

Théorie des probabilités

Sébastien MARTINEAU*

Table des matières

1	Vocabulaire des probabilités	5
1.1	Espace probabilisé	5
1.1.1	Intuition et formalisme	5
1.1.2	Pourquoi parler de mesure ?	6
1.2	Variables aléatoires	7
1.2.1	Pourquoi oublier l'univers ?	7
1.2.2	Loi d'une variable aléatoire	8
1.2.3	Le changement de variable	9
1.2.4	Rappels d'intégration	9
1.3	Premiers exercices	10
1.4	Indépendance	11
1.4.1	Mais qu'est-ce que ça veut dire ?	11
1.4.2	Tribu et information	12
1.4.3	Définition générale de l'indépendance	13
1.4.4	Attirail pratique et théorique	15
1.5	Les premières fonctions associées à une loi	19
1.5.1	La fonction génératrice	19
1.5.2	La fonction de répartition	20
1.6	Exercices, deuxième fournée	22
2	Intermède d'application des séries génératrices	26
2.1	Le processus dit de Galton-Watson	26
2.2	Un problème de dés	29
3	Convergences en théorie des probabilités	31
3.1	Des outils toujours utiles	31
3.1.1	Quelques inégalités	31
3.1.2	Lemmes de Borel-Cantelli	32
3.1.3	Loi du 0-1 de Kolmogorov	34
3.2	Convergence de variables définies sur le même espace probabilisé	35
3.3	Loi des grands nombres	38

*Bureau 420 Sud

3.3.1	Premières versions	38
3.3.2	Fonctions de répartition empiriques et théorème de Glivenko-Cantelli	41
3.3.3	Lois des grands nombres L^1	42
3.3.4	Méthode de Monte-Carlo	44
3.3.5	Preuve de la loi forte des grands nombres	45
3.4	Convergence en loi	47
3.5	Fonction de répartition, le retour	50
3.5.1	Un peu de théorie	51
3.5.2	Quelques exercices	52
3.5.3	Applications concrètes	53
3.5.4	Tension et compacité	56
3.6	Fonctions caractéristiques	58
3.7	Le théorème central limite unidimensionnel	63
3.7.1	Commentaires	63
3.7.2	Et concrètement	64
3.7.3	Démonstration	64
4	Les marches aléatoires	66
4.1	Retours en 0 et sommabilité	66
4.2	Marche aléatoire simple et fonctions caractéristiques	68
4.3	Le cas unidimensionnel sans calcul	70
4.4	Exercices	71
5	Supplément	73
5.1	Espérance conditionnelle	73
5.2	Compléments de cours	74
5.2.1	Transformée de Laplace et grandes déviations	74
5.2.2	Autres points et références	75
5.3	Matière à lire et à penser	76

Dans ce texte, les commentaires¹ ayant trait à l'intuition plutôt qu'aux mathématiques formelles seront en petits caractères.

Objectifs

L'objectif de ce cours est de présenter le corps commun de probabilités du programme de l'agrégation. Aucune connaissance des probabilités n'est prérequis. Cependant, on supposera acquise une certaine familiarité avec la théorie de l'intégration.

On ne démontrera pas tous les résultats qui seront énoncés. Certains seront admis, d'autres ne verront leur preuve qu'esquissée. Néanmoins, les définitions et théorèmes apparaîtront — sauf cas contraire explicite — dans un ordre mathématiquement cohérent du point de vue des démonstrations.

On essaiera d'insister sur les exemples et applications, et les exercices seront là pour vous familiariser concrètement avec les objets et raisonnements probabilistes.

1. parfois un peu trop digressifs

Deux résultats pour se motiver

Le cadre théorique des probabilités est, comme c'est souvent le cas en mathématiques, beaucoup plus compliqué que le concept intuitif auquel il se rapporte. Avant d'aller batifoler au pays des σ -algèbres, il me semble opportun de présenter deux jolis résultats qu'on sera capable de formuler mathématiquement et établir au cours du trimestre.

Marche aléatoire

Considérons \mathbb{Z}^d comme l'ensemble des sommets d'un graphe. Disons que deux sommets sont voisins s'ils diffèrent exactement selon une coordonnée, et d'exactly ± 1 ; bref, a et b sont voisins si $\|a - b\|_1 = 1$. Définissons informellement la marche aléatoire partant de 0. Au temps 0, le marcheur est à son point de départ, l'origine. Il est armé d'un dé à $2d$ faces, qui lui permet de tirer au sort équiprobablement un de ses $2d$ voisins quand il le souhaite. La procédure est alors simple : entre le temps n et le temps $n + 1$, le marcheur lance son dé et se déplace vers la destination indiquée.

Un théorème dû à Polya indique que

- si $d \leq 2$, avec probabilité 1, le marcheur visitera chaque sommet un nombre infini de fois.
- si $d \geq 3$, avec probabilité 1, le marcheur visitera chaque sommet un nombre fini de fois ; de plus, avec probabilité positive, le marcheur ne reviendra jamais en 0.

Ivre, l'homme toujours regagnera sa cour tandis que l'hippocampe de son foyer décampe.

Processus de Galton-Watson

Ce modèle très simple a été introduit pour comprendre l'extinction des noms de famille des nobles en Grande-Bretagne. On dispose d'un dé éventuellement biaisé, qui renvoie m en moyenne. On dispose aussi d'un noble, qui est le seul porteur de son nom de famille. Il jette le dé, a autant d'enfants que le nombre de points² qu'il a tiré puis meurt. Et ses enfants jouent au même jeu. La lignée a-t-elle une chance de survivre indéfiniment ?

Comme dans le théorème précédent, on a une dichotomie. Plus précisément,

- si $m < 1$, la lignée s'éteindra avec probabilité 1 ;
- si $m > 1$, la lignée a une probabilité strictement positive de ne jamais s'éteindre.

2. peut-être nul, il est très biaisé le dé

1 Vocabulaire des probabilités

L'objectif de cette partie est de définir les notions qu'on manipulera sans cesse par la suite et de construire le socle de base sur lequel se bâtissent les probabilités.

1.1 Espace probabilisé

DÉFINITION. On appelle *espace probabilisé* un espace muni d'une tribu et d'une mesure de probabilité (i.e. une mesure de masse totale 1) définie sur cette tribu. La notation standard est $(\Omega, \mathcal{F}, \mathbb{P})$. On appelle Ω l'*univers*. Un élément ω de Ω est appelé *éventualité*. Un élément de \mathcal{F} est appelé *événement*.

Aux premières définitions succèdent les premières digressions.

1.1.1 Intuition et formalisme

Essayons de concilier intuition et formalisme. Ceci est souvent une étape délicate, susceptible d'embrouiller ; aussi, si cette étape vous semble acquise, peut-être gagneriez-vous à sauter ce paragraphe.

Mettons-nous tout d'abord d'accord sur la notion intuitive de probabilité. Je jette un dé. Je dis que le résultat est aléatoire car si je jette vingt fois le dé, je n'arriverai toujours pas à prédire de façon convainquante le prochain résultat (réussir à prédire "une fois sur six" n'est pas convainquant à mon sens). En revanche, sur un grand nombre de lancers, je constate empiriquement que la proportion de 1 — de même que celle de n'importe quel autre chiffre — tend vers $1/6$. Je définis ce nombre comme la probabilité (au sens intuitif) d'obtenir 1 — ou n'importe quel autre chiffre. Une fois ce concept assimilé, je ne me ramène pas toujours à de longues séries d'expériences pour trouver les valeurs des probabilités ; souvent, quand le mécanisme sous-jacent me semblera présenter suffisamment de symétries pour que toutes les issues jouent le même rôle, je partirai du principe que ces issues sont équiprobables. Ce principe sera généralement en accord avec la réalité. Bref, la notion intuitive de probabilité n'est que le constat empirique de la convergence de certaines fréquences.

Comment penser intuitivement $(\Omega, \mathcal{F}, \mathbb{P})$? De même qu'un jet de dé donne un nombre aléatoire entre 1 et 6, je tire ici un élément aléatoire de Ω . C'est \mathbb{P} qui me définit la procédure selon laquelle cet élément est tiré, qui définit les probabilités intuitives associées à cette expérience aléatoire. Précisons comment. Si F est un élément de \mathcal{F} , je dirai que l'événement F a lieu si l'élément aléatoire tiré tombe dans cet ensemble F ⁽³⁾. La procédure selon laquelle l'élément est tiré est de telle sorte que la probabilité intuitive d'occurrence de l'événement F soit $\mathbb{P}[F]$.

Ainsi, $(\{1, \dots, 6\}, \text{tribu pleine}, \text{équiprobabilité})$ modélise le jet de dé.

Si, dépités, vous estimez qu'on ne sait toujours pas ce que veut dire aléatoire, rassurez-vous ; il n'y a rien de plus à comprendre que la notion intuitive floue ⁽⁴⁾, la définition mathématique formelle et le dictionnaire permettant de passer de l'une

3. on identifie ensemble et fonction indicatrice booléenne en quelque sorte

4. provenant du constat empirique de la validité de la loi des grands nombres

à l'autre. Quand Euclide formalisait la géométrie, il ne disait pas ce qu'était un point, ou une droite, mais comment ces objets interagissaient ; c'est un travail de nature semblable qui nous attend ici, puisqu'on donne des noms à des objets.

Commentaire. Dans le monde mathématique, l'ordre naturel de construction diffère souvent de celui dans lequel se bâtit l'esprit. Aussi, la loi des grands nombres ne sera pas en mathématiques une définition mais un théorème. N'est-ce pas absurde ? Cette loi des grands nombres qui donne leur sens aux définitions, comment pourrait-elle être considérée comme à démontrer ? Il y a deux réponses à apporter à cela. La première est que, dans le langage que nous allons développer, la loi des grands nombres pourra être énoncée de manière formelle ; elle dira *grosso modo* que “avec probabilité intuitive 1, la convergence des ratios a lieu”. Pourtant cet énoncé sera mathématiquement loin d'être évident ! Il faudra donc le démontrer, ne serait-ce que pour vérifier “à la chimiste” que notre théorie tient la route⁵. D'autre part, on peut aussi ne pas croire en la loi des grands nombres empirique, auquel cas on ne considérera la théorie des probabilités que comme une sous-branche de la théorie de la mesure où des concepts étranges sont définis et étudiés. La loi des grands nombres est alors pensée comme un énoncé mathématique qui ne fait intervenir aucun postulat de nature probabiliste ; il perd alors une part de son sens, mais pas de sa puissance mathématique. On peut par exemple utiliser la loi des grands nombres pour montrer qu'il existe des nombres qui sont univers en toute base.

Autre commentaire. Il n'est pas forcément clair que, étant donnée une mesure de probabilité \mathbb{P} , il existe une procédure concrète telle que pour tout F , la probabilité de tomber dans F soit $\mathbb{P}[F]$. (Dans ce commentaire, mise à part l'expression “mesure de probabilités”, les mots sont à comprendre au sens naïf, intuitif.) C'est sûr que sans les hypothèses disant que \mathbb{P} est une mesure de probabilités, c'est impossible, mais ces hypothèses suffisent-elles ? On démontrera plus tard que si on sait tirer une suite infinie de piles ou faces, une telle procédure existe toujours⁶. (Si ce cours est terminé et que vous le relisez, pensez à l'astuce de la fonction de répartition inverse.)

1.1.2 Pourquoi parler de mesure ?

Moralement, \mathbb{P} mesure et définit les probabilités des éventualités. Pourquoi donc \mathbb{P} est-elle définie sur les événements ?

Dès que l'univers est indénombrable, par exemple $([0, 1], \mathbf{Bor}([0, 1]))$, définir la probabilité de chaque éventualité suffit très rarement à dire quoi que ce soit d'intéressant ; on ne saurait alors pas différencier une uniforme sur $[0, 1]$ d'une sur $[0, \frac{1}{2}]$. Mesurer les probabilités des événements permet de contourner ce problème. Dans le cas “ $[0, 1]$ vs $[0, \frac{1}{2}]$ ”, l'événement $[0, \frac{1}{2}]$ différencie tout de suite les deux mesures évoquées.

Maintenant qu'on a compris, en passant par l'indénombrable, pourquoi on aura besoin du cadre de la théorie de la mesure, il est bon de voir comment ce cadre se concilie avec notre intuition dans le cadre dénombrable⁷.

5. Sous notre hypothèse, démontrons notre hypothèse !

6. pourvu que (Ω, \mathcal{F}) soit raisonnable : on le fera pour $(\mathbb{R}, \mathbf{Bor}(\mathbb{R}))$, mais ça marche en fait pour tout espace polonais muni de sa tribu borélienne

7. dans ces notes, fini sera inclus dans le cadre dénombrable

Soit \mathfrak{N} un ensemble dénombrable. Il sera implicitement muni de la tribu $P(\mathfrak{N}) = \text{Parties}(\mathfrak{N})$. Une probabilité est alors exactement la donnée d’une famille de poids $p(i)$ positifs sur \mathfrak{N} sommable de somme 1. Une analyse-synthèse directe indique que $\mathbb{P}(E) = \sum_{i \in E} p(i)$ et $p(i) = \mathbb{P}(\{i\})$ définissent une bijection entre ces deux mondes. La dénombrabilité est essentielle dans cet argument, les mesures étant *dénombrablement* additive (on dit plus souvent σ -additive).

Comment choisir la mesure de probabilité ?

Comment déterminer la mesure de probabilité ? La branche des mathématiques dévouée à la résolution de cette question s’appelle la *statistique*. Pour faire court, disons que des outils mathématiques sont mis en œuvre mais qu’on ne coupe pas à une part de bon sens, puisqu’il y a une phase de modélisation. Dans le reste de ce cours, cette question de l’adéquation entre \mathbb{P} et la réalité sera ignorée. Les théorèmes seront toujours justes, c’est seulement leur rapport à la réalité qui pourra être sujet à discussion. On pourra penser le choix de la mesure de probabilité comme une hypothèse.

1.2 Variables aléatoires

DÉFINITION. Une application mesurable de notre espace probabilisé vers un espace muni d’une tribu sera appelée une *variable aléatoire*. Dans le cas où la variable aléatoire est à valeurs complexes (on parlera de variables aléatoires complexes pour insister sur l’intuition “objet aléatoire”), on notera $\mathbb{E}[X]$ son intégrale (lorsqu’elle existe), et on l’appellera *espérance*. On dira qu’une condition a lieu *presque sûrement* si elle est satisfaite en tout point d’un événement de probabilité 1.

Remarque. L’appellation “variable” est naturelle si on pense son argument ω comme variable, indécis, tiré aléatoirement. Une fonction mangeant un argument indécis renverra généralement un résultat tout aussi incertain.

Sitôt défini, sitôt oublié : la majeure partie du temps, le probabiliste ne se soucie pas de l’univers dans lequel il travaille ! ⁽⁸⁾ Ce qui lui importe est la loi de la variable aléatoire qu’il étudie. Cette notion est introduite après le paragraphe de digressions. Il sera ensuite expliqué en quoi cette notion retient l’essence de la variable aléatoire et permet d’oublier l’espace probabilisé.

1.2.1 Pourquoi oublier l’univers ?

Il y a beaucoup de procédés naïfs différents pour obtenir un bit valant équivalemment 0 ou 1 : tirer à pile ou face, jeter un dé équilibré et considérer la

8. C’est pour ça que la notation $\mathbb{E}[X]$ ne fait apparaître ni Ω , ni \mathbb{P} .

parité du nombre de points, jeter n pièces et ne regarder que la dernière... Chacun correspondra à un univers différent (à 2, 6 ou 2^n points). Si on considère comme déterministe la dynamique d'évolution de la pièce mais comme aléatoires les influx nerveux, les intensités des contractions musculaires et la force du vent, la taille de l'univers considéré change encore considérablement.

Dans la pratique, si j'ai un bit aléatoire, son procédé de fabrication m'importe peu, seul m'importe de savoir s'il est biaisé ou non. ⁽⁹⁾ Dans le langage qu'on est en train de développer, on dira qu'on ne s'intéresse pas à l'espace probabilisé mais à la loi de la variable aléatoire considérée.

1.2.2 Loi d'une variable aléatoire

NOTATION. Souvent, on notera $\mathbb{P}[X \text{ vérifie quelque chose}]$ au lieu de

$$\mathbb{P}[\{\omega : X(\omega) \text{ vérifie quelque chose}\}]$$

De même, si X est une variable aléatoire et f une fonction mesurable définie là où X prend ses valeurs, on écrira souvent $f(X)$ au lieu de $f \circ X$. Cette notation nous incite à penser X comme une valeur indéterminée, une variable aléatoire.

DÉFINITION. La *loi* d'une variable aléatoire X est la mesure image de \mathbb{P} par elle (qui a un sens parce qu'on a imposé les variables aléatoires mesurables). C'est la mesure définie par

$$\mu(A) := \mathbb{P}[X^{-1}(A)]$$

C'est une mesure de probabilité. En cours de probabilités, on écrira $\mu(A) = \mathbb{P}[X \in A]$.

Remarque. Si je choisis une mesure de probabilité \mathbb{P} sur un espace (Ω, \mathcal{F}) , c'est bien la loi d'une variable aléatoire : il suffit de prendre pour espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et de considérer l'application identité !

La plupart des théorèmes de probabilité parleront de variables aléatoires et ne feront intervenir d'hypothèses que sur leur loi ; encore une fois, c'est la loi qui importe, pas l'univers. Ceci est si profond que les noms donnés aux variables aléatoires ne dépendent que de leur loi.

DÉFINITION. Une variable aléatoire telle que $\mathbb{P}[X = 1] = p = 1 - \mathbb{P}[X = 0]$ est appelé une variable aléatoire **de Bernoulli** de paramètre p . Une variable aléatoire à valeurs dans $[0, 1]$ de loi la mesure de Lebesgue est dite **uniforme**.

9. Son procédé de fabrication peut me renseigner à ce sujet

1.2.3 Le changement de variable

Tout d'abord, rappelons la notion de mesure image, essentiellement redéfinie lors de l'introduction de la notion de loi.

DÉFINITION. Soit (X, μ) un espace mesuré et $f : X \rightarrow Y$ une application mesurable. La *mesure image* de μ par f est la mesure sur Y définie par

$$A \mapsto \mu(f^{-1}(A))$$

On la notera $f_{\#}\mu$. Cette notation n'est pas universelle ; si vous l'employez, redéfinissez-la.

Le théorème qui suit, qu'on peut maintenant énoncer, illustre la puissance de la connaissance de la mesure image, de la loi.

THÉORÈME 1.1 (CHANGEMENT DE VARIABLES ABSTRAIT). Soit (X, μ) un espace mesuré et $f : X \rightarrow Y$ une application mesurable. Notons ν la mesure image de μ par f . Si $\phi : Y \rightarrow \mathbb{R}$ est mesurable, alors $\int_Y \phi(y) d\nu(y)$ a un sens ssi $\int_X \phi(f(x)) d\mu(x)$ en a un et, lorsque c'est le cas, ces quantités sont égales.

Esquisse de preuve. Lorsque ϕ est une fonction indicatrice, ceci est vrai par définition. On passe aux fonctions étagées par linéarité, puis aux fonctions mesurables positives par approximation et convergence monotone, et enfin aux fonctions intégrables en découpant en parties positive et négative. Pour les détails, consulter votre ouvrage d'intégration préféré. \square

Remarque. On appellera une fonction valant 0 ou 1 une fonction indicatrice d'un ensemble, et pas une fonction caractéristique. En effet, cette terminologie est déjà prise en probabilités, et correspond à une transformée de Fourier.

Remarque. Appliqué dans le cadre des probabilités, ce théorème nous indique que pour calculer l'espérance d'une fonction de X , la connaissance de la loi de X suffit. Ceci est suffisant pour traiter la majorité des questions qu'on se pose en probabilités.

1.2.4 Rappels d'intégration

Le théorème de changement de variable abstrait est essentiel ; le théorème de changement de variable C^1 , plus concret, l'est tout autant. Ce paragraphe a pour but de l'énoncer ; c'est un paragraphe d'intégration, non de probabilités.

NOTATIONS. La fonction indicatrice d'un ensemble $A \subset B$, valant 1 sur A et 0 sur $B \setminus A$, sera notée 1_A .

La mesure de Lebesgue sur \mathbb{R}^n (ou une partie borélienne de \mathbb{R}^n) sera désignée par λ_n ou λ . On écrira parfois dx au lieu de λ_1 .

DÉFINITION. Soit (X, μ) un espace mesuré et $g : X \rightarrow \mathbb{R}_+$ une application mesurable. On désigne par $g\mu$ la mesure sur X définie par

$$A \mapsto \int_X 1_A(x)g(x)d\mu(x) = \int_A g(x)d\mu(x)$$

On s'autorisera l'abus de notation consistant à parfois écrire $g(x)\mu$ au lieu de $g\mu$.

PROPOSITION 1.2. Soit (X, μ) et $g : X \rightarrow \mathbb{R}_+$ comme au-dessus. Soit $\phi : X \rightarrow \mathbb{C}$ mesurable. Alors, $\int_X \phi g d\mu$ a un sens ssi $\int_X \phi d(g\mu)$ en a un et, lorsque c'est le cas, ces quantités sont égales.

Cette proposition se prouve en suivant les mêmes étapes que dans la preuve du théorème de changement de variables abstrait.

THÉORÈME 1.3 (CHANGEMENT DE VARIABLES C^1). Soit $f : U \rightarrow V$ un difféomorphisme de classe C^1 entre deux ouverts de \mathbb{R}^n . On a

$$f_{\#}\lambda|_U = \frac{1}{\left|\det D_{f^{-1}(x)}f\right|} \cdot \lambda|_V$$

Ce théorème, très utile, est plutôt difficile à démontrer. Claude Danthony détient un polycopié très éclairant expliquant la preuve de ce résultat.

1.3 Premiers exercices

Remarque. Les exercices font partie du cours. Par exemple, ce n'est pas parce que la variance est définie dans cette sous-section que la connaissance de sa définition est facultative.

DÉFINITION. Soit $p \in [1, \infty]$. Une variable aléatoire complexe X est dite (dans) L^p si $\mathbb{E}[|X|^p] < \infty$.

PROPOSITION 1.4. Soit $p < q$ tels que $p \geq 1$. Si X est une variable aléatoire complexe L^q , alors elle est aussi L^p . En d'autres termes, pour les espaces de probabilité, $L^q \subset L^p$.

EXERCICE 1. Prouver cette proposition.

DÉFINITION. Soit X une variable aléatoire réelle L^2 . Notons m son espérance. On appelle *variance* l'espérance de $(X - m)^2$.

EXERCICE 2. 1. Calculer l'espérance et la variance des variables aléatoires que vous connaissez.

Dans les questions qui suivent, X est une variable aléatoire réelle L^2 .

2. Montrer que X est de variance nulle si et seulement si sa loi est une masse de Dirac (elle est concentrée sur un singleton).

3. Prouver que $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. Cette formule est à connaître.

4. Regarder la formule $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ droit dans les yeux jusqu'à voir les variables muettes qui devraient apparaître si on écrivait les intégrales et les variables d'intégration.

5. Montrer que $a \mapsto \mathbb{E}[(X - a)^2]$ est minimisée en $\mathbb{E}[X]$.

1.4 Indépendance

Dans cette sous-section, on travaillera dans un espace probabilisé fixé.

DÉFINITIONS. Soit A et B deux événements.

Si $\mathbb{P}[B] > 0$, on appelle *probabilité conditionnelle* de A par rapport à B (¹⁰) la quantité

$$\mathbb{P}[A|B] := \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

A et B sont dits *indépendants* si $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$.

Remarques. Dans les remarques suivantes, B est un événement de probabilité positive. (¹¹)

A et B sont indépendants ssi $\mathbb{P}[A|B] = \mathbb{P}[A]$. La symétrie de la relation d'indépendance, évidente par définition, est moins claire avec ce point de vue.

$A \mapsto \mathbb{P}[A|B]$ définit une mesure de probabilité donnant masse 1 à B .

Comme d'habitude, un petit peu de digression pour ceux qui veulent prendre le temps de digérer les définitions.

1.4.1 Mais qu'est-ce que ça veut dire ?

Commençons par une simple paraphrase. La probabilité de A conditionnellement à B , c'est la probabilité, sachant que j'ai observé B , d'avoir aussi observé A . La phrase qui suit définit plus précisément le concept intuitif de probabilité conditionnelle. Si dans la loi des grands nombres qui définit les probabilités intuitives, je ne retiens que les expériences où B a lieu, la proportion des expériences où A sera satisfait convergera vers la probabilité de A conditionnellement à B . Quand

10. on dit aussi la "probabilité de A conditionnellement à B ", et on lit souvent "probabilité de A sachant B "

11. en probabilités, on dit abusivement qu'un événement est de probabilité positive lorsque sa probabilité est strictement positive

je calcule cette proportion, je divise le nombre d'expériences où A et B ont lieu par le nombre d'expériences où B a lieu. Ceci justifie la formule de la définition de probabilité conditionnelle.

Si ces mots entrent en résonance avec vos images mentales, vous pouvez aussi penser la probabilité conditionnelle en termes de restriction, de projection ou d'exploitation exacte de l'information disponible.

Passons maintenant à l'indépendance. Deux événements sont indépendants au sens intuitif lorsque savoir que l'un a eu lieu ne donne pas d'information sur la probabilité d'occurrence de l'autre. La remarque suivant la définition formelle montre bien que le formalisme s'accorde avec cette définition intuitive.

Il est amusant de constater que le prédicat mathématique de base est l'indépendance, pas la dépendance ; ceci indique bien qu'il y a quelque chose à comprendre... Plus précisément, on aurait pu imaginer que deux événements étaient indépendants si, par définition, ils n'avaient "rien à voir". Par exemple deux jets de dé. Mais l'indépendance n'est pas toujours une hypothèse, dans le sens où elle peut surgir d'un mécanisme d'apparence dépendante. Un exemple concret rendra cette phrase plus claire. Tirons un pile ou face, et appelons A l'événement "pile". Si A a lieu, je tire un second pile ou face ; sinon, je jette un dé. Je définis l'événement B comme "soit j'ai tiré deux piles, soit j'ai tiré 1,4 ou 6 au dé". L'événement B , par sa définition, dépend de A ; il en est pourtant indépendant. C'est parce qu'elle englobe ce genre de considérations que la notion d'indépendance qu'on développe ici est beaucoup plus riche que l'absence, par définition, de dépendance.

