme sur $[-1, 1]$.

10.3 Le Théorème de convergence de Lévy

On note, pour $b > 0$, \mathcal{M}_b l'ensemble des mesures positives telles que $\mu(\mathbb{R}^d) \leq b$.

Proposition 10.3.1. ([26], proposition 14.2) \mathcal{M}_b est métrisable et compact pour la topologie faible (ou vague), c'est-à-dire la topologie $\sigma(C_0(\mathbb{R}^d)', C_0(\mathbb{R}^d))$.

Proposition 10.3.2. ([26], proposition 14.3) Soit μ_n une suite de mesures de \mathcal{M}_b . Si μ_n converge vaguement ou faiblement vers μ , alors $\mu_n \rightarrow \mu$ étroitement si et seulement si $\lim_n \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d)$.

Proposition 10.3.3. ([26], corollaire 14.7) Si $\mu_n \in \mathcal{M}_b$ converge étroitement vers $\mu \in \mathcal{M}_b$, elle est tendue, i.e.

$$\forall \varepsilon > 0, \exists K \text{ compact } \subset \mathbb{R}^d \text{ tel que } \sup_n \mu_n(K^c) \leq \varepsilon.$$

Exercice 10.3.1. Soit $\varepsilon > 0$. Montrer en utilisant le lemme de Fubini qu'il existe une fonction $f_\varepsilon \in C_0(\mathbb{R}^d)$ telle que pour toute mesure bornée $\nu \in \mathbb{R}^d$ on ait

$$\int_{[0, \varepsilon]^d} \hat{\nu}(t) dt = \int_{\mathbb{R}^d} f_\varepsilon d\nu.$$

Solution

On a $f_\varepsilon(x) = \prod_{j=1}^d g_\varepsilon(x_j)$ où $g_\varepsilon(u) = \frac{e^{i\varepsilon u} - 1}{iu}$ si $u \neq 0$, $= \varepsilon$ si $u = 0$ est bien continue.

Exercice 10.3.2. Notre but est de montrer le théorème de Lévy :

Théorème 10.3.1. *Théorème de convergence de Lévy*

(a) Si $\mu_n \in \mathcal{M}_b$ converge étroitement vers μ , alors $\hat{\mu}_n \rightarrow \hat{\mu}$ simplement.

(b) Inversement, si $\hat{\mu}_n \rightarrow \varphi$ simplement et si φ est continue en 0, alors il existe une unique mesure $\mu \in \mathcal{M}_b$ telle que $\varphi = \hat{\mu}$. De plus, $\mu_n \rightarrow \mu$ étroitement.

1) Montrer (a). A partir de la question suivante on s'attache à démontrer (b) et on prend les hypothèses de (b).

2) En utilisant les propositions précédentes, montrer que μ_n admet une sous-suite faiblement convergente, que l'on notera $\mu_{\psi(n)}$. On appellera μ sa limite.

Solution

L'application de la proposition 10.3.1 est immédiate car μ_n est bornée puisque

$$\|\mu_n\|_{C'_0} = \sup_{\|\varphi\| \leq 1} \int \varphi d\mu_n \leq \mu_n(\mathbb{R}^d) = b.$$

3) Montrer en appliquant le théorème de Lebesgue que $\lim_n \int_{[0,\varepsilon]^d} \mu_{\psi(n)}(t) dt = \int_{[0,\varepsilon]^d} \varphi(t) dt$.

Solution

Par la proposition mentionnée, il suffit de montrer $\lim_k \mu_k(\mathbb{R}^d) = \mu(\mathbb{R}^d)$. Or $\lim_k \mu_k(\mathbb{R}^d) = \lim_k \hat{\mu}_k(0) = \varphi(0)$. et $\mu(\mathbb{R}^d) = \hat{\mu}(0)$. Il suffit donc de montrer $\hat{\mu}(0) = \varphi(0)$.

4) Soit f_ε définie dans l'exercice 10.3.1. Montrer successivement en utilisant le 3) que

$$\begin{aligned} \lim_k \int f_\varepsilon d\mu_{\psi(k)} &= \int f_\varepsilon d\mu, \\ \lim_k \int_{[0,\varepsilon]^d} \hat{\mu}_{\psi(k)}(t) dt &= \int_{[0,\varepsilon]^d} \hat{\mu}(t) dt, \\ \frac{1}{\varepsilon^d} \int_{[0,\varepsilon]^d} \hat{\mu}(t) dt &= \frac{1}{\varepsilon^d} \int_{[0,\varepsilon]^d} \varphi(t) dt. \end{aligned}$$

Solution du 4)

La première relation découle de la convergence faible de μ_k et de $f_\varepsilon \in C_0(\mathbb{R}^d)$. La seconde découle du lemme utile car

$$\int_{[0,\varepsilon]^d} \hat{\mu}_k(t) dt = \int f_\varepsilon d\mu_k \rightarrow \int f_\varepsilon d\mu = \int_{[0,\varepsilon]^d} \hat{\mu}(t) dt.$$

La troisième relation vient immédiatement de la précédente et du 3).

5) a) Dédire de la question précédente que $\hat{\mu}(0) = \phi(0)$ et remarquer que $\lim_n \hat{\mu}_{\psi(n)}(0) = \phi(0) = \hat{\mu}(0)$.

b) En utilisant la proposition 10.3.2, montrer que $(\mu_{\psi(n)})_n$ converge étroitement vers μ . En déduire que $\hat{\mu} = \phi$.

c) En déduire que $(\mu_n)_n$ converge faiblement vers μ puis que $(\mu_n)_n$ converge étroitement vers μ .

Solution du 5)

On passe à la limite dans la dernière relation du 4), φ étant continue par hypothèse, et $\hat{\mu}$ comme transformée de Fourier de mesure bornée. Donc $\hat{\mu}(0) = \varphi(0)$. Enfin

$$\lim_n \mu_{\psi(n)}(\mathbb{R}^d) = \lim_n \hat{\mu}_{\psi(n)}(0) = \varphi(0) = \hat{\mu}(0) = \mu(\mathbb{R}^d).$$

Donc $\mu_{\psi(n)}$ converge étroitement vers l'unique μ telle $\hat{\mu} = \varphi$. Mais remarque que si $\mu_{\psi(n)}$ tend étroitement vers μ , alors par le (a) $\hat{\mu}_{\psi(n)}$ tend simplement vers $\hat{\mu}$. Mais comme par hypothèse $\hat{\mu}_{\psi(n)}$

tend simplement vers φ , on déduit que $\hat{\mu} = \varphi$. Ce raisonnement s'appliquant à toute sous-suite extraite, toutes les sous-suites faiblement convergentes de μ_n ont la même limite par le théorème d'injectivité de la Transformée de Fourier. Donc la suite μ_n de l'espace métrique compact $\mathcal{M}(b)$ a une unique limite et donc toute la suite converge.

6) Sous l'hypothèse (a) ou (b), montrer que la suite des fonctions $\hat{\mu}_n$ est équicontinue : on utilisera la proposition 10.3.3 et l'inégalité (que l'on démontrera) $|e^{ix.t} - e^{ix.t'}| \leq |x.(t - t')|$, où $x.t$ désigne le produit scalaire dans \mathbb{R}^d de x et de t . En déduire que la suite $\hat{\mu}_n(t)$ converge uniformément vers $\hat{\mu}$ sur tout compact.

10.4 Fonctions caractéristiques : ce qu'il faut absolument savoir

- 1) La définition d'une fonction caractéristique.
- 2) Retrouver rapidement les propriétés élémentaires sur les fonctions caractéristiques.
- 3) Connaître par coeur la fonction caractéristique de la Gaussienne et connaître une méthode pour la retrouver. Savoir aussi comment calculer les fonctions caractéristiques des lois usuelles.
- 4) Savoir que la fonction caractéristique caractérise complètement la loi d'une v.a. (injectivité de la transformée de Fourier) .
- 5) Etre au courant de la formule d'inversion de Levy. Aller y jeter un oeil, ou même les deux ([30], pp. 175-177).
- 6) Savoir que si X et Y sont indépendantes alors la fonction caractéristique de (X, Y) s'écrit comme le produit direct des fonctions caractéristiques de X et Y et réciproquement.
- 7) Connaître le théorème de Lévy.
- 8) Savoir que si X possède un moment d'ordre n alors la fonction caractéristique de X est n -fois dérivable et le théorème de dérivation sous le signe somme s'applique. Etre au courant aussi que la réciproque est fautive : On peut avoir une fonction caractéristique dérivable en 0 sans que la variable aléatoire X ait un moment d'ordre 1 (Voir [26] ou [16] par exemple).

Chapitre 11

Les gaussiennes

Dans ce chapitre, nous rappelons quelques notions de base sur les lois gaussiennes (Chapitre 13 de [26]). On utilisera les notations suivantes :

La base canonique sur \mathbb{R}^d est notée (e_1, \dots, e_d) et le produit scalaire usuel $\langle \cdot, \cdot \rangle$. La base du dual est notée $(e_i^* = \langle \cdot, e_i \rangle)_{i=1 \dots d}$. Pour un vecteur $x \in \mathbb{R}^d$, x_1, \dots, x_d désigneront les composantes de x dans la base canonique.

Si X est une variable (ou vecteur) aléatoire à valeurs dans \mathbb{R}^d dont les composantes sur la base canonique sont les variables aléatoires réelles X_1, \dots, X_d alors l'espérance de X est le vecteur de \mathbb{R}^d égal à $(\mathbb{E}(X_1), \dots, \mathbb{E}(X_d))$ et la matrice de covariance de X est la matrice symétrique carrée de taille d dont le coefficient (i, j) est donné par la covariance $Cov(X_i, X_j)$ de X_i et X_j . La variance de X est la forme quadratique associée à la matrice de covariance de X .

Si X et Y sont deux vecteurs aléatoires de \mathbb{R}^d alors leur matrice d'intercovariance (ou plus simplement de covariance) est la matrice carrée de taille d dont le coefficient (i, j) est donné par la covariance $Cov(X_i, Y_j)$ de X_i et Y_j . Cette matrice n'est pas obligatoirement symétrique. On dira que X et Y sont non corrélées si et seulement si leur matrice d'intercovariance est nulle.

Une probabilité μ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ (ou plus généralement sur un espace euclidien de dimension finie E muni de sa tribu borélienne) est dite gaussienne si et seulement si la mesure image de μ par toute forme linéaire sur \mathbb{R}^d est une mesure gaussienne sur \mathbb{R} .

Une mesure μ est dite gaussienne sur \mathbb{R} ssi c'est la masse de Dirac en un point ou si c'est une probabilité admettant une densité par rapport à la mesure de Lebesgue de la forme

$$f_{m, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \quad (11.0.1)$$

où $m \in \mathbb{R}$ et $\sigma^2 > 0$. Les masses de Dirac apparaissent comme des cas dégénérés des mesures gaussiennes à densité pour lesquelles le paramètre σ^2 dans la définition précédente est nul.

Une variable aléatoire X est dite gaussienne ssi sa loi est gaussienne.

1) Montrer que si X est gaussienne réelle non dégénérée (i.e. sa loi n'est pas une masse de Dirac) dont la densité de la loi est donnée par (11.0.1), alors X admet une moyenne égale à m et une

variance égale à σ^2 . (Pour le deuxième calcul, intégrer par parties ou utiliser le développement de Taylor en zéro de la fonction caractéristique de la gaussienne).

2) Montrer que la matrice de covariance d'une variable aléatoire est une forme quadratique positive.

$$\sum_{ij} \mathbb{E}(X_i - \bar{X}_i)(X_j - \bar{X}_j)x_i x_j = \mathbb{E}\left(\sum_i x_i (X_i - \bar{X}_i)^2\right) \geq 0.$$

3) a) Montrer que si X est gaussienne alors chacune de ses composantes est une variable aléatoire réelle gaussienne.

3) b) Montrer que si X est gaussienne sur \mathbb{R}^d et si f est une transformation affine de \mathbb{R}^d dans \mathbb{R}^k alors la variable aléatoire $f(X)$ est gaussienne. Pour une transformation linéaire c'est évident. Pour une translation, on remarque que $a.(X + b) = a.X + a.b$ est une gaussienne dont la moyenne est translatée de $a.b$.

4) Montrer que si X est gaussienne alors elle admet des moments de tout ordre (on rappelle que $X \in \mathbb{R}$ admet un moment d'ordre n ssi chacune de ses composantes sur la base canonique est une variable aléatoire admettant un moment d'ordre n ou ssi $\|X\|$ est une variable aléatoire réelle admettant un moment d'ordre n).

5) Gaussiennes et fonctions caractéristiques

On se donne une variable aléatoire X sur \mathbb{R}^d . Le but de la question est de montrer que X est gaussienne ssi sa fonction caractéristique ϕ_X s'écrit

$$\phi_X(t) = \exp \left[i \langle m, t \rangle - \frac{1}{2} \langle Ct, t \rangle \right]$$

où $m \in \mathbb{R}^d$ et C est une matrice symétrique positive.

5) a) Commencer par montrer que dans ce cas la moyenne de X est m et sa matrice de covariance est C .

Solution

On peut procéder par identification en effectuant les développements de Taylor à l'ordre 2 des exponentielles. D'une part si X est une variable aléatoire quelconque ayant ses moments d'ordre 2 et 3 bornés on a

$$\begin{aligned} \phi_X(t) &= \mathbb{E}(e^{it \cdot X}) = \mathbb{E}\left(1 + it \cdot X - \frac{1}{2}(t \cdot X)^2\right) + o(t^2) = \\ &= 1 + it \cdot \mathbb{E}X - \frac{1}{2} \sum_{ij} t_i t_j \mathbb{E}(X_i X_j) + o(t^2) = \\ &= 1 + it \cdot \mathbb{E}X - \frac{1}{2} \sum_{ij} t_i t_j \mathbb{E}(X_i - \bar{X}_i)(X_j - \bar{X}_j) - \frac{1}{2} \left(\sum_i t_i \mathbb{E}X_i\right)^2 + o(t^2). \end{aligned}$$

En développant de même $\phi_X(t)$ à l'ordre 2 pour notre variable particulière :

$$\phi_X(t) = 1 + it \cdot m - \frac{1}{2} \langle Ct, t \rangle - \frac{1}{2} (t \cdot m)^2 + o(t^2).$$

On conclut immédiatement par identification.

5) b) Ensuite, on utilise le théorème d'injectivité de la transformée de Fourier. Pour calculer la fonction caractéristique d'une variable d -dimensionnelle, on pourra montrer et utiliser l'égalité suivante :

$$\phi_X(\alpha t) = \Phi_{\langle X, t \rangle}(\alpha)$$

où $t \in \mathbb{R}^d$ et $\alpha \in \mathbb{R}$, ce qui permet de se ramener au cas unidimensionnel pour le calcul de la fonction caractéristique d'une gaussienne. **La moyenne et la matrice de covariance caractérisent complètement la loi d'une gaussienne : on les appelle les paramètres de la loi gaussienne**

6) Existence de lois gaussiennes

Etant donné un vecteur $m \in \mathbb{R}^d$ et une matrice symétrique positive C , montrer qu'il existe une unique mesure gaussienne de paramètre (m, C) .

Solution

La question de l'unicité a été réglée dans la question précédente. Pour l'existence, on remarquera qu'il suffit de prouver l'existence d'une mesure gaussienne de paramètres $(0, Id_d)$ et de considérer la variable aléatoire $X = m + BX_0$ où X_0 est une variable gaussienne sur \mathbb{R}^d de paramètres $(0, Id_d)$ (gaussienne standard) et B est une matrice symétrique positive dont le carré est C (théorème spectral des opérateurs auto-adjoints). X est alors une gaussienne de paramètres (m, C) (calculer la fonction caractéristique de X). L'existence de X_0 est assurée car on sait exhiber la densité d'une gaussienne standard, à savoir

$$f(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\|x\|^2}{2}\right)$$

On vérifie que la fonction caractéristique associée à cette loi est une gaussienne standard.

7) Pourquoi ne pas avoir démontré le résultat de la question 6) en exhibant comme pour la gaussienne standard une densité pour une gaussienne quelconque ?

Solution

Pour une gaussienne quelconque, l'existence d'une densité par rapport à la mesure de Lebesgue n'est pas forcément assurée (cf. la question suivante pour une formulation précise). Remarquer en effet que même si X_0 est à densité, il n'en est pas forcément de même pour BX_0 (ce n'est en fait le cas que si B est inversible).

8) Soit X une variable gaussienne de paramètre (m, C) .

a) Montrer que si C est définie positive alors X admet une densité par rapport à la mesure de Lebesgue donnée pour tout $x \in \mathbb{R}^d$ par

$$f_{m,C}(x) = \frac{1}{\sqrt{(2\pi)^d \det(C)}} \exp\left(-\frac{1}{2} \langle C^{-1}(x - m), (x - m) \rangle\right)$$

b) Montrer que si C n'est pas définie positive alors X prend presque sûrement ses valeurs dans l'espace affine $m + \text{Im}(C)$ et n'admet en conséquence pas de densité par rapport à la mesure de Lebesgue.

Solution

Pour la question a), on utilisera le fait que l'on connaît la densité d'une loi gaussienne standard et la transformation introduite dans la question 6) puis on effectuera un changement de variables. Pour la question b), en utilisant la décomposition et les notations de la question 6), remarquer que l'image de B est la même que l'image de C et que X prend ses valeurs dans le sous-espace affine $m + \text{Im}(B)$. En fait, on pourrait montrer que X restreint à ce sous-espace admet une densité par rapport à la mesure de Lebesgue sur $m + \text{Im}(B)$.

9) Propriétés d'indépendance des gaussiennes

Soit X une variable aléatoire gaussienne à valeurs dans \mathbb{R}^d et $\sum_{j=1}^n E_j$ une décomposition de \mathbb{R}^d en somme directe. Si $x \in \mathbb{R}^d$, on note x^1, \dots, x^n les éléments respectifs de E_1, \dots, E_n tels que $x = x^1 + \dots + x^n$. Ainsi, la variable aléatoire X s'écrit comme $X^1 + \dots + X^n$ où X^j est une variable aléatoire à valeurs dans E_j .

Alors, les variables aléatoires X^j sont gaussiennes.

De plus, les X^j sont indépendantes si et seulement si elles sont non corrélées deux à deux.

Solution

Pour la première implication, utiliser la question 3b).

On se concentre maintenant sur la deuxième proposition. Pour la condition nécessaire, remarquer que si X^j et X^k sont indépendantes, le coefficient (p, q) de la matrice d'intercovariance de X^j et X^k est donné par $\text{Cov}(e_p^*(X^j), e_q^*(X^k))$ et que $e_p^*(X^j)$ et $e_q^*(X^k)$ sont indépendantes.

Pour la condition suffisante, soient (u^1, \dots, u^n) des éléments de $E_1 \times \dots \times E_n$. Par hypothèse, puisque X est gaussienne, $Y = (\langle X^1, u^1 \rangle, \dots, \langle X^n, u^n \rangle)$ est gaussienne sur \mathbb{R}^n (utiliser la question 4b) avec une fonction affine bien choisie). Remarquer alors que la matrice de covariance de Y est diagonale (utiliser le fait que les X^j sont non corrélées deux à deux). On note $Y = (Y_1, \dots, Y_n)$ la décomposition de Y sur la base canonique de \mathbb{R}^n . Montrer que la fonction caractéristique de Y s'écrit comme produit direct des fonctions caractéristiques des Y_i : $\phi_Y(u) = \prod_{j=1}^n \phi_{Y_j}(u_j)$ et en déduire que les $(Y_j)_{j=1, \dots, n} = (\langle X^j, u^j \rangle)_{j=1, \dots, n}$ sont indépendantes. En remarquant que

$$\phi_{X^j}(u^j) = \Phi_{Y_j}(1)$$

terminer la démonstration.

Remarque 11.0.1. *Attention, dans ce théorème, on suppose que X est gaussienne. Sans cette hypothèse, le théorème est faux. On peut très bien avoir deux variables aléatoires réelles gaussiennes de covariance nulle telles que (X, Y) n'est pas gaussien. C'est cependant vrai si l'on ajoute l'hypothèse X et Y indépendantes. Voici un exemple classique, que l'on vérifiera : On pose $Z = (X, Y)$ avec X gaussienne normale centrée et $Y = U.X$ où U est une Bernouilli centrée de valeurs $(1, -1)$ et paramètre $\frac{1}{2}$. Vérifier que Y est normale centrée. En remarquant que $|Y| = |X|$, montrer que Y et X ne sont pas indépendantes. Montrer finalement qu'elles ont pourtant décorréélées.*

Chapitre 12

Théorème limite central

Exercice 12.0.1. Développement limité de la fonction caractéristique ([26] p. 323)

Soit X une variable aléatoire à valeurs dans \mathbb{R} admettant un moment d'ordre 2. Soit ϕ_X sa fonction caractéristique.

1) Montrer que pour tout réel x

$$\exp(ix) = 1 + ix - x^2 \int_0^1 (1-u) \exp(iux) du$$

2) En déduire

$$|\exp(ix) - (1 + ix - x^2/2)| \leq x^2$$

3) En appliquant la même méthode (Taylor avec reste intégral à l'ordre 3), montrer que

$$|\exp(ix) - (1 + ix - x^2/2)| \leq |x|^3/6$$

4) En déduire

$$|\exp(ix) - (1 + ix - x^2/2)| \leq \inf(|x|^2, |x|^3/6)$$

puis

$$\left| \phi_X(t) - \left(1 + it\mathbb{E}(X) - \frac{t^2}{2}\mathbb{E}(X^2) \right) \right| \leq t^2 \mathbb{E} [\inf(|X|^2, |t||X|^3/6)]$$

Remarquer en particulier le développement limité en $o(t^2)$ ainsi obtenu.

Théorème 12.0.1. ([26] p. 325, grande application du théorème de Lévy)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans \mathbb{R}^d , i.i.d. et admettant un moment d'ordre 2 (c'est-à-dire dont le carré de la norme euclidienne est intégrable). Alors la suite de terme général Y_n défini par

$$Y_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - \mathbb{E}(X_j))$$

converge en loi vers une gaussienne centrée dont la matrice de covariance C est celle des X_i .

1) Montrer que la fonction caractéristique de Y_n est donnée par

$$\phi_{Y_n}(t) = \left[\phi_{\langle X_1 - \mathbb{E}(X_1), t \rangle} \left(\frac{1}{\sqrt{n}} \right) \right]^n.$$

2) En déduire par le lemme précédent le développement asymptotique

$$\phi_{Y_n}(t) = \left[1 - \frac{1}{2n} \mathbb{E}(\langle X_1 - \mathbb{E}(X_1), t \rangle^2) + o\left(\frac{1}{n}\right) \right]^n.$$

3) Montrer que pour tout $z \in \mathbb{C}$,

$$\lim_{n \rightarrow +\infty} \left(1 + \frac{z}{n} + o\left(\frac{1}{n}\right) \right)^n = \exp(z)$$

(On utilisera le log complexe et un développement limité à l'ordre 1).

Solution

On procède comme suit. D'abord on définit le logarithme complexe pour $|z| < 1$ par $\log(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} + \dots$. Pour $z \in]-1, 1[$ cette fonction holomorphe coïncide avec $\log(1+x)$ et on a $e^{\log(1+x)} = 1+x$, $x \in]-1, 1[$. Par le théorème des zéros isolés, on a donc $e^{\log(1+z)} = 1+z$ pour $|z| < 1$. Enfin on a en considérant le développement en série $n \log(1 + \frac{z}{n} + o(\frac{1}{n})) \rightarrow z$. Donc en prenant l'exponentielle

$$e^{n \log(1 + \frac{z}{n} + o(\frac{1}{n}))} = \left(1 + \frac{z}{n} + o\left(\frac{1}{n}\right) \right)^n \rightarrow e^z$$

quand $n \rightarrow \infty$.

4) En déduire

$$\lim_{n \rightarrow +\infty} \phi_{Y_n}(t) = \exp \left[-\frac{1}{2} \langle Ct, t \rangle \right]$$

5) Conclure par le théorème de Lévy.

Exercice 12.0.2. Théorème central limite "local"

Dans cet exercice, nous suivons [18], pp. 195-196.

Théorème 12.0.2. Soient $(X_n)_n$ des variables aléatoires réelles i.i.d., centrées et de variance égale à 1, admettant une fonction densité intégrable g_1 . On note $\phi(t) = \phi_X(t)$ la fonction caractéristique commune des X_n . On suppose que cette fonction est intégrable.

Alors la fonction densité g_n de $U_n = \frac{X_1 + \dots + X_n}{\sqrt{n}}$ existe et satisfait

$$g_n(x) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

quand $n \rightarrow \infty$, et la convergence est uniforme pour $x \in \mathbb{R}$.

On utilisera systématiquement les résultats de l'exercice précédent.

1) Montrer que g_n existe et que $g_n(x) = \sqrt{n}(g_1 * \dots * g_1)(x\sqrt{n})$. Suggestion : Pour bien apprécier la rapidité de la convergence, il est conseillé de simuler sous Matlab cette convolution itérée, avec par exemple $g_1 = \frac{1}{2}\mathbb{1}_{[-1,1]}$, ou n'importe quelle autre fonction positive, paire, et d'intégrale égale à 1. On vérifiera pratiquement qu'une forme gaussienne parfaite apparaît pour g_n au bout de très peu d'itérations (moins d'une dizaine). Cela permet un développement numérique très simple et très convaincant. Montrer que $g_1 = \frac{1}{2}\mathbb{1}_{[-1,1]}$ ne vérifie pas les hypothèses du théorème, mais que $g_1 * g_1$ les vérifie. En généralisant cet exemple, déduire une forme plus générale du théorème 12.0.2. On pourra consulter [16] pour avoir une idée plus précise de la vitesse de convergence.

2) On note $\psi_n(t)$ la fonction caractéristique de U_n . Montrer que $\psi_n(t) = \phi(\frac{t}{\sqrt{n}})^n$ puis que g_n est donnée par $g_n(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \psi_n(t) dt$.

3) Montrer que $|\phi(t)| < 1$ pour $t \neq 0$. *Indications : supposer que cela soit faux pour un certain t , poser $\phi(t) = e^{i\alpha}$ et regarder la partie réelle de l'intégrale de Fourier définissant $\phi(0) - e^{-i\alpha}\phi(t)$.*

4) Montrer qu'il existe δ tel que $|\phi(t)| \leq e^{-\frac{1}{4}t^2}$ pour $|t| \leq \delta$. On utilisera le fait que $1 - t \leq e^{-t}$ pour tout réel t .

5) Montrer que pour tout $a > 0$, $\phi(\frac{t}{\sqrt{n}})^n \rightarrow e^{-\frac{1}{2}t^2}$ uniformément pour $t \in [-a, a]$

6) Montrer que

$$|g_n(x) - \frac{1}{2\pi} e^{-\frac{1}{2}x^2}| \leq \frac{1}{2\pi} \int_{\mathbb{R}} |\phi(\frac{t}{\sqrt{n}})^n - e^{-\frac{1}{2}t^2}| dt = I_n.$$

7) Montrer que

$$\int_{a \leq |t| \leq \delta\sqrt{n}} |\phi(\frac{t}{\sqrt{n}})^n - e^{-\frac{1}{2}t^2}| dt \leq 2 \int_a^\infty 2e^{-\frac{1}{4}t^2} dt \rightarrow 0$$

quand $a \rightarrow \infty$.

8) On traite la partie de l'intégrale I_n pour laquelle $|t| > \delta\sqrt{n}$. On pose $\eta = \sup_{|t| \geq \delta} |\phi(t)| < 1$. Montrer que

$$\int_{|t| \geq \delta\sqrt{n}} |\phi(\frac{t}{\sqrt{n}})^n - e^{-\frac{1}{2}t^2}| dt \leq \eta^{n-1} \int_{\mathbb{R}} |\phi(\frac{t}{\sqrt{n}})| dt + 2 \int_{\delta\sqrt{n}}^\infty e^{-\frac{1}{2}t^2} dt \rightarrow 0$$

quand $n \rightarrow \infty$.

9) Conclure en utilisant les résultats des questions 5) à 8)

10) En reprenant la démonstration pas à pas et en signalant les (petites) différences dans les formules, montrer que le théorème central limite local est encore vrai pour des variables X_n à valeurs dans \mathbb{R}^d .

Exercice 12.0.3. L'utilisation du théorème de Lévy est la clef de voute de la preuve du théorème limite central. Dans le même genre d'idée, on peut prouver le théorème dit des événements rares ([26] pp. 321-323). La preuve que l'on trouve dans [26], sans être totalement fausse, n'est pas agréativement acceptable.

Théorème 12.0.3. Théorème des événements rares Soit pour tout $n \in \mathbb{N}^*$ une famille finie $\{A_{n,j} | 1 \leq j \leq M_n\}$ d'événements indépendants définis sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On pose $p_{n,j} = \mathbb{P}(A_{n,j})$ et on note

$$S_n = \sum_{j=1}^{M_n} \mathbb{1}_{A_{n,j}}$$

On suppose que la suite de terme général M_n tend en croissant vers $+\infty$ et que

$$\sup_{1 \leq j \leq M_n} p_{n,j} \xrightarrow{n \rightarrow +\infty} 0, \quad \sum_{j=1}^{M_n} p_{n,j} \xrightarrow{n \rightarrow +\infty} \lambda$$

où $\lambda > 0$. Alors la suite $(S_n)_{n \in \mathbb{N}^*}$ converge en loi vers la loi de Poisson $\mathcal{P}(\lambda)$ de paramètre λ .

Avant de montrer ce théorème, remarquer qu'il généralise la convergence de la loi binomiale de paramètres (n, p) vers la loi de Poisson de paramètre λ quand $np \rightarrow \lambda$. Ensuite, l'interprétation est la même, mais elle est plus souple et plus adaptée à la modélisation économique. Par exemple, $p_{n,j}$ représente la probabilité que le client potentiel n se présente à une queue de supermarché. Les paramètres n et λ sont observables car ils peuvent être estimés comme le nombre de clients ayant passé à la caisse à un certain moment de la journée, ce nombre étant calculé sur une période très longue. Et λ représente le nombre moyen de clients observés à la caisse à un moment donné de la journée. Bien sûr $p_{n,j}$ n'est pas observable, mais le but est de justifier la validité de la loi de Poisson du nombre de clients passant à la caisse.

1) Montrer que la fonction caractéristique de S_n est

$$\phi_{S_n}(t) = \prod_{j=1}^{M_n} [1 + p_{n,j}(\exp(it) - 1)]$$

2) On pose $z = \exp(it) - 1$. Montrer qu'il existe $N \geq 1$ tel que pour $n \geq N$, et $j \in \{1, \dots, M_n\}$,

$$1 + p_{n,j}z = \exp\{\log[1 + p_{n,j}z]\}$$

où \log désigne une détermination principale du logarithme complexe sur le plan fendu de Cauchy.

3) Montrer que

$$\log(1 + z) = z - z^2 \int_0^1 (1 - u) \frac{1}{(1 + uz)^2} du, \quad |z| < 1$$

4) Montrer que pour tout $n \geq N$,

$$\left| \sum_{j=1}^{M_n} p_{n,j}^2 \int_0^1 \frac{1 - u}{(1 + up_{n,j}z)^2} du \right| \leq 2 \left[\sup_{1 \leq j \leq M_n} p_{n,j} \right] \left[\sum_{j=1}^{M_n} p_{n,j} \right]$$

5) En déduire que $\sum_{j=1}^{M_n} \log(1 + p_{n,j}z) \xrightarrow{n \rightarrow +\infty} \lambda z$.

6) Conclure par le théorème de Lévy. Regarder attentivement la démonstration de [26] et repérer le passage qui n'est pas très clair.

Chapitre 13

Grandes déviations

Source : [18]

Exercice 13.0.4. Fonction génératrice On appelle transformée de Laplace d'une variable aléatoire X de moyenne μ la fonction $M : \mathbb{R} \rightarrow \mathbb{R}^+$ définie par $M(t) = \mathbb{E}(e^{tX})$. Dans toute la suite, on supposera que pour les variables aléatoires considérées, $M(t)$ est définie et bornée dans un intervalle ouvert contenant l'origine.

1) En utilisant le développement en série entière de l'exponentielle, vérifier que

$$M(t) = \sum_{k=0}^{\infty} \frac{\mathbb{E}(X^k)}{k!} t^k$$

sur l'intervalle ouvert où elle est bornée et que

$$\mathbb{E}X = M'(0), \quad \mathbb{E}(X^k) = M^{(k)}(0), \quad M'(t) = \mathbb{E}(Xe^{tX}), \quad M''(t) = \mathbb{E}(X^2e^{tX}).$$

Indication. Soit μ_X la loi de X . Écrire

$$\mathbb{E}(e^{tX}) = \int_{\mathbb{R}} \sum_k \frac{t^k x^k}{k!} d\mu_X = \int_{\mathbb{R}^+} \sum_k \frac{t^k x^k}{k!} d\mu_X + \int_{\mathbb{R}^-} \sum_k \frac{t^k x^k}{k!} d\mu_X.$$

On peut intervertir intégration et sommation dans le premier terme de droite grâce au théorème de la convergence monotone (Beppo-Levi) et on peut intervertir dans le second terme parce que la série est alternée et donc uniformément convergente.

2) On pose $\Lambda(t) = \log M(t)$ la Log-Laplace de X . Montrer que

$$\Lambda(0) = \log M(0) = 0, \quad \Lambda'(0) = \frac{M'(0)}{M(0)} = \mu.$$

3) Sur son intervalle de définition, $\Lambda(t)$ est convexe. Plus précisément, montrer en utilisant l'inégalité de Cauchy-Schwartz que

$$\Lambda''(t) = \frac{M(t)M''(t) - M'(t)^2}{M(t)^2} \geq 0.$$

4) On rappelle que l'inégalité de Cauchy-Schwartz $(E(XY))^2 \leq E(X^2)E(Y^2)$ ne devient une égalité que s'il existe λ tel que $X = \lambda Y$. En déduire que si X est de variance non nulle, alors $\Lambda(t)$ est strictement convexe sur son intervalle de définition.

5) On définit la transformée de Fenchel-Legendre de Λ par

$$\Lambda^*(a) = \sup_{t \in \mathbb{R}} (at - \Lambda(t)), \quad a \in \mathbb{R}.$$

Montrer que Λ^* est convexe.

Exercice 13.0.5. Notre but est de montrer l'estimation de grande déviation suivante :

Théorème 13.0.4. Soient X_1, X_2, \dots des variables i.i.d. de moyenne μ et telles que $M(t) = \mathbb{E}(e^{tX})$ soit fini dans un intervalle ouvert contenant 0. On pose $S_n = (X_1 + \dots + X_n)$. Soit $a > \mu$ tel que $\mathbb{P}(X > a) > 0$. Alors $\Lambda^*(a) > 0$ et

$$-\frac{1}{n} \log \mathbb{P}(S_n > an) \rightarrow \Lambda^*(a) \quad \text{quand } n \rightarrow \infty.$$

Faisons le point : la loi des grands nombres nous dit que $\frac{1}{n}S_n \rightarrow \mu$ presque sûrement. Le théorème central limite nous dit que les déviations typiques de $S_n - n\mu$ sont de l'ordre de $c\sqrt{n}$. Donc, les probabilités $\mathbb{P}(S_n - \mu) \geq n^\alpha$ avec $\alpha > \frac{1}{2}$ sont petites et le théorème nous donne une estimation asymptotique précise de cette petitesse pour $\alpha = 1$.

PREMIÈRE PARTIE

1) Montrer que l'on peut supposer sans perte de généralité que $\mu = 0$: on supposera le théorème démontré pour $\mu = 0$ et si ce n'est pas le cas, on remplacera X_i par $X_i - \mu$.

2) En changeant X_i en $-X_i$ et en appliquant le théorème, donner aussi la limite de $\frac{1}{n} \log \mathbb{P}(S_n < na)$, pour $a < \mu$.

3) On va montrer que $\Lambda^*(a) > 0$. Soit $\sigma^2 = \text{var}(X)$. Montrer que

$$at - \Lambda(t) = \log\left(\frac{e^{at}}{M(t)}\right) = \log \frac{1 + at + o(t)}{1 + \frac{1}{2}\sigma^2 t^2 + o(t^2)}$$

pour t petit et en déduire que $\Lambda^*(a) > 0$.

4) Déduire de $\Lambda'(0) = E(X) = 0$ que

$$\Lambda^*(a) = \sup_{t > 0} (at - \Lambda(t)), \quad a > 0.$$

5) Dans cette question et les suivantes, on procède comme pour l'inégalité de Hoeffding. Remarquer que $e^{tS_n} > e^{nat} \mathbb{1}_{S_n > na}$ et en déduire que

$$\mathbb{P}(S_n > na) \leq e^{-n(at - \Lambda(t))}.$$

6) En déduire que

$$\frac{1}{n} \log \mathbb{P}(S_n > na) \leq -\sup_{t > 0} (at - \Lambda(t)) = -\Lambda^*(a).$$

Cette majoration donne la "moitié" du théorème; il va falloir maintenant minorer.

DEUXIÈME PARTIE Dans cette partie, on va faire une hypothèse sur Λ^* qui est vérifiée en pratique.

Définition 13.0.1. On dira qu'on est dans le cas "régulier" en a si le maximum définissant $\Lambda^*(a)$ est atteint dans l'intérieur de l'intervalle de définition de $M(t)$. On notera $\tau = \tau(a)$ cette valeur de t .

Soit $T := \sup\{t, M(t) < \infty\}$.

1) Vérifier que

$$\Lambda^*(a) = a\tau - \Lambda(\tau), \quad \Lambda'(\tau) = a.$$

2) On note $F(u) = \mathbb{P}(X \leq u)$ la fonction distribution de X . On va lui associer une fonction auxiliaire par un "changement exponentiel de distribution" défini comme suit. On pose $d\tilde{F}(u) = \frac{e^{\tau u}}{M(\tau)} dF(u)$, ou en d'autres termes,

$$\tilde{F}(y) = \frac{1}{M(\tau)} \int_{-\infty}^y e^{\tau u} dF(u).$$

On considère $\tilde{X}_1, \tilde{X}_2, \dots$ des variables aléatoires indépendantes ayant toutes \tilde{F} comme distribution et on pose $\tilde{S}_n = \tilde{X}_1 + \dots + \tilde{X}_n$.

3) En utilisant la formule générale $M(t) = \mathbb{E}(e^{tX}) = \int_{\mathbb{R}} e^{tu} dF(u)$, montrer que

$$\tilde{M}(t) = \frac{M(t + \tau)}{M(\tau)}.$$

4) En déduire que

$$\begin{aligned} \mathbb{E}(\tilde{X}_i) &= \tilde{M}'(0) = a, \\ \text{var}(\tilde{X}_i) &= \mathbb{E}(\tilde{X}_i^2) - \mathbb{E}(\tilde{X}_i)^2 = \tilde{M}''(0) - \tilde{M}'(0)^2 = \Lambda''(\tau). \end{aligned}$$

5) Montrer que la transformée de Laplace de \tilde{S}_n est

$$\left(\frac{M(t + \tau)}{M(\tau)} \right)^n = \frac{1}{M(\tau)^n} \int_{\mathbb{R}} e^{(t+\tau)u} dF_n(u),$$

où F_n est la fonction de distribution de S_n . En déduire que la fonction de distribution \tilde{F}_n de \tilde{S}_n vérifie

$$d\tilde{F}_n(u) = \frac{e^{\tau u}}{M(\tau)^n} dF_n(u).$$

6) Soit $b > a$. Déduire de la question précédente que

$$\mathbb{P}(S_n > na) \geq e^{-n(\tau b - \Lambda(\tau))} \mathbb{P}(na < \tilde{S}_n < nb).$$

7) En utilisant le fait que la moyenne des \tilde{X}_i est a et le théorème central limite et la loi des grands nombres, montrer que $\mathbb{P}(na < \tilde{S}_n < nb) \rightarrow \frac{1}{2}$. Déduire de ce fait et de la question précédente que

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n > na) \geq -(\tau b - \Lambda(\tau)) \rightarrow -\Lambda^*(a) \quad \text{quand } b \rightarrow a.$$

TROISIÈME PARTIE

1) On se place sous les mêmes hypothèses que dans la deuxième partie, sauf que l'on considère maintenant le cas non régulier. Pour se ramener au cas régulier, on fixe $c > a$ et on pose $X^c = \inf(X, c)$. On note $M^c(t)$ la transformée de Laplace associée et on pose $\Lambda^c(t) = \log M^c(t)$.

2) Montrer que $M^c(t) < \infty$ pour tout $t > 0$, que $\mathbb{E}(X^c) \rightarrow 0$ quand $c \rightarrow \infty$ et que $\mathbb{E}(X^c) \leq 0$.

3) Montrer qu'il existe $b \in]a, c[$ tel que $\mathbb{P}(X > b) > 0$ et en déduire que

$$at - \Lambda^c(t) \leq at - \log(e^{tb}\mathbb{P}(X > b)) \rightarrow -\infty \text{ quand } t \rightarrow \infty.$$

4) En conclure que la suite X_i^c est dans le "cas régulier" de la partie précédente c'est à dire que le sup de $at - \Lambda^c(t)$ pour $t > 0$ est atteint en une valeur $\tau^c \in]0, \infty[$, et finalement que

$$\frac{1}{n} \log \mathbb{P}\left(\sum_{i=1}^n X_i^c > na\right) \rightarrow -(\Lambda^c)^*(a) \text{ quand } t \rightarrow \infty.$$

(On pose $\Lambda^{c*} := \sup_{t>0}(at - \Lambda^c(t)) = a\tau^c - \Lambda^c(\tau^c)$).

5) Montrer que $\Lambda^{c*}(a) \downarrow \Lambda^{\infty*}$ quand $c \rightarrow +\infty$. et que $0 \leq \Lambda^{\infty*} < \infty$.

6) Montrer que

$$\frac{1}{n} \log \mathbb{P}(S_n > na) \geq \frac{1}{n} \log \mathbb{P}\left(\sum_{i=1}^n X_i^c > a\right).$$

7) Expliquer pourquoi il suffit, pour conclure le théorème dans le cas irrégulier, de montrer que $\Lambda^{\infty*} \leq \Lambda^*(a)$.

8) Pour montrer cette dernière relation, procéder comme suit : Montrer que l'ensemble $I_c = \{t \geq 0, at - \Lambda^c(t) \geq \Lambda^{\infty*}\}$ est non vide. Montrer que I_c est un intervalle fermé compact. Montrer que les intervalles I_c sont emboîtés quand c croît et en déduire qu'il existe $\zeta \in \cap_{c>a} I_c$. Montrer finalement que $\Lambda^c(\zeta) \rightarrow \Lambda(\zeta)$ quand $c \rightarrow \infty$ et que $a\zeta - \Lambda(\zeta) = \lim_{c \rightarrow \infty}(a\zeta - \Lambda^c(\zeta)) \geq \Lambda^{\infty*}$. Conclure.

Chapitre 14

Espérance conditionnelle

Dans ce chapitre, $(\Omega, \mathcal{F}, \mathbb{P})$ désignera un espace probabilisé et \mathcal{G} sera une sous-tribu de \mathcal{F} . Nous donnerons pour simplifier la notion d'espérance conditionnelle en dimension 1 (variables aléatoires réelles). Bien entendu, tout ceci se généralise sans peine au cas multidimensionnel (vecteurs aléatoires).

14.1 Introduction : variables aléatoires discrètes

On se donne deux variables aléatoires discrètes X et Z à valeurs dans \mathbb{Z} à support fini. L'idée intuitive de probabilité conditionnelle nous amène à prendre pour valeur de la "probabilité de $\{X = i\}$ sachant $\{Z = j\}$ " la quantité suivante

$$\mathbb{P}(X = i | Z = j) = \frac{\mathbb{P}(X = i, Z = j)}{\mathbb{P}(Z = j)}$$

pour j appartenant au support de Z . Si j n'appartient pas au support de Z , cette définition n'a pas de sens. Par suite, l'espérance de X sachant $\{Z = j\}$ sera définie comme

$$\mathbb{E}(X | Z = j) = \sum_{i \in \mathbb{Z}} i \mathbb{P}(X = i | Z = j)$$

Définissons maintenant l'espérance conditionnelle de X sachant Z comme la variable aléatoire $Y(\omega) = \mathbb{E}(X | Z)(\omega)$ donnée par

$$Y(\omega) = \sum_j \mathbb{1}_{Z^{-1}(\{j\})}(\omega) \mathbb{E}(X | Z = j)$$

En particulier, si ω n'appartient pas au support de Z , $Y(\omega) = 0$.

1) Remarquons que pour tout borélien $B \in \sigma(Z)$, c'est-à-dire $B \subset \mathcal{P}(Z^{-1}(\{z\}); z \in \text{Support}(Z))$, on a

$$\mathbb{E}(Y \mathbb{1}_B) = \mathbb{E}(X \mathbb{1}_B)$$

2) Y est $\sigma(Z)$ mesurable.

La question qui se pose est de savoir comment généraliser cette notion d'espérance conditionnelle lorsque Z n'est pas à valeurs discrètes. En particulier, on remarquera que si Z est à densité,

l'événement $\{Z = z\}$ est pour tout réel z de probabilité nulle si bien que la définition même de probabilité conditionnelle devient délicate. La première idée consisterait à approximer l'événement $\{Z = z\}$ par $\lim_{\varepsilon \rightarrow 0} \{Z \in (z - \varepsilon; z + \varepsilon)\}$ et l'on peut effectivement procéder ainsi mais cela devient assez vite compliqué. On va procéder tout autrement.

14.2 Théorème d'existence et d'unicité

Théorème 14.2.1. *Soit X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ intégrable. Il existe une variable aléatoire Y telle que :*

- 1) Y est \mathcal{G} -mesurable.
- 2) $\mathbb{E}(|Y|) < +\infty$.
- 3) $\forall G \in \mathcal{G}, \int_G Y d\mathbb{P} = \int_G X d\mathbb{P}$.

De plus, si \tilde{Y} est une autre variable aléatoire vérifiant ces trois propriétés alors $Y = \tilde{Y}$ \mathbb{P} p.s. On parlera alors de Y comme d'une version de l'espérance conditionnelle $\mathbb{E}(X|\mathcal{G})$ qui est un élément de $\mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$. Par abus de langage, on parlera de Y (ou toute autre version) comme de l'espérance conditionnelle de X sachant \mathcal{G} .

Avant d'entamer la preuve de l'existence de l'espérance conditionnelle, faisons quelques remarques.

Remarque 14.2.1. *Si l'on veut prouver 3), on peut se restreindre à montrer l'égalité pour des G qui appartiennent à un π -système contenant Ω et générant \mathcal{G} (Théorème de la classe monotone). D'autre part, si on a 3), par le théorème de la classe monotone, on a aussi que*

$$\mathbb{E}(YU) = \mathbb{E}(XU)$$

pour toute variable aléatoire U qui est \mathcal{G} -mesurable bornée.

Dans la démonstration, nous aurons besoin de considérer les espaces quotients $\mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ et $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. On a bien sûr l'inclusion triviale $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P}) \subset \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Cette inclusion passe au quotient puisque si $X, Y \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ définissent le même élément dans $\mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P})$, c'est que $\{Y \neq X\} \in \mathcal{G}$ est de \mathbb{P} probabilité nulle donc que X et Y définissent aussi le même élément dans $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. D'autre part, $\mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ est fermé dans $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.

Démonstration. (du théorème (14.2.1)).

Existence : On commence par traiter le cas où X admet un moment d'ordre 2. $\mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ est un sev complet de $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ pour la norme $\|\cdot\|_2$ dérivant du produit-scalaire noté $\langle \cdot, \cdot \rangle$. On désigne alors par Y la projection orthogonale de X sur $\mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P})$. Par définition de la projection orthogonale,

$$\forall Z \in \mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P}), \langle X - Y, Z \rangle = 0$$

donc $\mathbb{E}(X\mathbb{1}_G) = \mathbb{E}(Y\mathbb{1}_G)$ pour tout $G \in \mathcal{G}$. Ceci prouve l'existence de l'espérance conditionnelle pour les variables aléatoires ayant des moments d'ordre 2. Passons maintenant au cas général.

Soit $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. En l'écrivant comme différence de sa partie positive et de sa partie négative : $X = X^+ - X^-$, il est clair que l'on peut se ramener au cas où X est positive. On suppose donc maintenant que X est positive. Soit pour tout entier n , $X_n = X \wedge n \in \mathcal{L}^2$. On peut donc par ce qui précède prendre son espérance conditionnelle par rapport à \mathcal{G} , $Y_n = \mathbb{E}(X_n|\mathcal{G})$. D'autre part, X_n tend simplement en croissant vers X et on va voir que de même, $(Y_n)_n$ tend en croissant vers une

certaine limite Y .

Lemme 14.2.1. $0 \leq Y_n \leq Y_{n+1}$, \mathbb{P} p.s.

Démonstration. On prouve que si $U \geq 0$ admet un moment d'ordre 2 alors $W = \mathbb{E}(U|\mathcal{G}) \geq 0$. Si $\mathbb{P}(W < 0) > 0$ alors il existe n tel que $\mathbb{P}(W < -1/n) > 0$ et $\{W < -1/n\} \in \mathcal{G}$ donc

$$0 \leq \mathbb{E}(U \mathbb{1}_{W < -1/n}) = \mathbb{E}(W \mathbb{1}_{W < -1/n}) \leq -1/n$$

ce qui est exclu.

On en déduit immédiatement le lemme (par linéarité de l'espérance conditionnelle sur $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$).

□

On pose $Y = \limsup Y_n$ qui est \mathcal{G} -mesurable. Pour tout $G \in \mathcal{G}$,

$$\begin{aligned} \mathbb{E}(Y \mathbb{1}_G) &= \lim_{n \rightarrow \infty} \mathbb{E}(Y_n \mathbb{1}_G) \text{ par convergence monotone} \\ &= \lim_{n \rightarrow \infty} \mathbb{E}(X_n \mathbb{1}_G) \\ &= \mathbb{E}(X \mathbb{1}_G) \text{ par convergence monotone} \end{aligned}$$

En particulier, avec $G = \Omega$, $\mathbb{E}(Y) = \mathbb{E}(X) < +\infty$.

Unicité : Soit Y et \tilde{Y} deux variables aléatoires vérifiant 1), 2) et 3). On veut montrer que $Y = \tilde{Y}$, \mathbb{P} presque sûrement. On a donc :

$$\forall G \in \mathcal{G}, \mathbb{E}(\mathbb{1}_G(Y - \tilde{Y})) = 0$$

Si $\mathbb{P}(Y = \tilde{Y}) < 1$ alors quitte à permuter Y et \tilde{Y} , on a

$$\mathbb{P}(Y > \tilde{Y}) > 0$$

Or $\lim_{n \rightarrow \infty} \mathbb{P}(Y - \tilde{Y} > 1/n) = \mathbb{P}(Y > \tilde{Y})$ et il existe donc n tel que $\mathbb{P}(Y - \tilde{Y} > 1/n) > 0$. Mais $G_n = \{Y - \tilde{Y} > 1/n\} \in \mathcal{G}$ par 1) donc

$$\mathbb{E}(\mathbb{1}_{G_n}(Y - \tilde{Y})) = 0 \Rightarrow \frac{1}{n} \mathbb{P}(Y - \tilde{Y} > 1/n) \leq 0 \quad \text{: exclu}$$

Ceci conclut la preuve de l'unicité et du théorème. □

Remarque 14.2.2. 1) Quand on considère au lieu de \mathcal{G} la tribu engendrée par une variable aléatoire Z (resp. une famille de variables aléatoires $(Z_i)_{i \in I}$), il est d'usage de noter l'espérance conditionnelle de X sachant $\sigma(Z)$ par $\mathbb{E}(X|Z)$ (resp. $\mathbb{E}(X|Z_i, i \in I)$).

2) On remarquera que $\mathbb{E}(X|Z)$ est $\sigma(Z)$ -mesurable donc elle s'écrit sous la forme $f(Z)$ où f est une certaine fonction borélienne.

14.3 Variables à densité

Soit X et Z deux variables aléatoires réelles (pour simplifier les écritures) telles qu'elles aient une densité jointe $f_{X,Z}(x, z)$. On rappelle que X (resp. Z) a alors une densité notée $f_Z(z) = \int f_{X,Z}(x, z) dx$ (resp. $f_X(x) = \int f_{X,Z}(x, z) dz$). On pose alors

$$f_{X|Z}(x|z) = \mathbb{1}_{f_Z(z) \neq 0} \frac{f_{X,Z}(x, z)}{f_Z(z)}$$

Soit h une fonction borélienne (bornée pour simplifier) ; quelle est l'espérance conditionnelle de $h(X)$ sachant Z ?

Si g est une fonction borélienne bornée, par le théorème de Fubini, on a la suite d'égalités suivantes :

$$\begin{aligned}\mathbb{E}(h(X)g(Z)) &= \int h(x)g(z)f_{X,Z}(x,z)dx dz \\ &= \int_{\mathbb{R}^2} g(z)f_Z(z)f_{X|Z}(x,z)h(x)dx dz \\ &= \int_{\mathbb{R}} g(z)f_Z(z) \left(\int_{\mathbb{R}} f_{X|Z}(x,z)h(x)dx \right) \\ &= \mathbb{E}(g(Z)\theta(Z))\end{aligned}$$

où $\theta(z) = \int_{\mathbb{R}} f_{X|Z}(x,z)h(x)dx$. On a donc prouvé que

$$\mathbb{E}(h(X)|Z)(\omega) = \theta(Z(\omega)) = \int_{\mathbb{R}} f_{X|Z}(x, Z(\omega))h(x)dx$$

14.4 Variables discrètes

En se reportant à la section 1, on vérifiera sans peine que si X et Z sont deux variables aléatoires discrètes à valeurs dans \mathbb{Z} , on a :

$$\mathbb{E}(X|Z) = \sum_j j \left(\sum_i \mathbb{1}_{Z=i} \mathbb{P}(X = j|Z = i) \right)$$

Le terme

$$\sum_j j \mathbb{1}_{\{Z=i\}} \mathbb{P}(X = j|Z = i)$$

est souvent noté

$$\mathbb{E}(X|Z = i)$$

14.5 Noyaux et lois conditionnelles

Ce paragraphe est tiré de [26] et donné sans la démonstration des propriétés. On s'y reportera pour de plus amples détails (mais ce qui est ici doit être su). La notion de loi conditionnelle est plus profonde que celle d'espérance conditionnelle qui n'est d'ailleurs que "l'espérance par rapport à la loi conditionnelle".

(E, \mathcal{E}) et (F, \mathcal{F}) sont deux espaces probabilisés (qui seront ici en fait des $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$).

Définition 14.5.1. $\nu : E \times \mathcal{F} \rightarrow [0, 1]$ est un noyau (ou probabilité) de transition si il satisfait :

1. $\forall x \in E, \nu(x, \cdot)$ est une probabilité sur (F, \mathcal{F}) .
2. $\forall B \in \mathcal{F}, \nu(\cdot, B)$ est \mathcal{E} -mesurable.

Définition 14.5.2. Si $X \in \mathbb{R}^d$ et $Y \in \mathbb{R}^k$ sont deux variables aléatoires, on appelle loi conditionnelle de Y sachant X un noyau ν sur $(\mathbb{R}^d \times \mathcal{B}(\mathbb{R}^k))$ tel que :

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \cdot \nu$$

c'est-à-dire

$$\mathbb{P}(X \in A; Y \in B) = \int_{x \in A} \nu(x, B) d\mathbb{P}_X(x)$$

On notera, **et ce n'est qu'une notation**, un tel noyau ν de la manière suivante :

$$\nu(x, B) = \mathbb{P}(Y \in B | X = x) \text{ ou } \nu(x, B) = \mathbb{P}_Y(B | X = x)$$

Remarquer par exemple que si X est à densité, $\{X = x\}$ est un ensemble de probabilité nulle. Il est donc hors de question de la définir par :

$$\mathbb{P}(Y \in B | X = x) = \frac{\mathbb{P}(Y \in B; X = x)}{\mathbb{P}(X = x)}$$

Cependant cette formule est vraie dans le cas où $\mathbb{P}(X = x) \neq 0$.

Il reste à savoir quand on a l'existence de lois conditionnelles. La réponse à cette question est apportée par le théorème non trivial suivant.

Théorème 14.5.1. : théorème de Jirina

Si X et Y sont deux variables aléatoires à valeurs dans des $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ alors il existe toujours une loi conditionnelle de Y sachant X .

Le théorème suivant est une généralisation du théorème de Fubini dans le cadre conditionnel.

Théorème 14.5.2. *Soit (X, Y) une variable aléatoire à valeurs dans un espace probablisable quelconque $(E \times F, \mathcal{E} \otimes \mathcal{F})$ telle qu'existe une loi conditionnelle $\mathbb{P}(Y \in \cdot | X = \cdot)$ de Y sachant X . Soit f une application mesurable de l'espace probablisable $(E \times F, \mathcal{E} \otimes \mathcal{F})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

a) Si f est positive, l'application $x \rightarrow \int_F f(x, y) d\mathbb{P}_Y(dy | X = x)$ est \mathcal{E} -mesurable et on a :

$$\int_{E \times F} f(x, y) d\mathbb{P}_{(X, Y)}(x, y) = \int_E \left[\int_F f(x, y) d\mathbb{P}_Y(dy | X = x) \right] d\mathbb{P}_X(x)$$

b) Si f est de signe quelconque et $\mathbb{P}_{(X, Y)}$ -intégrable, pour \mathbb{P}_X presque tout x , l'application partielle $f(x, \cdot)$ est $\mathbb{P}_Y(\cdot | X = x)$ -intégrable, et l'application définie \mathbb{P}_X -presque sûrement par $x \rightarrow \int_F f(x, y) d\mathbb{P}_Y(dy | X = x)$ est \mathbb{P}_X -intégrable et l'égalité précédente est encore vraie.

Il est alors aisé d'établir le corrolaire suivant.

Proposition 14.5.1. *(Propriété de transfert conditionnel)*

Si $X \in \mathbb{R}^p$ et $Y \in \mathbb{R}^k$ sont deux variables aléatoires et $f : \mathbb{R}^{p+k} \rightarrow \mathbb{R}^q$ borélienne, alors

$$\mathbb{P}(f(X, Y) \in \cdot | X = x) = \mathbb{P}(f(x, Y) \in \cdot | X = x)$$

En particulier, si X et Y sont indépendantes, on a

$$\mathbb{P}(f(X, Y) \in \cdot | X = x) = \mathbb{P}(f(x, Y) \in \cdot)$$

Terminons par une définition avant de faire le lien avec l'espérance conditionnelle.

Définition 14.5.3. Soit Y une variable aléatoire et (X_1, \dots, X_n) une suite finie de variables aléatoires. On dira que relativement à $Y = y$ les variables aléatoires (X_1, \dots, X_n) ont pour loi μ_y (probabilité sur \mathbb{R}^n dépendant de y) ssi il existe une loi conditionnelle de (X_1, \dots, X_n) relativement à Y telle que \mathbb{P}_Y p.s.

$$\mathbb{P}_{(X_1, \dots, X_n)}(\cdot | Y = y) = \mu_y(\cdot)$$

En particulier, on dira que (X_1, \dots, X_n) sont indépendantes conditionnellement à $Y = y$ ssi la loi conditionnelle de (X_1, \dots, X_n) relativement à Y est une loi produit.

L'espérance conditionnelle de la variable aléatoire X relativement à une autre variable aléatoire Y est

$$\mathbb{E}(X|Y)(\omega) = \int_{\mathbb{R}} x d\mathbb{P}_X(dx | Y = Y(\omega))$$

Il suffit en effet de remarquer que la fonction définie par le membre de droite vérifie les propriétés caractéristiques de l'espérance conditionnelle. L'espérance conditionnelle apparaît donc comme "l'espérance relativement à la loi conditionnelle".

14.6 TD Probabilités : Espérance conditionnelle

Exercice 14.6.1. Les propriétés fondamentales : les lire ([30] p. 88 et la dernière page)

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. On suppose que X , \mathcal{F} -mesurable, vérifie $\mathbb{E}|X| < \infty$. \mathcal{G} et \mathcal{H} sont des sous-tribus de \mathcal{F} . On rappelle qu'il existe une unique v.a. \mathcal{G} -mesurable et sommable, notée $\mathbb{E}(X|\mathcal{G})$ telle que

$$\forall G \in \mathcal{G}, \int_G \mathbb{E}(X|\mathcal{G}) d\mathbb{P} = \int_G X d\mathbb{P}. \quad (14.6.1)$$

(L'unicité est vraie modulo un ensemble de mesure nulle). De manière plus condensée :

$$\forall G \in \mathcal{G}, \mathbb{E}(\mathbb{E}(X|\mathcal{G}); G) = E(X; G),$$

où $E(X; G) = \int_G X d\mathbb{P}$.

- (a) On pose $Y = \mathbb{E}(X|\mathcal{G})$, alors $\mathbb{E}(Y) = \mathbb{E}(X)$.
- (b) Si X est \mathcal{G} -mesurable, alors $\mathbb{E}(X|\mathcal{G}) = X$ p.p.
- (c) **Linéarité** $\mathbb{E}(aX_1 + a_2X_2|\mathcal{G}) = a_1\mathbb{E}(X_1|\mathcal{G}) + a_2\mathbb{E}(X_2|\mathcal{G})$, p.p.
- (d) **Positivité** Si $X \geq 0$, alors $\mathbb{E}(X|\mathcal{G}) \geq 0$, p.p.
- (e) **C-Convergence monotone** Si $0 \leq X_n \uparrow X$, alors $\mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$, p.p.
- (f) **C-Fatou** Si $X_n \geq 0$, alors $\mathbb{E}(\liminf X_n|\mathcal{G}) \leq \liminf \mathbb{E}(X_n|\mathcal{G})$, p.p.
- (g) **C-Convergence dominée** Si $\forall n |X_n(\omega)| \leq V(\omega)$, $\mathbb{E}V < \infty$ et $X_n \rightarrow X$ p.p., alors

$$\mathbb{E}(|X_n - X| |\mathcal{G}) \rightarrow 0, \text{ p.p.}$$

(Par Jensen juste après, on déduit que $\mathbb{E}(X_n |\mathcal{G}) \rightarrow \mathbb{E}(X |\mathcal{G})$, p.p.)

