UNIVERSITÉ
PARIS-EST
MARNE-LA-VALLÉE

Probabilités
Préparation à l'agrégation interne

Djalil Chafaï et Pierre-André Zitt

Copyright © 2014 (sauf certaines images)
Livre électronique enregistré à l'AFNIL
ISBN 978-2-9541710-2-9 EAN 9782954171029

Probabilités - Préparation à l'agrégation interne de mathématiques (1/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Probabilités - Préparation à l'agrégation interne de mathématiques (2/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Table des matières

Table des matières **1**

1 Modélisation d'une expérience **5**
1.1 Univers . . . 5
1.2 Événements, tribus . . . 6

2 Mesures de probabilité **9**
2.1 Définitions et premières propriétés . . . 9
2.2 Équiprobabilité discrète et combinatoire . . . 10
2.3 Quelques exemples remarquables . . . 13
2.4 Expériences répétées . . . 14
2.5 Conditionnement . . . 15
2.6 Indépendance . . . 17
2.7 Équiprobabilité continue : la mesure de Lebesgue . . . 20

3 Variables aléatoires et intégration **23**
3.1 Variables aléatoires réelles . . . 23
3.2 Fonction de répartition et loi . . . 24
3.3 Espérance — définition et propriétés générales . . . 30
3.4 Espérance des variables aléatoires discrètes . . . 33
3.5 Espérance des variables aléatoires à densité . . . 34
3.6 Variance . . . 36
3.7 Inégalités de Markov et de Bienaymé-Tchebychev . . . 38
3.8 Fonction génératrice . . . 39

4 Vecteurs aléatoires **45**
4.1 Définition, loi d'un vecteur aléatoire . . . 45
4.2 Indépendance de variables aléatoires . . . 47
4.3 Moyenne et matrice de covariance . . . 50
4.4 Loi normale multivariée, vecteurs gaussiens . . . 53
4.5 Fonctions génératrices et variables indépendantes . . . 54
4.6 Retour sur l'approximation binomiale/Poisson . . . 55

5 Théorèmes limites **57**
5.1 Loi faible des grands nombres . . . 57
5.2 Loi forte des grands nombres . . . 59
5.3 Théorème limite central . . . 61
5.4 Approximation de la loi binomiale par la loi normale . . . 66

Appendices **69**

A Biais par la taille **71**
A.1 Un cas simple . . . 71
A.2 Estimation dans une file d'attente . . . 71

Probabilités - Préparation à l'agrégation interne de mathématiques (1/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

B Jeu de pile ou face **73**
B.1 Récapitulatif . . . 73
B.2 Temps d'attente des succès successifs . . . 73
B.3 Fluctuations non asymptotiques . . . 74
B.4 Lien avec la loi uniforme . . . 74

C Collectionneur d'images **77**
C.1 Définition et formules exactes . . . 77
C.2 Une décomposition et ses conséquences . . . 78
C.3 Étude fine et fluctuations . . . 79

D Marche aléatoire simple et ruine du joueur **81**
D.1 Ruine du joueur . . . 81
D.2 Premier retour en zéro de la marche aléatoire . . . 83

E Lois exponentielles et durées de vie **87**
E.1 Définition, premières propriétés . . . 87
E.2 Modélisation des durées de vie . . . 90

F Extrêmes **93**
F.1 Modélisation des phénomènes extrêmes . . . 93
F.2 Quatre exemples simples . . . 94
F.3 Un résultat général . . . 95

G Familles sommables et intégrales de Riemann **97**
G.1 Familles sommables . . . 97
G.2 Intégration de Riemann . . . 98
G.3 Les deux théorèmes fondamentaux admis . . . 98

H Convergences pour des suites de variables aléatoires **99**
H.1 Convergences de variables et de lois . . . 99
H.2 Relations entre les convergences. . . . 101
H.3 Passer à la limite dans une espérance . . . 102

I Fonctions caractéristiques et vecteurs gaussiens **105**
I.1 Fonction caractéristique . . . 105
I.2 Application aux vecteurs gaussiens . . . 106
I.3 Applications en statistiques . . . 108

J Combinatoire, loi de Poisson, et partitions **111**

Bibliographie **113**

Table des figures **113**

Liste des tableaux **114**

Index **115**

Probabilités - Préparation à l'agrégation interne de mathématiques (2/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Avant propos

Ces notes de cours couvrent les notions de probabilités au programme de l'agrégation interne de mathématiques. Elles ne constituent pas des modèles de leçons d'oral. Elles forment un livre électronique, possédant un ISBN fourni par l'AFNIL. Ce livre est donc utilisable le jour de l'oral par les candidats. L'image de couverture provient d'Internet, son propriétaire est inconnu.