1.4.2 Tribu et information

DÉFINITION. La *tribu* (ou σ -algèbre) engendrée par une variable aléatoire $X : (\Omega, \mathcal{F}, P) \rightarrow (\Sigma, \mathcal{G})$ est

$$\sigma(X) := \{X^{-1}(G); G \in \mathcal{G}\}$$

On peut étendre cela aux événements en identifiant un événement et sa fonction indicatrice.

Remarque. Il faut penser la tribu engendrée par une variable aléatoire X comme toute l'information qu'on peut distinguer en ne regardant que la valeur que X prend. En effet, si je ne peux regarder que la valeur prise par X , les seuls événements que je m'autorise à voir sont les

$$\{\omega \in \Omega : X(\omega) \in G\}$$

où G est mesurable, c'est-à-dire précisément les éléments de $\sigma(X)$.

Jusqu'à présent, les σ -algèbres semblaient n'être qu'un moyen de contourner l'impossibilité de définir la mesure de Lebesgue sur toutes les parties de \mathbb{R}^n ; par dépit, on se contentait alors de mesurer une large classe de parties, la tribu borélienne (ou lebesguienne, plus vaste encore). En particulier, on voulait une tribu la plus grande possible. Ici, la notion de tribu trouve un

second souffle en encodant de l'information. On aura maintenant affaire à des tribus de toutes tailles, et les petites seront ni plus ni moins intéressantes que les grandes.

Lorsqu'on parlera de telles tribus, elles seront toujours supposées constituées d'événements; en d'autres mots, on considérera toujours des sous-tribus de la σ -algèbre \mathcal{F} de l'univers.

Exemple. La tribu engendrée par un événement A est $\{\emptyset, A, \Omega \setminus A, \Omega\}$.

PROPOSITION 1.5. *Toute tribu sur un espace dénombrable est engendrée par une partition. Plus précisément, si \mathcal{F} est une σ -algèbre sur un ensemble Ω dénombrable, alors il existe une unique partition $\mathfrak{N} = \bigsqcup_{i \in \mathcal{I}} A_i$ de Ω en ensembles non-vides telle que*

$$\mathcal{F} = \left\{ \bigsqcup_{i \in I} A_i ; I \subset \mathcal{I} \right\}$$

EXERCICE 3. 1. Prouvez la proposition précédente.

2. Si X est une variable aléatoire à valeurs dans un espace dénombrable, prouver en l'exhibant qu'il existe une partition dénombrable de Ω engendrant $\sigma(X)$.

1.4.3 Définition générale de l'indépendance

DÉFINITION. Soit $(\mathcal{A}_i)_{i \in \mathcal{I}}$ une famille d'ensembles d'événements (la majeure partie du temps, chaque \mathcal{A}_i sera une σ -algèbre). On dit que les \mathcal{A}_i sont *indépendantes* si pour toute suite finie d'indices distincts $i_1, \dots, i_n \in \mathcal{I}$, pour tout n -uplet d'événements $(A_1, \dots, A_n) \in \prod_{k=1}^n \mathcal{A}_{i_k}$, on a

$$\mathbb{P}[A_1 \cap \dots \cap A_n] = \prod_{k=1}^n \mathbb{P}[A_k]$$

Remarques. Cette définition est abstraite. Tâchons d'entamer sa démystification et espérons que la pratique fera le reste du travail.

Tout d'abord, il ne s'agit de comprendre l'indépendance définie ci-dessus que dans le cas des σ -algèbres. La définition est donnée dans le cadre général car, pour des raisons techniques dues à la machinerie complexe de l'intégration, il sera parfois pratique de passer par d'autres classes, mais il ne s'agit pas d'étendre l'intuition jusqu'à ce cadre très général.

Dans le cas des tribus, les \mathcal{A}_i ne sont pas dépourvus de contenu intuitif; on a vu qu'il s'agissait d'y penser comme à des encodages d'informations¹². La formule du produit des probabilités ne sort pas non plus de nulle part;

12. on gardera à l'esprit, comme exemples, les tribus engendrées par des événements ou des variables aléatoires

c'est la généralisation directe de $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$. Pour vraiment comprendre la définition, c'est cette généralisation qu'il s'agit de saisir. En d'autres termes, il faut comprendre la définition qui suit.

DÉFINITION. On dit que des événements A_1, \dots, A_n sont *indépendants* si, pour toute suite d'indices distincts $1 \leq i_1, \dots, i_m \leq n$,

$$\mathbb{P}[A_{i_1} \cap \dots \cap A_{i_m}] = \prod_{k=1}^m \mathbb{P}[A_{i_k}]$$

Explication. On peut imaginer plusieurs façons de généraliser la définition intuitive d'indépendance de 2 à n événements; nombre d'entre elles ne porte pas de sens profond. La bonne définition intuitive est la suivante : pour tout indice i , la connaissance des occurrences et non-occurrences des événements A_j pour $j \neq i$ doit ne donner aucune information nouvelle sur la probabilité que A_i ait lieu. Ainsi, la définition mathématiquement naturelle de l'indépendance de A_1, \dots, A_n est la suivante.

Disons, jusqu'à la fin de cette explication, que si A est un événement, A^0 désignera A et A^1 son complémentaire. Les événements A_1, \dots, A_n sont indépendants au sens intuitif si, pour tout i , pour toute fonction $\epsilon : \{1, \dots, n\} \setminus \{i\} \rightarrow \{0, 1\}$, A_i est indépendant de $\bigcap_{j \neq i} A_j^{\epsilon(j)}$. Vérifier que cette définition est équivalente à celle donnée plus haut est un exercice.

On pourra également démontrer, à titre d'exercice, que les A_i sont indépendants si et seulement si, pour tous I et J parties disjointes de $\{1, \dots, n\}$, pour toute fonction $\epsilon : \{1, \dots, n\} \rightarrow \{0, 1\}$, les événements $\bigcap_{i \in I} A_i^{\epsilon(i)}$ et $\bigcap_{j \in J} A_j^{\epsilon(j)}$ sont indépendants.

Digression finale : il s'agit de noter que la définition intuitive d'indépendance de n événements qu'on vient d'introduire, prise pour $n = 2$, ne donne pas exactement la définition intuitive introduite en 1.4.1. Ceci n'est en aucun cas problématique dans la mesure où ces deux notions s'équivalent ; il suffit pour le voir de remarquer que

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B] \Rightarrow \mathbb{P}[(\Omega \setminus A) \cap B] = \mathbb{P}[\Omega \setminus A] \cdot \mathbb{P}[B]$$

□

DÉFINITION. Des variables aléatoires (resp. des événements) sont dits *indépendantes*¹³ si les tribus que ces variables aléatoires (resp. événements) engendrent sont indépendantes.

EXERCICE 4. Vérifier que les différents sens donnés à l'indépendance d'événements sont en fait les mêmes.

EXERCICE 5. Soit $\mathcal{G}_1, \dots, \mathcal{G}_n$ des tribus. Montrer qu'elles sont indépendantes si et seulement si pour tous $A_1 \in \mathcal{G}_1, \dots, A_n \in \mathcal{G}_n$,

$$\mathbb{P}[A_1 \cap \dots \cap A_n] = \prod_{i=1}^n \mathbb{P}[A_i]$$

13. resp. *indépendants* !

Comprendre ce qu'est l'indépendance passe aussi par comprendre ce qu'elle n'est pas. La notion suivante, par exemple, ne coïncide pas avec celle d'indépendance.

DÉFINITION. Des variables aléatoires sont *indépendantes deux à deux* si, lorsqu'on prend deux quelconques d'entre elles, elles sont indépendantes. On a bien sûr une définition analogue pour les événements.

Des variables aléatoires indépendantes sont en particulier indépendantes deux à deux. La réciproque est fausse, comme l'indique le contre-exemple qui suit.

Contre-exemple. Travaillons ici avec l'univers $\{-1, 1\}^3$ muni de la tribu pleine et de la probabilité uniforme. Pour $i = 1, 2, 3$, posons

$$X_i : \{-1, 1\}^3 \rightarrow \{-1, 1\}$$

l'application $i^{\text{ème}}$ coordonnée. Posons aussi

$$Y_i = \frac{\prod_{j=1}^3 X_j}{X_i}$$

On peut vérifier¹⁴ que les Y_i sont indépendantes deux à deux, mais pas indépendantes.

1.4.4 Attirail pratique et théorique

Vous avez été patients, vous êtes arrivés jusqu'ici ; faites l'effort de lire les paragraphes à venir et vous pourrez enfin vous attaquer à des exercices relevant plus des probabilités que de la théorie de la mesure. Évidemment, ceci requiert qu'on forge des résultats probabilistes utilisables, ce qui devra se faire dans le fer de la théorie de la mesure.

Ce paragraphe est inspiré de la partie 1.4. du livre *Probability: theory and examples* de Rick Durrett — on s'y référera pour les preuves. On traitera successivement les trois questions suivantes :

- Comment vérifie-t-on que des variables aléatoires sont indépendantes ?
- Si X et Y sont indépendantes, en quoi cela aide-t-il à prouver quoi que ce soit ?
- “Mais en fait, est-ce que ça existe des variables aléatoires indépendantes ?”

Quelques critères pratiques pour vérifier l'indépendance

Rappelons qu'on travaille dans un espace probabilisé noté $(\Omega, \mathcal{F}, \mathbb{P})$.

14. faites-le !

DÉFINITION. Un π -système est un ensemble d'événements stable par intersection finie. Autrement dit, il s'agit d'un ensemble $\mathcal{P} \subset \mathcal{F}$ tel que

$$\forall (A, B) \in \mathcal{P}^2, A \cap B \in \mathcal{P}$$

DÉFINITION. Un λ -système est un ensemble d'événements \mathcal{L} tel que

1. $\Omega \in \mathcal{L}$;
2. $\forall A, B \in \mathcal{L}, A \subset B \Rightarrow B \setminus A \in \mathcal{L}$;
3. si (A_n) est une suite croissante d'éléments de \mathcal{L} , alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{L}$.

Ces notions sont utiles en théorie de la mesure. On a notamment le lemme suivant.

LEMME 1.6. *Le λ -système engendré par un π -système est une σ -algèbre.*

Ce lemme permet de démontrer que si deux mesures de probabilité μ et ν coïncident sur un π -système, elles coïncident sur toute la tribu engendrée ; en effet, l'ensemble des événements ayant même mesure pour μ et ν est un λ -système !

Le lemme 1.6 est l'outil central de la preuve du théorème suivant (voir chez Durrett).

THÉORÈME 1.7. *Si des π -systèmes sont indépendants¹⁵, les σ -algèbres qu'ils engendrent le sont également.*

On peut en déduire les trois résultats utiles qui suivent.

COROLLAIRE 1.8. *Soit X_1, \dots, X_n des variables aléatoires réelles¹⁶. Si, pour tous $x_1, \dots, x_n \in \mathbb{R}$,*

$$\mathbb{P}[X_1 \leq x_1, \dots, X_n \leq x_n] = \prod_{i=1}^n \mathbb{P}[X_i \leq x_i]$$

alors les X_i sont indépendantes.

Remarque. Dans la formule apparaissant dans le corollaire précédent, la formule $\mathbb{P}[X_1 \leq x_1, \dots, X_n \leq x_n]$ se lit, à travers des lunettes d'un théoricien de la mesure, $\mathbb{P}[\{\omega \in \Omega : X_1(\omega) \leq x_1 \text{ et } \dots \text{ et } X_n(\omega) \leq x_n\}]$.

Dans les deux énoncés qui suivent, $m(i)$ est un élément de $\mathbb{N} \cup \{\infty\}$ dépendant d'un paramètre i . Quand $m(i) = \infty$, $\mathbb{C}^{m(i)}$ désignera $\mathbb{C}^{\mathbb{N}}$, qui sera muni de la tribu cylindrique ; il s'agit de la tribu engendrée par les

$$B_1 \times \dots \times B_n \times \prod_{k > n} \mathbb{C}$$

pour $n \in \mathbb{N}$, B_1, \dots, B_n boréliens de \mathbb{C} .

¹⁵. en tant qu'ensembles d'événements

¹⁶. comme d'habitude implicitement supposées définies sur l'espace probabilisé qu'on s'est fixé pour cette partie

COROLLAIRE 1.9. Soit $(\mathcal{F}_{i,j})_{i \in \mathcal{I}, 1 \leq j \leq m(i)}$ une famille de tribus indépendantes. Notons \mathcal{G}_i la tribu engendrée par les $\mathcal{F}_{i,j}$ pour $1 \leq j \leq m(i)$. Alors les \mathcal{G}_i sont indépendantes.

COROLLAIRE 1.10. Soit $X_{i,j}$ une famille de variables aléatoires complexes indépendantes, les indices i et j se comportant comme dans le corollaire précédent. Pour chaque i , soit f_i une application mesurable de $\mathbb{C}^{m(i)}$ vers \mathbb{C} . Alors les $f_i(X_{i,1}, \dots, X_{i,m(i)})$ sont indépendantes.

Remarques. Pour un théoricien de la mesure, $f_i(X_{i,1}, \dots, X_{i,m(i)})$ devrait s'écrire $\omega \mapsto f_i(X_{i,1}(\omega), \dots, X_{i,m(i)}(\omega))$. Pour $m(i) = 1$, c'est une composition.

Les deux derniers corollaires sont naturels. Si Paul regarde deux jets de dés et produit un résultat avec, il sera indépendant de celui que Marie-Gertrude produira à partir de trois lancers de pièces. Mais rappelons-nous que des variables indépendantes ne sont pas nécessairement des variables qui, par construction, n'ont rien à voir ; elles doivent juste ne pas apporter d'information l'une sur l'autre. Une fois ceci remarqué, on sera peut-être moins étonné du fait que les preuves de ces résultats soient relativement techniques. ⁽¹⁷⁾

Ces résultats seront très pratiques quand on manipulera concrètement des variables aléatoires indépendantes.

Loi et indépendance

NOTATION. Si X_1, \dots, X_n sont des variables aléatoires, on note (X_1, \dots, X_n) la variable aléatoire

$$\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$$

Rappelez-vous : quand un probabiliste pose une question sur une variable aléatoire X , la réponse à sa question ne dépend que de la loi de X . De même, quand il se pose des questions sur plusieurs variables aléatoires, X_1, \dots, X_n , la réponse ne dépendra que de la loi μ de (X_1, \dots, X_n) . Cette loi est appelée *loi jointe* des variables aléatoires X_i . On appelle *$i^{\text{ème}}$ (loi) marginale* de μ la mesure image de μ par l'application $i^{\text{ème}}$ coordonnée — en d'autres termes, la loi de X_i .

Nous allons voir ici, en particulier, que si on connaît la loi de chaque X_i et qu'on sait ces variables indépendantes, on connaît la loi du n -uplet aléatoire. (Sans hypothèse d'indépendance, la donnée des lois des X_i est loin de suffire à déterminer celle du vecteur aléatoire.)

EXERCICE 6. Se convaincre de la remarque entre parenthèses ci-dessus.

17. dans le sens où elles utilisent le théorème 1.7 donc le lemme 1.6

THÉORÈME 1.11. Soit X_1, \dots, X_n des variables aléatoires définies sur un même espace probabilisé. Ces variables aléatoires sont indépendantes et de lois respectives μ_1, \dots, μ_n ssi la loi de la variable aléatoire (X_1, \dots, X_n) est la loi produit $\mu_1 \otimes \dots \otimes \mu_n$.

EXERCICE 7. Réfléchir au théorème précédent dans le cas où les X_i prennent une quantité finie ou dénombrable de valeurs.

Du théorème de Fubini, on déduit la proposition suivante.

PROPOSITION 1.12. Soit X_1, \dots, X_n des variables aléatoires indépendantes. On fait l'une des hypothèses suivantes :

- les X_i sont à valeurs dans \mathbb{R}_+ ;
- les X_i sont des variables aléatoires complexes L^1 .

Alors les membres de l'égalité suivante ont un sens et cette égalité a lieu :

$$\mathbb{E} \left[\prod_{i=1}^n X_i \right] = \prod_{i=1}^n \mathbb{E}[X_i]$$

Remarque. Notant que si les X_i sont indépendants, les $\phi_i(X_i)$ le sont aussi, on augmente instantanément la portée de la proposition précédente.

Le problème de l'existence...

Soit μ_1 et μ_2 deux mesures de probabilités sur (X_1, \mathcal{F}_1) et (X_2, \mathcal{F}_2) . Existe-il deux variables aléatoires indépendantes de lois respectives μ_1 et μ_2 ? Oui, il suffit de prendre

$$(\Omega, \mathcal{F}, \mathbb{P}) = (X_1 \times X_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$$

et poser X_i la projection sur la $i^{\text{ème}}$ coordonnée (cf paragraphe précédent). Le même argument fonctionne pour un nombre fini de variables. Que se passe-t-il quand le nombre de variables devient infini ?

Soit $(\mu_n)_{n \in \mathbb{N}}$ une suite de mesures de probabilité définies sur des espaces polonais (penser $\mathbb{R}, \mathbb{C}, \mathbb{C}^d$) dépendant éventuellement de n . Existe-t-il un espace probabilisé et des variables aléatoires X_n définies sur cet espace probabilisé tels que les X_n soient indépendantes et chaque X_n ait pour loi μ_n ?

THÉORÈME 1.13. Oui.

En plus de ce résultat d'existence, on peut démontrer¹⁸ qu'une telle suite de X_n est unique en loi. Soyons plus précis. Pour fixer les idées, on supposera

18. en utilisant le lemme des π - λ -systèmes

les variables aléatoires complexes. Rappelons que $\mathbb{C}^{\mathbb{N}}$ est munie de la tribu cylindrique, qu'engendrent les

$$B_1 \times \cdots \times B_n \times \prod_{k>n} \mathbb{C}$$

pour $n \in \mathbb{N}$, B_1, \dots, B_n boréliens de \mathbb{C} .

La loi qui est uniquement déterminée est celle de la variable aléatoire

$$\omega \mapsto (n \mapsto X_n(\omega))$$

J'ai insisté dessus à maintes reprises : souvent, en probabilités, on estime un problème bien posé lorsque la loi est spécifiée. Si je considère des variables aléatoires et que je spécifie seulement la loi de chacune d'entre elles, le problème peut prendre de multiples formes ; si j'impose les variables aléatoires indépendantes, il y a existence et unicité de la loi (jointe) qui nous intéresse. N'oubliez pas de préciser l'hypothèse d'indépendance lorsque vous la faites. Dans le cas où les variables aléatoires, que vous introduisez, en sus d'être indépendantes, ont même loi, vous parlerez de *variables aléatoires indépendantes identiquement distribuées*¹⁹, ou *v.a.i.i.d.* .

Pour en revenir au théorème 1.13, il est difficile, et semblable au théorème de Carathéodory permettant de construire la mesure de Lebesgue. Si vous cherchez à en trouver la vraie preuve, renseignez-vous sur le théorème d'extension ou de prolongement de Kolmogorov, qui établit un résultat plus fort encore.

Si on admet construite la mesure de Lebesgue, on peut contourner le problème. Dans le cadre de la préparation à l'agrégation, c'est éthique et responsable puisque ladite construction est admise ; mathématiquement, c'est un peu de la triche, vu qu'on admet de façon cachée le théorème de Carathéodory, moralement équivalent à celui qu'on veut établir. On verra comment faire ce tour de passe-passe au paragraphe 1.5.2.

1.5 Les premières fonctions associées à une loi

1.5.1 La fonction génératrice

Soit μ une mesure de probabilité sur \mathbb{N} . La *série génératrice* de μ est

$$\sum_{n \geq 0} \mu(\{n\}) z^n$$

Sur son domaine de convergence, qui contient le disque unité, cette série définit la *fonction génératrice* de μ . À l'intérieur de ce domaine, cette fonction coïncide avec

$$z \mapsto \mathbb{E} \left[z^X \right]$$

19. en anglais, la loi s'appelle la distribution

où X désigne une variable aléatoire de loi μ .

(On rappelle la convention “ $0^0 = 1$ ” qui garantit la continuité de z^0 comme fonction de z .)

Deux mesures de probabilité sur \mathbb{N} ont mêmes fonctions génératrices ssi elles sont égales. Autrement dit, les lois sont caractérisées par leur fonction génératrice.

La proposition suivante découle sans peine de ce qu’on a vu sur l’indépendance.

PROPOSITION 1.14. Si X et Y sont deux variables aléatoires indépendantes à valeurs dans \mathbb{N} , alors la série génératrice de $X + Y$ est le produit des séries génératrices de X et de Y .

On peut utiliser la proposition précédente pour déterminer la loi de la somme de deux variables de Poisson indépendantes — ceci est fait dans la partie consacrée aux exercices. Ce genre de considération s’avérera aussi très utile dans l’étude du processus de Galton-Watson.

EXERCICE 8. Soit μ une mesure de probabilité sur \mathbb{N} et f la fonction génératrice de μ restreinte à l’intervalle $] -1, 1[$. Soit X une variable aléatoire de loi μ . Soit enfin $f^{(k)}$ la dérivée $k^{\text{ème}}$ de f . Démontrer que

$$f^{(k)}(1 - \epsilon) \xrightarrow{n \rightarrow \infty} \mathbb{E}[X(X-1) \dots (X-k+1)] \in [0, +\infty]$$

Ceci montre que les *moments* $\mathbb{E}[X^k]$ de la loi μ se calculent assez facilement à partir de sa fonction génératrice.

1.5.2 La fonction de répartition

DÉFINITION. Soit μ une loi sur \mathbb{R} . Sa *fonction de répartition* est

$$F_\mu : x \mapsto \mu([-\infty, x])$$

Remarque. Si X est une variable aléatoire de loi μ , alors $F_\mu : x \mapsto \mathbb{P}[X \leq x]$.

PROPOSITION 1.15. Deux lois avec la même fonction de répartition sont égales.

Preuve. Les $] -\infty, x]$ forment un π -système engendrant la tribu borélienne et sur lequel les deux mesures considérées coïncident ; on conclut avec la remarque suivant l’énoncé du lemme 1.6. \square

PROPOSITION 1.16. Toute fonction de répartition est croissante, continue à droite, tend vers 0 en $-\infty$ et vers 1 en $+\infty$.

EXERCICE 9. Le vérifier.

PROPOSITION 1.17. Soit F une fonction de \mathbb{R} vers $[0, 1]$ qui est croissante, continue à droite, qui tend vers 0 en $-\infty$ et vers 1 en $+\infty$. Alors F est la fonction de répartition d'une mesure de probabilité sur \mathbb{R} .

Esquisse de preuve. Soit U une variable uniforme dans $[0, 1]$. On définit une fonction inverse généralisée de F :

$$F^{-1} : x \mapsto \inf\{y : F(y) \geq x\}$$

J'affirme que la loi μ de la variable aléatoire $X := F^{-1}(U)$ a pour fonction de répartition F . En effet,

$$F_{\mu}(x) = \mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x)$$

Où a-t-on utilisé les hypothèses me direz-vous ? La croissance et la continuité à droite servent à montrer que

$$F^{-1}(U) \leq x \Leftrightarrow U \leq F(x)$$

tandis que les hypothèses asymptotiques garantissent que X prend des valeurs infinies avec probabilité nulle. \square

Pour résumer, on sait que $\mu \mapsto F_{\mu}$ est injective, et on connaît son image. Mais il y a d'autres choses à dire. Le point de départ des observations à venir est la formule $F^{-1}(U)$; deux choses en elle peuvent nous frapper :

- l'inverse n'était pas pris au sens usuel,
- la fonction de répartition inverse généralisée permet de simuler n'importe quelle variable aléatoire réelle à partir d'une variable aléatoire uniforme sur $[0, 1]$.

Chacune de ces remarques mérite un approfondissement ; un paragraphe est consacré à chacune d'elle.

Existence de variables aléatoires indépendantes de lois prescrites

Soit μ_i une suite de lois (mesures de probabilité) sur \mathbb{R} . Nous allons montrer qu'il existe un espace probabilisé et une suite de variables aléatoires indépendantes définies sur cet univers, de lois respectives μ_i . On suppose connue l'existence de la mesure de Lebesgue sur $[0, 1]$.

Soit donc U une variable aléatoire uniforme sur $[0, 1]$. Posant X_k ses chiffres en base 2, on a une suite de variables aléatoires indépendantes identiquement distribuées de loi de Bernoulli de paramètre 1/2. Puisque \mathbb{N} est en bijection avec $\mathbb{N} \times \mathbb{N}$, quitte à jouer avec les indices des X_k , on a une suite de suites de v.a.i.i.d. Bernoulli(1/2). On la note $Y_{i,j}$. En posant $U_i := \sum_j 2^{-j} Y_{i,j}$, on obtient une suite de v.a.i.i.d. de loi la mesure de Lebesgue sur $[0, 1]$. La suite $Z_i := F_{\mu_i}^{-1}(U_i)$ convient.

EXERCICE 10. Vérifier les détails.

Remarque. Notez la portée algorithmique des lignes précédentes : à partir de bits équirépartis indépendants, on peut former une variable aléatoire réelle de n'importe quelle loi prescrite ! (Par ailleurs, ces lignes rassèrent ceux qui croyaient que certaines mesures de probabilité sur \mathbb{R} auraient pu n'être loi d'aucune variable aléatoire concrète.)

Discontinuités et plateaux de la fonction de répartition

Le défaut de bijectivité de $F_\mu : \mathbb{R} \rightarrow [0, 1]$ peut prendre deux formes : il y a les discontinuités et les plateaux, les images réciproques vides et les trop pleines. Quelles en sont les interprétations ?

PLATEAUX : Si F_μ est constante sur $]a, b[$, alors $\mu(]a, b[) = 0$. Si X a loi μ , cela s'écrit $\mathbb{P}[X \in]a, b[] = 0$.