- (h) **C-Jensen** Si $c : \mathbb{R} \rightarrow \mathbb{R}$ est convexe et $\mathbb{E}|c(X)| < \infty$, alors

$$\mathbb{E}(c(X)|\mathcal{G}) \geq c(\mathbb{E}(X|\mathcal{G})), \text{ p.p.}$$

Corollaire : $\|\mathbb{E}(X|\mathcal{G})\|_p \leq \|X\|_p$.

(i) **Emboitement** Si \mathcal{H} est une sous-tribu de \mathcal{G} , alors

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H}), \text{ p.p.}$$

(j) **”Sortir ce qui est connu”** Si Z est \mathcal{G} -mesurable et bornée, alors

$$\mathbb{E}(ZX|\mathcal{G}) = Z\mathbb{E}(X|\mathcal{G}), \text{ p.p. .}$$

Généralisations : encore vrai si $X \in L^p$ et $Z \in L^{p'}$, ou si $X \in (m\mathcal{F})^+$, $Z \in (m\mathcal{G})^+$, $\mathbb{E}(X) < \infty$ et $\mathbb{E}(ZX) < \infty$.

(k) **Rôle de l’indépendance** Si \mathcal{H} est indépendant de $\sigma(\sigma(X), \mathcal{G})$, alors

$$\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{H})) = \mathbb{E}(X|\mathcal{G}), \text{ p.p.}$$

En particulier, si X est indépendant de \mathcal{H} , alors $\mathbb{E}(X|\mathcal{H}) = E(X)$.

Exercice 14.6.2. Les propriétés fondamentales : les démontrer ([30] p. 89, 90)

(a) Provient de la relation fondamentale (14.6.1) appliquée à $G = \Omega$.

(b) Découle de l’unicité de la fonction \mathcal{G} -mesurable satisfaisant (14.6.1), i.e. de l’unicité de l’espérance conditionnelle.

(c) De même (expliquer!).

(d) Considérer l’ensemble $G = \{\mathbb{E}(X|\mathcal{G}) < -\frac{1}{n}\}$ et appliquer la relation fondamentale.

(e) Poser $Y = \limsup \mathbb{E}(X_n|\mathcal{G})$, montrer qu’elle est \mathcal{G} -mesurable, qu’elle est la limite presque partout des $\mathbb{E}(X_n|\mathcal{G})$. Montrer que $\lim_n \mathbb{E}(\mathbb{E}(X_n|\mathcal{G}); G) = \mathbb{E}(Y; G)$ et conclure.

(f) Reprendre la démonstration qui dérive le lemme de Fatou du théorème de convergence monotone pour déduire **C-Fatou** de **C-convergence monotone**. (Poser $Y_n = \mathbb{E}(X_n, \mathcal{G})$, puis $Z_n = \inf_{k \geq n} Y_k$ et lui appliquer **C-monotone**.)

(g) Reprendre la démonstration qui déduit le théorème de convergence monotone du lemme de Fatou pour déduire **C-convergence dominée** de **C-Fatou**.

(h)

(h1) Soit $c : \mathbb{R} \rightarrow \mathbb{R}$ convexe. Montrer qu’il existe une suite dénombrable $(a_n, b_n)_n \in \mathbb{R}^2$ telle que

$$c(x) = \sup_n (a_n x + b_n), \quad x \in \mathbb{R}.$$

(Remarquer que c est continue, d’épigraphe donc fermé, et utiliser par exemple Hahn-Banach : Il y a un hyperplan séparant tout point n’appartenant pas à un convexe fermé de ce convexe).

(h2) Déduire que

$$\mathbb{E}(c(X)|\mathcal{G}) \geq a_n \mathbb{E}(X|\mathcal{G}) + b_n$$

et conclure. Pour montrer l’extension de (h), prendre $c(x) = |x|^p$.

(i) A vous!

(j) La propriété est linéaire : Appliquer la "machine standard", c'est-à-dire montrer que la propriété est vraie quand Z est une fonction indicatrice d'élément G de \mathcal{G} , puis une fonction en escalier positive, puis une fonction \mathcal{G} -mesurable positive quelconque, puis une fonction \mathcal{G} -mesurable quelconque telle que $\mathbb{E}(|ZX|) < \infty$. Vérifier que la preuve s'applique aux généralisations indiquées.

(k) Supposer s.p.d.g. que $X \geq 0$ et $\mathbb{E}(X) < \infty$.

(k1) Montrer que si $G \in \mathcal{G}$ et $H \in \mathcal{H}$, alors

$$\mathbb{E}(X; G \cap H) = E(X \mathbb{1}_G) \mathbb{P}(H).$$

(k2) Poser $Y = \mathbb{E}(X|\mathcal{G})$. Montrer que

$$\mathbb{E}((Y \mathbb{1}_G) \mathbb{1}_H) = \mathbb{E}(Y \mathbb{1}_G) \mathbb{P}(H).$$

(k3) Dédire que les mesures $F \rightarrow \mathbb{E}(X; F)$ et $F \rightarrow \mathbb{E}(Y; F)$ définies sur $\sigma(\mathcal{G}, \mathcal{H})$ coïncident sur le π -système des ensembles de la forme $G \cap H$ ($G \in \mathcal{G}$, $H \in \mathcal{H}$) et donc sur $\sigma(\mathcal{G}, \mathcal{H})$. Conclure.

Exercice 14.6.3. Comprendre de quoi on parle

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et A_1, \dots, A_n une partition de Ω en ensembles mesurables et \mathcal{G} la sous-tribu engendrée par ces ensembles. Décrire cette tribu et calculer $E(X|\mathcal{G})$, pour X \mathcal{F} -mesurable.

Exercice 14.6.4. Lois de Bernoulli et uniforme ([26] Ex. 11.2 p.171)

Soient n variables aléatoires X_1, X_2, \dots, X_n indépendantes de même loi de Bernoulli de paramètre $p \in (0, 1)$. On note X le vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ à valeurs dans $\{0, 1\}^n$ et $S_n = \sum_{i=1}^n X_i$. Montrer que conditionnellement à $S_n = s$, le vecteur X est réparti selon une loi uniforme sur l'hyperplan $\{(k_1, k_2, \dots, k_n) | \sum_{i=1}^n k_i = s\}$ pour $0 \leq s \leq n$.

Exercice 14.6.5. Application importante

Soient X_1, X_2, \dots, X_n des variables aléatoires i.i.d. sommables et $S_n = X_1 + \dots + X_n$. On va montrer que $\mathbb{E}(X_1|S_n) = \dots = \mathbb{E}(X_n|S_n) = \frac{S_n}{n}$. On pose

$$\mathcal{G}_n = \sigma(S_n, S_{n+1}, \dots) = \sigma(S_n, X_{n+1}, X_{n+2}, \dots).$$

1) Montrer en utilisant (k) que

$$\mathbb{E}(X_1|\mathcal{G}_n) = \mathbb{E}(X_1|S_n).$$

2) Montrer que $\mathbb{E}(X_1|S_n) = \dots = \mathbb{E}(X_n|S_n)$.

(On montrera que pour tout borélien B , $\mathbb{E}(X_1; S_n \in B) = \dots = \mathbb{E}(X_n; S_n \in B)$.)

3) En déduire que

$$\mathbb{E}(X_1|S_n) = \dots = \mathbb{E}(X_n|S_n) = \frac{1}{n} \mathbb{E}(X_1 + \dots + X_n|S_n) = \frac{1}{n} S_n.$$

Exercice 14.6.6. Densités conditionnelles 1

Théorème 14.6.1. (rappel ([30] p. 87) et application : [7] p. 212). Soient X et Z deux v.a. qui ont une densité de probabilité jointe $f_{X,Z}(x, z)$. Alors $f_Z(z) = \int_{\mathbb{R}} f_{X,Z}(x, z) dx$ est une densité de probabilité pour Z . On définit la densité conditionnelle $f_{X|Z}$ de X sachant Z par

$$f_{X|Z}(x|z) := \frac{f_{X,Z}(x, z)}{f_Z(z)} \text{ si } f_Z(z) \neq 0, = 0 \text{ sinon.}$$

On pose alors

$$g(z) := \int_{\mathbb{R}} h(x) f_{X|Z}(x|z) dx.$$

Alors l'espérance conditionnelle de $h(X)$ sachant $\sigma(Z)$ est égale à $g(Z)$:

$$\mathbb{E}(h(X)|\sigma(Z)) = g(Z) = \int_{\mathbb{R}} h(x) f_{X|Z}(x|Z) dx.$$

Soient X, Y, Z trois variables aléatoires réelles telles que

- 1) X est uniformément répartie sur $[0, 1]$;
- 2) Sachant que $X = x$, Y admet une densité conditionnelle $f_{Y|X}(y|x)$ donnée par

$$f_{Y|X}(y|x) = \begin{cases} (y-x)e^{-(y-x)} & \text{si } y > x \\ 0 & \text{sinon} \end{cases}$$

- 3) Sachant que $X = x$ et $Y = y$, Z admet une densité conditionnelle $f_{Z|(X,Y)}(z|(x, y))$ définie par

$$f_{Z|(X,Y)}(z|(x, y)) = \begin{cases} (y-x)e^{-z(y-x)} & \text{si } z > 0 \\ 0 & \text{sinon} \end{cases}$$

pour $(x, y) \in A = \{(x, y) \in \mathbb{R}^2 | x \in [0, 1] \text{ et } y > x\}$.

- 1) Montrer que (X, Y, Z) admet une densité sur \mathbb{R}^3 donnée par

$$f_{(X,Y,Z)}(x, y, z) = (y-x)^2 e^{-(y-x)(z+1)} \mathbb{1}_B(x, y, z)$$

où $B = \{(x, y, z) \in \mathbb{R}^3 | x \in [0, 1], y > x, z > 0\}$.

- 2) Montrer que la loi de Z est donnée par la densité

$$f_Z(z) = \frac{2}{(z+1)^3} \mathbb{1}_{z>0}$$

- 3) Montrer que la loi conditionnelle de (X, Y) sachant $Z = z$ est donnée par la densité conditionnelle

$$f_{(X,Y)|Z}((x, y)|z) = \frac{1}{2}(z+1)^3 (y-x)^2 e^{-(y-x)(z+1)} \mathbb{1}_A(x, y)$$

- 4) Calculer $\mathbb{E}(\sqrt{Y-X}|Z=z)$, puis $\mathbb{E}(\sqrt{Y-X})$.
- 5) On pose $U = Y - X$ et $V = Z(Y - X)$. Montrer que (X, U, V) admet pour densité

$$f_{(X,U,V)}(x, u, v) = \mathbb{1}_{[0,1]}(x) u e^{-u} \mathbb{1}_{u>0, v>0}$$

- 6) X, U, V sont-elles indépendantes ?

Exercice 14.6.7. Densités conditionnelles 2

Soient X_1, \dots, X_n des variables aléatoires i.i.d. de densité commune $f(x)$ et de fonction de répartition $F(x) = \int_{-\infty}^x f(u)du$. On pose

$$Y = \max(X_1, \dots, X_n) \text{ et } Z = \min(X_1, \dots, X_n).$$

1) En remarquant que $F_{Y,Z}(y, z) = \mathbb{P}(Y \leq y) - \mathbb{P}(Y \leq y, Z > z)$, montrer que

$$F_{Y,Z}(y, z) = F(y)^n - (F(y) - F(z))^n \text{ si } y \geq z,$$

$$\text{et } F_{Y,Z}(y, z) = F(y)^n \text{ si } y < z.$$

2) En déduire que

$$f_{Y,Z}(y, z) = n(n - 1)(F(y) - F(z))^{n-2} f(y)f(z) \mathbb{1}_{y \geq z}.$$

3) En déduire une formule intégrale pour $\mathbb{E}(Z|Y)$.

4) Si les X_n sont équiréparties sur $[0, 1]$, en déduire que

$$f_{Z|Y}(z, y) = (n - 1) \frac{(y - z)^{n-2}}{y^n} \mathbb{1}_{0 \leq z \leq y \leq 1}.$$

Exercice 14.6.8. Processus de Poisson ([26] exercice 11.3, pp. 171-177)

Soit $(W_n)_{n \in \mathbb{N}^*}$ une suite croissante de variables aléatoires positives telles que $W_0 = 0$. Soit, pour $n \in \mathbb{N}^*$, la variable aléatoire $T_n = W_n - W_{n-1}$. On suppose que les variables aléatoires T_n forment une famille de variables aléatoires indépendantes de même loi exponentielle $\exp(\lambda)$ où λ est un réel strictement positif. On pose $X_0 = 0$ et pour tout $t > 0$,

$$X_t = \sum_{n \in \mathbb{N}^*} 1_{\{W_n \leq t\}}$$

La famille de variables aléatoires $(X_t)_{t \in \mathbb{R}^+}$ est appelée processus de Poisson d'intensité λ .

1) Montrer que \mathbb{P} presque sûrement $t \rightarrow X_t$ est une fonction càdlàg (continue à droite avec des limites à gauche) croissante et que les points de discontinuités sont les W_n .

2) Soit s, t tels que $0 \leq s < t$. Calculer l'intégrale définie pour tout $n \in \mathbb{N}^*$ par

$$I_n(s, t) = \int_{\mathbb{R}^n} 1_{(s \leq x_1 \leq \dots \leq x_n \leq t)} dx_1 \dots dx_n$$

On pourra soit procéder par récurrence comme Ouvrard, soit remarquer que si \mathcal{S}_n désigne le groupe des permutations sur $\{1, \dots, n\}$, on a

$$\forall \sigma \in \mathcal{S}_n, I_n(s, t) = \int_{\mathbb{R}^n} 1_{(s \leq x_{\sigma(1)} \leq \dots \leq x_{\sigma(n)} \leq t)} dx_1 \dots dx_n$$

et

$$\sum_{\sigma \in \mathcal{S}_n} 1_{(s \leq x_{\sigma(1)} \leq \dots \leq x_{\sigma(n)} \leq t)} = 1_{(x_1, \dots, x_n) \in [s, t]^n}$$

3) Calculer pour tout $n \in \mathbb{N}^*$ et toute famille $(f_j)_{j \in \{1, \dots, n\}}$ de fonctions positives mesurables bornées sur \mathbb{R} la quantité

$$\mathbb{E} \left[1_{(X_t = n)} \prod_{j=1}^n f_j(W_j) \right]$$

En déduire la loi de X_t et une loi conditionnelle de (W_1, \dots, W_n) sachant $\{X_t = n\}$.

Commencer par remarquer que $\{X_t = n\} = \{W_n \leq t\} \cap \{W_{n+1} > t\}$ et remplacer dans la quantité à calculer les W_k par $\sum_{i=1}^k T_i$. En utilisant la loi des T_i et leur indépendance, montrer que

$$\begin{aligned} \mathbb{E} \left[1_{(X_t = n)} \prod_{j=1}^n f_j(W_j) \right] &= \int_{(\mathbb{R}_+)^{n+1}} \left[1_{(\sum_{i=1}^n t_i \leq t) \cap (\sum_{i=1}^{n+1} t_i > t)} \prod_{j=1}^n f_j \left(\sum_{i=1}^j t_i \right) \right] \\ &\times \lambda^{n+1} \exp \left(-\lambda \sum_{i=1}^{n+1} t_i \right) dt_1 \dots dt_{n+1} \end{aligned}$$

Faire le changement de variables $w_1 = t_1, w_2 = t_1 + t_2, \dots, w_{n+1} = t_1 + \dots + t_{n+1}$ et en déduire

$$\begin{aligned} \mathbb{E} \left[1_{(X_t = n)} \prod_{j=1}^n f_j(W_j) \right] &= \lambda^n \exp(-\lambda t) \\ &\times \left[\int_{\mathbb{R}^n} \prod_{j=1}^n f_j(w_j) 1_{(0 \leq w_1 \leq w_2 \leq \dots \leq w_n \leq t)} dw_1 \dots dw_n \right] \end{aligned}$$

En faisant dans cette formule, $f_j = 1$, montrer que X_t suit une loi de Poisson de paramètre λt . En déduire que (W_1, \dots, W_n) admet une densité conditionnelle sachant $(X_t = n)$ donnée par

$$f_{(W_1, \dots, W_n)}(w_1, \dots, w_n | X_t = n) = \frac{n!}{t^n} 1_{(0 \leq w_1 \leq w_2 \leq \dots \leq w_n \leq t)} \quad (\text{loi de Dirichlet})$$

4) Montrer que le processus $(X_t)_{t \in \mathbb{R}_+}$ est à accroissements indépendants, c'est-à-dire que pour toute suite finie de réels positifs ordonnés $0 = t_0 \leq t_1 \leq \dots \leq t_k = t$, $(X_{t_{i+1}} - X_{t_i})_{i=0, \dots, k-1}$ est une famille de variables aléatoires indépendantes et donner la loi de $X_t - X_s$ pour $s \leq t$.

Soient $\alpha_1, \alpha_2, \dots, \alpha_k$ des entiers positifs quelconques. Notons n leur somme. On pose $l_j = \alpha_1 + \dots + \alpha_j$. En utilisant la question précédente, montrer

$$\mathbb{P} \left[\bigcap_{j=1}^k (X_{t_j} - X_{t_{j-1}} = \alpha_j) \right] = \lambda^n \exp(-\lambda t) \psi(t_1, \dots, t_k)$$

où

$$\Psi(t_1, \dots, t_k) = \int_{\mathbb{R}^n} \prod_{j=1}^{k-1} 1_{(w_{l_j} \leq t_j) \cap (w_{l_{j+1}} > t_j)} \times 1_{(0 \leq w_1 \leq w_2 \leq \dots \leq w_n \leq t)} dw_1 \dots dw_n$$

En déduire

$$\mathbb{P} \left[\bigcap_{j=1}^k (X_{t_j} - X_{t_{j-1}} = \alpha_j) \right] = \prod_{j=1}^k \exp(-\lambda(t_j - t_{j-1})) \frac{(\lambda(t_j - t_{j-1}))^{\alpha_j}}{\alpha_j!}$$

Cette formule démontre alors deux choses. D'une part l'indépendance des variables aléatoires $(X_{t_{i+1}} - X_{t_i})_{i=0, \dots, k-1}$ et d'autre part que la loi de $X_t - X_s$ est une loi de Poisson de paramètre $\lambda(t - s)$.

Chapitre 15

Martingales et inégalités maximales

15.1 Williams ou Ouvrard ?

Ouvrard ([26]) et Williams ([30]) démontrent le théorème de convergence des martingales bornées dans \mathbb{L}^2 et le théorème de convergence des martingales bornées dans \mathbb{L}^1 de deux manières différentes.

Williams commence par montrer le théorème général de convergence presque sûre des martingales bornées dans \mathbb{L}^1 , qui se prouve facilement dès que l'on a établi le lemme de passage. Quand il considère une martingale bornée dans \mathbb{L}^2 , il montre alors que c'est une suite de Cauchy dans \mathbb{L}^2 (mettant à profit l'orthogonalité des accroissements) puis utilise le lemme de Fatou conjugué au théorème de convergence des martingales bornées dans \mathbb{L}^1 .

Ouvrard opère différemment. Il ne démontre pas le théorème de convergence des martingales bornées dans \mathbb{L}^1 . Il commence par prouver le théorème d'inégalité maximale de Doob. Il en déduit l'inégalité de Doob qui dit que si $(X_n)_n$ est une martingale bornée dans \mathbb{L}^2 alors $X^* = \sup_n |X_n|$ est dans \mathbb{L}^2 et de plus on a l'inégalité de Doob

$$\|X^*\|_{\mathbb{L}^2} \leq 2 \sup_n \|X_n\|_{\mathbb{L}^2}$$

Le théorème de convergence pour les martingales bornées dans \mathbb{L}^2 apparaît alors comme un corolaire de cette inégalité. Plus exactement, comme Williams, il montre que la martingale est de Cauchy dans \mathbb{L}^2 , et il démontre qu'il y a convergence presque sûre de la martingale grâce à l'inégalité de Doob.

Il démontre ensuite le théorème de convergence presque sûre des martingales bornées dans \mathbb{L}^1 et le voit comme une conséquence du théorème de convergence des martingales bornées dans \mathbb{L}^2 . La démonstration est assez longue et fastidieuse.

En résumé, Williams commence par \mathbb{L}^1 pour en déduire ce qui se passe dans \mathbb{L}^2 alors qu'Ouvrard fait le contraire. Votons Williams !

15.2 Martingales et temps d'arrêt

15.3 Martingales

Dans la suite, $(\Omega, \mathcal{F}, \mathbb{P})$ désignera un espace probabilisé.

15.3.1 Filtrations et processus adaptés

Définition 15.3.1. On dit que $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ est un espace filtré sssi :

1) $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé.

2) $(\mathcal{F}_n)_n$ est une suite de sous-tribus de \mathcal{F} , croissante (au sens de l'inclusion).

On dit alors que $(\mathcal{F}_n)_n$ est une filtration et on pose

$$\mathcal{F}_\infty = \sigma(\cup_{n \geq 1} \mathcal{F}_n) \subset \mathcal{F}$$

Exemple fondamental : On se donne une suite de variables aléatoires $(W_n)_{n \geq 0}$. On pose alors $\mathcal{F}_n = \sigma\{W_0, \dots, W_n\}$. $(\mathcal{F}_n)_n$ est appelée la filtration naturelle associée à la suite de variables aléatoires $(W_n)_{n \geq 0}$.

Définition 15.3.2. Un processus sur $(\Omega, \mathcal{F}, \mathbb{P})$ est une suite $(W_n)_n$ de variables aléatoires sur Ω . Un processus sera dit adapté par rapport à une filtration $(\mathcal{F}_n)_n$ ssi pour tout entier n , W_n est \mathcal{F}_n -mesurable.

Remarque 15.3.1. Un processus $(W_n)_n$ est toujours adapté par rapport à sa filtration naturelle.

Ces notions sont souvent utilisés pour modéliser un certain nombre de processus physiques, biologiques, économiques... On peut en effet voir n comme le temps (discrétisation du temps mais les notions introduites se généralisent dans un cadre continu). \mathcal{F}_n est l'information que l'on connaît au temps n (grâce aux observations). Par exemple, si $(\mathcal{F}_n)_n$ est la filtration naturelle associée à un processus $(W_n)_n$ et si l'on peut observer $(W_n)_n$ alors on connaît \mathcal{F}_n et donc $f(W_0, \dots, W_n)$ pour toute fonction borélienne f . Dire qu'un processus $(X_n)_n$ est adapté par rapport à une filtration $(\mathcal{F}_n)_n$, c'est dire que l'au temps n , on connaît la valeur de X_n .

Rappel : X_n est $\sigma(W_0, \dots, W_n)$ -mesurable ssi il existe une fonction $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ borélienne telle que $X_n = f(W_0, \dots, W_n)$.

Définition 15.3.3. Un processus $(X_n)_n$ sera dit prévisible par rapport à une filtration $(\mathcal{F}_n)_n$ ssi pour tout $n \geq 1$, X_n est \mathcal{F}_{n-1} mesurable.

Remarque 15.3.2. A la différence d'un processus adapté, un processus prévisible est un processus que l'on peut prévoir (d'où son nom). En effet, si pour simplifier, on se place dans le cadre d'une filtration naturelle associée à un processus $(W_n)_n$ alors $(X_n)_n$ est prévisible ssi il existe une fonction borélienne f telle que

$$X_n = f(W_0, \dots, W_{n-1})$$

Autrement dit, on connaît X_n dès que l'on a pris connaissance du passé strict (c'est-à-dire W_0, \dots, W_{n-1}).

15.3.2 Martingale, Sur-martingale, Sous-martingale

Définition 15.3.4. Un processus $(X_n)_n$ est une $(\mathcal{F}_n)_n$ -martingale ssi :

- 1) $(\mathcal{F}_n)_n$ est une filtration ;
- 2) $(X_n)_n$ est adapté.
- 3) $\forall n \in \mathbb{N}, \mathbb{E}(|X_n|) < +\infty$
- 4) $\forall n \in \mathbb{N}, \mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$.

Un processus X sera une sur-martingale par rapport à $(\mathcal{F}_n)_n$ ssi il vérifie 1),2),3) et

$$4') \forall n \in \mathbb{N}, \mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq X_n$$

Un processus X sera une sous-martingale par rapport à $(\mathcal{F}_n)_n$ ssi il vérifie 1),2),3) et

$$4'') \forall n \in \mathbb{N}, \mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$$

Remarque 15.3.3. 1. X est une sur-martingale ssi $-X$ est une sous-martingale.

2. X martingale ssi c'est à la fois une sur et une sous-martingale.

3. X est une martingale ssi X_0 est intégrable et $X - X_0$ est une martingale.

4. L'ensemble des martingales a une structure d'espace vectoriel (ce qui est bien entendu faux pour les surmartingales et les sous martingales).

Proposition 15.3.1. Si X est une martingale (resp. sur-martingale, sous-martingale) alors $(\mathbb{E}(X_n))_n$ est une suite constante (resp. décroissante, resp. croissante).

15.3.3 Exemples

a) Somme de variables aléatoires indépendantes centrées

Soit $(X_i)_i$ une suite de variables aléatoires indépendantes, intégrables et centrées. On pose $S_0 = 0$ et $S_n = X_1 + \dots + X_n$. Soit $(\mathcal{F}_n)_n$ la filtration naturelle associée aux $(X_i)_i$ et $\mathcal{F}_0 = \{0, \emptyset\}$.

$(S_n)_n$ est une martingale puisque pour tout $n \geq 1$, on a :

$$\mathbb{E}(S_n|\mathcal{F}_{n-1}) = \mathbb{E}(S_{n-1}|\mathcal{F}_{n-1}) + \mathbb{E}(X_n|\mathcal{F}_{n-1}) = S_{n-1} + \mathbb{E}(X_n) = S_{n-1}$$

Bien entendu, si les X_i ne sont pas centrées mais de moyenne a alors $(S_n - na)_n$ est une martingale.

b) Produits de variables aléatoires indépendantes de moyenne 1

Soit X_1, \dots, X_n, \dots une suite de variables aléatoires indépendantes de moyenne 1 et soit $X_0 = 1$. Soit $(\mathcal{F}_n)_n$ la filtration naturelle associée à $(X_n)_n$. On pose

$$M_n = \prod_{k=0}^n X_k$$

$(M_n)_n$ est une martingale car pour tout $n \geq 1$, on a

$$\mathbb{E}(M_n|\mathcal{F}_{n-1}) = \mathbb{E}(X_n M_{n-1}|\mathcal{F}_{n-1}) = M_{n-1} \mathbb{E}(X_n) = M_{n-1}$$

Là aussi, dans le cas où les X_i ne sont pas de moyenne 1 mais de moyenne $a \neq 0$, on considèrera $M_n = a^{-n} \prod_{k=0}^n X_k$.

c) *Martingale obtenue par filtrage*

On se donne $\zeta \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ et une filtration $(\mathcal{F}_n)_n$. Pour tout $n \geq 0$, on pose

$$\mathbb{E}(\zeta | \mathcal{F}_n) = \zeta_n$$

$(\zeta_n)_n$ est une martingale.

Questions : Soit $(M_n)_n$ une martingale.

- 1) Est-ce qu'il existe une variable aléatoire M_∞ telle que $\lim_{n \rightarrow +\infty} M_n = M_\infty$? Dans quel sens a lieu cette convergence? (presque sûre, en probabilité, en loi, dans $\mathcal{L}^1, \mathcal{L}^2, \mathcal{L}^p$)
- 2) Si on a convergence, a-t-on $\mathbb{E}(M_0) = \mathbb{E}(M_\infty)$?
- 3) Dans le cas c), est-ce que l'on a

$$\zeta_\infty = \mathbb{E}(\zeta | \mathcal{F}_\infty) \text{ si } \lim_{n \rightarrow \infty} \zeta_n = \zeta_\infty$$

15.3.4 Illustration

On rappelle que l'on avait vu qu'une interprétation possible des notions de filtration, processus adaptés... était la suivante : n représente le temps et \mathcal{F}_n l'information dont on dispose au temps n . $(X_n)_n$ est $(\mathcal{F})_n$ -adapté signifie qu'au temps n , je connais X_n et $(X_n)_n$ est prévisible signifie qu'au temps $n-1$, je connais X_n . Avant de passer à l'interprétation de la notion de martingale, faisons un bref rappel sur la notion d'espérance conditionnelle.

Si X est dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ et Y est une variable aléatoire alors $\mathbb{E}(X | \sigma(Y))$ est la meilleure fonction $f(Y)$ de Y qui approche X au sens \mathcal{L}^2 .

Supposons donc maintenant que $X_n - X_{n-1}$ représente le profit que l'on fait au temps n en misant 1 euro dans un jeu de hasard (X_n représente donc le profit cumulé au temps n). Il est assez naturel de supposer que $X_n - X_{n-1}$ est \mathcal{F}_n -mesurable (au temps n , on a vu le résultat du jeu et l'on sait le profit que l'on a obtenu).

Si le jeu est parfaitement équitable, on ne doit rien pouvoir prévoir. On peut traduire cette hypothèse de la manière suivante. L'approximation de $(X_n - X_{n-1})$ au vu de l'information disponible au temps $n-1$, qui compte-tenu de ce qui précède est égale à $\mathbb{E}(X_n - X_{n-1} | \mathcal{F}_{n-1})$, doit être nulle. Autrement dit, on demande à $(X_n)_n$ d'être une martingale. Si $(X_n)_n$ est une sur-martingale alors le jeu est favorable au casino et si c'est une sous-martingale, le jeu est favorable au joueur.

Admettons que le casino se débrouille pour faire de $(X_n)_n$ une martingale (ou une sur-martingale). Est-ce que je peux me débrouiller pour gagner de l'argent en variant à chaque instant la mise (au lieu de mettre un euro à chaque fois)? Appelons C_n la mise placée au temps n . L'hypothèse naturelle à faire sur $(C_n)_n$ est de supposer que c'est un processus prévisible : on mise au vu de l'information dont on dispose au temps précédent de celui de la mise. Au temps n , on gagne donc

$$Y_n = \sum_{k=1}^n C_k (X_k - X_{k-1})$$

On appelle Y la transformée de X par C et l'on note $Y_n = (C \cdot X)_n$ (c'est la version discrétisée de l'intégrale stochastique relativement à une martingale).

Proposition 15.3.2. *Supposons pour simplifier que X soit une sur-martingale et C un processus prévisible borné (pour tout n , $|C_n(\omega)| \leq K$). Alors $C \cdot X$ est une sur-martingale.*

Ainsi, au temps n , mon esérance de gain est $\mathbb{E}(Y_n) \leq \mathbb{E}(Y_0) = 0$ et l'on est toujours perdant !

15.3.5 Temps d'arrêt

Définition 15.3.5. *Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ un espace probabilisé. Une application T de Ω dans $\bar{\mathbb{N}} = \{0, 1, \dots, \infty\}$ est un temps d'arrêt ssi*

$$\forall n \in \bar{\mathbb{N}}, \{T \leq n\} \in \mathcal{F}_n$$

Remarque 15.3.4. *Un peu de manipulation ensembliste montre que cette condition est équivalente à l'une des conditions suivantes :*

- 1) $\forall n \in \bar{\mathbb{N}}, \{T = n\} \in \mathcal{F}_n$.
- 2) $\forall n \in \bar{\mathbb{N}}, \{T \leq n\} \in \mathcal{F}_n$.
- 3) $\forall n \in \bar{\mathbb{N}}, \{T = n\} \in \mathcal{F}_n$.

Si on reprend à nouveau la modélisation précédente, on comprend aisément l'introduction de la notion de temps d'arrêt. On a vu que l'on ne pouvait gagner en variant la mise. Peut-être peut on gagner en s'arrêtant non plus à un temps donné n mais à un temps aléatoire. Cependant on doit faire un certain nombre d'hypothèses sur ce temps aléatoire T . En effet, il est réaliste de supposer que je m'arrêterai au temps n au vu de l'information dont je dispose au temps n (et non pas au temps $n + 1$ par exemple...). Autrement dit, on doit imposer à T de satisfaire $\{T = n\} \in \mathcal{F}_n$, c'est-à-dire d'être un temps d'arrêt.

Exemple : Soit $(A_n)_n$ un processus adapté et $B \in \mathcal{B}(\mathbb{R})$. On définit $T_B = \inf\{n \geq 0; A_n \in B\}$ comme le premier temps d'entrée dans B (éventuellement infini si l'ensemble est vide). T_B est un temps d'arrêt car

$$\{T \leq n\} = \cup_{k \leq n} \{A_k \in B\} \in \mathcal{F}_n$$

Contre-exemple : Le temps de sortie de B noté $L_B = \sup\{n, A_n \in B\}$ n'est pas un temps d'arrêt (en général).

15.3.6 Sur-martingales arrêtées

Théorème 15.3.1. *Si X est une martingale (resp. sur-martingale, sous-martingale) et T un temps d'arrêt alors $X^T = (X_{T \wedge n})_{n \geq 0}$ est appelée martingale (resp. sur-martingale, sous-martingale) arrêtée en T . C'est une martingale (resp. sur-martingale, sous-martingale). En particulier, on a pour tout entier n ,*

$$\mathbb{E}(X_n^T) = \mathbb{E}(X_{T \wedge n}) = \mathbb{E}(X_0), \quad (\text{resp. } \leq, \geq)$$

Remarque 15.3.5. *Ce théorème dit en outre que pour tout $n \in \mathbb{N}$, $X_{T \wedge n}$ est dans \mathcal{L}^1 . Il ne suppose aucune condition sur le temps d'arrêt.*

Démonstration. On a juste à montrer le théorème pour X surmartingale puisque X sous-martingale équivaut à $-X$ sur-martingale et X est une martingale ssi c'est à la fois une sur-martingale et une sous-martingale.

1) $X_{T \wedge n}$ est \mathcal{F}_n -mesurable car

$$X_{T \wedge n}(\omega) = \sum_{k=0}^{n-1} X_k(\omega) 1_{T(\omega)=k} + X_n(\omega) 1_{\{T(\omega) \geq n\}}$$

C'est une somme de variables aléatoires \mathcal{F}_n -mesurables.

2) Pour tout $n \in \mathbb{N}$, $X_{T \wedge n}$ est intégrable car $|X_{T \wedge n}| \leq \left(\sum_{k=0}^{n-1} |X_k| \right) + |X_n|$.

3) On a :

$$\begin{aligned} \mathbb{E}(X_{T \wedge n} | \mathcal{F}_{n-1}) &= \sum_{k=0}^{n-1} \mathbb{E}(X_k 1_{T=k} | \mathcal{F}_{n-1}) + \mathbb{E}(X_n 1_{T \geq n} | \mathcal{F}_{n-1}) \\ &= \sum_{k=0}^{n-1} 1_{T=k} \mathbb{E}(X_k | \mathcal{F}_{n-1}) + 1_{T \geq n} \mathbb{E}(X_n | \mathcal{F}_{n-1}) \\ &\leq \sum_{k=0}^{n-1} 1_{T=k} X_k + 1_{T \geq n} X_{n-1} = X_{T \wedge n-1} \end{aligned}$$

□

Si $(X_n)_n$ est une martingale et si T est un temps d'arrêt alors pour tout entier n , on a

$$\mathbb{E}(X_{T \wedge n}) = \mathbb{E}(X_0)$$

A-t-on $\mathbb{E}(X_T) = \mathbb{E}(X_0)$? Un contre exemple est le suivant. Soit $(X_n)_n$ une marche aléatoire telle que $X_0 = 0$ et $\mathbb{E}(X_{n+1} - X_n) = 0$. $(X_n)_n$ est une martingale. Soit $T = \inf\{n \in \mathbb{N}; X_n = 1\}$ le temps d'entrée en 1. On a vu que c'était un temps d'arrêt (par rapport à la filtration naturelle de X). Pourtant, on a

$$\mathbb{E}(X_T) = 1 \neq \mathbb{E}(X_0) = 0$$

15.3.7 Théorème d'arrêt de Doob

Théorème 15.3.2. (Théorème d'arrêt de Doob)

a) Soit T un temps d'arrêt et X une sur-martingale. Alors X_T est bien définie et intégrable avec

$$\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$$

dans l'une des 3 conditions suivantes :

- 1) T est bornée (i.e. : $\exists N \geq 0, \forall \omega \in \Omega |T(\omega)| \leq N$).
- 2) X est bornée (i.e. $\exists K \in \mathbb{R}_+, \forall n \in \mathbb{N}, \forall \omega \in \Omega, |X_n(\omega)| \leq K$) et T est \mathbb{P} p.s. fini.
- 3) $\mathbb{E}(T) < +\infty$ (donc T fini p.s.) et il existe $K \in \mathbb{R}_+$ tel que

$$\forall n \in \mathbb{N}, \forall \omega \in \Omega, |X_{n+1}(\omega) - X_n(\omega)| \leq K$$

b) En particulier, si 1), 2) ou 3) a lieu et si X est une martingale alors $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

Démonstration. b) est évident compte-tenu du fait que X est une martingale ssi X et $-X$ sont des sur-martingales.

a) On sait que $(X_{T \wedge n})_n$ est une sur-martingale donc

$$\mathbb{E}(X_{T \wedge n} - X_0) \leq 0$$

Pour 1), on prend $n = N$ et on a le résultat.

Pour 2), on utilise le théorème de convergence dominée vu que

$$|X_{T \wedge n}| \leq K \text{ et } \lim_{n \rightarrow +\infty} X_{T \wedge n} = X_T \text{ (} T \text{ fini)}$$

Pour 3), on remarque que

$$|X_{T \wedge n} - X_0| \leq \sum_{k=1}^{T \wedge n} |X_k - X_{k-1}| \leq TK$$

et $\mathbb{E}(T) < +\infty$ d'où le résultat par le théorème de convergence dominée. \square

Corollaire 15.3.1. *Soit M une martingale à incréments bornés (i.e. $\forall n \geq 1, |M_n - M_{n-1}| \leq K$ pour une constante $K > 0$) et soit C un processus prévisible borné par une constante et T un temps d'arrêt intégrable $\mathbb{E}(T) < +\infty$. On a alors*

$$\mathbb{E}((C \cdot M)_T) = 0$$

Démonstration. On a vu que $Y_n = (C \cdot M)_n$ était une martingale et d'autre part, on a

$$|Y_n - Y_{n-1}| \leq |C_n(M_n - M_{n-1})| \leq K_2 K_1$$

On conclut par le cas 3) du théorème précédent. \square

Proposition 15.3.3. *Soit X une sur-martingale positive et T un temps d'arrêt \mathbb{P} p.s. fini alors*

$$\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$$

Démonstration. $(X_n)_n$ est une sur-martingale de \mathcal{L}_+^1 et $x \in \mathbb{R}_+ \rightarrow x \wedge K$ est concave croissante bornée. Il est facile de montrer que si $(X_n)_n$ est une surmartingale positive et $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est croissante concave bornée alors $(\phi(X_n))_n$ est une surmartingale positive bornée (utiliser le théorème de Jensen version conditionnelle). On en déduit que $(X_n \wedge K)_n$ surmartingale positive bornée et d'après le théorème de Doob, cas 2),

$$\mathbb{E}(X_T \wedge K) \leq \mathbb{E}(X_0)$$

On conclut par le théorème de convergence monotone en faisant tendre K vers l'infini. \square

Application

Soit $(X_n)_n$ une suite de Bernoulli de paramètre 1/2 indépendantes : $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = 1/2$. On pose $S_0 = 0$ et $S_n = X_1 + \dots + X_n$ et $\mathcal{F}_n = \sigma(X_1, \dots, X_n) = \sigma(S_0, \dots, S_n)$. Soit enfin $T = \inf\{n \geq 0; S_n = 1\}$ qui est un temps d'arrêt. On cherche à connaître la distribution de T .

On montrera aisément que pour tout θ ,

$$M_n^\theta = \left(\frac{1}{\cosh \theta} \right)^n \exp(\theta S_n)$$

définit une martingale bornée de moyenne 1.

Soit $\theta > 0$. Pour tout $n \in \mathbb{N}$, on a

$$\mathbb{E}(M_{T \wedge n}^\theta) = 1 = \mathbb{E} \left(\left(\frac{1}{\cosh \theta} \right)^{T \wedge n} \exp(\theta S_{T \wedge n}) \right)$$

On remarque que \mathbb{P} presque sûrement,

$$\lim_{n \rightarrow \infty} M_{T \wedge n}^\theta = 1_{\{T < +\infty\}} \frac{e^\theta}{(\cosh \theta)^T}$$

donc d'après le théorème de convergence dominée,

$$\mathbb{E} \left[1_{\{T < +\infty\}} \frac{1}{(\cosh \theta)^T} \right] = e^{-\theta}$$

En faisant tendre θ vers 0 et en utilisant le théorème de convergence monotone, on déduit que $\mathbb{P}(T < +\infty) = 1$. On peut donc oublier l'indicatrice dans l'égalité précédente. Effectuant alors le changement de variables $\alpha = 1/\cosh(\theta)$, on obtient

$$\mathbb{E}(\alpha^T) = \frac{1}{\alpha} \left[1 - \sqrt{1 - \alpha^2} \right]$$

En particulier, on a

$$\mathbb{P}(T = 2m) = 0 \text{ et } \mathbb{P}(T = 2m - 1) = (-1)^{m+1} C(1/2, m)$$

15.4 TD : Jeux, ruine, temps d'arrêt

Exercice 15.4.1. Tribu engendrée par un nombre fini de v.a. discrètes Pour modéliser le jeu de pile ou face et tous les jeux répétés, pour les chaînes de Markov, pour les marches aléatoires, etc., on est amené à considérer une tribu engendrée par un nombre fini de variables aléatoires discrètes, $\mathcal{F} = \sigma(X_1, \dots, X_n)$. Les X_k sont à valeurs dans un ensemble de valeurs fini ou dénombrable. Pour fixer les notations et sans perte de généralité, on supposera que l'ensemble de ces valeurs est \mathbb{N} .

On considère tous les événements atomiques du type $A_m = [X_1 = m_1 \dots X_n = m_n]$, où $m = (m_1, \dots, m_n) \in \mathbb{N}^n$. Remarquer que ceux de ces événements qui sont non-vides constituent une partition de Ω et montrer que tout autre événement \mathcal{F} -mesurable est une union finie ou dénombrable d'atomes.

Exercice 15.4.2. Probabilité conditionnelle Soit A un événement de $(\Omega, \mathcal{F}, \mathbb{P})$ et \mathcal{G} une sous-tribu de \mathcal{F} . On appelle *probabilité conditionnelle de A sachant \mathcal{G}* et on note $\mathbb{P}(A|\mathcal{G})$ la variable aléatoire définie par

$$\mathbb{P}(A|\mathcal{G}) = \mathbb{E}(\mathbb{1}_A|\mathcal{G}).$$

1) On suppose que \mathcal{G} est engendrée par une partition finie ou dénombrable de Ω . Il existe donc I fini ou dénombrable tel que tout élément B de \mathcal{G} puisse s'écrire sous la forme $B = \cup_{i \in J} B_i$ où $J \subset I$; les B_i forment une partition dénombrable de Ω . Montrer que

$$\mathbb{P}(A|\mathcal{G}) = \sum_i \mathbb{P}(A|B_i) \mathbb{1}_{B_i}. \quad (15.4.1)$$

1) Réponse : On pose $X = \mathbb{P}(A|\mathcal{G})$ et $Y = \sum_i \mathbb{P}(A|B_i) \mathbb{1}_{B_i}$. Y est évidemment \mathcal{G} -mesurable. Par unicité de l'espérance conditionnelle, il suffit de montrer que $\forall B \in \mathcal{G}$, $E(X; B) = E(Y; B)$. Or par hypothèse, on peut écrire $B = \cup_{i \in J} B_i$, où J est fini ou dénombrable.

Par les définitions de l'espérance conditionnelle et de la probabilité conditionnelle, on a, puisque B est \mathcal{G} -mesurable,

$$\int_B X = \int_B \mathbb{E}(\mathbb{1}_A | \mathcal{G}) = \int_B \mathbb{1}_A = \sum_{i \in J} \mathbb{P}(A; B_i) = \int_B \sum_{j \in I} \frac{\mathbb{P}(A; B_i)}{\mathbb{P}(B_i)} \mathbb{1}_{B_i} = \int_B Y.$$

Exercice 15.4.3. "Tout ce qui a une chance raisonnable d'arriver se produira tôt ou tard" ([30], exercice E10.5 p. 233)

Soit \mathcal{F} une filtration (avec $\mathcal{F}_0 = \{\emptyset, \Omega\}$) et T un temps d'arrêt tels que pour un certain $N \in \mathbb{N}$ et un certain $\varepsilon > 0$,

$$\forall n, \mathbb{P}(T \leq n + N | \mathcal{F}_n) > \varepsilon, \text{ p.s.}$$

1) Montrer par récurrence en utilisant $\mathbb{P}(T > kN) = \mathbb{P}(T > kN; T > (k-1)N)$ que pour $k = 1, 2, \dots$,

$$\mathbb{P}(T > kN) \leq (1 - \varepsilon)^k.$$

1) Réponse : On procède par récurrence. Quand $n = 0$, on a bien la propriété par hypothèse car $\mathcal{F}_0 = \{\emptyset, \Omega\}$ et donc $\mathbb{P}(T > N) \mathbb{1}_\Omega = \mathbb{P}(T > N | \mathcal{F}_0) = 1 - \mathbb{P}(T \leq N | \mathcal{F}_0) \leq 1 - \varepsilon$. De plus,

$$\mathbb{P}(T > kN) = \mathbb{P}(T > kN; T > (k-1)N) = \int_{T > (k-1)N} \mathbb{1}_{T > kN} = \int_{T > (k-1)N} \mathbb{E}(\mathbb{1}_{T > kN} | \mathcal{F}_{(k-1)N})$$

(car l'ensemble $[T > (k-1)N]$ est $\mathcal{F}_{(k-1)N}$ -mesurable)

$$\leq (1 - \varepsilon) \mathbb{P}(T > (k-1)N) \text{ par hypothèse.}$$

2) En déduire que $\mathbb{E}(T) < \infty$.

2) Réponse : On a

$$\begin{aligned} \mathbb{E}(T) &= \sum_n n \mathbb{P}(T = n) \leq \sum_n n \mathbb{P}(T > n-1) \\ &\leq \sum_k \sum_{(k-1)N \leq n \leq kN} n \mathbb{P}(T > n-1) \\ &\leq N \sum_k kN \mathbb{P}(T > (k-1)N - 1) \\ &\leq N \sum_k kN \mathbb{P}(T > (k-2)N) \\ &\leq N^2 \sum_k k(1 - \varepsilon)^{k-2} < \infty \end{aligned}$$

Exercice 15.4.4. ABRACADABRA : une très longue attente (Grimmett-Stirzaker One thousand exercises in probability, exercice 16 p. 124 et [30], exercice E10.6 p. 233) On va résoudre par martingale le problème suivant :

Quelle est l'attente moyenne au jeu de pile ou face pour qu'une séquence préfixée se produise ? Prenons l'exemple de la séquence PPP (pile trois fois).

Un très grand casino contient une infinité de joueurs G_1, G_2, \dots qui disposent chacun d'une fortune de 1 euro. Un croupier tire à pile ou face (probabilités p et $q = 1 - p$ toutes les secondes. Au temps

n , le joueur G_n se met à parier de la manière suivante : il place 1 sur Pile. Le casino étant équitable, il touche en cas de succès $\frac{1}{p}$ (expliquer pourquoi). Il place alors à nouveau cette fortune sur Pile. Il continue ainsi à parier toute sa fortune sur Pile jusqu'à ce qu'il ait gagné trois fois de suite (PPP) ou qu'il perde tout. Dans les deux cas, il quitte alors le casino.

1) Soit S_n le profit (ou la perte) cumulé(e) du casino après le n -ième tirage. Montrer que S_n est une martingale.

1) Réponse : Soit Y_n la v.a. associée au n -ième tirage ($Y_n = F$ ou P) et $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$. Soit X_n la somme des gains et pertes des joueurs après le n -ième coup. Comme X_n est une fonction déterministe des résultats Y_n des n coups précédents, elle est \mathcal{F}_n -mesurable. Comme le jeu est équitable, la moyenne de X_n est nulle. De plus, le nombre de joueurs est plus petit que 3 et leur enjeu plus petit que p^{-3} . Donc X_n est sommable. S_n est donc l'exemple classique de martingale, à savoir une somme de v.a. sommables indépendantes et de moyenne nulle X_n .

2) Soit T le nombre de tirages effectués avant la première apparition de PPP . Montrer que T est un temps d'arrêt, montrer que $E(T) < \infty$. Utiliser le résultat de l'exercice 15.4.3.

2) Réponse : La décision $T = n$ est une fonction déterministe des résultats de Y_1, \dots, Y_n et est donc \mathcal{F}_n -mesurable. C'est donc un temps d'arrêt. Pour montrer que $E(T) < \infty$, il suffit de vérifier que l'hypothèse de l'exercice 15.4.3 est vérifiée : $\mathbb{P}(T \leq n + N | \mathcal{F}_n) > \varepsilon$, a.s.. On le montre pour $N = 3$. Notons $\Omega_n = \{P, F\}^n$ l'ensemble des résultats possibles pour les n premiers tirages. La tribu \mathcal{F}_n est engendrée par la partition en événements atomiques $B_i = [(Y_1, \dots, Y_n) = i]$, i décrivant Ω_n . Par la formule (15.4.1), on a pour tout événement A , $\mathbb{P}(A | \mathcal{F}_n) = \sum_{i \in \Omega_n} \mathbb{P}(A | B_i) \mathbb{1}_{B_i}$. On choisit $A = [T \leq n + 3]$. En effet, une séquence PPP peut se produire aux trois coups suivants avec probabilité $\varepsilon = p^3$ et on a

$$\mathbb{P}(T \leq n + 3 | \mathcal{F}_n) \geq \sum_{i \in \Omega_n} \mathbb{P}([(X_{n+1} X_{n+2} X_{n+3}) = (PPP)] | B_i) \mathbb{1}_{B_i} = p^3 \sum_i \mathbb{1}_{B_i} = p^3.$$

3) En déduire que $\mathbb{E}(T) = p^{-1} + p^{-2} + p^{-3}$.

3) Réponse : On applique le théorème de temps d'arrêt de Doob (10.10 (c) de Williams page 100-101). Donc $E(S_T) = 0$. Mais, au moment où le jeu s'arrête, les joueurs G_1, \dots, G_n ont misé chacun -1 et seuls G_{n-2} , G_{n-1} et G_n ont gagné respectivement p^{-3} , p^{-2} et p^{-1} . Donc $E(S_T) = 0$ donne $p^{-1} + p^{-2} + p^{-3} - \mathbb{E}T = 0$.

4) Adapter le raisonnement pour calculer le temps moyen d'attente de PF .

4) Réponse : on trouve $\mathbb{E}(N) = p^{-1} + p^{-2}q^{-1}$.

5) Dans le même esprit : le casino possède un singe qui tape au hasard sur les 26 touches majuscules à la vitesse de 60 caractères par minute. Montrer que le temps moyen d'attente de la séquence ABRACADABRA est $26^{11} + 26^4 + 26$. Donner un ordre de grandeur en années du temps d'attente.

6) Un paradoxe ? Les calculs précédents prouvent que le temps moyen d'attente de PP (deux fois Pile) dans le cas $p = \frac{1}{2}$ est égal à $2 + 4 = 6$ alors que le temps d'attente de PF est de 4. Ceci peut paraître contreintuitif, puisque les séquences PP et PF sont équiprobables ! De même, on vient de voir que le temps d'attente de ABRACADABRA est supérieur au temps d'attente de, disons, ABCDEFGHIJK, qui est une séquence de même longueur et donc équiprobable. Discuter ce paradoxe. Vous convaincre par une simulation de pile ou face (utiliser directement une monnaie) que l'expérience confirme bien la différence de temps d'attente pour PP et PF .

Exercice 15.4.5. La ruine du joueur : quelles chances de gagner et au bout de combien de temps ?

Soient X_1, X_2, \dots des v.a.i.i.d. avec $P(X = 1) = p$, $P(X = -1) = q = 1 - p$ et $0 < p < 1$. Supposons que a et b soient deux entiers avec $0 < a < b$. On définit

$$S_n := a + X_1 + \dots + X_n, \quad T := \inf\{n, S_n = 0 \text{ ou } S_n = b\}.$$

La fortune initiale du joueur de pile ou face est représentée par a et il compte s'arrêter s'il a atteint b . On peut aussi interpréter ce modèle comme un jeu à deux : alors la fortune initiale du premier joueur est a et celle du second est $b - a$. $S_n - a$ représente alors le gain cumulé du premier joueur et $a - S_n$ les pertes cumulées du second. Chacun des joueurs stoppe quand il est ruiné, i.e. quand $S_n = 0$ ou $S_n = b$. Soit $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ et $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

1) Vérifier que la condition de l'exercice 15.4.3 est vérifiée.

1) Réponse : On procède comme au 2) de l'exercice 4, avec $N = b$. On note $\Omega_n = \{-1, +1\}^n$. Alors \mathcal{F}_n est engendrée par les événements atomiques $B_i = [(X_1, \dots, X_n) = i]$, $i \in \Omega_n$. On applique la formule (15.4.1) avec $A = [T \leq n + b]$:

$$\mathbb{P}(T \leq n + b | \mathcal{F}_n) = \sum_{i \in \Omega_n} P(T \leq n + b | B_i) \mathbb{1}_{B_i}. \quad (15.4.2)$$

On a $\mathbb{P}(T \leq n + b | B_i) = \mathbb{P}(T \leq n | B_i) + \mathbb{P}(n < T \leq n + b | B_i)$. Or : $T > n \Rightarrow a + i_1 + \dots + i_n \in]0, b[$. Donc $[n < T \leq n + b] \cap B_i \supset [X_{n+1} = \dots = X_{n+b} = 1] \cap B_i$. Donc

$$\mathbb{P}(T \leq n + b | B_i) \geq \mathbb{P}(X_{n+1} = \dots = X_{n+b} = 1 | B_i) = \mathbb{P}(X_{n+1} = \dots = X_{n+b} = 1) = p^b.$$

En effet, $B_i \in \mathcal{F}_n$ est indépendant des X_{n+k} , $k \geq 1$. Donc (15.4.2) donne $\mathbb{P}(T \leq n + b | \mathcal{F}_n) \geq p^b$. En fait, comme dans l'exercice 4, ce long raisonnement conduit à une inégalité évidente, mais il était bon de détailler sa preuve pour vérifier que nos impressions sur la probabilité conditionnelle sont justes.

2) On pose $M_n = (\frac{q}{p})^{S_n}$ et $N_n = S_n - n(p - q)$. Montrer que ce sont des martingales et déduire les valeurs de $\mathbb{P}(S_T = 0)$ et $E(T)$.

2) Réponse : M_n est un produit de v. a. indépendantes, positives et de moyenne 1. N_n est une somme de v.a. aléatoires indépendantes sommables et de moyenne nulle. La suite tout seul.

15.5 Le théorème de convergence des martingales

Dans la suite, $(\Omega, \mathcal{F}, \mathbb{P})$ désignera un espace probabilisé muni d'une filtration $(\mathcal{F}_n)_n$.

On considère un processus adapté X et $a < b$ deux réels, N un entier non nul. La suite finie de réels, pour ω donné, $\{X_0(\omega), \dots, X_N(\omega)\}$ est appelée la trajectoire de X entre les instants 0 et N . Nous introduisons quelques notations.

★) $U_N^{[a,b]}(\omega)$ est défini comme le plus grand entier $k \geq 1$ tel qu'il existe deux suites d'entiers non nuls (s_1, \dots, s_k) et (t_1, \dots, t_k) avec

$$0 \leq s_1 < t_1 < \dots < s_k < t_k \leq N$$

$$\forall i \in \{1, \dots, k\}, X_{s_i}(\omega) < a \text{ et } X_{t_i}(\omega) > b$$

$U_N^{[a,b]}$ apparaît comme le nombre de passages croissants de a à b par la trajectoire de X entre 0 et N .

★) On définit le processus prévisible $(C_n)_n \in \{0, 1\}^{\mathbb{N}}$ de la manière suivante :

$$\begin{aligned} C_1 &= 1_{\{X_0 < a\}} \\ C_n &= 1_{\{C_{n-1}=1\}} 1_{\{X_{n-1} \leq b\}} + 1_{\{C_{n-1}=0\}} 1_{\{X_{n-1} < a\}} \end{aligned}$$

★) On notera enfin par $Y = (C \cdot X)$ la transformée de X par C , c'est-à-dire le processus donné par

$$Y_n(\omega) = \sum_{k=1}^n C_k(\omega)(X_k(\omega) - X_{k-1}(\omega))$$

Le théorème de convergence des martingales bornées dans \mathcal{L}^1 est basé sur la remarque suivante facile à démontrer :

$$Y_N(\omega) \geq (b-a)U_N^{[a,b]}(\omega) - [X_N(\omega) - a]^-$$

Lemme 15.5.1. (de passage de Doob)

Si X est une sur-martingale alors, avec les notations introduites précédemment, on a

$$\mathbb{E}(U_N^{[a,b]}) \leq \frac{\mathbb{E}((X_N - a)^-)}{b-a}$$

Démonstration. C est un processus prévisible positif borné et X est une sur-martingale donc $Y = C \cdot X$ est une sur-martingale (cf la section précédente). Donc

$$\mathbb{E}(Y_N) \leq \mathbb{E}(Y_0) = 0$$

On reporte dans l'inégalité précédente et on obtient le résultat. □

Corollaire 15.5.1. Soit X une surmartingale bornée dans \mathcal{L}^1 dans le sens où $M = \sup_n \mathbb{E}(|X_n|) < +\infty$. On note

$$U_\infty^{[a,b]} = \lim_{n \rightarrow \infty} U_n^{[a,b]}$$

On a : $(b-a)\mathbb{E}_\infty^{[a,b]} \leq |a| + M$ et par suite

$$\mathbb{P}(U_\infty^{[a,b]} = \infty) = 0$$

On obtient alors le théorème principal de cette section sur la convergence des martingales.

Théorème 15.5.1. (de convergence de Doob)

Soit X une sur-martingale bornée dans \mathcal{L}^1 . \mathbb{P} presque sûrement, la limite notée $X_\infty(\omega)$ de $(X_n(\omega))_n$ existe et est finie. De plus, la variable aléatoire X_∞ est \mathcal{F}_∞ -mesurable et dans \mathcal{L}^1 .

Démonstration. Soit $\Lambda = \{\omega \in \Omega; (X_n(\omega))_n \text{ ne converge pas dans } \mathbb{R}\}$. On veut montrer que cet ensemble est de probabilité nulle.

Or $\Lambda = \bigcup_{\substack{a,b \in \mathbb{Q} \\ a < b}} \Lambda_{a,b}$ où

$$\Lambda_{a,b} = \{\omega \in \Omega; \liminf X_n(\omega) < a < b < \limsup X_n(\omega)\}$$

et

$$\Lambda_{a,b} \subset \{\omega \in \Omega; U_{\infty}^{[a,b]}(\omega) = \infty\}$$

donc $\mathbb{P}(\Lambda_{a,b}) = 0$ par le lemme de passage de Doob et par σ -additivité de \mathbb{P} , $\mathbb{P}(\Lambda) = 0$.
D'autre part, par le lemme de Fatou,

$$\mathbb{E}(|X_{\infty}|) = \mathbb{E}(\liminf |X_n|) \leq \liminf \mathbb{E}(|X_n|) < +\infty$$

□

Remarque 15.5.1. *Pour l'instant, rien ne garantit qu'il y ait convergence dans \mathcal{L}^1 .*

Corollaire 15.5.2. *Si X est une sur-martingale positive alors $X_{\infty} = \lim X_n$ existe \mathbb{P} p.s.*

Démonstration. $0 \leq \mathbb{E}(|X_n|) = \mathbb{E}(X_n) \leq \mathbb{E}(X_0) < +\infty$ et il suffit d'appliquer le théorème de convergence de Doob. □

Chapitre 16

Martingales bornées dans \mathcal{L}^2 .

16.1 Le théorème de base de convergence des martingales de \mathcal{L}^2

Dans la suite, $(\Omega, \mathcal{F}, \mathbb{P})$ désignera un espace probabilisé muni d'une filtration $(\mathcal{F}_n)_n$. Dans cette section, nous considérerons des martingales dans \mathcal{L}^2 . La structure particulière d'espace de Hilbert de \mathcal{L}^2 facilitera de beaucoup notre étude.

16.1.1 Martingales de \mathcal{L}^2 : orthogonalité des accroissements

Définition 16.1.1. On dira qu'une martingale $(M_n)_n$ est bornée dans \mathcal{L}^2 ssi $\sup_n \|M_n\|_2 < +\infty$.

Notation : On note $\langle X, Y \rangle = \mathbb{E}(XY)$ le produit scalaire sur $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ et $\|\cdot\|_2$ la norme associée.

Une propriété fondamentale des martingales est *l'orthogonalité de ses accroissements*, c'est-à-dire que si l'on a des entiers tels que $s \leq t \leq u \leq v$ alors

$$\langle M_t - M_s, M_v - M_u \rangle = 0$$

En effet, on a

$$\begin{aligned} \mathbb{E}((M_t - M_s)M_v) &= \mathbb{E}(\mathbb{E}((M_t - M_s)M_v | \mathcal{F}_u)) \\ &= \mathbb{E}((M_t - M_s)\mathbb{E}(M_v | \mathcal{F}_u)) = \mathbb{E}((M_t - M_s)M_u) \end{aligned}$$

et la formule

$$M_n = M_0 + \sum_{k=1}^n (M_k - M_{k-1})$$

exprime M_n comme une somme d'éléments orthogonaux. Le théorème de Pythagore donne donc

$$\|M_n\|_2^2 = \|M_0\|_2^2 + \sum_{k=1}^n \|M_k - M_{k-1}\|_2^2$$

Théorème 16.1.1. Soit M une martingale de \mathcal{L}^2 , c'est-à-dire que pour tout $n \in \mathbb{N}$, $M_n \in \mathcal{L}^2$. $(M_n)_n$ est uniformément bornée dans \mathcal{L}^2 ssi

$$\sum_{k=1}^{\infty} \|M_k - M_{k-1}\|_2^2 < +\infty$$

Dans ce cas, $\lim_{n \rightarrow \infty} M_n = M_\infty$ \mathbb{P} p.s. et dans \mathcal{L}^2 .