Repères

L'ouvrage est divisé en deux parties :

— un cours, partant des notions de base sur les probabilités discrètes pour arriver aux deux théorèmes limites fondamentaux que sont la *loi des grands nombres* et le *théorème limite central*, illustré en particulier par quelques exemples récurrents.

— des compléments, qui illustrent les notions de cours ou les prolongent.

Les **nouveaux termes** sont composés en gras au moment de leur définition ; un index situé à la fin de l'ouvrage recense ces définitions.

Certains passages particuliers du texte sont repérés graphiquement de différentes manières.

Il est possible d'illustrer de très nombreuses notions du programme à partir d'exemples fondamentaux. Un des choix les plus courants est le *jeu de pile ou face*. Toutes les illustrations des notions du cours à partir de ce « fil rouge » seront indiqués par un liseré rouge.

Un autre exemple reviendra souvent : celui du *sondage simple*, c'est-à-dire de l'échantillonage sans remise dans une population. Les passages correspondants sont repérés par un liseré bleu.

Enfin, les notions fondamentales dans le cas continu seront illustrées par un modèle de tir uniforme sur une cible ronde.

La théorie des probabilités regorge de pièges pour l'intuition. Les erreurs courantes sont indiquées par un panneau « virage dangereux » pour éviter les sorties de route !

Le programme de l'agrégation externe exclut (implicitement) la notion d'intégration de Lebesgue, et les preuves de plusieurs résultats sont hors programme. Nous en avons pourtant inclus quelques-unes pour les curieux ; elles sont indiquées par le panneau « haute tension ».

Probabilités - Préparation à l'agrégation interne de mathématiques (3/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Remerciements

De nombreux agrégatifs ont souffert pour la préparation de ce livre, en essayant de comprendre des passages cryptiques, voire en découvrant des erreurs pendant une présentation de leçon — qu'ils en soient remerciés ! Merci aussi à deux autres repéreurs de coquilles : J.-B. Bardet et F. Malrieu.

Djalil Chafaï et Pierre-André Zitt
Marne-la-Vallée
Été 2014

Probabilités - Préparation à l'agrégation interne de mathématiques (4/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Chapitre 1

Modélisation d'une expérience

1.1 Univers

On modélise une expérience aléatoire en introduisant l'ensemble Ω qui code tous les résultats possibles de l'expérience, appelé **univers**. Les résultats possibles sont parfois appelés **épreuves** ou **événements élémentaires**. Voici quelques exemples concrets :

Expérience	Univers
un jet de dé à 6 faces	$\{1,2,3,4,5,6\}$
un jet de deux dés à 6 faces de $\neq$ couleurs	$\{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$
nombre d'étoiles observables la nuit	$\mathbb{N}$
cote de popularité de Nicolas	$[0,100]$
durée de vie d'une ampoule	$\mathbb{R}_+$
temps de désintégration d'un noyau radioactif	$\mathbb{R}_+$
poids d'un être humain	$[0,500]$
point d'impact au jeu de fléchettes	$\{(x,y) \in \mathbb{R}^2 : x^2+y^2 \leq 1\}$
température Celcius de la soupe du soir	$[-273,+\infty]$
position du moustique	$\mathbb{R}^3$
état d'un jeu de n cartes après battage	$\mathscr{S}_n$ (groupe symétrique)
cours d'une action en bourse sur la période $[0,T]$	$\mathscr{C}([0,T],\mathbb{R}_+)$

Le choix de Ω est assez naturel pour les phénomènes discrets, faciles à coder, mais semble plus discutable pour les situations continues. Il règne là un arbitraire typique de l'étape de modélisation, qu'on ne peut pas évacuer complètement. On dit souvent à ce propos que *tous les modèles sont faux, mais que certains sont plus utiles que d'autres.* Cet arbitraire de la modélisation n'est pas spécifique aux probabilités. Il se trouve simplement que les modélisations liées à l'analyse (analyse numérique par exemple), à l'algèbre (cryptographie par exemple), ou à la géométrie (cartographie par exemple) ne figurent pas au programme. La mathématisation du réel est une question épineuse — les nombres réels sont-ils bien réels ? L'infini existe-t-il dans la nature ?