DISCONTINUITÉS : Posons $F_\mu^-(x) := \lim_{\epsilon \rightarrow 0+} F_\mu(x - \epsilon)$. La fonction F_μ est continue en x ssi $F_\mu^-(x) = F_\mu(x)$. Il s'agit donc d'interpréter probabilistement $F_\mu(x) - F_\mu^-(x)$, qui n'est rien d'autre que $\mathbb{P}[X = x]$. Un point de discontinuité de F_μ sera appelé un *atome* de μ . Une loi a au plus une quantité dénombrable d'atomes.

Pour conclure ce paragraphe sur la fonction de répartition, considérons le cas où $\mu = f(x) dx$. On dira que la loi est à *densité* et f sera appelée "la" *densité* de μ .⁽²⁰⁾ On a la proposition suivante, dont la vérification est laissée en exercice.

PROPOSITION 1.18. *Si μ admet une densité f continue, alors F_μ est une primitive de f . Réciproquement, si F_μ est C^1 , alors μ admet pour densité la dérivée de F_μ .*

1.6 Exercices, deuxième fournée

Rappelons-le : ce n'est pas parce que des résultats sont proposés en exercice qu'il ne faut pas les connaître. En l'occurrence, à l'exception des trois derniers, tous les exercices de cette fournée doivent être faits et les résultats qui y sont démontrés retenus.

EXERCICE FIL ROUGE. Dressez, sur une feuille à part que vous conserverez, un tableau à double entrée qui contiendra toutes les informations que vous aurez accumulées sur les lois usuelles (définition, espérance, variance, fonction génératrice, propriétés particulières...).

Calculez les espérances et variances de toutes les lois usuelles.

20. La densité est uniquement définie à égalité presque partout près.

EXERCICE 11. Rappeler la définition d'une variable aléatoire de Bernoulli. En exprimer l'espérance et la variance en fonction du paramètre.

DÉFINITION. Soit $p \in [0, 1]$. La loi **géométrique** de paramètre p est la probabilité sur \mathbb{N} donnant masse $p^k \times (1 - p)$ au point k .

EXERCICE 12. Soit X_n une suite infinie de variables aléatoires indépendantes de loi de Bernoulli de paramètre $p \in [0, 1]$.

1. Vérifier que l'application associant à $\omega \in \Omega$ le plus petit indice n tel que $X_n(\omega) = 0$ est une variable aléatoire (à valeurs dans $\mathbb{N} \cup \{\infty\}$).

2. Prouver que sa loi est une géométrique de paramètre p .

DÉFINITION. La loi **binomiale** de paramètres $n \in \mathbb{N}$ et $p \in [0, 1]$ est la mesure sur $\{0, \dots, n\}$ donnant masse $\binom{n}{k} p^k (1 - p)^{n-k}$ au point k .

EXERCICE 13. Soit X_1, \dots, X_n des variables aléatoires L^2 définies sur un même espace probabilisé.

1. On sait que $\mathbb{E}[\sum_i X_i] = \sum_i \mathbb{E}[X_i]$. Montrer qu'on n'a pas systématiquement une telle égalité pour la variance.

2. Prouver que, si les X_i sont indépendantes, on a

$$\sum_i \text{Var}(X_i) = \text{Var}\left(\sum_i X_i\right)$$

3. Montrer que si X_1, \dots, X_n sont des variables aléatoires indépendantes de Bernoulli de paramètre p , alors leur somme suit une loi binomiale de paramètres n et p .

4. Utiliser les questions précédentes pour calculer sans peine l'espérance et la variance d'une variable aléatoire de loi binomiale de paramètres n et p .

5. Soit X et Y deux variables aléatoires indépendantes, de loi respective $\text{Binom}(n, p)$ et $\text{Binom}(m, p)$. Calculer, en utilisant une question précédente, la loi de $X + Y$.

6. Retrouver ce dernier résultat par les fonctions génératrices.

DÉFINITION. Une variable aléatoire réelle est dite **exponentielle** de paramètre $\lambda \in \mathbb{R}_+^*$ si sa loi est $\lambda e^{-\lambda x} 1_{x>0} dx$.

EXERCICE 14. Soit X une variable aléatoire exponentielle de paramètre 1.

1. Calculer l'espérance et la variance de X .

2. Soit $\alpha > 0$. Calculer la loi de αX . En déduire l'espérance et la variance d'une exponentielle de paramètre quelconque.

Dans ce qui suit, Y est une variable aléatoire exponentielle de paramètre $\lambda > 0$.

3. Calculer la fonction de répartition de Y . Utiliser cela pour redémontrer que αX est une variable aléatoire de loi exponentielle dès que $\alpha > 0$.

4. Montrer que $\mathbb{P}[Y > s + t | Y > s] = \mathbb{P}[Y > t]$. On dit que Y a la propriété de *perte de mémoire*. Montrer, réciproquement, que si une variable aléatoire à valeurs dans \mathbb{R}_+^* a la propriété de perte de mémoire, alors c'est une variable aléatoire exponentielle

5. Soit Z une variable aléatoire exponentielle de paramètre $\mu > 0$. On la suppose définie sur le même espace probabilisé que Y . On suppose de plus ces deux variables aléatoires indépendantes. Calculer la loi de la variable aléatoire $\min(Y, Z)$.

DÉFINITION. On appelle loi de **Poisson** de paramètre $\lambda > 0$ la probabilité sur \mathbb{N} donnant masse $e^{-\lambda} \cdot \frac{\lambda^k}{k!}$ à k .

EXERCICE 15. Soit X une variable aléatoire de Poisson de paramètre λ .

1. Calculer l'espérance et la variance de X .
2. Calculer sa fonction génératrice. Retrouver de la sorte les résultats de la première question.
3. Soit Y une variable aléatoire de Poisson de paramètre μ , indépendante de X . Calculer la loi de $X + Y$. (On pourra s'aider de la question précédente.)

EXERCICE 16. (convolée et somme indépendante)

1. Se rappeler la définition de la convolée de deux mesures sur \mathbb{R} .
2. Vérifier que si X et Y sont indépendantes, de lois respectives μ et ν , $X + Y$ a pour loi la convolée $\mu * \nu$.
3. Si vous avez peur, écrivez la formule dans le cas discret où les lois de X et Y sont des combinaisons linéaires finies de masses de Dirac.

EXERCICE 17. (calcul de l'intégrale de Gauss et lois normales)

1. En effectuant un changement de variables polaire, calculer

$$\int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} d\lambda_2$$

(On désigne par λ_2 la mesure de Lebesgue sur \mathbb{R}^2 .)

2. En déduire la valeur de $\int_{\mathbb{R}} e^{-x^2/2} dx$.
3. En déduire que, si m et σ désignent des réels, la mesure $\frac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} dx$ sur \mathbb{R} est de probabilité. C'est la loi **normale** ou **gaussienne** de moyenne m et de variance σ^2 . Vérifier qu'elle porte bien son nom, à savoir que sa moyenne vaut m et sa variance σ^2 .
4. Vérifier que la somme de deux variables aléatoires gaussiennes indépendantes est aussi gaussienne (de paramètres que vous pouvez deviner à l'avance).
5. Soit $\alpha, \beta \in \mathbb{R}$. Soit X une variable aléatoire **gaussienne centrée réduite**, i.e. de paramètres $m = 0$ et $\sigma^2 = 1$. Calculer la loi de $\alpha X + \beta$. Faire le rapprochement avec la question 3.

EXERCICE 18. (formule de la baignoire²¹) Soit X une variable aléatoire à valeurs positives. Montrer que $\mathbb{E}[X] = \int_{\mathbb{R}_+} \mathbb{P}[X \geq x] dx$, et que, dans le cas où les valeurs de X sont entières, on peut écrire $\mathbb{E}[X] = \sum_{n \geq 1} \mathbb{P}[X \geq n]$.

EXERCICE 19. (formule de Bayes) Soit $\Omega = \bigsqcup_{i=1}^N A_i$ une partition finie de l'univers en événements de mesure strictement positive. Soit B un événement lui aussi de mesure strictement positive. Exprimer les $\mathbb{P}[A_i|B]$ en fonction des $\mathbb{P}[B|A_j]$ et $\mathbb{P}[A_j]$.

EXERCICE 20. Montrer que toute variable aléatoire réelle à densité est sans atome, mais que la réciproque est fausse.

EXERCICE 21. Soit $(X_i)_{1 \leq i \leq \infty}$ une suite indépendante de variables aléatoires uniformément tirées dans $\mathbb{Z}/2\mathbb{Z}$.

1. Montrer que X_1 et $X_2 + X_3$ sont indépendantes.
2. Montrer que X_1 et $X_1 + X_2$ sont indépendantes.
3. Notons $T(\omega)$ l'infimum des $n \in \mathbb{N}$ tel que $X_n(\omega) = 0$. Montrer que

$$X_T : \omega \mapsto X_{T(\omega)+1}(\omega)$$

est une variable aléatoire, et qu'elle est indépendante de T .

EXERCICE 22. (ruine du joueur, pas évident) Soit X_i une suite de v.a.i.i.d. de loi $\frac{\delta_{-1} + \delta_1}{2}$. Soit $(M, N) \in \mathbb{N}^2$. Posons $S_n := M + \sum_{i=1}^n X_i$. Notons T le premier instant où $S_n \notin \{1, \dots, M + N - 1\}$. Montrer que T est presque sûrement fini et calculer $\mathbb{P}[S_T = 0]$. (Ce problème revient à calculer la probabilité de ruine d'un joueur au jeu qui suit. Le joueur a initialement M jetons et la banque N . À chaque étape, si le jeu n'est pas terminé, on tire à pile ou face pour déterminer qui prendra un jeton à l'autre. Le jeu s'arrête quand le joueur ou la banque a tous les jetons ; l'autre est ruiné.)

EXERCICE 23. (plus difficile) Soit X_1, \dots, X_n des variables aléatoires réelles indépendantes identiquement distribuées, de loi sans atome. On pose

$$A_k := \{\omega : \forall i < k, X_k(\omega) > X_i(\omega)\}$$

On pourra penser les données du problème de la sorte : un athlète dont la valeur intrinsèque est constante fait n sauts en longueurs ; les X_i correspondent à leur longueur. L'événement A_k s'interprète alors comme "l'athlète, au $k^{\text{ème}}$ essai, a battu son record".

Montrer que les A_k sont indépendants !

La correction est dans le livre de Durrett ; il s'agit de l'exemple 6.2, page 51, dans la troisième édition.

21. appellation non-standard

2 Intermède d'application des séries génératrices

Voici deux problèmes que nous sommes dès maintenant en mesure de formaliser et traiter. Le premier parle de survie de populations et le second de dés.

2.1 Le processus dit de Galton-Watson

On trouvera les grandes lignes de ce qui suit dans le livre *Promenade aléatoire* de Benaïm et El Karoui.

Le processus de Galton et Watson fut introduit afin d'étudier l'extinction des noms des nobles en Grande-Bretagne (²²). La transmission par le père du nom de famille justifie le fait qu'on ne considère qu'un parent, comme si les individus se divisaient et mouraient. Pour la même raison, Galton et Watson ne comptabilisaient pas la progéniture féminine dans le nombre d'enfants. Le modèle peut s'appliquer indifféremment aux noms de famille et aux cellules.

Soit μ une loi de probabilité sur \mathbb{N} . On appellera μ la *loi de reproduction*. On pose $(\xi_i^{(n)})$ une suite de v.a.i.i.d. de loi μ . On pensera à $\xi_i^{(n)}$ comme au nombre d'enfants du $i^{\text{ème}}$ individu de la $n^{\text{ème}}$ génération, s'il existe.

On pose :

$$— Z_0 = 1 ;$$

$$— Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i^{(n)} = \sum_{i \geq 1} 1_{i \leq Z_n} \xi_i^{(n)}.$$

La variable aléatoire Z_n représente le nombre d'individus que comporte la $n^{\text{ème}}$ génération. Posons $m := \mathbb{E}[Z_1]$.

THÉORÈME 2.1. *Si $m < 1$, alors l'extinction est presque sûre : cela signifie que, presque sûrement, Z_n stationne en 0.*

Si $m > 1$, on a survie avec probabilité positive : $\mathbb{P}[\liminf Z_n \geq 1] > 0$.

Si $m = 1$, à moins que $\mu = \delta_1$, l'extinction est presque sûre.

Démontrons ce théorème. Posons

$$g \left| \begin{array}{ll} [0, 1] & \longrightarrow [0, 1] \\ x & \longmapsto \mathbb{E}[x^{Z_1}] \end{array} \right.$$

C'est la fonction génératrice de Z_1 . De même, on pose g_n la fonction génératrice de Z_n . On rappelle qu'on a adopté la convention $0^0 = 1$: on veut que $x \mapsto x^0$ soit une brave fonction continue.

22. Plus précisément, Galton et Watson l'ont introduit pour cela. Mais Bienaymé l'avait déjà introduit longtemps auparavant. Sachant qu'en plus, Galton et Watson étaient arrivés à un théorème faux, on peut se demander pourquoi c'est leurs noms qu'on accole à ce processus.

FAIT 2.2. Pour $n \geq 0$, $g_n = \underbrace{g \circ \cdots \circ g}_{n \text{ fois}}$.

Preuve. On va prouver cela par récurrence. L'initialisation ne présente pas de difficulté. Supposons donc le fait connu au rang n et établissons-le au rang suivant.

$$\begin{aligned}
g_{n+1}(x) &= \mathbb{E} \left[x^{Z_{n+1}} \right] \\
&= \mathbb{E} \left[x^{\xi_1^{(n)} + \cdots + \xi_{Z_n}^{(n)}} \right] \\
&= \mathbb{E} \left[\sum_k 1_{Z_n=k} \cdot x^{\xi_1^{(n)} + \cdots + \xi_k^{(n)}} \right] \\
&= \sum_k \mathbb{E} \left[1_{Z_n=k} \cdot x^{\xi_1^{(n)} + \cdots + \xi_k^{(n)}} \right] \\
&\stackrel{\text{indép}}{=} \sum_k \mathbb{E} [1_{Z_n=k}] \prod_{i=1}^k \mathbb{E} \left[x^{\xi_i^{(n)}} \right] \\
&\stackrel{\text{m loi}}{=} \sum_k \mathbb{P}[Z_n = k] g(x)^k \\
&= g_n(g(x))
\end{aligned}$$

□

Pour raisonner dans le monde de l'analyse, on aura besoin de trois remarques :

1. Dérivant (correctement) sous l'espérance, on voit que $g'(1) = m$ (le théorème de la limite de la dérivée fonctionne même pour les limites infinies). Notons qu'il s'agit d'une dérivée à gauche.
2. Par ailleurs, g est croissante et convexe, comme combinaison linéaire (infinie) positive de $x \mapsto x^n$. En effet, n'oublions pas que

$$g(x) = \sum_n \mu(\{n\}) x^n.$$

3. Enfin, $\mathbb{P}[Z_n = 0] = g_n(0) = \underbrace{g \circ \cdots \circ g}_{n \text{ fois}}(0)$. Il résulte de cela, de

$$Z_n = 0 \Rightarrow Z_{n+1} = 0$$

et du principe de convergence monotone²³ que la probabilité de l'événement d'extinction est $\lim_{n \rightarrow \infty} g_n(0)$. Ce nombre est le plus petit point fixe de g .

23. Je pense ici au fait que la probabilité d'une union croissante d'événements est la limite de leurs probabilités. L'appellation "principe de convergence monotone" n'est pas standard.

On s'est ramené à étudier la suite définie en itérant l'application croissante g à partir de 0. Je vous invite fortement à suivre les prochaines lignes en dessinant le graphe de g .

Quand $m < 1$, par convexité et comme $g(1) = 1$, on a l'inégalité suivante

$$\forall x \in [0, 1], g(x) \geq 1 + m(x - 1)$$

Ainsi, 1 est l'unique point fixe de g et l'extinction est presque sûre.

Quand $m > 1$, le graphe de g est sous la diagonale au voisinage de 1. Par ailleurs, $g(0) \geq 0$. Par continuité, g a un point fixe différent de 1, qui est en fait unique par convexité (on n'en a pas vraiment besoin). Ce point fixe est la probabilité d'extinction, qui est donc strictement moins que 1.

Quand $m = 1$ et $\mu \neq \delta_1$, il existe un $n \geq 2$ qui est chargé par μ . Ceci garantit la stricte convexité de g , et ainsi l'inexistence d'un point fixe de g différent de 1.

Le théorème de Bienaymé-Galton-Watson est établi.

Commentaire. Je pense qu'à l'agrégation, il faut opter pour la présentation que vous venez de lire. Cependant, cette modélisation ne semble pas être la plus honnête de l'univers. En effet, on parle du $i^{\text{ème}}$ individu, sans dire comment on les numérote. Le fait que la numérotation est indifférente est un corollaire du fait que tout est indépendant et identiquement distribué, mais ce n'est pas très rigoureux de mélanger tout cela dans l'étape de modélisation. J'estime la construction suivante plus satisfaisante.

On identifie un individu (existant ou non) à sa lignée ascendante. Ainsi, l'individu souche est étiqueté par le mot vide. Ses fils sont (1), (2), (3)... Les fils de (4) sont (4, 1), (4, 2)... Plus généralement, (n_1, \dots, n_l) a pour fils virtuels tous les (n_1, \dots, n_l, k) .

Formellement, on appelle *individu* un mot fini sur l'alphabet \mathbb{N} . Un *ascendant* d'un individu est un préfixe de ce dernier. Enfin, on appellera *numéro* d'un individu sa dernière lettre.

On tire une variable aléatoire de loi μ par individu, et on les prend toutes indépendantes. Un individu est dit *en vie* si son numéro est inférieur à la variable de son père et si c'est le cas de tous ses aïeux. (On dira que l'individu souche est en vie par définition.)

À partir de ce modèle et d'un choix raisonnable de procédure pour énumérer la $n^{\text{ème}}$ génération, on pourrait définir les $\xi_i^{(n)}$ de la présentation précédente et vérifier que ces variables sont i.i.d. de loi μ . Une autre possibilité, plus intéressante et moins fastidieuse, serait de ré-écrire directement le théorème et sa démonstration dans ce formalisme plus propre (mais un peu plus lourd).

2.2 Un problème de dés

Appelons *dé* une mesure de la forme $\frac{1}{6} \sum_{i=1}^6 \delta_{n(i)}$, où les $n(i)$ sont des entiers strictement positifs. On appellera aussi *dé* une variable aléatoire dont la loi est un *dé*.

Évidemment, le *dé* classique correspond à $n(i) = i$. Prenons D_1 et D_2 deux tels dés indépendants. La variable aléatoire $D_1 + D_2$, pertinente dans beaucoup de jeux, suit une certaine loi, à laquelle on peut accéder en posant une table d'addition.

Y a-t-il d'autres dés, D'_1 et D'_2 , tels que D'_1 et D'_2 soient indépendants et $D'_1 + D'_2$ ait même loi que $D_1 + D_2$?

THÉORÈME 2.3. *Il existe exactement trois tels couples de dés : celui usuel, $(1, 2, 2, 3, 3, 4)/(1, 3, 4, 5, 6, 8)$ et son symétrique $(1, 3, 4, 5, 6, 8)/(1, 2, 2, 3, 3, 4)$.*

Prouvons-cela sans sortir quoi que ce soit du chapeau. Ou plutôt, au lieu d'en sortir ces chiffres, sortons-en des concepts !

Procédons par analyse/synthèse. Soit (D'_1, D'_2) un couple tel que recherché. Notons f_1 et f_2 les fonctions génératrices de D'_1 et D'_2 . Comme les D'_i sont des dés, les $P_i := 6f_i$ sont des polynômes à coefficients entiers positifs ! De plus, $P_1(1) = P_2(1) = 6$, puisqu'on considérerait des mesures de probabilités. La propriété recherchée sur nos dés s'exprime précisément comme suit

$$P_1(X)P_2(X) = (X + X^2 + X^3 + X^4 + X^5 + X^6)^2$$

On peut ré-écrire cela sous la forme

$$\begin{aligned} P_1(X)P_2(X) &= X^2(1 + X + X^2 + X^3 + X^4 + X^5)^2 \\ &= X^2(\Phi_2(X)\Phi_3(X)\Phi_6(X))^2 \end{aligned}$$

où $\Phi_i(X)$ désigne le $i^{\text{ème}}$ polynôme cyclotomique.

Par factorialité de $\mathbb{Z}[X]$, il doit exister une façon de répartir les facteurs de droite pour obtenir, d'une part, P_1 et, d'autre part, P_2 . (La positivité de $P_i(1)$ indique qu'on n'aura pas à utiliser d'inversible de \mathbb{Z} .) Chaque P_i doit recevoir un X , car les dés sont supposés renvoyer des entiers naturels *non-nuls*. Par ailleurs, ayant,

$$\Phi_2(X) = X + 1 \quad \Phi_3(X) = X^2 + X + 1 \quad \Phi_6(X) = X^2 - X + 1$$

on a $\Phi_2(1) = 2$, $\Phi_3(1) = 3$ et $\Phi_6(1) = 1$. Puisque les $P_i(1)$ valent 6, les $\Phi_2(X)$ et $\Phi_3(X)$ doivent être équitablement répartis entre P_1 et P_2 . Il y a donc au plus deux possibilités non-triviales : celle correspondant à

$$P_1(X) = \Phi_2(X)\Phi_3(X) \quad P_2(X) = \Phi_2(X)\Phi_3(X)\Phi_6(X)^2$$

et la symétrie. Pour vérifier que ces polynômes correspondent bien à des dés, et les connaître, il faut faire les calculs²⁴. On trouve

$$P_1(X) = X + 2X^2 + 2X^3 + X^4$$

$$P_2(X) = X + X^3 + X^4 + X^5 + X^6 + X^8$$

24. La seule obstruction possible au fait que nos polynômes correspondent à des dés qui conviennent serait qu'un $P_i(X)$ ait au moins un coefficient négatif. Il correspondrait alors à une mesure *signée* de masse +1.

3 Convergences en théorie des probabilités

Les notions de convergence sont essentielles en analyse : à l'instar des inégalités, elles permettent de passer du calcul exact au qualitatif. Dans ce chapitre, on étudiera les différentes notions de convergences qui interviennent en théorie des probabilités ainsi que deux théorèmes majeurs de convergence — la loi des grands nombres et le théorème central limite. Mais avant ça, étoffons notre boîte à outils.

3.1 Des outils toujours utiles

3.1.1 Quelques inégalités

N'ayant pas envie d'énoncer un résultat juste après deux titres de parties, je glisse subtilement en ces lignes la présente phrase vide de contenu, afin qu'elle occupe l'espace.

INÉGALITÉ (MARKOV). Si $a > 0$ et X est une variable aléatoire positive, alors

$$\mathbb{P}[X \geq a] \leq a^{-1} \mathbb{E}[X]$$

Preuve. On note que $a \cdot 1_{X \geq a} \leq X$ et passe cette inégalité à l'espérance. \square

INÉGALITÉS (CHEBYSHEV, OU BIENAYMÉ-CHEBYSHEV). Si X est une variable aléatoire réelle de carré intégrable, i.e. si X est dans L^2 , et si de plus $a \in \mathbb{R}_+^$, alors les deux inégalités suivantes sont vérifiées :*

1. $\mathbb{P}[|X| \geq a] \leq a^{-2} \mathbb{E}[X^2]$,
2. $\mathbb{P}[|X - \mathbb{E}[X]| \geq a] \leq a^{-2} \text{Var}(X)$.

Preuve. La première inégalité s'établit comme suit

$$\mathbb{P}[|X| \geq a] = \mathbb{P}[X^2 \geq a^2] \stackrel{\text{Markov}}{\leq} a^{-2} \mathbb{E}[X^2]$$

tandis que la seconde n'est que l'application de la première à $X - \mathbb{E}[X]$. \square

Remarque. Les inégalités de Markov et de Chebyshev permettent, vous l'aurez deviné, de transformer un contrôle sur l'espérance en un contrôle sur des probabilités.

INÉGALITÉ (JENSEN). Soit φ une fonction convexe de $I \subset \mathbb{R}$ vers \mathbb{R} . Soit X une variable aléatoire L^1 à valeurs dans I . Alors,

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$

Preuve. Notons $m := \mathbb{E}[X]$. Le réel m appartient à l'intervalle I . La fonction φ étant convexe, il existe f une fonction affine partout inférieure à φ et telle que $f(m) = \varphi(m)$. Écrivons

$$f : x \mapsto \varphi(m) + \alpha(x - m)$$

On a

$$\mathbb{E}[f(X)] = \varphi(m) + \alpha \cdot \mathbb{E}[X - m] = \varphi(m) + 0 = \varphi(m)$$

On conclut en passant à l'espérance l'inégalité $f(X) \leq \varphi(X)$. \square

Remarques. On retrouve l'inégalité de convexité usuelle en prenant B une Bernoulli de paramètre λ et posant $X := y + (x - y)B$. Ceci permet, en particulier, de se rappeler quel membre est plus petit que l'autre dans l'inégalité de Jensen.

On retrouve bien entendu aussi l'inégalité de convexité à n points ; je trouve cette preuve plus élégante qu'une récurrence facile.

Où sont les applications, dans tout ça ? Pour Markov/Chebyshev, soyez patients, ça arrive dans quelques pages. L'inégalité de Jensen, quant à elle, loin d'être inutile, est tout de même bien plus dispensable que les deux autres ; voyons-la simplement comme la généralisation de l'inégalité de convexité. La sous-section, en toute rigueur, devrait s'intituler *Des outils toujours utiles, plus Jensen*.

3.1.2 Lemmes de Borel-Cantelli

Le lemme suivant est très important. Il mange une hypothèse de nature calculatoire (sommabilité de probabilités) et renvoie l'occurrence presque sûre d'un certain événement qui a du sens ; il fait passer du quantitatif au qualitatif. On l'utilisera souvent par la suite.