Démonstration. La première partie du théorème est évidente d'après ce qui précède. \mathcal{L}^2 est complet et le théorème de Doob assure que $\lim_{n \rightarrow +\infty} M_n = M_\infty$ \mathbb{P} p.s. D'autre part,

$$\mathbb{E}((M_{n+r} - M_n)^2) \leq \sum_{k=n+1}^{n+r} \mathbb{E}((M_k - M_{k-1})^2)$$

et par le lemme de Fatou, quand r tend vers $+\infty$, on a

$$\mathbb{E}((M_\infty - M_n)^2) \leq \sum_{k=n+1}^{\infty} \mathbb{E}((M_k - M_{k-1})^2)$$

et ce reste de série convergente (première partie du théorème) tend vers 0 quand n tend vers $+\infty$, ce qui prouve la deuxième partie du théorème. \square

16.1.2 Décomposition de Doob

Définition 16.1.2. Un processus $(A_n)_n$ sera dit prévisible nul en 0 ssi $A_0 = 0$ et pour tout $n \geq 1$, A_n est \mathcal{F}_{n-1} -mesurable.

Théorème 16.1.2. (de décomposition de Doob)

a) Soit $(X_n)_{n \geq 1}$ un processus adapté dans \mathcal{L}^1 . Alors X a une décomposition de Doob :

$$X = X_0 + M + A$$

où M est une martingale nulle en 0 et A un processus prévisible nul en 0.

De plus cette décomposition est unique dans le sens où si $X = X_0 + \tilde{M} + \tilde{A}$, \tilde{M} martingale nulle en 0 et \tilde{A} processus prévisible nul en 0, alors :

$$\mathbb{P}(\forall n \in \mathbb{N}, M_n = \tilde{M}_n, A_n = \tilde{A}_n) = 1$$

b) X est une sous-martingale ssi le processus A est croissant dans le sens suivant :

$$\forall n \in \mathbb{N}, A_n \leq A_{n+1} \text{ } \mathbb{P} \text{ p.s.}$$

Démonstration. \star) Si la décomposition existe, on doit avoir

$$\begin{aligned} \mathbb{E}(X_n - X_{n-1} | \mathcal{F}_{n-1}) &= \mathbb{E}(M_n - M_{n-1} | \mathcal{F}_{n-1}) + \mathbb{E}(A_n - A_{n-1} | \mathcal{F}_{n-1}) \\ &= 0 + A_n - A_{n-1} \end{aligned}$$

donc $A_n = \sum_{k=1}^n \mathbb{E}(X_k - X_{k-1} | \mathcal{F}_{k-1})$ \mathbb{P} p.s.

et $M_n = X_n - X_0 - A_n = X_n - X_0 - \sum_{k=1}^n \mathbb{E}(X_k - X_{k-1} | \mathcal{F}_{k-1})$, $n \geq 1$.

Ceci prouve l'unicité de la décomposition.

★) Pour ce qui est de l'existence de la décomposition, on définit A_n et M_n par les formules précédentes. Il est facile de vérifier que M est une martingale nulle en 0 et A un processus prévisible nul en 0.

★) X est une sous-martingale ssi

$$\forall n \geq 1, \mathbb{E}(X_n | \mathcal{F}_{n-1}) \geq X_{n-1}$$

ce qui est équivalent, d'après la formule définissant A , à

$$\forall n \geq 1, A_n - A_{n-1} \geq 0$$

□

16.1.3 Le crochet d'une martingale $\langle M \rangle$

Proposition 16.1.1. *Soit M une martingale de \mathcal{L}^2 nulle en 0. Alors M^2 est une sous-martingale.*

Démonstration. Utiliser la propriété de Jensen version conditionnelle. □

On écrit la décomposition de Doob de la sous-martingale M^2 sous la forme $M^2 = N + A$ où N martingale nulle en 0 et A processus prévisible croissant nul en 0. On posera $A_\infty = \lim_{n \rightarrow \infty} A_n$.

Définition 16.1.3. *Le processus prévisible $(A_n)_n$ est appelé le crochet de M et noté $\langle M \rangle$.*

Proposition 16.1.2. *Une martingale M de \mathcal{L}^2 est uniformément bornée dans \mathcal{L}^2 ssi $\mathbb{E}(\langle M \rangle_\infty) < +\infty$.*

Remarque 16.1.1. *On a $A_n - A_{n-1} = \mathbb{E}(M_n^2 - M_{n-1}^2 | \mathcal{F}_{n-1}^2) = \mathbb{E}((M_n - M_{n-1})^2 | \mathcal{F}_{n-1})$.*

Lien entre la convergence de M et la finitude de $\langle M \rangle_\infty$

Soit M une martingale de \mathcal{L}^2 nulle en 0. Notons le crochet de M par $\langle M \rangle = A$. Le théorème suivant éclaircit le lien qui existe entre bornitude du crochet et la convergence de la martingale.

Théorème 16.1.3. *a) $\lim_{n \rightarrow \infty} M_n(\omega)$ existe pour \mathbb{P} p.s. tous les ω tels que $A_\infty(\omega) < +\infty$.*

b) Si M a des accroissements uniformément bornés (i.e. $\forall n \geq 1, \forall \omega \in \Omega, |(M_n(\omega) - M_{n-1}(\omega))| \leq K$).

Alors : $\lim_{n \rightarrow \infty} M_n(\omega)$ existe ssi $A_\infty(\omega) < +\infty$.

Démonstration. a) A est prévisible donc $\forall k \in \mathbb{N}, S_k = \inf\{n; A_{n+1} > k\}$ est un temps d'arrêt. D'autre part, A^{S_k} est un processus prévisible nul en 0 (le vérifier) borné par k . $(M^2 - A)^{S_k}$ est une martingale en tant que martingale arrêtée. Puisque $\sup_n \mathbb{E}((M_n^{S_k})^2) = \sup_n \mathbb{E}(A_n^{S_k}) \leq k$, M^{S_k} est bornée dans \mathcal{L}^2 et $\lim_{n \rightarrow \infty} M_{n \wedge S_k}$ existe \mathbb{P} p.s. et dans \mathcal{L}^2 . Or on a

$$\{A_\infty < +\infty\} = \cup_k \{S_k = \infty\}$$

d'où sur $\{A_\infty < +\infty\}$, il existe k (fixé) tel que $S_k = \infty$ et $\lim M_{n \wedge S_k} = \lim M_n$ existe donc.

b) Si l'on avait $\mathbb{P}(A_\infty = \infty, \sup_n |M_n| < +\infty) > 0$ alors pour un $c > 0$, $\mathbb{P}(T_c = \infty, A_\infty = \infty) > 0$ où l'on a posé $T_c = \inf\{r; |M_r| > c\}$.

En effet,

$$\{A_\infty = \infty, \sup_n |M_n| < \infty\} = \cup_{c \in \mathbb{N}} \{A_\infty = \infty, T_c = \infty\}$$

donc

$$0 < \mathbb{P} \left[A_\infty = \infty, \sup_n |M_n| < \infty \right] \leq \sum_{c \in \mathbb{N}} \mathbb{P}(A_\infty = \infty; T_c = \infty)$$

et il existe un $c \in \mathbb{N}$ tel que $\mathbb{P}(T_c = \infty, A_\infty = \infty) > 0$.

Or $\mathbb{E}(M_{T_c \wedge n}^2 - A_{T_c \wedge n}) = 0$ et M^{T_c} est bornée par $c + K$ d'où

$$\mathbb{E}(A_{T_c \wedge n}) \leq (c + K)^2$$

Par le théorème de convergence monotone pour le processus croissant A , on aurait

$$+\infty \cdot \mathbb{P}(T_c = \infty, A_\infty = \infty) \leq (c + K)^2$$

ce qui est bien entendu exclu. En conclusion, on a

$$\mathbb{P}(A_\infty = \infty, \sup_n |M_n| < \infty) = 0$$

En utilisant a), on finit aisément la preuve. □

16.1.4 Application

Théorème 16.1.4. Soit $(X_k)_k$ une suite de variables aléatoires indépendantes telles que pour tout $k \in \mathbb{N}$, on ait

$$\mathbb{E}(X_k) = 0, \sigma_k^2 = \text{Var}(X_k) < +\infty$$

a) Alors $\sum_k \sigma_k^2 < +\infty \Rightarrow \sum_k X_k$ converge \mathbb{P} p.s.

b) Si les $(X_k)_k$ sont bornées par une constante $K \geq 0$ tel que $|X_k| \leq K$ pour tout k , \mathbb{P} p.s. alors

$$\left(\sum_k X_k \right) \text{ converge p.s. } \Rightarrow \sum_k \sigma_k^2 < +\infty$$

Démonstration. On rappelle que la loi du 0–1 nous dit que $\mathbb{P}(\sum_k X_k \text{ converge}) \in \{0, 1\}$. Soit la filtration $(\mathcal{F}_n)_n$ associée au processus $X = (X_0, X_1, \dots, X_n, \dots)$ et $M_n = X_1 + \dots + X_n$, $M_0 = 0$. M est une martingale. La décomposition de Doob de la sous-martingale M^2 est donnée par

$$M_n^2 = N_n + A_n$$

où N est une martingale nulle en 0 et $A_n = \sum_{k=1}^n \sigma_k^2$. Le théorème précédent donne le résultat. □

A titre d'exemple, on se donne une suite de signes aléatoires ϵ , c'est-à-dire une suite de variables aléatoires indépendantes $(\epsilon_n)_n$ telles que $\mathbb{P}(\epsilon_n = 1) = \mathbb{P}(\epsilon_n = -1) = 1/2$ et une suite déterministe de réels $(a_n)_n$. Le théorème précédent montre que $\sum_n \epsilon_n a_n$ converge \mathbb{P} p.s. ssi $\sum_n a_n^2 < +\infty$.

16.2 TD Probabilités : Inégalités maximales

Ce td est composé d'un théorème relatif à une inégalité maximale qu'il faut connaître et savoir démontrer. Le deuxième théorème est une conséquence du premier (à savoir retrouver). Le troisième théorème concerne la loi du log-itéré (dont on trouvera l'énoncé complet dans [30] en appendice) dans le cas "facile" des lois normales. On trouvera dans [16] un énoncé semblable avec cette fois-ci des lois de Bernoulli au lieu des lois normales. Enfin, le td se termine par un exercice modélisant un portefeuille d'actions tiré de [26] qui peut aussi servir à illustrer (y réfléchir) les exposés sur la loi des grands nombres et le TCL.

Théorème 16.2.1. Inégalité maximale de Doob : [30], chap. 14

Soit $(Z_n)_n$ une sous martingale positive par rapport à une filtration $(\mathcal{F}_n)_n$. Alors pour tout $c > 0$, on a

$$c\mathbb{P}\left(\sup_{0 \leq k \leq n} Z_k \geq c\right) \leq \mathbb{E}(Z_n)$$

1) Soit pour tout $k \in \{0, \dots, n\}$, $F_k = \{Z_0 < c\} \cap \{Z_1 < c\} \cap \dots \cap \{Z_{k-1} < c\} \cap \{Z_k \geq c\}$ avec $F_0 = \{Z_0 \geq c\}$. Montrer que $F = \{\sup_{0 \leq k \leq n} Z_k \geq c\}$ est la réunion disjointe des F_k et que pour tout k , $F_k \in \mathcal{F}_k$.

2) Utiliser la propriété de sous-martingale pour montrer que $c\mathbb{P}(F_k) \leq \mathbb{E}(Z_n; F_k)$ et conclure.

Solution

1) Pour tout ω , il y a une première valeur de k telle que $Z_k(\omega) \geq c$. Donc cet ω est dans F_k . Les Z_i sont \mathcal{F}_k -mesurables pour $i \leq k$. Donc les ensembles $Z_i < c$ et $Z_k \geq c$ sont \mathcal{F}_k mesurables et il en est donc de même de F_k .

2) En vertu de la propriété de sous-martingale et comme F_k est \mathcal{F}_k -mesurable, $\mathbb{E}(Z_n; F_k) \geq \mathbb{E}(Z_k; F_k) \geq c\mathbb{P}(F_k)$. On somme sur tous les $k = 1, \dots, n$ et on obtient le résultat parce que F est l'union disjointe des F_k et que $\mathbb{E}(Z_n; F) \leq \mathbb{E}Z_n$.

Exercice 16.2.1. Inégalité de Kolmogorov : [30], chap. 14

Théorème 16.2.2. Soit $(X_n)_n$ une suite de variables aléatoires indépendantes centrées et de variances respectives σ_n^2 finies. On pose $S_n = X_1 + \dots + X_n$ et $V_n = \text{Var}(S_n) = \sigma_1^2 + \dots + \sigma_n^2$. Alors pour tout c strictement positif, on a :

$$c^2\mathbb{P}\left(\sup_{1 \leq k \leq n} |S_k| \geq c\right) \leq V_n$$

1) Montrer que S_n^2 est une sous-martingale.

2) Plus généralement, montrer que si M est une martingale et c une fonction convexe telle que $\forall n \mathbb{E}(|c(M_n)|) < \infty$, alors $c(M)$ est sous-martingale.

3) Appliquer maintenant le théorème précédent. Comparer avec l'inégalité qu'on peut obtenir directement pour $|S_n|$ par Tchebycheff et en déduire que l'inégalité est bien un raffinement.

Solution

1) On remarque que $S_n - S_k$ et S_k sont indépendantes et centrées. On a donc $\mathbb{E}(S_n^2 | \mathcal{F}_k) = \mathbb{E}((S_n - S_k)^2 + S_k^2 + 2S_k(S_n - S_k) | \mathcal{F}_k)$
 $\geq \mathbb{E}(S_k^2 | \mathcal{F}_k) + 2\mathbb{E}(S_k(S_n - S_k) | \mathcal{F}_k)$
 $= \mathbb{E}(S_k^2 | \mathcal{F}_k) + S_k \mathbb{E}(S_n - S_k | \mathcal{F}_k)$ (par la propriété (j) de l'espérance conditionnelle)
 $= \mathbb{E}(S_k^2 | \mathcal{F}_k) + S_k \mathbb{E}(S_n - S_k | \mathcal{F}_k) = \mathbb{E}(S_n - S_k) = 0$ p.s. par la propriété (k).

2) Cela vient de la propriété (h) de l'espérance conditionnelle : si $c : \mathbb{R} \rightarrow \mathbb{R}$ est convexe et si $\mathbb{E}|c(X)| < \infty$, alors $\mathbb{E}(c(X) | \mathcal{G}) \geq c(\mathbb{E}(X | \mathcal{G}))$. On a donc pour $k \leq n$, $\mathbb{E}(c(M_n) | \mathcal{G}_k) \geq c(\mathbb{E}(M_n | \mathcal{G}_k)) = c(M_k)$.

3) On a $\mathbb{E}(S_n^2) = \text{Var}(S_n) = \sigma_1^2 + \dots + \sigma_n^2$ et donc l'application du 1) est directe. Pour $|S_n|$, on peut directement appliquer l'inégalité de Tchebycheff : $\mathbb{P}(|S_k| \geq c) \leq \frac{\text{Var}(S_k)}{c^2}$ et donc $c^2 \sup_k \mathbb{P}(|S_k| \geq c) \leq V_n$ pour tout k . L'inégalité de Doob est donc une amélioration nette : elle nous autorise à commuter sup et \mathbb{P} dans la formule précédente.

Théorème 16.2.3. Un cas particulier de la loi du log itéré : [30], chap. 14

Soit $(X_n)_n$ une suite i.i.d. de normales standard. On pose $S_n = X_1 + \dots + X_n$. Alors, presque sûrement, on a :

$$\limsup \frac{S_n}{(2n \log \log n)^{1/2}} = 1$$

et

$$\liminf \frac{S_n}{(2n \log \log n)^{1/2}} = -1$$

1) Montrer que la deuxième partie du théorème est équivalente à la première en considérant $-X_i$ au lieu de X_i .

2) On pose $\phi(y) = (2\pi)^{-1/2} \exp(-y^2/2)$ et $\Phi(x) = \int_{-\infty}^x \phi(y) dy$.

a) En remarquant que $\phi'(y) = -y\phi(y)$, montrer que pour tout x strictement positif, $x^{-1}\phi(x) \geq (1 - \Phi(x))$.

b) En remarquant que $(y^{-1}\phi(y))' = -(1 + y^{-2})\phi(y)$, montrer que pour tout $x > 0$, $\phi(x) \leq (x + x^{-1})(1 - \Phi(x))$.

c) En déduire un encadrement de $\mathbb{P}(X > x)$ pour $x > 0$.

3) Montrer que $\mathbb{P}(S_n > \sqrt{2n \log \log n}) \leq \frac{1}{\log n \sqrt{4\pi \log \log n}}$. Montrer que cette inégalité est aussi une équivalence. Comparer avec le résultat du théorème.

4) Obtention d'une borne exponentielle. a) Soit $(\mathcal{F}_n)_n$ la filtration naturelle associée aux X_i . Montrer que $\exp(\theta S_n)$ est une sous-martingale relativement à cette filtration pour tout θ réel.

b) En déduire, en utilisant l'inégalité maximale de Doob, que pour $\theta > 0$,

$$\mathbb{P} \left(\sup_{1 \leq k \leq n} S_k \geq c \right) \leq e^{-\theta c + \frac{1}{2} \theta^2 n}$$

c) En déduire

$$\mathbb{P} \left(\sup_{1 \leq k \leq n} S_k \geq c \right) \leq e^{-\frac{1}{2} c^2 / n}$$

5) Obtention d'une borne sup.

On se donne un $K > 1$ proche de 1 et on pose $c_n = Kh(K^{n-1})$ où $h(x) = \sqrt{2x \log \log x}$.

a) En utilisant BC1, montrer que presque sûrement, pour n assez grand, si k est un entier compris entre K^{n-1} et K^n , $S_k \leq Kh(k)$.

b) En déduire que $\limsup h(k)^{-1} S_k \leq 1$.

6) Obtention d'une borne inf.

Soit N un entier très grand. Soit ϵ un réel strictement positif petit. Soit $F_n = \{S_{N^{n+1}} - S_{N^n} > (1 - \epsilon)h(N^{n+1} - N^n)\}$.

a) En utilisant la question 1), montrer que

$$\mathbb{P}(F_n) \geq (y + y^{-1})^{-1} \phi(y)$$

où $y = (1 - \epsilon)\{2 \log \log(N^{n+1} - N^n)\}^{1/2}$.

b) Utiliser BC2 pour montrer que presque sûrement, une infinité de F_n ont lieu.

c) En déduire que l'on a presque sûrement pour une infinité de n

$$S_{N^{n+1}} > (1 - \epsilon)h(N^{n+1} - N^n) + S_{N^n}$$

d) **attention : étape non expliquée dans [30]**. En appliquant le résultat du 5)b) aux $-X_i$, on a presque sûrement, pour tout $\epsilon > 0$, $S(N^n) \geq -(1 + \epsilon)h(N^n)$ pour n assez grand.

e) Montrer alors que $\limsup h(k)^{-1} S_k \geq (1 - \epsilon)(1 - N^{-1})^{1/2} - 2N^{-1/2}$. Conclure.

Solution

1) Immédiat

2) a) En intégrant entre x et $+\infty$ la relation différentielle, on obtient $-\phi(x) = -\int_x^{+\infty} y\phi(y) dy \leq -x \int_x^{+\infty} \phi(y) dy = -x(1 - \Phi)$. Donc $(1 - \Phi) = \int_x^{+\infty} \phi(y) dy \leq \frac{\phi(x)}{x}$.

b) En intégrant la relation proposée, on a $-x^{-1}\phi(x) = \int_x^{+\infty} (1 + y^{-2})\phi(y) dy \geq -(1 + x^{-2})(1 - \Phi)$. En combinant a) et b), on obtient

$$\frac{x^{-1}}{1 + x^2} e^{-\frac{x^2}{2}} \leq \int_x^{+\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi} \mathbb{P}(X > x) \leq x^{-1} e^{-\frac{x^2}{2}}.$$

3) Comme S_n est une somme de n variables i.i.d.n.c., elle est gaussienne centrée de variance n . Donc, en utilisant la majoration donnée au 2),

$$\mathbb{P}(S_n > \sqrt{2n \log \log n}) = \int_{\sqrt{2n \log \log n}}^{+\infty} \frac{1}{\sqrt{2\pi n}} e^{-\frac{x^2}{2n}} dx$$

$$\int_{\sqrt{2 \log \log n}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \leq \frac{1}{\sqrt{2\pi}} \frac{e^{-\log \log n}}{\sqrt{2 \log \log n}} = \frac{1}{\log n \sqrt{4\pi \log \log n}}.$$

Les deux résultats sont compatibles et complémentaires mais aucun des deux n'implique l'autre, car on vient d'obtenir un ordre de grandeur pour la probabilité que ne peut donner la loi du log itéré.

4) a) Ceci découle immédiatement de la convexité de l'exponentielle et du résultat de l'exercice 16.2.1, question 2).

4)b) On a en utilisant l'inégalité de Doob

$$\mathbb{P}(\sup_{k \leq n} S_k \geq c) = \mathbb{P}(\sup_{k \leq n} e^{\theta S_k} \geq e^{\theta c}) \leq e^{-\theta c} \mathbb{E}(e^{\theta S_n}).$$

Comme les X_i sont indépendantes, on a $\mathbb{E}e^{\theta S_n} = (\mathbb{E}e^{\theta X})^n$. De plus, $\mathbb{E}e^{\theta X} = \int e^{\theta x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = e^{\frac{\theta^2}{2}}$, ce qui donne le résultat.

4) c) Il suffit de choisir la valeur de θ qui donne la meilleure inégalité, à savoir $\theta = \frac{c}{n}$.

5) On a

$$\mathbb{P}(\sup_{k \leq K^n} S_k \geq c_n) \leq e^{-\frac{c_n^2}{2K^n}} = (n-1)^{-K} (\log K)^{-K}.$$

Cette série étant convergente, Borel-Cantelli 1 nous dit que presque sûrement, pour n assez grand et $K^{n-1} \leq k \leq K^n$,

$$S_k \leq \sup_{k \leq K^n} S_k \leq c_n = Kh(K^{n-1}) \leq Kh(k).$$

Donc pour $K > 1$,

$$\limsup_k h(k)^{-1} S_k \leq K, \text{ p.s.,}$$

ce qui donne la première majoration en faisant tendre K vers 1.

6)a) On écrira $S(n)$ pour S_n pour des raisons typographiques. $S(N^{n+1}) - S(N^n)$ est une variable gaussienne dont on note la variance $J = N^{n+1} - N^n$. Par l'inégalité prouvée au 1)b), on a donc

$$\mathbb{P}(F_n) = \int_{(1-\varepsilon)\sqrt{2K \log \log K}}^{+\infty} \frac{1}{\sqrt{2\pi K}} e^{-\frac{x^2}{2K}} = \int_y^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \geq \frac{y + y^{-1}}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}.$$

6) b)c) On a pour une constante C , $\mathbb{P}(F_n) \geq \frac{C}{\log \log N^{n+1}} \frac{1}{[\log(N^{n+1} - N^n)]^{(1-\varepsilon)^2}}$, clairement une série divergente. Or, les F_n sont indépendants. Donc, presque sûrement, un nombre infini de F_n se produit, c'est-à-dire

$$S(N^{n+1}) > (1 - \varepsilon)h(N^{n+1} - N^n) + S(N^n)$$

pour une infinité de n .

6) e) On obtient donc

$$\limsup_n \frac{S(N^{n+1})}{h(N^{n+1})} \geq \limsup_n \frac{(1 - \varepsilon)h(N^{n+1} - N^n) - 2(1 + \varepsilon)h(N^n)}{h(N^{n+1})} =$$

$$\limsup_n \frac{(1 - \varepsilon)\sqrt{N^{n+1} - N^n} \sqrt{\log \log(N^{n+1} - N^n)} - 2(1 + \varepsilon)\sqrt{2N^n \log \log N^n}}{\sqrt{N^{n+1} \log \log N^{n+1}}} \\ \approx (1 - \varepsilon)(1 - N^{-\frac{1}{2}})^{\frac{1}{2}} - 2(1 + \varepsilon)N^{-\frac{1}{2}}.$$

Il suffit alors de faire N arbitrairement grand et ε arbitrairement petit.

Exercice 16.2.2. Un modèle de portefeuille d'actions : [26] pp. 405-410

$S = (S_n)_{n \geq 0}$ est un processus stochastique décrivant l'évolution temporelle (le temps est discret) d'une action boursière avec bruit. Plus exactement, la suite de v.a. $(S_n)_n$ est définie par :

$$S_0 = s_0 > 0, \text{ et si } n \in \mathbb{N}^*, S_n = (1 + \mu)S_{n-1} + \sigma S_{n-1} \epsilon_n$$

où $(\epsilon_n)_n$ est un processus de bruits qui est ici une suite de Bernoulli indépendantes prenant les valeurs ± 1 avec même probabilité. μ et σ sont deux paramètres vérifiant $|\sigma| < 1 + \mu$. On note $\lambda = \sqrt{(1 + \mu)^2 - \sigma^2}$. Et enfin soit \mathcal{F}_n la filtration naturelle associée au processus S .

1) Soit f la fonction définie sur \mathbb{R}^2 par $f(x, y) = (1 + \mu)x + \sigma xy$. Montrer que pour tout $x \geq 0$, $f(x, 1) \geq 0$ et $f(x, -1) \geq 0$.

2) a) Calculer $\mathbb{E}(S_n | \mathcal{F}_{n-1})$ et montrer que S est une martingale pour $\mu = 0$, une sous-martingale pour $\mu > 0$ et une sur-martingale pour $\mu < 0$.

b) On se place dans le cas $\mu < 0$. Montrer que $(S_n)_n$ converge presque sûrement. Etablir une relation de récurrence sur $\mathbb{E}(S_n)$ et en déduire que la limite de $(S_n)_n$ est nulle.

3) Vérifier que S_n est de carré intégrable et montrer que $\mathbb{E}(S_n^2) = s_0^2 [(1 + \mu)^2 + \sigma^2]^n$.

4) Soit Z le processus $\log(S)$. Montrer que selon que λ est supérieur, égal ou inférieur à 1, Z est une sous-martingale, une martingale, ou une sur-martingale.

a) Montrer que pour tout entier $n \geq 1$,

$$Z_n = Z_0 + \sum_{j=1}^n \log(1 + \mu + \sigma \epsilon_j)$$

b) En déduire, en utilisant la loi des grands nombres, que $(Z_n)_n$ converge presque sûrement vers $+\infty$ et $\frac{Z_n}{n}$ converge presque sûrement vers 0 si $\lambda > 1$, 0 si $\lambda = 1$ et $-\infty$ si $\lambda < 1$.

5) On se place dans le cas particulier $|\sigma| < 2^{-1/2}$ et $(1 + \mu)^2 + \sigma^2 < 1$.

a) Vérifier que l'on est dans le cas où $-S$ est une sous-martingale.

b) Ecrire la décomposition de Doob de $-S = M + A$ avec M martingale de carré intégrable et A processus croissant prévisible nul en 0 de carré intégrable. Plus exactement, montrer que :

$$A_n = -\mu \sum_{j=0}^{n-1} S_j$$

$$M_0 = s_0, \text{ et si } n \geq 1, M_n = -S_n + \mu \sum_{j=1}^{n-1} S_j$$

c) Montrer que le crochet de M s'écrit :

$$\langle M \rangle_n = \sigma^2 \sum_{j=0}^{n-1} S_j^2$$

d) En utilisant la question 3), montrer que $(M_n)_n$ converge presque sûrement.

6) On définit le processus R par $R_n = \lambda^{-\sqrt{n}} S_n^{1/\sqrt{n}}$ et on pose $\rho^2 = \frac{1}{4} \left\{ \log \left(\frac{1 + \mu + \sigma}{1 + \mu - \sigma} \right) \right\}^2$.

a) Montrer que

$$\log(R_n) = \frac{1}{\sqrt{n}} \left[Z_0 + \sum_{j=1}^n (\log(1 + \mu + \sigma \epsilon_j) - \log(\lambda)) \right]$$

b) Montrer à l'aide du TCL que la suite des lois $(\mathbb{P}_{\log R_n})_n$ converge étroitement vers une loi $\mathcal{N}(0, \rho^2)$.

c) En déduire que $(R_n)_n$ converge étroitement vers la probabilité de densité par rapport à la mesure de Lebesgue donnée par :

$$1_{y>0} \frac{1}{|\rho|y\sqrt{2\pi}} \exp\left(-\frac{(\log y)^2}{2\rho^2}\right)$$

C'est la loi Log-normale de paramètres 0 et ρ^2 .

Chapitre 17

Martingales uniformément intégrables

Bibliographie : [30] chapitres 13 et 14, [5], section 1.3.

17.1 L'uniforme intégrabilité

Remarque 17.1.1. Fatou inverse : Si $f_n \geq 0$ mesurables, $f_n \leq g$ et $g \in \mathbb{L}^1$, alors $\int \limsup f_n \geq \limsup \int f_n$. *Démonstration* : appliquer Fatou à $(g - f_n)$.

Théorème 17.1.1. Soit $X \in \mathbb{L}^1 = \mathbb{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Alors pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que pour tout $F \in \mathcal{F}$, $\mathbb{P}(F) < \delta$ implique $\mathbb{E}(|X|; F) < \varepsilon$.

Démonstration. Sinon, il existe $\varepsilon_0 > 0$ et une suite F_n d'éléments de \mathcal{F} tels que $\mathbb{P}(F_n) < 2^{-n}$ et $\mathbb{E}(|X|; F_n) \geq \varepsilon_0$. Soit $H := \limsup F_n$. Alors Borel Cantelli implique que $\mathbb{P}(H) = 0$. En utilisant Fatou inverse :

$$0 = \mathbb{E}(|X|; H) \geq \limsup \mathbb{E}(|X|; F_n) \geq \varepsilon_0, \quad \text{contradiction.}$$

□

Corollaire 17.1.1. Si $X \in \mathbb{L}^1$ et $\varepsilon > 0$. Alors il existe $K \in [0, \infty)$ tel que

$$\mathbb{E}(|X|; |X| > K) < \varepsilon.$$

Démonstration. On a $\mathbb{P}(|X| > K) \leq \frac{\mathbb{E}(|X|)}{K}$. Donc pour K assez grand on peut avoir $\mathbb{P}(|X| > K) < \delta$, où on fixe δ comme dans le théorème 17.1.1. □

Définition 17.1.1. Une classe \mathcal{C} de variables aléatoires est dite uniformément intégrable (UI) si pour tout $\varepsilon > 0$, il existe $K \in [0, \infty)$ tel que

$$\mathbb{E}(|X|; |X| > K) < \varepsilon, \quad \forall X \in \mathcal{C}.$$

Remarque 17.1.2. On a alors avec K_1 associé à $\varepsilon = 1$, $\mathbb{E}(|X|) \leq \mathbb{E}(|X|; |X| > K_1) + \mathbb{E}(|X|; |X| \leq K_1) \leq 1 + K_1$ et donc une famille UI est aussi bornée dans \mathbb{L}^1 . Il n'est pas vrai qu'une famille bornée dans \mathbb{L}^1 est UI : prendre sur $[0, 1]$ muni de la mesure de Lebesgue, $f_n = n\mathbb{1}_{0, \frac{1}{n}}$.

Proposition 17.1.1. *Si \mathcal{C} est bornée dans \mathcal{L}^p pour un $p > 1$, alors elle est uniformément intégrable. Si \mathcal{C} a un chapeau intégrable Y elle est uniformément intégrable.*

Démonstration.

$$\mathbb{E}(|X|; |X| > K) \leq K^{1-p} \mathbb{E}(|X|^p; |X| > K) \leq K^{1-p} \sup_{\mathcal{C}} \mathbb{E}(|X|^p);$$

$$\mathbb{E}(|X|; |X| > K) \leq \mathbb{E}(Y; |X| > K) \leq \mathbb{E}(Y; |Y| > K)$$

et on applique le corollaire 17.1.1. □

Voici un excellent développement : Le théorème de de la Vallée-Poussin donne une caractérisation intuitive et facile à montrer des familles UI de \mathbb{L}^1 . ([5] p. 10).

Théorème 17.1.2. *\mathcal{C} est UI si et seulement s'il existe une fonction mesurable croissante $G : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ telle que*

$$\lim_{t \rightarrow \infty} \frac{G(t)}{t} \rightarrow +\infty \text{ et } \sup_{X \in \mathcal{C}} \mathbb{E}(G(|X|)) < \infty.$$

Il est aussi important de savoir donner une condition d'uniforme intégrabilité adaptée à \mathbb{R}^N . C'est la suivante :

Définition 17.1.2. *On dit que $\mathcal{C} \subset \mathbb{L}^1(\mathbb{R}^N)$ est uniformément intégrable si pour tout $\varepsilon > 0$ il existe $R > 0$ et $K > 0$ tels que pour tout $f \in \mathcal{C}$,*

$$\int_{|x| \geq R} f(x) dx < \varepsilon \text{ et } \int_{|f| \geq K} f(x) dx < \varepsilon.$$

La plupart des théorèmes qui suivent se généralisent avec cette définition de l'UI sur \mathbb{R}^N .

17.1.1 Les espérances conditionnelles d'une même variable X sont équiintégrables

Théorème 17.1.3. *Soit $X \in \mathbb{L}^1$. Alors la classe $\{\mathbb{E}(X|\mathcal{G}) : \mathcal{G} \text{ est une sous-tribu de } \mathcal{F}\}$ est uniformément intégrable.*

Démonstration. On fixe $\varepsilon > 0$ et choisit δ tel que pour $F \in \mathcal{F}$, on ait

$$\mathbb{P}(F) < \delta \Rightarrow \mathbb{E}(|X|; F) < \varepsilon.$$

Enfin, on choisit K tel que $\frac{\mathbb{E}(|X|)}{K} < \delta$. Soit \mathcal{G} une sous-algèbre et Y une version quelconque de $\mathbb{E}(X|\mathcal{G})$; Par Jensen appliquée à l'espérance conditionnelle ,

$$|Y| = |\mathbb{E}(X|\mathcal{G})| \leq \mathbb{E}(|X| | \mathcal{G}), \text{ p.p.} \tag{17.1.1}$$

Donc $\mathbb{E}(|Y|) \leq \mathbb{E}(|X|)$ et

$$K\mathbb{P}(|Y| > K) \leq \mathbb{E}(|Y|) \leq \mathbb{E}(|X|),$$

et donc $\mathbb{P}(|Y| > K) < \delta$. Mais $\{|Y| > K\} \in \mathcal{G}$, et donc par la définition de probabilité conditionnelle et par (17.1.1),

$$\mathbb{E}(|Y|; |Y| \geq K) \leq \mathbb{E}(\mathbb{E}(|X| | \mathcal{G}); |Y| \geq K) = \mathbb{E}(|X|; |Y| \geq K) < \varepsilon.$$

□

17.1.2 Convergence en probabilité

Lemme 17.1.1. (Vitali) Soit C un borné de \mathbb{R}^N et f_k est une suite de fonctions équi-intégrable dans $\mathbb{L}^1(C)$ qui converge presque partout vers une fonction f . Alors f_k converge vers f dans $\mathbb{L}^1(C)$.

On peut faire mieux avec la convergence en probabilité et donner une équivalence complète!

Définition 17.1.3. On dit que $X_n \rightarrow X$ en probabilité si pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

Lemme 17.1.2. Si $X_n \rightarrow X$ p.p. alors $X_n \rightarrow X$ en probabilité.

Démonstration. Supposons que $X_n \rightarrow X$ p.p. (rappel : les lemmes Fatou, Beppo-Levy, convergence monotone, s'appliquent aux fonctions caractéristiques d'ensembles avec les correspondances suivantes :

- $\mathbb{P}(A) = \mathbb{E}(\mathbb{1}_A)$,
- à la limite sup de fonctions caractéristiques $\mathbb{1}_{A_n}$ d'ensembles A_n correspond l'événement $[A_n, \text{ infiniment souvent (i.s.)}]$
- à leur limite inf : $[A_n, \text{ finalement (eventually) }].$)

Par le lemme de Fatou appliqué aux ensembles,

$$0 = \mathbb{P}((X_n \not\rightarrow X)) \geq \mathbb{P}(|X_n - X| > \varepsilon \text{ i.s.}) = \mathbb{P}(\limsup |X_n - X| > \varepsilon) \geq \limsup \mathbb{P}(|X_n - X| \geq \varepsilon).$$

□

17.1.3 Théorème de convergence bornée

Théorème 17.1.4. Soit X_n une suite de v.a. et X une v.a. Si $X_n \rightarrow X$ en probabilité et s'il existe une borne $K \in \mathbb{R}^+$, telle que $|X_n| \leq K$ pour tout n , alors

$$\mathbb{E}(|X_n - X|) \rightarrow 0.$$

Démonstration. On vérifie que $\mathbb{P}(|X| \leq K) = 1$. Or pour tout $k \in \mathbb{N}^*$,

$$\mathbb{P}(|X| > K + k^{-1}) \leq \mathbb{P}(|X - X_n| > k^{-1}) \rightarrow 0.$$

Donc $\mathbb{P}(|X| > K) = \mathbb{P}(\cup_k [|X| > K + k^{-1}]) = 0$. On a alors pour tout $\varepsilon > 0$,

$$\mathbb{E}(|X_n - X|) = \mathbb{E}(|X_n - X|; |X_n - X| > \varepsilon) + \mathbb{E}(|X_n - X|; |X_n - X| \leq \varepsilon) \leq 2K\mathbb{P}(|X_n - X| > \varepsilon) + \varepsilon < 2\varepsilon$$

pour n assez grand.

□

17.1.4 Condition nécessaire et suffisante de convergence \mathbb{L}^1 .

Théorème 17.1.5. Soit (X_n) une suite de \mathbb{L}^1 et $X \in \mathbb{L}^1$. Alors $X_n \rightarrow X$ dans \mathbb{L}^1 si et seulement si

- (i) $X_n \rightarrow X$ en probabilité.
- (ii) La suite X_n est UI.

Démonstration. On tronque X_n pour la rendre bornée en posant $X_n^k = \max(-k, \min(k, X_n))$. Par définition de UI, on peut choisir K tel que

$$\mathbb{E}(|X_n^k - X_n|) < \varepsilon, \quad \mathbb{E}(|X^k - X|) < \varepsilon.$$

Comme $|X_n^k - X^k| \leq |X_n - X|$, $X_n^k \rightarrow X^k$ en probabilité. Par le théorème de convergence bornée, on déduit que $X_n^k \rightarrow X^k$ dans \mathbb{L}^1 . L'inégalité triangulaire permet de conclure que $X_n \rightarrow X$ dans \mathbb{L}^1 .

Réciproque : Si $X_n \in \mathbb{L}^1$, $X_n \rightarrow X$ dans \mathbb{L}^1 , alors $X_n \rightarrow X$ en probabilité et X_n est UI. La première relation est une conséquence de l'inégalité de Markov,

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}(|X_n - X|).$$

Montrons que X_n est UI. On choisit N assez grand pour que $\mathbb{E}(|X_n - X|) < \varepsilon$. On remarque que toute famille finie dans \mathbb{L}^1 est UI et donc il nous suffit de régler le sort des X_n pour $n \geq N$. On fixe δ tel que $\mathbb{P}(F) < \delta \Rightarrow \mathbb{E}(X; F) < \varepsilon$. Par l'inégalité de Markov à nouveau

$$\mathbb{P}(|X_n| > K) \leq K^{-1} \sup_n \mathbb{E}(|X_n|) < \delta$$

pour K assez grand. Donc pour $n \geq N$,

$$\mathbb{E}(|X_n|; |X_n| > K) \leq \mathbb{E}(|X|; |X_n| > K) + \mathbb{E}(|X - X_n|) < 2\varepsilon.$$

□

17.2 Martingales UI

Définition 17.2.1. On dit que M est une martingale UI si c'est une martingale dans le cadre habituel $(\Omega, \mathcal{F}_n, \mathbb{P})$ et si M_n est une suite UI. Comme M est UI, elle est bornée dans \mathbb{L}^1 et donc $M_\infty := \lim M_n$ existe presque sûrement : $\mathbb{E}(|M_n - M_\infty|) \rightarrow 0$.

Théorème 17.2.1. Soit M une martingale UI. Alors $M_\infty := \lim M_n$ vérifie

$$M_n = \mathbb{E}(M_\infty | \mathcal{F}_n).$$

Démonstration. Pour $F \in \mathcal{F}_n$, et $r \geq n$, on a

$$\mathbb{E}(M_r; F) = \mathbb{E}(E(M_r | \mathcal{F}_n); F) = \mathbb{E}(M_n; F).$$

Mais

$$|\mathbb{E}(M_r; F) - \mathbb{E}(M_\infty; F)| \leq \mathbb{E}(|M_r - M_\infty|; F) \leq \mathbb{E}(|M_r - M_\infty|) \rightarrow 0.$$

Donc

$$\mathbb{E}(M_\infty; F) = \mathbb{E}(M_n; F). \tag{17.2.1}$$

□

17.2.1 Théorème "montant" de Lévy.

Théorème 17.2.2. Soit $\xi \in \mathbb{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, et soit $M_n := \mathbb{E}(\xi | \mathcal{F}_n)$. Alors M est une martingale UI et

$$M_n \rightarrow \xi := \mathbb{E}(\xi | \mathcal{F}_\infty), \quad \text{p.s. et dans } \mathbb{L}^1.$$

Démonstration. M est une martingale par la propriété des tours : si $r \geq n$,

$$\mathbb{E}(M_r | \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(\xi | \mathcal{F}_r) | \mathcal{F}_n) = \mathbb{E}(\xi | \mathcal{F}_n) = M_n.$$

On sait que M est UI (Théorème 17.1.3) donc $M_\infty := \lim M_n$ existe p.s. et dans \mathbb{L}^1 . Il reste à montrer que $M_\infty = \eta$ p.s. où $\eta := \mathbb{E}(\xi | \mathcal{F}_\infty)$. On suppose SPDG que $\xi \geq 0$ et on considère les mesures Q_1 et Q_2 telles que

$$Q_1(F) := \mathbb{E}(\eta; F), \quad Q_2(F) = \mathbb{E}(M_\infty; F), \quad F \in \mathcal{F}_\infty.$$

Si $F \in \mathcal{F}_n$, comme $\mathbb{E}(\eta | \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(\xi | \mathcal{F}_\infty) | \mathcal{F}_n) = \mathbb{E}(\xi | \mathcal{F}_n) (= M_n)$ par la propriété des tours, on a en utilisant (17.2.1),

$$\mathbb{E}(\eta; F) = \mathbb{E}(M_n; F) = \mathbb{E}(M_\infty; F).$$

Donc Q_1 et Q_2 coïncident sur le π -système $\cup_n \mathcal{F}_n$ et donc aussi sur $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$.

Mais η et M_∞ sont \mathcal{F}_∞ -mesurables (poser $M_\infty := \limsup M_n$). Donc $F := [\eta > M_\infty] \in \mathcal{F}_\infty$, et comme $Q_1(F) = Q_2(F)$,

$$\mathbb{E}(\eta - M_\infty; \eta > M_\infty) = 0.$$

Donc $\mathbb{P}(\eta > M_\infty) = 0$ et de même $\mathbb{P}(M_\infty > \eta) = 0$. □

Exemple 1. Prendre $M \in \mathbb{L}^1([0, 1]^2)$ et \mathcal{F}_n la tribu engendrée par la partition de $[0, 1]^2$ en carrés dyadiques d'arête 2^{-n} . Alors M_n est la fonction en escalier égale à la moyenne de M sur chaque cube dyadique. Dans ce cas, \mathcal{F}_∞ est la tribu des Boréliens, et $M_n \rightarrow M$. Remarque : si $f \in \mathbb{L}^2$, on savait déjà qu'on a convergence \mathbb{L}^2 par le théorème des projections dans un Hilbert : pourquoi ?

17.2.2 La loi du 0-1 par martingale

Théorème 17.2.3. Soit X_1, X_2, \dots une suite de v.a. indépendantes. On pose

$$\mathcal{T}_n := \sigma(X_{n+1}, X_{n+2}, \dots), \quad \mathcal{T} = \cap_n \mathcal{T}_n.$$

Alors si $F \in \mathcal{T}$, $\mathbb{P}(F) = 0$ ou 1.

Démonstration. Soit $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$. Soit $F \in \mathcal{T}$ et $\eta := \mathbb{1}_F$. Comme η est sommable, on a par le théorème "montant" de Lévy que

$$\eta = \mathbb{E}(\eta | \mathcal{F}_\infty) = \lim \mathbb{E}(\eta | \mathcal{F}_n), \quad \text{p.s.}$$

Mais pour tout n , η est \mathcal{T}_n -mesurable et donc indépendant de \mathcal{F}_n . Donc par la propriété (k) de l'espérance conditionnelle,

$$\mathbb{E}(\eta | \mathcal{F}_n) = \mathbb{E}(\eta) = \mathbb{P}(F), \quad \text{p.s.}$$

Donc $\eta = \mathbb{P}(F)$, p.s. et comme η ne prend que les valeurs 0 et 1, c'est fini. □

17.2.3 Théorème de Lévy "descendant"

Théorème 17.2.4. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ et $\{\mathcal{G}_{-n}; n \in \mathbb{N}\}$ une suite de sous-tribus de \mathcal{F} telles que

$$\mathcal{G}_{-\infty} := \bigcap_k \mathcal{G}_{-k} \subset \cdots \subset \mathcal{G}_{-(n+1)} \subset \cdots \subset \mathcal{G}_{-1}.$$

Soit $\gamma \in \mathbb{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ et posons

$$M_n := \mathbb{E}(\gamma | \mathcal{G}_n).$$

Alors

$$M_{-\infty} := \lim M_{-n} \quad \text{existe p.s. et dans } \mathbb{L}^1,$$

et

$$M_{-\infty} = \mathbb{E}(\gamma | \mathcal{G}_{-\infty}), \quad \text{p.s.}$$

Démonstration. En appliquant le lemme de passage de Doob à la martingale $(M_k, \mathcal{G}_k, -N \leq k \leq -1)$, on montre exactement comme pour le théorème de convergence de Doob que M_{-n} converge presque partout. Du théorème 17.1.3, on déduit que M_{-n} est équiintégrable, et donc que M_{-n} converge dans \mathbb{L}^1 . Reste à montrer que $M_{-\infty} = \mathbb{E}(\gamma | \mathcal{G}_{-\infty})$. On pose $M_{-\infty} = \limsup M_{-n}$ pour avoir que M est $\mathcal{G}_{-\infty}$ -mesurable. Soit $G \in \mathcal{G}_{-\infty} \subset \mathcal{G}_{-r}$:

$$\mathbb{E}(\gamma; G) = \mathbb{E}(M_{-r}; G) \rightarrow \mathbb{E}(M_{-\infty}; G)$$

car $M_{-r} \rightarrow M_{-\infty}$ dans \mathbb{L}^1 . □

17.2.4 Preuve de la LGN par martingale

Théorème 17.2.5. Soient X_1, X_2, \dots des v.a.i.i.d. avec pour tout k , $\mathbb{E}(X_k) < \infty$. Soit $\mu = \mathbb{E}(X_k)$ et $S_n := X_1 + \cdots + X_n$. Alors

$$\frac{S_n}{n} \rightarrow \mu, \quad \text{p.s. et dans } \mathbb{L}^1.$$

Démonstration. On pose

$$\mathcal{G}_{-n} = \sigma(S_n, S_{n+1}, \dots), \quad \mathcal{G}_{-\infty} = \bigcap_n \mathcal{G}_{-n}.$$

On sait que

$$\mathbb{E}(X_1 | \mathcal{G}_n) = \frac{1}{n} S_n, \quad \text{p.s.}$$

Par le lemme de Lévy descendant, $L := \lim \frac{1}{n} S_n$ existe p.s. et dans \mathbb{L}^1 . On pose $L := \limsup \frac{1}{n} S_n$ pour avoir une définition partout. Or, pour tout k ,

$$L = \limsup \frac{X_{k+1} + \cdots + X_{k+n}}{n},$$

ce qui veut dire que L est \mathcal{T}_k mesurable, où $\mathcal{T}_k = \sigma(X_{k+1}, \dots)$. Par la loi du 0-1, il existe $c \in \mathbb{R}$ tel que $\mathbb{P}(L = c) = 1$. Reste à calculer c :

$$c = \mathbb{E}(L) = \lim \mathbb{E}\left(\frac{1}{n} S_n\right) = \mu. \quad \square$$

17.3 TD probabilité : Deuxième théorème d'arrêt : [26] p. 391

Définition 17.3.1. Une martingale M par rapport à une filtration $(\mathcal{F}_n)_n$ est dite fermable s'il existe une variable aléatoire M_∞ qui est \mathcal{F}_∞ -mesurable et intégrable et telle que pour tout n :

$$M_n = \mathbb{E}(M_\infty | \mathcal{F}_n)$$

Bien entendu, une martingale fermable est uniformément intégrable et donc converge presque sûrement et dans \mathbb{L}^1 vers M_∞ . La réciproque est vraie : si M est uniformément intégrable alors elle est fermable.

On va commencer par donner un résultat relatif aux martingales arrêtées. Soit donc M une martingale et T un temps d'arrêt.

1) On définit la tribu des événements antérieurs à T par

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty | \forall n \in \mathbb{N}, A \cap \{T = n\} \in \mathcal{F}_n\}$$

Montrer que c'est bien une tribu.

2) Montrer que si S et T sont deux temps d'arrêt bornés par $k \in \mathbb{N}$ tels que $S \leq T$ alors $\mathcal{F}_S \subset \mathcal{F}_T$.

Solution

Si $A \in \mathcal{F}_S$, on a $\forall n \in \mathbb{N}, A \cap \{S = n\} \in \mathcal{F}_n$. Alors, comme $S \leq T$,

$$A \cap \{T = n\} = \left(\bigcup_{k \leq n} A \cap \{S = k\} \right) \cap \{T = n\} \in \mathcal{F}_n \cap \mathcal{F}_n = \mathcal{F}_n.$$

3) On conserve les notations des questions précédentes. Soit $A \in \mathcal{F}_S$. On définit R par $R = S\mathbb{1}_A + k\mathbb{1}_{A^c}$.

a) Montrer que R est un temps d'arrêt borné par k .

Solution

$\{R = n\} = \{S = n\} \cap A$ si $n \neq k$, et $\{R = k\} = \{S = k\} \cap A \cup (A^c \cap \cup_{l < k} \{S = l\})$ car S est borné par k .

b) Appliquer le premier théorème d'arrêt (celui vu en cours pour les temps d'arrêt bornés) et en déduire $\mathbb{E}(M_R) = \mathbb{E}(M_0) = \mathbb{E}(M_k)$. En déduire $\mathbb{E}(\mathbb{1}_A M_S) = \mathbb{E}(\mathbb{1}_A M_k)$.

Solution

En effet, $\mathbb{E}M_R = \mathbb{E}(\mathbb{1}_A M_S) + \mathbb{E}(\mathbb{1}_{A^c} M_k) = \mathbb{E}M_k = \mathbb{E}(M_k \mathbb{1}_A) + \mathbb{E}(M_k \mathbb{1}_{A^c})$.

c) En déduire $X_S = \mathbb{E}(M_k | \mathcal{F}_S)$. Remarquer que de même, $M_T = \mathbb{E}(M_k | \mathcal{F}_T)$.

d) Conclure que $M_S = \mathbb{E}(M_T | \mathcal{F}_S)$.

Solution

Par la propriété des tours,

$$\mathbb{E}(M_T | \mathcal{F}_S) = \mathbb{E}(\mathbb{E}(M_k | \mathcal{F}_T) | \mathcal{F}_S) = \mathbb{E}(M_k | \mathcal{F}_S) = M_S.$$

On vient donc de démontrer le résultat suivant :

Théorème 17.3.1. *Soit M une martingale et $S \leq T$ deux temps d'arrêt bornés. On a alors*

$$M_S = \mathbb{E}(M_T | \mathcal{F}_S)$$

On se propose maintenant de démontrer le théorème suivant.

Théorème 17.3.2. *Soit M une martingale uniformément intégrable (donc fermable). On note M_∞ sa limite presque sûre et dans \mathbb{L}^1 . Soit T un temps d'arrêt (fini ou pas). On a le théorème d'arrêt de Doob :*

$$M_T \text{ est dans } \mathbb{L}^1 \text{ et } \mathbb{E}(M_T) = \mathbb{E}(M_0)$$

Remarque 17.3.1. *Il y a en fait équivalence, à savoir que si M est une martingale, elle est fermable ssi elle vérifie la propriété précédente pour tous les temps d'arrêt T .*

1) Le but de cette question est de montrer que la famille de variables aléatoires $\{M_{T \wedge n} | n \in \mathbb{N}\}$ est uniformément intégrable.

a) On rappelle que si X est une v.a. intégrable alors la famille $\{\mathbb{E}(X | \mathcal{G}) | \mathcal{G} \text{ sous-tribu de } \mathcal{F}\}$ est uniformément intégrable.

b) En se servant du théorème précédent, montrer que si T est un temps d'arrêt borné alors $M_T = \mathbb{E}(M_\infty | \mathcal{F}_T)$. (*ind.* : On remarquera pour cela que M étant fermable, on a pour tout k , $M_k = \mathbb{E}(M_\infty | \mathcal{F}_k)$ et d'autre part, par le théorème précédent, $M_T = \mathbb{E}(M_k | \mathcal{F}_T)$ pour k assez grand. On conclura ensuite en observant que $\mathcal{F}_T \subset \mathcal{F}_k$ et que donc par la propriété des tours, $M_T = \mathbb{E}(\mathbb{E}(M_\infty | \mathcal{F}_k) | \mathcal{F}_T) = \mathbb{E}(M_\infty | \mathcal{F}_T)$).

c) Montrer que $\{T \wedge n | n \in \mathbb{N}\}$ est une suite de temps d'arrêt bornés. En déduire que la famille de variables aléatoires $\{M_{T \wedge n} | n \in \mathbb{N}\}$ est uniformément intégrable. (Application directe du b)).

2) Montrer que la martingale arrêtée M^T converge presque sûrement et dans \mathbb{L}^1 vers M_T (*ind.* : on regardera ce qui se passe sur les événements $\{T < +\infty\}$ et $\{T = \infty\}$).

3) En déduire le théorème.

Solution

Par définition, $M_n^T = M_{T \wedge n}$. C'est une martingale UI par le 1) et donc elle converge p.s. et dans L^1 . Sur $\{T < \infty\}$, $M_{T \wedge n} \rightarrow M_T$ quand $n \rightarrow \infty$. Sur $\{T = \infty\}$, $M_{T \wedge n} = M_n \rightarrow M_\infty = M_T$ p.s.

17.4 TD Probabilités : Optimisation d'une vente ([18] p. 511)

Un étudiant ayant brillamment réussi l'agrégation désire se débarrasser de son Williams en en tirant le maximum de profit. Pour cela, il ouvre le Grimmitt et cherche si il n'y aurait pas moyen d'optimiser cette vente. Toutes les semaines, il passe donc une annonce dans "La Sauce" où il propose de le vendre. Il sait que chaque semaine, il lui sera fait une offre dont le prix est modélisé par une variable aléatoire X_n (n pour indiquer que c'est la n -ème semaine). On supposera les $(X_n)_n$ i.i.d. ayant une fonction de densité f et une moyenne finie $p > 0$. D'autre part, le fait de stocker ce livre sacré l'oblige à louer une pièce dans laquelle il l'entrepose et qui lui coûte chaque semaine un prix constant c . On supposera de plus que notre petit agrégé, en bon petit épargnant qu'il est,

place son argent dès qu'il en a sur un compte avec un taux d'intérêt par semaine constant égal à α .

1) Expliquer pourquoi il est raisonnable de vendre le livre la semaine T où T est un temps d'arrêt (par rapport à la filtration naturelle associée aux X_i) qui maximise :

$$\mu(T) = \mathbb{E} \left((1 + \alpha)^{-T} X_T - \sum_{n=1}^T (1 + \alpha)^{-n} c \right)$$

Solution

Si l'étudiant avait vendu son livre dès l'agrégation en poche, il aurait à la date n empoché, grâce à l'épargne, une somme égale à $X_0(1 + \alpha)^n$. Supposons maintenant qu'il ait attendu pour le vendre la date n : la première semaine, il aura non seulement perdu le prix de la location de la pièce mais aussi l'intérêt qu'il aurait pu gagner en épargnant cet argent soit $c(1 + \alpha)$; au bout de la deuxième semaine, il aura de la même manière perdu $c(1 + \alpha)$ mais aussi l'argent de la première semaine qu'il aurait placé depuis le début et qui lui aurait donc rapporté $c(1 + \alpha)^2$; ... et finalement à la semaine n , (vu qu'il ne louera pas la pièce cette semaine), tout se passe comme s'il avait perdu $\sum_{k=0}^{n-1} c(1 + \alpha)^k$. D'un autre côté, il aura gagné X_n . Pour estimer quantitativement l'intérêt de vendre au temps n plutôt qu'au temps 0, on calcule le bénéfice fait au temps n si la vente s'est faite au temps n . Mais ce bénéfice n'est exercé qu'au temps n . Pour comparer les différentes opérations possibles, les bénéfices sont actualisés au temps 0, à savoir que gagner la quantité C au temps n représente une valeur actualisée au temps 0 de $(1 + \alpha)^{-n}C$. C'est en effet ce dernier capital qui rendrait C au temps n pour un taux constant de α . La valeur actualisée à maximiser est donc

$$\frac{X_n + \sum_{k=0}^{n-1} c(1 + \alpha)^k}{(1 + \alpha)^n} = (1 + \alpha)^{-n} X_n - \sum_{k=1}^n (1 + \alpha)^{-k} c.$$

2) On pose $Z_n = X_n + c/\alpha$. Montrer qu'il existe un unique γ ayant la propriété

$$\alpha\gamma = \int_{\gamma}^{\infty} \mathbb{P}(Z_n > y) dy$$

Solution

La fonction $h : \gamma \rightarrow \alpha\gamma - \int_{\gamma}^{\infty} \mathbb{P}(Z_n > y) dy$ est continue et strictement croissante avec $h(0) = -\mathbb{E}(Z_n)$ (calcul classique consistant à écrire $\mathbb{P}(Z_n > y) = \mathbb{E}(1_{Z_n > y})$ et à utiliser Fubini). Donc $h(0) < 0$ puisque les X_i sont positives. D'autre part, h tend vers $+\infty$ en $+\infty$ (puisque toujours en utilisant Fubini, $\int_{\gamma}^{\infty} \mathbb{P}(Z_n > y) dy = \mathbb{E}((Z_n - \gamma)^+)$). Le théorème des valeurs intermédiaires permet de conclure.

Remarquer que tout revient maintenant avec ces notations à maximiser sur les temps d'arrêt finis (on suppose que l'on veut absolument vendre le Williams) la quantité

$$\mathbb{E} \{ (1 + \alpha)^{-T} Z_T \}$$

Cette quantité a toujours un sens dès que T est fini (mais elle peut-être infinie a priori).

Solution

On a

$$(1 + \alpha)^{-T} X_T - \sum_{k=1}^T (1 + \alpha)^{-k} c = (1 + \alpha)^{-T} (X_T + \frac{c}{\alpha}) - \frac{c}{\alpha}.$$

Donc

$$\mu(T) = E((1 + \alpha)^{-T} (X_T + \frac{c}{\alpha})) - \frac{c}{\alpha} = E(Z_T (1 + \alpha)^{-T}) - \frac{c}{\alpha}.$$

3) Montrer que $(V_n)_n$ définie par $V_n = (1 + \alpha)^{-n} \sup(Z_n, \gamma)$ est une surmartingale.

Solution

Remarquons que les $(V_n)_n$ sont indépendantes et qu'il suffit donc de prouver que les V_n vérifient $\mathbb{E}(V_{n+1}) \leq V_n$. Or $\mathbb{E}(\sup(Z_n, \gamma)) = \gamma + \int_{\gamma}^{\infty} \mathbb{P}(Z_n > y) dy$ (toujours Fubini), ce qui est égal par choix de γ à $(1 + \alpha)\gamma$. Donc $E(V_n) = (1 + \alpha)^{-n+1} \gamma$ qui est bien décroissante avec n . Détail du calcul :

$$\begin{aligned} \gamma + \int_{\gamma}^{+\infty} \mathbb{P}(Z_n > y) dy &= \gamma + \int_{\gamma}^{+\infty} \mathbb{E} \mathbb{1}_{Z_n > y} dy = \\ \gamma + \mathbb{E} \int_{\gamma}^{+\infty} \mathbb{1}_{Z_n > y} dy &= \gamma + \mathbb{E}(Z_n - \gamma)^+ = \mathbb{E} \sup(Z_n, \gamma). \end{aligned}$$

4) Remarquer que $\frac{c}{\alpha}$ est le coût du loyer jusqu'à la fin des temps. Une stratégie sera dite bonne si elle consiste à prendre un prix déterministe fixé τ dans le support de Z_1 (c'est-à-dire $\tau - c/\alpha$ dans le support de X_1) et prendre pour T_τ le temps d'arrêt qui consiste à vendre dès que le prix offert dépasse $\tau - c/\alpha$ (on suppose $\tau \geq c/\alpha$), c'est-à-dire dès que Z_n dépasse τ . Trouver le τ pour lequel parmi toutes les bonnes stratégies, $\mu(T_\tau)$ est maximal.