La « mise bout à bout » de deux expériences se traduit, en termes d'univers, par un produit cartésien, comme par exemple pour le jet de deux dés, et plus généralement quand on répète une même expérience plusieurs fois.

Pile ou face : univers. On joue à pile ou face de façon répétée. On peut choisir comme univers les ensembles suivants :

— pour un lancer, $\Omega = \{0,1\}$;

— pour deux lancers consécutifs, $\Omega = \{0,1\} \times \{0,1\} = \{0,1\}^2 = \{(0,0),(0,1),(1,0),(1,1)\}$;

— pour n lancers : $\Omega = \{0,1\}^n$;

— pour une infinité de lancers de pièce de monnaie consécutifs : $\Omega = \{0,1\}^{\mathbb{N}}$.

Le dernier cas peut sembler peu naturel mais permet de traiter de façon unifiée des expériences dont la durée n'est pas fixée à l'avance (par exemple, l'attente du premier pile).

L'ensemble Ω peut être fini, infini dénombrable, ou même infini non dénombrable.

Probabilités - Préparation à l'agrégation interne de mathématiques (5/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

	Mathématiques	Français
Ensembles et événements	Ensembles	Événements
	Ω	tous les résultats possibles, événement certain
	A sous-ensemble de Ω	événement
	$\varnothing$	événement impossible
	$A \cap B$	A et B, c'est-à-dire à la fois A et B
	$A \cup B$	A ou B, c'est-à-dire au moins A ou B
	A^c (aussi noté $\overline{A}$ ou $\Omega \setminus A$))	contraire de A
Relations entre événéments	$A \cap B = \varnothing$ (A et B disjoints)	A et B sont incompatibles
	$A \subset B$	A implique B

Table 1.1 – Approche ensembliste des probabilités. On écrit *événement* ou *évènement*.

1.2 Événements, tribus

Une fois Ω choisi, on s'intéresse à l'éventuelle réalisation d'événements : le dé est-il tombé sur 6 ? La soupe est-elle gelée ? La fléchette est-elle tombée sur la zone de score maximal ? Le moustique est-il dans ma tente ? Notons que pour l'instant on ne modélise que des questions *qualitatives*, dont la réponse est oui ou non.

Ces événements s'identifient naturellement à des *sous-ensembles* de l'univers Ω : pour les deux premiers cas respectivement, $\{6\} \subset \{1,2,3,4,5,6\}$ et $]-273,0] \subset]-273,\infty[$. L'événement *impossible* s'identifie à l'ensemble $\varnothing$ tandis que l'événement *certain* s'identifie à l'ensemble Ω tout entier. Deux événements sont **incompatibles** lorsqu'ils sont disjoints en tant qu'ensembles. Les opérations ensemblistes (union, intersection, complémentaire) ont une traduction sur les phrases décrivant les événements : ces liens sont rappelés dans le tableau 1.1. Rappelons à cette occasion que si $(A_i)_{i\in I}$ est une famille d'événements et B un événement alors on a :

$$\text{Formules de Morgan} \begin{cases} \left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c \\ \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c \end{cases} \qquad \text{Distribuvité} \begin{cases} B \cap \left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} (B \cap A_i) \\ B \cup \left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} (B \cup A_i) \end{cases}$$

Pour diverses raisons[1] on se restreint souvent à considérer seulement *certaines* des parties de Ω. La collection de ces parties est donc un ensemble $\mathscr{F} \subset \mathscr{P}(\Omega)$. Il est naturel d'imposer à $\mathscr{F}$ quelques propriétés de stabilité, qui en font une *tribu* (on dit parfois également *σ-algèbre*[2]).

Définition 1.1 (Tribu)**.** *On dit qu'une collection $\mathscr{F} \subset \mathscr{P}(\Omega)$ constitue une **tribu** lorsque*

i) $\Omega \in \mathscr{F}$ *;*

ii) stabilité par complémentaire : *pour tout $A \in \mathscr{F}$ on a $A^c \in \mathscr{F}$;*

iii) stabilité par union dénombrable : *pour toute suite (A_n) d'éléments de $\mathscr{F}$ on a $\cup_n A_n \in \mathscr{F}$.*

On déduit de ces axiomes les propriétés suivantes.