LEMME 3.4 (BOREL-CANTELLI). *Soit (A_n) une suite d'événements définis sur un même espace de probabilité. Si $\sum_n \mathbb{P}[A_n] < +\infty$, alors, \mathbb{P} -presque sûrement, seul un nombre fini des A_n a lieu.*

Autrement dit, si les $\mathbb{P}[A_n]$ sont sommables, l'événement $\limsup A_n$ défini par $1_{\limsup A_n} = \limsup 1_{A_n}$ a probabilité nulle.

Preuve. Soit $N \in \mathbb{N}$. Il résulte de $\limsup A_n \subset \bigcup_{n \geq N} A_n$ que

$$\mathbb{P}[\limsup A_n] \leq \sum_{n \geq N} \mathbb{P}[A_n]$$

Le membre de droite de cette inégalité tendant vers 0 quand N tend vers l'infini, le lemme est établi. \square

Contre-exemple. L'exemple suivant montre que la réciproque du lemme de Borel-Cantelli est incorrecte (sans hypothèse additionnelle). Considérons

$$(\Omega, \mathcal{F}, \mathbb{P}) := ([0, 1], \mathbf{Bor}([0, 1]), dx)$$

et posons $A_n := \{\omega \leq \frac{1}{n}\}$. Les probabilités des A_n ne sont pas sommables, pourtant $\limsup A_n$ est vide.

Une autre façon, plus probabiliste, de rédiger le même argument, est la suivante : considérons U une variable aléatoire uniforme et désignons par A_n l'événement $U \leq \frac{1}{n}$; les probabilités des A_n ne sont pas sommables, pourtant presque sûrement, $U \neq 0$, i.e. seul un nombre fini des A_n a lieu.

Ces deux arguments sont les mêmes, parce que seule la loi importe et que toute probabilité est loi d'une variable aléatoire²⁵.

LEMME 3.5 (BOREL-CANTELLI, VERSION INDÉPENDANTE). *Soit (A_n) une suite d'événements indépendants.*

Si $\sum_{n \geq 0} \mathbb{P}[A_n] = +\infty$, alors $\mathbb{P}[\limsup A_n] = 1$.

Preuve. Supposons que $\mathbb{P}[\limsup A_n] < 1$ et montrons que la série des $\mathbb{P}[A_n]$ est sommable. Dans les calculs qui suivent, A_n^c désigne le complémentaire de A_n .

$$\begin{aligned} \underbrace{1 - \mathbb{P}[\limsup A_n]}_{>0} &\stackrel{\text{déf.}}{=} \mathbb{P} \left[\bigcup_N \bigcap_{n \geq N} A_n^c \right] \\ &\stackrel{\text{mon.}}{=} \lim_{N \rightarrow \infty} \mathbb{P} \left[\bigcap_{n \geq N} A_n^c \right] \\ &\stackrel{\text{mon.}}{=} \lim_{N \rightarrow \infty} \lim_{K \rightarrow \infty} \mathbb{P} \left[\bigcap_{n=N}^{N+K} A_n^c \right] \\ &\stackrel{\text{indép.}}{=} \lim_{N \rightarrow \infty} \lim_{K \rightarrow \infty} \prod_{n=N}^{N+K} (1 - \mathbb{P}[A_n]) \end{aligned}$$

Ainsi, pour n_0 assez grand, $\prod_{n=n_0}^{\infty} (1 - \mathbb{P}[A_n])$ est strictement positif. La série $\sum_{n \geq n_0} \log(1 - \mathbb{P}[A_n])$ est donc sommable si n_0 est assez grand.

(Ceci ne veut pas dire que $\sum_{n \geq 0} \log(1 - \mathbb{P}[A_n])$ est sommable car il peut y avoir parmi les premiers termes de la série des termes infinis. Mais cette subtilité n'intervient pas pour les $\mathbb{P}[A_n]$, et il suffit de démontrer que $\sum_{n \geq n_0} \mathbb{P}[A_n]$ est sommable.)

On peut maintenant conclure de deux façons différentes :

25. en particulier, les variables aléatoires uniformes existent

1. De la sommabilité précédente, il résulte que $\mathbb{P}[A_n]$ tend vers 0, donc que $\mathbb{P}[A_n] \sim \log(1 - \mathbb{P}[A_n])$. On conclut en appliquant un lemme de comparaison de séries.
2. On applique l'inégalité de convexité $x \leq |\log(1 - x)|$, valable sur $[0, 1]$.

□

EXERCICE 24. Soit N un entier supérieur ou égal à 2. Soit (X_n) une suite de variables aléatoires indépendantes uniformes sur $\{0, \dots, N - 1\}$. Soit (k_1, \dots, k_m) un m -uplet d'éléments de $\{0, \dots, N - 1\}$.

1. Montrer que, presque sûrement, il existe n tel que $(X_{n+1}, \dots, X_{n+m}) = (k_1, \dots, k_m)$.
2. On rappelle qu'un nombre est dit *univers en base N* si toute succession finie de chiffres apparaît (une infinité de fois) dans son développement en base N . Dédurre de la question précédente que presque tout nombre réel (pour la mesure de Lebesgue) est univers en base N .
3. En déduire l'existence d'un nombre réel univers en toute base.

Remarque. De même que le théorème de Baire, la méthode probabiliste permet de prouver que des objets pathologiques existent en montrant qu'ils forment le cas générique : on cherchait du foin dans une botte de foin... (Merci à Maxime Bourrigan pour la somptueuse expression.)

EXERCICE 25. Soit (X_n) une suite de variables aléatoires uniformes sur $[0, 1]$ définies sur un même espace probabilisé. Notez qu'on ne les a pas supposées indépendantes.

1. Montrer que, presque sûrement, $n^2 X_n \xrightarrow{n \rightarrow \infty} +\infty$.
2. Montrer que, si les X_n sont indépendantes, alors, presque sûrement, $\liminf n X_n = 0$.
3. Montrer que la conclusion précédente n'a pas forcément lieu si les X_n ne sont pas indépendantes.

3.1.3 Loi du 0-1 de Kolmogorov

Soit $(\mathcal{F}_n)_{n \geq 0}$ une suite de tribus *indépendantes*. On appelle *tribu queue*, ou *tribu asymptotique*, l'intersection des tribus

$$\mathcal{G}_N := \sigma(\mathcal{F}_n; n \geq N)$$

LOI DU 0-1 DE KOLMOGOROV. Tout événement de la tribu queue a probabilité 0 ou 1.

Preuve. Notons \mathcal{F}_∞ la tribu queue. Si on sait que \mathcal{F}_∞ est indépendante d'elle-même, on a pour tout $A \in \mathcal{F}_\infty$,

$$\mathbb{P}[A] = \mathbb{P}[A \cap A] = \mathbb{P}[A]^2,$$

A étant indépendant de lui-même, et la proposition est prouvée.

Vérifions donc que \mathcal{F}_∞ est indépendante d'elle-même. Posons $\mathcal{H}_N := \sigma(\mathcal{F}_n; n < N)$. D'après le corollaire 1.9, pour tout N , \mathcal{G}_N et \mathcal{H}_N sont indépendants. En particulier, pour tout N , les tribus \mathcal{F}_∞ et \mathcal{H}_N sont indépendantes (en effet, \mathcal{F}_∞ est une sous-tribu de \mathcal{G}_N).

$\bigcup_N \mathcal{H}_N$ est un π -système, comme union croissante de tribus. En vertu du théorème 1.7, la tribu qu'il engendre est indépendante de \mathcal{F}_∞ . Or cette tribu est \mathcal{G}_0 , qui contient \mathcal{F}_∞ . Ceci achève la démonstration de la loi du 0-1 de Kolmogorov. \square

Remarque. On pourrait penser avoir prouvé un peu plus que la loi du 0-1 : il nous suffisait d'avoir l'indépendance de \mathcal{F}_∞ avec elle-même et on a établi son indépendance avec la plus grosse tribu \mathcal{G}_0 . En réalité, ce n'est pas un résultat plus fort : si une tribu ne contient que des événements de probabilité 0 ou 1, elle est indépendante de n'importe quelle tribu !

EXERCICE 26. Vérifier l'appartenance à la tribu queue des événements considérés dans les deux applications suivantes de la loi du 0-1.

Application. Si (A_n) est une suite d'événements indépendants, $\limsup A_n$ a probabilité 0 ou 1. (On avait déjà vu cela lorsqu'on avait établi le lemme de Borel-Cantelli indépendant ; cela n'empêche pas la loi du 0-1 d'éclairer ce fait d'un nouveau jour.)

Application. Si (X_n) est une suite de variables aléatoires complexes indépendantes, l'événement " $\sum_{n \geq 0} X_n(\omega)$ converge" a probabilité 0 ou 1.

3.2 Convergence de variables définies sur le même espace probabilisé

Comme cela a déjà été indiqué en début de chapitre, les notions de convergence ont ceci d'essentiel en mathématiques qu'elles permettent de passer du calcul exact au raisonnement qualitatif ; on comprendra donc aisément pourquoi il est crucial pour le probabiliste de pouvoir parler de convergence de variables aléatoires.

De même qu'on a plusieurs notions de convergence pour les suites de fonctions, on aura plusieurs notions de convergence pour les variables aléatoires²⁶.

DÉFINITION. Soit (X_n) une suite de variables aléatoires définies sur un même espace probabilisé et prenant leur valeur dans un même espace topologique. On dit que X_n converge *presque partout* si, presque sûrement, $X_n(\omega)$ converge.

26. qui, on ne l'a pas oublié, sont elles mêmes des fonctions

Remarque. À peu de choses près, c'est une convergence simple.

DÉFINITION. Soit (X_n) une suite de variables aléatoires complexes L^p définies sur un même espace probabilisé. Soit X une variable aléatoire complexe L^p définie sur ce même espace probabilisé. On dit que X_n converge vers X dans L^p si

$$\mathbb{E}[|X_n - X|^p] \xrightarrow{n \rightarrow \infty} 0$$

Remarque. C'est la convergence L^p avec laquelle vous êtes déjà familiers ; elle a pour cadre l'espace fonctionnel L^p .

DÉFINITION. Soit (X_n) une suite de variables aléatoires complexes définies sur un même espace probabilisé. Soit X une variable aléatoire complexe définie sur ce même espace probabilisé. On dit que X_n converge *en probabilité* vers X si

$$\forall \epsilon > 0, \mathbb{P}[|X_n - X| < \epsilon] \xrightarrow{n \rightarrow \infty} 1$$

On note

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

Remarques. Converger en ce sens signifie que, pour tout seuil de précision, si n est grand, X_n sera proche de X avec grande probabilité.

On peut en fait définir cette convergence dès que les variables aléatoires considérées, définies sur le même espace probabilisé, prennent leurs valeurs dans un même espace métrique, qu'on supposera toujours "sympathique" (localement compact) ; il suffit de remplacer $|X_n - X|$ par $d(X_n, X)$.

Remarque. Si X_n converge en un des trois sens précédents vers X et aussi vers Y , alors X et Y sont égales presque partout. (On suppose que la convergence vers Y a lieu dans le même sens que celle vers X .)

EXERCICE 27. Tracer un diagramme, que vous conserverez, avec toutes les implications/absences d'implications reliant les différentes notions de convergence probabiliste que vous connaissez.

PROPOSITION 3.7. *Si X_n converge vers X presque partout, ou dans un L^p , alors X_n converge vers X en probabilité. En particulier, si X_n converge presque partout et dans L^p , c'est vers la même limite.*

Preuve. Si X_n converge presque partout, appliquer le théorème de convergence dominée à $1_{|X_n - X| < \epsilon} \leq 1$.

Si X_n converge dans L^p , appliquer l'inégalité de Markov à $|X_n - X|^p$. À moins que p ne soit infini, auquel cas le résultat est immédiat. \square

Remarque. On déduit de la proposition précédente que si X_n converge vers X en un des trois sens précédents et si X_n converge vers Y en un (possible-ment autre) des trois sens précédents, alors $X = Y$ presque sûrement.

Contre-exemple. Soit U une variable aléatoire uniforme sur $[0, 1]$. Posons $X_n := n1_{U \leq 1/n}$. La suite (X_n) converge presque sûrement vers 0, chaque X_n est dans L^1 . Pourtant, X_n ne converge pas vers 0 dans L^1 . Ceci implique que X_n ne converge pas dans L^1 (utiliser la proposition précédente).

Cependant, rappelons le

THÉORÈME 3.8 (CONVERGENCE DOMINÉE). *On suppose $p < \infty$. Si X_n converge presque sûrement vers X , et s'il existe Y dans L^p telle que $\forall n, |X_n| \leq |Y|$, alors X_n converge vers X dans L^p .*

Contre-exemple. Soit A_n une suite de boréliens de $[0, 1]$ telle que

1. la mesure de Lebesgue de A_n tende vers 0,
2. $\limsup A_n = [0, 1]$.

Posons $X_n := 1_{A_n}$. La suite (X_n) ne converge pas presque partout (cette suite converge presque nulle part), pourtant X_n tend vers 0 dans tout L^p .

Évidemment, pour que cela soit un contre-exemple, il faut prouver qu'il existe une telle suite A_n . Voici deux choix possibles :

- $A_n := \left\{ x : \exists m \in \mathbb{N}, \sum_{i=1}^n i \leq x + m \leq \sum_{i=1}^{n+1} i \right\}$,
- pour $i \in \mathbb{N}$ et $j \in \{0, \dots, 2^i - 1\}$, poser $B_{i,j} := \left[\frac{j}{2^i}, \frac{j+1}{2^i} \right]$, puis ré-arranger ces termes en une suite A_n .

Néanmoins, on a la proposition suivante.

PROPOSITION 3.9. *Si la suite (X_n) converge en probabilité vers X , alors il en existe une sous-suite qui converge presque sûrement vers X . En particulier, c'est le cas si X_n converge vers X dans un L^p .*

Preuve. Pour $k \geq 1$ et $n \geq 0$, posons $A_n^k := \{\omega : |X_n - X| \geq 1/k\}$. A k fixé, $\mathbb{P}[A_n^k]$ tend vers 0 quand n tend vers l'infini (par hypothèse). Toujours à k fixé, quitte à extraire, on peut supposer les $\mathbb{P}[A_n^k]$ sommables. Par procédé d'extraction diagonale, on peut même supposer que, pour tout k , les $\mathbb{P}[A_n^k]$ sont sommables en n .

Travaillons avec une telle suite extraite, qu'on continue de noter (X_n) . On veut démontrer que presque sûrement, pour tout k , à partir d'un certain rang, on a $|X_n - X| < 1/k$. Par dénombrabilité, il suffit de démontrer que pour tout k , presque sûrement, à partir d'un certain rang, on a $|X_n - X| < 1/k$. Mais cela découle du lemme de Borel-Cantelli, puisque les $\mathbb{P}[A_n^k]$ sont sommables à k fixé ! \square

PROPOSITION 3.10. Soit X_n et X des variables aléatoires complexes définies sur un même espace probabilisé. Soit $f : \mathbb{C} \rightarrow \mathbb{C}$ continue. Si X_n converge en probabilité vers X , alors $f(X_n)$ converge en probabilité vers $f(X)$.

Remarque. Ce résultat est immédiat si f est uniformément continue.

Preuve. Soit $\epsilon > 0$. Soit $\epsilon' > 0$. Soit K un compact tel que $\mathbb{P}[X \in K] > 1 - \epsilon'$. (Il en existe un par principe de convergence monotone.) Soit δ donné par l'uniforme continuité de f sur le 1-voisinage de K , pour ϵ . Soit n_0 tel que

$$\forall n \geq n_0, \mathbb{P}[|X_n - X| < \min(1, \delta)] > 1 - \epsilon'$$

(Ce n_0 existe par convergence en probabilité.)

Pour $n \geq n_0$,

$$\mathbb{P}[X \in K \text{ et } |X_n - X| \leq \min(1, \delta)] > 1 - 2\epsilon'.$$

Or, sur cet événement, $|f(X_n) - f(X)| \leq \epsilon$. La proposition en découle. \square

3.3 Loi des grands nombres

La loi des grands nombres est un résultat fondamental de théorie des probabilités. Rappelons que la définition informelle des probabilités provenait de l'observation empirique de cette loi des grands nombres ; ayant pris le parti d'axiomatiser la théorie à partir de la notion de probabilité, il est important, à un moment, de vérifier la cohérence du tout en établissant formellement la convergence des ratios.

Par ailleurs, si on ne croit pas en l'aléatoire, on voit la théorie des probabilités comme une simple partie de la théorie de l'intégration et la loi des grands nombres est alors un théorème à part entière ne demandant aucun acte de foi initial. La loi des grands nombres peut servir à démontrer des théorèmes dont l'énoncé ne fait pas intervenir de variables aléatoires. L'existence d'un nombre normal est un exemple de tel résultat.

3.3.1 Premières versions

VOCABULAIRE. Une variable aléatoire réelle est dite *centrée* si son espérance est (bien définie et) nulle.

Soit (X_i) une famille de variables aléatoires réelles L^2 définies sur un même espace probabilisé. Elles sont dites *décorrélées* si pour tous i et j distincts, $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i] \cdot \mathbb{E}[X_j]$.

Remarques. Des variables aléatoires L^2 centrées sont décorréliées si et seulement si elles sont orthogonales dans L^2 .

La décorrélation est un affaiblissement de l'indépendance deux à deux. Des événements sont indépendants deux à deux ssi leurs indicatrices sont décorréliées.

Soit X et Y deux variables aléatoires à valeurs dans un espace E . Elles sont indépendantes ssi pour toutes f, g applications mesurables bornées de E vers \mathbb{R} , $f(X)$ et $g(Y)$ sont décorrélées.

PROPOSITION 3.11 (LOI FAIBLE DES GRANDS NOMBRES L^2). Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires décorrélées de même espérance m et de variance majorée par σ^2 . Notons

$$S_n := X_1 + \cdots + X_n$$

Alors $\frac{S_n}{n}$ converge en probabilité vers m .

Remarque. Quand on parle de loi faible ou de loi forte des grands nombres, l'épithète "faible" ou "forte" ne se rapporte pas aux hypothèses faites mais à la convergence apparaissant dans la conclusion : "faible" pour une convergence en probabilités, "forte" pour une convergence presque sûre.

Preuve. Soit $\epsilon > 0$. On a

$$\begin{aligned} \mathbb{P} \left[\left| \frac{S_n}{n} - m \right| \geq \epsilon \right] &\leq \epsilon^{-2} \mathbb{V}\text{ar} \left(\frac{S_n}{n} \right) \\ &\leq (\epsilon n)^{-2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{\epsilon^2 n} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

où la première ligne découle de l'inégalité de Chebyshev et la seconde de l'hypothèse de décorrélation. Et c'est tout ! \square

Remarques. Remarquez que si vous avez un contrôle explicite sur la variance commune σ^2 , par exemple si vous savez majorer les $\mathbb{E}[X_i^2]$, alors la preuve vous fournit un contrôle explicite sur la vitesse de convergence. Ceci vous sera utile si vous voulez utiliser la méthode de Monte-Carlo de façon parfaitement éthique et responsable.

Avec une preuve plus brève encore, on peut démontrer, sous les hypothèses précédentes, la conclusion au sens de la convergence L^2 , qui implique la convergence en probabilités. (Mais pour montrer que la convergence L^2 implique celle en probabilités, on utilise l'argument de Chebyshev présent dans la preuve précédente.)

PROPOSITION 3.12 (LOI FORTE DES GRANDS NOMBRES L^4). Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires complexes indépendantes identiquement distribuées et dans L^4 . Notons

$$S_n := X_1 + \cdots + X_n \quad \text{et} \quad m := \mathbb{E}[X_1]$$

Alors $\frac{S_n}{n}$ converge presque sûrement vers m .

Preuve. Quitte à remplacer les X_i par $X_i - m$, on peut supposer $m = 0$. Quitte à remplacer les X_n par leurs parties réelle et imaginaire, on peut les supposer réelles.

Soit $\epsilon > 0$. Posons $A_n := \left\{ \left| \frac{S_n}{n} \right| > \epsilon \right\}$. Il s'agit de prouver que $\limsup A_n$ a probabilité nulle. (On peut intervertir presque partout et le "soit ϵ " car on peut prendre ϵ rationnel.) Aussi, par le lemme de Borel-Cantelli, il nous suffit de montrer que la série des $\mathbb{P}[A_n]$ est sommable.

On cherche un contrôle sur des probabilités et a des hypothèses d'intégrabilité : l'inégalité de Markov nous appelle !

$$\mathbb{P}[A_n] = \mathbb{P} \left[\left| \frac{S_n}{n} \right| > \epsilon \right] = \mathbb{P} \left[S_n^4 > (\epsilon n)^4 \right] \leq \frac{\mathbb{E}[S_n^4]}{\epsilon^4 n^4}$$

$S_n^4 = (X_1 + \dots + X_n)^4$ se développe sous la forme

$$\sum_i X_i^4 + 6 \sum_{i \neq j} X_i^2 X_j^2 + \dots,$$

où " \dots " est une somme de termes s'écrivant sous la forme

$$X_i \times \text{produit de } X_{\text{autre chose que } i}.$$

L'espérance de chaque tel terme est nulle, par indépendance et car X_i est centrée. Ainsi,

$$\begin{aligned} \mathbb{E}[S_n^4] &\stackrel{\text{indép.}}{=} \sum_{i=1}^n \mathbb{E}[X_i^4] + 6 \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i^2] \mathbb{E}[X_j^2] \\ &\stackrel{\text{m loi}}{=} \sum_{i=1}^n \mathbb{E}[X_1^4] + 6 \sum_{1 \leq i < j \leq n} \mathbb{E}[X_1^2]^2 \\ &= \mathcal{O}(n^2) \end{aligned}$$

Par conséquent, $\mathbb{P}[A_n] = \mathcal{O}(n^{2-4})$ et la série des $\mathbb{P}[A_n]$ est sommable, ce qui met un terme à la démonstration. \square

Remarques. Les hypothèses L^2 et L^4 ne sont pas optimales, mais elle englobent la majeure partie des cas qui nous intéresseront (par exemple toutes les lois du programme). De plus, n'oublions pas que L^∞ est inclus dans tous les L^p (on fait un cours de probabilités).

On peut par exemple dès maintenant appliquer la loi des grands nombres aux variables aléatoires de Bernoulli. N'oubliez pas que les variables aléatoires de Bernoulli sont inintéressantes : elles sont essentielles ! Si A est un événement, 1_A est une variable aléatoire de Bernoulli de paramètre $\mathbb{P}[A]$. Ainsi, on dérive de la loi des grands nombres L^4 le résultat qui suit. Soit

(X_n) une suite de v.a.i.i.d., et soit B un borélien de l'espace où les X_n prennent leurs valeurs. Alors, presque sûrement,

$$\frac{\text{Card}(\{k \leq n : X_k \in B\})}{n} \xrightarrow[n \rightarrow \infty]{} \mu(B),$$

où μ est la loi de l'une quelconque des X_n .

Voici dès maintenant une application de la loi forte des grands nombres pour les variables aléatoires de Bernoulli.

3.3.2 Fonctions de répartition empiriques et théorème de Glivenko-Cantelli

Soit (X_n) une suite de v.a.i.i.d. sur \mathbb{R} . Notons F_n la fonction de répartition de la “mesure aléatoire” $\frac{\delta_{X_1} + \dots + \delta_{X_n}}{n}$. Puisque j’ai écrit le mot aléatoire dans un contexte inhabituel, il faut que je précise une certaine condition de mesurabilité. Ici, on aura juste besoin de remarquer que, à $(n, x) \in \mathbb{N}^* \times \mathbb{R}$ fixé, $\omega \mapsto F_n(\omega, x)$ est mesurable.

THÉORÈME 3.13 (GLIVENKO-CANTELLI). *Reprenons les notations des lignes précédentes. Notons F la fonction de répartition de l’une quelconque des X_i . Alors, il existe un événement de probabilité 1 sur lequel F_n converge uniformément vers F .*

Remarque. La conclusion est ainsi formulée car la condition “ F_n converge uniformément vers F ” n’a a priori aucune raison d’être mesurable.

Preuve. Soit $x \in \mathbb{R}$ et $Y_i^x := 1_{X_i \leq x}$. Les Y_i^x sont i.i.d. de loi de Bernoulli de paramètre $F(x)$. Ainsi, $F_n(x) = \frac{Y_1^x + \dots + Y_n^x}{n}$ converge presque sûrement vers $F(x)$. Aussi, presque sûrement, cette convergence a lieu en tout point x rationnel (\mathbb{Q} étant dénombrable). Si F est continue, ceci suffit pour conclure, puisqu’on sait F monotone.

Si F n’est pas continue, il faut travailler un peu plus. Ce cas est géré par Durrett : c’est, dans la troisième édition, l’exemple 7.2 du chapitre 1. \square

Remarque. Le théorème de Glivenko-Cantelli est porteur de sens. Il dit que si on a accès à un échantillon infini de v.a.i.i.d., on est capable d’en retrouver la loi. Plus précisément, il existe une façon d’attribuer à un \mathbb{N} -uplet de réels une mesure de probabilité sur \mathbb{R} de telle sorte que quelques soient

- μ est une mesure de probabilité sur \mathbb{R} ,
- (X_n) est une suite de v.a.i.i.d. de loi μ ,

notre procédure renvoie presque sûrement la loi μ . En effet, à partir des X_n , on peut construire les F_n , qui convergent presque sûrement uniformément vers F . Or on sait que la fonction de répartition caractérise la loi. Notons

cependant que cette conséquence du théorème de Glivenko-Cantelli peut être établie plus directement : à partir des X_i , on peut presque sûrement, par loi des grands nombres, retrouver $F|_{\mathbb{Q}}$, c'est-à-dire les $\mathbb{P}[X_1 \leq x]$ pour x rationnel, ce qui suffit à retrouver la loi de X_1 . Bref, les ingrédients clés sont :

- la loi forte des grands nombres pour les variables de Bernoulli,
- le fait qu'une mesure de probabilité sur \mathbb{R} peut être caractérisée par une quantité dénombrable de valeurs de probabilités.

3.3.3 Lois des grands nombres L^1

Je ne vous ferai pas languir plus longtemps : voici venir les stars.