Solution

Soit G la fonction de répartition des Z_n . Par définition de T_τ et puisque les X_i sont i.i.d. la loi de T_τ est donnée par :

$$\mathbb{P}(T = n) = G(\tau)^n (1 - G(\tau))$$

(ne pas oublier que l'on commence à la semaine 0 et $\{T = n\} = \{Z_0 \leq \tau, \dots, Z_{n-1} \leq \tau, Z_n > \tau\}$)

Remarquer que T n'est jamais infini car τ a été pris dans le support de Z_1 si bien que $G(\tau) \neq 1$.

$$\begin{aligned}
\mu(T_\tau) + c/\alpha &= \sum_{n=0}^{\infty} \mathbb{E} \left((1 + \alpha)^{-n} Z_n \mathbb{1}_{T=n} \right) \\
&= \sum_{n=0}^{\infty} (1 + \alpha)^{-n} \mathbb{E} \left(Z_n \mathbb{1}_{Z_n > \tau} \mathbb{1}_{\{Z_0 \leq \tau, \dots, Z_{n-1} \leq \tau\}} \right) \\
&= \sum_{n=0}^{\infty} (1 + \alpha)^{-n} \mathbb{E} \left(Z_n \mathbb{1}_{\{Z_n > \tau\}} \right) \mathbb{P}(Z_0 \leq \tau) \dots \mathbb{P}(Z_{n-1} \leq \tau) \\
&= \sum_{n=0}^{\infty} (1 + \alpha)^{-n} G(\tau)^n \mathbb{E} \left(Z_1 \mathbb{1}_{\{Z_1 > \tau\}} \right) \\
&= \frac{1 + \alpha}{1 + \alpha - G(\tau)} \left\{ \tau(1 - G(\tau)) + \int_{\tau}^{\infty} (1 - G(y)) dy \right\} (= \varphi(\tau)).
\end{aligned}$$

La dernière égalité résulte une fois de plus de l'utilisation de Fubini en écrivant :

$$\mathbb{E}(Z_1 \mathbb{1}_{\{Z_1 > \tau\}}) = \mathbb{E} \left(\int_0^{\infty} (1_{y < Z_1} 1_{\tau < Z_1}) dy \right)$$

En effet

$$\begin{aligned}
\mathbb{E}(Z_1 \mathbb{1}_{Z_1 > \tau}) &= \mathbb{E}(\sup(Z_1, \tau)) - \tau \mathbb{P}(Z_1 \leq \tau) = \\
&\tau + \int_{\tau}^{+\infty} \mathbb{P}(z_1 > y) dy - \tau \mathbb{P}(z_1 \leq \tau) = \tau(1 - G(\tau)) + \int_{\tau}^{+\infty} (1 - G(y)) dy.
\end{aligned}$$

La fonction $\varphi(\tau)$ malgré son aspect repoussant s'étudie bien. On pose $G' = f$. On a

$$\begin{aligned}
\varphi'(\tau) &= \frac{1 + \alpha}{(1 + \alpha - G(\tau))^2} f(\tau) (\tau(1 - G(\tau)) + \int_{\tau}^{+\infty} (1 - G(y)) dy) + \frac{1 + \alpha}{1 + \alpha - G(\tau)} (1 - G(\tau) - \tau f(\tau) - 1 + G(\tau)) = \\
&\frac{1 + \alpha}{(1 + \alpha - G(\tau))^2} f(\tau) (\tau(1 - G(\tau)) + \int_{\tau}^{+\infty} (1 - G(y)) dy - \tau(1 + \alpha - G(\tau))) = \\
&= -\frac{1 + \alpha}{(1 + \alpha - G(\tau))^2} f(\tau) h(\tau).
\end{aligned}$$

Le signe de la dérivée est le même que celui de $-f(\tau)h(\tau)$, où h est définie dans le 2), et donc la fonction est maximale pour $\tau = \gamma$. On a alors

$$\mu(T_\gamma) = \gamma(1 + \alpha) - c/\alpha.$$

En effet,

$$\begin{aligned}
\frac{c}{\alpha} + \mu(T_\gamma) &= \frac{1 + \alpha}{1 + \alpha - G(\gamma)} (\gamma(1 - G(\gamma)) + \int_{\gamma}^{+\infty} \mathbb{P}(Z_1 > y) dy) = \\
&\frac{1 + \alpha}{1 + \alpha - G(\gamma)} (\gamma(1 - G(\gamma)) + \alpha\gamma) = \gamma(1 + \alpha).
\end{aligned}$$

Donc $\mu(T_\gamma) = \gamma(1 + \alpha) - \frac{c}{\alpha}$.

5) Considérer maintenant une stratégie quelconque telle que $T < +\infty$. Montrer en utilisant le théorème de convergence des surmartingales UI que cette stratégie est moins bonne que la meilleure des bonnes stratégies et conclure.

Solution

Considérons une stratégie générale avec un temps d'arrêt T fini. On sait que $(V_{T \wedge m})_m$, définie à la question 3), est une surmartingale et que $\mathbb{E}(V_{T \wedge m}) \leq \mathbb{E}(V_0) = \gamma(1 + \alpha)$ car V est une surmartingale arrêtée. Mais cette surmartingale arrêtée est en fait UI car :

$$|V_{T \wedge m}| \leq \sum_{i=0}^{\infty} |V_i|, \quad \sum_i \mathbb{E}(|V_i|) < +\infty$$

Donc $\lim_{m \rightarrow \infty} V_{T \wedge m} = V_T$ dans \mathcal{L}^1 d'où :

$$\mathbb{E}(V_T) \leq \mathbb{E}(V_0)$$

Or $\mu(T) \leq \mathbb{E}(V_T) - c/\alpha \leq \mu(T_\gamma)$.

La stratégie de la question précédente est donc la meilleure.

17.5 TD Probabilités : Théorème de Radon-Nykodim

Le but de ce TD est de donner une démonstration via les martingales du théorème de Radon-Nykodim (une autre référence pour ce théorème qui peut être démontré sans les martingales est bien-sûr [27]).

Théorème 17.5.1. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et \mathbb{Q} une mesure finie sur (Ω, \mathcal{F}) absolument continue par rapport à \mathbb{P} , c'est-à-dire :*

$$F \in \mathcal{F}, \quad \mathbb{P}(F) = 0 \implies \mathbb{Q}(F) = 0$$

Il existe alors une v.a. dans $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ telle que $\mathbb{Q} = X\mathbb{P}$, c'est-à-dire :

$$\forall F \in \mathcal{F}, \quad \mathbb{Q}(F) = \int_F X d\mathbb{P}$$

X est appelée une version de la dérivée de Radon-Nykodym de \mathbb{Q} par rapport à \mathbb{P} . Deux versions de X sont égales \mathbb{P} -p.s. et X est notée :

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = X$$

Nous supposons la tribu \mathcal{F} séparable, c'est-à-dire engendrée par un nombre dénombrable d'événements $(F_n)_{n \geq 0}$. [30] montre comment, une fois ceci connu, on passe au cas général (grosso modo, on approxime la tribu par des tribus séparables) : s'y reporter pour plus de détails. Dans la suite, \mathbb{E} désignera l'espérance relativement à \mathbb{P} .

1) Montrer en raisonnant par l'absurde et en utilisant BC1 que pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que pour tout événement $F \in \mathcal{F}$,

$$\mathbb{P}(F) < \delta \implies \mathbb{Q}(F) < \epsilon$$

Solution

Si tel n'était pas le cas, on pourrait trouver une suite $(G_n)_n$ d'événements tels que $\mathbb{P}(G_n) \leq 2^{-n}$ et cependant $\mathbb{Q}(G_n) \geq \epsilon$. On considère $\bar{\lim} G_n$ qui par BC1 est de \mathbb{P} -probabilité 0 mais par Fatou de \mathbb{Q} -probabilité supérieure à ϵ d'où la contradiction.

2) On définit maintenant la filtration $\mathcal{F}_n = \sigma\{F_1, \dots, F_n\}$. Un atome A de \mathcal{F}_n est un élément de \mathcal{F}_n qui s'écrit $H_1 \cap H_2 \cap \dots \cap H_n$ ou H_i est soit F_i , soit F_i^c . Montrer que l'ensemble des $r(n)$ atomes de \mathcal{F}_n , notés $A_{n,1}, \dots, A_{n,r(n)}$ forment une partition de Ω et que \mathcal{F}_n est constituée des $2^{r(n)}$ réunions (disjointes) possibles d'atomes.

Solution

Il est bien évident que deux atomes distincts sont disjoints (puisqu'il existe un indice i pour lequel l'un intersecte F_i et l'autre F_i^c). D'autre part, tout élément ω de Ω est contenu dans un atome (composer l'atome de la manière suivante : pour chaque i , prendre F_i si $\omega \in F_i$ et prendre F_i^c sinon).

3) On définit alors une fonction X_n de Ω dans $[0, \infty)$ de la manière suivante :

Si $\omega \in A_{n,k}$ alors : $X_n(\omega) = \mathbb{Q}(A_{n,k})/\mathbb{P}(A_{n,k})$ si $\mathbb{P}(A_{n,k}) > 0$ et $X_n(\omega) = 0$ sinon.

a) Montrer que $X_n \in \mathcal{L}^1(\Omega, \mathcal{F}_n, \mathbb{P})$ et que l'on a pour tout événement F de \mathcal{F}_n

$$\mathbb{E}(X_n 1_F) = \mathbb{Q}(F) \tag{17.5.1}$$

Solution

Il suffit d'écrire F comme la réunion disjointe d'événements atomiques et d'utiliser la linéarité de l'espérance. On montre que X_n est intégrable en appliquant ceci à l'ensemble $F = \Omega$.

b) Remarquer que X_n est la dérivée de Radon-Nikodym de \mathbb{Q} restreinte à \mathcal{F}_n relativement à \mathbb{P} restreinte à \mathcal{F}_n .

4) Montrer que X_n est une martingale positive relativement à la filtration \mathcal{F}_n et prouver l'existence d'une v.a.r. X_∞ telle que \mathbb{P} p.s. :

$$\lim_{n \rightarrow +\infty} X_n = X_\infty$$

Solution

Seule la preuve que (X_n) est une martingale est à donner, le reste résultant directement du théorème de convergence des martingales bornées dans \mathcal{L}^1 (car elle est positive). D'après (17.5.1), on a pour tout $F \in \mathcal{F}_{n-1}$, puisque F appartient à \mathcal{F}_n :

$$\mathbb{E}(X_n 1_F) = \mathbb{Q}(F) = \mathbb{E}(X_{n-1} 1_F)$$

5) Prouver que X est une martingale uniformément intégrable.

Solution

On choisit K tel que $K^{-1}\mathbb{Q}(\Omega) < \delta$ (car \mathbb{Q} est de masse finie). Une inégalité de Markov, suivie de l'utilisation de (17.5.1) permet alors d'écrire :

$$\mathbb{P}(X_n > K) \leq K^{-1}\mathbb{E}(X_n) = K^{-1}\mathbb{Q}(\Omega) < \delta$$

En utilisant à nouveau (17.5.1) avec $F = \{X_n > K\}$ combiné avec le résultat de la question 1), on obtient alors :

$$\mathbb{E}(X_n 1_{\{X_n > K\}}) < \epsilon$$

et la martingale X est UI.

6) Montrer que la convergence de X_n vers X_∞ se fait dans \mathcal{L}^1 .

7) Conclure dans le cas où \mathcal{F} est séparable.

Solution

Les deux mesures \mathbb{Q} et $F \rightarrow \mathbb{E}(X_\infty 1_F)$ coïncident sur le π -système $\cup \mathcal{F}_n$ qui engendre \mathcal{F} et sont de même masse : elles coïncident donc sur \mathcal{F} .

8) Prouver l'unicité \mathbb{P} p.s. de la dérivée de Radon-Nikodym.

Solution

C'est la machine standard .

Chapitre 18

Théorie financière

Cette section est un résumé des chapitres 1 et 2 de [24].

18.1 Le problème des options

L'exposé est centré sur le problème des options. Une option est un titre donnant à son détenteur le droit, et non l'obligation, d'acheter ou de vendre (selon qu'il s'agit d'une option d'achat ou de vente) une certaine quantité d'un actif financier, à une date convenue et à un prix fixé d'avance. La description précise d'une option se fait à partir des éléments suivants :

- la nature de l'option : on parle de *call* pour une option d'achat et de *put* pour une option de vente.
- l'actif sous-jacent, sur lequel porte l'option : dans la pratique, il peut s'agir d'une action, d'une obligation, d'une devise etc.
- le montant, c'est-à-dire la quantité d'actif sous-jacent à acheter ou à vendre.
- l'échéance ou date d'expiration, qui limite la durée de vie de l'option ; si l'option peut être exercée à n'importe quel instant précédent l'échéance, on parle d'option *américaine*, si l'option ne peut être exercée qu'à l'échéance, on parle d'option *européenne*.
- le prix d'exercice, qui est le prix (fixé d'avance) auquel se fait la transaction en cas d'exercice de l'option.

L'option elle-même a un prix appelé la prime. Lorsque l'option est cotée sur un marché organisé, la prime est donnée par le marché. En l'absence de cotation, le calcul de la prime se pose.

Examinons, pour fixer les idées, le cas d'un call européen, d'échéance T , sur une action, dont le cours à la date t est donné par S_t . Soit K le prix d'exercice. On voit qu'à l'échéance, la valeur du call est donnée par la quantité :

$$(S_T - K)_+ = \sup(S_T - K, 0)$$

Pour le vendeur de l'option, il s'agit, en cas d'exercice, d'être en mesure de fournir une action au prix K et, par conséquent de pouvoir produire à l'échéance une richesse égale à $(S_T - K)_+$. Au moment de la vente de l'option, que l'on prendra pour origine des temps, le cours S_T est inconnu et deux questions se posent :

- 1) Combien faut-il faire payer à l'acheteur de l'option, autrement dit comment évaluer à l'instant

$t = 0$ une richesse $(S_T - K)_+$ disponible à la date T ? C'est le problème du *pricing*.

2) Comment le vendeur, qui touche la prime à l'instant 0, parviendra-t-il à produire la richesse $(S_T - K)_+$ à la date T ? C'est le problème de la *couverture*.

18.2 Le formalisme des modèles discrets

L'hypothèse de base, retenue dans tous les modèles, est que, dans un marché suffisamment fluide, il est impossible de faire des profits sans prendre de risques. Nous traduirons dans la suite cette hypothèse en termes mathématiques. Dans la suite, $(\Omega, \mathcal{F}, \mathbb{P})$ désignera un espace probabilisé fini que l'on munit d'une filtration (finie) $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_N = \mathcal{F} = \mathcal{P}(\Omega)$. \mathcal{F}_n représente l'information dont on dispose au temps n . La date N est la date d'échéance des options. On fera l'hypothèse que

$$\forall \omega \in \Omega, \mathbb{P}(\{\omega\}) > 0$$

On suppose d'autre part qu'il y a sur le marché $d + 1$ actifs financiers dont les prix à l'instant n sont donnés par des variables aléatoires $S_n^0, S_n^1, \dots, S_n^d$ qui sont \mathcal{F}_n -mesurables. Le vecteur $S_n = (S_n^0, S_n^1, \dots, S_n^d)$ est le vecteur des prix à l'instant n . L'actif numéroté 0 représente les placements sans risque avec un taux d'intérêt r et on posera $S_0^0 = 1$. Ainsi $S_n^0 = (1 + r)^n$ et $\beta_n = 1/S_n^0$ est appelé le *coefficient d'actualisation* : c'est la somme d'argent qui investie à l'instant 0 dans l'actif sans risque permet de disposer de 1 euro à l'instant n . Les autres actifs sont appelés les actifs à risque. Les actifs actualisés sont définis par

$$\tilde{S}_n = \beta_n S_n$$

18.3 Les stratégies

Une stratégie de gestion est définie par un processus aléatoire $\phi = ((\phi_n^0, \dots, \phi_n^d))_{0 \leq n \leq N}$ à valeurs dans \mathbb{R}^{d+1} . A chaque instant n , elle donne les quantités $\phi_n^0, \dots, \phi_n^d$ des différents actifs détenues en portefeuille. On supposera que le processus ϕ est prévisible (le portefeuille ϕ_n à la date n est constitué au vu des informations disponibles à la date $(n - 1)$ et conservé tel quel au moment des cotations à la date n).

La valeur du portefeuille à l'instant n est donnée par le produit scalaire :

$$V_n(\phi) = \phi_n \cdot S_n = \sum_{i=0}^{d+1} \phi_n^i S_n^i$$

Une stratégie sera dite *auto-financée* si il n'y a ni apports, ni retraits de fonds. La variation de la valeur du portefeuille est donc uniquement due aux gains procurés par l'agitation des cours. Cette condition d'autofinancement se traduit par

$$\phi_n \cdot S_n = \phi_{n+1} \cdot S_n$$

18.4 Marché viable

Pour l'instant, rien n'interdit aux quantités d'actifs ϕ_n^i d'être négatives, c'est-à-dire que l'on a des dettes libellées en actifs à risques. Les emprunts et ventes à découverts sont autorisés dans

notre modèle mais on impose néanmoins qu'à tout instant, la valeur du portefeuille soit positive.

Une stratégie sera dite *admissible* si elle est autofinancée et si à tout instant, la valeur du portefeuille correspondant est positif ou nul. Une *stratégie d'arbitrage* est une stratégie admissible de valeur initiale nulle et de valeur finale non nulle.

D'après l'hypothèse de base qui consiste à dire que l'on ne peut pas gagner sans prendre de risque, on doit refuser aux stratégies d'arbitrage le droit à l'existence (ce qui est une hypothèse de droite). Un tel marché est alors appelé un *marché viable*. Cette hypothèse maintenant mathématiquement formulée conduit assez simplement au théorème suivant :

Théorème 18.4.1. *Le marché est viable ssi il existe une probabilité \mathbb{P}^* équivalente à \mathbb{P} sous laquelle les prix actualisés des actifs \tilde{S} sont des martingales.*

Démonstration. La démonstration de ce théorème est élémentaire et géométrique mais demande un peu de travail. \square

18.5 Marché complet et évaluation des options

On définit une option européenne d'échéance N comme une variable aléatoire $h \geq 0$ \mathcal{F}_N -mesurable et qui représente le profit que permet l'exercice de l'option. Ainsi pour un call sur une unité d'actif 1, au prix d'exercice K , on a

$$h = (S_N^1 - K)_+$$

et pour une option de vente ou put sur une unité d'actif 1 au prix d'exercice K , on a

$$h = (K - S_N^1)_+$$

On dira que l'actif conditionnel défini par h est *simulable* (ou atteignable) s'il existe une stratégie admissible dont la valeur à l'instant N est égale à h . On dira que le marché est *complet* si tout actif conditionnel est simulable. Imposer qu'un marché est complet est une hypothèse restrictive dont la justification économique n'est pas claire. Le modèle de Cox-Ross-Rubinstein, que nous étudierons en détail, est un exemple de marché viable et complet.

Théorème 18.5.1. *Un marché viable est complet si, et seulement si, il existe une seule probabilité \mathbb{P}^* équivalente à \mathbb{P} sous laquelle les prix actualisés des actifs soient des martingales.*

On suppose le marché viable et complet et on note \mathbb{P}^* l'unique probabilité sous laquelle les prix actualisés des actifs sont des martingales. Soit un actif conditionnel défini par une variable aléatoire h positive et \mathcal{F}_N -mesurable et soit ϕ une stratégie admissible simulant h , c'est-à-dire vérifiant

$$\forall n \in \{0, \dots, N\}, V_n(\phi) \geq 0, \phi_n \cdot S_n = \phi_{n+1} \cdot S_n, V_N(\phi) = h$$

La suite des valeurs de portefeuilles actualisées $(\tilde{V}_n)_{0 \leq n \leq N}$ définie par

$$\tilde{V}_n(\phi) = \beta_n V_n(\phi)$$

est une martingale sous \mathbb{P}^* (en tant que transformée de martingale). On a donc

$$V_n(\phi) = (1+r)^{n-N} \mathbb{E}^*(h|\mathcal{F}_n), \quad n = 0, \dots, N$$

Il est alors naturel d'appeler $V_n(\phi)$ la valeur de l'option au temps n : c'est la richesse qui, détenue à l'instant n , permet, en suivant la stratégie ϕ à partir de l'instant n , de produire exactement la richesse h à l'instant N .

Si à l'instant 0, un investisseur vend l'option au prix

$$\mathbb{E}^*((1+r)^{-N}h)$$

il a la possibilité, en suivant une stratégie simulante ϕ , de restituer la richesse promise h à l'instant N , c'est-à-dire qu'il peut se couvrir parfaitement.

18.6 Le modèle de Cox, Ross et Rubinstein

C'est une version discrétisée du modèle de Black-Scholes (modèle continu) dans lequel il y a un seul actif à risque, de prix S_n à l'instant n et un actif sans risque de rendement certain r sur une période si bien que $S_n^0 = (1+r)^n$. On fait les hypothèses suivantes sur l'évolution du cours de l'actif risqué : entre deux périodes consécutives, la variation relative des cours est soit a , soit b , avec $-1 < a < b$:

$$S_{n+1} = \begin{cases} S_n(1+a) \\ S_n(1+b) \end{cases}$$

Le cours initial S_0 est donné. On a ici $\Omega = \{1+a, 1+b\}^N$, chaque N -uplet représentant les valeurs successives de $T_n = S_n/S_{n-1}$. $\mathcal{F}_n = \sigma(S_1, \dots, S_n)$. On suppose que la probabilité considérée dans ce modèle \mathbb{P} assure une probabilité strictement positive à tous les singletons.

1) Sous l'hypothèse de viabilité, on montre que le modèle dépend en fait d'un paramètre $r \in (a, b)$ et que le marché est complet. La valeur d'un call d'échéance N et de prix d'exercice K à l'instant n peut être calculé et est donné par $C_n = c(n, S_n)$ où

$$c(n, x) = (1+r)^{(N-n)} \sum_{j=0}^{N-n} C_{N-n}^j p^j (1-p)^{N-n-j} (x(1+a)^j (1+b)^{(N-n-j)} - K)_+$$

Ici $p = (b-r)/(b-a)$. On peut aussi établir la relation dite de parité entre un call C_n et un put P_n de prix d'exercice K et d'échéance N

$$C_n - P_n = S_n - K(1+r)^{-(N-n)}$$

2) On peut aussi calculer la stratégie de couverture parfaite d'un call. Elle est définie par une quantité d'actif risqué $H_n = \Delta(n, S_{n-1})$ à détenir à l'instant n , où Δ est la fonction définie par

$$\Delta(n, x) = \frac{c(n, x(1+b)) - c(n, x(1+a))}{x(b-a)}$$

3) On peut “rendre continu” ce modèle. Pour cela, on fait tendre N vers l’infini en imposant les relations suivantes

$$\begin{cases} r_N = RT/N \\ \log((1 + a_N)/(1 + r_N)) = -\sigma/\sqrt{N} \\ \log((1 + b_N)/(1 + r_N)) = \sigma/\sqrt{N} \end{cases}$$

R s’interprète comme le taux d’intérêt instantané entre les instants 0 et T et σ^2 comme la variance limite de $\log(S_N)$ quand N tend vers l’infini, S_N représentant le cours de l’action à la date T . On obtient alors une valeur limite pour la valeur du put à l’instant 0 donnée par

$$P_0 = Ke^{-RT}F(-d_2) - S_0F(-d_1)$$

où F est la fonction de répartition d’une gaussienne standard et d_1, d_2 sont deux fonctions connues dépendant de K, R, T et σ . On obtient une formule similaire pour le call ainsi qu’une version continue de la relation de parité call-put. On remarquera que le seul paramètre non directement observable sur le marché est σ . Son évaluation se fait souvent par des méthodes statistiques.

18.7 Enveloppes de Snell et options américaines

Le but de cette section est de traiter l’évolution et la couverture des options américaines. Un lien avec le problème d’arrêt optimal sera mis en évidence. On se place dans le cadre d’un marché complet et viable et on note \mathbb{P}^* l’unique probabilité sous laquelle les cours actualisés \tilde{S}_n sont des martingales. A la différence des options européennes, une option américaine peut être exercée à tout moment entre les instants 0 et N . Nous la définissons comme une suite Z_n positive et adaptée à la filtration $(\mathcal{F}_n)_n$. Z_n représente le profit que permet l’exercice de l’option à l’instant n . Par exemple, pour un call américain au prix d’exercice K sur une unité d’actif dont le cours est donné par $(S_n)_n$, on a $Z_n = (S_n - K)_+$. Quelle est la valeur de l’option américaine associée ? Raisonnons par récurrence en marche arrière. Au temps N , $U_N = Z_N$. Pour calculer U_{N-1} , mettons nous à la place de celui qui vend l’option. Si le client exerce l’option au temps $N - 1$, le vendeur doit être capable de fournir une richesse Z_{N-1} . Si par contre, le client exerce au temps N , le vendeur doit être capable de fournir une richesse Z_N . Le vendeur doit donc encaisser au temps $N - 1$ une richesse au moins égale à Z_{N-1} et à partir de cette richesse être capable de fournir l’instant d’après une richesse égale à Z_N . La somme qui, disponible à l’instant $N - 1$, permet d’obtenir à l’instant N une richesse Z_N , c’est la valeur à l’instant $N - 1$ d’une stratégie admissible de valeur finale Z_{N-1} , c’est-à-dire $S_{N-1}^0 \mathbb{E}^*(Z_N/S_N^0 | \mathcal{F}_{N-1})$. La valeur de l’option à l’instant $N - 1$ est donc donnée par :

$$U_{N-1} = \max \left(Z_{N-1}, S_{N-1}^0 \mathbb{E}^*(\tilde{Z}_N | \mathcal{F}_{N-1}) \right)$$

où $\tilde{Z}_N = Z_N/S_N^0$ est la valeur actualisée de Z_N . On rappelle que $S_n^0 = (1 + r)^n$. En raisonnant de même aux temps antérieurs, on en déduit que la valeur de l’option au temps n est donnée par :

$$U_{n-1} = \max \left(Z_{n-1}, S_{n-1}^0 \mathbb{E}^*(U_n/S_n^0 | \mathcal{F}_n) \right)$$

On est alors naturellement amené à définir la notion d’enveloppe de Snell. On passe dans un cadre un peu plus abstrait et on suppose désormais que $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{0 \leq n \leq N}, \mathbb{P})$ est un espace filtré sur un ensemble Ω fini avec \mathcal{F}_0 la tribu triviale et $\mathcal{F}_N = \mathcal{F}$.

Définition 18.7.1. Soit $(Z_n)_{0 \leq n \leq N}$ une suite adaptée à la filtration $(\mathcal{F})_n$. On appelle enveloppe de Snell de la suite $(Z_n)_n$ le processus adapté $(U_n)_{n \leq n \leq N}$ défini par :

$$\begin{cases} U_N = Z_N \\ U_n = \max(Z_n, \mathbb{E}(U_{n+1} | \mathcal{F}_n)) \end{cases} \quad (18.7.1)$$

Proposition 18.7.1. L'enveloppe de Snell $(U_n)_n$ d'une suite $(Z_n)_n$ est la plus petite surmartingale majorant $(U_n)_n$.

Démonstration. Par définition, $(U_n)_n$ est une surmartingale majorant $(Z_n)_n$. Si $(T_n)_n$ est une autre surmartingale majorant $(Z_n)_n$ alors $T_N \geq U_N$ et si $T_n \geq U_n$, on a :

$$T_{n-1} \geq \mathbb{E}(T_n | \mathcal{F}_{n-1}) \geq \mathbb{E}(U_n | \mathcal{F}_{n-1})$$

et donc

$$T_{n-1} \geq \sup(Z_{n-1}, \mathbb{E}(U_n | \mathcal{F}_{n-1}))$$

ce qui démontre la proposition. \square

Ceci suggère que si l'on arrête convenablement la suite $(U_n)_n$, on obtient une martingale. C'est effectivement le cas puisque l'on a la propriété suivante.

Proposition 18.7.2. La variable aléatoire ν_0 définie par $\nu_0 = \inf\{n \geq 0 | U_n = Z_n\}$ est un temps d'arrêt et la suite arrêtée $(U_{n \wedge \nu_0})_{0 \leq n \leq N}$ est une martingale.

Démonstration. TRès facile. \square

Dans la suite, on notera $\mathcal{T}_{n,N}$ l'ensemble des temps d'arrêt qui prennent leurs valeurs dans $\{n, n+1, \dots, N\}$.

Corollaire 18.7.1. (arrêt optimal)

Le temps d'arrêt ν_0 vérifie :

$$U_0 = \mathbb{E}(Z_{\nu_0} | \mathcal{F}_0) = \sup_{\nu \in \mathcal{T}_{0,N}} \mathbb{E}(Z_\nu | \mathcal{F}_0)$$

On appelle alors temps d'arrêt optimal pour la suite $(Z_n)_n$ tout d'arrêt ν qui vérifie

$$\mathbb{E}(Z_\nu | \mathcal{F}_0) = \sup_{\mu \in \mathcal{T}_{0,N}} \mathbb{E}(Z_\mu | \mathcal{F}_0)$$

Démonstration. U^{ν_0} est une martingale et si $\nu \in \mathcal{T}_{0,N}$ alors U^ν est une surmartingale. On a donc

$$U_0 \geq \mathbb{E}(U_N^\nu | \mathcal{F}_0) = \mathbb{E}(U_\nu | \mathcal{F}_0) \geq \mathbb{E}(Z_\nu | \mathcal{F}_0)$$

Or $U_0 = U_0^{\nu_0} = \mathbb{E}(U_N^{\nu_0} | \mathcal{F}_0) = \mathbb{E}(Z_{\nu_0} | \mathcal{F}_0)$ et le résultat tombe. \square

On a le théorème suivant qui caractérise plus ou moins les temps d'arrêt optimaux.

Théorème 18.7.1. Un temps d'arrêt ν est optimal ssi :

$$\begin{cases} Z_\nu = U_\nu \\ U^\nu \text{ est une martingale} \end{cases} \quad (18.7.2)$$

Démonstration. Si la suite U^ν est une martingale alors $U_0 = \mathbb{E}(U_\nu|\mathcal{F}_0)$ et par conséquent si $Z_\nu = U_\nu$, on a $U_0 = \mathbb{E}(Z_\nu|\mathcal{F}_0)$ ce qui entraîne l'optimalité de ν par le corollaire précédent. Réciproquement, si ν est optimal, on a

$$U_0 = \mathbb{E}(Z_\nu|\mathcal{F}_0) \leq \mathbb{E}(U_\nu|\mathcal{F}_0)$$

Mais puisque U^ν est une surmartingale, on a aussi l'inégalité inverse et donc $\mathbb{E}(U_\nu|\mathcal{F}_0) = \mathbb{E}(Z_\nu|\mathcal{F}_0)$. Puisque $U_\nu \geq Z_\nu$, on a $U_\nu = Z_\nu$. D'autre part, puisque U^ν est une surmartingale, on a

$$U_0 \geq \mathbb{E}(U_n^\nu|\mathcal{F}_0) \geq \mathbb{E}(U_\nu|\mathcal{F}_0)$$

Or ce qui précède montre que $\mathbb{E}(U_\nu|\mathcal{F}_0) = U_0$. Il en résulte que

$$\mathbb{E}(U_n^{\nu\nu}|\mathcal{F}_0) = \mathbb{E}(U_\nu|\mathcal{F}_0) = \mathbb{E}(\mathbb{E}(U_\nu|\mathcal{F}_n)|\mathcal{F}_0)$$

Mais on a $U_n^\nu \geq \mathbb{E}(U_\nu|\mathcal{F}_n)$ et donc $U_n^\nu = \mathbb{E}(U_\nu|\mathcal{F}_n)$, ce qui prouve que U^ν est une martingale. \square

On rappelle la propriété suivante de décomposition des surmartingales (ou "décomposition de Doob"). Soit $(U_n)_{0 \leq n \leq N}$ une surmartingale. Il existe alors une martingale $(M_n)_{0 \leq n \leq N}$ et un processus prévisible croissant nul en 0 noté $(A_n)_{0 \leq n \leq N}$ tel que

$$U_n = M_n - A_n$$

Cette décomposition est de plus unique. Supposons maintenant que $(U_n)_n$ soit l'enveloppe de Snell d'une suite adaptée $(Z_n)_n$. Cette propriété permet de caractériser le plus grand temps d'arrêt optimal pour (Z_n) à l'aide du processus croissant $(A_n)_n$ intervenant dans la décomposition de Doob de $(U_n)_n$.

Proposition 18.7.3. *Le plus grand temps d'arrêt optimal pour la suite $(Z_n)_n$ est donné par :*

$$\nu_{max} = \begin{cases} N & \text{si } A_N = 0 \\ \inf\{n, A_{n+1} \neq 0\} & \text{si } A_N \neq 0 \end{cases} \quad (18.7.3)$$

Démonstration. Il est facile de voir que ν_{max} est un temps d'arrêt. Du fait que $A_j = 0$ pour $j \leq \nu_{max}$, on en déduit que $U^{\nu_{max}} = M^{\nu_{max}}$ ce qui entraîne que $U^{\nu_{max}}$ est une martingale. Pour avoir l'optimalité, il suffit de montrer l'égalité :

$$U_{\nu_{max}} = Z_{\nu_{max}}$$

Or

$$\begin{aligned} U_{\nu_{max}} &= \sum_{j=0}^{N-1} \mathbb{1}_{\{\nu_{max}=j\}} U_j + \mathbb{1}_{\{\nu_{max}=N\}} U_N \\ &= \sum_{j=0}^{N-1} \mathbb{1}_{\{\nu_{max}=j\}} \sup \{Z_j, \mathbb{E}(U_{j+1}|\mathcal{F}_j)\} + \mathbb{1}_{\{\nu_{max}=N\}} Z_N \end{aligned}$$

On a $\mathbb{E}(U_{j+1}|\mathcal{F}_j) = M_j - A_{j+1}$ et sur l'ensemble $\{\nu_{max} = j\}$, $A_j = 0$ et $A_{j+1} > 0$ donc $U_j = M_j$ et $\mathbb{E}(U_{j+1}|\mathcal{F}_j) = M_j - A_{j+1} < U_j$. Par suite $U_j = Z_j$ sur cet ensemble. On a donc démontré l'égalité

voulue.

Pour démontrer que c'est le plus grand temps d'arrêt optimal, on utilise le fait que si $\mathbb{P}(\nu > \nu_{max}) > 0$ alors on a

$$\mathbb{E}(U_\nu) = \mathbb{E}(M_\nu) - \mathbb{E}(A_\nu) = \mathbb{E}(U_0) - \mathbb{E}(A_\nu) < \mathbb{E}(U_0)$$

et par conséquent U^ν ne peut pas être une martingale. \square

Voyons maintenant comment ces résultats sur le contrôle optimal donnent des informations sur l'exercice et la couverture des options américaines. Pour simplifier les notations, on note \mathbb{P} l'unique probabilité sous laquelle les actifs actualisés sont des martingales. Nous avons vu que la valeur actualisée $(\tilde{U}_n)_n$ d'une option américaine décrite par une suite $(Z_n)_n$ était l'enveloppe de Snell de la suite actualisée $(\tilde{Z}_n)_n$.

Question : Comment le vendeur de l'option peut-il se couvrir parfaitement ? Réponse : On peut écrire \tilde{U} sous la forme $\tilde{M} - \tilde{A}$ où \tilde{M} est une martingale et \tilde{A} est un processus croissant prévisible nul en 0. Puisque le marché est complet, il existe une stratégie autofinancée ϕ telle que :

$$V_N(\phi) = S_N^0 \tilde{M}_N$$

c'est-à-dire $\tilde{V}_N(\phi) = \tilde{M}_N$. Comme la suite $(\tilde{V}_N(\phi))$ est une martingale (en tant que transformée de martingale), on a

$$\tilde{V}_n(\phi) = \tilde{M}_n$$

et donc

$$U_n = V_n(\phi) - A_n$$

A partir de cette expression, on voit comment le vendeur de l'option peut se couvrir. A l'instant 0, il doit encaisser la prime $U_0 = V_0(\phi)$ et suivre la stratégie ϕ . A l'instant n , il est capable de produire une richesse $V_n(\phi)$ qui majore U_n donc Z_n .

Dans le cadre du modèle de Cox-Ross-Rubinstein, on peut calculer assez facilement la stratégie ϕ dans le cas d'un put américain (exercice 4, Ch.2, p.37-38 de [24]).

Question : Quelle est la date d'exercice optimale pour l'acheteur de l'option ?

Réponse : La date d'exercice est à choisir parmi les temps d'arrêt. Le détenteur de l'option n'a pas intérêt à exercer à un temps n où $U_n > Z_n$ car il perdrait un actif de valeur U_n (l'option) contre une richesse égale à Z_n . Donc une date τ d'exercice optimal est telle que $U_\tau = Z_\tau$. Par ailleurs, il n'a pas intérêt à exercer après l'instant ν_{max} car à cet instant, en vendant l'option, il peut se constituer une richesse égale à $U_{\nu_{max}} = V_{\nu_{max}}(\phi)$ et, en suivant à partir de cet instant la stratégie ϕ , il se constitue un portefeuille dont la valeur est strictement plus grande que celle de l'option aux instants $\nu_{max} + 1, \dots, N$. La seconde condition est donc $\tau \leq \nu_{max}$, donc \tilde{U}_τ est une martingale. En conclusion, τ doit être un temps d'arrêt optimal.

Exercice 18.7.1. Soit E un ensemble fini et $(X_n)_{n \geq 0}$ une chaîne de Markov sur E de matrice de transition P . Soit $(Z_n)_n$ une suite adaptée définie par $Z_n = \psi(n, X_n)$ où ψ est une fonction de $\mathbb{N} \times E$ dans \mathbb{R} . Montrer que l'enveloppe de Snell $(U_n)_n$ de la suite $(Z_n)_n$ est donnée par $U_n = u(n, X_n)$ où u est la fonction définie par les relations suivantes :

$$u(N, x) = \psi(N, x)$$

$$u(n, \cdot) = \max(\psi(n, \cdot), Pu(n+1, \cdot)), \quad n \leq N-1$$

Exercice 18.7.2. (exercice 4, Ch. 2, [24])

On considère le modèle de Cox-Ross-Rubinstein et on adopte les notations de la précédente section.

1) Montrer que le prix \mathcal{P}_n à l'instant n du put américain d'échéance N , de prix d'exercice K sur une action peut s'écrire :

$$\mathcal{P}_n = P_{am}(n, S_n)$$

où $P_{am}(n, x)$ est définie par $P_{am}(N, x) = (K - x)_+$ et, pour $n \leq N - 1$,

$$P_{am}(n, x) = \max \left((K - x)_+, \frac{f(n+1, x)}{(1+r)} \right)$$

avec $f(n+1, x) = pP_{am}(n+1, x(1+a)) + (1-p)P_{am}(n+1, x(1+b))$ et $p = \frac{b-r}{b-a}$.

2) Montrer que la fonction $P_{am}(0, \cdot)$ peut se mettre sous la forme :

$$P_{am}(0, x) = \sup_{\nu \in \mathcal{T}_{0,N}} \mathbb{E}^*((1+r)^{-\nu}(K - xV_\nu)_+)$$

où la suite de variables aléatoires $(V_n)_{0 \leq n \leq N}$ est définie par : $\mathcal{L}V_0 = 1$ et, pour $n \geq 1$, $V_n = \prod_{i=1}^n U_i$ où les U_i sont des variables aléatoires dont on précisera la loi conjointe sous \mathbb{P}^* .

3) A partir de la formule de la question précédente, montrer que la fonction $x \rightarrow P_{am}(0, x)$ est convexe et décroissante.

4) On suppose $a < 0$. Montrer qu'il existe un réel $x^* \in [0, K]$ tel que pour $x \leq x^*$, $P_{am}(0, x) = (K - x)_+$ et pour $x \in (x^*, K/(1+a)^N)$, $P_{am}(0, x) > (K - x)_+$.

5) Un agent détient le put américain à l'instant 0. Pour quelles valeurs du cours spot S_0 a-t-il intérêt à exercer immédiatement son option ?

6) Montrer que la stratégie de couverture du put américain est définie par une quantité d'actif risqué $H_n = \Delta(n, S_{n-1})$ à détenir à l'instant n où Δ est une fonction que l'on exprimera à partir de la fonction P_{am} .

Chapitre 19

Chaînes de Markov

19.1 Analyse bibliographique brève

Ce sujet est exposé dans beaucoup de livres et à des niveaux très différents, avec des longueurs très différentes. Les trois critères que nous avons retenus sont :

1. Formalisation correcte : formalisation totale des chaînes, construction formelle correcte, preuve d'existence et preuve propre de la propriété de Markov forte.
2. Longueur d'exposition sous contrôle (les ouvrages consacrent de 12 à 100 pages à ce thème.)
3. Exemples applicatifs intéressants et en suffisance.
4. Lien avec les techniques de martingales développé.

En fait, aucun ouvrage ne répond à tous les critères comme le montre la liste suivante (nous mettons les numéros des critères qui sont respectés).

- [18], Grimmett et Stirzaker Probability and Random Processes 3, 4
- [14], Durrett Probability, Theory and Examples 3, 4
- [5], Borkar Probability theory 1, 2
- [9], Brzezniak et Zastawniak (Basic Stochastic Processes) 2, 3
- [17], Foatta et Fuchs (Processus stochastiques) 2, 3
- [29], Varadhan Probability theory 1, 3, 4
- [3], Billingsley Probability and Measure : pas de chaînes de Markov
- [28], Taylor (Introduction to measure and probability) : très peu de chaînes de Markov
- [15], Durrett Essentials of Stochastic Processes 2, 3. Très bien pour les exemples
- [2], Bauer Probability Theory : 1
- [8], Brémaud Markov chains, Gibbs fields, Monte Carlo simulations and queues. 3, 4
- [12], Dacunha-Castelle-Duflo Probability and Statistics 1, 2
- [1], Baldi, Mazliak, Priouret Martingales et Chaînes de Markov 1, 2, 4.
- [26], Ouvrard 1, 3, 4

Les ouvrages les plus recommandables pour la correction sont [1] et [26]. Les ouvrages les meilleurs pour les applications sont [14], [15] et [8]. Pour le cours, nous suivrons [1], qui réunit les avantages d'être écrit dans notre langue, d'être formellement correct et très court (20 pages). Il contient des exercices applicatifs en suffisance, mais n'est pas à la hauteur de [8], [14] et [15] pour un exposé amenant à des applications.

19.2 Préliminaires

19.2.1 Théorème de Kolmogorov

Théorème 19.2.1. Soit $\mu_n, n \in \mathbb{N}$ une famille de probabilités sur les espaces produits $(E^{n+1}, \sigma(\mathcal{E}^{n+1}))$ vérifiant

$$\forall n, \quad \mu_{n-1}(A_0 \times A_1 \times \cdots \times A_{n-1}) = \mu_n(A_0 \times A_1 \times \cdots \times A_{n-1} \times E). \quad (T)$$

Alors il existe une unique probabilité \mathbb{P} sur l'espace canonique (Ω, \mathcal{F}) définie par

$$\Omega = E^{\mathbb{N}}, \quad \omega = (\omega_n)_{n \in \mathbb{N}}, \quad X_n(\omega) = \omega_n, \quad \mathcal{F}_n = \sigma(X_k, k \leq n), \quad \mathcal{F} = \sigma(X_k, k \geq 0), \quad (19.2.1)$$

et telle que le processus $(\Omega, \mathcal{F}, \mathcal{F}_n, X_n, \mathbb{P})$ ait les μ_n comme lois de répartition finies, c'est-à-dire

$$\mu_n(A_0 \times A_1 \times \cdots \times A_{n-1} \times A_n) = \mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n).$$

Ce théorème est la réciproque de la remarque suivante sur les processus adaptés : Soient $X = (\Omega, \mathcal{F}, \mathcal{F}_n, X_n, \mathbb{P})$, $n \in \mathbb{N}$, une filtration et un processus adapté à valeurs dans (E, \mathcal{E}) . Alors on note μ_n la loi de la variable aléatoire (X_0, X_1, \dots, X_n) définie sur $(E^{n+1}, \sigma(\mathcal{E}^n))$. Vu que

$$\mu_n(A_0 \times A_1 \times \cdots \times A_n) = \mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n),$$

on a bien la relation (T).

19.3 Temps d'arrêt et tribu des événements antérieurs

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbb{P})$ une filtration. On pose $\mathcal{F}_\infty = \sigma(\cup_{n \geq 0} \mathcal{F}_n)$.

Définition 19.3.1. (i) on appelle temps d'arrêt de la filtration toute application $\nu : \Omega \rightarrow \bar{\mathbb{N}}$ telle que pour tout $n \geq 0$, on ait

$$\{\nu \leq n\} \in \mathcal{F}_n.$$

(ii) On appelle tribu des événements antérieurs à ν , notée \mathcal{F}_ν , la tribu

$$\mathcal{F}_\nu = \{A \in \mathcal{F}_\infty, \forall n \geq 0, A \cap \{\nu \leq n\} \in \mathcal{F}_n\}.$$

Soit X un processus adapté. On veut définir le "processus arrêté à l'instant ν ", i.e. $X_\nu(\omega) = X_{\nu(\omega)}(\omega)$. Comme il y a un problème lorsque $\nu(\omega) = +\infty$, une manière de procéder est de poser $X_\nu = \delta$ sur $\{\nu = +\infty\}$, où δ est un élément isolé adjoint à E que l'on appelle cimetière.

Avec cette convention, on définit $X_\nu = X_n$ sur $\{\nu = n\}$, $n \in \bar{\mathbb{N}}$. Alors X_ν est \mathcal{F}_ν -mesurable car

$$\{X_\nu \in B\} \cap \{\nu = n\} = \{X_n \in B\} \cap \{\nu = n\} \in \mathcal{F}_n, \quad n \in \bar{\mathbb{N}}.$$

En résumé, "le processus arrêté est mesurable par rapport à la tribu des événements antérieurs".

La proposition suivante nous sera utile.

Proposition 19.3.1. Soient X une v.a. positive ou intégrable et ν un temps d'arrêt. On pose pour tout $n \in \bar{\mathbb{N}}$, $X_n = \mathbb{E}(X|\mathcal{F}_n)$. Alors sur $\{\nu = n\}$, $\mathbb{E}(X|\mathcal{F}_\nu) = \mathbb{E}(X|\mathcal{F}_n)$, ce qui veut dire :

$$\mathbb{E}(X|\mathcal{F}_\nu) = X_\nu, \quad p.s. \quad (19.3.1)$$

Démonstration. On vérifie (19.3.1) pour $X \geq 0$. Soit $Z \geq 0$ et \mathcal{F}_ν -mesurable. Alors $Z\mathbb{1}_{\nu=n}$ est \mathcal{F}_n -mesurable et on a donc

$$\mathbb{E}(ZX_\nu) = \sum_{n \in \bar{\mathbb{N}}} \mathbb{E}(ZX_n \mathbb{1}_{\nu=n}) = \sum_{n \in \bar{\mathbb{N}}} \mathbb{E}(Z \mathbb{1}_{\nu=n} X) = \mathbb{E}(ZX).$$

Remarquer que $\mathbb{1}_{\nu=+\infty}$ est \mathcal{F}_∞ -mesurable, puisque $\{\nu = +\infty\} = (\cup_{n \in \mathbb{N}} \{\nu = n\})^c$. \square

Remarque 19.3.1. Pour montrer que $Y = \mathbb{E}(X|\mathcal{F})$, il faut et il suffit de montrer que $\mathbb{E}(YZ) = \mathbb{E}(XZ)$ pour toute fonction $Z \geq 0$ et \mathcal{F} -mesurable : c'est comme cela que nous venons de procéder. En effet, par définition de l'espérance conditionnelle, $Y = \mathbb{E}(X|\mathcal{F})$ si et seulement si $\mathbb{E}(YZ) = \mathbb{E}(XZ)$ pour tout $Z = \mathbb{1}_A$ avec A \mathcal{F} -mesurable. Cette dernière relation est encore vraie par la machine standard pour tout $Z \geq 0$, \mathcal{F} -mesurable.

19.4 Chaînes de Markov : la construction

19.4.1 Matrices de transition

Dans la suite, E est un ensemble d'états, fini ou dénombrable et \mathcal{E} la tribu de toutes les parties de E , tribu engendrée par les atomes. On note

- \mathcal{E}^+ : applications $E \rightarrow \mathbb{R}^+$. Exemple : les fonctions caractéristiques de sous-ensembles de E , ce sont des vecteurs colonne. $\mathbb{1}_E$ est un tel vecteur colonne valant 1 sur les lignes de $x \in E$, 0 ailleurs.)
- \mathcal{M}^+ : mesures positives sur E notées $\mu = (\mu(x))_{x \in E}$. Exemple : les lois initiales sur E . Ce sont des vecteurs ligne et on peut faire les produits $\mu.f$ pour $\mu \in \mathcal{M}^+$, $f \in \mathcal{E}^+$.
- P, Q, R , matrices positives ($P(x, y) \in \bar{\mathbb{R}}^+$), $x, y \in E$. Le produit de deux matrices se définit par $QR(x, y) = \sum_{z \in E} Q(x, z)R(z, y)$, avec la convention importante : $+\infty \times 0 = 0$.
- Si $f \in \mathcal{E}^+$, $(Qf)(x) = \sum_{y \in E} Q(x, y)f(y)$.
- Si $\mu \in \mathcal{M}^+$, $(\mu Q)(y) = \sum_{x \in E} \mu(x)Q(x, y)$.

Définition 19.4.1. On appelle *matrice de transition sur E* une famille $(P(x, y))_{x, y \in E}$ de réels telle que pour tous $x, y \in E$,

$$P(x, y) \geq 0, \quad \sum_{y \in E} p(x, y) = 1.$$

Ceci implique que $\forall x \in E$, l'application $A \rightarrow P(x, A) = \sum_{y \in A} P(x, y)$ est une probabilité sur E .

$P \geq 0$ est une matrice de transition si et seulement si $P\mathbb{1} = \mathbb{1}$ (où $\mathbb{1} := \mathbb{1}_E$.) En conséquence, le produit de deux matrices de transition est une matrice de transition. Si μ est une probabilité et P une matrice de transition, μP est une probabilité.

Définition 19.4.2. Soient μ une probabilité et P une matrice de transition sur E . On appelle *chaîne de Markov de loi initiale μ et de matrice de transition P* , un processus aléatoire $(\Omega, \mathcal{F}, \mathcal{F}_n, X_n, \mathbb{P})$ à valeurs dans E et tel que

$$(i) \mathbb{P}(X_0 \in A) = \mu(A), \quad \forall A \subset E$$

$$(ii) \mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = P(X_n, A) \text{ p.s.}, \quad \forall A \subset E, n \geq 0.$$

Remarque 19.4.1. La relation (ii) implique en particulier que $\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in A | X_n)$. En effet, la v.a. $P(X_n, A)$ est X_n -mesurable et par la propriété des tours :

$$\mathbb{P}(X_{n+1} \in A | X_n) = \mathbb{P}(\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) | X_n) = \mathbb{P}(P(X_n, A) | X_n) = P(X_n, A).$$

Proposition 19.4.1. Si $\mathcal{F}_n = \mathcal{F}_n^0 = \sigma(X_0, X_1, \dots, X_n)$ est la "filtration naturelle du processus", la tribu \mathcal{F}_n^0 est engendrée par la partition de Ω en événements $\{X_0 = a_0, X_1 = a_1, \dots, X_n = a_n\}$, $(a_0, a_1, \dots, a_n) \in E^n$. Alors (ii) peut être remplacée par

$$(ii1) \quad \mathbb{P}(X_{n+1} = b | X_n = a_n, X_{n-1} = a_{n-1}, \dots, X_0 = a_0) = P(a_n, b),$$

pour tous $n \geq 0$, tous $(a_0, a_1, \dots, a_n, b) \in E^{n+2}$ et tels que $\mathbb{P}(X_n = a_n, \dots, X_0 = a_0) > 0$.

Démonstration. Il suffit de se rappeler que quand \mathcal{F} est engendrée par des événements d'une partition dénombrable Ω_i , on a

$$\mathbb{P}(X \in A | \mathcal{F}) = \sum_i \mathbb{P}(A | \Omega_i) \mathbb{1}_{\Omega_i},$$

et comme l'ensemble des valeurs est aussi dénombrable, cela revient à écrire

$$\forall b \in E, \mathbb{P}(X = b | \mathcal{F}) = \sum_i \mathbb{P}(X = b | \Omega_i) \mathbb{1}_{\Omega_i}.$$

Par (ii), $\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = P(X_n, A)$ p.s. $\Leftrightarrow \forall b \in E, \mathbb{P}(X_{n+1} = b | \mathcal{F}_n) = P(X_n, b)$ p.s. En appliquant la relation précédente avec $\Omega_i = \{(a_0, \dots, a_n)\}$, cette dernière relation est équivalente à (ii1). En effet,

$$\begin{aligned} (ii) &\Leftrightarrow (\forall b, \mathbb{P}(X_{n+1} = b | \mathcal{F}_n) = P(X_n, b)) \\ &\Leftrightarrow \sum_i \mathbb{P}(X_{n+1} = b | \Omega_i) \mathbb{1}_{\Omega_i} = P(X_n, b) \\ &\Leftrightarrow (\forall b, \forall (a_0, \dots, a_n), \mathbb{P}(X_{n+1} = b | X_0 = a_0, \dots, X_n = a_n) = P(a_n, b)). \end{aligned}$$

□

Théorème 19.4.1. (i) Soit $X = (\Omega, \mathcal{F}, \mathcal{F}_n, X_n, \mathbb{P})$ une chaîne de Markov à valeurs dans E de loi initiale μ et de matrice de transition P . On a, pour tout $(a_0, \dots, a_n) \in E^{n+1}$,

$$\mathbb{P}(X_0 = a_0, X_1 = a_1, \dots, X_n = a_n) = \mu(a_0)P(a_0, a_1) \dots P(a_{n-1}, a_n). \quad (19.4.1)$$

(ii) Réciproquement, si un processus $X = (\Omega, \mathcal{F}, X_n, \mathbb{P})$ à valeurs dans E vérifie (19.4.1) pour une probabilité μ et une matrice de transition P sur E , alors X est une chaîne de Markov de loi initiale μ et de matrice de transition P par rapport à la filtration naturelle $(\mathcal{F}_n^0)_{n \geq 0}$.

Démonstration. (i)

$$\begin{aligned} &\mathbb{P}(X_0 = a_0, \dots, X_{n-1} = a_{n-1}, X_n = a_n) = \\ &\mathbb{E}(\mathbb{1}_{X_0=a_0, \dots, X_{n-1}=a_{n-1}} \mathbb{1}_{X_n=a_n}) = \text{(définition de l'espérance conditionnelle)} \\ &\mathbb{E}(\mathbb{1}_{X_0=a_0, \dots, X_{n-1}=a_{n-1}} \mathbb{E}(\mathbb{1}_{X_n=a_n} | \mathcal{F}_{n-1})) = \text{(définition chaîne de Markov, (ii))} \\ &\mathbb{E}(\mathbb{1}_{X_0=a_0, \dots, X_{n-1}=a_{n-1}} P(X_{n-1}, a_n)) = \\ &\mathbb{P}(X_0 = a_0, \dots, X_{n-1} = a_{n-1})P(a_{n-1}, a_n). \end{aligned}$$

(ii) Réciproque. Comme on considère la filtration naturelle \mathcal{F}_n^0 , il suffit de vérifier la "forme simple" de (ii), à savoir (ii1). Or, en utilisant (19.4.1),

$$\begin{aligned} \mathbb{P}(X_{n+1} = b | X_n = a_n, \dots, X_0 = a_0) &= \frac{\mathbb{P}(X_0 = a_0, \dots, X_n = a_n, X_{n+1} = b)}{\mathbb{P}(X_0 = a_0, \dots, X_n = a_n)} = \\ &= \frac{\mu(a_0)P(a_0, a_1) \dots P(a_{n-1}, a_n)P(a_n, b)}{\mu(a_0)P(a_0, a_1) \dots P(a_{n-1}, a_n)} = P(a_n, b). \end{aligned}$$

□

19.4.2 Conséquences de la relation (19.4.1)

1. Soit $f : E^{n+1} \rightarrow \mathbb{R}$ positive ou bornée. Alors (il suffit de sommer)

$$\mathbb{E}(f(X_0, \dots, X_n)) = \sum_{a_0, \dots, a_n \in E} f(a_0, \dots, a_n) \mu(a_0) P(a_0, a_1) \dots P(a_{n-1}, a_n) \quad (19.4.2)$$

2. Si on choisit $f(x_0, \dots, x_n) = \mathbb{1}_{A_0}(x_0) \dots \mathbb{1}_{A_n}(x_n)$, $A_i \in E$,

$$\mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n) = \sum_{a_0 \in A_0, \dots, a_n \in A_n} \mu(a_0) P(a_0, a_1) \dots P(a_{n-1}, a_n) \quad (19.4.3)$$

Si, dans (19.4.3), on choisit $A_i = \{a_i\}$ pour certains indices et $A_j = E$ pour d'autres, on a pour $l_1, \dots, l_k \in \mathbb{N}$,

$$\mathbb{P}(X_0 = a_0, X_{l_1} = a_1, \dots, X_{l_1 + \dots + l_k} = a_k) = \mu(a_0) (P^{l_1})(a_0, a_1) \dots (P^{l_k})(a_{k-1}, a_k). \quad (19.4.4)$$

Démonstration de (19.4.4). Le mieux est de traiter un exemple, le cas général se faisant par récurrence sur le même principe. On a par la relation (19.4.3),

$$\begin{aligned} \mathbb{P}(X_0 = \alpha_0, X_1 \in E, X_2 \in E, X_3 = \alpha_3) &= \sum_{a_1, a_2 \in E} \mu(\alpha_0) P(\alpha_0, a_1) P(a_1, a_2) P(a_2, \alpha_3) = \\ &= \mu(\alpha_0) \sum_{a_2 \in E} P^2(\alpha_0, a_2) P(a_2, \alpha_3) = \mu(\alpha_0) P^3(\alpha_0, \alpha_3). \end{aligned}$$

3. Conséquence de (19.4.1) : si $\mathbb{P}(X_0 = a_0, \dots, X_n = a_n) > 0$,

$$\mathbb{P}(X_{n+1} = b_1, \dots, X_{n+k} = b_k | X_0 = a_0, \dots, X_n = a_n) = P(a_n, b_1) \dots P(b_{k-1}, b_k). \quad (19.4.5)$$

4. Par application de (19.4.4), $\mathbb{P}(X_{n+k} = b | X_0 = a_0, \dots, X_n = a_n) = P^k(a_n, b)$.

19.5 Construction de l'espace canonique d'une matrice de transition

Théorème 19.5.1. Soient μ une probabilité et P une matrice de transition sur E . Alors il existe une chaîne de Markov à valeurs dans E , de loi initiale μ et de matrice de transition P . Plus précisément :

$$\Omega = E^{\mathbb{N}}, \quad \omega = (\omega_n)_{n \geq 0}, \quad X_n(\omega) = \omega_n, \quad \mathcal{F}_n = \sigma(X_k, k \leq n), \quad \mathcal{F} = \sigma(X_k, k \geq 0)$$

et il existe une unique probabilité \mathbb{P} sur cet espace canonique vérifiant :

$$\mathbb{P}(X_0 = a_0, \dots, X_n = a_n) = \mu_n(a_0, \dots, a_n) := \mu(a_0) P(a_0, a_1) \dots P(a_{n-1}, a_n). \quad (19.5.1)$$

Démonstration. Par le théorème de Kolmogorov, il suffit de vérifier que

- (a) μ_n est une probabilité sur E^{n+1}
 (b) La condition de compatibilité (T) est vérifiée :

$$\mu_{n-1}(A_0 \times A_1 \times \cdots \times A_{n-1}) = \mu_n(A_0 \times A_1 \times \cdots \times A_{n-1} \times E).$$

Preuve de (a) : $\mu_n(a_0, \dots, a_n) = \mu(a_0)P(a_0, a_1) \dots P(a_{n-1}, a_n)$. On somme successivement sur $a_n \in E$, puis $a_{n-1} \in E$, etc. jusqu'à $a_0 \in E$ et on trouve que $\sum_{a_0, \dots, a_n \in E^{n+1}} \mu_n(a_0, \dots, a_n) = 1$.

Preuve de (b). Il suffit de montrer que

$$\mu_{n-1}(a_0, \dots, a_{n-1}) = \mu_n(\{a_0\} \times \cdots \times \{a_{n-1}\} \times E),$$

et c'est immédiat en reprenant la relation (19.5.1) et en sommant par rapport à a_n : on obtient ainsi

$$\mu_n(\{a_0\} \times \cdots \times \{a_{n-1}\} \times E) = \mu(a_0)P(a_0, a_1) \dots P(a_{n-2}, a_{n-1}) \times 1 = \mu_{n-1}(a_0, \dots, a_{n-1}).$$

□

19.6 La chaîne de Markov canonique associée à une matrice de transition

Soient P une matrice de transition sur E et $\mu \in \mathcal{M}^+$, $\mu.1 = 1$ une mesure de probabilité sur E . On appelle \mathbb{P}_μ la probabilité construite sur l'espace canonique Ω par le théorème de Kolmogorov. Si $\mu = \delta_x$, on note $\mathbb{P}_\mu = \mathbb{P}_x$. En utilisant de nouveau (19.4.1), on a

$$\mathbb{P}_\mu(A) = \sum_{x \in E} \mu(x) \mathbb{P}_x(A) \text{ si } A = \{X_0 = a_0, \dots, X_n = a_n\}. \quad (19.6.1)$$

Par le théorème de la classe monotone, on déduit que (19.6.1) est encore vraie pour tout $A \in \mathcal{F}$.

Définition 19.6.1. Soit P une matrice de transition sur E . On appelle chaîne de Markov canonique de matrice de transition P le quintuplet $(\Omega, \mathcal{F}, \mathcal{F}_n, X_n, (\mathbb{P}_x)_{x \in E})$ où $(\Omega, \mathcal{F}_n, \mathcal{F}, X_n)$ est l'espace canonique associé à E et, pour chaque x , \mathbb{P}_x est une probabilité sur (Ω, \mathcal{F}) vérifiant pour tout (a_0, \dots, a_n) ,

$$\mathbb{P}_x(X_0 = a_0, \dots, X_n = a_n) = \mathbb{1}_x(a_0)P(a_0, a_1) \dots P(a_{n-1}, a_n). \quad (19.6.2)$$

La relation (19.6.1) permet de définir \mathbb{P}_μ et alors $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, X_n, \mathbb{P}_\mu)$ est une chaîne de Markov de loi initiale μ et de matrice de transition P .

Proposition 19.6.1.

$$\mathbb{P}_x(X_n = y) = P^n(x, y), \quad \mathbb{E}_x(f(X_n)) = (P^n f)(x) \quad (19.6.3)$$

$$\mathbb{E}_\mu(f(X_n)) = \sum_{x \in E} \mu(x)(P^n f)(x) = \mu(P^n f) = (\mu P^n) f. \quad (19.6.4)$$

Les derniers produits peuvent être vus comme des produits matriciels. La relation (19.6.4) veut dire : "Si μ est la loi de X_0 , μP^n est la loi de X_n ".