Théorème 1.2 (Propriétés des tribus)**.** *Si $\mathscr{F}$ est une tribu sur Ω et $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathscr{F}$ alors…*

i) Vide : $\varnothing \in \Omega$ *;*

ii) stabilité par intersection dénombrable : *l'intersection $\cap_n A_n$ est dans $\mathscr{F}$;*

iii) stabilité par *limite inférieure* et *limite supérieure* : *les deux sous-ensembles suivants appartiennent également à $\mathscr{F}$:*

$$\underline{\lim}\, A_n = \bigcup_n \bigcap_{m\geq n} A_m = \{\omega \in \Omega \ \textit{t.q.}\ \omega \in A_n \ \textit{à partir d'un certain rang}\}$$

1. La première raison est technique (et hors programme) : pour des univers Ω non-dénombrables, on ne peut pas attribuer de manière cohérente une probabilité à tous les sous-ensembles de Ω. La deuxième raison, plus satisfaisante et également hors-programme, est que les tribus permettent de représenter une information partielle sur l'expérience aléatoire.

2. en anglais *σ-field*, d'où l'usage de la lettre $\mathscr{F}$.

Probabilités - Préparation à l'agrégation interne de mathématiques (6/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

et

$$\overline{\lim}\, A_n = \bigcap_n \bigcup_{m\geq n} A_m = \{\omega \in \Omega \text{ t.q. } \omega \in A_n \text{ pour une infinité de valeurs de } n\}.$$

De plus

$$(\underline{\lim}\, A_n)^c = \overline{\lim}\, A_n^c \quad \text{et} \quad (\overline{\lim}\, A_n)^c = \underline{\lim}\, A_n^c.$$

iv) Intersection de tribus : *Si $(\mathscr{F}_i)_{i\in I}$ est une famille quelconque de tribus sur Ω alors $\cap_{i\in I}\mathscr{F}_i$ est une tribu sur Ω.*

Exemple 1.3 (Exemples de tribus)**.**

Tribu triviale : *c'est la famille $\{\varnothing, \Omega\}$.*

Tribu grossière : *l'ensemble $\mathscr{P}(\Omega)$ de (toutes) les parties de Ω vérifie naturellement toutes les propriétés de stabilité. On parle de **tribu grossière**. Quand on étudie des phénomènes finis ou dénombrables (tirages de cartes, lancers de dés,...) on utilise sytématiquement cette tribu.*

Tribu engendrée : *si $A \subset \Omega$ alors $\{\varnothing, A, A^c, \Omega\}$ est une tribu : c'est la plus petite tribu qui contient A. Plus généralement, si $\mathscr{A}$ est une collection de parties, on peut définir (grâce à la propriété d'intersection vue au-dessus) la **tribu engendrée** par $\mathscr{A}$ comme intersection de toutes les tribus contenant $\mathscr{A}$; c'est la plus petite tribu qui contient tous les éléments de $\mathscr{A}$.*

Tribu produit : *si $\Omega = \Omega_1 \times \Omega_2$ et si $\mathscr{F}_1$ et $\mathscr{F}_2$ sont des tribus sur Ω_1 et Ω_2, on munit généralement Ω de la **tribu produit**, notée $\mathscr{F}_1 \otimes \mathscr{F}_2$, engendrée par les produits $A_1 \times A_2$ où $A_1 \in \mathscr{F}_1$ et $A_2 \in \mathscr{F}_2$.*

Pile ou face : tribus. Pour un lancer à pile ou face, $\Omega = \{0,1\}$, on choisit la tribu grossière $\mathscr{F} = \{\varnothing, \{0\}, \{1\}, \Omega\}$.

Pour n lancers, $\Omega = \{0,1\}^n = \Omega_1 \times \cdots \times \Omega_n$, on prend aussi la tribu grossière $\mathscr{F} = \mathscr{P}(\Omega)$. C'est également la tribu produit des tribus grossières sur Ω_1, ..., Ω_n.

Le programme se limite aux exemples de tribus pour Ω au plus dénombrable, pour lesquels $\mathscr{F}$ est presque toujours la tribu grossière. Si Ω est indénombrable, la tribu grossière est en général trop grosse, comme on le verra dans le chapitre suivant. Les deux exemples suivants couvrent les cas qui apparaissent « en creux » dans le programme.