THÉORÈME 3.14 (LOI FAIBLE DES GRANDS NOMBRES). *Soit (X_n) une suite de variables complexes aléatoires intégrables indépendantes identiquement distribuées. Notons*

$$S_n := X_1 + \cdots + X_n$$

Alors $\frac{S_n}{n}$ converge en probabilité vers m , l'espérance de X_1 .

THÉORÈME 3.15 (LOI FORTE DES GRANDS NOMBRES). *Soit (X_n) une suite de variables complexes aléatoires intégrables indépendantes identiquement distribuées. Notons*

$$S_n := X_1 + \cdots + X_n$$

Alors $\frac{S_n}{n}$ converge presque sûrement vers m , l'espérance de X_1 .

Commentaire. Avoir la loi forte des grands nombres sous les seules hypothèses d'indépendance et d'intégrabilité est essentiel : cela permet de concilier l'intuition pratique qu'on peut avoir de l'espérance et sa définition formelle.

Reportons au paragraphe 3.3.5 la preuve de la loi forte des grands nombres, qui implique la faible, et tâchons de mettre en relief ces énoncés, par quelques commentaires et applications. Si les questions d'optimalité des énoncés ne vous intéressent guère, vous pouvez vous rendre directement aux lignes consacrées aux applications, en 3.3.4.

Commentaires. La loi forte des grands nombres demeure valide lorsqu'on affaiblit l'hypothèse d'indépendance des X_n en indépendance deux à deux. Ce n'est pas du tout évident à partir de la preuve que je vais en donner ; pour une démonstration, se référer à l'ouvrage de Durrett.

Les lois des grands nombres sont bien sûr fausses sans hypothèse d'indépendance. (Prendre les X_i égales et non-déterministes.)

On déduit facilement ²⁷ de la loi forte des grands nombres, par troncature, que si $\mathbb{E}[X_-] < \infty$ et $\mathbb{E}[X_+] = +\infty$ et si les X_i , indépendantes, ont même loi que X , alors, presque sûrement, $\frac{S_n}{n}$ tend vers $+\infty$.

Dans l'énoncé de la loi forte des grands nombres, on peut remplacer convergence presque sûre par convergence L^1 . C'est un résultat moins difficile que celui de la convergence presque sûre ; pourtant, il ne se déduit pas de la loi forte des grands nombres. Pour une preuve, vous pourrez par exemple consulter le livre de Toulouse. Attention, l'énoncé dont il est question devient faux lorsqu'on remplace l'hypothèse d'indépendance par celle d'indépendance deux à deux. Le contre-exemple est le suivant.

Contre-exemple. (²⁸) Soit (ξ_n) une suite de variables aléatoires indépendantes identiquement distribuées de loi $\frac{\delta_{-1} + \delta_1}{2}$. On va définir à partir de ces variables des variables aléatoires X_i indépendantes deux à deux de loi $\frac{\delta_{-1} + \delta_1}{2}$ telles que $\frac{X_1 + \dots + X_n}{n}$ ne converge pas vers 0 dans L^1 .

Posons

- $X_1 = \xi_1$,
- si $m = 2^{n-1} + j$ avec $0 < j \leq 2^{n-1}$, alors $X_m = X_j \xi_{n+1}$,
- $S_n := X_1 + \dots + X_n$.

On peut vérifier que $\mathbb{P}[|S_n| = 2^n] = \mathbb{P}[S_n \neq 0] = 2^{-(n-1)}$ et en déduire la propriété désirée.

Commentaires. On peut aussi se demander ce qu'on peut dire lorsqu'on ne fait aucune hypothèse d'intégrabilité. Par exemple, si S_n/n converge, peut-on en déduire l'intégrabilité des X_i ? La réponse est oui. Plus précisément, si (X_n) est une suite de variables aléatoires indépendantes identiquement distribuées et si $\frac{S_n}{n}$ converge presque sûrement vers une constante m finie, alors les X_i sont intégrables (donc d'espérance m). On peut même seulement supposer que $\frac{S_n}{n}$ converge, sur un événement de probabilité positive, vers une limite finie pouvant être aléatoire, la loi du 0-1 nous permettant de ramener ce cas au précédent.

On pourrait aussi se poser la question de la convergence du ratio $\frac{S_n}{n}$ quand les $(X_i)_-$ et $(X_i)_+$ ne sont pas intégrables. Dans ce cas, deux comportements, et deux seulement, peuvent se présenter : soit les ratios convergent presque sûrement vers $+\infty$ (ou presque sûrement vers $-\infty$), soit, presque sûrement, leur limite inférieure est $-\infty$ et leur limite supérieure $+\infty$. Voir l'article *The strong law of large numbers when the mean is undefined*, de Bruce Erickson.

²⁷. Faites-le !

²⁸. chapitre 2/exemple 4.5 dans la troisième édition de *Probability : theory and examples*, de Rick Durrett

3.3.4 Méthode de Monte-Carlo

Soit f une fonction intégrable définie, disons, sur un borélien borné de \mathbb{R}^d . On aimerait estimer son intégrale. Si f est régulière, son domaine gentil et la dimension d assez petite, les méthodes d'analyse numérique sont assez efficaces. Lorsque ce n'est pas le cas, on doit chercher d'autres approches. Celle qu'on va présenter est la méthode dite de **Monte Carlo**. Elle sert en pratique quand on travaille en grande dimension.

Le domaine D étant borné, quitte à y appliquer une translation et une dilatation²⁹, on peut le supposer inclus dans $[0, 1]^d$. On peut même supposer $D = [0, 1]^d$, quitte à prolonger f par 0 en dehors de son domaine de définition initial (ceci ne change pas la valeur de l'intégrale). On supposera donc désormais $D = [0, 1]^d$. Tirons des v.a.i.i.d. uniformes sur D , i.e. tirons indépendamment des coordonnées uniformes³⁰ pour chaque vecteur. Notons-les U_i . La méthode de Monte Carlo consiste à faire l'approximation :

$$\int_D f d\lambda_n \simeq \frac{\sum_{i=1}^n f(U_i)}{n}$$

La loi *faible* des grands nombres garantit que si n est très grand, la probabilité que je commette une erreur significative est très faible. Le théorème central limite nous renseignera plus finement sur les fluctuations de l'erreur commise.

La méthode de Monte-Carlo peut notamment s'appliquer à des fonctions indicatrices, pour mesurer des volumes.

Remarques. Avoir la loi des grands nombres dans le cadre général permet d'appliquer la méthode de Monte-Carlo dès que la question de la valeur de l'intégrale de f a un sens. (En effet, on a vu que la loi des grands nombres continuait d'être valable pour les variables positives non-intégrables...)

Si vous utilisez numériquement la méthode de Monte-Carlo, après avoir choisi votre ϵ , vous devez choisir un n grand, de telle sorte que la probabilité d'avoir un écart supérieur à ϵ entre la valeur de l'intégrale et sa valeur approchée soit faible. Comment choisir ce n ? Pour le faire proprement, il faut une estimée quantitative sur la convergence " $\mathbb{P}[\text{écart} \geq \epsilon] \xrightarrow{n \rightarrow \infty} 0$ ". Si f est L^2 et si vous avez accès à un majorant de l'intégrale de f^2 , la preuve de la loi faible des grands nombres L^2 vous fournit l'estimée quantitative dont vous avez besoin.

Pour d'autres applications de la loi des grands nombres, se référer au livre de Durrett.

29. ce qui change la valeur de l'intégrale par un facteur qu'on connaît bien

30. elles-mêmes obtenues en tirant uniformément et indépendamment leurs chiffres en base 2

3.3.5 Preuve de la loi forte des grands nombres

La preuve suivante de la loi forte des grands nombres peut faire l'objet d'un développement. Les références correspondantes sont les ouvrages de Billingsley, Ouvrard (tome 2) et Toulouse.

Avant de prouver ladite loi forte des grands nombres, on prouve deux lemmes, le premier servant à établir le second.

LEMME (INÉGALITÉ MAXIMALE DE KOLMOGOROV). *Soit $(X_i)_{1 \leq i \leq n}$ une famille de variables aléatoires réelles indépendantes, L^2 et centrées. Soit de plus $\epsilon > 0$. Notons $S_k := \sum_{i=1}^k X_i$. Alors*

$$\mathbb{P} \left[\max_{1 \leq k \leq n} |S_k| \geq \epsilon \right] \leq \frac{1}{\epsilon^2} \cdot \sum_{i=1}^n \mathbb{E}[X_i^2]$$

Remarque. Notez que cette inégalité ressemble fort à une inégalité type Markov/Chebyshev, et que ce type d'inégalité constituait la moitié de la preuve de la loi forte des grands nombres L^4 .

Preuve. Posons, pour k entre 1 et n ,

$$A_k := \{\omega : \forall i < k, |S_i| < \epsilon \text{ et } |S_k| \geq \epsilon\}$$

En mots, l'événement A_k dit “ $|S_i|$ excède ϵ pour la première fois pour $i = k$ ”. Cette description rend clair le fait que les A_k sont disjoints. En tant que somme de variables aléatoires centrées, S_n est centrée, de même que $S_n - S_k$, pour $k \leq n$.

Ces définition et remarques étant faites, on peut mener les calculs suivants :

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}[X_i^2] &= \sum_i \text{Var}(X_i) && \text{car les } X_i \text{ sont centrées} \\ &= \text{Var}(S_n) && \text{par indépendance} \\ &= \mathbb{E}[S_n^2] && S_n \text{ étant centrée} \\ &\geq \mathbb{E}[S_n^2 1_{\bigcup_k A_k}] && S_n^2 \text{ étant positive} \\ &= \sum_k \mathbb{E}[S_n^2 1_{A_k}] && \text{les } A_k \text{ étant disjoints} \\ &= \sum_k \mathbb{E} \left[((S_n - S_k) + S_k)^2 1_{A_k} \right] \\ &= \sum_k \underbrace{\mathbb{E}[(S_n - S_k)^2 1_{A_k}]}_{\geq 0} + \mathbb{E}[S_k^2 1_{A_k}] + \underbrace{\mathbb{E}[2(S_n - S_k)S_k 1_{A_k}]}_{=\mathbb{E}[S_n - S_k]\mathbb{E}[2S_k 1_{A_k}]=0} \\ &\geq \sum_k \underbrace{\mathbb{E}[S_k^2 1_{A_k}]}_{\geq \epsilon^2 1_{A_k}} \geq \epsilon^2 \mathbb{P} \left[\bigcup_k A_k \right] \end{aligned}$$

□

LEMME (LOI DES GRANDS NOMBRES DE KOLMOGOROV). *Travaillons avec les mêmes notations et hypothèses que dans l'inégalité maximale de Kolmogorov. Supposons de plus que*

$$\sum_{i \geq 1} i^{-2} \mathbb{E}[X_i^2] < \infty$$

Alors $\frac{S_n}{n}$ converge presque sûrement vers 0.

Preuve. D'après le lemme de Borel Cantelli³¹, il suffit de montrer que

$$\forall \epsilon > 0, \sum_{k \geq 0} \underbrace{\mathbb{P} \left[\max_{2^k < n \leq 2^{k+1}} \frac{|S_n|}{n} \geq \epsilon \right]}_{=: p_{k,\epsilon}} < +\infty$$

Soit donc $\epsilon > 0$. Pour $k \geq 0$,

$$\begin{aligned} p_{k,\epsilon} &\leq \mathbb{P} \left[\max_{1 \leq n \leq 2^{k+1}} |S_n| \geq \epsilon 2^k \right] \\ &\leq \frac{1}{\epsilon^2 2^{2k}} \sum_{i=1}^{2^{k+1}} \mathbb{E}[X_i^2] \quad \text{par l'inégalité maximale} \end{aligned}$$

Ainsi,

$$\begin{aligned} \sum_{k \geq 0} p_{k,\epsilon} &\leq \sum_{k \geq 0} \frac{1}{\epsilon^2 2^{2k}} \sum_{i=1}^{2^{k+1}} \mathbb{E}[X_i^2] \\ &\leq \epsilon^{-2} \sum_{i \geq 1} \mathbb{E}[X_i^2] \sum_{k: 2^{k+1} \geq i} 2^{-2k} \quad \text{intersion possible : termes positifs} \end{aligned}$$

Il suffit pour conclure de montrer que $\sum_{k: 2^{k+1} \geq i} 2^{-2k} \leq \frac{C}{i^2}$, ce qu'établit le calcul ci-dessous. On note $k_0 := \min\{k \geq 0 : 2^{k+1} \geq i\}$.

$$\sum_{k: 2^{k+1} \geq i} 2^{-2k} = \sum_{k \geq k_0} 2^{-2k} = \sum_{k \geq 0} 2^{-2(k_0+k)} = 2^{-2k_0} \sum_{k \geq 0} 4^{-k} \leq \frac{4}{i^2} \sum_{k \geq 0} 4^{-k}$$

□

On peut maintenant établir la loi forte des grands nombres.

Preuve. Quitte à considérer les parties réelle et imaginaire de nos variables et à y retrancher leur espérance, on peut supposer les variables aléatoires X_i réelles et centrées. Pour $i \geq 1$, posons

$$U_i := X_i 1_{|X_i| < i} \quad \text{et} \quad U'_i := X_i - U_i$$

31. Tiens donc, le second ingrédient est arrivé. Dans le cadre L^4 , on avait besoin de Markov et de Borel-Cantelli. Le cas L^1 , bien que plus compliqué, suit la même philosophie.

Pour tout n , on peut écrire

$$S_n = \underbrace{\sum_{i=1}^n (U_i - \mathbb{E}[U_i])}_{S_{1,n}} + \underbrace{\sum_{i=1}^n U'_i}_{S_{2,n}} + \underbrace{\sum_{i=1}^n (-\mathbb{E}[U'_i])}_{S_{3,n}}$$

On va montrer que, pour $j \in \{1, 2, 3\}$, $\frac{S_{j,n}}{n}$ converge presque sûrement vers 0. (Rappelez-vous qu'on a supposé nos variables centrées.)

$S_{1,n}$: Pour $i \geq 1$, $\mathbb{E}[(U_i - \mathbb{E}[U_i])^2] \leq \mathbb{E}[U_i^2] = \mathbb{E}[X_1^2 1_{|X_1| < i}]$.

Or, pour $x \in [2, +\infty[$, $\sum_{i \geq x} i^{-2} \leq \int_{x-1}^{+\infty} t^{-2} dt = \frac{1}{x-1} \leq \frac{2}{x}$.

Ainsi,

$$\begin{aligned} \sum_{i \geq 1} i^{-2} \mathbb{E}[X_1^2 1_{|X_1| < i}] &\leq 1 + \mathbb{E} \left[X_1^2 \sum_{i \geq \max(2, |X_1|)} i^{-2} \right] \\ &\leq 1 + 2\mathbb{E}[|X_1|] < +\infty \end{aligned}$$

Ci-dessus, l'interversion somme/espérance est rendue licite par le théorème de Beppo-Levi. Si vous craignez l'expression $\sum_{i \geq \max(2, |X_1|)} i^{-2}$ qui est une “somme sur un ensemble aléatoire”, écrivez $\sum_{i \geq 2} i^{-2} 1_{i \geq |X_1|}$ à la place.

Il résulte de ce qu'on vient d'établir et de la loi des grands nombres de Kolmogorov que $\frac{S_{1,n}}{n}$ converge presque sûrement vers 0.

$S_{2,n}$: Utilisant la “formule de la baignoire”, on a

$$\sum_{i \geq 1} \mathbb{P}[U'_i \neq 0] = \sum_{i \geq 1} \mathbb{P}[|X_1| \geq i] = \mathbb{E}[|X_1|] < +\infty$$

Aussi, d'après le lemme de Borel-Cantelli, presque sûrement, seul un nombre fini de U'_i sont non-nuls. En particulier, $\frac{S_{2,n}}{n}$ converge presque sûrement vers 0.

$S_{3,n}$: La dernière suite à contrôler — $S_{3,n}$ — est déterministe.

$$\frac{S_{3,n}}{n} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_1 \cdot 1_{|X_1| \geq i}] \xrightarrow{n \rightarrow \infty} 0$$

La convergence vers 0 découle du lemme de Cesàro et de la convergence de $\mathbb{E}[X_1 \cdot 1_{|X_1| \geq i}]$ vers 0. (Cette dernière convergence découle du théorème de convergence dominée et de la nullité de $\mathbb{E}[X_1]$.) \square

3.4 Convergence en loi

DÉFINITION. Soit $(\mu_n)_{0 \leq n \leq \infty}$ une suite de mesures de probabilité sur un même espace topologique sympathique E (un espace fini discret, \mathbb{N} , \mathbb{R} , \mathbb{C} ,

$\mathbb{R}^d, \mathbb{C}^d, \dots$). On dit que μ_n converge *étroitement* vers μ_∞ si, pour toute fonction continue bornée de E vers \mathbb{R} (ou \mathbb{C}),

$$\int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu_\infty$$

On note parfois $\mu_n \xrightarrow[n \rightarrow \infty]{} \mu_\infty$.

Remarque. Rappelez-vous qu’une mesure de probabilité sur un espace symplectique E est entièrement déterminée par la donnée de la forme linéaire $f \mapsto \int f d\mu$ définie sur l’espace des fonctions continues bornées de E vers \mathbb{R} .

DÉFINITION. Soit $(X_n)_{0 \leq n \leq \infty}$ une suite de variables aléatoires, éventuellement définies sur des espaces probabilisés différents, et à valeurs dans un même espace symplectique $(\mathbb{R}, \mathbb{C}, \mathbb{R}^d, \mathbb{C}^d, \dots)$. On dit que X_n converge *en loi* (ou *en distribution*³²) vers X_∞ si la loi de X_n converge étroitement vers celle de X_∞ .

(Il serait probablement plus propre de dire que X_n converge vers μ_∞ , la loi de $X_\infty \dots$)

Remarques. “Converger en loi” est avant tout un abus de langage, alors que la convergence étroite est vraiment une convergence, dans le sens où elle provient d’une topologie sur l’espace des mesures de probabilité sur un espace symplectique.

La convergence en loi diffère énormément des autres convergences qu’on a rencontrées : la convergence d’une suite de variables aléatoires ne dépend plus de comment ces variables s’agencent entre elles ! Pour les autres notions, la convergence de X_n vers X_∞ dépendait de la loi de $(X_n)_{0 \leq n \leq \infty}$; ici, la convergence ne dépend que de la donnée de la loi de chacune de ces variables — on n’a plus besoin de la loi jointe ! On n’a même pas besoin de considérer des variables définies sur un même espace probabilisé³³ !

Il s’agit d’une convergence qui peut troubler, au début ; aussi, je commence à l’illustrer par un contre-exemple, qui a pour but de dissiper ce trouble. Rassérénés, vous pourrez ensuite passer à des exemples “plus positifs”.

Contre-exemple. Soit B une variable aléatoire de Bernoulli de paramètre $1/2$. Posons

$$X_n = \begin{cases} B & \text{si } n \text{ est pair} \\ 1 - B & \text{sinon} \end{cases}$$

La suite X_n ne converge pas en probabilité, pas plus qu’elle ne converge presque sûrement ou dans L^p . Pourtant, X_n converge en loi vers B . Elle converge d’ailleurs tout autant vers $1 - B$!

32. ici, le mot “distribution” est à prendre comme synonyme de “loi”

33. auquel cas la notion de loi jointe n’est plus du tout définie

Exemple. Si les μ_n ($0 \leq n \leq \infty$) sont des probabilités sur \mathbb{N} , la convergence étroite de μ_n vers μ_∞ équivaut au fait que, pour tout $A \subset \mathbb{N}$,

$$\mu_n(A) \xrightarrow{n \rightarrow \infty} \mu_\infty(A)$$

EXERCICE 28. Soit $(\mu_n)_{0 \leq n \leq \infty}$ une suite de mesures de probabilité sur \mathbb{N} . Montrer que les conditions suivantes s'équivalent :

1. μ_n converge étroitement vers μ_∞ ,
2. pour tout $A \subset \mathbb{N}$, $\mu_n(A) \xrightarrow{n \rightarrow \infty} \mu_\infty(A)$,
3. pour tout $k \in \mathbb{N}$, $\mu_n(\{k\}) \xrightarrow{n \rightarrow \infty} \mu_\infty(\{k\})$.

Quand le mot “tension” aura été ajouté à votre lexique probabiliste, vous rejetterez un coup d’œil à cet exercice (pourtant faisable dès maintenant).

Remarque. Cet exercice semble suggérer une définition bien plus naturelle de convergence pour les probabilités, la convergence simple en tant qu’application associant à un mesurable sa mesure. Mais ce n’est pas une bonne notion, car on veut se rappeler la topologie de l’espace qu’on munit de probabilités. Les exemples qui suivent devraient éclairer la situation.

Exemple. $\delta_{1/n} \xrightarrow{n \rightarrow \infty} \delta_0$

Exemple. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue. Soit X_n et X des variables aléatoires réelles. Si X_n converge en loi vers X , alors $f(X_n)$ converge en loi vers $f(X)$.

Exemple. Pour chaque n , tirons M_n une variable aléatoire uniforme sur $\{1, \dots, n\}$ et posons $X_n := \frac{M_n}{n}$. La suite (X_n) converge en loi vers une variable aléatoire uniforme sur $[0, 1]$. En d’autres termes,

$$\frac{1}{n} \cdot (\delta_{1/n} + \dots + \delta_{n/n}) \xrightarrow{n \rightarrow \infty} \text{Leb}([0, 1])$$

Cet énoncé paraphrase précisément le fait que l’intégrale de Riemann coïncide avec celle de Lebesgue sur les fonctions continues définies sur un segment !

La proposition à venir caractérise la convergence étroite en des termes très proches de ceux employés dans la dernière remarque. Aussi, si la définition de la convergence étroite ne vous parle pas, vous pouvez tout à fait voir cette proposition comme une définition (qui, je l’espère, sera pour vous chargée de sens) et voir la définition qu’on a vue comme une caractérisation.

PROPOSITION 3.18. Soit $(\mu_n)_{0 \leq n \leq \infty}$ une suite de probabilités sur un même espace sympathique E (espace fini discret, \mathbb{N} , \mathbb{R} , \mathbb{C} , \mathbb{R}^d , $\mathbb{C}^d \dots$). Les quatre conditions suivantes s'équivalent

1. $\mu_n \xrightarrow[n \rightarrow \infty]{} \mu_\infty$,
2. pour tout ouvert O de E , $\liminf \mu_n(O) \geq \mu_\infty(O)$,
3. pour tout fermé F de E , $\limsup \mu_n(F) \leq \mu_\infty(F)$,
4. pour tout borélien B de E , si la frontière de B est μ_∞ -négligeable, alors $\mu_n(B) \xrightarrow[n \rightarrow \infty]{} \mu_\infty(B)$.

EXERCICE 29. Prouver cette proposition, ou lire sa preuve dans le livre de Billingsley (section 29 dans la troisième édition).

Concluons enfin cette présentation de la convergence en loi en expliquant le rapport que la convergence en probabilité entretient avec elle.

PROPOSITION 3.19. *La convergence en probabilité implique la convergence en loi.*

Preuve. Jetez un coup d'œil à la proposition 3.10, et rappelez-vous que les fonctions test considérées sont bornées. . . \square

PROPOSITION 3.20. *Soit (X_n) une suite de variables aléatoires définies sur un même espace probabilisé et à valeurs dans un même \mathbb{R}^n . Supposons qu'elle converge en loi vers une variable aléatoire constante, de valeur c . Alors, X_n converge en probabilité vers c (plus précisément vers l'application constante de l'univers des X_n vers $\{c\}$).*

Preuve. Pour simplifier légèrement les notations, supposons les X_n à valeurs réelles. Posons $f : x \mapsto \frac{|x-c|}{1+|x-c|}$. Soit $\epsilon > 0$. Pour tout n , on a

$$\frac{\epsilon}{1+\epsilon} 1_{|X_n-c|>\epsilon} \leq f(X_n)$$

Passant à l'espérance, $\frac{\epsilon}{1+\epsilon} \mathbb{P}[|X_n - c| > \epsilon] \leq \mathbb{E}[f(X_n)] \xrightarrow[n \rightarrow \infty]{} \int f d\delta_c = 0$ et la preuve s'achève. \square

On ne peut pas espérer mieux (le cas où la limite ne serait pas constante). En effet,

- la convergence en loi ne dépend que de la loi,
- si une probabilité μ sur un espace sympathique n'est pas une masse de Dirac, je peux trouver deux variables aléatoires de loi μ définies sur un même espace probabilisé et telles qu'elles diffèrent avec probabilité positive.

3.5 Fonction de répartition, le retour

Revenons sur la notion de fonction de répartition, cette fois pour la mettre en lien avec celle de convergence en loi.

3.5.1 Un peu de théorie

DÉFINITION. Soit $(F_n)_{n \leq \infty}$ une suite de fonctions de répartition. On écrira

$$F_n \xrightarrow[n \rightarrow \infty]{} F_\infty$$

si, pour tout x point de continuité de F_∞ , on a $F_n(x) \xrightarrow[n \rightarrow \infty]{} F_\infty(x)$.

Remarque. Comme le suggère la notation, cette convergence sera bientôt mise en lien avec la convergence en loi. On peut percevoir dès maintenant une saveur commune. Dans la proposition 3.18, il y a une condition de μ_∞ -négligeabilité du borélien B considéré. Si on pense ici à $B =]-\infty, x]$ et remarque que x est un point de continuité de F_∞ ssi $\mu_\infty(\partial B) = 0$, la définition précédente devient nettement moins suspecte.

LEMME 3.21. Soit $(F_n)_{n \leq \infty}$ une suite de fonctions de répartition. Notons μ_n la mesure de probabilité sur \mathbb{R} ayant pour fonction de répartition F_n . Si $F_n \xrightarrow[n \rightarrow \infty]{} F_\infty$, alors il existe des variables aléatoires X_n définies sur un même espace probabilisé telles que

1. pour tout n , X_n est de loi μ_n ,
2. X_n converge presque sûrement vers X_∞ .