Démonstration. La première relation de (19.6.3) découle de (19.4.4) : $\mathbb{P}_x(X_n = y) = \mathbb{P}_x(X_0 = x, X_n = y) = P^n(x, y)$. Pour la seconde relation, on applique (19.4.2) :

$$\begin{aligned}\mathbb{E}_x(f(X_n)) &= \sum_{a_0, \dots, a_n \in E} f(a_n) \mathbb{1}_x(a_0) P(a_0, a_1) \dots P(a_{n-1}, a_n) \\ &= \sum_{a_1, \dots, a_n \in E} f(a_n) P(x, a_1) \dots P(a_{n-1}, a_n) \\ &= \sum_{a_n \in E} f(a_n) P^n(x, a_n) = (P^n f)(x).\end{aligned}$$

La relation (19.6.4) découle de (19.6.1) et de la relation précédente. \square

19.6.1 Canonisation d'un processus de Markov

Une chaîne de Markov n'est pas en général définie sur l'espace canonique. Soit $X' = (\Omega', \mathcal{F}', \mathcal{F}'_n, X'_n, \mathbb{P}')$ une chaîne de Markov de loi initiale μ et de matrice de transition \mathbb{P} et soit la chaîne canonique $X = (\Omega, \mathcal{F}, \mathcal{F}_n, X_n, (\mathbb{P}_x)_{x \in E})$ associée à P . Alors pour tout $(a_0, \dots, a_n) \in E^{n+1}$, on a

$$\begin{aligned}\mathbb{P}'(X'_0 = a_0, \dots, X'_n = a_n) &= \mu(a_0) P(a_0, a_1) \dots P(a_{n-1}, a_n) \\ &= \mathbb{P}_\mu(X_0 = a_0, \dots, X_n = a_n).\end{aligned}$$

Donc, par le théorème de la classe monotone 2.4, on a pour toute $\varphi \in b\mathcal{F}$ ou \mathcal{F}^+ ,

$$\mathbb{E}'(\varphi(X'_0, X'_1, \dots, X'_n, \dots)) = \mathbb{E}_\mu(\varphi(X_0, X_1, \dots, X_n, \dots)).$$

Les deux processus X et X' ont donc même loi. Donc, tous les calculs d'espérance et de probabilité concernant X' peuvent se faire avec X .

19.7 Calculs sur la chaîne canonique et propriété de Markov forte

Définition 19.7.1. Sur l'espace canonique

$$(\Omega = E^{\mathbb{N}}, \omega = (\omega_n)_{n \geq 0}, X_n(\omega) = \omega_n, \mathcal{F}_n = \sigma(X_k, k \leq n), \mathcal{F} = \sigma(X_k, k \geq 0)),$$

on définit l'opérateur de translation $\theta : \Omega \rightarrow \Omega, \omega = (\omega_n)_{n \geq 0} \rightarrow (\omega_{n+1})_{n \geq 0}$. On pose $\theta_p = \theta^p$ et θ_p est mesurable de $(\Omega, \sigma(X_k, k \geq p))$ dans (Ω, \mathcal{F}) . On a :

$$X_n \circ \theta_p = X_{n+p}.$$

On peut interpréter θ_p comme un opérateur de remise à zéro des compteurs à l'instant p , i.e. d'oubli de tout ce qui s'est passé strictement avant p . On considère un temps d'arrêt τ . On pose $\Omega_\tau = \{\tau < +\infty\}$ et on définit θ_τ sur Ω_τ par $\theta_\tau = \theta_n$ sur $\{\tau = n\}$. On a alors sur Ω_τ :

$$\forall n \geq 0, \quad X_n \circ \theta_\tau = X_{n+\tau}.$$

Lemme 19.7.1. (4.10 page 90) On a :

$$\mathbb{P}_\mu(X_{n+1} = b_1, \dots, X_{n+k} = b_k | \mathcal{F}_n) = \mathbb{P}_{X_n}(X_1 = b_1, \dots, X_k = b_k).$$

(Dans le second membre, μ disparaît mais la dépendance de μ se fait via X_n).

Remarque 19.7.1. Une des notations précédentes peut paraître abusive : il s'agit de $\mathbb{P}_{X_n}(X_1 = b_0, \dots, X_k = b_k)$. En effet, X_n apparaît dans cette notation comme le point de départ et aussi comme un élément de la suite X_0, \dots, X_k . Il faut comprendre ce terme comme $\mathbb{P}_{X_n(\omega)}(\{\rho \in \Omega, \rho_0 = b_0, \dots, \rho_k = b_k\})$. On a donc fait un abus de notation du type $\int_0^x f(x)dx$, où x est utilisée comme variable explicite et muette à la fois.

Démonstration. Par la propriété (19.4.5),

$$\mathbb{P}_\mu(X_{n+1} = b_1, \dots, X_{n+k} = b_k | X_0 = a_0, \dots, X_n = a_n) = P(a_n, b_1) \dots P(b_{k-1}, b_k).$$

Mais, posant $A = \{X_{n+1} = b_1, \dots, X_{n+k} = b_k\}$, on a

$$\begin{aligned} \mathbb{P}_\mu(A | \mathcal{F}_n) &= \sum_{a_0, \dots, a_n \in E} \mathbb{P}_\mu(A | X_0 = a_0, \dots, X_n = a_n) \mathbb{1}_{X_0=a_0, \dots, X_n=a_n} = \\ &= \sum_{a_0, \dots, a_n \in E} P(a_n, b_1) \dots P(b_{k-1}, b_k) \mathbb{1}_{X_0=a_0, \dots, X_n=a_n} \\ &= P(X_n, b_1) \dots P(b_{k-1}, b_k) \sum_{a_0, \dots, a_{n-1} \in E} \mathbb{1}_{X_0=a_0, \dots, X_{n-1}=a_{n-1}} = \mathbb{P}_{X_n}(X_1 = b_1, \dots, X_k = b_k), \end{aligned}$$

□

Théorème 19.7.1. (Propriété de Markov, théorème 4.5) Soit X une chaîne de Markov canonique. On a pour tout $n \geq 0$, toute $\varphi \in b\mathcal{F}$ ou \mathcal{F}^+ et toute loi initiale μ ,

$$\mathbb{E}_\mu(\varphi \circ \theta_n | \mathcal{F}_n) = \mathbb{E}_{X_n}(\varphi), \text{ p.s.}, \quad (19.7.1)$$

(qui est une fonction de X_n) ou, de manière équivalente, pour tout $n \geq 0$, toute $\varphi \in b\mathcal{F}$, toute $\psi \in b\mathcal{F}_n$, (resp. $\varphi \in \mathcal{F}^+$, $\psi \in \mathcal{F}_n^+$) et toute loi initiale μ :

$$\mathbb{E}_\mu(\psi \varphi \circ \theta_n) = \mathbb{E}_\mu(\psi \mathbb{E}_{X_n}(\varphi)). \quad (19.7.2)$$

Démonstration. Prenons d'abord $\varphi = \mathbb{1}_{X_0=b_0, \dots, X_k=b_k}$. Alors

$$\mathbb{E}_\mu(\varphi \circ \theta_n | \mathcal{F}_n) = \mathbb{E}_\mu(\mathbb{1}_{X_n=b_0, \dots, X_{n+k}=b_k} | \mathcal{F}_n) =$$

$$\mathbb{1}_{X_n=b_0} \mathbb{P}_\mu(X_{n+1} = b_1, \dots, X_{n+k} = b_k | \mathcal{F}_n) = \mathbb{1}_{X_n=b_0} \mathbb{P}_{X_n}(X_1 = b_1, \dots, X_k = b_k) = \mathbb{E}_{X_n}(\varphi).$$

La troisième égalité découle du lemme 19.7.1. Expliquons la dernière égalité. Comme φ est une fonction caractéristique, on a immédiatement

$$\mathbb{E}_{X_n}(\varphi) = \mathbb{P}_{X_n}(X_0 = b_0, \dots, X_k = b_k) = 0 \text{ si } b_0 \neq X_n, = \mathbb{P}_{X_n}(X_1 = b_1, \dots, X_k = b_k) \text{ si } b_0 = X_n.$$

Nous avons donc prouvé (19.7.1) quand $\varphi = \mathbb{1}_{X_0=a_0, \dots, X_n=a_n}$ et on peut appliquer le théorème de la classe monotone, puisque les ensembles $\{X_0 = a_0, \dots, X_n = a_n\}$, $n \in \mathbb{N}$ forment un π -système de \mathcal{F} . (Théorème 2.5 page 19 du Baldi-Mazliak-Priouret). Montrons maintenant (19.7.2) : c'est juste la définition de l'espérance conditionnelle étendue de $\psi = \mathbb{1}_A$, A \mathcal{F}_n -mesurable à $b\mathcal{F}_n$ ou à \mathcal{F}_n^+ par la machine standard. □

Théorème 19.7.2. (Propriété de Markov forte, théorème 4.6). Soient X une chaîne de Markov canonique et τ un temps d'arrêt. On a pour tout $n \geq 0$, toute $\varphi \in b\mathcal{F}$ (ou \mathcal{F}^+), toute $\psi \in b\mathcal{F}_\tau$ (ou \mathcal{F}_τ^+) et toute loi initiale μ ,

$$\begin{aligned}\mathbb{E}_\mu(\mathbb{1}_{\tau < +\infty} \varphi \circ \theta_\tau | \mathcal{F}_\tau) &= \mathbb{1}_{\tau < +\infty} \mathbb{E}_{X_\tau}(\varphi) \text{ p.s. ,} \\ \mathbb{E}_\mu(\mathbb{1}_{\tau < +\infty} \psi \varphi \circ \theta_\tau) &= \mathbb{E}_\mu(\mathbb{1}_{\tau < +\infty} \psi \mathbb{E}_{X_\tau}(\varphi))\end{aligned}\tag{19.7.3}$$

(Comparer ces relations avec (19.7.1) et (19.7.1)).

Démonstration. On rappelle (Proposition 2.7 page 21) que si $X \geq 0$ est un processus, ν un temps d'arrêt, $X_n = \mathbb{E}(X | \mathcal{F}_n)$, alors sur $\{\nu = n\}$, on a par définition,

$$\mathbb{E}(X | \mathcal{F}_\nu) = \mathbb{E}(X | \mathcal{F}_n) \text{ p.s.}$$

Appliquons cette relation, puis la propriété de Markov (19.7.1) :

$$\mathbb{E}_\mu(\varphi \circ \theta_\tau | \mathcal{F}_\tau) = \mathbb{E}_\mu(\varphi \circ \theta_n | \mathcal{F}_n) = \mathbb{E}_{X_n}(\varphi), \quad \mathbb{P}_\mu - \text{p.s. sur } \{\tau = n\}.$$

Donc

$$\sum_n \mathbb{1}_{\tau=n} \mathbb{E}_\mu(\varphi \circ \theta_\tau | \mathcal{F}_\tau) = \sum_n \mathbb{1}_{\tau=n} \mathbb{E}_{X_n}(\varphi) = \mathbb{1}_{\tau < +\infty} \mathbb{E}_{X_\tau}(\varphi), \quad \mathbb{P}_\mu - \text{p.s.}$$

Mais $\sum_n \mathbb{1}_{\tau=n}$ est \mathcal{F}_τ -mesurable et donc

$$\sum_n \mathbb{1}_{\tau=n} \mathbb{E}_\mu(\varphi \circ \theta_\tau | \mathcal{F}_\tau) = \mathbb{E}_\mu(\mathbb{1}_{\tau < +\infty} \varphi \circ \theta_\tau | \mathcal{F}_\tau),$$

ce qui prouve la première relation annoncée. La seconde, (19.7.3) s'en déduit de la même manière que (19.7.2) de (19.7.1). \square

19.8 Première grande application : la probabilité de retour

Lemme 19.8.1. Soient τ et σ deux temps d'arrêt. Alors ρ défini par $\rho = \tau + \sigma \circ \theta_\tau$ sur $\{\tau < \infty\}$ et $\rho = +\infty$ sinon est aussi un temps d'arrêt. De plus,

$$\{\rho < +\infty\} = \{\tau < +\infty\} \cap \{\sigma \circ \theta_\tau < +\infty\}, \quad X_\rho = X_\sigma \circ \theta_\tau \text{ sur } \{\rho < +\infty\}.$$

La démonstration est immédiate : un ensemble $\rho = n$ s'exprime en fonction des ensembles $\tau = k$ et $\sigma \circ \theta_k = l$ pour de valeurs de k et l inférieures ou égales à n . Le lemme précédent nous permet de définir les *temps de passage par x successifs après θ* comme des temps d'arrêt. On pose

$$\begin{aligned}\sigma_x &:= \inf\{n \geq 1, X_n = x\} = \sigma_x^1, \\ \sigma_x^n &:= \sigma_x^{n-1} + \sigma_x \circ \theta_{\sigma_x^{n-1}} \text{ si } \sigma_x^{n-1} < +\infty, \\ \sigma_x^n &:= \sigma_x^{n-1} = +\infty \text{ sinon.}\end{aligned}$$

Proposition 19.8.1. Soit X une chaîne de Markov canonique de matrice de transition P . Alors

$$\mathbb{P}_x(\sigma_x^n < +\infty) = \mathbb{P}_x(\sigma_x < +\infty)^n.\tag{19.8.1}$$

Démonstration.

$$\begin{aligned}\mathbb{P}_x(\sigma_x^n < +\infty) &= \mathbb{P}_x(\sigma_x^{n-1} + \sigma_x \circ \theta_{\sigma_x^{n-1}} < +\infty) \\ &= \mathbb{P}_x(\sigma_x^{n-1} < +\infty, \sigma_x \circ \theta_{\sigma_x^{n-1}} < +\infty) \\ &= \mathbb{E}_x(\mathbb{1}_{\sigma_x^{n-1} < +\infty} \mathbb{1}_{\sigma_x < +\infty} \circ \theta_{\sigma_x^{n-1}}).\end{aligned}$$

Mais, dans ce dernier terme, on peut appliquer la propriété de Markov forte (19.7.3) avec $\varphi = \mathbb{1}_{\sigma_x < +\infty}$, et on obtient

$$\begin{aligned}\mathbb{P}_x(\sigma_x^n < +\infty) &= \mathbb{E}_x(\mathbb{1}_{\sigma_x^{n-1} < +\infty} \mathbb{E}_{X_{\sigma_x^{n-1}}}(\mathbb{1}_{\sigma_x < +\infty})) = \mathbb{E}_x(\mathbb{1}_{\sigma_x^{n-1} < +\infty} \mathbb{P}_x(\sigma_x < +\infty)), \\ (\text{car } X_{\sigma_x^{n-1}} &= x \text{ sur } \{\sigma_x^{n-1} < +\infty\}) \\ &= \mathbb{P}_x(\sigma_x^{n-1} < +\infty) \mathbb{P}_x(\sigma_x < +\infty).\end{aligned}$$

□

19.9 Opérateurs potentiels

Définition 19.9.1. Soient P une matrice de transition et X la chaîne canonique associée. On appelle matrice potentielle de P (ou de X) la matrice

$$U = I + P + \dots + P^n + \dots = \sum_{k \geq 0} P^k.$$

Lemme 19.9.1. Si $A \subset E$ et $N_A(\omega)$ désigne le nombre de visites de la trajectoire ω à A , on a

$$N_A = \sum_{k \geq 0} \mathbb{1}_A(X_k), \quad \mathbb{E}_x(N_A) = U(x, A), \quad E_x(N_y) = U(x, y). \quad (19.9.1)$$

Démonstration. La première formule est la définition du nombre de visites de A . On rappelle la formule (4.13) : $\mathbb{E}_x(f(X_n)) = (P^n f)(x)$. En appliquant cette formule à $f = \mathbb{1}_A$,

$$\begin{aligned}\mathbb{E}_x(N_A) &= \mathbb{E}_x\left(\sum_{k \geq 0} \mathbb{1}_A(X_k)\right) = \sum_{k \geq 0} \mathbb{E}_x(\mathbb{1}_A(X_k)) = \\ &= \sum_{k \geq 0} (P^k \mathbb{1}_A)(x) = (U \mathbb{1}_A)(x) = \delta_x U \mathbb{1}_A = U(x, A).\end{aligned}$$

La troisième formule est l'application de la précédente à $A = \{x\}$. □

19.10 Récurrence et transience

Théorème 19.10.1. Soit $x \in E$. On rappelle que $U(x, x) = E_x(N_x)$.

- (i) Si $U(x, x) = +\infty$, alors $\mathbb{P}_x(\sigma_x < +\infty) = 1$ et $\mathbb{P}_x(N_x = +\infty) = 1$. Le point x est dit récurrent.
- (ii) Si $U(x, x) < +\infty$, alors $\mathbb{P}_x(\sigma_x < +\infty) < 1$ et $\mathbb{P}(N_x < +\infty) = 1$. Le point x est dit transitoire (transient).

On passe donc par ce théorème de l'espérance à la probabilité : si le point est récurrent, presque toutes les trajectoires passent une infinité de fois par x . Si x est transitoire, presque toutes les trajectoires ne passent qu'un nombre fini de fois par x .

Démonstration. On vérifie que

$$N_x = \mathbb{1}_{\{x\}}(X_0) + \sum_{n \geq 1} \mathbb{1}_{\sigma_x^n < +\infty} \quad (19.10.1)$$

Deux cas sont possibles.

(i) $\mathbb{P}_x(\sigma_x < +\infty) = 1$. Alors pour tout $n \geq 0$, $\mathbb{P}_x(\sigma_x^n < +\infty) = 1$ par (19.8.1). Donc par (19.10.1), $N_x = +\infty$ \mathbb{P}_x -p.s. et $U(x, x) = \mathbb{E}_x(N_x) = +\infty$.

(ii) $\mathbb{P}_x(\sigma_x < +\infty) = a < 1$. Alors par (19.10.1), $U(x, x) = \mathbb{E}_x(N_x) = 1 + \sum_{n \geq 1} \mathbb{P}_x(\sigma_x^n < +\infty) = 1 + \sum_{n \geq 1} a^n < +\infty$. On en déduit que $N_x < +\infty$ \mathbb{P}_x -p.s. \square

Proposition 19.10.1. *Si $x \neq y$, alors $U(x, y) = \mathbb{P}_x(\sigma_y < +\infty)U(y, y)$. En particulier, $U(y, y) \geq U(x, y)$.*

Démonstration. On va utiliser la propriété de Markov forte, et les relations $X_{\sigma_y} = y$ sur $\{\sigma_y < +\infty\}$, $N_y = \mathbb{1}_{\sigma_y < +\infty} N_y \circ \theta_{\sigma_y}$, \mathbb{P}_x -p.s. (On a $x \neq y$, donc le compte de N_y commence en σ_y).

$$U(x, y) = \mathbb{E}_x(N_y) = \mathbb{E}_x(\mathbb{1}_{\sigma_y < +\infty} N_y \circ \theta_{\sigma_y}) = \mathbb{E}_x(\mathbb{1}_{\sigma_y < +\infty} \mathbb{E}_{X_{\sigma_y}}(N_y)) =$$

$$\mathbb{E}_x(\mathbb{1}_{\sigma_y < +\infty} \mathbb{E}_y(N_y)) = \mathbb{P}_x(\sigma_y < +\infty)U(y, y),$$

par (19.9.1). \square

Remarque 19.10.1. *Si E est fini, il existe au moins un état récurrent. En effet, $\sum_{y \in E} N_y = N_E = +\infty$ p.s., d'où*

$$\sum_{y \in E} U(x, y) \stackrel{(19.9.1)}{=} \mathbb{E}_x\left(\sum_{y \in E} N_y\right) = \mathbb{E}_x(N_E) = +\infty.$$

Il existe donc $y \in E$ tel que $U(x, y) = +\infty$. Par la proposition 19.10.1, $U(y, y) \geq U(x, y) = +\infty$ et donc y est récurrent.

19.10.1 Communications entre deux états

On note $\tau_x = \inf\{n \geq 0, X_n = x\}$. (Attention : la chaîne ne part pas nécessairement de x . Si $X_0 = x$, on a appelé σ_x le temps de premier retour, qui est donc distinct de $\tau_x = 0$. Si la chaîne ne part pas de x , on a $\sigma_x = \tau_x$.) On dit que x conduit à y et on note $x \rightarrow y$ si $\mathbb{P}_x(\tau_y < +\infty) > 0$. Donc $x \rightarrow x$ et pour $y \neq x$, on a $\mathbb{P}_x(\tau_y < +\infty) = \mathbb{P}_x(\sigma_y < +\infty)$. Puisque $\{\sigma_y < +\infty\} = \cup_{n \geq 1} \{X_n = y\}$, les trois conditions suivantes sont équivalentes :

- (a) $\mathbb{P}(\sigma_y < +\infty) > 0$.
- (b) Il existe $n \geq 1$ tel que $P^n(x, y) > 0$
- (c) $U(x, y) (= \mathbb{E}_x(N_y)) > 0$.

Lemme 19.10.1. *Si $x \rightarrow y$ et $y \rightarrow z$, on a $x \rightarrow z$*

Démonstration. En effet, comme il y a moins de chemins allant de x à z en passant par y que de chemins allant de x à z (par “moins”, on entend une inclusion entre deux ensembles de chemins),

$$\begin{aligned}\mathbb{P}_x(\tau_z < +\infty) &\geq \mathbb{P}_x(\tau_y + \tau_z \circ \theta_{\tau_y} < +\infty) = \\ \mathbb{E}_x(\mathbb{1}_{\tau_y < +\infty} \mathbb{1}_{\tau_z < +\infty} \circ \theta_{\tau_y}) &= \mathbb{E}_x(\mathbb{1}_{\tau_y < +\infty} \mathbb{P}_{X_{\tau_y}}(\tau_z < +\infty)) = \\ \mathbb{P}_x(\tau_y < +\infty) \mathbb{P}_y(\tau_z < +\infty) &> 0.\end{aligned}$$

□

Proposition 19.10.2. *Si x est récurrent et $x \rightarrow y$, alors*

- (i) $U(y, y) = +\infty$ et y est récurrent
- (ii) $y \rightarrow x$
- (iii) $\mathbb{P}_x(N_y = +\infty) = 1$ et, par symétrie, $\mathbb{P}_y(N_x = +\infty) = 1$.

Démonstration. (i) Par l'équivalence des trois propriétés (a), (b) et (c), il existe n tel que $P^n(x, y) > 0$. Donc en utilisant la proposition 19.10.1,

$$U(y, y) \geq U(x, y) = \sum_{k \geq 0} P^k(x, y) \geq \sum_{k \geq 0} P^{n+k}(x, y) \geq \sum_{k \geq 0} P^k(x, x) P^n(x, y) = U(x, x) P^n(x, y) = +\infty.$$

Donc $U(y, y) = +\infty$ et y est récurrent.

(ii) Puisque x est récurrent, la chaîne visite x \mathbb{P}_x -presque sûrement. Donc, sur $\{\sigma_y < +\infty\}$, on a $\sigma_x \circ \theta_{\sigma_y} < +\infty$ \mathbb{P}_x -p.s. et vu que $x \rightarrow y$,

$$0 < \mathbb{P}_x(\sigma_y < +\infty) = \mathbb{P}_x(\sigma_y < +\infty, \sigma_x \circ \theta_{\sigma_y} < +\infty) = \mathbb{P}_x(\sigma_y < +\infty) \mathbb{P}_y(\sigma_x < +\infty).$$

D'où $\mathbb{P}_y(\sigma_x < +\infty) = 1$ et $y \rightarrow x$. Par symétrie, $\mathbb{P}_x(\sigma_y < +\infty) = 1$.

(iii) Plus généralement,

$$\mathbb{P}_x(\sigma_y^n < +\infty) = \mathbb{P}_x(\sigma_y^{n-1} < +\infty, \sigma_y \circ \theta_{\sigma_y^{n-1}} < +\infty) =$$

$$\mathbb{P}_x(\sigma_y^{n-1} < +\infty) \mathbb{P}_y(\sigma_y < +\infty) = \mathbb{P}_x(\sigma_y^{n-1} < +\infty) = \dots = \mathbb{P}_x(\sigma_y < +\infty) = 1.$$

D'où $\mathbb{P}_x(N_y = +\infty) = 1$ et par symétrie $\mathbb{P}_y(N_x = +\infty) = 1$.

□

Corollaire 19.10.1. *Il existe une partition $(C_i)_{i \in I}$ des états récurrents de E telle que*

- (i) si $x, y \in C_i$, $U(x, y) = +\infty$ et $\mathbb{P}_x(N_y = +\infty) = 1$,
- (ii) si $x \in C_i$ et $y \in C_j$, $i \neq j$, $U(x, y) = 0$ et $\mathbb{P}_x(N_y = 0) = 1$.

Démonstration. On dit que x et y communiquent si $x \rightarrow y$ et $y \rightarrow x$, ce qu'on écrit $x \leftrightarrow y$, une relation d'équivalence. Soit C une classe d'équivalence pour cette relation. Si cette classe C contient un état récurrent, ils le sont tous par la proposition 19.10.2 et ne conduisent à aucun état qui ne soit dans C_i .

□

19.11 Lien entre martingale et chaîne de Markov

Définition 19.11.1. Soit P une matrice de transition. Une fonction $f \in \mathcal{E}^+$ est dite excessive si $f \geq Pf$, et invariante si $f = Pf$. Une mesure μ est dite excessive si $\mu P \leq \mu$, invariante si $\mu P = \mu$. (Par exemple, $f = 1$ est invariante).

Proposition 19.11.1. Si $f \in \mathcal{E}^+$ est une fonction excessive (resp. invariante), alors $Y_n = f(X_n)$ est une surmartingale (resp. une martingale) positive pour toute loi \mathbb{P}_μ associée à une distribution initiale μ .

Démonstration. Par la propriété de Markov, on a presque sûrement

$$\mathbb{E}_\mu^{\mathcal{F}_n}(Y_{n+1}) = \mathbb{E}_\mu^{\mathcal{F}_n}(f(X_1) \circ \theta_n) = \mathbb{E}_{X_n}(f(X_1)) = Pf(X_n) \leq f(X_n) = Y_n.$$

(Rappel, $\mathbb{E}_x(f(X_k)) = (P^k f)(x)$, formule (4.13) : la loi de X_k est $\delta_x.P^k$.) □

19.12 Chaînes irréductibles récurrentes

Définition 19.12.1. Soit P une matrice de transition sur E . Une partie F de E est dite close si pour tout $x \in F$, on a $\mathbb{P}_x(X_n \in F \text{ pour tout } n \geq 0) = 1$. Si $F = \{x\}$, l'état x est dit absorbant. Si E est la seule partie close, la chaîne (ou sa matrice de transition) est dite irréductible.

ATTENTION : LA PREUVE DE LA PROPOSITION QUI SUIT EST TRÈS SUCCINCTE DANS [1].

Proposition 19.12.1. On a les propriétés suivantes

1. F est close si et seulement si pour tout $x \in F$, $P(x, F) = 1$ ou si et seulement si, pour tout $x \in F$ et tout $y \notin F$, $U(x, y) = 0$.
2. P est irréductible si et seulement si $\forall x, y \in E$, $U(x, y) > 0$.
3. Dans une chaîne irréductible, ou bien tous les états sont récurrents (alors $U(x, y) = +\infty$ pour tous $x, y \in E$), ou bien tous les états sont transitoires (alors $0 < U(x, y) < +\infty$). On parle de chaîne récurrente dans le premier cas et de chaîne transitoire dans le second.
4. Une chaîne irréductible sur un ensemble d'états finis est récurrente.
5. Si X est irréductible et si X possède une probabilité invariante, alors X est récurrente.

Démonstration. On rappelle (19.9.1) : $U(x, y) = \mathbb{E}_x(N_y) = \sum_k \mathbb{P}_x(X_k = y)$ en conséquence, $U(x, y) > 0$ si et seulement si x communique avec y , i.e. si $\mathbb{P}_x(\sigma_y < +\infty) > 0$.

1. ($\forall x \in F$, $P(x, F) = 1$) équivaut à $(P\mathbb{1}_F \geq \mathbb{1}_F)$ et donc à $(\forall n, P^n\mathbb{1}_F \geq \mathbb{1}_F)$, soit à $(P^n(x, F) = 1 \text{ pour } x \in F)$, soit encore à $(\mathbb{P}_x(X_n \in F) = 1 \text{ pour tout } x \in F)$ et donc à " F est close".

Si F est close, on a $P^n(x, F) = 1$ pour tout $x \in F$. Donc $P^n(x, y) = 0$ si $y \notin F$ et donc $U(x, y) = \sum_n P^n(x, y) = 0$. Réciproquement, si $U(x, y) = 0$ pour $x \in F$ et $y \notin F$, on déduit que $P^n(x, y) = 0$ pour tout $x \in F$, $y \notin F$. Donc $\forall x \in F$, $P^n(x, F) = 1$ et donc F est close.

2. Si $\forall x, y$, $U(x, y) > 0$: s'il existait une partie close $F \subset E$ et $x \in F$, $y \in F^c$: comme $U(x, y) > 0$, par la définition de U il existe n tel que $\mathbb{P}_x(X_n = y) > 0$, et donc $\mathbb{P}_x(X_n \in F) < 1$, ce qui contredit la clôture de F .

Réciproquement, supposons que E soit irréductible et montrons que $U(x, y) > 0$ pour tous $x, y \in E$. Remarquons pour commencer que $U(x, y) > 0$ et $U(y, z) > 0$ impliquent qu'il existe n et p tels que $P^n(x, y) > 0$ et $P^p(y, z) > 0$ et donc $P^{n+p}(x, z) > 0$, ce qui donne $U(x, z) > 0$. Posons $E_x = \{y, U(x, y) > 0\}$. Alors si pour un $y \in E_x$ on a $U(y, z) > 0$, la remarque précédente entraîne que $z \in E_x$. Donc $U(y, z) = 0$ pour tous $y \in E_x$ et $z \notin E_x$, ce qui veut dire par le 1. que E_x est close. Comme E est irréductible on déduit que $E = E_x$, et donc on a $U(x, y) > 0$ pour tous $x, y \in E$.

3. Dans une chaîne irréductible, tous les états communiquent. Si un état est récurrent, alors tous le sont car on a montré que $x \rightarrow y$ et x récurrent impliquent y est récurrent. Donc l'alternative est : tous les états récurrents ou tous les états transitoires. On rappelle la proposition 19.10.1 : $U(x, y) = \mathbb{P}_x(\sigma_y < +\infty)U(y, y)$. On rappelle de plus que y est récurrent si et seulement si $U(y, y) = +\infty$. Donc si un état y est récurrent, on a $U(x, y) = +\infty$ pour tous $x, y \in F$. Si tous les états sont transitoires, on a $U(x, x) < +\infty$ pour tout $x \in F$ et donc $U(x, y) < +\infty$ pour tous $x, y \in F$.
4. Application : si la chaîne irréductible est à états finis, il y a au moins un état récurrent (remarque 19.10.1) et donc la chaîne est récurrente.
5. Si λ est une probabilité invariante, on a

$$\lambda(y) = (\lambda P^n)(y) = \sum_{x \in E} \lambda(x) P^n(x, y).$$

Si la chaîne était transiente, on aurait $U(x, y) < +\infty$ pour tous $x, y \in E$ et donc $P^n(x, y) \rightarrow 0$ quand $n \rightarrow \infty$. Comme $P^n(x, y) \leq 1$, on peut appliquer le théorème de Lebesgue dans la relation précédente et on obtiendrait $\lambda(y) = 0$, ce qui est absurde. □

Proposition 19.12.2. *Si X est une chaîne de Markov canonique irréductible et récurrente et si f est une fonction excessive, alors f est constante.*

Démonstration. $f(X_n)$, $n \geq 0$ est une surmartingale positive (Proposition 4) et donc converge p.s. pour toute loi initiale μ vers une variable aléatoire Z . Soit $x \neq y$. On a (p.s.) $X_n = x$ pour une infinité de x et $X_n = y$ pour une infinité de y . Donc $Z = f(x) = f(y)$. □

Proposition 19.12.3. *Soit P une chaîne canonique et μ une mesure invariante telle que $0 < \mu(x) < +\infty$ pour tout $x \in E$. On définit la "chaîne duale" \hat{P} par*

$$\hat{P}(x, y) = \frac{\mu(y)}{\mu(x)} P(y, x).$$

Alors \hat{P} est bien une matrice de transition. Si P est récurrente irréductible, alors la chaîne duale l'est aussi.

Démonstration.

$$\sum_{y \in E} \hat{P}(x, y) = \frac{1}{\mu(x)} \sum_{y \in E} \mu(y) P(y, x) = \frac{1}{\mu(x)} (\mu P)(x) = \frac{1}{\mu(x)} \mu(x) = 1.$$

Donc, \hat{P} est une matrice de transition.

$$\hat{P}^2(x, y) = \sum_z \hat{P}(x, z) \hat{P}(z, y) = \sum_z \frac{\mu(z)}{\mu(x)} P(z, x) \frac{\mu(y)}{\mu(z)} P(y, z) = \frac{\mu(y)}{\mu(x)} P^2(y, x).$$

De même, on a $\hat{P}^n(x, y) = \frac{\mu(y)}{\mu(x)} P^n(y, x)$ et donc

$$\hat{U}(x, y) = \frac{\mu(y)}{\mu(x)} U(y, x).$$

Donc $\hat{U}(x, y) = +\infty$ et donc la chaîne duale est aussi récurrente irréductible. \square

Théorème 19.12.1. *Soit X une chaîne irréductible récurrente. Alors X possède une mesure invariante μ unique à une constante multiplicative près. De plus, pour tout $y \in E$, $0 < \mu(y) < +\infty$. Enfin, pour $x \in E$, la mesure invariante μ_x telle que $\mu_x(x) = 1$ est donnée par*

$$\mu_x(y) = \mathbb{E}_x \left(\sum_{k=0}^{\sigma_x-1} \mathbb{1}_{\{y\}}(X_k) \right).$$

Démonstration. On fixe $x \in E$ et on pose

$$\mu'_x(y) = \mathbb{E}_x \left(\sum_{k=1}^{\sigma_x} \mathbb{1}_{\{y\}}(X_k) \right).$$

$\mu'_x(y)$ est donc le nombre moyen de passages par y entre $k = 1$ et le retour en x , alors que $\mu_x(y)$ est le nombre moyen de passages par y entre $k = 0$ et le retour en x , non compris. Si $y \neq x$, on en déduit immédiatement que $\mu_x(y) = \mu'_x(y)$. On a aussi $\mu_x(x) = \mu'_x(x) = 1$ et donc $\mu_x = \mu'_x$. De plus,

$$\mu_x(E) = \mu_x \cdot \mathbb{1}_E = \mathbb{E}_x(\sigma_x). \quad (19.12.1)$$

(nombre moyen de passages avant le retour à x).

(i) On montre que μ_x est invariante. On commence par calculer, pour $f \in \mathcal{E}^+$, $\mu_x \cdot f$:

$$\mu_x \cdot f = \sum_{y \in E} \mathbb{E}_x \left(\sum_{k=0}^{\sigma_x-1} \mathbb{1}_{\{y\}}(X_k) \right) f(y) = \mathbb{E}_x \left(\sum_{k=0}^{\sigma_x-1} \sum_y \mathbb{1}_{\{y\}}(X_k) f(y) \right) = \mathbb{E}_x \left(\sum_{k=0}^{\sigma_x-1} f(X_k) \right).$$

On obtient la formule utile

$$\mu_x \cdot f = \mathbb{E}_x \left(\sum_{k=0}^{\sigma_x-1} f(X_k) \right) \quad (19.12.2)$$

On en déduit que

$$\mu_x P \cdot f = \mu_x \cdot P f = \mathbb{E}_x \left(\sum_{k=0}^{\sigma_x-1} P f(X_k) \right) = \mathbb{E}_x \left(\sum_{k=0}^{\sigma_x-1} \mathbb{E}_{X_k}(f(X_1)) \right) = \sum_{k \geq 0} \mathbb{E}_x(\mathbb{1}_{\{k < \sigma_x\}} \mathbb{E}_{X_k}(f(X_1)))$$

(En effet, la relation (4.13), $\mathbb{E}_\mu(f(X_n)) = \mu.P^n f$, donne $Pf(X_k) = \delta_{X_k}.Pf = \mathbb{E}_{X_k}(f(X_1))$.) Vu que $\{k < \sigma_x\} \in \mathcal{F}_k$, en appliquant (en sens inverse pour une fois) la propriété de Markov (4.18), on obtient

$$\mu_x P.f = \sum_{k \geq 0} \mathbb{E}_x(\mathbb{1}_{k < \sigma_x} f(X_{k+1})) = \mathbb{E}_x\left(\sum_{k=1}^{\sigma_x} f(X_k)\right) = \mu'_x.f = \mu_x.f.$$

(ii) On montre que pour tout $y \in E$, on a $0 < \mu_x(y) < +\infty$.

Lemme 19.12.1. *Soit P irréductible et λ une mesure telle que $\lambda \geq \lambda.P$. Si, pour un $x \in E$, $\lambda(x) = 0$ (resp. $\lambda(x) < +\infty$), alors pour tout y , $\lambda(y) = 0$ (resp. $\lambda(y) < +\infty$).*

En effet, soit $x \neq y$. Il existe $n \geq 1$ tel que $P^n(y, x) > 0$. Comme $\lambda \geq \lambda P^n$, on a

$$\lambda(x) \geq \sum_{z \in E} \lambda(z) P^n(z, x) \geq \lambda(y) P^n(y, x).$$

(iii) On montre l'unicité de la mesure invariante μ_x . Soit λ une mesure excessive (en particulier invariante) avec $\lambda(y) < +\infty$ pour tout $y \in E$. On pose $f(y) = \frac{\lambda(y)}{\mu_x(y)}$. Alors f est une fonction excessive pour la chaîne duale \hat{P} , avec $\mu = \mu_x$. En effet,

$$\hat{P}f(y) = \sum_{z \in E} \hat{P}(y, z) \frac{\lambda(z)}{\mu_x(z)} = \sum_{z \in E} \frac{\mu_x(z)}{\mu_x(y)} P(z, y) \frac{\lambda(z)}{\mu_x(z)} = \frac{1}{\mu_x(y)} \lambda P(z) \leq \frac{1}{\mu_x(y)} \lambda(y) = f(y).$$

Par la proposition 19.12.2, cela implique que f est constante. □

Théorème 19.12.2. *Soit X une chaîne irréductible récurrente de mesure invariante μ . Il y a deux cas possibles,*

(i) $\mu(E) = +\infty$. Alors, pour tout $x \in E$, $\mathbb{E}_x(\sigma_x) = +\infty$. La chaîne est dite récurrente nulle.

(ii) $\mu(E) < +\infty$. Alors, pour tout $x \in E$, $\mathbb{E}_x(\sigma_x) < +\infty$. La chaîne est dite récurrente positive.

Dans ce cas l'unique probabilité invariante est donnée par

$$\pi(x) = \frac{1}{E_x(\sigma_x)}.$$

Démonstration. Soit on a $\mu(E) = +\infty$ pour toute mesure invariante μ , soit $\mu(E) < \infty$. Mais par (19.12.1), $\mu(E) = \mathbb{E}_x(\sigma_x)$ et donc $\mathbb{E}_x(\sigma_x) = +\infty$ pour tout x ou bien $\mathbb{E}_x(\sigma_x) < \infty$ pour tout x . Dans le second cas, il y a une unique mesure invariante $\pi(y) = \frac{\mu_x(y)}{\mu_x(E)} = \frac{\mu_x(y)}{\mathbb{E}_x(\sigma_x)}$ et en particulier $\pi(x) = \frac{\mu_x(x)}{\mathbb{E}_x(\sigma_x)} = \frac{1}{\mathbb{E}_x(\sigma_x)}$. □

Remarque 19.12.1. *Un exemple simple et générique de chaîne récurrente nulle. Si E est infini et si P est une matrice doublement stochastique, alors $(1, \dots, 1, \dots)$ est une mesure invariante (pourquoi?). On en déduit alors que la chaîne est récurrente nulle. Prenons comme exemple une marche aléatoire symétrique sur \mathbb{Z}^N . La matrice de la chaîne de Markov associée est bistochastique et a donc $(\dots, 1, 1, \dots, 1, \dots)$ comme mesure invariante. Or, on sait qu'une telle chaîne est récurrente si et seulement si $N \leq 2$. Si $N \geq 3$, elle est transiente, bien qu'ayant une mesure invariante.*

19.13 Théorèmes ergodiques

Théorème 19.13.1. Soit X une chaîne irréductible récurrente de mesure invariante μ . Soient $f, g \in L^1(\mu)$ avec $\mu.g \neq 0$. Alors

$$\frac{\sum_0^n f(X_k)}{\sum_0^n g(X_k)} \rightarrow \frac{\mu.f}{\mu.g}, \text{ p.s.}$$

Démonstration du théorème ergodique 19.13.1 On fixe $x \in E$ et on considère $\mu = \mu_x$ la mesure invariante valant 1 en x et $\sigma_x^p = \sigma^p$ le p -ième retour en x . Soit λ une loi initiale arbitraire. Soit $f \in L^1(\mu)$, i.e. vérifiant $\sum_{x \in E} |f(x)|\mu(x) < +\infty$, on pose

$$Z_0 = \sum_{k=0}^{\sigma^1-1} f(X_k)$$

$$Z_p = Z_0 \circ \theta_{\sigma^p} = \sum_0^{\sigma^1-1} f(X_k) \circ \theta_{\sigma^p} = \sum_{k=\sigma^p}^{\sigma^{p+1}-1} f(X_k).$$

Par la relation (19.12.2), $\mathbb{E}_x(Z_0) = \mu.f$ et par la propriété de Markov,

$$\mathbb{E}_\lambda(Z_p) = \mathbb{E}_\lambda(\mathbb{E}_{X_{\sigma^p}}(Z_0)) = \mathbb{E}_x(Z_0) = \mu.f.$$

Pour la loi \mathbb{P}_λ , les v.a. Z_1, \dots, Z_p, \dots sont indépendantes, de même loi, d'espérance $\mu.f$: voir le lemme 19.13.1 ci-dessous. Par la loi forte des grands nombres,

$$\frac{1}{n} \sum_0^{\sigma^n} f(X_k) = \frac{Z_0}{n} + \frac{1}{n}(Z_1 + \dots + Z_{n-1}) + \frac{f(x)}{n} \rightarrow \mu.f, \mathbb{P}_\lambda - \text{p.s.}$$

A $m \in \mathbb{N}$ on associe l'unique entier $\nu(m)$ tel que $\sigma^{\nu(m)} \leq m \leq \sigma^{\nu(m)+1}$, alors $\nu(m) \rightarrow +\infty$ et

$$\nu(m) \leq \sum_0^m \mathbb{1}_x(X_k) = \nu(m) + \mathbb{1}_x(X_0) \leq \nu(m) + 1.$$

Si on suppose $f \geq 0$,

$$\frac{\nu(m)}{\nu(m)+1} \frac{\sum_0^{\sigma^{\nu(m)}} f(X_k)}{\nu(m)} \leq \frac{\sum_0^m f(X_k)}{\sum_0^m \mathbb{1}_x(X_k)} \leq \frac{\nu(m)+1}{\nu(m)} \frac{\sum_0^{\sigma^{\nu(m)+1}} f(X_k)}{\nu(m)+1}$$

et, les termes latéraux tendant vers $\mu.f$ \mathbb{P}_λ -p.s., on a donc

$$\frac{\sum_0^m f(X_k)}{\sum_0^m \mathbb{1}_x(X_k)} \rightarrow \mu.f, \mathbb{P}_\lambda - \text{p.s.}$$

En écrivant $f = f^+ - f^-$, on a obtenu la relation précédente pour $f \in L^1(\mu)$. Si de plus $\mu.g \neq 0$, on peut écrire cette relation pour f et g et faire le quotient, ce qui prouve le théorème.

ATTENTION : LA PREUVE DU LEMME 19.13.1 EST LAISSEE EN EXERCICE DANS [1].

Lemme 19.13.1. *Pour la loi \mathbb{P}_λ , les v.a. Z_1, \dots, Z_p, \dots sont indépendantes, de même loi, d'espérance $\mu.f$.*

Démonstration. On a $Z_p = Z_0 \circ \theta_{\sigma^p}$. Par la propriété forte de Markov on en déduit pour toute g , fonction réelle positive ou bornée et tout $p \geq 1$:

$$\mathbb{E}_\lambda(g(Z_p)) = \mathbb{E}_\lambda(g \circ Z_0 \circ \theta_{\sigma^p}) = \mathbb{E}_\lambda(\mathbb{E}_{x_{\sigma^p}}(g(Z_0))) = \mathbb{E}_\lambda(\mathbb{E}_x(g(Z_0))) = \mathbb{E}_x(g(Z_0)).$$

(Remarquer que $\sigma^p < +\infty$ p.s. par hypothèse). Donc les Z_p , $p \geq 1$ ont toutes la même loi. Montrons ensuite que Z_{k+p} et Z_p sont indépendantes pour $p, k \geq 1$. Pour tous $s, t \in \mathbb{R}$ on a, en utilisant à la troisième ligne la propriété de Markov forte et à la quatrième $X_{\sigma^k} = x$:

$$\begin{aligned} \mathbb{P}_\lambda(Z_{k+p} \leq s, Z_k \leq t) &= \mathbb{E}_\lambda(\mathbb{1}_{Z_{k+p} \leq s} \mathbb{1}_{Z_k \leq t}) \\ &= \mathbb{E}_\lambda(\mathbb{1}_{Z_k \leq t} \mathbb{1}_{\{Z_p \leq s\}} \circ \theta_{\sigma^k}) \\ &= \mathbb{E}_\lambda(\mathbb{1}_{Z_k \leq t} \mathbb{E}_{X_{\sigma^k}}(\mathbb{1}_{Z_p \leq s})) \\ &= \mathbb{P}_\lambda(Z_k \leq t) \mathbb{P}_x(Z_p \leq s). \end{aligned}$$

Pour déduire que Z_{k+p} et Z_p sont indépendantes, il suffit donc de vérifier que $\mathbb{P}_x(Z_p \leq s) = \mathbb{P}_\lambda(Z_p \leq s)$. Mais, en utilisant de nouveau la propriété de Markov forte,

$$\mathbb{P}_\lambda(Z_p \leq s) = \mathbb{E}_\lambda(\mathbb{1}_{Z_p \leq s}) = \mathbb{E}_\lambda(\mathbb{1}_{Z_0 \leq s} \circ \theta_{\sigma^p}) = \mathbb{E}_\lambda(\mathbb{E}_{X_{\sigma^p}}(\mathbb{1}_{Z_0 \leq s})) = \mathbb{P}_x(Z_0 \leq s).$$

Donc

$$\mathbb{P}_x(Z_p \leq s) = \mathbb{P}_x(Z_0 \leq s) = \mathbb{P}_\lambda(Z_p \leq s).$$

□

19.14 Conséquences du théorème ergodique

Proposition 19.14.1.

$$\frac{\text{nombre de visites de } X \text{ en } y \text{ avant } n}{\text{nombre de visites de } X \text{ en } x \text{ avant } n} \rightarrow \frac{\mu(y)}{\mu(x)}, \quad p.s.$$

Démonstration. Prendre $f = \mathbb{1}_{\{y\}}$ et $g = \mathbb{1}_{\{x\}}$. □

Proposition 19.14.2. *Soit X une chaîne irréductible, récurrente positive, de probabilité invariante π . Alors, pour toute $f \in L^1(\pi)$,*

$$\frac{1}{n} \sum_0^{n-1} f(X_k) \rightarrow \pi.f, \quad p.s.$$

Démonstration. Prendre $g = 1$ et $\mu = \pi$, la probabilité invariante (on a $\pi.g = 1$). □

Proposition 19.14.3. *(i) Si X est une chaîne irréductible, récurrente positive, de probabilité invariante π , alors, pour tout $x \in E$,*

$$\frac{1}{n} \sum_0^{n-1} \mathbb{1}_x(X_k) \rightarrow \pi(x) \quad p.s.$$

$\pi(x)$ est donc la fréquence de passage de la chaîne par x .

(ii) Si X est une chaîne irréductible récurrente nulle, alors, pour tout $x \in E$,

$$\frac{1}{n} \sum_0^{n-1} \mathbb{1}_x(X_k) \rightarrow 0, \text{ p.s.}$$

Démonstration. (i) découle de la proposition antérieure. Montrons (ii). Par hypothèse (théorème 19.12.2) on a $\mu(E) = +\infty$. Pour $F \subset E$ fini, on a par le théorème 19.13.1

$$\frac{\sum_0^n \mathbb{1}_x(X_k)}{n+1} \leq \frac{\sum_0^n \mathbb{1}_x(X_k)}{\sum_0^n \mathbb{1}_F(X_k)} \rightarrow \frac{\mu(x)}{\mu(F)} = \frac{1}{\mu(F)} \text{ p.s.}$$

pour tout F fini, arbitrairement grand. Donc

$$\limsup_n \frac{\sum_0^n \mathbb{1}_x(X_k)}{n+1} \leq \frac{1}{\mu(F)} \text{ p.s.}$$

Comme $\mu(E) = +\infty$, on peut prendre $\mu(F)$ arbitrairement grand. □

Corollaire 19.14.1. (i) Si X est récurrente positive de probabilité invariante π ,

$$\frac{1}{n} \sum_{k=0}^{n-1} P^k(y, x) \rightarrow_{n \rightarrow \infty} \pi(x),$$

(ii) si X est récurrente nulle,

$$\frac{1}{n} \sum_{k=0}^{n-1} P^k(y, x) \rightarrow_{n \rightarrow \infty} 0.$$

Démonstration. Appliquer \mathbb{E}_y aux formules de la proposition 19.14.3. Comme le terme de gauche de ces formules est borné par 1, on peut appliquer le théorème de Lebesgue, i.e. intervertir \mathbb{E}_y et la sommation en k . Mais $\mathbb{E}_y(\mathbb{1}_x(X_k)) = P^k(y, x)$ (formule (19.6.4)) ce qui donne le résultat. □

19.15 Périodicité

L'exemple suivant montre que la convergence en moyenne de Cesaro de $P^n(y, x)$ vers $\pi(x)$ n'implique pas une convergence ordinaire. Il suffit de prendre

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

On vérifie aisément que $P^{3n} = I$, $P^{3n+1} = P$, $P^{3n+2} = P^2$ et que la suite $P^n(1, 2)$ vaut 0, 1, 0, 0, 1, 0, 0, 1, ... est périodique et ne converge pas.

Définition 19.15.1. Soit P une matrice de transition. Pour $x \in E$, on pose

$$I(x) = \{n > 0, P^n(x, x) > 0\}$$

Le p.g.c.d. de l'ensemble $I(x)$ s'appelle la période de x et se note $a(x)$.

Proposition 19.15.1. *Si X est une chaîne irréductible, tous les états ont même période. La chaîne est dite apériodique si cette période est 1.*

Démonstration. Soient x et y deux états de périodes $a(x)$ et $a(y)$ respectivement. Comme x communique avec y , il existe k tel que $P^k(x, y) > 0$ et de même il existe l tel que $P^l(y, x) > 0$. Donc $k+l \in I(x)$ et de même $l+k \in I(y)$. On déduit que $a(y)$ divise $k+l$ et aussi que $l+a(x)+k \in I(y)$. Donc $a(y)$ divise aussi $a(x)$. Réciproquement, $a(x)$ divise $a(y)$ et donc les deux périodes sont égales. \square

Théorème 19.15.1. *Soit X une chaîne irréductible, apériodique, récurrente positive de probabilité invariante π . Alors $P^n(x, y) \rightarrow \pi(y)$ pour tous $x, y \in E$.*

Théorème 19.15.2. *Soit X une chaîne irréductible, récurrente positive de probabilité invariante π , de période d . Alors pour tout $x, y \in E$, il existe $r, 0 \leq r < d$, avec $r = 0$ si $x = y$, tel que, pour tout $s, 0 \leq s < d$,*

$$\lim_{n \rightarrow \infty} P^{nd+r}(x, y) = d\pi(y), \quad P^{nd+s}(x, y) = 0 \text{ pour } s \neq r.$$

Théorème 19.15.3. *Soit X une chaîne irréductible, récurrente nulle. Alors on a $P^n(x, y) \rightarrow 0$ quand $n \rightarrow \infty$, pour tout $x, y \in E$.*

Ce résultat est vrai que la chaîne soit périodique ou non.

19.16 Chaînes de Markov à espaces d'états finis ([8] pp. 195-199)

Quand on considère une chaîne de Markov à espace d'états finis, l'étude du comportement asymptotique de la chaîne se ramène à l'étude de la convergence de la suite $(P^n)_n$ des itérés de la matrice de transition P . P est une matrice stochastique, c'est-à-dire à coefficients positifs avec la somme des éléments de chaque ligne valant 1. Ces matrices ont des propriétés spectrales très particulières que nous allons décrire maintenant.

Définition 19.16.1. *Une matrice $A = \{a_{i,j}\}_{1 \leq i,j \leq r}$ à coefficients réels est dite positive (resp. strictement positive) ssi ses coefficients sont positifs (resp. strictement positifs). Une matrice positive est dite stochastique ssi $\sum_{j=1}^r a_{i,j} = 1$ pour tout i et sous-stochastique si $\sum_{j=1}^r a_{i,j} \leq 1$ pour tout i avec une inégalité stricte pour au moins un i .*

On notera $A \geq 0$ (resp. $A > 0$) pour signifier que A est positive (resp. strictement positive).

Définition 19.16.2. *Le graphe de communication d'une matrice positive A est le graphe orienté avec pour espace d'états $E = \{1, \dots, r\}$ et pour ensemble d'arêtes les couples (i, j) tels que $a_{i,j} > 0$.*

Définition 19.16.3. *Une matrice positive A est dite irréductible (resp. irréductible aperiodique) ssi elle a meme graphe de communication qu'une matrice stochastique irréductible (resp. aperiodique irréductible). Elle est dite primitive s'il existe un entier k tel que $A^k > 0$.*

Proposition 19.16.1. *Une matrice est irréductible aperiodique ssi elle est primitive.*

On rappelle que si M est une matrice carrée et λ une valeur propre de M , la multiplicité algébrique de M est la multiplicité de λ dans le polynôme caractéristique de M et la multiplicité géométrique de λ est la dimension du sous-espace propre de M associé à λ .

Théorème 19.16.1. Théorème de Perron-Frobenius

Soit A une matrice primitive de taille r . Il existe une valeur propre réelle $\lambda_1 > 0$ de multiplicité algébrique et géométrique égales à 1. Les autres valeurs propres $\lambda_2, \dots, \lambda_r$ de A sont de module strictement inférieurs à λ_1 . D'autre part, le vecteur à gauche u_1 (vecteur ligne) et le vecteur à droite v_1 (vecteur colonne) associés à λ_1 peuvent être choisis strictement positifs et tels que $u_1 v_1 = 1$.

On suppose que : $\lambda_1 > |\lambda_2| \geq \dots \geq |\lambda_r|$

Si $|\lambda_2| = |\lambda_j|$ pour un $j \geq 3$ alors $m_2 \geq m_j$ où m_j est la multiplicité algébrique de λ_j . Aussi, on a :

$$A^n = \lambda_1^n v_1 u_1 + O(n^{m_2-1} |\lambda_2|^n)$$

Si de plus A est stochastique (resp. sous-stochastique) alors $\lambda_1 = 1$ (resp. $\lambda_1 < 1$).

Si A est stochastique mais non irréductible les multiplicités géométriques et algébriques de la valeur propre $\lambda_1 = 1$ sont égales au nombre de classes de communication.

Si la matrice A est stochastique, irréductible avec une période $d > 1$, il y a exactement d valeurs propres distinctes de module 1, les racines d -ièmes de l'unité, et les autres valeurs propres sont de module strictement inférieur à 1.

Aussi, si P est la matrice de transition d'une chaîne de Markov irréductible aperiodique sur $E = \{1, \dots, r\}$ (donc primitive) alors :

$$v_1 = 1, \quad u_1 = \pi$$

où π est l'unique probabilité stationnaire. Ainsi,

$$P^n = 1\pi + O(n^{m_2-1} |\lambda_2|^n)$$

On remarquera que la vitesse de convergence est géométrique. Pour des problèmes de simulation, il est important de disposer de majorants de la seconde plus grande valeur propre (SLEM). C'est l'objet d'une bonne partie du chapitre 6 de [8].

19.17 TD Probabilités : Chaînes de Markov I**Exercice 19.17.1. Marches aléatoires sur \mathbb{Z} : [1], exercice 4.16**

On considère une marche aléatoire $X_0 = 0, X_n = Z_1 + \dots + Z_n$ où les Z_i sont des variables aléatoires de loi de Bernoulli indépendantes à valeurs dans $\{-1, 1\}$ prenant la valeur 1 avec probabilité $p \in (0, 1)$.

- 1) Montrer que $(X_n)_n$ est une chaîne de Markov dont on précisera la matrice de transition.
- 2) Montrer que la chaîne est irréductible et de période 2 (ind. : on pourra montrer que si P est la matrice de transition de la chaîne, $P^2(i, i) \geq p(1-p)$)
- 3) On suppose $p \neq 1/2$. En utilisant la loi des grands nombres, montrer que la chaîne est transiente.
- 4) On suppose $p = 1/2$.

a) On pose $Y_n = \frac{1}{2}(Z_n + 1)$ et $T_n = \frac{1}{2}(X_n + n)$. Donner la loi de Y_n et de T_n . En déduire que si n est impair alors $P^n(0, 0) = 0$ et si n est pair alors $P^n(0, 0) = C_n^{n/2}(1/2)^n$.

b) En déduire que la chaîne est récurrente.

c) Trouver une mesure invariante pour la chaîne de masse infinie et en déduire que la chaîne est récurrente nulle.

Exercice 19.17.2. Processus de naissance et de mort discret : [26], p. 502-504

On considère une population dans laquelle à chaque instant n un individu peut disparaître ou apparaître avec une probabilité dépendant uniquement de la taille de la population à l'instant n . La taille de la population à l'instant n est notée X_n . On modélise cette population par une chaîne de Markov à valeurs dans \mathbb{N} dont la chaîne canonique est notée $(\Omega, \{\mathcal{F}_n\}_n, (X_n)_n, \mathbb{P}_x)$. La matrice de transition M est donnée pour $x \in \mathbb{N}$ par :

$$M(x, x + 1) = p_x, \quad M(x, x - 1) = q_x, \quad M(x, x) = r_x$$

avec $p_x, q_x, r_x \geq 0$, $p_x + q_x + r_x = 1$

Soit a, b deux entiers tels que $0 \leq a < b$. On notera $T_x = \inf\{n \geq 1; X_n = x\}$ le temps d'entrée en x (éventuellement infini).

1) Démontrer que pour tout x , on a

$$\mathbb{P}_x(T_a < T_b) = M(x, a) + \sum_{z \neq a, b} \mathbb{P}_z(T_a < T_b) M(x, z)$$

2) On suppose que $a = 0$ et que a et b sont des barrières, c'est-à-dire que $q_0 = p_b = 0$ et qu'elles sont élastiques dans la mesure où $r_a, r_b > 0$. On suppose que $p_x > 0$ pour tout $x \in (0, b)$. On pose enfin $f(x) = \mathbb{P}_x(T_0 < T_b)$.

a) En utilisant la question précédente, montrer que f est solution de l'équation du second ordre avec conditions aux limites (*problème de Dirichlet discret*) :

$$f(x) = p_x f(x + 1) + q_x f(x - 1) + r_x f(x) \quad \text{si } 1 < x < b$$

$$f(1) = p_1 f(2) + q_1 + r_1 f(1)$$

$$f(b - 1) = q_{b-1} f(b - 2) + r_{b-1} f(b - 1)$$

b) Montrer que si l'on pose $f(0) = 1, f(b) = 0$ alors

$$f(x + 1) - f(x) = \frac{q_x}{p_x} (f(x) - f(x - 1))$$

c) En posant $a_0 = 1$ et pour $x \in (0, b), a_x = \frac{q_1 \cdots q_x}{p_1 \cdots p_x}$, montrer que :

$$f(x) = \frac{\sum_{y=x}^{b-1} a_y}{\sum_{y=0}^{b-1} a_y}$$

3) En supposant que $p_x = p$ et $q_x = q$ soient constants pour $x \in (0, b)$, étudier le comportement de f selon les valeurs de p et q .

Exercice 19.17.3. [1], exercice 4.13 p. 112

Le but de cet exercice est de classer suivant les valeurs des nombres $p_k, q_k \geq 0$ les états de la chaîne de Markov sur \mathbb{N} de matrice de transition Q donnée par

$$\begin{aligned} Q(0, 0) &= \alpha, \quad Q(0, 1) = 1 - \alpha, \quad 0 < \alpha < 1 \\ Q(1, 2) &= \beta, \quad Q(1, 3) = 1 - \beta, \quad 0 < \beta < 1 \\ Q(k, 1) &= p_k, \quad Q(k, k+2) = q_k = 1 - p_k, \quad 0 < p_k < 1, \quad k \geq 2 \end{aligned}$$

- 1) Montrer que $F = \{1, 2, \dots\}$ est une classe close et que 0 est transient.
- 2) Montrer que tous les états de F communiquent entre eux et qu'ils sont donc ou tous transients, ou tous récurrents.
- 3) On note $\sigma_1 = \inf\{n \geq 1 | X_n = 1\}$.

a) Montrer que pour tout $n \geq 2$, on a

$$\mathbb{P}_1(\sigma_1 = n) = \beta q_2 q_4 \dots q_{2(n-2)} p_{2(n-1)} + (1 - \beta) q_3 q_5 \dots q_{2n-3} p_{2n-1}$$

et $\mathbb{P}_1(\sigma_1 = 1) = 0$.

b) En déduire que

$$\mathbb{P}_1(\sigma_1 < +\infty) = 1 - \beta \prod_{k=1}^{\infty} q_{2k} - (1 - \beta) \prod_{k=1}^{\infty} q_{2k+1}$$

et conclure.

Exercice 19.17.4. [1], exercice 4.8 p. 109

Pour modéliser l'évolution des configurations génétiques dans une population, on est amené à considérer la chaîne de Markov suivante. Soit P la matrice de transition sur $E = \{0, 1, \dots, N\}$ définie par

$$P(i, j) = C_N^j \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$$

- 1) Calculer $P(0, j)$ et $P(N, j)$. La chaîne est-elle irréductible? Quels sont les états récurrents?
- 2) Soit $X = (\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, (X_n)_{n \geq 0}, (\mathbb{P}_x)_{x \in E})$ la chaîne de Markov canonique associée. Montrer que pour tout $x \in E$, X est une \mathbb{P}_x -martingale et que la limite

$$\lim_{n \rightarrow \infty} X_n = X_\infty$$

existe \mathbb{P}_x -p.s.. Déterminer la loi de X_∞ .