Tribu borélienne : Pour le cas $\Omega = \mathbb{R}^d$, on choisit en général la **tribu borélienne** $\mathscr{B}(\mathbb{R}^d)$ qui est la tribu engendrée par les pavés de $\mathbb{R}^d$ (i.e. par les intervalles lorsque $d = 1$). On admet que la tribu borélienne sur $\mathbb{R}^d$ est la tribu produit des tribus boréliennes.

Tribu cylindrique : pour pile ou face avec une infinité de lancers, $\Omega = \{0,1\}^{\mathbb{N}}$ n'est pas dénombrable (par l'argument diagonal de Cantor). On choisit usuellement la **tribu cylindrique** engendrée par les **cylindres**

$$A_0 \times A_1 \times A_2 \times \cdots$$

où $A_0, A_1, A_2, \ldots \subset \{0,1\}$ et $A_n = \{0,1\}$ à partir d'un certain rang sur n. Plus généralement, si $\mathscr{F}'$ est une tribu sur un ensemble Ω', on munit $\Omega = (\Omega')^{\mathbb{N}}$ de la tribu engendrée par les cylindres $A_0 \times A_1 \times A_2 \times \cdots$ où $A_0, A_1, A_2, \ldots \in \mathscr{F}'$ et $A_n = \Omega'$ à partir d'un certain rang sur n.

La *fonction indicatrice* d'un événement $A \subset \Omega$ est la fonction booléenne

$$\mathbf{1}_A : \omega \in \Omega \mapsto \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

Les fonctions indicatrices servent à compter : la somme $\sum_n \mathbf{1}_{A_n}$ est par exemple égale au nombre d'événement A_n qui ont lieu, dans le sens où :

$$\sum_n \mathbf{1}_{A_n}(\omega) = \#\{n \in \mathbb{N} : \omega \in A_n\}.$$

Voici d'autres exemples :

Théorème 1.4 (Indicatrices)**.** *Si $A, B \in \mathscr{F}$ et si les $(A_n)_{n\geq 1}$ sont dans la tribu $\mathscr{F}$ alors...*

1. $\mathbf{1}_{A\cap B} = \mathbf{1}_A \mathbf{1}_B$ *et* $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A\cap B}$ *;*

Probabilités - Préparation à l'agrégation interne de mathématiques (7/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

2. $\underline{\lim}\, A_n = \left\{\sum_n \mathbf{1}_{A_n^c} < \infty\right\}$ *et* $\overline{\lim}\, A_n = \left\{\sum_n \mathbf{1}_{A_n} = \infty\right\}$ *;*

3. $\mathbf{1}_{\underline{\lim}\, A_n} = \underline{\lim}\, \mathbf{1}_{A_n}$ *et* $\mathbf{1}_{\overline{\lim}\, A_n} = \overline{\lim}\, \mathbf{1}_{A_n}$.

Notons que $\mathbf{1}_A^2 = \mathbf{1}_A$ pour tout $A \in \mathscr{F}$. Les fonctions indicatrices vont jouer un rôle important après l'introduction des notions de probabilité $\mathbb{P}$ et d'espérance $\mathbb{E}$ car $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}[A]$ pour tout $A \in \mathscr{F}$.

Probabilités - Préparation à l'agrégation interne de mathématiques (8/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Chapitre 2

Mesures de probabilité

2.1 Définitions et premières propriétés

Les mesures de probabilité permettent de comparer l'importance des événements. Donnons d'abord une définition formelle.

Définition 2.1 (Probabilité)**.** *Soit $\mathscr{F}$ une tribu sur un univers Ω. Une* ***mesure de probabilité*** *ou* ***loi de probabilité*** *sur $(\Omega, \mathscr{F})$ est une application $\mathbb{P} : \mathscr{F} \to [0,1]$ vérifiant les deux conditions suivantes :*

i) Normalisation : $\mathbb{P}[\Omega] = 1$ *;*

ii) Σ-additivité : *si $(A_n)_{n\geq 1}$ est une suite d'événements deux à deux disjoints*[1] *alors*

$$\mathbb{P}\left[\bigcup_n A_n\right] = \sum_n \mathbb{P}[A_n].$$

On dit que le triplet $(\Omega, \mathscr{F}, \mathbb{P})$ est un ***espace probabilisé****. On réservera le terme d'****événements*** *aux parties de Ω qui appartiennent à la tribu $\mathscr{F}$: ce sont les parties auxquelles on a attribué une probabilité*[2]*.*

La notion de mesure positive s'obtient en autorisant des valeurs dans $\mathbb{R}_+$ et en renonçant à la propriété de normalisation.