Preuve. Soit U une variable aléatoire uniforme sur $[0, 1]$. On notera abusivement F_n^{-1} la fonction inverse généralisée de F_n . Posons $X_n := F_n^{-1}(U)$. On vérifiera en exercice³⁴ que cette suite de variables aléatoires convient. (Indication : on pourra remarquer que tout réel est accumulation de points de continuité de F , par la gauche et par la droite.) \square

Le lemme précédent définit ce qu'on appelle un *couplage*. Plus précisément, si (μ_i) est une famille de lois, un couplage des μ_i est le choix de variables aléatoires X_i définies sur un même espace probabilisé de telle sorte que

1. pour tout i , X_i est de loi μ_i ,
2. les X_i s'agencent entre elles de façon agréable. (On parle de couplage même quand cette condition informelle n'est pas satisfaite, mais la notion perd alors une part de son intérêt.)

On peut expliquer ceci en termes de loi. Un couplage de lois μ_i définies sur des espaces E_i n'est rien d'autre qu'une mesure de probabilité (agréable) sur $\prod_i E_i$ de marginales les μ_i .

EXERCICE 30. Soit μ_1 et μ_2 deux mesures de probabilité sur \mathbb{R} . Montrer que les deux conditions suivantes s'équivalent :

³⁴. en dernier recours, vous trouverez les quelques lignes ici omises dans le livre de Durrett, section 2.2.b de la troisième édition

1. pour tout $x \in \mathbb{R}$, $F_{\mu_1}(x) \leq F_{\mu_2}(x)$;
2. il existe un couplage de μ_1 et μ_2 donnant masse nulle à

$$\{(x, y) \in \mathbb{R}^2 : x < y\}$$

(Indication : pour $1 \Rightarrow 2$, penser aux fonctions de répartition inverses généralisées.)

Pour saisir concrètement l'intérêt des couplages, lisez la preuve du théorème suivant, où vous verrez le couplage du lemme 3.21 à l'action.

THÉORÈME 3.22. *Avec les notations du lemme précédent, $\mu_n \xrightarrow[n \rightarrow \infty]{} \mu_\infty$ équivaut à $F_n \xrightarrow[n \rightarrow \infty]{} F_\infty$.*

Preuve. Si $F_n \xrightarrow[n \rightarrow \infty]{} F_\infty$, on peut prendre les X_n fournis par le lemme 3.21. Soit f fonction continue bornée de \mathbb{R} vers \mathbb{R} . Par continuité, $f(X_n)$ converge presque sûrement vers $f(X_\infty)$. Puisque f est bornée, il découle du théorème de convergence dominée que $\mathbb{E}[f(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[f(X_\infty)]$. En d'autres termes, μ_n converge étroitement vers μ_∞ .

Réciproquement³⁵, supposons que μ_n converge étroitement vers μ_∞ . Soit x un point de continuité de F_∞ . Appliquant la définition de la convergence en loi à une fonction f continue de \mathbb{R} vers $[0, 1]$ valant 1 sur $] -\infty, x - \epsilon]$ et 0 sur $[x, +\infty[$, on a

$$\liminf F_n(x) \geq \liminf \int f d\mu_n = \int f d\mu_\infty \geq F(x - \epsilon) \xrightarrow[\epsilon \rightarrow 0]{} F(x)$$

La dernière convergence a bien entendu lieu car x est supposé un point de continuité de F_∞ . On prouve de même que $\limsup F_n(x) \leq F(x)$, ce qui achève la preuve du théorème. \square

Remarque. Ce théorème est un critère pratique de vérification de convergences en loi.

De ce théorème et celui de Glivenko-Cantelli découle le résultat suivant.

THÉORÈME 3.23. *Soit (X_n) une suite de variables aléatoires réelles indépendantes de même loi μ . Il existe un événement de probabilité 1 en tout point duquel $\frac{\delta_{X_1} + \dots + \delta_{X_n}}{n}$ converge en loi vers μ .*

3.5.2 Quelques exercices

EXERCICE 31. Soit $\mu_n \xrightarrow[n \rightarrow \infty]{} \mu_\infty$ et $\nu_n \xrightarrow[n \rightarrow \infty]{} \nu_\infty$ deux suites de mesures de probabilité sur \mathbb{R} convergeant étroitement.

1. Montrer que $\mu_n \otimes \nu_n \xrightarrow[n \rightarrow \infty]{} \mu_\infty \otimes \nu_\infty$.
2. En déduire que $\mu_n * \nu_n \xrightarrow[n \rightarrow \infty]{} \mu_\infty * \nu_\infty$, où $*$ désigne le produit de convolution.

³⁵. On propose ici une preuve n'utilisant pas la boîte noire que constitue la proposition 3.18, laquelle permet de trivialisier la question en prenant $B =] -\infty, x]$.

Maintenant qu'on a bien compris ce qu'est la convergence en loi, illustrons ce concept.

EXERCICE 32. Soit $\lambda > 0$.

1. Montrer que

$$\text{Bin}(n, \min(\lambda/n, 1)) \xrightarrow[n \rightarrow \infty]{} \mathcal{P}(\lambda)$$

où $\mathcal{P}(\lambda)$ désigne la loi de Poisson de paramètre λ .

2. Le résultat que vous venez d'établir s'appelle la **loi des événements rares**. Pourquoi ?

3. Voir en quoi cela éclaire le fait suivant : “la somme de deux variables aléatoires de Poisson indépendantes est toujours une variable aléatoire de Poisson.” On pourra utiliser l'exercice 31.

EXERCICE 33. Notons (X_n) une suite de variables aléatoires telles que X_n ait une loi géométrique de paramètre $1 - 1/n$.

1. Montrer que X_n/n converge en loi vers une variable aléatoire exponentielle de paramètre 1.
2. Utilisez cela pour éclairer d'un jour nouveau la propriété de perte de mémoire de l'exponentielle.

EXERCICE 34. Soit g une fonction mesurable de \mathbb{R} vers \mathbb{R} .

1. Montrer que si X_n est une suite de variables aléatoires réelles convergeant en loi vers X_∞ , et si la probabilité que X_∞ soit un point de discontinuité de g est nulle, alors $g(X_n)$ converge en loi vers $g(X_\infty)$.
2. Montrer que si de plus g est bornée, alors $\mathbb{E}[g(X_n)]$ converge vers $\mathbb{E}[g(X_\infty)]$.

3.5.3 Applications concrètes

Voyons à quoi peu servir le théorème 3.22 dans des situations concrètes.

THÉORÈME 3.24 (DE MOIVRE-LAPLACE). Soit (X_n) une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre p . Notons

$$S_n := X_1 + \cdots + X_n$$

Alors, $\sqrt{n} \left(\frac{S_n}{n} - p \right) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, p(1-p))$.

Remarques. Le théorème central limite, qu'on établira bientôt, est la généralisation du théorème ci-dessus à des variables aléatoires autres que de Bernoulli.

Vu ce qu'on sait, on aurait pu écrire l'hypothèse du théorème de de Moivre et Laplace, sous la forme “ S_n suit une loi binomiale de paramètres

n et p ". Celle-ci est évidemment beaucoup moins propice à la généralisation que celle pour laquelle on a opté.

La loi des grands nombres nous disait que $\frac{S_n}{n} - p$ convergeait, en tout sens, vers 0. Le théorème de de Moivre et Laplace nous dit à quelle vitesse : "moralement, $\frac{S_n}{n} - p \sim \sqrt{n}$ ". (J'ai bien mis les guillemets.) Si on avait choisi un terme correctif — je parle du \sqrt{n} — croissant plus vite, on n'aurait plus eu convergence ; on dira plus tard qu'on perdrait la tension. Si on avait choisi ce terme correctif croissant plus lentement, la limite aurait été une masse de Dirac en 0.

Preuve. Pour établir le théorème de de Moivre et Laplace, il suffit, grâce au théorème 3.22, de vérifier que pour tout x réel,

$$\mathbb{P} \left[\sqrt{n} \left(\frac{S_n}{n} - p \right) \leq x \right] \xrightarrow{n \rightarrow \infty} \int_{-\infty}^x \frac{1}{\sqrt{2\pi p(1-p)}} e^{-t^2/(2p(1-p))} dt$$

Le membre de gauche, en tant que probabilité qu'une binomiale vérifie quelque chose, s'écrit sous la forme

$$\sum_{k \in I(n, x, p)} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Puis on se dépatouille laborieusement, la somme se transforme en intégrale et Stirling se charge de ramener $\sqrt{\pi}$ et l'exponentielle. \square

PROPOSITION 3.25 (PARADOXE DES ANNIVERSAIRES). *Etant donné un entier naturel N non-nul, prenons $(X_n)_{1 \leq n \leq N+1}$ une suite de variables aléatoires indépendantes uniformes sur $\{1, \dots, N\}$ et posons*

$$T_N := \min\{n : \exists m < n, X_m = X_n\}$$

Notons μ_N la loi de $\frac{T_N}{\sqrt{N}}$. Alors μ_n converge étroitement vers une loi de fonction de répartition $F : x \mapsto (1 - e^{-x^2/2})1_{x \geq 0}$.

Preuve. On a $\mathbb{P}[T_N > n] = \prod_{m=2}^n \left(1 - \frac{m-1}{N}\right)$ et il s'agit de montrer que $\mathbb{P}[T_N/\sqrt{N} > x] \xrightarrow{n \rightarrow \infty} e^{-x^2/2}$. Il n'y a donc que des calculs à faire, et je les laisse en exercice. \square

Remarque. Pourquoi "le paradoxe des anniversaires" ? Pensez N comme le nombre de jours dans une année (disons 365). La variable X_n représente l'anniversaire du $n^{\text{ième}}$ individu d'une classe, énumérée dans un ordre quelconque. Si on demande successivement³⁶ aux individus de la classe la date de leur anniversaire, T_N représente le premier instant où on aura obtenu deux fois la même réponse. Notamment, si la classe ne compte en réalité que

36. suivant leur indice

k individus, $\{T_N \leq k\}$ est l'événement “deux personnes au moins dans la classe ont le même anniversaire”, qu'on notera A .

Le fait que T_N converge en loi vers une loi non-triviale (pas une masse de Dirac en 0) indique que T_N est de l'ordre de \sqrt{N} ; que si une classe comporte $N^{1/2+\epsilon}$ individus, la probabilité de A sera très proche de 1, alors qu'avec $N^{1/2-\epsilon}$ individus, elle sera très proche de 0. Une autre façon de dire cela est que le principe des tiroirs probabiliste n'a pas besoin de $N+1$ mais de l'ordre de \sqrt{N} “tirages de chaussettes”.

Vous avez compris pourquoi le terme “anniversaire” apparaît, mais pas le terme “paradoxe”? Les données numériques suivantes vous éclaireront peut-être à ce sujet : pour $N = 365$, la probabilité de A dépasse 50% à partir de 23 élèves, et 99% à partir de $k = 57$.

Voici un petit supplément pour ceux qui rétorqueraient à raison que les anniversaires ne s'équirépartissent pas dans l'année; on ne veut plus supposer les X_i de loi *uniforme* sur $\{1, \dots, N\}$.

PROPOSITION 3.26 (ANNIVERSAIRES, LE PIRE CAS). *Soit $N \geq 2$. Si $\nu = \sum_{i=1}^N p_i \delta_i$ est une mesure de probabilité sur $\{1, \dots, N\}$, on tire (X_n) une suite de variables aléatoires indépendantes de loi ν et appelle μ_{p_1, \dots, p_n} la loi de*

$$\min\{n : \exists m < n, X_m = X_n\}$$

Posons $\vec{p}_0 = (\frac{1}{n}, \dots, \frac{1}{n})$ et prenons $\vec{p} = (p_1, \dots, p_N)$ un vecteur quelconque de nombres positifs se sommant à 1. Alors $\mu_{\vec{p}_0}$ majore stochastiquement³⁷ $\mu_{\vec{p}}$. Cela signifie l'une quelconque des deux assertions équivalentes suivantes est satisfaite

- *pour tout k , $\mu_{\vec{p}}(\{0, \dots, k\}) \geq \mu_{\vec{p}_0}(\{0, \dots, k\})$;*
- *qu'on peut trouver deux variables aléatoires K_1 et K_2 définies sur un même univers, avec K_1 de loi $\mu_{\vec{p}}$, K_2 de loi $\mu_{\vec{p}_0}$, et telles que, (presque) partout, $K_1 \leq K_2$.*

Remarque. Cette proposition affirme que le paradoxe a encore *plus* lieu quand la répartition des anniversaires n'est pas uniforme (c'est relativement intuitif : on peut commencer par penser au cas où le jour est déterministe). En particulier, le paradoxe doit avoir lieu en pratique, sous peine de porter un violent coup à notre foi en l'applicabilité des probabilités.

Preuve. L'équivalence des deux formes de la conclusion a été vue dans l'exercice 30 : on ne démontrera que la première d'entre elles. Fixons dès maintenant $k \leq N$. Pour \vec{p} quelconque (à coefficients positifs se sommant à 1),

$$f(\vec{p}) := 1 - \mu_{\vec{p}}(\{0, \dots, k\}) = \sum_i \prod_{j=1}^k p_{i_j}$$

où la somme porte sur les injections i de $\{1, \dots, k\}$ dans $\{1, \dots, N\}$. Regroupant convenablement les termes, on peut écrire cette quantité sous la forme

$$a(p_3, \dots, p_N) + (p_1 + p_2) \cdot b(p_3, \dots, p_N) + p_1 \cdot p_2 \cdot c(p_3, \dots, p_N)$$

37. autre mot, d'origine grecque, pour “aléatoirement”

où a , b et c sont à valeurs positives. Ainsi, la transformation

$$\Phi : (p_1, \dots, p_N) \mapsto \left(\frac{p_1 + p_2}{2}, \frac{p_1 + p_2}{2}, p_3, \dots, p_N \right)$$

ne peut qu'augmenter (au sens large) la valeur prise par f .

f est une application continue à valeurs réelles définie sur l'espace compact non-vidé des \vec{p} admissibles (i.e. à coefficients positifs se sommant à 1). Elle admet donc un maximum : on en notera \vec{p}_1 un point d'atteinte.

On commence par montrer qu'on peut supposer toutes les coordonnées de \vec{p}_1 non-nulles. Ses coefficients se sommant à 1, il y en a un non-nul ; disons que c'est le premier. En appliquant successivement les Φ_j (analogues de Φ changeant les première et $j^{\text{ème}}$ coordonnées en $\frac{p_1 + p_j}{2}$), on obtient un vecteur \vec{p}_1' qui maximise toujours f , et dont on sait toutes les coordonnées non-nulles. On rebaptise ce vecteur \vec{p}_1 . Supposons-le différent de \vec{p}_0 et tâchons de mettre à jour une contradiction.

Sans perdre en généralité, on peut supposer que les deux premières coordonnées de \vec{p}_1 diffèrent. La forme de c indique qu'elle prend des valeurs strictement positives quand ses arguments sont tous non-nuls. On en déduit l'inégalité stricte

$$f(\Phi(\vec{p}_1)) > f(\vec{p}_1) = \max f,$$

absurdité closant la démonstration. \square

3.5.4 Tension et compacité

Les lemmes qui suivent s'avéreront utiles pour démontrer le théorème de Lévy, duquel découle le théorème central limite. Comme bien souvent, vous pourrez lire les détails dans le Durrett.

LEMME 3.27 (LEMME DE SÉLECTION DE HELLY). *Soit (F_n) une suite de fonctions de répartition. Il en existe une suite extraite $F_{n(k)}$, ainsi qu'une fonction F_∞ croissante continue à droite telle que $F_{n(k)}(x)$ converge vers $F_\infty(x)$ dès que x est un point de continuité de F_∞*

Preuve. À x fixé, les $F_n(x)$ vivent dans le compact $[0, 1]$. On peut donc, toujours à x fixé, extraire de $F_n(x)$ une suite convergente. Par un procédé diagonal, on peut assurer que pour tout x rationnel, $F_n(x)$ converge vers un certain $G(x)$. Ayant à l'esprit qu'on cherche une fonction continue à droite, on pose

$$F_\infty : x \mapsto \inf \{ G(q) ; q \in \mathbb{Q} \cap]x, +\infty[\}$$

On vérifie ensuite que ce F_∞ convient, ce qui n'est pas difficile mais doit être écrit comme il faut... et est laissé en exercice. \square

Un second lemme — dit de tension — viendra bientôt compléter le lemme 3.27 en établissant une condition suffisante pour qu'on puisse garantir que le F_∞ obtenu est une fonction de répartition, i.e. que F_∞ tend vers 0 en $-\infty$ et 1 en $+\infty$. Avant de définir cette condition (la tension), un petit aparté topologique est bienvenu.

DÉFINITION. Notons $\mathcal{M}_f(\mathbb{R}^d)$ l'espace des mesures de masse finie sur \mathbb{R}^d , $\text{Prob}(\mathbb{R}^d)$ le sous-espace des mesures de probabilité et $\mathcal{M}_{\leq 1}(\mathbb{R}^d)$ celui des mesures de masse au plus 1. (Ces trois notations ne sont pas universelles.) Si C est un espace de fonctions mesurables bornées de \mathbb{R}^d vers \mathbb{R} , associons-lui la plus petite topologie sur $\mathcal{M}_f(\mathbb{R}^d)$ rendant continues les $\mu \mapsto \int f d\mu$.

- la topologie de la convergence *étroite* est celle associée aux fonctions continues bornées ;
- la topologie de la convergence *faible* est celle associée aux fonctions continues tendant vers 0 à l'infini ;
- la topologie de la convergence *vague* est celle associée aux fonctions continues à support compact.

Remarques. Il est malheureux de constater que, dans la littérature anglo-saxonne, “weak topology” est généralement la traduction de “topologie étroite”.

Ces topologies interviennent comme auxiliaires dans l'élaboration de la théorie, mais c'est avant tout la convergence étroite qu'il faut retenir pour la pratique. De toute façon, on a la propriété suivante !

PROPOSITION 3.28. *Sur $\text{Prob}(\mathbb{R}^d)$, les topologies vague, faible et étroite coïncident et sont métrisables.*

Cependant, des mesures de probabilité peuvent converger faiblement ou vaguement vers des mesures qui ne sont pas de probabilité. (Dans ces cas, on ne peut pas avoir convergence étroite.)

Exemple. δ_n converge faiblement et vaguement vers la mesure nulle car si f continue tend vers 0 à l'infini, $f(n)$ tend vers 0 !

Exemple. De même, la mesure de probabilité uniforme sur $[0, n]$ converge faiblement et vaguement vers la mesure nulle.

EXERCICE 35. Montrer que $\text{Prob}(\mathbb{R}^d)$ est étroitement fermé, tandis que ses adhérences faible et vague sont $\mathcal{M}_{\leq 1}(\mathbb{R}^d)$.

En fait, $\mathcal{M}_{\leq 1}(\mathbb{R}^d)$ est métrisable et compact pour la topologie faible. Une CNS pour qu'une suite de mesures de probabilité ait toutes ses valeurs d'adhérence (faibles) dans $\text{Prob}(\mathbb{R}^d)$ est la tension. Cette propriété caractérise également la relative compacité pour la topologie étroite. (C'est plus ou moins ce que nous dira le lemme de tension.)

DÉFINITION. Une suite (μ_n) de mesures de probabilité sur \mathbb{R}^d est dite *tendue* si, pour tout $\epsilon > 0$, il existe un compact $K_\epsilon \subset \mathbb{R}^d$ tel que

$$\forall n, \mu_n(K_\epsilon) \geq 1 - \epsilon$$

Remarque-exemple. Le fait qu'une suite constante de mesures de probabilité est tendue découle du théorème de convergence monotone et de la σ -compacité de \mathbb{R}^d .

EXERCICE 36. Montrer que toute suite de mesures de probabilité sur \mathbb{R}^d convergeant étroitement est tendue.

Dans la suite du cours, on n'utilisera de cet aparté topologique que la définition de la tension.

LEMME 3.29 (LEMME DE TENSION). Si (μ_n) est une suite tendue de mesures de probabilité sur \mathbb{R} , et si $F_{n(k)}$ et F_∞ sont donnés par le lemme 3.27, alors F_∞ tend vers 0 en $-\infty$ et vers 1 en $+\infty$. Autrement dit, vu le lemme de sélection, F_∞ est une fonction de répartition et $F_{n(k)} \xrightarrow[k \rightarrow \infty]{} F_\infty$.

Preuve. Soit $\epsilon > 0$. On va montrer que

$$1 - F_\infty(+\infty) + F_\infty(-\infty) \leq \epsilon,$$

ce qui suffit. (En effet, on aura alors $1 \leq 1 + F_\infty(-\infty) \leq F_\infty(+\infty) \leq 1$.)

Soit K_ϵ un compact donné par la définition de la tension de (μ_n) ; on peut le choisir sous la forme $[-M_\epsilon, M_\epsilon]$. Soit r un point de continuité de F_∞ strictement inférieur à $-M_\epsilon$. Soit s un point de continuité de F_∞ supérieur ou égal à M_ϵ . (Il existe de tels r et s .)

$$\begin{aligned} 1 - F_\infty(s) + F_\infty(r) &= \lim_{k \rightarrow \infty} 1 - F_{n(k)}(s) + F_{n(k)}(r) \\ &\leq \limsup_{n \rightarrow \infty} \mu_n(\mathbb{R} \setminus [-M_\epsilon, M_\epsilon]) \\ &\leq \epsilon \end{aligned}$$

□

3.6 Fonctions caractéristiques

DÉFINITION. Soit X une variable aléatoire à valeurs dans \mathbb{R}^d , muni de sa structure euclidienne canonique. La *fonction caractéristique* de X est

$$\varphi_X \left| \begin{array}{ll} \mathbb{R}^d & \longrightarrow \mathbb{C} \\ t & \longmapsto \mathbb{E} \left[e^{i \langle t, X \rangle} \right] \end{array} \right.$$

Terminologie. Puisqu'on appellera les φ_X fonctions caractéristiques, on prendra garde à ne pas se référer ainsi aux 1_A : ces dernières applications seront qualifiées de fonctions *indicatrices* — comme depuis le début de ce cours.

Remarques. Les fonctions caractéristiques ne sont rien d'autre que des transformées de Fourier (de mesures de probabilité sur \mathbb{R}^d).

Toute fonction caractéristique est continue, par théorème de continuité sous l'espérance (ou, ce qui revient essentiellement au même, par convergence dominée).

Exemple. Si X est une variable aléatoire de Bernoulli de paramètre p ,

$$\varphi_X : t \mapsto (1 - p) + p \cdot e^{it}$$

EXERCICE 37. Par calcul direct, calculer φ_X quand

1. X est une variable aléatoire binomiale de paramètres n et p ,
2. X est une variable aléatoire géométrique de paramètre p ,
3. X est une variable aléatoire de Poisson de paramètre λ ,
4. X est une variable aléatoire exponentielle de paramètre λ ,
5. X est une variable aléatoire uniforme sur $[0, 1]$.

Exemple. On peut vérifier que si X est une variable aléatoire gaussienne de moyenne m et variance σ^2 , alors $\varphi_X : t \mapsto e^{imt - \sigma^2 t^2 / 2}$.

PROPOSITION 3.30. Soit X et Y deux variables aléatoires indépendantes à valeurs dans \mathbb{R}^d . Pensons \mathbb{R}^{2d} comme $\mathbb{R}^d \times \mathbb{R}^d$. Alors,

$$\varphi_{(X,Y)} : (t_1, t_2) \mapsto \varphi_X(t_1) \cdot \varphi_Y(t_2)$$

Preuve. Deux ingrédients : le théorème de Fubini et la propriété de morphisme de l'exponentielle. \square

Remarque. Cette proposition reste bien entendu valable avec plus de deux variables aléatoires (pourvu qu'on ait l'hypothèse d'indépendance globale, pas seulement deux à deux). On n'est pas non plus obligé de supposer que ces variables aléatoires prennent leurs valeurs dans le même \mathbb{R}^d . Si vous voulez à tout prix un énoncé formel, attendez la proposition 3.32, qui donnera bien plus encore.

EXERCICE 38. En utilisant l'idée de la preuve de la proposition 3.30, recalculer la fonction caractéristique d'une variable aléatoire binomiale de paramètres n et p .

Vous vous en êtes rendus compte : la fonction caractéristique de X ne dépend que de la loi de X . En fait, elle la *caractérise* — d'où la dénomination.

THÉORÈME 3.31. Si deux variables aléatoires à valeurs dans \mathbb{R}^d ont même fonction caractéristique, alors elles ont même loi. (En d'autres termes, l'application qui à une mesure de probabilité sur \mathbb{R}^d associe sa transformée de Fourier est injective.)

Commentaires bibliographiques. Je ne prouverai pas ce théorème. On trouvera dans l'ouvrage de Rick Durrett la démonstration du cas unidimensionnel ; il donne essentiellement une formule pour retrouver la loi à partir de sa transformée de Fourier. Le cas général (dimension quelconque) est traité dans *Probabilités 2*, de Jean-Yves Oувrard. Il procède en convolant par un noyau gaussien qu'il fait tendre vers une masse de Dirac.

PROPOSITION 3.32 (CRITÈRE D'INDÉPENDANCE). *Soit X_1, \dots, X_n des variables aléatoires à valeurs dans $\mathbb{R}^{d_1}, \dots, \mathbb{R}^{d_n}$. Supposons-les définies sur le même espace probabilisé et pensons à $\mathbb{R}^{d_1+\dots+d_n}$ comme à $\mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_n}$.*

Alors, les X_i sont indépendantes si et seulement si

$$\forall t_1, \dots, t_n, \varphi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \prod_{i=1}^n \varphi_{X_i}(t_i)$$

Preuve. Regarder droit dans les yeux le théorème 1.11, la proposition 3.30 et le théorème 3.31. \square

Le théorème qui suit sera essentiellement l'unique ingrédient de la preuve du théorème central limite.