Exercice 19.17.5. Lien entre martingale et chaîne de Markov, [1], exercice 4.5

Le lien qui existe entre martingales et chaînes de Markov est beaucoup plus important que celui mis en évidence ici. On pourra consulter le chapitre de [1] relatif à ce sujet ou encore *citeBr*. Soit E un ensemble dénombrable et $(X_n)_n$ une chaîne de Markov de matrice de transition P sur E .

On note \mathcal{H} l'espace vectoriel des fonctions bornées sur E et $\{\mathcal{F}_n\}_n$ la filtration naturelle associée. On définit l'opérateur A sur \mathcal{H} par $A = P - I$. Montrer que pour tout $f \in \mathcal{H}$,

$$f(X_n) - \sum_{k=0}^{n-1} Af(X_k)$$

est une $(\mathcal{F}_n)_n$ -martingale.

Chapitre 20

Chaînes de Markov et télécommunications

20.1 TD Probabilités : Un exemple de télécommunication

Exercice 20.1.1. Critère de Dynkin, [1], ex. 4.35

Il est en général faux que si X est une chaîne de Markov alors $\psi(X)$ est une chaîne de Markov. Le critère suivant donne une condition suffisante pour qu'il en soit ainsi.

Soit $((X_n)_n, (\mathcal{F}_n)_n, P)$ une chaîne de Markov sur E de matrice de transition P . Soit $\psi : E \rightarrow F$ une application surjective telle que, pour tout $j \in F$, on ait :

$$P(x, \psi^{-1}(j)) = P(y, \psi^{-1}(j)) \text{ si } \psi(x) = \psi(y)$$

On définit alors pour $i, j \in F$,

$$Q(i, j) = P(x, \psi^{-1}(j))$$

où x est un état quelconque de E tel que $\psi(x) = i$.

a) Montrer que Q est une matrice de transition et que si $Y_n = \psi(X_n)$ alors Y est une $(\mathcal{F}_n)_n$ chaîne de Markov.

b) Montrer que si π est une probabilité stationnaire pour P , la loi image $\pi \circ \psi^{-1}$ est une probabilité stationnaire pour Q .

Exercice 20.1.2. Produit de chaînes de Markov, [1], ex. 4.42

On considère deux chaînes de Markov canoniques

$$X = (\Omega, \mathcal{F}, \mathcal{F}_n, (X_n)_n, (\mathbb{P}_x)_x)$$

et

$$X' = (\Omega', \mathcal{F}', \mathcal{F}'_n, (X'_n)_n, (\mathbb{P}'_x)_x)$$

à valeurs dans E et E' de matrices de transitions respectives Q et Q' . On pose :

$$\bar{\Omega} = \Omega \times \Omega', \bar{\mathcal{F}} = \mathcal{F} \otimes \mathcal{F}', \bar{X} = (X, X'), \bar{\mathbb{P}}_{(x, x')} = \mathbb{P}_x \otimes \mathbb{P}'_{x'}$$

1) Montrer que $\bar{X} = X \otimes X' = (\bar{\Omega}, \bar{\mathcal{F}}, (\bar{X}_n)_n, \bar{\mathbb{P}}_{(x, x')})$ est une chaîne de Markov à valeurs dans $\bar{E} = E \times E'$ et de matrice de transition \bar{Q} définie par $\bar{Q}((x, x'), (y, y')) = Q(x, y)Q'(x', y')$.

2) Calculer \overline{Q}^n pour tout n .

3) Montrer que si ν (resp. ν') est une mesure invariante pour X (resp. X') alors $\nu \otimes \nu'$ en est une pour \overline{X} .

Remarque 20.1.1. *On pourra démontrer que si (a, a') est un état récurrent pour \overline{X} alors a est récurrent pour X et a' l'est pour X' mais que la réciproque est fautive (considérer des marches aléatoires simples sur \mathbb{Z} et les coupler habilement en mettant à profit le rôle de la dimension dans le caractère transient d'une marche aléatoire simple).*

Exercice 20.1.3. Un problème de télécommunications, [1], Problème 4.14

Un canal de transmission de données peut recevoir des messages à partir de m sources indépendantes. Chacune des sources peut être active (1) ou inactive (0). Si on les observe à des intervalles de temps fixés, chacune d'elles passe de l'état 0 à l'état 1 avec probabilité α et de 1 à 0 avec probabilité β et persiste dans les états 0 et 1 avec probabilités $1 - \alpha$ et $1 - \beta$ respectivement. On suppose que les comportements des sources sont indépendantes et $0 < \alpha, \beta < 1$.

On modélise le comportement de chaque source par une chaîne de Markov à espace d'états $\{0, 1\}$ et de matrice de transition

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

On note X la chaîne de Markov canonique sur $\{0, 1\}$ de matrice de transition P .

1) Montrer que l'unique probabilité stationnaire de cette chaîne est donnée par :

$$\mu_0 = \frac{\beta}{\alpha + \beta}, \quad \mu_1 = \frac{\alpha}{\alpha + \beta}$$

2) On considère l'ensemble $E = \{0, 1\}^m$ des vecteurs binaires de longueur m et sur E la chaîne produit $Y = \otimes^m X$.

a) Montrer que $\mu^{\otimes m}$ est une probabilité stationnaire pour Y .

b) Montrer que Y est irréductible et en déduire que Y ne possède qu'une seule probabilité stationnaire.

c) Montrer que X est aperiodique et en déduire qu'il en est de même de Y .

3) Y modélise les m sources indépendantes. Soit $Z = Y_1 + \dots + Y_m$ le nombre de sources actives.

a) Montrer que $(Y_n)_n$ converge en loi vers $\mu^{\otimes m}$.

b) En déduire que $(Z_n)_n$ converge en loi vers une loi binomiale de paramètre $(m, (\alpha + \beta)^{-1}\alpha)$.

c) En déduire que le nombre moyen de sources actives pour n grand est $m(\alpha + \beta)^{-1}\alpha$.

Remarque 20.1.2. *En utilisant le critère de Dynkin, on peut montrer que Y est une chaîne de Markov, ce qui se comprend facilement puisque les sources sont indépendantes (cf. [1]).*

L'exercice suivant utilise à nouveau le théorème de Foster vu dans le td sur le protocole de transmission Aloha.

Exercice 20.1.4. : Un modèle de file d'attente, problème 4.8, [1]

Soit $(Y_n, Z_n)_{n \geq 0}$ une suite de v.a. indépendantes telles que chacune des suites $(Y_n)_n$ et $(Z_n)_n$ soient composées de variables intégrables à valeurs dans \mathbb{N} intégrables. Z_n représente le nombre de clients servis dans le n -ième intervalle de temps alors que Y_n représente le nombre de clients qui arrivent dans la queue à la fin de l'intervalle de temps. On pose $\mathcal{F}_n = \sigma(Y_i, Z_i; i \leq n)$ et on suppose qu'il existe un entier $K > 1$ tel que :

$$\mathbb{P}(Z_1 > K) = 0, \mathbb{P}(Y_1 \geq K) > 0$$

et

$$\mathbb{P}(Z_1 = 1) > 0, \mathbb{P}(Y_1 = 0) > 0, \mathbb{E}(Y_1), \mathbb{E}(Z_1)$$

On pose enfin :

$$X_0 = i \in \mathbb{N}, X_{n+1} = (X_n - Z_{n+1})^+ + Y_{n+1}$$

1) Montrer que X est une $\{\mathcal{F}\}_n$ chaîne de Markov dont la matrice de transition est :

$$P(x, y) = \mathbb{P}(Y_1 + (x - Z_1)^+ = y)$$

2) Montrer que la condition de Foster est satisfaite pour $F = \{0, \dots, K\}$ et $h(x) = x$.

3) Montrer que X est irréductible et en déduire qu'elle est récurrente positive. Pour cela, on montrera que chaque état communique à son voisin de gauche et que si $k > K$ est tel que $\mathbb{P}(Y_1 = k) > 0$ alors x communique avec $x + k - 1$.

20.2 TD Probabilités : Système de transmission Aloha

Tout ce td est tiré de [8].

Exercice 20.2.1. Exercice 2.7.2 de [8] non corrigé

Soit $(X_n)_n$ une chaîne de Markov homogène canonique sur E , dénombrable. Soit F un sous-ensemble fini de l'espace des états E . Soit $\tau_1 = \inf\{n \geq 1 | X_n \in F\}$, $\tau_2 = \inf\{n \geq \tau_1 + 1 | X_n \in F\}$, ... le temps successifs de retour dans F . On suppose que pour tout $j \in F$, $\mathbb{P}_j(\tau_1 < +\infty) = 1$. Alors

1. Pour tout n , τ_n est fini presque sûrement.

2. Le processus $(Y_n)_n = (X_{\tau_n})_n$ est une chaîne de Markov homogène sur F .

Solution

On travaille avec la chaîne canonique et on remarque que :

$$\tau_n = \tau_{n-1} + \tau_1 \circ \theta_{\tau_{n-1}} \text{ si } \tau_{n-1} < +\infty \text{ et } +\infty \text{ sinon}$$

La propriété forte de Markov permet alors d'écrire, pour une fonction Φ bornée et pour $x \in F$:

$$\mathbb{E}_x [1_{\{\tau_n < +\infty\}} \Phi(X_{\tau_n}) | \mathcal{F}_{\tau_{n-1}}] = 1_{\{\tau_{n-1} < +\infty\}} \mathbb{E}_{X_{\tau_{n-1}}} (\Phi(X_{\tau_1}))$$

Appliquant ceci à $\Phi = 1$, on a $\mathbb{P}_x(\tau_n < +\infty) = \mathbb{P}_x(\tau_{n-1} < +\infty) = \dots = \mathbb{P}_x(\tau_1 < +\infty) = 1$ ce qui établit la première affirmation.

Pour la deuxième, il suffit de remarquer que l'on peut maintenant ôter les indicatrices et en posant $\mathcal{G}_n = \mathcal{F}_{\tau_n}$, on a :

$$\mathbb{E}_x(\Phi(Y_n)|\mathcal{G}_{n+1}) = \mathbb{E}_{Y_{n-1}}(\Phi(Y_1))$$

ce qui prouve que Y est bien une \mathcal{G}_n chaîne de Markov homogène sur F .

20.2.1 Description du modèle

On modélise un système de transmission satellitaire de la manière suivante. Des utilisateurs transmettent des messages dont la taille est bornée par une certaine constante K à travers un canal à des moments $t_n, n \in \mathbb{N}$ fixés par avance tels que $t_{n+1} - t_n \geq K$. Dans la suite, on prendra $t_n = n$ et $K = 1$. On suppose de plus que dès que deux messages sont émis au même moment, ils s'annulent et l'on doit recommencer (un seul message peut-être émis à la fois). D'autre part, l'utilisateur sait si son message a été transmis ou pas et décide donc le cas échéant de tenter de le renvoyer avec une certaine probabilité $\nu \in (0, 1)$, le message passant alors dans un état dit "semi-bloqué". Un message émis pour la première fois (c'est-à-dire non semi-bloqué) arrivant alors qu'aucun autre message "semi-bloqué" ne tente de rémettre est immédiatement transmis. Sinon, il passe dans l'état de message "semi-bloqué". Ces règles montrent en particulier que deux messages émis pour la première fois mais arrivant en même temps deviendront aussitôt des messages semi-bloqués. Les trois règles fondamentales à retenir sont donc les suivantes :

1. Les transmissions et les retransmissions ne peuvent se faire qu'aux instants $1, 2, \dots$
2. Tous les messages semi-bloqués (c'est-à-dire ceux ayant déjà essayé d'être transmis mais sans succès) tentent de retransmettre indépendamment des autres avec probabilité ν à chaque instant. On dit que la retransmission se fait selon une loi de Bernoulli.
3. Les messages émis pour la première fois, s'ils arrivent alors que le canal est libre (c'est-à-dire pas de messages semi-bloqués en train d'essayer de retransmettre) sont immédiatement transmis.

On notera X_n le nombre de messages semi-bloqués à l'instant n et A_n est le nombre de messages qui arrivent à l'instant n (ce sont donc des messages qui tentent d'être émis pour la première fois). On suppose que $(A_n)_{n \geq 0}$ est i.i.d. telle que :

$$\mathbb{P}[A_n = i] = a_i \quad , \quad \lambda = \mathbb{E}(A_n)$$

On supposera de plus que $a_0 + a_1 \in (0, 1)$ et $a_0 > 0$. λ est appelée l'intensité du trafic.

Ce système de transmission est connu sous le nom de système ALOHA.

20.2.2 Instabilité de ALOHA

- 1) Expliquer pourquoi si $X_n = k$, alors la probabilité que i parmi eux tentent de rémettre est donnée par :

$$b_i(k) = C_k^i \nu^i (1 - \nu)^{k-i}$$

- 2) La modélisation mathématique de ce système de transmission consiste à voir $(X_n)_n$ comme une chaîne de Markov à valeurs dans \mathbb{N} avec pour matrice de transition :

$$\begin{aligned}
p_{i,j} &= b_1(i)a_0 \quad \text{si } j = i - 1, \\
p_{i,j} &= [1 - b_1(i)]a_0 + b_0(i)a_1 \quad \text{si } j = i, \\
p_{i,j} &= [1 - b_0(i)]a_1 \quad \text{si } j = i + 1, \\
p_{i,j} &= a_{j-i} \quad \text{si } j \geq i + 2.
\end{aligned}$$

a) Expliquer pourquoi la probabilité de transition est celle donnée ci-dessus.

b) Montrer que cette chaîne est irréductible.

3) Le but de la question est de prouver que le système ALOHA est instable dans le sens suivant : “la chaîne n’est pas récurrente positive”.

a) Expliquer pourquoi ce critère est mathématiquement acceptable. Quels autres critères pourrait on demander ?

b) Montrer que l’instabilité est la négation de l’existence d’une distribution stationnaire π .

4) Montrer que si une distribution stationnaire pour la chaîne existait alors elle devrait satisfaire :

$$\begin{aligned}
\pi(i) &= \pi(i)\{[1 - b_1(i)]a_0 + b_0(i)a_1\} + \pi(i - 1)[1 - b_0(i - 1)]a_1 \\
&+ \pi(i + 1)b_1(i + 1)a_0 + \sum_{l=2}^{\infty} \pi(i - l)a_l
\end{aligned}$$

avec $\pi(j) = 0$ si $j < 0$.

5) On pose $P_N = \sum_{i=0}^N \pi(i)$.

a) Montrer que

$$P_N(1 - a_0) = \pi(N)b_0(N)a_1 + \pi(N + 1)b_1(N + 1)a_0 + \sum_{l=1}^N a_l P_{N-l}$$

b) Montrer que

$$\sum_{l=1}^N a_l P_{N-l} \leq P_{N-1}(1 - a_0)$$

c) En déduire

$$\frac{\pi(N + 1)}{\pi(N)} \geq \frac{(1 - a_0) - (1 - \nu)^N a_1}{(N + 1)\nu(1 - \nu)^N a_0}$$

6) Conclure.

20.2.3 Quelques outils théoriques : fonctions de Liapounov

Nous allons tenter de résoudre ce problème d’instabilité. Commençons par donner un critère de récurrence positive plus maniable que l’existence d’une probabilité stationnaire.

Lemme 20.2.1. Soit $(X_n)_n$ une chaîne de Markov irréductible homogène par rapport à sa filtration naturelle $(\mathcal{F}_n)_n$. Soit F un sous-ensemble fini de l'espace des états E . Soit $\tau = \inf\{n \geq 1 | X_n \in F\}$ le premier temps de retour dans F . Si pour tout $j \in F$, $\mathbb{E}_j[\tau] < +\infty$ alors la chaîne est récurrente positive.

1) On définit comme dans l'exercice les temps successifs de retour en F de X , τ_n , et la chaîne de Markov $Y_n = X_{\tau_n}$. Montrer que Y_n est irréductible.

2) En déduire que Y est positive récurrente.

3) On pose T_i le premier temps de retour en $i \in F$ de la chaîne X et \tilde{T}_i celui de la chaîne Y . La question 2) permet d'écrire $\mathbb{E}_i(\tilde{T}_i) < +\infty$. On pose alors $S_0 = \tau_1$ et $S_k = \tau_{k+1} - \tau_k$ pour $k \geq 1$. Montrer que

$$T_i = \sum_{k=0}^{\infty} S_k 1_{\{k < \tilde{T}_i\}}$$

4) En déduire $\mathbb{E}_i[T_i] = \sum_{k=0}^{\infty} \mathbb{E}_i[S_k 1_{\{k < \tilde{T}_i\}}]$.

5)

a) Montrer que $\{k < \tilde{T}_i\} \in \mathcal{F}_{\tau_k}$.

b) En utilisant la propriété forte de Markov, montrer que pour tout $l \in F$, on a

$$\mathbb{E}_i[S_k 1_{\{k < \tilde{T}_i\}} 1_{\{X_{\tau_k} = l\}}] = \mathbb{E}_i[S_k | X_{\tau_k} = l] \mathbb{P}_i(k < \tilde{T}_i, X_{\tau_k} = l)$$

6)

a) Montrer que $\mathbb{E}_i[S_k | X_{\tau_k} = l] = \mathbb{E}_l[\tau]$.

b) En déduire

$$\mathbb{E}_i(T_i) \leq \left(\sup_{l \in F} \mathbb{E}_l(\tau) \right) \sum_{k=0}^{\infty} \mathbb{P}_i(\tilde{T}_i > k) = \left(\sup_{l \in F} \mathbb{E}_l(\tau) \right) \mathbb{E}_i(\tilde{T}_i) < +\infty$$

et conclure.

Théorème 20.2.1. Théorème de Foster

Soit P une matrice de transition irréductible sur un espace d'état dénombrable E associée à une chaîne de Markov X . On suppose qu'il existe une fonction $h : E \rightarrow \mathbb{R}$ telle que $\inf_i h(i) > -\infty$ et vérifiant :

$$\sum_{k \in E} p_{ik} h(k) = \mathbb{E}_i(h(X_1)) < +\infty \text{ pour tout } i \in F,$$

$$\sum_{k \in E} p_{ik} h(k) = \mathbb{E}_i(h(X_1)) \leq h(i) - \epsilon \text{ pour tout } i \notin F,$$

pour un $\epsilon > 0$ et un ensemble fini F .

Prouvons ce théorème.

1) Montrer que l'on peut se ramener au cas où $h \geq 0$.

2) On appelle $\tau = \inf\{n \geq 1 | X_n \in F\}$ le premier temps de retour dans F et $\{\mathcal{F}_n\}_n$ la filtration naturelle associée à X . On définit $Y_n = h(X_n)1_{\{n < \tau\}}$.

a) Montrer que l'on a

$$\mathbb{E}_i(Y_{n+1} | \mathcal{F}_n) = \mathbb{E}_i[Y_{n+1}1_{\{n < \tau\}} | \mathcal{F}_n] + \mathbb{E}_i[Y_{n+1}1_{\{n \geq \tau\}} | \mathcal{F}_n]$$

b) Puis

$$\mathbb{E}_i(Y_{n+1} | \mathcal{F}_n) = \mathbb{E}_i[h(X_{n+1})1_{\{n < \tau\}} | \mathcal{F}_n] = 1_{\{n < \tau\}} \mathbb{E}_i[h(X_{n+1}) | \mathcal{F}_n]$$

c) Et enfin

$$\mathbb{E}_i(Y_{n+1} | \mathcal{F}_n) \leq 1_{\{n < \tau\}} h(X_n) - \epsilon 1_{\{n < \tau\}}$$

3)

a) Montrer que $\sum_{k=0}^{\infty} \mathbb{P}_i(\tau > k) = \mathbb{E}_i(\tau)$.

b) Dédire de la question 2c) l'inégalité suivante

$$0 \leq \mathbb{E}_i(Y_{n+1}) \leq \mathbb{E}_i(Y_n) - \epsilon \mathbb{P}(\tau > n)$$

4) En remarquant que Y est positive, montrer

$$\mathbb{E}_i(\tau) \leq \epsilon^{-1} h(i) < +\infty \text{ pour } i \notin F$$

5) Si $j \in F$, montrer en utilisant la propriété forte de Markov que

$$\mathbb{E}_j[\tau] = 1 + \sum_{i \notin F} p_{ji} \mathbb{E}_i[\tau]$$

6) Dédire de 4) et 5) que si $j \in F$, alors $\mathbb{E}_j(\tau) < +\infty$.

7) Utiliser le lemme précédent et conclure.

Ceci termine la preuve du théorème.

Corollaire 20.2.1. Lemme de Pake

Soit $(X_n)_n$ une chaîne de Markov irréductible sur \mathbb{N} telle que pour tout $n \geq 0$ et tout $i \in E$,

$$\mathbb{E}[X_{n+1} | X_n = i] < +\infty$$

et

$$\limsup_{i \rightarrow \infty} \mathbb{E}[X_{n+1} - X_n | X_n = i] < 0$$

Alors $(X_n)_n$ est récurrente positive.

Appliquer le théorème de Foster avec $2\epsilon = -\limsup_{i \rightarrow \infty} \mathbb{E}[X_{n+1} - X_n | X_n = i], h(i) = i$ et $F = \{i; i \leq i_0\}$ où i_0 est un entier suffisamment large.

Remarque 20.2.1. La fonction h dans le théorème de Foster est appelée une fonction de Lyapounov car elle joue le même rôle que celui joué par les fonctions de Lyapounov “classiques” dans la théorie de la stabilité pour les équations différentielles. h a en effet tendance à décroître le long des trajectoires du processus, du moins en dehors d’un certain ensemble d’états (F). Puisque h est positive (cf. la démonstration), cette fonction ne va pas pouvoir éternellement décroître et va donc un jour ou l’autre rentrer dans F . (p. 169)

20.2.4 Stabilisation du protocole Aloha

On a vu dans la première section que le protocole ALOHA était, avec une retransmission de probabilité ν , instable. Nous allons voir comment, en autorisant la probabilité de retransmission $\nu(\cdot)$ à dépendre du nombre de messages semi-bloqués et à condition que l’intensité du trafic soit suffisamment faible, on peut rendre le protocole stable si $a_0 > a_1$, hypothèse que nous faisons dans la suite. La description est donc toujours la même que dans la première section mais maintenant, ν n’est plus une constante mais une fonction du nombre de messages semi-bloqués. Si $X_n = k$ est le nombre de messages semi-bloqués, on note toujours $b_i(k)$ la probabilité que i parmi eux soient réémis.

Remarque 20.2.2. Dans le cas où l’on ne suppose pas $a_0 > a_1$, on peut montrer que le protocole reste instable même si l’on autorise une probabilité de retransmission non constante.

1) Montrer que :

$$\mathbb{E}[X_{n+1} - X_n | X_n = i] = \lambda - b_1(i)a_0 - b_0(i)a_1$$

2) En utilisant le lemme de Pake, montrer qu’il suffit de trouver une fonction ν telle que :

$$\lambda \leq \lim_{i \rightarrow \infty} (b_1(i)a_0 + b_0(i)a_1) - \epsilon$$

où $\epsilon > 0$.

3) Montrer que cette condition est équivalente à :

$$\lambda \leq g_i(\nu(i)) - \epsilon$$

où $g_k(\nu) = (1 - \nu)^k a_1 + k\nu(1 - \nu)^{k-1} a_0$.

4) En déduire qu’une condition de stabilité est :

$$\lambda < a_0 \exp \left\{ \frac{a_1}{a_0} - 1 \right\}$$

5) Critiquer ce protocole.

La principale critique que l’on puisse faire à ce protocole est que pour le rendre stable, notre stratégie suppose que chaque utilisateur connaisse le nombre de messages semi-bloqués. Ce n’est

pratiquement pas faisable. Ce qui est réaliste, c'est de concevoir une probabilité de retransmission qui se ferait en fonction de l'information suivante : "y a-t-il collision, transmission d'un message ou pas?". On peut montrer que ces stratégies sont effectivement stables.

Remarque 20.2.3. *Pour un autre protocole, voir [8] pp.174-178. Un autre exemple intéressant traité par [8] est une marche aléatoire 2-dimensionnelle avec des barrières réfléchissantes sur l'axe des ordonnées et des abscisses. Le processus reste donc toujours dans le quadrant positif. Le critère de Foster est aussi donné p. 135, problème corrigé 4.8 dans [1] (en particulier, l'exercice du début y est démontré). Une application y est donnée pour un modèle (simple) de file d'attente.*

Chapitre 21

Formulaire des chaînes de Markov

L'espace canonique $\Omega = \mathbb{E}^{\mathbb{N}}$, $\omega = (\omega_n)_{n \in \mathbb{N}}$, $X_n(\omega) = \omega_n$, $\mathcal{F}_n = \sigma(X_k, k \leq n)$, $\mathcal{F} = \sigma(X_k, k \geq 0)$
Temps d'arrêt de la filtration $\nu : \Omega \rightarrow \overline{\mathbb{N}}$ t.q. $\forall n, \{\nu \leq n\} \in \mathcal{F}_n$.
Tribu des événements antérieurs à ν , $\mathcal{F}_\nu = \{A \in \mathcal{F}_\infty, \forall n \geq 0, A \cap \{\nu \leq n\} \in \mathcal{F}_n\}$.
Si X une v.a. positive ou intégrable et ν un temps d'arrêt,

$$\mathbb{E}(X|\mathcal{F}_\nu) = X_\nu, \text{ p.s. (19.3.1)}$$

E est un ensemble d'états, fini ou dénombrable et \mathcal{E} la tribu de toutes les parties de E .

\mathcal{E}^+ : applications $E \rightarrow \overline{\mathbb{R}}^+$, ce sont des vecteurs colonne.

\mathcal{M}^+ : mesures positives sur E notées $\mu = (\mu(x))_{x \in E}$. Ce sont des vecteurs ligne et on peut faire les produits $\mu.f$ pour $\mu \in \mathcal{M}^+$, $f \in \mathcal{E}^+$.

P, Q, R , matrices positives ($P(x, y) \in \overline{\mathbb{R}}^+$), $x, y \in E$. Produit : $(QR)(x, y) = \sum_{z \in E} Q(x, z)R(z, y)$.
Convention : $+\infty \times 0 = 0$.

Si $f \in \mathcal{E}^+$, $(Qf)(x) = \sum_{y \in E} Q(x, y)f(y)$. Si $\mu \in \mathcal{M}^+$, $(\mu Q)(y) = \sum_{x \in E} \mu(x)Q(x, y)$.

Matrice de transition sur E : $P(x, y) \geq 0$, $\sum_{y \in E} p(x, y) = 1$. $P \geq 0$ est une matrice de transition si et seulement si $P1 = 1$

Une chaîne de Markov de loi initiale μ et de matrice de transition P est un processus aléatoire $(\Omega, \mathcal{F}, \mathcal{F}_n, X_n, \mathbb{P})$ à valeurs dans E et tel que

- (i) $\mathbb{P}(X_0 \in A) = \mu(A)$, $\forall A \subset E$
- (ii) $\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = P(X_n, A)$ p.s., $\forall A \subset E$, $n \geq 0$.

$\mathcal{F}_n = \mathcal{F}_n^0 = \sigma(X_0, X_1, \dots, X_n)$ est la "filtration naturelle du processus",

- (iii) Si $\mathbb{P}(X_n = a_n, \dots, X_0 = a_0) > 0$ alors

$$\mathbb{P}(X_{n+1} = b | X_n = a_n, X_{n-1} = a_{n-1}, \dots, X_0 = a_0) = P(a_n, b)$$

$$\mathbb{P}(X_0 = a_0, X_1 = a_1, \dots, X_n = a_n) = \mu(a_0)P(a_0, a_1) \dots P(a_{n-1}, a_n), \quad (19.4.1)$$

$$\mathbb{E}(f(X_0, \dots, X_n)) = \sum_{a_0, \dots, a_n \in E} f(a_0, \dots, a_n) \mu(a_0)P(a_0, a_1) \dots P(a_{n-1}, a_n) \quad (19.4.2)$$

$$\mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n) = \sum_{a_0 \in A_0, \dots, a_n \in A_n} \mu(a_0)P(a_0, a_1) \dots P(a_{n-1}, a_n) \quad (19.4.3)$$

$$\mathbb{P}(X_0 = a_0, X_{l_1} = a_1, \dots, X_{l_1+\dots+l_k} = a_k) = \mu(a_0)(P^{l_1})(a_0, a_1) \dots (P^{l_k})(a_{k-1}, a_k) \quad (19.4.4)$$

$$\mathbb{P}(X_{n+1} = b_1, \dots, X_{n+k} = b_k \mid X_0 = a_0, \dots, X_n = a_n) = P(a_n, b_1) \dots P(b_{k-1}, b_k) \quad (19.4.5)$$

$$\mathbb{P}(X_0 = a_0, \dots, X_n = a_n) = \mu_n(a_0, \dots, a_n) := \mu(a_0)P(a_0, a_1) \dots P(a_{n-1}, a_n) \quad (19.5.1)$$

$$\mathbb{P}_\mu(A) = \sum_{x \in E} \mu(x)\mathbb{P}_x(A) \text{ si } A = \{X_0 = a_0, \dots, X_n = a_n\} \quad (19.6.1)$$

$$\mathbb{P}_x(X_0 = a_0, \dots, X_n = a_n) = \mathbb{1}_x(a_0)P(a_0, a_1) \dots P(a_{n-1}, a_n)$$

$$\mathbb{P}_x(X_n = y) = P^n(x, y), \quad \mathbb{E}_x(f(X_n)) = (P^n f)(x) \quad (19.6.3)$$

$$\mathbb{E}_\mu(f(X_n)) = \sum_{x \in E} (P^n f)(x) = \mu(P^n f) = (\mu P^n)f. \quad (19.6.4)$$

”Si μ est la loi de X_0 , μP^n est la loi de X_n ”.

Opérateur de translation $\theta : \Omega \rightarrow \Omega$, $\omega = (\omega_n)_{n \geq 0} \rightarrow (\omega_{n+1})_{n \geq 0}$. On pose $\theta_p = \theta^p$. $X_n \circ \theta_p = X_{n+p}$.

Soit un temps d'arrêt τ . Sur $\{\tau < +\infty\} : \forall n \geq 0, X_n \circ \theta_\tau = X_{n+\tau}$.

Propriété de Markov. Si $\varphi \in b\mathcal{F}$ ou \mathcal{F}^+ $\psi \in b\mathcal{F}_n$ ou \mathcal{F}_n^+ ,

$$\mathbb{E}_\mu(\varphi \circ \theta_n \mid \mathcal{F}_n) = \mathbb{E}_{X_n}(\varphi), \text{ p.s.}, \quad (19.7.1)$$

$$\mathbb{E}_\mu(\psi \varphi \circ \theta_n) = \mathbb{E}_\mu(\psi \mathbb{E}_{X_n}(\varphi)). \quad (19.7.2)$$

Propriété de Markov forte. Pour $\varphi \in b\mathcal{F}$ (ou \mathcal{F}^+), toute $\psi \in b\mathcal{F}_\tau$ (ou \mathcal{F}_τ^+),

$$\mathbb{E}_\mu(\mathbb{1}_{\tau < +\infty} \varphi \circ \theta_\tau \mid \mathcal{F}_\tau) = \mathbb{1}_{\tau < +\infty} \mathbb{E}_{X_\tau}(\varphi) \text{ p.s.},$$

$$\mathbb{E}_\mu(\mathbb{1}_{\tau < +\infty} \psi \varphi \circ \theta_\tau) = \mathbb{E}_\mu(\mathbb{1}_{\tau < +\infty} \psi \mathbb{E}_{X_\tau}(\varphi)). \quad (19.7.3)$$

Temps de passage par x successifs après 0 :

$$\begin{cases} \sigma_x := \inf\{n \geq 1, X_n = x\} = \sigma_x^1, \\ \sigma_x^n := \sigma_x^{n-1} + \sigma_x \circ \theta_{\sigma_x^{n-1}} \text{ si } \sigma_x^{n-1} < +\infty, \\ \sigma_x^n := \sigma_x^{n-1} = +\infty \text{ sinon.} \end{cases}$$

$$\mathbb{P}_x(\sigma_x^n < +\infty) = \mathbb{P}_x(\sigma_x < +\infty)^n. \quad (19.8.1)$$

Matrice potentielle de P (ou de X) : $U = I + P + \dots + P^n + \dots = \sum_{k \geq 0} P^k$.

$N_A(\omega)$ désigne le nombre de visites de la trajectoire ω à A .

$$N_A = \sum_{k \geq 0} \mathbb{1}_A(X_k), \quad \mathbb{E}_x(N_A) = U(x, A), \quad E_x(N_y) = U(x, y). \quad (19.9.1)$$

(i) Si $U(x, x) = +\infty$, alors $\mathbb{P}_x(\sigma_x < +\infty) = 1$ et $\mathbb{P}_x(N_x = +\infty) = 1$. Le point x est dit récurrent.

(ii) Si $U(x, x) < +\infty$, alors $\mathbb{P}_x(\sigma_x < +\infty) < 1$ et $\mathbb{P}(N_x < +\infty) = 1$. Le point x est dit transitoire (transient).

$$N_x = \mathbb{1}_{\{x\}}(X_0) + \sum_{n \geq 1} \mathbb{1}_{\sigma_x^n < +\infty} \quad (19.10.1)$$

Si $x \neq y$, alors $U(x, y) = \mathbb{P}_x(\sigma_y < +\infty)U(y, y)$. On dit que x conduit à y et on note $x \rightarrow y$ si $\mathbb{P}_x(\tau_y < +\infty) > 0$.

0. Les trois conditions suivantes sont équivalentes : (1) $\mathbb{P}(\sigma_y < +\infty) > 0$, (2) $\exists n \geq 1$ tel que $P^n(x, y) > 0$ et (3) $U(x, y) (= \mathbb{E}_x(N_y)) > 0$.

1. F est close si et seulement si pour tout $x \in F$, $P(x, F) = 1$ ou si et seulement si, pour tout $x \in F$ et tout $y \notin F$, $U(x, y) = 0$.
2. P est irréductible si et seulement si $\forall x, y \in E$, $U(x, y) > 0$.
3. Dans une chaîne irréductible, ou bien tous les états sont récurrents, ou bien tous les états sont transitoires. On parle de chaîne récurrente dans le premier cas et de chaîne transitoire dans le second. Dans le premier cas $U(x, y) = +\infty$ pour tous $x, y \in E$ et dans le second, $0 < U(x, y) < +\infty$.

Soit X une chaîne irréductible récurrente. Alors X possède une mesure invariante μ unique à une constante multiplicative près. De plus, pour tout $y \in E$, $0 < \mu(y) < +\infty$. Enfin, pour $x \in E$, la mesure invariante μ_x telle que $\mu_x(x) = 1$ est donnée par

$$\begin{aligned} \mu_x(y) &= \mathbb{E}_x\left(\sum_{k=0}^{\sigma_x-1} \mathbb{1}_{\{y\}}(X_k)\right). \\ \mu_x \cdot f &= \mathbb{E}_x\left(\sum_{k=0}^{\sigma_x-1} f(X_k)\right). \end{aligned} \quad (19.12.2)$$

Soit X une chaîne irréductible récurrente de mesure invariante μ .

- (i) $\mu(E) = +\infty$. Alors, pour tout $x \in E$, $\mathbb{E}_x(\sigma_x) = +\infty$. La chaîne est dite récurrente nulle.
- (ii) $\mu(E) < +\infty$. Alors, pour tout $x \in E$, $\mathbb{E}_x(\sigma_x) < +\infty$. La chaîne est dite récurrente positive. Dans ce cas l'unique probabilité invariante est donnée par

$$\pi(x) = \frac{1}{E_x(\sigma_x)}.$$

Chapitre 22

Processus de Poisson

22.1 Définitions et propriétés

Nous présentons ici les définitions et propriétés élémentaires relatifs au processus de Poisson telles qu'on peut les trouver dans [8]. On peut dire beaucoup plus de choses sur le processus de Poisson que ce qui est dit ici. On trouvera de nombreuses informations dans [22]. Les processus de Poisson sont souvent mal présentés dans les livres de probabilités élémentaires. [8] fait à peu près exception...

Définition 22.1.1. *Un processus de pointage sur le demi axe positif \mathbb{R}^+ est une suite $(T_n)_{n \geq 0}$ de variables aléatoires strictement positives (\mathbb{P} p.s.) telles que presque sûrement :*

- i) $T_0 = 0$
- ii) $0 < T_1 < T_2 < \dots$
- iii) $\lim_{n \rightarrow \infty} T_n = +\infty$

La suite $(S_n)_n$ définie par $S_n = T_n - T_{n-1}$ est appelée la suite des inter-événements ou des inter-arrivées (terme clair dans le contexte des files d'attente). Pour tout interval $(a, b]$ de \mathbb{R}^+ , on définit la variable aléatoire

$$N((a, b]) = \sum_{n \geq 1} \mathbb{1}_{(a, b]}(T_n)$$

et on simplifie souvent l'écriture en notant $N(a, b)$ au lieu de $N((a, b])$. Dans le cas particulier où $(a, b] = (0, b]$, on pose $N(b) = N((0, b])$. En particulier, on a $N(0) = 0$ et $N(a, b) = N(b) - N(a)$.

Proposition 22.1.1. *Soit $(T_n)_n$ un processus de pointage sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Pour \mathbb{P} presque tout $\omega \in \Omega$, la trajectoire $t \rightarrow N(t, \omega)$ est croissante, continue à droite avec des limites à gauche (càdlàg).*

La famille de variables aléatoires $\{N(t)\}_{t \geq 0}$ est appelée le processus de comptage associé au processus de pointage $(T_n)_n$.

Définition 22.1.2. *Un processus de pointage N sur le demi-axe positif est appelé un processus de Poisson homogène avec intensité $\lambda > 0$ si :*

- i) *Pour tout temps t_i , $i \in \{1, \dots, k\}$ tels que $0 \leq t_1 \leq t_2 \leq \dots \leq t_k$, les variables aléatoires $N(t_i, t_{i+1}]$, $i \in \{1, \dots, k-1\}$ sont indépendantes.*
- ii) *Pour tout interval $(a, b] \subset \mathbb{R}^+$, $N(a, b]$ est une variable aléatoire de Poisson avec moyenne*

$\lambda(a - b)$.

Ainsi pour tout entier $k \geq 0$,

$$\mathbb{P}(N(a, b) = k) = e^{-\lambda(b-a)} \frac{[\lambda(b-a)]^k}{k!}$$

et en particulier

$$\mathbb{E}[N(a, b)] = \lambda(b - a)$$

Dans ce sens, λ s'interprète comme une "densité de points".

La condition i) porte le nom de "propriété d'indépendance des accroissements". En particulier, d'après le théorème des coalitions, $N(a, b)$ est indépendant de $(N(s), s \in (0, a])$. Pour cette raison, on dit souvent que le processus de Poisson est dans mémoire. Passons maintenant au premier théorème important concernant les processus de Poisson.

Théorème 22.1.1. *Soit N un processus de Poisson d'intensité λ . La suite d'interévénements $(S_n)_n$ est une suite de variables aléatoires i.i.d. de loi exponentielle de paramètre λ .*

Démonstration. On reproduit la démonstration de [18] mais sans la faute (d'expression)! On considère $S_1 = T_1$ et on va montrer que S_1 est distribuée selon une loi exponentielle de paramètre λ . En effet, on a

$$\mathbb{P}(S_1 > t) = \mathbb{P}(N(t) = 0) = e^{-\lambda t}$$

Maintenant, nous allons conditionner relativement à S_1 . La preuve de [18] est la suivante :

$$\mathbb{P}(S_2 > t | S_1 = t_1) = \mathbb{P}(\text{pas d'arrivée dans } (t_1, t_1 + t] | S_1 = t_1)$$

L'événement $\{S_1 = t_1\}$ est relié aux arrivées durant l'intervalle de temps $[0, t_1]$ tandis que l'événement pas d'arrivées dans $(t_1, t_1 + t]$ est relié aux arrivées après le temps t_1 . Ces événements sont *indépendants* ... Arrêtons nous! Que signifie " $\{S_1 = t_1\}$ est indépendant de ..." puisque cet événement est de probabilité nulle? En fait cela ne signifie pas grand chose si c'est énoncé tel quel (bien qu'il soit clair que ce soit clair dans l'esprit de [18]). La bonne formulation (jusqu'à preuve du contraire) doit se faire via les lois conditionnelles. On rappelle la propriété suivante :

Proposition 22.1.2. *(Propriété de transfert conditionnel)*

Si $X \in \mathbb{R}^p$ et $Y \in \mathbb{R}^k$ sont deux variables aléatoires et $f : \mathbb{R}^{p+k} \rightarrow \mathbb{R}^q$ borélienne, alors

$$\mathbb{P}(f(X, Y) \in \cdot | X = x) = \mathbb{P}(f(x, Y) \in \cdot | X = x)$$

En particulier, si X et Y sont indépendantes, on a

$$\mathbb{P}(f(X, Y) \in \cdot | X = x) = \mathbb{P}(f(x, Y) \in \cdot)$$

Ce que l'on peut dire est : la variable aléatoire S_1 est indépendante de la variable aléatoire $N(t_1 + t) - N(t)$ et par la propriété de transfert conditionnel, on a :

$$\mathbb{P}(S_2 > t | S_1 = t_1) = \mathbb{P}(\mathbb{1}_{\{0\}}(N(S_1 + t) - N(S_1)) | S_1 = t_1) = \mathbb{P}(\mathbb{1}_{\{0\}}(N(t_1 + t) - N(t_1))) = e^{-\lambda t}$$

ce qui prouve que S_2 est indépendant de S_1 et a une distribution exponentielle de paramètre λ . La preuve se poursuit par récurrence en conditionnant l'événement $\{S_{n+1} > t\}$ relativement à $\{S_1 = t_1, \dots, S_n = t_n\}$ □

On passe maintenant à un théorème assez simple et naturel. Que se passe-t-il si l'on met en compétition deux processus de Poisson indépendants ?

Avant d'énoncer le théorème, commençons par une remarque élémentaire. On se donne deux processus de Poisson indépendants N^1 et N^2 d'intensités respectives λ_1 et λ_2 et définis via leurs temps d'arrivée $(T_n^1)_n$ et $(T_n^2)_n$. Il est facile de voir que $(T_n^1)_n$ et $(T_n^2)_n$ n'ont pas de points en commun (ormis 0). En effet, cela résulte du théorème précédent puisque pour tout $k, \ell \geq 1$, T_k^1 et T_ℓ^2 ont une densité par rapport à la mesure de Lebesgue et sont indépendantes si bien que $\mathbb{P}(T_k^1 = T_\ell^2) = 0$.

Théorème 22.1.2. *Sous les conditions précédentes, le processus de pointage $(T_n)_n$ défini par la réunion de $\{T_n^1; n \geq 1\}$ et $\{T_n^2; n \geq 1\}$ est un processus de pointage associé à un processus de Poisson de paramètre $\lambda = \lambda_1 + \lambda_2$.*

Démonstration. Il suffit de remarquer que le processus de comptage N associé au processus de pointage $(T_n)_n$ vérifie $N(a, b) = N^1(a, b) + N^2(a, b)$. Cependant, a priori, il n'est pas évident que $\lim_{n \rightarrow \infty} T_n = +\infty$. On vérifie sans peine que $\mathbb{E}(N(a)) = \lambda a < +\infty$ donc que presque sûrement pour tout $a > 0$, $N(a) < +\infty$, ce qui est équivalent à $\lim_{n \rightarrow \infty} T_n = +\infty$. \square

On peut généraliser le théorème de superposition précédent avec un nombre infini (dénombrable) de processus de Poisson indépendants.

Théorème 22.1.3. *Soit $\{N^i\}_i$ une suite de processus de Poisson indépendants avec des intensités respectives $\lambda_i > 0$ telles que $\lambda = \sum_i \lambda_i < +\infty$. Soit $(T_n)_n$ le processus de pointage obtenu par réunion des processus de pointage des $(N^i)_i$ (remarquer qu'il n'y a pas risque de superposition d'un T_k^i avec un T_ℓ^j pour $i \neq j$). Le processus de comptage associé à $(T_n)_n$ est un processus de Poisson d'intensité λ .*

D'autre part, dénotons par $Z = T_1$ le premier temps d'arrivée de $N = \sum_i N^i$ et par J l'indice du processus de Poisson responsable du premier saut de N . Alors Z et J sont indépendants; la loi de Z est une exponentielle de paramètre λ et la loi de J est donnée par $\mathbb{P}(J = i) = \frac{\lambda_i}{\lambda}$.

Démonstration. La première partie du théorème se démontre comme dans le théorème précédent. En particulier, la loi de Z est une exponentielle de paramètre λ . Le fait que J soit bien définie résulte du fait que la probabilité pour que deux processus N^i et N^j aient un saut en commun est nulle. On va d'abord considérer le cas d'un nombre fini K de processus de Poisson indépendants. On a à montrer que si (X_1, \dots, X_K) sont K variables aléatoires indépendantes de loi exponentielle de paramètres respectifs $\lambda_1, \dots, \lambda_K$, si $Z_K = \inf(X_1, \dots, X_K)$ et si J_K est définie par $X_{J_K} = Z_K$ alors

$$\mathbb{P}(J_K = i, Z_K \geq a) = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_K} \exp\{-(\lambda_1 + \dots + \lambda_K)a\}$$

Cette égalité se montre facilement. Pour étendre ce résultat au cas $K = +\infty$, on remarque que l'événement $\{J_K = i, Z_K \geq a\}$ décroît avec K vers l'événement $\{J = 1, Z \geq a\}$. \square

22.2 TD Probabilités : Processus de Poisson

On considère un processus de Poisson d'intensité λ . On note $(T_k)_{k=0 \dots \infty}$ les instants de sauts avec $T_0 = 0$ et $S_k = T_k - T_{k-1}$ les temps d'inter-arrivées qui sont donc des variables i.i.d. de loi

exponentielle de paramètre λ .

Théorème 22.2.1. *Montrer que pour tout $n \geq 1$, la loi de (T_1, \dots, T_n) conditionnellement à $\{N(t) = n\}$ est celle d'une loi de Dirichlet de paramètre n sur $(0, t)$.*

On pourra trouver cette propriété démontrée dans [21] ou dans [26].

Exercice 22.2.1. Un modèle de compteur ([21] pp. 128-131)

Soit $N(t)$ un processus de Poisson de paramètre λ dont les temps de sauts sont notés t_i (les variables aléatoires $s_i = t_i - t_{i-1}$ sont donc i.i.d. de loi exponentielle de paramètre λ). Des signaux arrivent alors aux instants t_i avec une amplitude X_i et leurs effets décroissent de manière exponentielle une fois arrivés. Le détecteur est d'autre part linéaire et par suite, les amplitudes des signaux s'ajoutent. Aussi, au temps t , l'amplitude détectée par le détecteur est :

$$\eta(t) = \sum_{i=1}^{N(t)} X_i \exp(-\alpha(t - t_i))$$

On suppose les X_i indépendantes du processus de Poisson des arrivées et i.i.d. ayant pour loi une loi à densité h supportée par \mathbb{R}^+ et supposée continue pour simplifier. Soit ψ la fonction caractéristique des X_i et ϕ_t celle de η_t .

1) Soit (τ_1, \dots, τ_n) n variables aléatoires indépendantes uniformes sur $(0, t)$ indépendantes de (X_1, \dots, X_n) . On note $(\tau_1^*, \dots, \tau_n^*)$ les variables aléatoires définies par :

$$\tau_1^* = \inf\{\tau_1^*, \dots, \tau_n^*\}, \tau_{k+1}^* = \inf\{\{\tau_1, \dots, \tau_n\} \setminus \{\tau_1^*, \dots, \tau_k^*\}\}$$

a) Soit $X'_k = X_j$ où $\tau_k^* = \tau_j$ pour tout $k = 1, \dots, n$. Montrer que les X'_k sont des variables aléatoires bien définies (à un ensemble de probabilité nulle près).

Solution

Les variables aléatoires $(\tau_i)_i$ sont des variables aléatoires indépendantes à densité (par rapport à la mesure de Lebesgue). Par suite, la probabilité qu'il existe un couple d'entiers distincts (i, j) avec $\tau_i = \tau_j$ est nulle. Les variables aléatoires $(X'_i)_i$ sont donc bien définies.

b) Montrer que les $(X'_k)_{k=1, \dots, n}$ forment une suite de variables aléatoires indépendantes identiquement distribuées de même loi que X_1 et indépendantes des $(\tau_k)_{k=1, \dots, n}$.

Solution

Soit H, f_1, \dots, f_n des fonctions boréliennes bornées. On a :

$$\begin{aligned} & \mathbb{E} \left[H(\tau_1, \dots, \tau_n) f_1(X'_1) \dots f_n(X'_n) \right] \\ &= \sum_{\sigma \in \Sigma_n} \mathbb{E} \left[\mathbb{1}_{\{\tau_{\sigma(1)} < \dots < \tau_{\sigma(n)}\}} H(\tau_1, \dots, \tau_n) f_1(X_{\sigma(1)}) \dots f_n(X_{\sigma(n)}) \right] \\ &= \sum_{\sigma \in \Sigma_n} \mathbb{E} \left[\mathbb{1}_{\{\tau_{\sigma(1)} < \dots < \tau_{\sigma(n)}\}} H(\tau_1, \dots, \tau_n) \right] \mathbb{E} [f_1(X_{\sigma(1)}) \dots f_n(X_{\sigma(n)})] \quad (22.2.1) \\ &= \sum_{\sigma \in \Sigma_n} \mathbb{E} \left[\mathbb{1}_{\{\tau_{\sigma(1)} < \dots < \tau_{\sigma(n)}\}} H(\tau_1, \dots, \tau_n) \right] \mathbb{E} [f_1(X_1) \dots f_n(X_n)] \\ &= \mathbb{E} [H(\tau_1, \dots, \tau_n)] \mathbb{E} [f_1(X_1)] \dots \mathbb{E} [f_n(X_n)] \end{aligned}$$

L'égalité entre la deuxième et troisième ligne découle de l'indépendance entre les τ_i et les X_i . L'égalité entre la troisième et la quatrième provient de l'équidistribution et de l'indépendance des X_i . Ceci prouve le résultat annoncé.

2)a) Montrer que conditionnellement à $\{N_t = n\}$, $(X_1, \dots, X_n, t_1, \dots, t_n)$ sont distribuées comme $(X_1, \dots, X_n, \tau_1^*, \dots, \tau_n^*)$.

Solution

Ceci provient du théorème rappelé au-dessus.

b) Montrer que pour toute fonction borélienne à valeurs complexes f , on a :

$$\mathbb{E} \left(f \left(\sum_{i=1}^n X_i \exp(-\alpha(t - \tau_i^*)) \right) \right) = \mathbb{E} \left(f \left(\sum_{i=1}^n X'_i \exp(-\alpha(t - \tau_i^*)) \right) \right)$$

Solution

On vient de voir que les $(X'_i)_i$ étaient indépendantes entre elles de même loi que les $(X_i)_i$ et d'autre part indépendantes des $(\tau_i)_i$ (comme les $(X_i)_i$) donc des $(\tau_i^*)_i$. On a par conséquent que la loi de $(X_1, \dots, X_n, \tau_1^*, \dots, \tau_n^*)$ est la même que celle de $(X'_1, \dots, X'_n, \tau_1^*, \dots, \tau_n^*)$, ce qui donne immédiatement le résultat.

b) Montrer que

$$\mathbb{E} \left(f \left(\sum_{i=1}^n X'_i \exp(-\alpha(t - \tau_i^*)) \right) \right) = \mathbb{E} \left(f \left(\sum_{i=1}^n X_i \exp(-\alpha(t - \tau_i)) \right) \right)$$

Solution

Les deux sommes à l'intérieur des espérances sont égales à une réindexation près.

Les dernières questions sont des trivialités.

3) On pose $Y_t(i) = X_i \exp(-\alpha(t - \tau_i))$.

a) Montrer que les $(Y_i)_{i=1 \dots n}$ sont i.i.d.. On notera θ_t leur fonction caractéristique.

b) Montrer que :

$$\theta_t(s) = \frac{1}{t} \int_0^t \psi(se^{-\alpha v}) dv$$

4) En déduire que :

$$\phi_t(w) = \exp \left(-\lambda \int_0^t (1 - \psi(we^{-\alpha v})) dv \right)$$

22.3 Processus de Poisson (version élégante)

Cette introduction conclue à la hâte est tirée de [22].

22.3.1 Généralités

Bien que pouvant être développée sur des espace plus généraux, nous restreindrons notre étude aux espaces multidimensionnels \mathbb{R}^d . On notera $\mathcal{B}(\mathbb{R}^d)$ la tribu des boréliens de \mathbb{R}^d . Dans toute la suite $(\Omega, \mathcal{F}, \mathbb{P})$ désignera un espace probabilisé fixé.

On dira qu'une fonction Π de Ω dans $\mathcal{B}(\mathbb{R}^d)$ est un ensemble aléatoire dénombrable si et seulement si :

- i) Pour tout $\omega \in \Omega$, $\Pi(\omega)$ est un borélien dénombrable (bien entendu, ici, le fait que Π soit à valeurs dans $\mathcal{B}(\mathbb{R}^d)$ est superfétatoire).
- ii) Pour tout borélien A , la fonction $N(A) : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$ définie par $N(A)(\omega) = \text{Card}(\Pi(\omega) \cap A)$ est mesurable.

Définition 22.3.1. *Un processus de Poisson sur \mathbb{R}^d est un ensemble aléatoire (sur \mathbb{R}^d) dénombrable tel que :*

- i) *Pour toute séquence finie d'ensembles disjoints, les variables aléatoires $N(A_1), \dots, N(A_n)$ sont indépendantes.*
- ii) *$N(A)$ a pour loi une loi de Poisson de paramètre $\mu(A) \in [0, +\infty]$.*

Remarque 22.3.1. *La convention adoptée est la suivante. Une loi de Poisson de paramètre 0 est la masse de Dirac en 0. Une loi de Poisson de paramètre infini est la masse de Dirac concentrée sur $+\infty$.*

Il est aisé de remarquer que $\mu : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, +\infty]$ est en fait une mesure (de masse non nécessairement finie) que l'on appelle la *mesure moyenne* du processus. Toute mesure μ ne peut être une mesure moyenne. En effet, supposons que μ soit une mesure moyenne associée à un processus de Poisson Π et qu'elle soit atomique. Il existe alors $x \in \mathbb{R}^d$ tel que $m = \mu(\{x\}) > 0$ et $\mu(\{x\}) = \mathbb{E}(N(\{x\})) \leq 1$ car $N(\{x\}) \leq 1$. Or on a

$$\mathbb{P}(N(\{x\}) \geq 2) = 1 - e^{-m} - me^{-m} > 0$$

et il y a contradiction avec le fait que $N(\{x\}) \leq 1$.

Ainsi, une mesure moyenne doit être non atomique.

Dans le cas où la mesure moyenne du processus de Poisson est absolument continue par rapport à la mesure de Lebesgue avec une densité $\lambda(x)$, on dit que λ est l'intensité du processus (en dimension 1, on parle de taux). Dans le cas où λ est une fonction constante, on parle de processus homogène ou uniforme.

22.4 Superposition, restriction et transformation

Théorème 22.4.1. *(de superposition)*

Soit Π_1, Π_2, \dots une collection dénombrable de processus de Poisson sur \mathbb{R}^d tels que Π_n ait pour mesure moyenne μ_n . Alors la superposition Π des Π_n définie par

$$\Pi = \cup_{n=1}^{\infty} \Pi_n$$

est un processus de Poisson de mesure moyenne

$$\mu = \sum_{n=1}^{\infty} \mu_n$$

Théorème 22.4.2. (de restriction)

Soit Π un processus de Poisson avec mesure moyenne μ et soit S un borélien de \mathbb{R}^d . Alors l'ensemble aléatoire dénombrable

$$\Pi_1 = \Pi \cap S$$

est un processus de Poisson sur \mathbb{R}^d avec mesure moyenne

$$\mu_1(A) = \mu(A \cap S)$$

et un processus de Poisson sur S de mesure moyenne la restriction de μ à S .

Théorème 22.4.3. (de transformation)

Soit Π un processus de Poisson avec une mesure moyenne μ σ -finie sur \mathbb{R}^d et soit $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ une fonction mesurable telle que la mesure image μ^* de μ par f soit non atomique. Alors $f(\Pi)$ est un processus de Poisson de mesure moyenne μ^* .

22.4.1 Théorème d'existence

Nous passons maintenant à la question de l'existence d'un processus de Poisson associé à une mesure donnée (non atomique) μ .

Théorème 22.4.4. Soit μ une mesure non atomique sur \mathbb{R}^d qui peut s'écrire

$$\mu = \sum_{n=1}^{\infty} \mu_n, \quad \mu_n(\mathbb{R}^d) < \infty$$

Alors il existe un processus de Poisson sur \mathbb{R}^d ayant μ pour mesure moyenne.

Remarque 22.4.1. La condition sur μ est moins restrictive que la condition de σ -finitude qu'elle implique.

La démonstration de l'existence est en fait basée sur une description du processus de Poisson qui est la suivante. Soit Π un processus de Poisson associée à une mesure moyenne μ supposée de masse finie. On cherche à déterminer la loi de Π conditionnellement à $N(\mathbb{R}^d) = n$, qui est donc un ensemble aléatoire à n points. Soit p la probabilité définie par $\pi(\cdot) = \mu(\cdot)/\mu(\mathbb{R}^d)$ et X_1, \dots, X_n n variables aléatoires indépendantes de loi p . Du fait que la mesure μ est non atomique, la probabilité pour qu'il existe $i \neq j$ tels que $X_i = X_j$ est nulle.

Conditionnellement à $N(\mathbb{R}^d) = n$, Π est distribué comme $\{X_1, \dots, X_n\}$ (voir Kingman pour une formulation précise).

22.4.2 Processus de Poisson marqués

Théorème 22.4.5. (théorème de la boîte de peinture)

Soit Π un processus de Poisson sur \mathbb{R}^d de mesure moyenne μ . Supposons que les points de Π soient colorés aléatoirement avec k couleurs, la probabilité de colorer un point avec la couleur i étant p_i . On suppose que le coloriage se fait de manière indépendante et indépendamment des points de Π . Soit Π_i l'ensemble des points de Π qui ont été coloriés avec la couleur i . Alors les Π_i sont des processus de Poisson indépendants avec pour mesure moyennes respectives $\mu_i = p_i\mu$.

Nous passons maintenant à une généralisation du théorème précédent connu sous le nom de théorème de marquage. L'idée est la suivante. On se donne un processus de Poisson Π et en chaque point $X \in \Pi$, on attache une marque m_X . La loi de m_X peut dépendre de X mais non des autres points de Π et donc les m_X sont indépendantes. Cette formulation demeure trop floue. En effet, les m_X sont donc indexés par les point X de Π mais cet ensemble est aléatoire. On pourrait être tenté de se donner dès le départ une famille $(m_x)_x$ de variables aléatoires indexées par $x \in \mathbb{R}^d$ indépendantes entre elles et indépendante de Π mais se pose alors l'existence d'une telle famille. La manière la plus simple de décrire ce que l'on veut est d'utiliser la notion de noyaux de transition que l'on a vu dans le chapitre sur l'espérance conditionnelle.

Soit Π un processus de Poisson sur \mathbb{R}^d avec mesure moyenne μ . On se donne un autre espace mesuré \mathbb{R}^k muni de la tribu des boréliens et qui sera l'espace des marques. On considère d'autre part une probabilité de transition $p : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^k) \rightarrow [0, 1]$, c'est-à-dire une application telle que pour tout $x \in \mathbb{R}^d$, $p(x, \cdot)$ soit une probabilité et pour tout borélien B de \mathbb{R}^k , l'application $p(\cdot, B)$ soit mesurable. Un marquage de Π est un ensemble aléatoire sur $\mathbb{R}^d \times \mathbb{R}^k$ de la forme

$$\Pi^* = \{(X, m_X); X \in \Pi\}$$

tel que

- i) La projection de Π^* sur \mathbb{R}^d soit Π .
- ii) Conditionnellement à Π , la distribution des $(m_X)_{X \in \Pi}$ est une suite de variables aléatoires indépendantes de distributions respectives $\pi(X, \cdot)$.

Théorème 22.4.6. L'ensemble aléatoire dénombrable Π^* est un processus de Poisson sur $\mathbb{R}^d \times \mathbb{R}^k$ avec mesure moyenne μ^* donnée par

$$\mu^*(C) = \int_C \mu(dx)p(x, dm)$$

Un corrolaire immédiat est le théorème de déplacement

Théorème 22.4.7. (de déplacement) Soit Π un processus de Poisson sur \mathbb{R}^d avec une intensité $\lambda(x)$. On suppose que les points de Π sont déplacés de manière aléatoire et chaque déplacement est exécuté indépendamment des autres. On suppose que la loi du déplacement d'un point de Π situé en $x \in \mathbb{R}^d$ est une loi à densité $\rho(x, \cdot)$.

Alors les point déplacés forment un processus de Poisson Π' avec intensité λ' donnée par

$$\lambda'(y) = \int_{\mathbb{R}^d} \lambda(x)\rho(x, y)dx$$

En particulier, si l'intensité est une fonction constante égale à λ et si $\rho(x, y)$ est une fonction de $y - x$ alors $\lambda(y) = \lambda$ pour tout y .

Remarque 22.4.2. *La formulation de ce théorème peut paraître floue. Pour en donner une formulation rigoureuse, il suffit de paraphraser ce théorème à l'aide de noyaux de transition comme on l'a fait pour le théorème de marquage.*

On trouvera dans [22] les preuves et des applications de ces théorèmes à des modèles écologiques, de trafic routier ...

Chapitre 23

Limites hydrodynamiques

Ce td est issu du chapitre 1 de [23]. Il pourra servir à illustrer les leçons concernant la loi des grands nombres et le théorème limite central, les variables de Poisson et la transformée de Laplace.