Théorème 2.2 (Propriétés immédiates)**.** *Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace probabilisé et A, B deux éléments de $\mathscr{F}$. Alors :*

1. $\mathbb{P}[\varnothing] = 0$ *;*
2. $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$ *;*
3. $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$ *;*
4. *si $A \subset B$ alors $\mathbb{P}[B] - \mathbb{P}[A] = \mathbb{P}[A^c \cap B]$, en particulier $\mathbb{P}[A] \leq \mathbb{P}[B]$;*
5. $\mathbb{P}[A]\mathbb{P}[B] \leq \min(\mathbb{P}[A], \mathbb{P}[B])$ *;*
6. ***probabilités totales*** *: si (A_n) est une partition de Ω alors $\sum_n \mathbb{P}[A_n] = 1$.*

Démonstration. Toutes ces propriétés se montrent relativement aisément ; il ne faut pas manquer de vérifier que tous les ensembles introduits sont bien dans la tribu $\mathscr{F}$. Pour la troisième propriété par exemple, on utilise les partitions

$$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A) \quad \text{et} \quad A = (A \setminus B) \cup (A \cap B).$$

Tous les ensembles apparaissant dans ces partitions sont dans $\mathscr{F}$, et toutes les unions sont disjointes. Il suffit alors d'appliquer la Σ-additivité. □

1. Cela signifie que $A_i \cap A_j = \varnothing$ si $i \neq j$. À ne pas confondre avec $\cap_n A_n = \varnothing$.
2. Dans les cas discrets, Ω est muni de la tribu grossière et toutes les parties de Ω sont des événements. Si Ω n'est pas dénombrable, on peut montrer qu'il n'est pas possible de construire une application $\mathbb{P}$ normalisée et Σ-additive définie sur tout $\mathscr{P}(\Omega)$; c'est la raison pour laquelle on se restreint à des tribus particulières (cylindriques, boréliennes, etc.). Certains sous-ensembles de Ω n'ont alors pas de probabilité : les choix de tribus sont faits pour que tous les ensembles « raisonnables » en aient une.

Probabilités - Préparation à l'agrégation interne de mathématiques (9/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Il est commode d'interpréter $\mathbb{P}[A]$ comme la surface (ou le cardinal) du patatoïde A dessiné sur le plan (ou sur le réseau $\mathbb{Z}^2$). La propriété 3 du théorème se généralise à une union finie : c'est la formule du *crible de Poincaré*, qui nous dit que la surface d'une union est égale à la somme des surfaces, moins la surface des intersections deux à deux, plus la surface des intersections trois à trois, etc, ce qui correspond à inclure et exclure alternativement.

Théorème 2.3 (Formule du crible, principe d'inclusion-exclusion)**.** *Pour tous* $A_1, \dots, A_r \in \mathscr{F}$

$$\mathbb{P}\left[\bigcup_{1\le i\le r} A_i\right] = \sum_{k=1}^{r} (-1)^{k+1} S_k \quad \text{où} \quad S_k = \sum_{1\le i_1<\dots<i_k\le r} \mathbb{P}\left[A_{i_1}\cap\dots\cap A_{i_r}\right].$$

Pour $r=2$, on retrouve $\mathbb{P}[A_1\cup A_2] = \mathbb{P}[A_1]+\mathbb{P}[A_2]-\mathbb{P}[A_1\cap A_2]$, et pour $r=3$,

$$\begin{aligned}
\mathbb{P}[A_1\cup A_2\cup A_3] &= \mathbb{P}[A_1]+\mathbb{P}[A_2]+\mathbb{P}[A_3] && \text{(inclusion)}\\
&\quad -\mathbb{P}[A_1\cap A_2]-\mathbb{P}[A_2\cap A_3]-\mathbb{P}[A_1\cap A_3] && \text{(exclusion)}\\
&\quad +\mathbb{P}[A_1\cap A_2\cap A_3] && \text{(inclusion).}
\end{aligned}$$

Le principe d'inclusion-exclusion est rarement utilisé pour $r>2$. Il l'est cependant dans l'étude de la fluctuation asymptotique du collectionneur de coupons (théorème C.8).