THÉORÈME 3.33 (LÉVY). *Soit (μ_n) une suite de mesures de probabilité sur \mathbb{R}^d , ainsi que μ_∞ une mesure de probabilité sur \mathbb{R}^d . Notons φ_n la fonction caractéristique de μ_n . Alors,*

1. *si $\mu_n \xrightarrow{n \rightarrow \infty} \mu_\infty$, φ_n converge simplement vers φ_∞ ;*
2. *si φ_n converge simplement vers une fonction φ continue en 0 (ce qui est automatique si φ est une fonction caractéristique), alors μ_n converge étroitement vers une certaine mesure de probabilité ν (et, par le point précédent, φ est la fonction caractéristique de ν).*

Remarque. Puisque la fonction caractéristique caractérise la loi, le second point est une réciproque forte du premier.

Contre-exemple. Dans la réciproque, on ne peut pas s'affranchir de l'hypothèse de continuité en 0 de la limite ponctuelle. Pour le voir, considérer une suite de gaussiennes centrées de variance tendant vers l'infini. (Notez que cette suite n'est pas tendue.)

Preuve. On ne traitera dans ces notes que le cas unidimensionnel. La preuve qui suit est issue du livre de Durrett et peut faire l'objet d'un développement.

Le point 1 résulte de la définition de la convergence en loi et du fait que, pour tout t , $x \mapsto e^{itx}$ est une application continue bornée.

Pour prouver le point 2, le travail consistera à établir que la suite (μ_n) est tendue. Voyons comment conclure si on suppose cela acquis. La tension permet d'appliquer le lemme de tension (lemme 3.29) : de toute sous-suite

de (μ_n) on peut donc extraire une suite qui converge. La limite ν de cette suite extraite sera indépendante des choix de sous-suite et extraction : en effet, d'après le premier point, sa fonction caractéristique est φ , et la fonction caractéristique caractérise la loi. Cela suffit pour avoir la convergence étroite des μ_n vers ν : en effet, le critère de “convergence via extraction” qu'on veut utiliser est vrai dans \mathbb{R} , et une convergence étroite n'est jamais qu'une convergence de suites de nombres réels (les suites $(\int f d\mu_n)_{n \in \mathbb{N}}$, pour f continue bornée).

L'hypothèse de continuité en 0 ne sert qu'à établir la tension de (μ_n) . On va maintenant mener les calculs magiques remplissant cet office. Soit $u > 0$. Notant “sinc” la fonction sinus cardinal ³⁸, on a, pour x réel,

$$\int_{-u}^u (1 - e^{itx}) dt = 2u - \int_{-u}^u (\cos(tx) + i \sin(tx)) dt = 2u(1 - \text{sinc}(ux))$$

Ré-écrivons cela sous la forme

$$\forall x \in \mathbb{R}, \quad 2(1 - \text{sinc}(ux)) = u^{-1} \int_{-u}^u 1 - e^{itx} dt$$

Intégrant par rapport à μ_n cette égalité entre fonctions bornées (donc intégrables) de x , il vient

$$\begin{aligned} 2 \int_{\mathbb{R}} 1 - \text{sinc}(ux) d\mu_n(x) &= u^{-1} \int_{\mathbb{R}} \int_{-u}^u 1 - e^{itx} dt d\mu_n(x) \\ &\stackrel{\text{Fubini}}{=} u^{-1} \int_{-u}^u \int_{\mathbb{R}} 1 - e^{itx} d\mu_n(x) dt \\ &= u^{-1} \int_{-u}^u 1 - \varphi_n(t) dt \end{aligned}$$

Par ailleurs,

$$\begin{aligned} 2 \underbrace{\int_{\mathbb{R}} \underbrace{1 - \text{sinc}(ux)}_{|\dots| \leq 1} d\mu_n(x)}_{\geq 0} &\geq 2 \int_{|x| > \frac{2}{u}} \underbrace{1 - \frac{1}{|ux|}}_{\geq 1/2} d\mu_n(x) \\ &\geq \mu_n \left(\left\{ x : |x| > \frac{2}{u} \right\} \right) \end{aligned}$$

N'oublions pas qu'on veut établir une tension : il est temps de prendre ϵ strictement positif. Prenons u_ϵ tel que

$$\forall t \in]-u_\epsilon, u_\epsilon[, \quad |1 - \varphi(t)| < \frac{\epsilon}{2}$$

38. La fonction sinus cardinal prolonge par continuité en 0 la fonction $x \mapsto \frac{\sin(x)}{x}$

Son existence est garantie par l'hypothèse de continuité en 0. Ce qu'on a prouvé pour u quelconque valant pour u_ϵ , la quantité

$$\limsup_{n \rightarrow \infty} \mu_n(\{x : |x| > 2/u_\epsilon\})$$

est majorée par

$$u_\epsilon^{-1} \int_{-u_\epsilon}^{u_\epsilon} 1 - \varphi(t) dt + \underbrace{\limsup_{n \rightarrow \infty} u_\epsilon^{-1} \left(\int_{-u_\epsilon}^{u_\epsilon} \varphi(t) - \varphi_n(t) dt \right)}_{=0 \text{ par cv. dom.}},$$

c'est-à-dire par $u_\epsilon^{-1} \int_{-u_\epsilon}^{u_\epsilon} |1 - \varphi(t)| dt < \epsilon$.

(N'ayez crainte : dans les inégalités précédentes, les intégrandes étaient complexes mais les intégrales bien réelles.)

Aussi, pour n assez grand, la μ_n -masse du compact $K_\epsilon := [-2/u_\epsilon, 2/u_\epsilon]$ est majorée par ϵ . Ceci suffit pour obtenir la tension de (μ_n) donc le théorème de Lévy. \square

Remarque. Ce théorème permet de prouver que certaines fonctions sont bien des fonctions caractéristiques (en tant que limites simples continues en 0 de fonctions caractéristiques).

EXERCICE 39. En utilisant la loi des événements rares (exercice 32), la formule de la fonction caractéristique d'une variable aléatoire binomiale et le sens facile du théorème de Lévy, retrouver la formule de la fonction caractéristique d'une variable aléatoire de Poisson.

Inversement, à l'aide des formules des fonctions caractéristiques des variables aléatoires binomiales et de Poisson ainsi que du théorème de Lévy, ré-établir la loi des événements rares.

Pour clore notre panorama des fonctions caractéristiques, voici une brève présentation du lien unissant les moments d'une variable aléatoire (définis ci-dessous) et les dérivées de sa fonction caractéristique.

VOCABULAIRE. Soit X une variable aléatoire réelle et soit $n \geq 0$. On dira que X admet un *moment* d'ordre n si elle est dans L^p pour $p = n$. Lorsque c'est le cas, si $k \leq n$, on appelle $k^{\text{ème}}$ moment de X la quantité $\mathbb{E}[X^k]$.

En intervertissant espérance et dérivation, on peut démontrer la proposition suivante.

PROPOSITION 3.34. Soit X une variable aléatoire réelle ayant un moment d'ordre n . Alors φ_X est de classe C^n et

$$\forall k \leq n, \varphi_X^{(k)}(0) = i^k \cdot \mathbb{E}[X^k]$$

Il existe une forme de réciproque à cette proposition, mais son énoncé est assez tordu (cf *Probabilités 2* de Jean-Yves Oувrard). Sa preuve fait intervenir des formules de Taylor.

3.7 Le théorème central limite unidimensionnel

Sans plus attendre, voici la star, dans sa version unidimensionnelle.

THÉORÈME 3.35 (THÉORÈME CENTRAL LIMITE). *Soit (X_n) une suite de variables aléatoires indépendantes identiquement distribuées avec un moment d'ordre 2. On note m leur espérance et σ^2 leur variance. Alors*

$$\sqrt{n} \left(\frac{S_n}{n} - m \right) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, \sigma^2)$$

Remarque. On peut s'amuser à faire se balader le σ^2 dans l'énoncé ; l'écriture suivante de la conclusion est correcte dès que $\sigma^2 > 0$:

$$\frac{S_n - nm}{\sqrt{n\sigma^2}} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1)$$

En revanche, on ne joue pas avec l'espérance : la limite est centrée, point barre !

3.7.1 Commentaires

On ne peut pas espérer, dans l'énoncé du théorème central limite, une convergence plus forte qu'en loi. Ainsi, on ne peut pas pousser plus loin le développement asymptotique de S_n entamé avec la loi des grands nombres et le théorème central limite.

EXERCICE 40. En utilisant la loi du 0-1 de Kolmogorov, montrer que la convergence en loi du théorème central limite ne peut être remplacée par une convergence en probabilités que si chaque X_i a pour loi une masse de Dirac.

- \sqrt{n} est le bon facteur de renormalisation des $\frac{S_n}{n} - m$. En effet,
- si $a_n = o(\sqrt{n})$, la suite des $a_n \left(\frac{S_n}{n} - m \right)$ converge en loi vers la variable nulle ;
 - si $\sqrt{n} = o(a_n)$, la suite des $a_n \left(\frac{S_n}{n} - m \right)$ n'est pas tendue, donc ne converge pas en loi.

Soit X une variable aléatoire de loi $\mathcal{N}(0, \sigma^2)$. On a les deux propriétés suivantes :

1. X admet un moment d'ordre 2,

2. si Y est une copie indépendante de X , $X + Y$ a même loi que $\sqrt{2}X$.
(Ceci se voit très bien avec les fonctions caractéristiques.)

En fait, la réciproque est vraie. On peut la voir comme un corollaire du théorème central limite. Mais supposons ce théorème inconnu. On peut prouver la sus-évoquée réciproque “à la main”. De cette façon, on peut prédire que si un théorème central limite existe (avec une renormalisation en \sqrt{n} qu’on aurait devinée), la loi limite sera nécessairement une gaussienne centrée (de variance elle-aussi prédictible). On peut même construire une preuve du théorème central limite à partir de cela : Vincent Tassion³⁹ m’a dit en connaître une !

Tout comme pour la “loi des grands nombres à conclusion dans L^1 ”, on ne peut pas affaiblir l’hypothèse d’indépendance en indépendance deux à deux. Le contre-exemple est rigoureusement le même.

3.7.2 Et concrètement...

Concrètement, on peut utiliser le théorème central limite pour affiner notre compréhension de l’erreur commise dans la méthode de Monte-Carlo. En reprenant les notations précédemment introduites, si f est de carré sommable,

$$\mathbb{P} \left[\left| \frac{f(X_1) + \dots + f(X_n)}{n} - m \right| \geq \frac{\epsilon}{\sqrt{n}} \right] \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\epsilon}^{\epsilon} e^{-x^2/2\sigma^2} dx$$

On dit que $\frac{f(X_1) + \dots + f(X_n)}{n}$ “fluctue en $\frac{1}{\sqrt{n}}$ ” autour de m .

Notons que le membre de droite est explicitement contrôlable dès qu’on a une majoration de l’espérance de f^2 , par exemple quand f est explicitement bornée.

On peut aussi, par une simple application du théorème central limite, montrer que si X_n désigne une variable aléatoire de Poisson de paramètre n , alors $(X_n - n)/\sqrt{n}$ converge en loi vers une gaussienne centrée réduite.

3.7.3 Démonstration

Contrairement à la loi des grands nombres, le théorème central limite aura une preuve courte. Bien entendu, c’est parce qu’on a déjà beaucoup travaillé. Récoltons le fruit de notre dur labeur...

Preuve. On établit ici la version unidimensionnelle du théorème central limite en s’appuyant sur le théorème de Lévy.

39. bureau sud Lilas 1

Sans perdre en généralité, on peut supposer $m = 0$. Notons $X := X_1$ (ou un X_k quelconque). D'après la proposition 3.34 et le théorème de Taylor-Young, on a le développement limité suivant en 0 :

$$\varphi_X(t) = 1 - \sigma^2 t^2 / 2 + o(t^2)$$

Ainsi, pour $t \in \mathbb{R}$ fixé,

$$\begin{aligned} \varphi_{S_n/\sqrt{n}}(t) &= \mathbb{E} \left[e^{it(X_1 + \dots + X_n)/\sqrt{n}} \right] \\ &= \mathbb{E} \left[\prod_{k=1}^n e^{iX_k t/\sqrt{n}} \right] \\ &\stackrel{\text{indép.}}{=} \prod_{k=1}^n \mathbb{E} \left[e^{iX_k t/\sqrt{n}} \right] \\ &= \left(1 - \frac{\sigma^2 t^2}{2n} + o\left(\frac{1}{n}\right) \right)^n \end{aligned}$$

De cette écriture, on peut déduire que, pour tout t , $\varphi_{S_n/\sqrt{n}}(t)$ converge simplement vers $e^{-t^2 \sigma^2 / 2}$.

(Convincez-vous en, tout en ayant à l'esprit que les $\varphi_X(t/\sqrt{n})$ sont complexes. Durrett et Toulouse proposent des justifications détaillées.)

Puisque $t \mapsto e^{-t^2 \sigma^2 / 2}$ est la fonction caractéristique de $\mathcal{N}(0, \sigma^2)$, l'emploi du théorème de Lévy achève la preuve. \square

4 Les marches aléatoires

Ce chapitre est consacré à l'étude des marches aléatoires. Les marches aléatoires ne sont pas explicitement au programme, mais elles constituent une excellente application des outils probabilistes. De multiples parties de ce chapitre peuvent être dissociées des autres et transformées en développements.

Qu'est-ce qu'une marche aléatoire ? On se place en dimension $d \in \mathbb{N}^*$ et se donne une suite $(X_i)_{i \geq 1}$ de variables aléatoires indépendantes identiquement distribuées de loi uniforme sur $\{x \in \mathbb{Z}^d : \|x\|_1 = 1\}$. Pour $n \in \mathbb{N}$, on pose

$$S_n := \sum_{i=1}^n X_i$$

On pensera la variable X_i comme désignant le $i^{\text{ème}}$ pas du marcheur et S_n comme la position du marcheur au $n^{\text{ème}}$ instant. Cette marche aléatoire est appelée la *marche aléatoire simple*. Quand le contexte ne précisera pas le contraire, on considérera toujours cette marche aléatoire.

Ce chapitre tourne autour du théorème suivant. Ce théorème sera démontré et, pour certains points, plusieurs démonstrations seront présentées.

THÉORÈME 4.1 (POLYA). *Si $d \leq 2$, la marche aléatoire est récurrente : presque sûrement, tout site de \mathbb{Z}^d est visité une infinité de fois par la marche.*

Si $d \geq 3$, la marche aléatoire est transiente, ou transitoire : presque sûrement, tout site de \mathbb{Z}^d n'est visité qu'un nombre fini⁴⁰ de fois par la marche. De plus, pour tout $x \in \mathbb{Z}^d$, on a $\mathbb{P}[\forall n > 0, S_n \neq x] > 0$.

4.1 Retours en 0 et sommabilité

Voici une proposition ramenant une bonne partie de l'étude de la récurrence/transience à une question de sommabilité. La référence pour ce paragraphe est Durrett.

PROPOSITION 4.2. *Considérons la marche aléatoire simple en dimension d .*

Si les $\mathbb{P}[S_n = 0]$ sont sommables, alors presque sûrement, le marcheur ne revient qu'un nombre fini (peut-être nul) de fois en 0.

Si les $\mathbb{P}[S_n = 0]$ ne sont pas sommables, alors, presque sûrement, le marcheur revient une infinité de fois en 0.

Remarque. La première partie de cette proposition résulte d'une application directe du lemme de Borel-Cantelli.

Preuve. Notons τ_i les temps de retours successifs en 0.

40. peut-être nul

- $\tau_0 = 0$;
- $\tau_{i+1} = \min\{j > \tau_i : S_j = 0\} \in \mathbb{N} \cup \{+\infty\}$.

On va montrer que $\mathbb{P}[\tau_j < +\infty] = \mathbb{P}[\tau_1 < +\infty]^j$. Ceci permet de conclure. En effet, si $\mathbb{P}[\tau_1 < +\infty]$ est strictement plus petit que 1,

$$\sum_n \mathbb{P}[S_n = 0] = \text{espérance du nombre de retours en 0} = \sum_j \mathbb{P}[\tau_1 < +\infty]^j$$

est fini, et le nombre de retours en 0 est fini presque sûrement par Borel-Cantelli. Par ailleurs, si $\mathbb{P}[\tau_1 < +\infty] = 1$, alors, pour tout j , $\mathbb{P}[\tau_j < +\infty] = 1^j = 1$, si bien que $\mathbb{P}[\forall j, \tau_j < +\infty] = 1$. Bref, dans ce second cas, le nombre de retours en 0 est infini presque sûrement et $\sum_n \mathbb{P}[S_n = 0] = \sum_n 1 = +\infty$.

L'initialisation de $\mathbb{P}[\tau_j < +\infty] = \mathbb{P}[\tau_1 < +\infty]^j$ est tautologique. Prouvons donc que

$$\mathbb{P}[\tau_{j+1} < +\infty] = \mathbb{P}[\tau_j < +\infty]\mathbb{P}[\tau_1 < +\infty]$$

Notons que si les lignes qui suivent peuvent paraître un peu techniques, elles n'en encodent pas moins une morale très simple⁴¹ : la marche aléatoire, quittant 0 après son $j^{\text{ème}}$ retour, est exactement la même marche que celle qui nous intéresse.

Formellement, on a

$$\begin{aligned} \mathbb{P}[\tau_{j+1} < +\infty] &= \mathbb{P}[\exists k \in \mathbb{N}, \exists k' \in \mathbb{N}^*, \tau_j = k \text{ et } \tau_{j+1} = k + k'] \\ &\stackrel{\exists!}{=} \sum_{k \in \mathbb{N}} \sum_{k' \in \mathbb{N}^*} \mathbb{P}[\tau_j = k \text{ et } \tau_{j+1} = k + k'] \\ &= \sum_{k \in \mathbb{N}} \sum_{k' \in \mathbb{N}^*} \mathbb{P}\left[\underbrace{\tau_j = k}_{\in \sigma(X_1, \dots, X_k)} \text{ et } \underbrace{\tau_1(i \mapsto X_{k+i}) = k'}_{\in \sigma(X_{k+1}, \dots, X_{k+k'})}\right] \\ &\stackrel{\text{indép}}{=} \sum_{k \in \mathbb{N}} \sum_{k' \in \mathbb{N}^*} \mathbb{P}[\tau_j = k] \mathbb{P}[\tau_1(i \mapsto X_{k+i}) = k'] \\ &= \sum_{k \in \mathbb{N}} \sum_{k' \in \mathbb{N}^*} \mathbb{P}[\tau_j = k] \mathbb{P}[\tau_1 = k'] \\ &= \left(\sum_{k \in \mathbb{N}} \mathbb{P}[\tau_j = k] \right) \times \left(\sum_{k' \in \mathbb{N}^*} \mathbb{P}[\tau_1 = k'] \right) \\ &= \mathbb{P}[\tau_j < +\infty] \mathbb{P}[\tau_1 < +\infty] \end{aligned}$$

□

Si $d = 1$, revenir en 0 en $2n$ pas équivaut à faire précisément n pas vers la droite et n vers la gauche. La probabilité que cela arrive est

$$\binom{2n}{n} 2^{-2n} \sim \frac{1}{\sqrt{\pi n}}$$

41. qui est à la base du concept de chaîne de Markov

L'équivalent provient de la formule de Stirling : $n! \sim (n/e)^n \sqrt{2\pi n}$.

Ainsi, en dimension 1, les probabilités de retour en 0 ne sont pas sommables. Par la proposition 4.2, presque sûrement, le marcheur aléatoire unidimensionnel revient une infinité de fois en 0. (Ce qui est une forme affaiblie de l'énoncé de récurrence.)

En dimension 2, on peut regarder les coordonnées du marcheur dans le repère $((1/2, -1/2), (1/2, 1/2))$: elles suivent simultanément deux marches aléatoires simples indépendantes sur \mathbb{Z} ! Ainsi, en dimension 2

$$\mathbb{P}[S_{2n} = 0] = \left(\binom{2n}{n} 2^{-2n} \right)^2 \sim \frac{1}{\pi n}$$

La série harmonique divergeant, la marche aléatoire simple bidimensionnelle revient elle aussi presque sûrement une infinité de fois en 0.

Le miracle de la marche aléatoire bidimensionnelle qui se ramène à deux marches unidimensionnelles indépendantes n'a pas lieu en dimension supérieure⁴². A moins qu'on ne choisisse d'étudier un autre marcheur, avec des pas plus adaptés à cette astuce ! Considérons cette fois des pas Y_i indépendants identiquement distribués de loi uniforme sur $\{x \in \mathbb{Z}^d : \forall i, |x_i| = 1\}$. La position du marcheur au temps n est alors définie comme valant $M_n := \sum_{i=1}^n Y_i$. Dans ce modèle, chaque coordonnée suit une marche aléatoire simple sur \mathbb{Z} , et ceci indépendamment les unes des autres. On a donc

$$\mathbb{P}[M_{2n} = 0] = \left(\binom{2n}{n} 2^{-2n} \right)^d \sim \left(\frac{1}{\pi n} \right)^{d/2}$$

Ainsi, pour ce modèle modifié, on a la dichotomie suivante :

- si $d \leq 2$, le marcheur revient p.s. une infinité de fois en 0.
- si $d \geq 3$, le nombre de retour en 0 est fini presque sûrement.

4.2 Marche aléatoire simple et fonctions caractéristiques

L'objectif de ce paragraphe est d'utiliser les fonctions caractéristiques pour déterminer aisément si les $\mathbb{P}[S_n = 0]$ sont sommables ou non. Avec le paragraphe précédent, on obtiendra donc la dichotomie suivante (pour la marche aléatoire usuelle sur \mathbb{Z}^d).

- En dimension $d \leq 2$, le marcheur revient p.s. une infinité de fois en 0.
- En dimension $d \geq 3$, le nombre de retour en 0 est fini presque sûrement.

Une référence possible pour ce paragraphe est *Fourier series and integrals*, de Dym et Mc Kean.

42. Il faut avoir $2d = 2^d$.

Les termes impairs étant nuls, on pourra les omettre : il s'agit de déterminer la sommabilité des $\mathbb{P}[S_{2n} = 0]$. Notons

$$\varphi \left| \begin{array}{ccc} \mathbb{R}^d & \longrightarrow & \mathbb{R} \\ t & \longmapsto & \mathbb{E} \left[e^{i\langle t, X_1 \rangle} \right] = \frac{1}{d} \sum_{i=1}^d \cos(t_i) \end{array} \right.$$

La fonction caractéristique de S_n peut s'écrire de deux façons :

$$\varphi(t)^n = \varphi_{S_n}(t) = \sum_{p: \|p\|_1 \leq n} \mathbb{P}[S_n = p] e^{i\langle t, p \rangle}$$

De cette expression, il résulte que

$$\mathbb{P}[S_{2n} = 0] = \frac{1}{(2\pi)^d} \int_{[-\pi/2, 3\pi/2]^d} \varphi(t)^{2n} dt$$

Puisque $\varphi^{2n} \geq 0$, on peut intervertir licitement somme et intégrale :

$$\sum_n \mathbb{P}[S_{2n} = 0] = \frac{1}{(2\pi)^d} \int_{[-\pi/2, 3\pi/2]^d} \frac{1}{1 - \varphi(t)^2} dt$$

Étudiant localement la fonction cosinus au voisinage de 0 et π , on ramène (proprement) la question de l'intégrabilité de $\frac{1}{1-\varphi^2}$ à celle de $\|t\|^{-2}$ en 0.

Voici deux façons différentes de voir que l'intégrabilité de $\|t\|^{-2}$ au voisinage de 0 équivaut à $d > 2$.

1. On peut faire un “changement de variable polaire” et se ramener à des intégrales unidimensionnelles connues.
2. On peut partitionner $[-1, 1]^d \setminus \{0\}$ en des 2^{-k} -homothétiques de

$$[-1, 1]^d \setminus [-1/2, 1/2]^d,$$

utiliser un changement de variable linéaire et réfléchir deux minutes.

Ainsi, la dichotomie annoncée en début de paragraphe est établie. Notons que si on ne s'intéresse qu'au cas “ $d \geq 3$ ”, on n'a pas besoin de la proposition 4.2 : on n'en utilise que le sens facile, qui est un simple Borel-Cantelli !

Commentaire. Se ramener aux instants pairs a été utile pour justifier notre interversion somme/espérance : φ^2 est positive ! On peut se demander si l'interversion était possible sans cette astuce. Eh bien, oui, elle est possible : on a bien

$$\sum_{n=0}^{\infty} \mathbb{P}[S_n = 0] = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \frac{1}{1 - \varphi(t)} dt$$

Mais ce n'est pas évident ! Voir la preuve de l'interversion ci-dessous.

Preuve.

$$\begin{aligned}
\sum_{n=0}^{\infty} \mathbb{P}[S_n = 0] &= \frac{1}{(2\pi)^d} \sum_{n \geq 0} \int_{[-\pi, \pi]^d} \varphi(t)^n dt \\
&= \frac{1}{(2\pi)^d} \sum_{n \geq 0} \lim_{\epsilon \rightarrow 1^-} \int_{[-\pi, \pi]^d} (\epsilon \varphi(t))^n dt \\
&\stackrel{\text{cv. mon.}}{=} \frac{1}{(2\pi)^d} \lim_{\epsilon \rightarrow 1^-} \sum_{n \geq 0} \int_{[-\pi, \pi]^d} (\epsilon \varphi(t))^n dt \\
&\stackrel{\text{cv. dom.}}{=} \frac{1}{(2\pi)^d} \lim_{\epsilon \rightarrow 1^-} \int_{[-\pi, \pi]^d} \sum_{n \geq 0} (\epsilon \varphi(t))^n dt \\
&= \frac{1}{(2\pi)^d} \lim_{\epsilon \rightarrow 1^-} \int_{[-\pi, \pi]^d} \frac{1}{1 - \epsilon \varphi(t)} dt \\
&= \frac{1}{(2\pi)^d} \lim_{\epsilon \rightarrow 1^-} \int_{[-\pi, \pi]^d} \frac{1_{0 \leq \varphi(t) \leq 1}}{1 - \epsilon \varphi(t)} dt + \dots \\
&\quad \dots \frac{1}{(2\pi)^d} \lim_{\epsilon \rightarrow 1^-} \int_{[-\pi, \pi]^d} \frac{1_{\varphi(t) < 0}}{1 - \epsilon \varphi(t)} dt \\
&= \underbrace{\frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \frac{1_{0 \leq \varphi(t) \leq 1}}{1 - \varphi(t)} dt}_{\text{par cv. monotone}} + \dots \\
&\quad \dots \underbrace{\frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \frac{1_{\varphi(t) < 0}}{1 - \varphi(t)} dt}_{\text{par cv. dominée par 1}} \\
&= \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \frac{1}{1 - \varphi(t)} dt
\end{aligned}$$

□

4.3 Le cas unidimensionnel sans calcul

Dans cette partie, on se place dans le cas où $d = 1$. On va établir sans calcul la récurrence de la marche aléatoire. Cette preuve est légère : elle n'utilise que la loi du 0-1 de Kolmogorov. Les autres paragraphes de ce chapitre ne sont pas utilisés.