Dans ce qui suit, nous modéliserons un fluide par un ensemble de particules indistinguables qui évoluent de manière aléatoire et indépendante. Cette description est donc en premier lieu microscopique. Un état d'équilibre est un état statistique, c'est-à-dire décrit par une probabilité sur l'espace des phases des particules constituant le fluide. Dans l'étude des systèmes thermodynamiques hors équilibre (c'est-à-dire que les particules ne sont pas distribuées selon une mesure d'équilibre), une hypothèse fondamentale appelée l'hypothèse de l'équilibre local est souvent admise. Celle-ci consiste à supposer qu'à chaque instant t , on peut diviser le système étudié, qui dans son ensemble est hors équilibre, en "cellules" de taille intermédiaire - assez petites pour que les propriétés thermodynamiques du système y varient peu, mais assez grandes pour pouvoir être traitées comme des sous-systèmes thermodynamiques en contact avec leur environnement. On peut alors définir des grandeurs thermodynamiques locales, uniformes à l'intérieur de chaque cellule, mais variant d'une cellule à l'autre. Notre fluide ne sera décrit au niveau macroscopique que par une seule variable thermodynamique, la densité de particules. D'autre part, à chaque valeur de la densité correspondra un équilibre. Partant d'une densité non uniforme (dans l'espace) sur ce système, on s'attend à ce qu'au cours du temps, la densité varie. Il faut en fait pour comprendre le concept d'équilibre local introduire deux échelles de temps distinctes. Le premier, et le plus petit de ceux-ci, est la durée (moyenne) d'une collision τ_0 . Le deuxième temps, beaucoup plus grand est le temps de relaxation vers un équilibre local $\tau_R \gg \tau_0$. Celui-ci est en effet le résultat de très nombreuses collisions. Pour que l'hypothèse d'équilibre local soit valide, il faut donc que $\tau_0 \ll \tau_R$. Nous verrons dans la suite comment cette hypothèse apparaît naturellement et nous démontrerons (dans un cas très simple) l'hypothèse d'équilibre local.

Si l'on veut faire rentrer le modèle proposé dans les thèmes du programme, on pourra toujours assimiler les particules à des individus, des bactéries où des paquets d'informations qui transitent d'un site à un autre...

On utilisera les notations suivantes. Pour $N \geq 1$, $\mathbb{T}_N = (\mathbb{Z}/N\mathbb{Z})$ est le tore à N points, $\mathbb{T}_N^d = (\mathbb{T}_N)^d$ le tore à N points d -dimensionnel et \mathbb{T}^d le tore d -dimensionnel $[0, 1)^d$. On se fixe une probabilité de transition p sur \mathbb{Z}^d à support fini et supposée de moyenne $m = \sum_{x \in \mathbb{Z}^d} p(x)x \neq 0$ (tout ce qui suit peut être établi avec $m = 0$ mais le résultat obtenu est alors sans intérêt).

23.1 Marche aléatoires à temps continu

Une manière conventionnelle de se représenter une marches aléatoire à temps continu sur \mathbb{Z}^d est de considérer une famille $(Y_n)_{n \geq 0}$ de variables aléatoires indépendantes sur \mathbb{Z}^d de loi $\{p(z)\}_{z \in \mathbb{Z}^d}$, un processus de Poisson indépendants $(N_t)_{t \geq 0}$ de paramètre 1 et indépendant des Y_n et enfin une variable aléatoire indépendante du reste de variables aléatoires X_0 représentant la position initiale de la marche aléatoires. Le processus $(X_t)_{t \geq 0}$ est alors donné par :

$$X_t = X_0 + \sum_{k=1}^{N_t} Y_k$$

On note $p_t(x)$ la probabilité de transition de la marche aléatoire, définie par

$$p_t(x) = \mathbb{P}_0(X_t = x)$$

Ici la notation \mathbb{P}_z désigne la loi de la marche aléatoire partant de z (c'est-à-dire $X_0 = z$).

1) Montrer que

$$p_t(x) = \sum_{k=0}^{\infty} e^{-t} \frac{t^k}{k!} p^{*k}(x)$$

où p^{*k} est le produit de convolution k -ième de p .

2) Le but de cette question est d'établir la loi (faible) des grands nombres pour les marches aléatoires à temps continu. On se propose donc de montrer que

$$\lim_{t \rightarrow \infty} \frac{X_t}{t} = m \text{ p.s.}$$

a) Montrer que cette propriété impliquera que pour tout $\epsilon > 0$, et tout $t > 0$ fixé,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left[\left| \frac{X_{tN}}{N} - mt \right| \leq \epsilon \right] = 1$$

b) Montrer que \mathbb{P} p.s. $\lim_{t \rightarrow +\infty} N_t = +\infty$.

c) En utilisant la loi des grands nombres, montrer que \mathbb{P} p.s.,

$$\lim_{t \rightarrow \infty} \frac{N_t}{t} = 1$$

d) Conclure par la loi des grands nombres.

3) On considère N un entier fixé. Soit $(Z_t)_t$ une marche aléatoire partant de $z \in \mathbb{Z}^d$ de probabilité de transition p . Une marche aléatoire $(X_t)_t$ sur \mathbb{T}_N^d est définie par

$$X_t = Z_t \text{ mod } N$$

Montrer que pour une marche aléatoire $(X_t)_t$ sur le tore \mathbb{T}_N^d partant de 0, la probabilité de transition p_t^N est donnée par

$$p_t^N(z) = \mathbb{P}_0(X_t = z) = \sum_{u \in \mathbb{Z}^d} p_t(z + Nu)$$

23.2 Système de particules libres

Le processus que l'on cherche à modéliser est un système de particules évoluant de manière indépendantes selon des marches aléatoires sur \mathbb{T}_N^d . Les particules sont supposées indistinguables et seules leur localisation importe. La dynamique est construite de manière à conserver le nombre de particules mais rien ne nous empêche de considérer au départ un nombre aléatoire de particules. Soit donc K une variable aléatoire à valeurs dans \mathbb{N} . On se donne $(X^i, \dots, X_0^n, \dots)_{n \geq 1}$ une suite de marches aléatoires indépendantes de K à valeurs dans \mathbb{T}_N^d de même probabilité de transition p_t^N . Le processus $(\eta_t^K)_{t \geq 0}$ est alors défini par

$$\eta_t^K(x) = \sum_{i=1}^K \mathbb{1}_{\{X_t^i = x\}}$$

L'espace des configurations (ou états) de ce processus est $\mathbb{N}^{\mathbb{T}_N^d}$. A chaque instant t , et pour chaque site x de \mathbb{T}_N^d , $\eta_t(x)$ est le nombre de particules situées sur le site x . Nous appellerons ce processus un système de particules libres sur le tore.

1) Montrer que la loi du nombre de particules $\sum_{x \in \mathbb{T}_N^d} \eta_t(x)$ est une constante dans le temps.

Ainsi, si initialement, on a un nombre déterministe K de particules alors au temps t , on aura encore K particules (la dynamique ne connaît ni création, ni destruction de particules).

23.3 Mesures invariantes

Définition 23.3.1. Pour une fonction $\rho : \mathbb{T}^d \rightarrow \mathbb{R}_+$ lisse, on appelle mesure de Poisson sur $\mathbb{N}^{\mathbb{T}_N^d}$ associée à la fonction ρ la probabilité produit $\nu_\rho^N = \otimes_{x \in \mathbb{T}_N^d} \mu_x$ sur $\mathbb{N}^{\mathbb{T}_N^d}$ telle que la marginale μ_x soit une loi de Poisson de paramètre $\rho(x/N)$.

Pour tout $\alpha > 0$ fixé, on notera ν_α la mesure produit sur $\mathbb{N}^{\mathbb{Z}^d}$ dont les marginales sont des mesures de Poisson de paramètre α .

2) a) Montrer que la loi d'une variable aléatoire η à valeurs dans $\mathbb{N}^{\mathbb{T}_N^d}$ est caractérisée par sa transformée de Laplace multi-dimensionnel

$$L(\lambda) = \mathbb{E} \left[\exp \left\{ - \sum_{x \in \mathbb{T}_N^d} \lambda(x) \eta(x) \right\} \right]$$

où $\lambda : \mathbb{T}_N^d \rightarrow \mathbb{R}_+$ fonction quelconque.

b) Montrer que l'on a la propriété suivante :

Proposition 23.3.1. *Si les particules sont initialement distribuées selon une mesure de Poisson associée à une fonction constante $\alpha > 0$ alors au temps $t \geq 0$, elles sont distribuées selon la même mesure de Poisson associée à la fonction constante α . Autrement dit, les mesures de Poisson sont invariantes pour le système de particules.*

ind. : on calculera la transformée de Laplace de $(\eta_t(x))_{x \in \mathbb{T}_N^d}$

3) Quelle est la densité de particules lorsque le système est à l'équilibre sous une mesure de Poisson associée à la fonction α ?

4) Existe-t-il d'autres mesures invariantes pour le système de particules libres (N fixé) ?

La réponse nécessite un peu de connaissance sur les chaînes de Markov à temps continu. Le processus des particules libres est en fait un processus de Markov (à temps continu!) à espace d'états dénombrable. Supposons que p soit irréductible, c'est-à-dire que $\{z; p(z) > 0\}$ génère (en tant que groupe additif) tout \mathbb{Z}^d (sinon la réponse à la question est délicate). Alors, vu que le nombre de particules est conservées, partant d'une configuration initiale déterministe fixée, le nombre de particules K est une constante (non aléatoire fixée). Il y a un ensemble dénombrable de classe irréductibles (au moins) indexé par $K \in \mathbb{N}$ (représentant le nombre de particules) que l'on note $(\mathcal{C}_K)_K$.

Puisque p est supposée irréductible, sur chacune des classes précédentes \mathcal{C}_K , le système de particules libres est lui-même un processus markovien à espace d'états fini et irréductible (l'espace d'état du processus est alors en fait $\{0, \dots, K\}^{\mathbb{T}_N^d}$ qui est fini). Chaque classe irréductible \mathcal{C}_K possède donc une seule mesure invariante μ_K .

Les mesures invariantes du système sont donc données par les combinaisons convexes des $(\mu_K)_{K \in \mathbb{N}}$. Les mesures de Poissons précédentes ne sont que des cas particuliers. Cependant, l'ensemble des mesures invariantes forment un ensemble convexe et on peut montrer que les point extrémaux de cet ensemble sont les mesures de Poisson.

23.4 Equivalence des ensembles

On a vu dans la section précédente que les mesures de Poisson étaient invariantes pour la dynamique. Néanmoins, puisque le nombre de particules est conservées, les mesures de Poissons conditionnées à un nombre de particules fixé sont aussi invariantes. Néanmoins, les deux points de vue sont équivalents lorsque l'on regarde ce qui se passe pour N tend vers l'infini.

Cette équivalence est connue, dans un contexte plus général, sous le terme de ce que les physiciens appellent l'équivalence des ensembles. L'équivalence des ensembles établit une correspondance entre les mesures dites grand-canonique (les mesures de Poisson pour nous) et les mesures canoniques (les mesures de Poisson conditionnées à un nombre fixé de particules).

On considère une fonction constante notée α égale à α et on considère la mesure $\mu_{\alpha,K}^N$ sur $\{\eta \in \mathbb{N}^{\mathbb{T}_N^d}; \sum_{x \in \mathbb{T}_N^d} \eta(x) = K\}$ donnée par

$$\mu_{\alpha,K}^N = \nu_{\alpha}^N(\cdot | \sum_{x \in \mathbb{T}_N^d} \eta(x) = K)$$

1) Montrer que cette mesure ne dépend pas de α et est en fait la mesure multinomiale sur $\{0, \dots, K\}^{\mathbb{T}_N^d}$.

2) Montrer que c'est une mesure invariante du système de particules libres (c'est en fait ce que l'on a noté μ_K^N).

3) Soit $\beta > 0$ un paramètre positif fixé. Montrer que pour tout entier $r > 0$ fixé, pour toute suite $(k_1, \dots, k_r) \in \mathbb{N}^r$ et $(x_1, \dots, x_r) \in \mathbb{Z}^r$,

$$\lim_{N \rightarrow \infty} \mu_{\alpha, N^d \beta}(\eta(x_1) = k_1, \dots, \eta(x_r) = k_r) = \nu_{\beta}(\eta(x_1) = k_1, \dots, \eta(x_r) = k_r)$$

On a un résultat similaire assez simple d'équivalence des ensemble dans un cadre continu (ici entre les mesures micro-canoniques et les mesures canoniques). On considère des particules libres $1, \dots, x, \dots, N$ indiscernables. La particule x possède la quantité de mouvement p_x , et n'interagit pas avec les autres. L'énergie de l'ensemble des particules est donc uniquement donnée par son énergie cinétique $\sum_{x=1}^N p_x^2$ qui est conservée lors du mouvement des particules. Supposons que notre système soit isolé et à l'équilibre. Soit E l'énergie du système. Le principe d'équipartition de l'énergie nous dit que les vitesses $(p_x)_{x=1, \dots, N}$ se répartissent de manière uniforme sur la surface d'énergie. On cherche à savoir ce qui se passe pour une observable locale à la limite thermodynamique (i.e. N tend vers l'infini). En termes mathématiques, la question est la suivante.

Soit $f : \mathbb{R}^k \rightarrow \mathbb{R}$ une fonction continue bornée avec k un entier fixé et $E > 0$ un réel. Pour tout N plus grand que k , on introduit $\nu_{N,E}$ la mesure uniforme sur la sphère S_E^{N-1} de \mathbb{R}^N de rayon \sqrt{NE} :

$$S_E^{N-1} = \{(x_1, \dots, x_N); \sum_{i=1}^N x_i^2 = NE\}$$

On rappelle que cette mesure est l'unique probabilité sur S_E^{N-1} stable par les rotations de \mathbb{R}^N .

1) Soit $(\zeta_1, \dots, \zeta_N)$ N variables normales standard uniformes. Montrer à l'aide de cette caractérisation

que la loi de $Y = \frac{\sqrt{NE}}{\sqrt{\sum_{i=1}^N \zeta_i^2}}(\zeta_1, \dots, \zeta_N)$ est $\nu_{N,E}$.

2) A l'aide de la loi des grands nombres, montrer que

$$\lim_{N \rightarrow \infty} \nu_{N,E}(f(x_1, \dots, x_k)) = \mu_{k,E}(f)$$

où $\mu_{k,E}$ est une loi Gaussienne sur \mathbb{R}^k de moyenne nulle et de matrice de covariance $E^{-1}Id$ (i.e. la distribution de Maxwell avec température E).

23.5 Equilibre local

Définition 23.5.1. Soit $(\mu_N)_{N \geq 1}$ une suite de probabilités sur $\mathbb{N}^{\mathbb{T}_N^d}$. On dit que c'est un équilibre local de profil $\rho_0 : \mathbb{T}^d \rightarrow \mathbb{R}_+$ (fonction lisse) ssi pour tout $u \in \mathbb{T}^d$, et toute fonction $f : \mathbb{N}^{\mathbb{Z}^d} \rightarrow \mathbb{R}$ ne dépendant que d'un nombre fini de sites,

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\mu^N} [f(\tau_{[uN]}\eta)] = \mathbb{E}_{\nu_{\rho_0(u)}}(f)$$

On se propose de démontrer le théorème suivant, dit de conservation de l'équilibre local :

Théorème 23.5.1. On considère une fonction $\rho_0 : \mathbb{T}^d \rightarrow \mathbb{R}_+$ lisse. Soit $(\mu^N)_{N \geq 1}$ un équilibre local associé au profil ρ_0 . Pour tout $t \geq 0$, on note μ_t^N la loi du processus η_t initialement distribué selon μ^N . Pour tout $t \geq 0$, $(\mu_{tN}^N)_{N \geq 1}$ est un équilibre local de profil $\rho_t = \rho(t, \cdot)$ où ρ est solution de l'équation aux dérivées partielles suivantes

$$\partial_t \rho + m \cdot \nabla \rho = 0$$

avec condition initiale ρ_0 .

1) Montrer que

$$\mathbb{E}_{\nu_{\rho_0}^N} \left(\exp - \sum_{x \in \mathbb{T}_N^d} \lambda(x) \eta_t(x) \right) = \exp \left(\sum_{y \in \mathbb{T}_N^d} (e^{-\lambda(y)} - 1) \psi_{N,t}(y) \right)$$

où

$$\Psi_{N,t}(y) = \sum_{x \in \mathbb{T}_N^d} p_t^N(y-x) \rho_0(x/N)$$

2) Montrer que

$$\lim_{N \rightarrow \infty} \psi_{N,t}([uN]) = \rho_0(u)$$

Cette propriété montre que le temps microscopique est trop court pour atteindre un équilibre local et est à relier à l'hypothèse $t \sim \tau_0 \ll \tau_R$.

3) Montrer que pour tout $\epsilon > 0$,

$$\lim_{N \rightarrow \infty} \sum_{x; |x/N - mt| \leq \epsilon} p_{tN}(x) = 1$$

4) Conclure.

23.6 Quelques remarques

Nous avons ici traité le cas de marches aléatoires de moyenne non nulle, non pas parce que pour $m = 0$, le résultat est faux mais parce que pour $m = 0$, il ne dit rien d'intéressant (le profil ne change pas dans l'échelle de temps tN). Pour le cas $m = 0$, il faut considérer une échelle de temps plus grande en tN^2 . Dans ce cas, le profil qui apparaît est décrit par une équation aux dérivées partielles du second ordre. Plus exactement, on introduit la matrice σ définie par

$$\sigma_{i,j} = \sum_{x \in \mathbb{Z}^d} x_i x_j p(x)$$

En utilisant un théorème limite centrale pour les marches aléatoires à temps continu, on obtient le théorème suivant.

Théorème 23.6.1. *Supposons qu'un système de particules évolue selon des marches aléatoires indépendantes avec une moyenne $m = 0$ et qu'initialement on ait un équilibre local associé à un profil lisse ρ_0 . Dans une échelle de temps diffusive (c'est-à-dire en tN^2), il y a conservation de l'équilibre local et l'évolution du profil est donnée par la solution de l'équation aux dérivées partielles suivantes :*

$$\begin{cases} \partial_t \rho = \sum_{1 \leq i, j \leq d} \sigma_{i,j} \partial_{u_i, u_j}^2 \rho \\ \rho(0, u) = \rho_0(u) \end{cases}$$

Dans la théorie des limites hydrodynamiques, l'équation qui apparaît ici est appelée l'équation hydrodynamique du système. On peut considérer des cas plus complexes que celui-ci. Par exemple, on peut imposer aux particules d'évoluer comme des marches aléatoires à temps continu mais refuser que plus d'une particule occupe un site (on parle alors du processus d'exclusion simple). Là aussi, une équation hydrodynamique peut-être établie. Selon que p est à moyenne nulle ou pas, les échelles de temps à considérer ne sont pas les mêmes (échelle diffusive en N^2 ou Eulérienne en N) et les équations hydrodynamiques obtenues sont différentes (équation type équation de la chaleur ou équation de Burgers).

Chapitre 24

Simulations de variables aléatoires et applications

24.1 La méthode MCMC

Ce cours-td est issu de quatre sources ([1], [8],[19] et [25]). Le meilleur exposé est sans nul doute [19]. Néanmoins, pour ce qui est des motivations de l'étude du modèle d'Ising, [25] est nécessaire même si [8] l'aborde d'une manière moins agréable. Ce qui concerne les mesures de Gibbs en volume infini est donné sans preuves (cf. [25] pour de plus amples détails). La seule raison d'exposer ces résultats est de donner une motivation non dénuée d'intérêt à la simulation d'une mesure de Gibbs. D'autre part, la "convergence" de l'algorithme de Propp-Wilson n'est pas donné dans [19] mais il l'est dans [1], exercice 4.38 p.123.

24.1.1 Les méthodes classiques

On peut distinguer trois grandes méthodes de simulations de variables aléatoires.

1. Méthode de l'inverse : On cherche à simuler une variable aléatoire Z de fonction de répartition F . Pour ce faire, on simule une variable aléatoire uniforme sur $[0, 1]$. Soit F^{-1} la pseudo inverse de F alors $F^{-1}(U)$ a pour loi F .
2. Méthode de rejet : On cherche à simuler une variable aléatoire Z ayant une densité f . On suppose que l'on est capable de générer une suite i.i.d. de variables aléatoires $(Y_n)_n$ ayant une loi à densité g telle que

$$\frac{f(x)}{g(x)} \leq c$$

Soit $(U_n)_n$ une suite i.i.d. de variables aléatoires uniformément distribuées sur $[0, 1]$. On définit τ comme le premier instant n pour lequel on a

$$U_n \leq \frac{f(Y_n)}{cg(Y_n)}$$

et

$$Z = Y_\tau$$

Alors Z est une loi ayant pour densité f .

3. Méthode MCMC (Monte Carlo Markov Chain) : c'est celle que nous allons décrire plus en détail

maintenant.

Avant de décrire la méthode MCMC (dans le cas d'une loi discrète), commençons par donner quelques critiques en ce qui concerne les deux premières méthodes. Les méthodes de l'inverse et de rejet ont des contre-parties quand Z est une variable discrète de loi π sur un ensemble fini E supposé très grand.

La méthode de l'inverse consiste à générer une variable uniforme U sur $[0, 1]$ et à poser $Z = i$ ssi $\sum_{l=1}^{i-1} \pi_l \leq U < \sum_{l=1}^i \pi_l$. Quand la taille de E est très grande, on a une partition de l'intervalle $[0, 1]$ constituée de petits intervalles. Savoir où U se trouve exactement demande donc un grand coût en précision de calculs. Dans une simulation où E n'est pas une suite d'entiers mais un espace de configurations plus ou moins compliqué, il est nécessaire de définir un codage de l'espace des configurations (et de coder et décoder) ce qui peut être assez lourd.

La méthode de rejet dans le cas discret utilise une distribution p sur E telle que $\frac{\pi_i}{p_i} \leq c$ pour tout $i \in E$. Le problème vient du fait que bien souvent π est connue à une constante de normalisation près si bien que la méthode de rejet (tout comme la méthode de l'inverse) devient inutilisable.

24.1.2 Le principe de la méthode MCMC

La méthode MCMC est la suivante. On construit une chaîne de Markov irréductible aperiodique sur E avec π pour distribution stationnaire. Puisque E est fini, la chaîne est ergodique et pour tout $i \in E$ et toute distribution initiale μ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}_\mu(X_n = i) = \pi_i$$

Aussi, quand n est "grand, on peut considérer que X_n a une distribution proche de π . Bien entendu, il se pose deux problème majeur dans cette méthode :

- 1) Comment construire une chaîne X ayant π pour distribution stationnaire ?
- 2) Comment évaluer si la distribution de X_n est proche de π ? Ou plus exactement, comment obtenir une estimation de la vitesse de convergence de la loi de X_n vers π en fonction de n ?

Un élément de réponse à la première question est fournie par les algorithmes de Métropolis (1953) et Barker (1965). On remarquera au passage que ce dernier algorithme ne nécessite la connaissance de π qu'à un facteur multiplicatif près. Pour ce qui est de la deuxième question, on sait que la vitesse de convergence est géométrique et donnée par la deuxième plus grande valeur propre de la matrice de transition. Tout revient alors à obtenir une bonne estimation de cette seconde valeur propre... On trouvera tous les compléments utiles dans le chapitre 6 de [8].

Exercice 24.1.1. Algorithme de Metropolis ([1] exercice 4.37)

Soit E un ensemble d'états fini et P une matrice de transition sur E symétrique et irréductible. Soit π une probabilité sur E telle que $\pi(x) > 0$ pour tout $x \in E$. On définit sur E une nouvelle matrice de transition Q par

$$Q(x, y) = \begin{cases} P(x, y) & \text{si } x \neq y \text{ et } \pi(y) \geq \pi(x) \\ P(x, y) \frac{\pi(x)}{\pi(y)} & \text{si } \pi(y) < \pi(x) \\ 1 - \sum_{z \neq x} Q(x, z) & \text{si } y = x \end{cases}$$

a) Montrer que la loi π est réversible pour Q et donc invariante.

b) Montrer que Q est irréductible.

c) On suppose que π n'est pas la distribution uniforme (sinon on aurait $Q = P$). On note M l'ensemble des états $x \in E$ tels que $\pi(x) = \sup_{y \in E} \pi(y)$.

c1) Montrer qu'il existe $x_0 \in M$ tel que $P(x_0, y) > 0$ pour un certain $y \notin M$. En déduire que $Q(x_0, z) < P(x_0, z)$ pour un $z \in E$ au moins. Montrer que $Q(x_0, x_0) > 0$.

c2) Montrer que Q est apériodique.

On a donc montré que π est réversible pour Q donc invariante et que Q est irréductible et apériodique. Par le théorème ergodique, on est assuré que pour n assez grand, la loi de la chaîne de Markov $(X_n)_n$ de matrice de transition Q est proche de π . Cette méthode de simulation est particulièrement intéressante lorsque l'on ne connaît π qu'à une constante multiplicative près (c'est ce qui arrive fréquemment en mécanique statistique où la fonction de répartition est inconnue, ou du moins si E est si grand que la calculer a un coût trop grand). Une excellente application, malheureusement en dehors des thèmes du programme et assez difficile mais faisable, de l'algorithme de Metropolis est l'illustration de la transition de phase dans le modèle d'Ising (Cf. [8], [25]).

24.1.3 Simulation d'une chaîne de Markov

La question est la suivante. Comment simuler une chaîne de Markov (X_0, \dots, X_n, \dots) à espace d'états fini $S = \{s_1, \dots, s_k\}$, de matrice de transition P , avec la distribution μ pour loi initiale. Nous admettons que nous pouvons simuler une suite (U_0, \dots, U_n, \dots) (en fait, une suite finie...) de variables aléatoires uniformes indépendantes sur $[0, 1]$. Les deux ingrédients principaux sont la fonction d'initiation ψ et la fonction de mise à jour ϕ . On les définit de la manière suivante.

$$\psi(x) = \begin{cases} s_1 & \text{pour } x \in [0, \mu(s_1)) \\ s_2 & \text{pour } x \in [\mu(s_1), \mu(s_1) + \mu(s_2)) \\ \cdot & \\ \cdot & \\ s_i & \text{pour } x \in [\sum_{j=1}^{i-1} \mu(s_j), \sum_{j=1}^i \mu(s_j)) \\ s_k & \text{pour } x \in [\sum_{j=1}^{k-1} \mu(s_j), 1] \end{cases} \quad (24.1.1)$$

et pour $i \in \{1, \dots, k\}$,

$$\phi(s_i, x) = \begin{cases} s_1 & \text{pour } x \in [0, P_{i,1}) \\ s_2 & \text{pour } x \in [P_{i,1}, P_{i,1} + P_{i,2}) \\ \cdot & \\ \cdot & \\ s_j & \text{pour } x \in [\sum_{l=1}^{j-1} P_{i,l}, \sum_{l=1}^j P_{i,l}) \\ s_k & \text{pour } x \in [\sum_{l=1}^{k-1} P_{i,l}, 1] \end{cases} \quad (24.1.2)$$

Soit (U_0, \dots, U_n, \dots) une suite de variables aléatoires uniformes sur $[0, 1]$ indépendantes. On

pose $X_0 = \psi(U_0), X_1 = \phi(X_0, U_1), \dots, X_n = \phi(X_{n-1}, U_n), \dots$.

1) Montrer que $(X_n)_n$ est une chaîne de Markov de matrice de transition P , de loi initiale μ . 2) En déduire un algorithme pour simuler une chaîne de Markov donnée.

24.1.4 Un modèle d'exclusion

Soit $\Lambda_m = \{1, \dots, m\}^2$ un carré dans \mathbb{Z}^2 de longueur m . Une configuration ζ sur Λ_m est une fonction de Λ_m dans $\{0, 1\}$. Une configuration ζ est dite faisable si pour tout x et y de Λ_m , $|x - y| = 1$ implique $\zeta(x)\zeta(y) = 0$ (autrement dit, on ne peut pas avoir deux sites adjacents occupés avec des 1). Soit Z_m le nombre de configurations faisables. On définit une probabilité μ_m sur l'espace $\Omega_m = \{0, 1\}^m$ des configurations par :

$$\mu_m(\zeta) = Z_m^{-1} \mathbb{1}_{\{\zeta \text{ faisable}\}} \quad (24.1.3)$$

1) Soit $n(\zeta)$ le nombre de 1 dans la configuration ζ . On s'intéresse à l'estimation de $\mu(n)$. Pour cela, on doit simuler une variable aléatoire X ayant pour loi (ou approximativement) la loi μ . Admettons que l'on dispose d'une procédure permettant de simuler une telle variable aléatoire. Comment faire pour estimer (en terme de simulation) la quantité $\mu(n)$?

2) On se propose maintenant de donner un algorithme MCMC pour ce modèle d'exclusion, i.e.

simuler par la méthode MCMC une variable aléatoire X ayant pour loi μ_m .

a) Pourquoi une méthode MCMC s'avère nécessaire pour m grand ?

b) Vérifier que l'algorithme suivant convient :

1. Choisir un site s dans Λ_m de manière aléatoire et uniforme.
 2. Simuler une variable aléatoire de Bernoulli U de paramètre $1/2$.
 3. Si $U = 1$ et si tous les voisins de s ont pour valeur 0 alors mettre la nouvelle valeur de s à 1, sinon mettre à 0.
 4. Pour les autres sites (i.e. différents de s), ne rien faire.
- c) Le simuler.

3) Donner une interprétation en termes de "modèle de télécommunication" et de "dynamique des populations" de cette mesure.

24.2 Critique de la méthode MCMC

On peut adresser deux grandes critiques à la méthodes MCMC.

A) L'argument théorique qui fonde la méthode MCMC est le théorème qui dit que la loi μ_n au temps n d'une chaîne de Markov irréductible aperiodique (partant de n'importe quel état) converge vers sa distribution stationnaire π (distribution que l'on cherche à simuler). Bien entendu, ce n'est qu'une convergence et cela n'implique pas l'existence d'un n pour lequel $\mu_n = \pi$. Il y a donc toujours une erreur lorsque l'on approxime π par μ_n .

B) Pour rendre cette erreur petite, on doit être capable d'une manière ou d'une autre d'estimer cette erreur (erreur que l'on mesure au moyen de la distance en variation).

Le problème soulevé en A) est un problème inhérent à toute simulation numérique et on s'en accomode aisément, d'autant plus que l'ordinateur lui-même ne donne que des résultats approximatifs (quelque soit la précision dont il dispose, \mathbb{R} lui reste terra incognita). Le second problème B) est quant à lui beaucoup plus sérieux. Il s'est avéré qu'estimer cette erreur (i.e. majorer $d(\mu_n, \pi)$) était très difficile. Nous disposons cependant de résultats dans ce sens (théorème 8.1 p.56 de [19] par exemple). En pratique, on opère en général de la manière suivante. On simule la chaîne de Markov jusqu'à un temps n de l'ordre 10^5 et on croise les doigts ! Dans la section suivante, nous présentons un algorithme "parfait" dans le sens où il simule une loi qui est exactement égale à la distribution π cherchée. Bien entendu, cet algorithme possède lui aussi des désavantages que nous expliciterons.

24.3 L'algorithme de Propp-Wilson

On adopte les notations classiques de ce chapitre. Soit une chaîne de Markov (X_0, \dots, X_n, \dots) à espace d'états fini $S = \{s_1, \dots, s_k\}$, de matrice de transition P , avec la distribution μ pour loi initiale. La fonction de mise à jour est notée $\phi : S \times [0, 1] \rightarrow S$.

On considère une suite de nombre $N_1 = 1, \dots, N_m = 2^{m-1}, \dots$. Soit $(U_0, U_{-1}, \dots, U_{-n}, \dots)$ une suite de variables aléatoires uniformes indépendantes sur $[0, 1]$. L'algorithme est le suivant :

1. Prendre $m = 1$.
2. Pour tous les $s \in \{s_1, \dots, s_k\}$, simuler la chaîne de Markov partant de $-N_m$ dans l'état s jusqu'au temps 0 en utilisant ϕ et les variables aléatoires $U_{-N_m+1}, \dots, U_{-1}, U_0$ (ce sont les mêmes pour les k trajectoires simulées).
3. Si les k chaînes dans l'étape 2 se terminent dans le même état s' au temps 0 alors l'algorithme se termine et donne s' . Autrement, il passe à l'étape 4.
4. Incrémenter m de 1 et continuer avec l'étape 2.

Quelques commentaires sur cet algorithme. Première chose, il est important de dire que à la m -ème étape, l'algorithme réutilise certaines variables aléatoires $(U_{-N_{m-1}+1}, \dots, U_0)$ de l'étape $m-1$. Ceci est économique en appel de simulation de variables aléatoires uniformes mais coûteux en mémoire puisque l'on doit conserver en mémoire les variables aléatoires du passé. Néanmoins, ceci est nécessaire pour que l'algorithme marche (un exemple est donné dans [19] où l'algorithme ne fonctionne pas si l'on ne réutilise pas les anciennes uniformes). Deuxièmement, a priori, l'algorithme contient une boucle infinie. Il peut très bien ne jamais terminer. En fait, un résultat (que nous ne prouverons pas mais que l'on peut trouver dans [1], exercice 4.38) dit que l'on peut choisir une bonne fonction ϕ , que nous dirons *valide*, telle que avec probabilité 1, l'algorithme termine toujours. Mais un mauvais choix de la fonction ϕ peut conduire à une boucle infinie (cf. le problème élémentaire 10.2 de [19]). Troisièmement, il peut sembler étrange de partir dans le "passé", en plus à des temps en -2^m . Boh ! Nous passons maintenant à la preuve du théorème suivant :

Théorème 24.3.1. *Soit P la matrice de transition d'une chaîne de Markov irréductible et aperiodique à espace d'états $S = \{s_1, \dots, s_k\}$ et ayant pour distribution stationnaire $\pi = (\pi_1, \dots, \pi_k)$. Soit ϕ une fonction de mise à jour valide pour P et considérons l'algorithme de Propp-Wilson. On suppose que l'algorithme se termine avec probabilité 1 et on note Y la sortie de l'algorithme. Alors pour tout $i \in \{1, \dots, k\}$,*

$$\mathbb{P}(Y = s_i) = \pi_i$$

1) Fixons $s_i \in S$. Il suffit de prouver que pour tout $\epsilon > 0$, on a

$$|\mathbb{P}(Y = s_i) - \pi_i| \leq \epsilon$$

Montrer qu'il existe M tel que

$$\mathbb{P}(\text{l'algorithme n'a pas besoin de faire plus de } M \text{ étapes}) \geq 1 - \epsilon$$

2) On fixe un tel M . On considère une chaîne de Markov \tilde{Y} partant du temps $-N_M$ avec la même fonction de mise à jour ϕ et la même suite d'uniformes U_{-N_M+1}, \dots, U_0 mais avec pour loi initiale (au temps $-N_M$) la probabilité stationnaire π . Montrer que

$$\mathbb{P}(Y \neq \tilde{Y}_0) \leq \epsilon$$

3) En déduire $|\mathbb{P}(Y = s_i) - \pi_i| \leq \epsilon$.

Passons maintenant à la critique de cet algorithme. Nous nous intéressons à la simulation de distributions π sur S lorsque le cardinal k de S est très grand. L'algorithme de Propp-Wilson propose une solution qui demande de simuler à chaque étape k trajectoires de la chaîne. L'algorithme semble donc a priori inutilisable... Cependant, nous allons voir que si la chaîne considérée possède certaines propriétés de monotonie alors on peut réduire considérablement le nombre de trajectoires à simuler (en appliquant le principe du sandwich). L'algorithme devient alors intéressant. Nous allons commencer par un cas instructif (mais sans aucun intérêt pratique) puis passer à un cas réellement intéressant (modèle d'Ising).

Exercice 24.3.1. On désire simuler la probabilité uniforme π sur $\{1, \dots, k\}$ avec l'algorithme de Propp-Wilson en appliquant le principe du sandwich.

1) Montrer que la marche aléatoire simple symétrique sur $\{1, \dots, k\}$ avec barrière réfléchissante en 1 et en k admet π pour probabilité stationnaire.

2) Montrer que la fonction ϕ définie par :

$$\begin{aligned} \phi(1, \cdot) &= \mathbb{1}_{[0,1/2)} + 2 \cdot \mathbb{1}_{[1/2,1]} \\ \phi(k, \cdot) &= (k-1) \cdot \mathbb{1}_{[0,1/2)} + k \cdot \mathbb{1}_{[1/2,1]} \\ \phi(i, \cdot) &= (i-1) \cdot \mathbb{1}_{[0,1/2)} + (i+1) \cdot \mathbb{1}_{[1/2,1]}, \quad i \in \{2, \dots, k-1\} \end{aligned}$$

est une fonction de mise à jour.

3) Montrer que la fonction ϕ vérifie

$$i \leq j \Rightarrow \phi(i, \cdot) \leq \phi(j, \cdot)$$

4) En déduire que toute trajectoire démarrant dans un état $i \in \{2, \dots, k-1\}$ reste entre les trajectoires partant respectivement de 1 et de k . Expliquer le terme "principe du sandwich".

5) Proposer une nouvelle version de l'algorithme de Propp-Wilson qui économise le nombre d'opérations.

6) Implémenter l'algorithme.

24.4 Le modèle d'Ising

Soit S un ensemble fini de \mathbb{Z}^2 . Les points $s \in S$ sont appelés les sites. On introduit aussi l'espace des états $\Lambda = \{\pm 1\}$ et l'espace des configurations $E = \Lambda^S$.

Définition 24.4.1. *Un champ aléatoire sur S est une collection $X = \{X(s)\}_{s \in S}$ de variables aléatoires à valeurs dans Λ , où si l'on préfère une variable aléatoire à valeurs dans Λ .*

On dira qu'un site $y \in S$ est un voisin de $x \in S$ si et seulement si la distance (euclidienne standard) de x à y est égale à 1 : $|x - y| = 1$. On notera alors $x \sim y$. L'ensemble des voisins de x dans S est noté $\mathcal{N}_x(S) = \{y \in S; |y - x| = 1\}$ et on pose $\tilde{\mathcal{N}}_x(S) = \mathcal{N}_x(S) \cup \{x\}$. On remarquera que si $S = \mathbb{Z}^2$ alors pour tout x , le cardinal de $\mathcal{N}_x(\mathbb{Z}^2)$ est 4.

Définition 24.4.2. *On dira que le champ aléatoire X est un champ aléatoire Markovien sur S si il vérifie : pour toute configuration $x \in E$ et pour tout site $s \in S$*

$$\mathbb{P}(X(s) = x(s) | X(S \setminus s) = x(S \setminus s)) = \mathbb{P}(X(s) = x(s) | X(\mathcal{N}_s) = x(\mathcal{N}_s)) \quad (24.4.1)$$

Dans ce cas, la fonction $\pi^s : E \rightarrow [0, 1]$ définie par $\pi^s(x) = \mathbb{P}(X(s) = x(s) | X(\mathcal{N}_s) = x(\mathcal{N}_s))$ est appelée spécification locale du champ.

Sous une condition de positivité assez faible sur la spécification locale, on peut montrer qu'il existe au plus une seule distribution de champ aléatoire Markovien pour une spécification locale donnée.

24.4.1 Distribution de Gibbs

Définition 24.4.3. *On appelle clique sur S tout singleton $\{s\}$ avec $s \in S$ ou bien toute paire $\{s, t\}$ de voisins $s, t \in S$. Une clique à deux éléments est dite maximale (ou bien non triviale).*

Définition 24.4.4. *Un potentiel V sur $E = \Lambda^S$ est une collection $\{V_C\}_{C \subset S}$ de fonctions de E dans \mathbb{R} telles que :*

i) $V_C = 0$ si C n'est pas une clique.

ii) V_C ne dépend de x qu'à travers les valeurs de $x(C)$ (i.e. V_C est à portée dans C).

La fonction énergie \mathcal{E} associé à un tel potentiel est définie par :

$$\mathcal{E}(x) = \sum_{C \subset S} V_C(x), \quad x \in E$$

Soit $T > 0$. La distribution π_T définie par

$$\pi_T(x) = Z_T^{-1} \exp\left(-\frac{\mathcal{E}(x)}{T}\right), \quad x \in E$$

où Z_T est une constante de normalisation (dite fonction de partition) est appelée distribution de Gibbs à température T associée au potentiel V .

Exemple : Modèle d'Ising. C'est le modèle le plus simple que l'on puisse concevoir. On prend $S = \{1, \dots, m\}^2$ et pour potentiel V ,

$$V_{\{s\}}(x) = -Hx(s), \quad V_{\{s,t\}}(x) = -Jx(s)x(t)$$

où $H \in \mathbb{R}$ est le champ magnétique externe et $J \in \mathbb{R}^+$ est l'énergie interne d'un dipôle magnétique.

Théorème 24.4.1. *Les distributions de Gibbs associées à un potentiel $V = \{V_C\}_{C \subset S}$ et à une température T sont des champs Markoviens. Plus exactement, soit X un champ aléatoire dont la loi est donnée par la distribution π_T de la définition précédente, alors X est un champ Markovien et ses spécifications locales sont données par :*

$$\pi^s(x) = \frac{\exp(-T^{-1} \sum_{C \ni x} V_C(x))}{\sum_{\lambda \in \Lambda} \exp(-T^{-1} \sum_{C \ni x} V_C(\lambda, x(S \setminus \{s\})))} \quad (24.4.2)$$

Remarque 24.4.1. *La réciproque est quasi-vraie (il faut rajouter la condition de positivité dont on a parlé succinctement). On pourra consulter [8] pour une démonstration (où l'on verra une application inattendue de la formule de Möbius...).*

24.4.2 Distribution de Gibbs en volume infini et transition de phase

On cherche à étendre la notion de mesure de Gibbs lorsque S est infini : $S = \mathbb{Z}^2$). Il est naturel, compte-tenu de la définition (24.4.1) de dire qu'une probabilité μ sur l'espace des configurations $\{\pm 1\}^S$ est une mesure de Gibbs associée au potentiel V à température $T > 0$ si elle vérifie les équations (dites DLR) suivantes :

$$\mu(X(s) = x(s) | X(S \setminus s) = x(S \setminus s)) = \frac{\exp(-T^{-1} \sum_{C \ni x} V_C(x))}{\sum_{\lambda \in \Lambda} \exp(-T^{-1} \sum_{C \ni x} V_C(\lambda, x(S \setminus \{s\})))}$$

On dénote par $\mathcal{G}(T)$ l'ensemble des mesures de Gibbs associée à ce potentiel à température T . Pour simplifier, on prendra le potentiel associé au modèle d'Ising avec le champ magnétique externe H égal à 0.

On cherche maintenant à caractériser les éléments de \mathcal{G} . Pour cela, nous avons besoin d'introduire, en volume fini (S fini), les mesures de Gibbs avec conditions aux bords qui généralisent les mesures de Gibbs introduites précédemment. Soit S une ensemble fini et W un sous-ensemble de S (en général l'intérieur de S). On se donne une configuration ζ sur $S \setminus W$, c'est-à-dire un élément de $\{0, 1\}^{S \setminus W}$. La mesure de Gibbs $\pi_{T, \zeta}$ avec conditions aux bords ζ sur $S \setminus W$ à température $T > 0$ est définie par :

$$\forall \eta \in \{-1, 1\}^W, \quad \pi_{T, \zeta}(\eta) = Z_{T, \zeta}^{-1} \exp\left(-\frac{\mathcal{E}(\eta^\zeta)}{T}\right)$$

where $Z_{T, \zeta}$ est une fonction de normalisation et η^ζ est la configuration sur $\{-1, +1\}^S$ obtenue en concaténant η et ζ .

Deux cas particulièrement importants de mesures de Gibbs avec conditions aux bords peuvent être considérés. Prenons $S_m = \{1, \dots, m\}^2$ où $m \geq 2$ et $W_m = \{2, \dots, m-1\}^2$. On prend ζ la configuration sur $\{-1, 1\}^{S_m \setminus W_m}$ composée uniquement de +1 (resp. de -1) et on note la mesure $\pi_{T, \zeta}$ obtenue $\pi_{T, m}^+$ (resp. $\pi_{T, m}^-$).

On peut alors démontrer le théorème suivant qui sert à caractériser l'ensemble des mesures de Gibbs en volume infini.

Théorème 24.4.2. *On considère le modèle d'Ising à température T .*

1. *La limite (faible) de $(\pi_{T,m}^+)_m$ (resp. de $(\pi_{T,m}^-)_m$) existe. On la note π_T^+ (resp. π_T^-).*
2. *π_T^\pm sont des éléments de \mathcal{G} .*

On dit qu'il y a transition de phases lorsque l'ensemble \mathcal{G}_T contient plus d'un élément. D'après le théorème précédent, on voit que si $\pi_T^+ \neq \pi_T^-$, il y a transition de phase et on peut même démontrer la réciproque. On peut aller encore plus loin avec le théorème suivant.

Théorème 24.4.3. *Il y a transition de phase si et seulement si $\pi_T^+(\eta(0) = 1) \neq \pi_T^-(\eta(0) = 1)$. D'autre part il existe une température critique T_c telle que si $T > T_c$, il n'y a pas transition de phase et si $T < T_c$, il y a transition de phase.*

Finissons par quelques remarques. Nous avons ici considéré le cadre de la dimension $d = 2$. Tout ceci se généralise en dimension d quelconque. On peut démontrer que pour $d = 1$, $T_c = 0$ (il n'y a jamais transition de phase). Pour $d = 2$, on connaît la valeur exacte de T_c , $T_c = 2 [\log(1 + \sqrt{2})]$ (pour $J = 1$). En dimension $d \geq 3$, c'est un problème ouvert. Le fait qu'à haute température, il y ait transition de phase est assez facile à montrer (argument de Peierls).

24.4.3 Simulation d'une distribution de Gibbs par la méthode MCMC

On considère un processus $(X_n)_{n \in \mathbb{N}}$ à valeurs dans l'espace des configurations E . Autrement dit, pour tout $n \in \mathbb{N}$, X_n est une champ aléatoire. Supposons qu'il existe une chaîne de Markov irréductible aperiodique $(X_n)_{n \in \mathbb{N}}$ à valeurs dans E telle que sa distribution stationnaire soit donnée par π_T . Alors on a

$$\lim_{n \rightarrow \infty} d_V(\mathbb{P}_{X_n}, \pi_T) = 0 \quad (24.4.3)$$

où $d_V(\mu, \nu)$ est la distance en variation entre les deux probabilités μ et ν définie par

$$d_V(\mu, \nu) = \sum_{i \in E} |\mu(i) - \nu(i)|$$

On trouvera une preuve de ceci dans [8] ou [19]. Cette propriété donne une idée de comment simuler une variable aléatoire X ayant la probabilité π_T . La difficulté qui se présente lorsque l'on désire simuler une variable aléatoire ayant pour loi π_T est que l'on ne connaît pas la fonction de partition Z_T et que le calcul de celle-ci, lorsque E est très grand (de l'ordre de $2^{|S|}$), s'avère coûteux en temps et en précision.

L'échantillonnage de Gibbs est le suivant. C'est en fait une version de l'algorithme de Metropolis. On se donne une probabilité $(q_s)_{s \in S}$ strictement positive (par exemple $q_s = 1/|S|$, i.e. la distribution uniforme sur S). La transition de probabilité qui permet de passer de $X_n = x$ à $X_{n+1} = y$ est obtenu de la manière suivante. On choisit un site $s \in S$ avec probabilité q_s . x est alors changée en y uniquement sur le site s , i.e. $x(S \setminus \{s\}) = y(S \setminus \{s\})$; sur le site s $x(s)$ est changé en $y(s) = \pm 1$ avec la probabilité

$$\pi(y(s)|x(S \setminus s))$$

La matrice de transition de la chaîne de Markov $(X_n)_n$ ainsi construite est

$$\mathbb{P}(X_{n+1} = y | X_n = x) = q_s \pi(y(s)|x(S \setminus s)) \mathbb{1}_{y(S \setminus s) = x(S \setminus s)}$$

La chaîne correspondante est irréductible et aperiodique (pourquoi?). Pour cette chaîne, π_T est la distribution stationnaire car elle satisfait la condition de "bilan détaillé". Le problème beaucoup

plus compliqué est de savoir à quel moment (i.e. pour quel n) on peut considérer que la loi de X_n est quasi égale à π_T . On consultera encore [?] à ce sujet. Un échantillonnage plus astucieux existe. Il consiste, au lieu de ne changer que la valeur d'un site à chaque unité de temps, de changer la valeur des $|S|$ sites (cf. problème 7.6.2 de [8]).

24.4.4 Simulation via l'algorithme de Propp-Wilson pour le modèle d'Ising

On considère le modèle d'Ising à température $\beta^{-1} > 0$. L'espace des configurations est $\Omega_m = \{-1, +1\}^{\Lambda_m}$ où $\Lambda_m = \{1, \dots, m\}^2$. La probabilité que l'on considère sur l'espace des configurations est donnée par :

$$\mu_{m,\beta}(\zeta) = Z_{m,\beta}^{-1} \exp \left(\beta \sum_{x \sim y} \zeta(x) \zeta(y) \right)$$

où $Z_{m,\beta}$ est une constante de normalisation (*fonction de partition*). Si x est un site de Λ_m , on dénote par $k_+(x, \zeta)$ (resp. $k_-(x, \zeta)$) le nombre de voisins de x qui ont pour valeur $+1$ (resp. -1).

1) Montrer que si ζ_0 est un élément de Ω_m fixé

$$\mu_{m,\beta}(\zeta(x) = 1 | \zeta(\Omega_m \setminus \{x\}) = \zeta_0(\Omega_m \setminus \{x\})) = \frac{\exp(2\beta(k_+(x, \zeta_0) - k_-(x, \zeta_0)))}{\exp(2\beta(k_+(x, \zeta_0) - k_-(x, \zeta_0))) + 1}$$

2) On introduit un ordre partiel sur Ω_m donné par $\zeta_0 \leq \eta_0$ si et seulement si pour tout site x , $\zeta_0(x) \leq \eta_0(x)$. Montrer que :

$$\mu_{m,\beta}(\zeta(x) = 1 | \zeta(\Omega_m \setminus \{x\}) = \zeta_0(\Omega_m \setminus \{x\})) \leq \mu_{m,\beta}(\zeta(x) = 1 | \zeta(\Omega_m \setminus \{x\}) = \eta_0(\Omega_m \setminus \{x\}))$$

3) Montrer qu'il existe une configuration maximale et une configuration minimale pour cet ordre.

4) Soit $(U_n)_n$ une suite de variables aléatoires uniformes sur $[0, 1]$ indépendantes. On définit le processus de Markov (X_n) sur Ω_m de la manière suivante. Pour passer de X_n à X_{n+1} , on choisit un site x de Λ_m de manière aléatoire et uniforme. On laisse inchangés les sites de X_n autres que x . $X_{n+1}(x)$ est égal à 1 si

$$U_{n+1} < \frac{\exp(2\beta(k_+(x, X_n) - k_-(x, X_n)))}{\exp(2\beta(k_+(x, X_n) - k_-(x, X_n))) + 1}$$

et -1 sinon.

a) Montrer que $(X_n)_n$ admet $\mu_{m,\beta}$ pour probabilité stationnaire et que la chaîne est aperiodique et irréductible.

b) Montrer que le principe du sandwich dans l'algorithme de Propp-Wilson s'applique pour cette chaîne.

c) Simuler une variable aléatoire de loi $\mu_{m,\beta}$ via l'algorithme de Propp-Wilson.

24.4.5 Simulation de la transition de phase

On propose d'illustrer par simulation informatique la transition de phase pour le modèle d'Ising $H = 0$ et $J = 1$.

- 1) Soit $m \geq 1$ et $T > 0$. A l'aide de l'algorithme de Propp-Wilson, écrire une procédure qui simule une variable aléatoire ayant pour loi $\pi_{T,m}^+$ (resp. $\pi_{T,m}^-$) sur le carré $\{1, \dots, m\}^2$. On note cette fonction $sim^+(m, T)$ (resp $sim^-(m, T)$). Ce sont des tableaux de longueur m^2 .
- 2) Ecrire une fonction $mom - magn^+(T, m)$ (resp. $mom - magn^-(T, m)$) qui approxime $\pi_{T,m}^+(\zeta(0))$ (resp. $\pi_{T,m}^-(\zeta(0))$). Pour calculer cette espérance, on fera appel un certain nombre de fois à $sim^+(m, T)$ (resp. $sim^-(m, T)$) et on effectuera une moyenne empirique. $mom - magn^\pm(T, m)$ aura pour sortie un nombre.
- 3) Sur un graphique, tracer $mom - magn^\pm(m, T)$ en fonction de T . On choisira m de taille assez grande pour considérer que $m \rightarrow \infty$ et pas trop grand pour que les simulations ne durent pas plus de 10 minutes.
- 4) Observer que les deux courbes sont approximativement égales pour $T > T_c$ à la valeur 1/2 mais pas pour $T < T_c$.
- 5) Attention, les simulations peuvent être assez longues. En particulier, on constatera que l'algorithme prend beaucoup plus de temps lorsque T est aux environs de T_c . On évitera donc de prendre des valeurs trop proches de T_c .

24.5 Méthode de gradient stochastique

Une application à des méthodes de simulation apparaît dans ce que l'on appelle la méthode de gradient stochastique. Soit E un ensemble fini et $U : E \rightarrow \mathbb{R}$ une fonction appelée fonction de coût et que l'on cherche à minimiser. Autrement dit, on cherche à déterminer les i_0 de E tels que

$$U(i_0) \leq U(i) \text{ pour tout } i \in E$$

Un tel élément est appelé un minimum global de la fonction de coût. L'algorithme de de gradient classique (ou déterministe) est le suivant et est grosso modo, une version discrète de la méthode du gradient. Pour chaque état $i \in E$, on se donne un ensemble (petit) $N(i)$ de sites de E appelés les voisins de i . Supposons qu'à une certaine étape, une solution i est examinée. A l'étape suivante, un site j est choisi dans l'ensemble $N(i)$ des voisins de i et on compare $U(i)$ avec $U(j)$. Si $U(i)$ est plus petit, on s'arrête là. Si $U(j)$ est plus petit, alors on repart de nouveau avec j qui joue maintenant le rôle de i . Puisque E est finie, l'algorithme s'arrêtera. Cependant, rien n'exclut qu'en fait de minimum global, ce soit un minimum local que l'on obtienne (c'est-à-dire que l'on tombe sur i tel que $U(i) \leq U(j)$ pour tout $j \in N(i)$).

L'idée de l'algorithme de gradient stochastique est de donner la possibilité (avec une petite probabilité) à chaque étape de ne pas choisir le minimum entre $U(i)$ et $U(j)$. On évite ainsi de rester coincé sur un minimum local.

On se donne une suite de voisinages $N(i)$ qui communiquent :

Pour tout couple $i, j \in E$ d'états, il existe une trajectoire de i à j , c'est-à-dire une suite finie d'états $i_1, \dots, i_m \in E$ tels que $i_1 \in N(i), i_2 \in N(i_1), \dots, j \in N(i_m)$.

On définit maintenant une matrice de transition irréductible $(Q_{i,j})_{i,j \in E}$ telle que $Q_{i,j} > 0$ si et seulement si $i = j$ ou $j \in N(i)$. On supposera de plus que Q est symétrique.

L'échantillonnage de Métropolis est le suivant. On se donne un paramètre $T > 0$ et on pose

$$\alpha_{i,j}(T) = \inf \left\{ 1, e^{-\frac{U(i) - U(j)}{T}} \right\}$$

On définit d'autre part une matrice de transition irréductible $P(T)$ par

$$P_{i,j}(T) = Q_{i,j} \alpha_{i,j}(T)$$

- 1) Montrer que si U n'est pas une constante alors $P(T)$ est irréductible et aperiodique.
- 2) Montrer que $P(T)$ admet une probabilité stationnaire donnée par :

$$\forall i \in E, \quad \pi_i(T) = \frac{e^{-U(i)/T}}{\sum_{k \in E} e^{-U(k)/T}}$$

- 3) Soit H l'ensemble des minimas globaux

$$H = \{i \in E; \forall j \in E, U(i) \leq U(j)\}$$

Montrer que

$$\lim_{T \rightarrow 0} \pi_i(T) = \begin{cases} \frac{1}{|H|} & \text{si } i \in H \\ 0 & \text{si } i \notin H \end{cases}$$

Ces résultats suggèrent la procédure suivante. On initialise l'algorithme avec une valeur du paramètre $T = T_0$ et on simule la chaîne de Markov de matrice de transition $P(T_0)$. Au bout d'un temps suffisamment long, la distribution de la chaîne sera proche de la probabilité $\pi(T_0)$. A l'étape suivante, on choisit une valeur du paramètre $T = T_1 < T_0$ et à nouveau on simule la chaîne de Markov de matrice de transition $P(T_1)$ partant de la distribution obtenue (qui est proche de $\pi(T_0)$ donc proche de $\lim_{T \rightarrow 0} \pi(T)$). On attend un temps suffisamment long pour que la distribution de la chaîne simulée soit proche de la distribution stationnaire $\pi(T_1)$. On continue ainsi de suite avec une suite de paramètres $(T_k)_{k \geq 0}$ qui tend vers 0.

Toute la difficulté consiste alors à savoir quel temps a_k il faut attendre entre deux changements de paramètres $T_k \rightarrow T_{k+1}$ pour être proche de la distribution stationnaire et quel doit être l'ordre de grandeur du passage de T_k à T_{k+1} . Une condition nécessaire et suffisante a été donnée par Hajek (1988) mais elle est en pratique inutilisable. Elle énonce qu'il existe une constante γ telle que l'algorithme précédent converge quel que soit l'état initial si et seulement si

$$\sum_{k=1}^{\infty} e^{-\frac{\gamma}{a_k}} = \infty$$

En particulier, on peut choisir a_k de la forme

$$a_k = \frac{a}{\log(k+1)}$$

pour $a \geq \gamma$. Bien entendu, on ne connaît pas la valeur de γ .

Un exemple élémentaire (Exemple 13.4 p. 104-105 de [19]) montre que si l'on choisit un refroidissement (i.e. une décroissance des T_k) trop rapide, alors l'algorithme peut defaillir. On considère un ensemble E à 4 éléments s_1, \dots, s_4 et une fonction f définie par :

$$f(s_1) = 1, \quad f(s_2) = 2, \quad f(s_3) = 0, \quad f(s_4) = 2$$

On veut appliquer la méthode du gradient stochastique pour trouver le minimum de f .

1) Montrer que la matrice de transition de la chaîne de Markov qui permet d'appliquer l'algorithme pour une température T est donnée par :

$$\begin{pmatrix} 1 - e^{-1/T} & 0.5e^{-1/T} & 0 & 0.5e^{-1/T} \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5e^{-2/T} & 1 - e^{-2/T} & 0.5e^{-2/T} \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix} \quad (24.5.1)$$

Soit (T_k) la suite des températures que l'on choisit dans l'algorithme et a_k les temps de simulation avec température T_k . On commence donc par prendre une valeur X_0 . On simule une chaîne de Markov partant de X_0 avec la température T_0 dans l'algorithme de Métropolis et au bout d'un certain temps a_0 , on arrête là avec la température T_0 . On a un nouvel état X_1 . On part de ce nouvel état et on simule via Metropolis une chaîne de Markov mais à la température T_1 pendant a_1 et on continue ainsi de suite. On suppose que l'on choisit s_1 pour valeur de X_0 . Soit A l'événement : "pour tout k , $X_k = s_1$ ".

1) Montrer que

$$\mathbb{P}(A) = \prod_{i=1}^{\infty} (1 - e^{-1/T_i})^{a_i}$$

2) Supposons que l'on choisisse d'appliquer l'algorithme avec $a_i = t_0 = 10^5$. Montrer que si $a_i = i^{-1}$ alors l'algorithme ne fonctionnera pas toujours.

Remarque : Dans [19], il semble que $a_i = 1$, ce qui est très étrange... Néanmoins, cela ne modifie en rien l'argumentation finale.

Chapitre 25

Le modèle de Cox-Ross-Rubinstein, Fanny Godet

Le modèle de Cox-Ross-Rubinstein est une version discrétisée du modèle de Black-Scholes dans laquelle on considère deux actifs : un actif sans risque et un actif risqué. Tous les programmes sont rédigés en Matlab.

Simulation d'une variable aléatoire discrète

Le programme ci-dessous simule une variable aléatoire discrète X lorsque qu'on lui donne le vecteur x des valeurs que peut prendre X et le vecteur p des probabilités associées. On utilise pour cela la fonction `rand` de Matlab qui permet de simuler une variable aléatoire de distribution uniforme sur $[0,1]$.

```
function z = simulationdiscrete(x,p)
%le vecteur p représente les probabilités des composantes x_i
if sum(p)~=1
    error('la somme des probabilités doit etre égale à 1')
elseif size(p)~=size(x)
    error('x et p doivent etre de meme longueur')
else
    [n,m]=size(x);
    u=rand(1,1);
    sp=cumsum(p);
    pp=[0 p(1:(n-1))];
    spp=cumsum(pp);
    k=find((u>=spp)&(u<sp));
    z=x(k(1));
end
```

Simulation du cours de l'actif risqué et des cours des call et put sur l'actif risqué

Le rendement de l'actif sans risque est noté r . Le cours de l'actif risqué est donné par la formule suivante avec $-1 < a < b$:

$$S_{n+1} = \begin{cases} S_n(1+a) & \text{avec une probabilité } p \\ S_n(1+b) & \text{avec une probabilité } 1-p \end{cases}$$

Pour obtenir que le marché est viable et complet, j'ai démontré durant mon exposé qu'il est nécessaire et suffisant de vérifier les conditions suivantes :

$$r \in]a, b[\text{ et } p = \frac{b-r}{b-a}$$

Dans les programmes suivants, j'ai appelé 'départ' la valeur initiale de l'actif risqué, 'longueur' le temps pendant lequel j'étudie les cours (qui correspond à N l'échéance des options) et enfin 'K' le prix d'exercice de l'option.