Démonstration. On procède par récurrence sur r, en observant que

$$\begin{aligned}
\mathbb{P}\left[\bigcup_{1\le i\le r+1} A_i\right] &= \mathbb{P}\left[\bigcup_{1\le i\le r} A_i\right] + \mathbb{P}[A_{r+1}] - \mathbb{P}\left[\left(\bigcup_{1\le i\le r} A_i\right)\cap A_{r+1}\right]\\
&= \mathbb{P}\left[\bigcup_{1\le i\le r} A_i\right] + \mathbb{P}[A_{r+1}] - \mathbb{P}\left[\bigcup_{1\le i\le r} (A_i\cap A_{r+1})\right]
\end{aligned}$$

ce qui permet d'utiliser l'hypothèse de récurrence (pour le premier et dernier terme). □

2.2 Équiprobabilité discrète et combinatoire

Soit $(\Omega, \mathscr{P}(\Omega), \mathbb{P})$ un espace probabilisé avec $\Omega = \{\omega_1, \omega_2, \dots \omega_n\}$ **fini**. La propriété de (Σ-)additivité montre que, si l'on connaît les n probabilités des atomes $(\mathbb{P}[\{\omega_1\}], \mathbb{P}[\{\omega_2\}], \dots, \mathbb{P}[\{\omega_n\}])$, l'application $\mathbb{P}$ est entièrement connue : la probabilité de n'importe quel ensemble s'obtient en sommant les probabilités des atomes qui le composent.

Le cas le plus simple correspond à donner à tous les atomes le même poids : on parle alors d'**équiprobabilité**. La normalisation impose alors :

$$\mathbb{P}[\{\omega\}] = \frac{1}{\mathrm{Card}(\Omega)} \qquad \text{pour tout } \omega\in\Omega,$$

et plus généralement, pour tout $A\subset\Omega$,

$$\mathbb{P}[A] = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}.$$

Il s'agit de la fameuse formule « nombre de cas favorables sur nombre de cas possibles » [3]. Elle n'est **valable qu'en cas d'équiprobabilité**, ou comme le dit Laplace quand nous sommes « également indécis sur l'existence [des cas possibles] », et réduit le calcul des probabilités à du dénombrement. Un lancer de pile ou face avec une pièce équilibrée se modélise avec la mesure de probabilité uniforme sur $\{0,1\}$, qui affecte la probabilité $\frac{1}{2}$ aux 2 atomes $\{0\}$ et $\{1\}$. Un jet de dé équilibré à six faces se modélise avec la mesure de probabilité sur $\{1,2,3,4,5,6\}$ qui affecte la probabilité $\frac{1}{6}$ aux 6 atomes $\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}$.

Équiprobabilité sur les univers infinis. Il ne peut pas y avoir d'équiprobabilité lorsque $(\Omega,\mathscr{F}) = (\mathbb{N},\mathscr{P}(\mathbb{N}))$ car la masse d'un atome serait nulle. Il n'y a donc pas de mesure de probabilité uniforme sur les ensembles infinis dénombrables. En revanche, si Ω est un pavé de $\mathbb{R}^d$ nous verrons plus loin que la mesure de Lebesgue normalisée joue le rôle de modèle d'équiprobabilité, à condition de remplacer le cardinal par le volume.

3. Dans les termes de Pierre-Simon de Laplace : « La théorie des hasards consiste à réduire tous les événements du même genre, à un certain nombre de cas également possibles, c'est-à-dire, tels que nous soyons également indécis sur leur existence ; et à déterminer le nombre de cas favorables à l'événement dont on cherche la probabilité ». *Théorie analytique des probabilités*, 1814, p. iv.

Probabilités - Préparation à l'agrégation interne de mathématiques (10/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

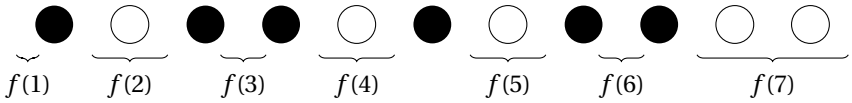


FIGURE 2.1 – Codage d'un multiensemble par une coloration.