Posons $S_{\infty}^+ := \limsup S_n$ et $S_{\infty}^- := \liminf S_n$. Les événements “ S_{∞}^+ est fini” et “ S_{∞}^- est fini” appartiennent à la tribu asymptotique des X_n . D'après la loi du 0-1 de Kolmogorov, chacun de ces événements a probabilité 0 ou 1.

Supposons que l'un d'entre eux a probabilité 1. C'est aussi le cas de l'autre, car S_{∞}^+ et $-S_{\infty}^-$ ont même loi, $(X_i)_{i \geq 1}$ et $(-X_i)_{i \geq 1}$ ayant même loi. Ainsi, presque sûrement, il existe M tel que, pour tout n , $S_n \in [-M, M]$. Par conséquent, il existe M tel que $\mathbb{P}[\forall n, S_n \in [-M, M]] > 0$. Or on sait

que ce n'est pas le cas. (En effet, les événements

$$"X_{(2M+2)k} = X_{(2M+2)k+1} = \dots = X_{(2M+2)k+2M+1} = 1"$$

sont indépendants et de même probabilité ϵ non-nulle. La probabilité qu'aucun de ces événements n'ait lieu vaut $\prod_{k=1}^{\infty} (1 - \epsilon) = 0$. Or l'occurrence d'un de ces événements suffit à garantir la non-occurrence de " $\forall n, S_n \in [-M, M]$ ".)

On vient donc de démontrer que, presque sûrement, S_{∞}^+ et S_{∞}^- étaient infinis. Puisque, pour $(\epsilon, \eta) \in \{-1, +1\}^2$, les événements $S_{\infty}^{\epsilon} = \eta\infty$ appartiennent à la tribu asymptotique, il doit exister η_+ et η_- tels que presque sûrement, $S_{\infty}^+ = \eta_+\infty$ et $S_{\infty}^- = \eta_-\infty$. Comme S_{∞}^+ et $-S_{\infty}^-$ ont même loi, $\eta_- = -\eta_+$. Comme, ponctuellement, $S_{\infty}^-(\omega) \leq S_{\infty}^+(\omega)$, on a $\eta_- \leq \eta_+$, si bien que $\eta_- = -1$ et $\eta_+ = +1$.

Mais toute trajectoire faisant des sauts de ± 1 et ayant pour limites inférieure et supérieure $-\infty$ et $+\infty$ doit visiter tout site de \mathbb{Z} une infinité de fois, si bien que la récurrence est établie dans le cas unidimensionnel.

4.4 Exercices

EXERCICE 41. — Biaisé implique transient

Soit $(X_k)_{k \geq 1}$ une famille de variables aléatoires bornées indépendantes identiquement distribuées à valeurs dans \mathbb{Z}^d . Soit $(S_n)_{n \geq 0}$ la marche aléatoire associée, définie par $S_n = \sum_{k=1}^n X_k$. Pour $k \geq 1$ et $1 \leq i \leq d$, on note $X_k^{(i)}$ la $i^{\text{ème}}$ coordonnée de X_k , et on suppose qu'il existe un i tel que $\mathbb{E}[X_1^{(i)}] \neq 0$. Montrer que la marche aléatoire est transiente, c'est-à-dire que, presque sûrement, tout site de \mathbb{Z}^d n'est visité qu'un nombre fini de fois.

EXERCICE 42. — Retours en 0 et passages ailleurs

On considère la marche aléatoire simple sur \mathbb{Z}^d , pour $d \geq 1$. On désigne par X_k le $k^{\text{ème}}$ pas et par S_n la position du marcheur à l'instant n . Pour $x \in \mathbb{Z}^d$, on note

$$T^{(x)} := T_1^{(x)} = \inf\{n > 0, S_n = x\} \in \mathbb{N} \cup \{+\infty\}$$

le premier temps (non-nul) d'atteinte du point x . Plus généralement, on définit par récurrence, pour $i \geq 2$,

$$T_i^{(x)} := \inf\{n > T_{i-1}^{(x)}, S_n = x\} \in \mathbb{N} \cup \{+\infty\}$$

le $i^{\text{ème}}$ temps d'atteinte du point x .

1. Montrer que, pour tous $n \geq 2$, $k \in \mathbb{N}^n$ et $x \in \mathbb{Z}^d$,

$$\mathbb{P}[\forall i \leq n, T_i^{(x)} = k_1 + \dots + k_i] = \mathbb{P}[T^{(x)} = k_1] \prod_{i=2}^n \mathbb{P}[T^{(0)} = k_i].$$

2. Démontrer que si $T^{(0)}$ est infini avec probabilité non-nulle, alors, presque sûrement, tout site n'est visité qu'un nombre fini (peut-être nul) de fois.
3. Démontrer que si $T^{(0)}$ est fini presque sûrement et si $x \in \mathbb{Z}^d$, alors $T^{(x)}$ est fini presque sûrement.
4. Démontrer que si $T^{(0)}$ est fini presque sûrement, alors, presque sûrement, tout site est visité une infinité de fois.
5. Dans cette question, on se place en dimension $d \geq 2$. On suppose également que la probabilité que $T^{(0)}$ soit infini est strictement positive. Démontrer que, pour tout x , la probabilité que $T^{(x)}$ soit infini est strictement positive.

EXERCICE 43. Le présent exercice propose une preuve alternative du fait qu'en dimension 1, la marche aléatoire revient presque sûrement au moins une fois en 0. Une fonction $f : \mathbb{Z} \rightarrow \mathbb{R}$ est dite *harmonique* sur l'ensemble $A \subset \mathbb{Z}$ si

$$\forall x \in A, f(x) = \frac{f(x-1) + f(x+1)}{2}$$

1. Démontrer que si f est bornée et harmonique sur $\mathbb{Z} \setminus \{0\}$, alors f est constante.
2. Démontrer que la fonction

$$f : n \mapsto \mathbb{P}[\exists k \in \mathbb{N}, n + S_k = 0]$$

est harmonique sur $\mathbb{Z} \setminus \{0\}$.

3. Conclure.

EXERCICE 44. — *Marche aléatoire sur $\mathbb{Z}/N\mathbb{Z}$*

Soit $N \geq 3$. On considère la marche aléatoire simple sur $\mathbb{Z}/N\mathbb{Z}$. Plus précisément, pour $n \geq 0$, on considère $S_n = \sum_{k=1}^n X_k$, où les X_k sont des variables aléatoires indépendantes à valeurs dans $\mathbb{Z}/N\mathbb{Z}$ telles que $\mathbb{P}[X_k = -1] = \mathbb{P}[X_k = 1] = 1/2$.

1. Sans utiliser les théorèmes de ce chapitre, démontrer que, presque sûrement, tout élément de $\mathbb{Z}/N\mathbb{Z}$ est visité par la marche. On notera $V(\omega)$ le dernier élément de $\mathbb{Z}/N\mathbb{Z}$ à être découvert.
2. Démontrer que V est une variable aléatoire uniforme sur $\mathbb{Z}/N\mathbb{Z} - \{0\}$.

5 Supplément

5.1 Espérance conditionnelle

Cette partie peut faire office de développement pertinent dans plusieurs leçons d'analyse (espaces fonctionnels, espace de Hilbert, espaces L^p). On pourra se référer au livre de Barbe et Ledoux.

THÉORÈME 5.1. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit X une variable aléatoire complexe intégrable (resp. à valeurs dans \mathbb{R}_+). Soit \mathcal{G} une sous-tribu de \mathcal{F} . Alors il existe une variable aléatoire complexe $\mathbb{E}[X|\mathcal{G}]$ intégrable (resp. à valeurs dans \mathbb{R}_+) \mathcal{G} -mesurable telle que*

$$\forall G \in \mathcal{G}, \int_G \mathbb{E}[X|\mathcal{G}](\omega) d\mathbb{P}(\omega) = \int_G X(\omega) d\mathbb{P}(\omega)$$

De plus, elle est unique modulo égalité presque partout. On appelle $\mathbb{E}[X|\mathcal{G}]$ l'espérance conditionnelle de X par rapport à \mathcal{G} .

Commentaire. En peu de mots, $\mathbb{E}[X|\mathcal{G}]$ est la meilleure approximation de X n'utilisant que l'information de \mathcal{G} .

Pour s'en convaincre, il peut être utile de commencer par aborder la situation discrète : pensons Ω dénombrable et \mathcal{F} la tribu pleine. La donnée de \mathcal{G} est celle de la partition qui l'engendre, qu'on note (P_i) . L'espérance conditionnelle de X par rapport à \mathcal{G} est une variable aléatoire constante sur les P_i qui, sur P_i , vaut

$$\mathbb{E}[X|P_i] := \frac{\mathbb{E}[X1_{P_i}]}{\mathbb{P}[P_i]} = \int X(\omega) d\mathbb{P}[\bullet|P_i](\omega)$$

Les deux conditions de ce théorème ne font que généraliser cette notion au cas continu.

Preuve. Unicité : Commençons par traiter le cas où X est intégrable. Soit Y et Z deux espérances conditionnelles de X . L'événement $E := \{Y \leq Z\}$ est \mathcal{G} -mesurable. Donc $\int_E Y d\mathbb{P} = \int_E X d\mathbb{P} = \int_E Z d\mathbb{P}$ si bien que $\int_E (Y - Z) d\mathbb{P}$ est nulle. Ainsi $Y \leq Z$ presque partout. Ayant l'autre inégalité par symétrie, on a l'égalité presque sûre.

Le cas positif se traite analogiquement à l'aide de $E := \{n \geq Y \geq Z\}$.

Existence : Soit $X \in L^2(\mathcal{F})$. L'inclusion de $\mathcal{L}^2(\mathcal{G})$ dans $\mathcal{L}^2(\mathcal{F})$ permet de voir $L^2(\mathcal{G})$ comme un sous-espace vectoriel complet de $L^2(\mathcal{F})$. (Ce n'est pas le cas à strictement parler car si X est dans $\mathcal{L}^2(\mathcal{G})$, l'ensemble des fonctions \mathcal{F} -mesurables lui étant presque partout égales est généralement plus grand que celui de celles qui sont \mathcal{G} -mesurables.)

Notons Q la projection orthogonale de $L^2(\mathcal{F})$ sur $L^2(\mathcal{G})$. On sait que $\forall Z \in L^2(\mathcal{G}), \langle X - Q(X), Z \rangle = \mathbb{E}[(X - Q(X))Z] = 0$. Ceci valant en particulier pour les indicatrices d'événements \mathcal{G} -mesurables, $Q(X)$ est une espérance conditionnelle de X .

Notons que si X est positive et intégrable, toute espérance conditionnelle a priori intégrable sera aussi positive presque partout (prendre G l'ensemble des points où $\mathbb{E}[X|\mathcal{G}]$ est négative) et toute espérance conditionnelle a priori positive sera aussi intégrable (prendre $G = \Omega$). La première partie de cette remarque servira à traiter le cas positif, la seconde à traiter le cas intégrable. De plus, cela garantit, sous réserve d'existence, qu'il n'y aura aucun conflit de définitions.

Traisons le cas où X prend ses valeurs dans \mathbb{R}_+ . Posons $X_n := \max(X, n)$; c'est un élément de $L^2(\mathcal{F})$. Pour n quelconque, ayant $X_{n+1} - X_n \geq 0$, on a

$$Q(X_{n+1}) - Q(X_n) = \mathbb{E}[X_{n+1} - X_n | \mathcal{G}] \geq 0 \text{ p.s.}$$

Puisque pour tout n , presque sûrement, $\mathbb{E}[X_{n+1} | \mathcal{G}] \geq \mathbb{E}[X_n | \mathcal{G}]$, la suite $\mathbb{E}[X_n | \mathcal{G}]$ converge presque sûrement vers une variable aléatoire Y \mathcal{G} -mesurable (a priori à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$). Pour $G \in \mathcal{G}$,

$$\int_G Y d\mathbb{P} \stackrel{\text{cv. mon.}}{=} \lim_{n \rightarrow \infty} \int_G \mathbb{E}[X_n | \mathcal{G}] d\mathbb{P} = \lim_{n \rightarrow \infty} \int_G X_n d\mathbb{P} \stackrel{\text{cv. mon.}}{=} \int_G X d\mathbb{P}$$

D'où l'existence dans le cas positif. (On vérifie que Y est finie presque partout en prenant $G = \{X \leq n\}$ avec n grand.)

Le cas intégrable s'en déduit en posant

$$\mathbb{E}[X | \mathcal{G}] := \mathbb{E}[X_+ | \mathcal{G}] - \mathbb{E}[X_- | \mathcal{G}]$$

□

Il existe une autre preuve de l'existence de l'espérance conditionnelle, qui fait appel au théorème de Radon-Nikodym.

Preuve. Définissons une mesure signée (resp. positive σ -finie) sur \mathcal{G} par $\mu : G \mapsto \int_G X d\mathbb{P}$. Elle est absolument continue par rapport à $\mathbb{P}|_{\mathcal{G}}$, donc a une dérivée de Radon-Nikodym Y par rapport à elle. Par définition, Y vérifie les conditions de l'espérance conditionnelle. □

5.2 Compléments de cours

Cette sous-section dit quelques mots sur les points du programme et de son adhérence qui n'ont pas encore été abordés. Souvent, ces mots se résument à la donnée d'une référence.

5.2.1 Transformée de Laplace et grandes déviations

La *transformée de Laplace* d'une variable aléatoire réelle X est l'application

$$t \mapsto \mathbb{E}[e^{-tX}]$$

Contrairement à la fonction caractéristique de X , son domaine de définition n'est pas toujours \mathbb{R} tout entier.

La transformée de Laplace a beaucoup de propriétés communes avec la fonction caractéristique, ce qui n'est pas surprenant vu à quel point les définitions se ressemblent. On pourra lire les énoncés précis et leurs preuves dans le chapitre 13 du livre *Calcul des probabilités* de Foata et Fuchs (sections 1, 2 et 3).

En pratique, la transformée de Laplace intervient en théorie des *grandes déviations*. Voici un problème typique dit de “grandes déviations”; vous constaterez que ce problème porte bien son appellation. Si on se place dans le cadre de la loi des grands nombres, on a convergence en probabilité de $\frac{S_n}{n}$ vers m . Ainsi, on sait que $\mathbb{P}\left[\frac{S_n}{n} > m + \epsilon\right]$ converge vers 0. Mais à quelle vitesse? En fait, modulo hypothèses, cette convergence se fait à vitesse exponentielle, i.e.

$$\mathbb{P}\left[\frac{S_n}{n} > m + \epsilon\right] = e^{-nh_X(\epsilon) + o(n)}$$

Et la transformée de Laplace dans tout ça? Eh bien elle intervient dans le calcul de h_X ! (En fait, h_X est la transformée de Cramer⁴³ du logarithme de la transformée de Laplace.)

Tout cela est expliqué dans le troisième chapitre du livre *Thèmes de probabilités et statistiques* de Toulouse. Je vous encourage à lire

- la section 1 de ce chapitre,
- la partie 2.3,
- l'énoncé du théorème 3.15,
- le lemme 3.17 et sa preuve.

5.2.2 Autres points et références

Les *vecteurs gaussiens* généralisent en dimension supérieure la loi normale unidimensionnelle. Ce sont les lois limites apparaissant dans le théorème central limite multidimensionnel. Pour une présentation claire des vecteurs gaussiens, lire le chapitre 16 de l'ouvrage *L'essentiel en théorie des probabilités* de Jacod et Protter. Le *théorème central limite multidimensionnel* est énoncé à la fin du chapitre 21.

Les *chaînes de Markov à espace d'états fini* sont étudiées dans le chapitre 10 du livre *Thèmes de probabilités et statistiques* de Toulouse.

Pour la notion de *loi conditionnelle*, vous pourrez par exemple vous référer à la section 11.1 du Ouvrard.

Enfin, vous pourrez aborder le problème de la *régression linéaire au sens des moindres carrés* chez Ouvrard, section 13.4.

43. cela se rapporte à une notion de dualité qui intervient en analyse convexe

5.3 Matière à lire et à penser

Cette sous-section contient pêle-mêle quelques suggestions de lecture et exercices vous invitant à aller plus loin.

EXERCICE 45. Considérez des variables aléatoires à densité. Comparez la convergence en loi à “la convergence” des densités.

EXERCICE 46. Un individu choisit deux nombres réels différents dans sa tête puis tire à pile ou face. Il vous révèle celui désigné par la pièce. Votre but est de deviner si le nombre annoncé à voix haute était le plus grand des deux. Trouvez une stratégie qui, quelque soit le choix de l’individu, vous garantisse une probabilité de réussite strictement supérieure à $\frac{1}{2}$.

EXERCICE 47. Comment simuleriez-vous, sur ordinateur, une variable aléatoire uniforme dans \mathfrak{S}_n ?

LECTURE :

- On peut se demander si une variable aléatoire réelle est caractérisée par ses moments. Cette question est traitée dans *Probability : theory and examples*, de Durrett. Chercher “Hamburger moment problem” dans l’index.
- S’il y a 50 figurines à collectionner dans les paquets de chocapics, combien de paquets devrai-je acheter (en moyenne) avant d’avoir une collection complète ? Chercher “coupon collector’s problem” dans le livre de Durrett.
- Soit (T_n) une suite d’instants aléatoires. On suppose que $n \mapsto T_n$ croît vers l’infini presque sûrement. On suppose aussi les accroissements *stationnaires*, ce qui signifie que la loi du nombre d’instants tombant dans la fenêtre temporelle $[0, t]$ est la même que pour la fenêtre $[s, s + t]$. Enfin, on suppose les accroissements *indépendants* : cela veut dire que si on considère des fenêtres temporelles disjointes, les nombre d’instants tombant dans chacune d’entre elles forment une famille de variables aléatoires indépendantes. Sous ces conditions, les $T_{n+1} - T_n$ sont i.i.d. de loi exponentielle. Ces processus, appelés *processus de Poisson*, sont étudiés dans *Processus stochastiques*, de Dominique Foata et Aimé Fuchs.
- On peut prouver la densité des fonctions polynomiales dans $C([0, 1], \mathbb{R})$ en utilisant les probabilités. Ceci est fait dans l’ouvrage *Analyse pour l’agrégation* de Hervé Queffélec et Claude Zuily. Chercher “Weierstrass (théorème d’approx. de)” dans l’index.

Annexe : méthodes

Avant d'expliquer comment y répondre, il s'agit de dire quelques mots sur les questions qu'on se pose en probabilités.

On en a déjà parlé : la réponse à une question sur la variable aléatoire X ne dépendra que de sa loi ; une autre portant sur des variables aléatoires X_i ne dépendra que de la loi de (X_i) . La **loi** est le cadre pertinent dans lequel se pose une question probabiliste.

En sus de la loi qui détermine uniquement le problème, nous disposons de diverses notions de **convergence**.

Enfin, qu'ils constituent des objectifs à part entière ou des étapes intermédiaires, des **calculs** de quantités jalonneront les sentiers que vous emprunterez.

Une grande parties des questions que vous vous poserez pourront être exprimées en termes d'interaction entre ces trois mots fondamentaux : “loi”, “convergence” et “calculs”.

Avertissement : dans ce qui suit, je ne me poserai pas systématiquement les questions de bonne définition ou autre. Pour la rigueur, se référer aux parties correspondantes du cours.

Je dispose de la loi et veux calculer une quantité.

Le calcul d'une quantité se ramène toujours à des calculs d'espérance, on ne traitera donc que le cas de l'espérance. Supposons donnée X une variable aléatoire, disons réelle. Commençons par le cas où sa loi est donnée explicitement.

- Si la loi de X est de la forme $\sum_n p_n \delta_{x_n}$, l'espérance de $g(X)$ est $\sum_n g(x_n) p_n$.
- Si la loi de X est de densité f par rapport à la mesure de Lebesgue, l'espérance de $g(X)$ est $\int g(x) f(x) dx$.

Ces deux cas englobent la majorité de ceux que vous rencontrerez.

La loi de X peut cependant être fournie de façon plus compliquée. Par exemple, la quantité X qui vous intéresse peut être donnée sous la forme $f(X_1, \dots, X_n)$, où la loi jointe des (X_i) est donnée. (Le théorème de changement de variable abstrait dit — abstraitement — que ce calcul peut être mené.) Très souvent, les X_i seront indépendants, vous n'aurez donc qu'à intégrer par rapport à une mesure produit, pour laquelle le théorème de Fubini ramène tout au cas unidimensionnel déjà traité.

Les mesures de probabilité sont parfois données par leur fonction de répartition. Lorsque la fonction de répartition est régulière, on peut la dériver pour retrouver la densité de la probabilité considérée par rapport à la mesure de Lebesgue (proposition 1.18). S'il s'agit d'une fonction en escalier, la loi associée est de la forme $\sum_n p_n \delta_{x_n}$, où les x_n désignent les positions des sauts

et les p_n les hauteurs des marches. Enfin, de façon générale, on peut aussi utiliser la formule de la baignoire (exercice 18) pour calculer directement des espérances à partir de la fonction de répartition.

Les rares autres cas qu'il serait envisageable d'évoquer sont trop marginaux selon moi pour être mentionnés ici.

Je veux déterminer la loi d'une variable aléatoire.

On se ramène au calcul d'une quantité : une fonction de répartition, une fonction génératrice, une fonction caractéristique, la loi elle-même ... vous avez le choix ! *Les fonctions génératrices et caractéristiques se comportent agréablement avec la somme indépendante.*

Je veux vérifier une indépendance.

On se référera aux critères pratiques de vérification d'indépendance de la sous-section 1.4.4. On peut aussi utiliser le critère d'indépendance 3.32 faisant intervenir les fonctions caractéristiques.

Enfin, pour les cas plus subtils, il s'agit de ruser et d'appliquer une technique de *découpage*, illustrée ci-dessous.

EXERCICE 48. Soit $(X_n)_{n \geq 1}$ une suite de v.a.i.i.d. de Bernoulli de paramètre p . Posons $T_0 = 0$, $T_1 := \min\{i : X_i = 0\}$ et $T_{n+1} := \min\{i > T_n : X_i = 0\}$. Ces variables sont a priori à valeurs dans $\{1, 2, \dots, +\infty\}$. Les T_i sont des variables aléatoires (elles sont mesurables). Montrer qu'elles prennent presque sûrement des valeurs finies et que les $T_{i+1} - T_i - 1$ forment une suite de variables aléatoires indépendantes identiquement distribuées, de loi géométrique de paramètre p .

Solution. Soit $k_1, \dots, k_n \in \mathbb{N}^*$.

$$\begin{aligned} \mathbb{P}[T_2 - T_1 = k_1, \dots, T_{n+1} - T_n = k_n] &= \mathbb{P}\left[\forall i \leq \sum_{i=1}^n k_i, X_i = 0 \Leftrightarrow \exists j, i = \sum_{i=1}^j k_i\right] \\ &= p^{\sum_{i=1}^n k_i - n} (1-p)^n \\ &= \prod_{i=1}^n (p^{k_i-1} (1-p)) \end{aligned}$$

Donc les $T_{i+1} - T_i - 1$ sont des variables aléatoires indépendantes de loi géométrique de paramètre p . \square

Je veux prouver une convergence.

Nous avons à notre disposition la loi des grands nombres, le théorème central limite, les définitions, les implications entre convergences. ... On peut penser à d'autres outils encore.

Pour montrer que quelque chose a *presque sûrement* lieu *asymptotiquement*, on pensera au **lemme de Borel-Cantelli**.

Pour les convergences en loi, on pourra utiliser les caractérisation en termes de fonction de répartition, de fonctions caractéristiques et de convergences de mesures de boréliens gentils. On notera aussi que si X_n converge en loi vers X et si f est continue, alors $f(X_n)$ converge en loi vers $f(X)$.

Autres techniques

Pour transformer un contrôle sur les espérances en contrôle sur des probabilités, utiliser **l'inégalité de Markov**.

Pour montrer que des objets pathologiques existent, on pourra penser à montrer qu'ils sont typiques au sens des probabilités. C'est une "généralisation du principe des tiroirs". L'exercice 24 illustre ce principe.

Bien entendu, pour montrer qu'un événement a probabilité 0 ou 1, on pourra songer à employer la loi du 0-1 de Kolmogorov.

Enfin, pour se faciliter la vie, il importe de savoir **réaliser intelligemment des variables aléatoires**.

EXERCICE 49. Soit X une variable aléatoire binomiale de paramètres n et p . Calculer sa variance sachant que

- la variance d'une variable de Bernoulli de paramètre p est $p(1-p)$;
- la somme de n variables aléatoires indépendantes de Bernoulli de paramètre p est une variable aléatoire binomiale de paramètres n et p .

Solution. La variance de X est celle de n'importe quelle variable aléatoire de même loi. Soit (X_1, \dots, X_n) un n -uplet de variables aléatoires de Bernoulli indépendantes de paramètre p (cela existe).

$$\mathbb{V}\text{ar}(X) = \mathbb{V}\text{ar}(X_1 + \dots + X_n) \stackrel{\text{indép}}{=} \mathbb{V}\text{ar}(X_1) + \dots + \mathbb{V}\text{ar}(X_n) = np(1-p)$$

□

Réaliser intelligemment des variables aléatoires peut aussi signifier "introduire un couplage". L'exercice sur la percolation vu en TD illustre cela.