La fonction trajectoire simule la valeur du cours de l'actif risqué.

```

fonction s=trajectoire(depart,a,b,r,longueur)
p=(b-r)/(b-a);
p=[p (1-p)];
x=[(1+a) (1+b)];
z=depart*ones(1,longueur);
for j=2:longueur
    z(j)=z(j-1)*simulationdiscrete(x,p);
end
s=z;

```

La fonction call simule le cours du call à l'instant n en fonction du cours de l'actif risqué à l'instant n noté s et de l'instant n .

$$\text{call}(n, s) = \sum_{j=0}^{N-n} C_{N-n}^j p^j (1-p)^{N-n-j} (s(1+a)^j (1+b)^{N-n-j} - K)_+$$

```

fonction c=valeurcall(n,s,r,N,a,b,K)
p=(b-r)/(b-a);
ta=(1+a)*ones(N-n,1);
ta=[1 ;ta(1:N-n)];
ta=cumprod(ta);
tb=(1+b)*ones(N-n,1);
tb=[1 ;tb(1:N-n)];
tb=cumprod(tb);
tb=[tb((N-n+1):-1:1)];
t=(s*ta.*tb)-K;
tpos=[(t>0)];
t=t.*tpos;
c=(1+r)^(n-N)*dbinom(0:(N-n),(N-n),p)*t;

```

La fonction put calcule la valeur du put à l'instant n. Pour cela j'ai utilisé la relation de parité call-put :

$$\text{call}(n, s) - \text{put}(n, s) = s - K(1 + r)^{n-N}$$

```
function p=valeurput(n,s,r,N,a,b,K)
p=valeurcall(n,s,r,N,a,b,K)-s+K*(1-r)^(n-N);
```

Les simulations

J'ai décidé d'étudier les cours au jour le jour pendant un an. J'ai choisi un rendement sans risque de 3% par an et un cours de départ de l'actif risqué à 100 euros.

```
longueur=365;
r=(1.03)^(1/365)-1;
depart=100;

K=120;
a=-0.02;
b=0.02;
s=trajectoire(depart,a,b,r,longueur);
c=zeros(longueur,1);
p=zeros(longueur,1);
for j=1: longueur
    c(j)=valeurcall(j-1,s(j),r,longueur,a,b,K);
    p(j)=valeurput(j-1,s(j),r,longueur,a,b,K);
end
hold on
ylabel('Cours des actions')
xlabel('t en jours')
plot(1:365,s','b',1:365,c,'r:',1:365,p,'g-.',1:365,K,'y-.');
hold off
print -eps dessous.eps
```

En trait plein est dessiné le cours de l'action, en pointillé le cours du put et en pointillé plus léger le cours du call. Le trait horizontal correspond au prix d'exercice de l'option.

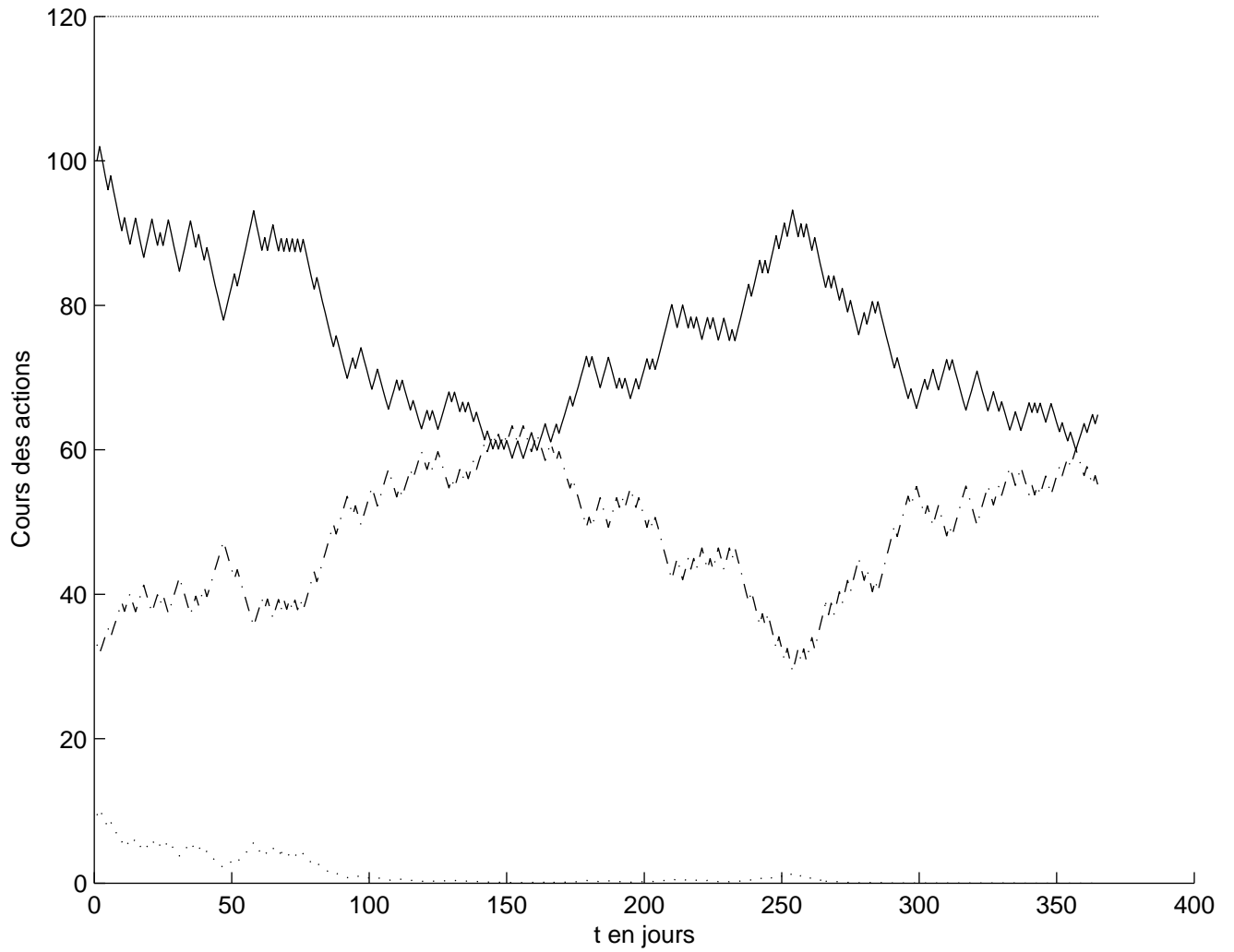


FIG. 25.1 – K est supérieur au cours de l'actif.

Ici près de la date d'échéance la valeur du put se rapproche de la différence entre K et le cours de l'actif alors que le cours du call tend vers 0. En effet il n'y a aucun intérêt à posséder le droit d'acheter l'actif à un prix plus élevé que celui pratiqué par le marché.

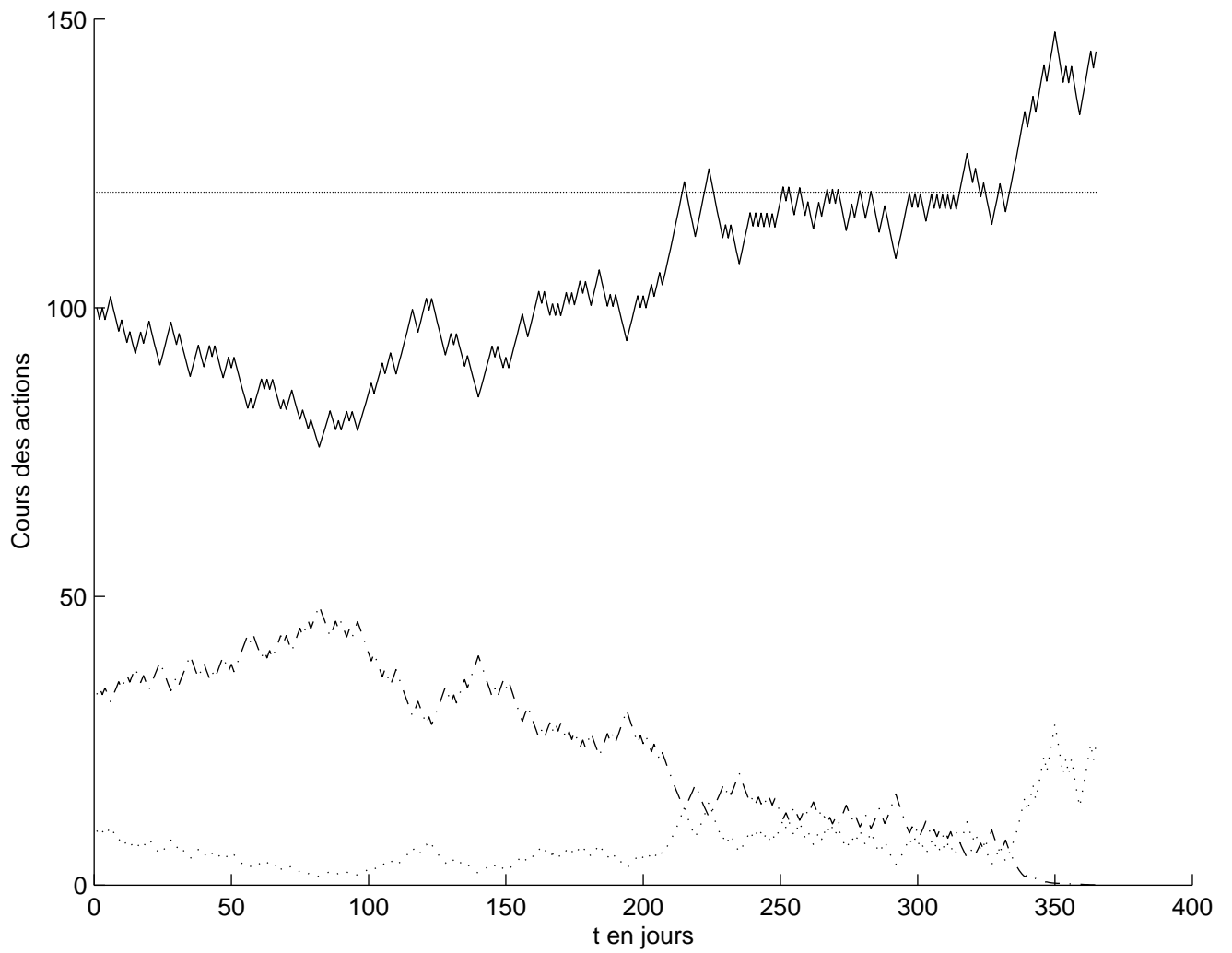


FIG. 25.2 – K est inférieur au cours de l'actif.

Dans ce cas on voit que près de la date d'échéance le cours du call vaut approximativement la différence entre le cours de l'action et K .

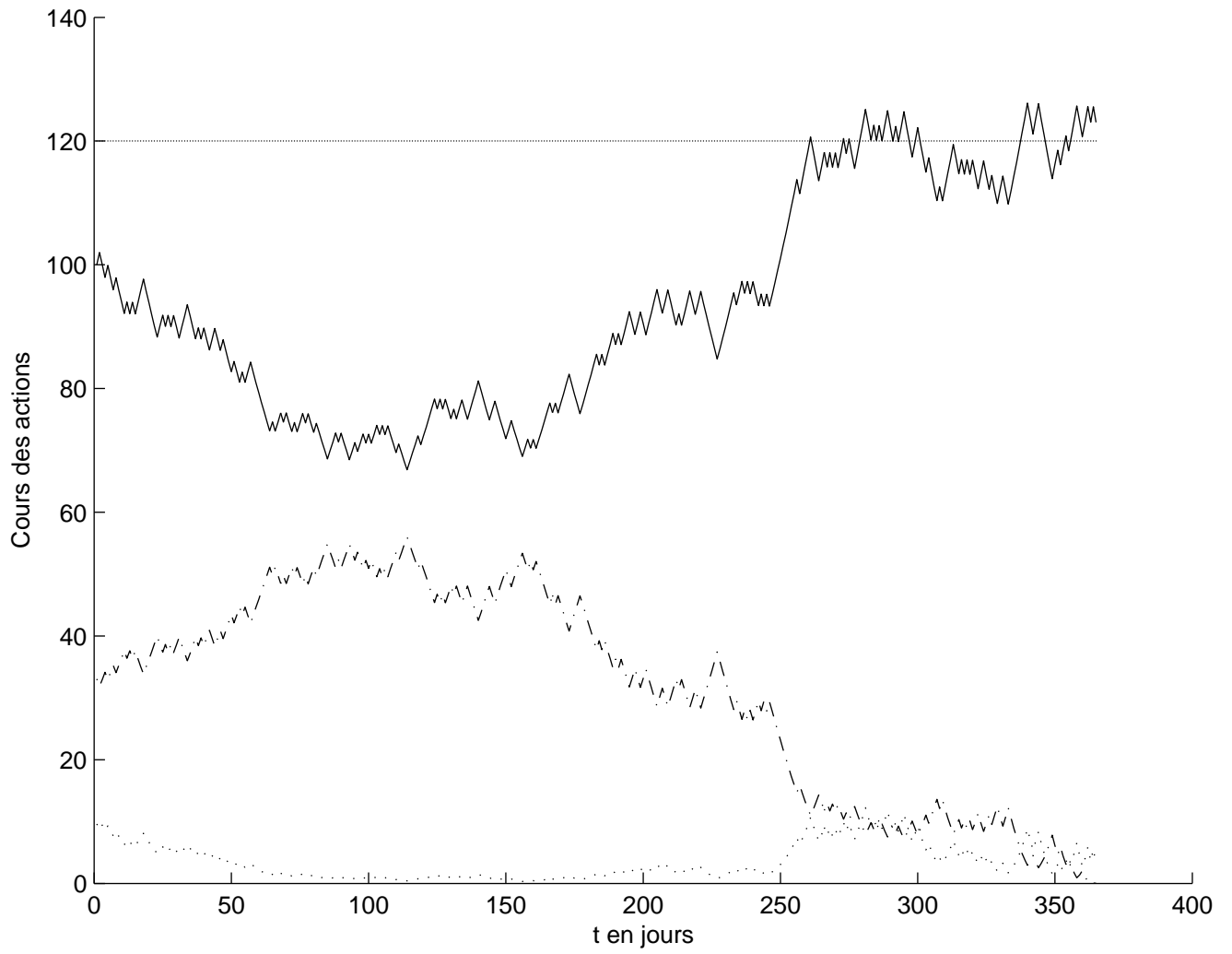


FIG. 25.3 – K est au niveau du cours de l'actif.

Chapitre 26

Théorie de l'information et de la communication

Sources : Claude E. Shannon A mathematical source of communication
Pierre Brémaud Introduction aux probabilités, Chapitre 5, Springer
Thomas M. Cover, Joy A. Thomas, Elements of information theory (pages 194-197),
Wiley, 1991.

26.1 Introduction

Ces notes, qui donnent toutes les démonstrations, parfois sous forme d'exercice, sont basées essentiellement sur le fameux livre de Shannon. Celui-ci a l'avantage d'être très pédagogique, ayant été écrit pour des ingénieurs. L'exposé étant parfois légèrement elliptique du point de vue mathématique, on le complète facilement à l'aide du Brémaud (Chapitre 5) et du Cover-Thomas (trois pages). Dans son livre, Shannon introduit la notion d'entropie en théorie des transmissions (cette notion était avant lui connue en physique statistique dans un contexte très différent). Le livre de Shannon contient tous les points de départ des thèmes qui seront ensuite développés en ce qu'on appellera plus tard la "théorie de l'information". Les nombreux livres sur la théorie de l'information suivent, cinquante ans après, exactement le même plan que celui de Shannon ! Shannon commence par réduire le problème de la transmission à celui de la transmission d'une série de symboles émis par une source aléatoire. La source émet des symboles (représentant par exemple des phrases en français). L'incertitude du récepteur concernant le symbole qui va être transmis est grande mais pas totale, puisque certains symboles (certaines phrases) sont plus probables que d'autres. On suppose connue la probabilité de chaque symbole.

Dans le cas du langage, Shannon montre que l'on peut évaluer cette probabilité en modélisant l'émission d'un texte comme la production d'une chaîne de Markov. Les probabilités de transition (lettre à lettre, ou plutôt mot à mot) de cette chaîne de Markov peuvent être apprises en analysant un texte de taille suffisante. Shannon montre par une simulation numérique très simple que si on respecte les probabilités de transition mot à mot, on peut déjà synthétiser un texte qui ressemble vraiment à de l'anglais ! On peut donc supposer connue, ou calculable, la probabilité de chaque symbole (phrase) possible. Alors, dans la communication, tout se passe comme si la source (l'émetteur) tirait au sort un symbole et l'envoyait au récepteur. Voici quelques exemples tirés du livre de Shannon.

1. Approximation d'ordre zéro (symboles indépendants et équiprobables).
XFOML RXKHRRJFFJUZ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIUBZLH-
JQD.
2. Approximation d'ordre un (symboles indépendants mais respectant les fréquences d'un texte anglais).
OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA
NAH BRL.
3. Approximation d'ordre 2 (fréquence des couples de lettres comme en anglais).
ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE
TUCOOWE AT TEATSONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.
4. Approximation d'ordre 3 (fréquence des triples de lettres correcte).
IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONS-
TURES OF THE REPTAGIN IS REGOACTIONA OF CRE.
5. Approximation d'ordre 1 en mots : plutôt que de passer aux tétragrammes, il est plus facile de sauter aux unités en mots. La fréquence des mots est correcte :
REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT
NATURAL HERE HE THE A CAME THE TO OF TO EXPERT GRAY COME TO FUR-
NISHES THE LINE MESSAGE HAD BE THESE ;
6. Approximation d'ordre deux : les probabilités de transition entre mots successifs sont correctes (et rien d'autre).
THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHA-
RACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS
THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED ;

Shannon remarque : "The resemblance to ordinary English text increases quite noticeably at each of the above steps. Note that these samples have reasonably good structure out to about twice the range that is taken into account in their construction. Thus in 3. the statistical process insures reasonable text for two-letter sequences, but four-letter sequences from the sample can usually be fitted into good sentences. In 6. sequences of four or more words can easily be placed in sentences without unusual or strained constructions."

26.1.1 L'entropie d'une source

Shannon définit une *source*, comme un ensemble \mathcal{X} de symboles possibles avec leurs probabilités $p(x)$ et cherche à définir son entropie comme une mesure de l'incertitude laissée au récepteur.

La définition formelle est

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x)$$

où X est la variable aléatoire dont la valeur est le symbole émis. (Sauf mention contraire, on note \log pour le logarithme en base 2 et \log_D pour le logarithme en base D). Pour expliquer cette définition de l'entropie de diverses manières, Shannon a recours à des arguments tirés de la loi des grands nombres. Aussi ne va-t-il pas considérer des symboles émis isolément, mais bien des suites de n symboles, où n est grand. Nous appellerons ces séquences des *messages*, et ces séquences sont composées de n symboles tous tirés de manière aléatoire indépendante selon la distribution de probabilité $p(x)$.

L'entropie d'une source s'interprétera alors de trois manières :

- l'entropie est la longueur moyenne minimale des mots codant tous les symboles dans un alphabet binaire : Shannon montre que l'on peut coder de manière optimale en associant à chaque symbole x de probabilité $p(x)$ un code binaire (nombre fait de 0 et 1) de longueur inférieure ou égale à $-\log p(x) + 1$.
- l'entropie est la capacité minimale moyenne d'un canal capable de transmettre les messages issus de la source
- l'entropie est la moyenne du logarithme du nombre de "messages typiques" quand la source envoie un nombre n assez grand de symboles successifs indépendants. En effet, toutes les séquences de symboles $x_1x_2 \dots x_n$ sont possibles, mais elles ne sont pas du tout équiprobables : par la loi des grands nombres, la fréquence d'apparition dans la séquence du symbole x tend vers $p(x)$ quand $n \rightarrow \infty$. Il y donc beaucoup moins de messages typiques que de messages possibles. Le nombre de messages typiques est 2^{nH} où H est l'entropie, alors que le nombre de messages réels est $(\text{Card } X)^n = 2^{n \log \text{Card}(X)}$.

Le principe du codage optimal est très intuitif : les messages sont codés par des codes d'autant plus courts qu'ils sont fréquents (codages de Elias-Fano-Shannon, ou de Huffman, voir le Brémaud, et le T.P. d'Agnès Desolneux). Dans tous les cas on retombe sur la même formule ; l'entropie est égale à $-\sum_{x \in \mathcal{X}} p(x) \log p(x)$.

26.1.2 Transmission dans un canal bruité

Ensuite, Shannon s'occupe de transmission dans un "canal bruité" : son problème est de savoir si celle-ci est réalisable, et à quel prix en termes de redondance dans l'envoi. Shannon s'inspire de ses expériences d'enfant avec le télégraphe : il sait qu'un message peut être corrigé par le récepteur à cause de la "redondance" du langage. Pour l'anglais, comme il le mesurera plus tard, cette redondance est de l'ordre de 50 pour cent : on peut supprimer jusqu'à la moitié des lettres, ou des mots, d'un message en anglais et qu'il soit encore reconstituable (essayer !). De plus, *plus le message est long et plus la reconstruction est facile*. Cette intuition va être transformée par Shannon en un théorème fondamental. Ce résultat de Shannon, obtenu pourtant par un argument mathématique non constructif, suscitera l'enthousiasme des ingénieurs. Il introduit la notion d'entropie relative, $H(Y|X)$, où X est la source de messages aléatoires et Y est la sortie du canal, également une variable aléatoire. $H(Y|X)$ mesure l'incertitude sur le message envoyé X sachant que l'on a reçu Y . C'est donc une mesure de l'incertitude produite par le canal, une mesure du "bruit" dans celui-ci.

Dans son théorème fondamental, Shannon montre qu'il est possible d'obtenir un taux d'erreur aussi bas que désiré. La transmission dans le bruit est donc possible et Shannon définit la capacité abstraite d'un canal comme $\max_X H(X) - H(X|Y)$ pour toutes les sources X utilisées comme entrées. (La capacité est la différence entre quantité d'information envoyée ($H(X)$) et l'incertitude générée sur ce qui est envoyé, $H(X|Y)$). Il montre ensuite que toute source d'entropie inférieure ou égale à cette capacité est transmissible avec un taux d'erreur arbitrairement bas.

26.1.3 Plan

Dans ces notes, qui combinent exercices élémentaires et démonstrations, nous allons (les numéros correspondent aux différents exercices) :

1. donner les formulaires de l'entropie et de l'entropie relative
2. caractériser l'entropie comme le logarithme du nombre moyen de séquences typiques de messages typiques quand la source émet une longue suite

3. définir la capacité d'un canal non bruité
4. montrer que toute source d'entropie inférieure ou égale à la capacité peut être transmise
5. compter les entrées et sorties typiques pour un canal bruité
6. définir la capacité d'un canal bruité
7. prouver le théorème fondamental de Shannon selon lequel la transmission est possible dans un canal bruité. Ce théorème est d'une élégance extrême. Il prouve l'existence d'un codage optimal par un argument de moyenne sur tous les codes possibles.

Ce dernier résultat a montré la possibilité de codes correcteurs d'erreur avant que de tels codages ne commencent à être trouvés. Leur théorie est loin d'être terminée.

26.2 Les étapes du codage : vocabulaire

- *symbole* : on peut penser comme unités significatives minimales aux phrases. Chaque phrase possible est dans le modèle de Shannon représentée par un symbole x dans un "alphabet" X qui est donc fini mais très grand. (A comparer au langage de Sancho Panza dans Don Quichotte, qui ne parle que par proverbes : chaque proverbe a une fréquence et c'est une concaténation de proverbes qui constitue un discours de Sancho Panza.)
- Chaque phrase x a une probabilité $p(x)$. Comment la calculer ? On modélise la langue comme le résultat d'une chaîne de Markov : on mesure pour chaque paire de mots M et N la probabilité $p(M|N)$ que M suive N , mesurée comme le rapport entre le nombre d'apparitions de NM divisé par le nombre d'apparitions de N . Il suffit pour cela de parcourir quelques milliers de documents dans la langue concernée. On interprète $p(M|N)$ comme la probabilité conditionnelle de lire M sachant qu'on a lu N . Le stockage en machine de ces probabilités est faisable : 10^5 mots usuels et donc 10^{10} probabilités conditionnelles. On stocke aussi les probabilités d'apparition de chaque mot. On peut alors générer une "phrase typique de longueur n ", $x = M_1M_2 \dots M_n$ de la langue en tirant au sort le premier mot M_1 avec la distribution de probabilité $p(M)$, le second avec la distribution de probabilité $p(M|M_1)$, etc. Chomsky et Schützenberger ont démontré qu'un modèle markovien était insuffisant pour rendre compte de la structure syntaxique récursive du langage. Toutefois, le modèle markovien donne des résultats très acceptables, comme le prouve le fait que les phrases générées, apprises de journaux, ressemblent assez à la production journalistique moyenne.
- *message* : il s'agit d'une concaténation de symboles tirés selon la distribution $p(x)$. Si chaque symbole représente une phrase, alors un message est une concaténation de symboles et représente donc un long texte. Chaque symbole est tiré indépendamment des autres, et un message de longueur n est donc simplement une suite de variables indépendantes et identiquement distribuées.
- *variations du calcul d'entropie* : Il est clair que le calcul de l'entropie d'un texte n'est pas le même si on utilise comme unité le caractère, ou le mot, ou la phrase. Shannon montre bien que plus les unités dont on calcule la probabilité sont longues et plus l'entropie décroît et serre de près la vraie incertitude sur le texte.
- *codage d'une source par une autre* : Soient deux sources différentes X_1 et X_2 . Les messages de longueur n émis par la source X_1 peuvent être associés à des messages de longueur m émis par la source X_2 , à condition que m soit assez long. Ainsi, les caractères de l'alphabet peuvent-être codés comme séquences de huit bits (mots composés de 0 et de 1). On appelle

une telle application un codage. Les codages optimaux permettront d'adapter la longueur des symboles à leur fréquence, les symboles fréquents étant codés par des mots très courts et les symboles infréquents par des codes longs. Cette propriété du codage est déjà patente dans le langage courant : "oui", "non", "et", "quoi?", "homme", "chat", "chien" sont courts et fréquents, "anticonstitutionnellement" et "rhinocéros" sont longs et rares.

Exercice 26.2.1. Formulaire de Shannon (source : Shannon) On définit l'entropie d'une v.a. discrète X dont les valeurs x sont dans l'alphabet fini \mathcal{X} , et idem pour un couple de variables aléatoires (X, Y) à valeurs dans $\mathcal{X} \times \mathcal{Y}$:

$$H(X) = - \sum_x p(x) \log p(x), \quad H(X, Y) = - \sum_{x,y} p(x, y) \log p(x, y).$$

Montrer que l'entropie est l'espérance de $g(X)$, avec $g(X) = \log \frac{1}{p(X)}$.

Exemple important : on prend pour X la variable de Bernoulli, $X = 1$ avec probabilité p , $X = 0$ avec probabilité $1 - p$. Alors

$$H(X) = -p \log p - (1 - p) \log(1 - p),$$

que l'on notera $H(p)$. Vérifier que $H(X) = 1$ bit quand $p = \frac{1}{2}$. Tracer le graphe de $H(p)$ et montrer qu'elle vaut 0 en 0 et 1 et est maximale en $p = \frac{1}{2}$. Interprétation : l'incertitude sur X est maximale quand $p = \frac{1}{2}$ et minimale quand X est déterministe. L'entropie est une mesure de l'incertitude sur la valeur de X .

Proposition 26.2.1. *L'entropie jointe de deux variables aléatoires est plus petite que la somme des entropies. Il y a égalité si et seulement si les deux variables aléatoires sont indépendantes*

$$H(X, Y) \leq H(X) + H(Y).$$

Lemme 26.2.1. *Soient p et q deux distributions de probabilité discrètes. Alors $\sum p(x) \log \frac{p(x)}{q(x)} \geq 0$. Il y a égalité si et seulement si $p(x) = q(x)$ pour tout x .*

Démonstration de la proposition 26.2.1 . On applique le lemme 26.2.1 aux distributions $p(x, y)$ et $p(x)p(y)$.

$$\sum p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \geq 0,$$

ce qui donne

$$- \sum_{x,y} p(x, y) \log p(x, y) \leq - \sum_{x,y} p(x, y) \log p(x) - \sum_{x,y} p(x, y) \log p(y),$$

l'égalité n'étant vérifiée que si $p(x, y) = p(x)p(y)$. Comme $\sum_y p(x, y) = p(x)$ et $\sum_x p(x, y) = p(y)$, on obtient le résultat annoncé. \square

Démonstration du lemme 26.2.1. On utilise la stricte concavité du logarithme.

$$-\sum p(x) \log \frac{p(x)}{q(x)} = \sum p(x) \log \frac{q(x)}{p(x)} \leq \log \left(\sum p(x) \frac{q(x)}{p(x)} \right) = \log 1 = 0,$$

l'inégalité étant une égalité si et seulement si toutes les valeurs $\frac{p(x)}{q(x)}$ sont égales entre elles, auquel cas la valeur commune est évidemment 1. \square

Entropie conditionnelle de Y sachant X : c'est "la moyenne de l'entropie de Y pour chaque valeur de X , pondérée par la probabilité d'avoir cette valeur particulière de X ". (Vérifier que la formule qui suit correspond à cette définition !)

$$H(Y|X) = - \sum_{x,y} p(x,y) \log p(y|x).$$

"This quantity measures how uncertain we are of Y on the average when we know X ." En utilisant la définition de la probabilité conditionnelle :

$$p(y|x) = \frac{p(x,y)}{\sum_y p(x,y)},$$

$$H(Y|X) = - \sum_{x,y} p(x,y) \log p(x,y) + \sum_{x,y} p(x,y) \log \sum_y p(x,y) = H(X,Y) - H(X).$$

Donc

$$H(X,Y) = H(X) + H(Y|X).$$

"The uncertainty (or entropy) of the joint event X,Y is the uncertainty of X plus the uncertainty of Y when X is known". Utilisons la proposition 26.2.1 :

$$H(X) + H(Y) \geq H(X,Y) = H(X) + H(Y|X).$$

Donc

$$H(Y) \geq H(Y|X).$$

"The uncertainty of Y is never increased by knowledge of X . It will be decreased unless X and Y are independent events, in which case it is not changed".

Exercice 26.2.2. Messages typiques (énoncé et démontré dans le Brémaud page 99 : essentiel!)
 Nous gardons les notations du Brémaud. En pratique $D = 2$. Soit X_n une suite de v.a.i.i.d. à valeurs dans un ensemble fini de symboles $\mathcal{X} = \{\infty, \epsilon, \dots, \|\}$ et telles que $P(X_n = i) = p_i, i = 1, \dots, k$. Soit \mathcal{X}^n l'ensemble des suites (messages) de longueur n : il y en a k^n . On considère l'entropie de la répartition $(p_i)_{1 \leq i \leq k}$,

$$H_D(p_1, \dots, p_k) = -\sum_{i=1}^k p_i \log_D p_i.$$

On définit l'ensemble des "messages typiques"

$$C_n = \{(i_1, \dots, i_n) \in M^n, D^{-n(H_D + \epsilon)} \leq p_{i_1} \dots p_{i_n} \leq D^{-n(H_D - \epsilon)}\}.$$

1) Montrer que $E(-\log_D p_{X_n}) = H_D$.

2) Dédurre que

$$P(\{|\frac{1}{n}\sum_{i=1}^n(-\log_D p_{X_i}) - H_D| \geq \varepsilon\}) = P((X_1, \dots, X_n) \in C_n^c) \leq \frac{Var(-\log_D p_{X_1})}{n\varepsilon^2}.$$

3) Montrer que

$$P((X_1, \dots, X_n) \in C_n) \geq D^{-n(H_D+\varepsilon)} Card(C_n).$$

Dédurre que $Card(C_n) \leq D^{(H_D+\varepsilon)n}$.

4) Réciproque. Supposons que l'on ait pu trouver $\tilde{C}_n \subset \mathcal{X}\mathcal{X}^\lambda$ tel que $\lim_{n \rightarrow +\infty} P((X_1, \dots, X_n) \in \tilde{C}_n) = 1$ et $Card(\tilde{C}_n) \leq D^{Kn}$. On va montrer que $K \geq H_D$.

4a) Montrer que $\lim_{n \rightarrow \infty} P((X_1, \dots, X_n) \in C_n \cap \tilde{C}_n) = 1$.

4b) Montrer que $P((X_1, \dots, X_n) \in C_n \cap \tilde{C}_n) \leq D^{-n(H_D-\varepsilon)} D^{Kn}$ et conclure.

5) On prend $D = 2$, $\mathcal{X}\mathcal{X} = \{l, \infty\}$, $p_1 = p$, $p_2 = (1-p)$. Si $p_1 = p_2 = \frac{1}{2}$, vérifier que $H_2 = 1$ et que le nombre de suites typiques est 2^n . (Pas de compression possible).

Cas général : étudier la forme de $H_2(p) = -p \log p - (1-p) \log(1-p)$ et en déduire le comportement des suites typiques.

Corrigé : Brémaud, Introduction aux probabilités, pages 97 à 102.

Exercice 26.2.3. Capacité d'un canal discret non bruité. Cette définition est donnée par Shannon page 37. Soit un canal transmettant des séquences de symboles sans erreur. On appelle capacité du canal la quantité (mesurée en bits/seconde)

$$C = \lim_{T \rightarrow \infty} \frac{\log N(T)}{T} \tag{26.2.1}$$

où $N(T)$ est le nombre maximal de messages distincts transmissibles pendant le temps T . Remarquer que cette définition ne se préoccupe pas de l'alphabet ou du type de codage discret ou physique dans lequel les messages sont transmis. On obtient une définition intrinsèque en considérant uniquement le nombre de messages distincts. Si par exemple le canal déterministe transmet des séquences arbitraires de 0 ou 1 au débit de un symbole par seconde, on a 2^T messages possibles et on a donc bien une transmission à un bit par seconde. Cette définition de la capacité est aussi une hypothèse, à savoir l'existence de la limite (26.2.1).

Exercice 26.2.4. Application : transmission d'une source par un canal déterministe.

L'énoncé qui suit et sa démonstration suivent très directement, et avec les mêmes notations, la démonstration originale de Shannon (page 59). Une deuxième démonstration proposée ensuite par Shannon donne un codage explicite, appelé depuis codage de Shannon-Elias-Fano. (Voir Brémaud pages 85-95). L'intérêt de la démonstration qui suit est qu'elle s'appuie entièrement sur l'évaluation du nombre de messages typiques grâce à l'entropie de la source. Le problème est de savoir comment lier la capacité d'un canal à l'entropie de la source dont on veut transmettre la production via ce canal. Comme la définition de la capacité est asymptotique (une limite quand des messages arbitrairement longs sont transmis), on considère aussi des messages arbitrairement longs émis par la source, c'est-à-dire des séquences de n symboles avec $n \rightarrow +\infty$.

Définition 26.2.1. Soit un système de transmission formé d'une source et d'un procédé de codage de cette source permettant son entrée dans un canal déterministe. On appelle débit maximal asymptotique du système la quantité $\limsup_{n \rightarrow \infty} \frac{n}{T_n}$, où T_n est le temps minimal garantissant la transmission d'une séquence de n symboles. On note aussi $E\tilde{T}_n$ l'espérance du temps pour transmettre n symboles, autrement dit le temps moyen de transmission de n symboles de la source. On appelle "débit moyen" du système de transmission, mesuré en symboles par seconde, la quantité

$$\liminf_{n \rightarrow \infty} \frac{n}{E\tilde{T}_n}.$$

Théorème 26.2.1. Soient une source d'entropie H (en bits par symbole) et un canal de capacité C (en bits par seconde). Alors pour tout $\varepsilon > 0$, il est possible de coder l'output de la source de telle sorte que la transmission dans le canal se fasse avec un débit moyen de $\frac{C}{H} - \varepsilon$ symboles par seconde. Il n'est pas possible de transmettre avec un débit maximal strictement supérieur à $\frac{C}{H}$. En d'autres termes,

$$\limsup_{n \rightarrow \infty} \frac{n}{T_n} \leq \frac{C}{H} \quad \text{et} \quad \liminf_{n \rightarrow \infty} \frac{n}{E\tilde{T}_n} \geq \frac{C}{H}.$$

Démonstration Montrons d'abord la deuxième partie : pour n grand, le cardinal de l'ensemble des messages de n symboles qui peuvent être émis par la source est supérieur ou égal à $2^{n(H-\varepsilon)}$ (exercice 2). Soit T_n le temps de transmission minimal requis pour être sûrs de transmettre l'un quelconque de ces messages. Par la définition de la capacité,

$$N(T_n) \geq 2^{n(H-\varepsilon)}.$$

Comme $T_n \rightarrow \infty$, on a pour n assez grand

$$\frac{\log N(T_n)}{T_n} \leq C + \varepsilon$$

En combinant ces deux inégalités, on obtient

$$C + \varepsilon \geq \frac{n(H - \varepsilon)}{T_n}$$

et donc

$$\frac{T_n}{n} \geq \frac{H - \varepsilon}{C + \varepsilon}.$$

D'où $\liminf_{n \rightarrow \infty} \frac{T_n}{n} \geq \frac{H}{C}$ et donc $\limsup_{n \rightarrow \infty} \frac{n}{T_n} \leq \frac{C}{H}$

Montrons maintenant la première partie : Pour n grand, on divise l'ensemble des messages de n symboles produits par la source en deux groupes, l'un contenant moins de $2^{n(H+\eta)}$ messages est l'ensemble des "messages typiques". L'autre, son complémentaire, contient moins de 2^{Rn} messages, où R est le logarithme du nombre de symboles différents. En effet, on a $\text{Card}(X)$ symboles, donc $\text{Card}(X)^n = 2^{n \log \text{Card}(X)}$ suites possibles de n symboles. La probabilité du second groupe (les messages non typiques) est plus petite que μ . Quand n tend vers l'infini, on peut faire tendre η et μ vers zéro.

Par ailleurs, on peut envoyer par le canal un nombre de messages différents de durée T supérieure

ou égale à $2^{(C-\theta)T}$, où θ est arbitrairement petit quand T est grand. Ceci découle de la définition de la capacité. On choisit alors

$$T = n\left(\frac{H}{C} + \lambda\right).$$

On a donc dans ce temps la possibilité d'envoyer $2^{n(C-\theta)(\frac{H}{C}+\lambda)}$ messages. On peut donc coder chacun des messages typiques de la source, à condition que $(H + \eta) \leq (C - \theta)(\frac{H}{C} + \lambda)$, soit

$$\lambda \geq \frac{1}{C - \theta}(\eta + \theta\frac{H}{C}),$$

condition réalisable avec λ arbitrairement petit pour n grand. Tous les messages typiques de longueur n sont donc transmissibles en temps T . Reste à transmettre les messages non typiques, qui sont au plus 2^{nR} . En reprenant le raisonnement précédent, on voit qu'on peut les transmettre en un temps $T_1 = n(\frac{R}{C} + \varphi)$, où φ peut être arbitrairement petit pour n assez grand. Les messages non typiques sont beaucoup plus nombreux que les messages typiques et demandent un temps de transmission plus long. Mais ils sont rares. On peut donc estimer le temps moyen de transmission d'un message de n symboles issu de la source, l'espérance étant calculée sur l'espace produit $\mathcal{X}\mathcal{X}^n$:

$$E\tilde{T}_n \leq (1 - \mu)T + \mu T_1 = n[(1 - \mu)(\frac{H}{C} + \lambda) + \mu(\frac{R}{C} + \varphi)].$$

Si on divise ce temps moyen par n on obtient le temps moyen de transmission par symbole et son inverse est le nombre de symboles transmis par unité de temps. Comme μ, λ, φ peuvent être rendus arbitrairement petits pour n grand, on obtient comme annoncé

$$\liminf \frac{n}{E\tilde{T}_n} \geq \frac{C}{H}.$$

□

Exercice 26.2.5. Messages et sorties typiques pour un canal bruité (Cover-Thomas 8.6 page 195) On considère une source, c'est-à-dire une variable aléatoire X discrète de loi $p(x)$, $x \in \mathcal{X}\mathcal{X}$. Un canal de transmission bruité transmet X avec erreur. Le résultat est une variable aléatoire Y de loi $p(y)$. On note $p(x, y)$ la loi jointe de x et de y . Dans le cas d'un canal non bruité, on aurait $p(x, x) = p(x)$ pour tout x , $p(x, y) = 0$ pour $x \neq y$. On note $H(X)$, $H(Y)$ et $H(X, Y)$ les entropies de X , de Y et de (X, Y) respectivement. Cette dernière entropie s'appelle l'entropie jointe de X et Y . Si la source émet une séquence de n messages indépendants et équidistribués, on note X^n la variable aléatoire correspondante et Y^n la sortie (*output*). Les valeurs de X^n se notent $x^n \in \mathcal{X}\mathcal{X}^n$ et celles de Y^n se notent $y^n \in \mathcal{Y}^n$.

Définition 26.2.2. On appelle ensemble de couples message-sortie typiques relativement à la distribution $p(x, y)$ l'ensemble A_ε^n des suites $\{(x^n, y^n)\}$ qui ont leurs entropies empiriques proches à ε près des entropies réelles, à savoir que $A_\varepsilon^n = B_\varepsilon^n \cap C_\varepsilon^n \cap D_\varepsilon^n$ avec

$$B_\varepsilon^n = \{(x^n, y^n) \in \mathcal{X}\mathcal{X}^n \times \mathcal{Y}^n : |-\frac{\infty}{n} \log \sqrt{(\S^\dagger)} - \mathcal{H}(\mathcal{X})| < \varepsilon\}, \quad (26.2.2)$$

$$C_\varepsilon^n = \{(x^n, y^n) \in \mathcal{X}\mathcal{X}^n \times \mathcal{Y}^n : |-\frac{\infty}{n} \log \sqrt{(\dagger^\dagger)} - \mathcal{H}(\mathcal{Y})| < \varepsilon\}, \quad (26.2.3)$$

$$D_\varepsilon^n = \{(x^n, y^n) \in \mathcal{X}\mathcal{X}^n \times \mathcal{Y}^n : |-\frac{\infty}{n} \log \sqrt{(\S^\dagger, \dagger^\dagger)} - \mathcal{H}(\mathcal{X}, \mathcal{Y})| < \varepsilon\}, \quad (26.2.4)$$

où $p(x^n) = \prod_{i=1}^n p(x_i)$, $p(y^n) = \prod_{i=1}^n p(y_i)$, $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$ en raison de l'indépendance des n messages successifs.

Lemme 26.2.2. Soit (X^n, Y^n) une suite de longueur n de v.a.i.i.d. selon la loi $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$. Alors

1. $\mathbb{P}((X^n, Y^n) \in A_\varepsilon^n) \rightarrow 1$ quand $n \rightarrow \infty$.
2. $\text{Card}(\{y^n, (x^n, y^n) \in A_\varepsilon^n\}) \leq 2^{n(H(Y|X)+2\varepsilon)}$
3. $\text{Card}(\{x^n, (x^n, y^n) \in A_\varepsilon^n\}) \leq 2^{n(H(X|Y)+2\varepsilon)}$

Démonstration Montrons la première relation. Notons $p_1 : (x^n, y^n) \rightarrow x^n$ et $p_2 : (x^n, y^n) \rightarrow y^n$ les projections sur X^n et Y^n respectivement. Le résultat de l'exercice 2 nous assure que quand $n \rightarrow \infty$,

$$\mathbb{P}(B_\varepsilon^n) \rightarrow 1 \quad (26.2.5)$$

$$\mathbb{P}(p_1(C_\varepsilon^n) \times Y^n) = \mathbb{P}(p_1(C_\varepsilon^n)) \rightarrow 1 \quad (26.2.6)$$

$$\mathbb{P}(X^n \times p_2(D_\varepsilon^n)) = \mathbb{P}(p_2(D_\varepsilon^n)) \rightarrow 1. \quad (26.2.7)$$

La relation 1 découle alors de la remarque générale que si des suites d'ensembles B_i^n , $i = 1, \dots, k$ vérifient $\mathbb{P}(B_i^n) \rightarrow 1$ quand $n \rightarrow \infty$, alors $\mathbb{P}(\cap_{i=1}^k B_i^n) \rightarrow 1$ aussi. On applique cette remarque à l'intersection des trois ensembles précédents,

$$A_\varepsilon^n = B_\varepsilon^n \cap (p_1(C_\varepsilon^n) \times Y^n) \cap (X^n \times p_2(D_\varepsilon^n)).$$

Montrons maintenant la relation 2. On a par hypothèse $\mathbb{P}(X^n = x^n) = p(x^n) = 2^{-n(H(X)\pm\varepsilon)}$ et $\mathbb{P}((X^n, Y^n) = (x^n, y^n)) = p(x^n, y^n) = 2^{-n(H(X,Y)\pm\varepsilon)}$. Donc

$$p(x^n) = \sum_{y^n} p(x^n, y^n) \geq \sum_{y^n, (x^n, y^n) \in A_\varepsilon^n} p(x^n, y^n) \geq \text{Card}(\{y^n, (x^n, y^n) \in A_\varepsilon^n\}) \inf_{(x^n, y^n) \in A_\varepsilon^n} p(x^n, y^n).$$

Donc

$$\text{Card}(\{y^n, (x^n, y^n) \in A_\varepsilon^n\}) \leq 2^{-n(-H(X,Y)+H(X)+2\varepsilon)} = 2^{n(H(Y|X)+2\varepsilon)}.$$

□

Exercice 26.2.6. Capacité d'un canal bruité On définit un canal discret comme un système disposant d'un alphabet d'entrée \mathcal{X} et d'un alphabet de sortie \mathcal{Y} et d'une matrice de probabilités de transition $p(y|x)$ donnant la probabilité d'observer le symbole y en sortie quand le symbole x a été envoyé. Un tel canal est dit "sans mémoire" car la distribution de probabilité de la sortie à chaque temps dépend uniquement de l'entrée et est indépendante des entrées et sorties précédentes. Shannon page 70 : On commence par définir le *rate of transmission* du canal (le débit) pour une source X et la sortie correspondante Y par

$$R = I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y).$$

The first defining expression has already been defined as the amount of information sent less the uncertainty of what was sent. The second measures the amount received less the part of this which is due to noise. The third is the sum of the two amounts less the joint entropy and therefore in a sense is the number of bits per second common to the two. Thus, all three expressions have a certain intuitive significance.

La capacité d'un canal va être définie comme le débit maximal pour toutes les sources possibles X .

Définition 26.2.3. On appelle capacité d'un canal discret sans mémoire la quantité

$$C = \max_{p(x)} I(X; Y),$$

le maximum étant pris sur toutes les sources, i.e. toutes distributions de probabilité possibles $p(x)$, $x \in \mathcal{X}$ à l'entrée.

Exemple 1 Si le canal transmet intégralement une entrée binaire sans erreur, la matrice de transition est l'identité. On a $Y = X$ et donc $I(X; X) = H(X) - H(X|X) = H(X)$. Donc la capacité est maximisée quand l'entropie de la source binaire est maximisée, ce qui donne comme attendu $p(0) = p(1) = \frac{1}{2}$ et $C = H(p) = 1$ bit.

Exemple 2 : canal binaire symétrique On prend $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ et

$$p(y = 1|x = 1) = p(y = 0|x = 0) = 1 - p, \quad p(y = 1|x = 0) = p(y = 0|x = 1) = p.$$

On rappelle que l'entropie d'un canal binaire (variable de Bernoulli $(p, 1-p)$), est $H(p) = -p \log p - (1-p) \log(1-p)$. On a

$$H(Y) - H(Y|X) = H(Y) - \sum p(x) H(Y|X=x) = H(Y) - \sum p(x) H(p) = H(Y) - H(p) \leq 1 - H(p).$$

On obtient l'égalité en prenant X uniforme, car alors Y l'est aussi et on a $H(Y) = 1$. Donc $C = 1 - H(p)$.

Exercice 26.2.7. Le théorème fondamental pour un canal discret bruité

Théorème 26.2.2. (Shannon pages 71, 72, 73) Soit un canal discret de capacité C et une source discrète d'entropie R .

(a) Si $R \leq C$ il existe un système de codage tel que la sortie de la source puisse être transmise sur le canal avec une fréquence arbitrairement petite d'erreurs.

(b) Si $R > C$ il est possible de coder la source de manière que l'incertitude sur les messages $H(X|Y)$ soit inférieure à $R - C + \varepsilon$ où ε est arbitrairement petit.

(c) Il n'y a pas de méthode de codage qui donnerait une incertitude sur les messages $H(X|Y)$ inférieure à $R - C$.

Démonstration (a) Soit une source X_0 de débit approchant la capacité maximale C . On va l'utiliser comme entrée dans le canal. Considérons toutes les suites possibles transmises et reçues, d'une durée assez longue n . Dans toute la suite, les ε , η , θ désignent des réels positifs qui tendent vers zéro quand le débit de la source X_0 se rapproche de la capacité maximale du canal ou quand la durée de transmission $n \rightarrow \infty$. Pour ne pas alourdir les notations, on continuera à noter ε pour $C\varepsilon$ quand C est une constante indépendante de n . Enfin on notera $2^{a \pm \varepsilon}$ un nombre quelconque b vérifiant $2^{a-\varepsilon} \leq b \leq 2^{a+\varepsilon}$.

Nous considérons l'ensemble A_ε^n des couples typiques (x^n, y^n) de couples typiques formés eux-mêmes de messages typiques envoyés et de messages typiques reçus (Définition 26.2.2).

1. Les suites transmises sont dans deux groupes : les messages typiques, dont le nombre est environ $2^{n(H(X) \pm \varepsilon)}$ et les autres, dont la probabilité totale est inférieure à η .

2. De même, les suites reçues forment un ensemble de suites typiques de nombre $2^{n(H(Y)\pm\varepsilon)}$ et de probabilité totale supérieure à $1 - \eta$. On note \mathcal{M}_0 cet ensemble de messages typiques.
3. Par le lemme 26.2.2, chaque sortie typique peut être produite par au plus $2^{n(H(X|Y)+\varepsilon)}$ entrées.
4. De même, chaque entrée typique peut produire au plus $2^{n(H(Y|X)\pm\varepsilon)}$ sorties (on ne va pas utiliser cette dernière propriété.)

Soit maintenant une autre source X émettant avec un débit $R < C$ (i.e. ayant une entropie inférieure à C .) On pose $R = C - \theta$. Dans le temps n , cette source peut avoir $2^{n(R\pm\varepsilon)}$ messages typiques. Nous noterons \mathcal{M} cet ensemble de messages typiques. Nous allons coder ceux-ci en les associant à des messages typiques de longueur n de la source X_0 . Ces messages seront donc utilisés comme codes. Chaque codage est donc une application $\mathbb{C} : \mathcal{M} \rightarrow \mathcal{M}_0$ obtenue en tirant au hasard pour chaque message dans \mathcal{M} (avec distribution uniforme) un élément de \mathcal{M}_0 qui devient son code. Les autres messages, non typiques, ne sont pas codés, ce qui donne une probabilité d'erreur plus petite que η . Supposons qu'on observe une sortie y_1 . On va maintenant évaluer la probabilité d'erreur : quelle est la probabilité que y_1 ait été associée à plus d'un message issu de \mathcal{M} ? Par (3), on sait que y_1 a pu être produite par au plus $2^{n(H(X|Y)+\varepsilon)}$ messages x_0 appartenant à \mathcal{M}_0 . La probabilité que chaque $x_0 \in \mathcal{M}_0$ soit un code est $2^{n(R-H(X)\pm\varepsilon)}$, car on a distribué $2^{n(R\pm\varepsilon)}$ messages uniformément dans $2^{n(H(X)\pm\varepsilon)}$ codes. Donc la probabilité que y_1 soit issu d'un message de X (en dehors du message qui l'a provoqué) est majorée de la manière suivante

$$\mathbb{P}(\text{erreur sur } y_1) \leq 2^{n(R-H(X)\pm\varepsilon)} 2^{n(H(X|Y)+2\varepsilon)}.$$

Mais $R = H(X) - H(X|Y) - \theta$, avec $\theta > 0$. Donc

$$\mathbb{P}(\text{erreur sur } y_1) \leq 2^{n(R-H(X)+\varepsilon+H(X|Y)+2\varepsilon)} = 2^{-n(\theta-3\varepsilon)}$$

Comme η a été fixé (aussi petit qu'on veut) et que l'on peut à η fixé prendre ε aussi petit qu'on veut pour n assez grand, on déduit que la probabilité d'erreur pour chaque message est arbitrairement petite, ce qui prouve (a) pour $R < C$.

(b) Si $R \geq C$, on peut encore appliquer la construction précédente en choisissant de ne coder que $2^{n(C-\varepsilon)}$ messages des $2^{n(R\pm\varepsilon)}$ typiques. On "jette" les autres. Ce n'est quand même pas très intéressant, puisque la plupart des messages typiques ne sont pas transmis!

(c) Supposons que l'on puisse transmettre par un canal de capacité C les messages d'une source X d'entropie $R = C + a$, avec $a > 0$ et que l'incertitude sur le message vérifie $H(Y|X) = a - \varepsilon$ avec $\varepsilon > 0$. Alors $H(X) - H(X|Y) = C + \varepsilon$ et cela contredit la définition de C comme le maximum de $H(X) - H(X|Y)$ pour toutes les sources en entrée.

La question de la fin.

Expliquer le raisonnement suivant de Shannon, qui conclut la démonstration précédente :

Actually more has been proved than was stated in the theorem. If the average of a set of positive numbers is within ε of zero, a fraction of at most $\sqrt{\varepsilon}$ can have values greater than $\sqrt{\varepsilon}$. Since ε is arbitrarily small we can say that almost all the systems are arbitrarily close to the ideal.

En d'autres termes, presque tous les codes considérés précédemment sont des codes correcteurs d'erreurs : on peut tirer un code au hasard et il fonctionnera.

Alors, pourquoi est-il difficile en pratique de fabriquer un code correcteur d'erreur ?

26.3 Entropie, entropie relative, information mutuelle

Conseils de lecture :

Claude E. Shannon et Warren Weaver The mathematical theory of communication University of Illinois Press 1998. (Très bon marché, commander par Amazon.com par exemple).

Thomas M. Cover et Joy A. Thomas Elements of Information Theory, Wiley Series Telecommunications, Chapitres 2, 5 et 8 (ces chapitres forment un tout et peuvent être lus indépendamment du reste)

Pierre Brémaud Introduction aux probabilités, Springer.

Exercice 26.3.1. La définition de l'entropie, opus cit. p. 49 Soit un ensemble fini d'événements dont les probabilités sont p_1, \dots, p_n . On connaît les p_i , mais on est dans l'incertitude sur quel événement va se produire. L'entropie $H(p_1, \dots, p_n)$ va être définie comme une mesure de l'incertitude qui nous est laissée. Afin de comprendre les axiomes qui vont mener à la définition de l'entropie, il faut réaliser qu'une présentation d'un ensemble d'événements disjoints peut faire l'objet de regroupements et être donc présenté en arbre : par exemple, on peut regrouper les k premiers événements en un seul événement de probabilité p'_1 et on obtient donc les deux suites de probabilités $p'_1 = \sum_{i=1}^k p_i, p_{k+1}, \dots, p_n$, le premier événement étant décomposé à son tour en π_1, \dots, π_k avec $\pi_i = \frac{p_i}{p_1 + \dots + p_k}$. Dans un cas on a n événements et dans l'autre un premier tirage entre $n - k + 1$ événements disjoints, suivi d'un deuxième tirage conditionnel entre k événements. On a donc deux présentations des mêmes événements :

(p_1, \dots, p_n) , ou bien

$((p'_1, \pi_1), \dots, (p'_1, \pi_k), p_{k+1}, \dots, p_n)$

On pose les axiomes intuitifs suivants.

1. H est continue
2. Supposons que les p_i sont égaux, $p_i = \frac{1}{n}$. Alors H doit être une fonction croissante de n
3. Si on recompose les n événements comme expliqué ci-dessous, il est clair que l'incertitude finale est la même. On exige donc

$$H(p_1, \dots, p_n) = H(p'_1, p_{k+1}, \dots, p_n) + p'_1 H(\pi_1, \dots, \pi_k).$$

Il s'agit de montrer que sous les axiomes 1, 2 et 3, on a pour une constante positive K ,

$$H(p_1, \dots, p_n) = -K \sum_{i=1}^n p_i \log p_i.$$

1) On pose $H(\frac{1}{n}, \dots, \frac{1}{n}) = A(n)$. Montrer d'abord en utilisant l'axiome 3 que $A(s^m) = mA(s)$. Si t est un autre entier, on peut trouver pour n aussi grand qu'on veut m tel que $s^n \leq t^n < s^{m+1}$. En déduire en utilisant la monotonie (axiome 2) que $\frac{m}{n} \leq \frac{A(t)}{A(s)} \leq \frac{m}{n} + \frac{1}{n}$ et finalement que $A(t) = K \log t$.

2) Supposons que $p_i = \frac{n_i}{\sum_1^n n_i}$ sont des probabilités commensurables. Montrer en utilisant à nouveau l'axiome 3 que

$$K \log \sum n_i = H(p_1, \dots, p_n) + K \sum p_i \log n_i.$$

3) Traiter ensuite le cas non commensurable en approchant les p_i par des rationnels et en utilisant l'axiome 2 de continuité.

Exercice 26.3.2. Entropie relative et information mutuelle

Soit deux distributions de probabilité $p(x)$ et $q(x)$. On appelle distance de Kullback Leibler, ou entropie relative des deux distributions $p(x)$ et $q(x)$ la quantité

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} = \mathbb{E}_p \log \frac{p(X)}{q(X)}.$$

L'entropie relative est une mesure de la distance entre deux distributions.

Soient deux variables aléatoires discrètes X et Y avec une distribution jointe $p(x, y)$. On note $p(x) = \sum_{y \in \mathcal{X}} p(x, y)$ et $p(y) = \sum_{x \in \mathcal{X}} p(x, y)$ les deux distributions marginales de X et Y respectivement. Alors on appelle information mutuelle $I(X, Y)$ l'entropie relative entre la distribution jointe $p(x, y)$ et la distribution produit $p(x)p(y)$,

$$I(X, Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = D(p(x, y)||p(x)p(y)) = \mathbb{E}_{p(x, y)} \log \frac{p(X, Y)}{p(X)p(Y)}.$$

Démontrer que

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X),$$

$$I(X, Y) = I(Y, X) = H(X) + H(Y) - H(X, Y), \quad I(X, X) = H(X).$$

Remarquer que $I(X, Y) = 0$ si X et Y sont indépendantes.

Soient X_1, X_2, \dots, X_n des variables aléatoires discrètes de distribution jointe $p(x_1, x_2, \dots, x_n)$. Montrer en utilisant uniquement (??) que

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1). \tag{26.3.1}$$

Montrer en utilisant les définitions de D et de la probabilité conditionnelle que

$$D(p(x, y)||q(x, y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x)).$$

Exercice 26.3.3. On rappelle l'inégalité de Jensen : Soit f une fonction convexe et X une variable aléatoire discrète, alors

$$\mathbb{E}f(X) \geq f(\mathbb{E}X).$$

Si f est strictement convexe, l'égalité n'est vraie que si $X = \mathbb{E}X$ presque sûrement, i.e. X est constante. Montrer cette inégalité pour une variable aléatoire discrète.

Soient $p(x)$ et $q(x)$, $x \in \mathcal{X}\mathcal{X}$ deux distributions de probabilité discrètes. En utilisant l'inégalité de Jensen appliquée au logarithme, qui est strictement concave, montrer que

$$D(p||q) \geq 0$$

et que l'égalité est vraie si et seulement si $p(x) = q(x)$ pour tout x .

En déduire que pour tout couple de variables aléatoires, (X, Y) , on a

$$I(X, Y) \geq 0$$

et égalité si et seulement si X et Y sont indépendantes.

En déduire aussi que $H(X) \leq \log |\mathcal{X}\mathcal{X}|$, où $|\mathcal{X}\mathcal{X}|$ représente le nombre d'éléments dans l'image de X et il y a égalité si et seulement si X a une distribution uniforme sur $\mathcal{X}\mathcal{X}$. (Exprimer que $D(p||u) \geq 0$, où $u(x) = \frac{1}{|\mathcal{X}\mathcal{X}|}$ est la distribution uniforme sur $\mathcal{X}\mathcal{X}$).

Déduire aussi de la positivité de l'entropie relative que $H(X|Y) \leq H(X)$.

Déduire finalement que

$$H(X_1, X_2, \dots, X_n) \leq \sum_{i=1}^n H(X_i),$$

et que l'on a égalité si et seulement si les X_i sont indépendantes. (Utiliser la question précédente et (26.3.1)).

Bibliographie

- [1] Baldi P., Mazliak L., Priouret P., *Martingales et chaînes de Markov*, Hermann, (2000)
- [2] Bauer, *Probability Theory*, (****)
- [3] Billingsley P., *Probability and Measure*, Wiley, (1979)
- [4] Bony J.M., *Cours d'Analyse*, Ellipses, (****)
- [5] Borkar V., *Probability Theory*, Springer, (****)
- [6] Brémaud P., *An Introduction to Probabilistic Modeling*, Springer, (1997)
- [7] Brémaud P., *Introduction aux probabilités : Modélisation des phénomènes aléatoires*, Springer, (1997)
- [8] Brémaud P., *Markov Chains : Gibbs fields, Monte Carlo simulation, and queues*, Springer Verlag, (1999)
- [9] Brzesniak et Zastawniak, *Basic Stochastic Processes*, (****)
- [10] Buchwalter H., *Le Calcul inégal*, Ellipses, (1991)
- [11] Chambert-Loir A., *Exercices d'analyse pour l'agrégation I/II/III*, Masson (1994/95/96)
- [12] Dacuhna, Castelle et Duflo, *Probability and Statistics*, (****)
- [13] J. Dieudonné, *Calcul infinitésimal*, Hermann, (1967)
- [14] Durrett R., *Probability : Theory and examples*, Duxburry, (1995)
- [15] Durrett R., *Essential of Stochastic Processes*, Springer, (1999)
- [16] Feller W., *An Introduction to Probability Theory and Its Applications*, Vol. I/II, Wiley and Sons, Third Edition (1970)
- [17] Foata D., Fuchs A., *Processus stochastiques : Processus de Poisson, chaînes de Markov et martingales*, Dunod, (2002)
- [18] Grimmett G., Stirzaker D., *Probability and Random Processes*, Oxford University Press, Third Edition, (2001)
- [19] Häggström O., *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, (2002)
- [20] Hauchecorne B., *Les contre-exemples en mathématiques*, Ellipses, (1988)
- [21] Karlin S., Taylor H.M., *A First Course in Stochastic Processes/ A Second Course in Stochastic Processes*, Springer, (1999)
- [22] Kingman J.F.C., *Poisson processes*, Oxford Science Publications, (1993)
- [23] Kipnis C., Landim C., *Scaling limits of interacting particle systems*, Springer, (1999)

- [24] Lamberton D., Lapeyre P., *Introduction au calcul stochastique appliqué à la finance*, Ellipses, (1999)
- [25] Liggett T., *Interacting Particle Systems*, Springer Verlag, (1985)
- [26] J.-Y. Oувrard, *Probabilités 2, maîtrise-agrégation*, Cassini, (2000)
- [27] W. Rudin, *Analyse réelle et complexe*, Masson, (1975)
- [28] Taylor, *Introduction to measure and probability*, (****)
- [29] S.R.S. Varadhan, *Probability Theory*, Courant Lecture Notes in Mathematics, (2001)
- [30] D. Williams, *Probability with martingales*, Cambridge University Press, (1991)