Pour $n = 7$ et $r = 5$, on aligne $n + r - 1 = 11$ boules et on en colore $n - 1 = 6$ en noir. Le nombre de boules blanches à gauche de la première noire donne $f(1)$ (ici 0), le nombre de blanches entre les deux premières noires donne $f(2)$ (ici 1), et ainsi de suite. Ici on a donc $f(1) = 0$, $f(2) = 1$, $f(3) = 0$, $f(4) = 1$, $f(5) = 1$ et $f(6) = 0$ et $f(7) = 2$.

En pratique, les modèles équiprobables sont les plus naturels, et constituent le socle sur lequel beaucoup d'autres modèles sont construits. Certaines de ces constructions sont abordées dans la suite. Les calculs de cardinaux dans les modèles d'équiprobabilité nécessitent bien souvent des formules combinatoires. Pour $0 \le r \le n$, on notera[4] $\binom{n}{r}$ le coefficient binomial

$$\binom{n}{r} := \frac{n(n-1)\cdots(n-r+1)}{r(r-1)\cdots 1} = \frac{n!}{r!(n-r)!}.$$

Rappelons que pour a et b dans un anneau commutatif, on a la formule du binôme de Newton :

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Les « briques de base » du dénombrement sont usuellement expliquées par des tirages de boules dans des urnes. On considère donc une urne contenant n boules numérotées de 1 à n (donc distinguables[5]) et on effectue le tirage de r boules dans l'urne. Le nombre de tirages possibles diffère selon que l'on remet ou non les boules dans l'urne, et que l'on s'intéresse ou non à l'ordre de tirage. Pour un tirage...

sans remise et ordonné, on parle d'**arrangement** ; r est nécessairement inférieur à n. Il y en a $A_{n,r} := n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!} = r!\binom{n}{r}$. Il s'agit du nombre de r-uplets $(b_1, \dots, b_r)$ constitués d'éléments $b_1, \dots, b_r$ de $\{1, \dots, n\}$ deux à deux distincts, ou encore du nombre d'injections de $\{1, \dots, r\}$ dans $\{1, \dots, n\}$.

Pour $r = n$ on trouve le nombre de manières de permuter $\{1, \dots, n\}$, soit $A_{n,n} = n!$. *Exemple :* nombre de tiercés avec n chevaux au départ ($r = 3$).

sans remise et non ordonné, on parle de **combinaison** ou de sous-ensemble ; on a encore la condition $r \le n$. Il y en a $\binom{n}{r}$. Il s'agit du nombre de sous-ensembles de cardinal r de $\{1, \dots, n\}$. *Exemple :* nombre de binômes possibles dans une classe de n élèves ($r = 2$).

avec remise et ordonné, on parle de r-uplets ; il y en a n^r. Il s'agit du nombre de r-uplets $(b_1, \dots, b_r)$ constitués d'éléments $b_1, \dots, b_r$ de $\{1, \dots, n\}^r$, c'est-à-dire le nombre d'applications de $\{1, \dots, r\}$ dans $\{1, \dots, n\}$. Bien entendu, $n^r \ge A_{n,r} \ge \binom{n}{r}$. *Exemple :* nombre de mots de r lettres ($n = 26$).

avec remise et non ordonnés : on parle de multi-ensemble, ou de combinaison avec répétition. Ce cas pourtant naturel est moins souvent abordé. Il y a $\binom{n+r-1}{r} = \binom{n+r-1}{n-1}$ multi-ensembles à r éléments parmi n. Il s'agit également du nombre de manières de placer r boules indistinguables dans n urnes distinguables, ou encore du nombre d'applications $f : \{1, \dots, n\} \to \{1, \dots, r\}$ vérifiant $f(1) + \cdots + f(n) = r$ (la quantité $f(i)$ correspondant au nombre de fois où la boule i est tirée). Pour obtenir la formule, on aligne $n + r - 1$ objets (disons blancs) et on en colore $n - 1$ en noir : à chaque coloriage correspond une (et une seule) application f vérifiant $\sum_i f(i) = r$ (voir la figure 2.1). Comme il y a $\binom{n+r-1}{n-1}$ choix possibles pour le coloriage, c'est aussi le nombre de multi-ensembles. *Exemple :* nombre de possibilités au jeu des chiffres et des lettres ($n = 26$ et $r = 9$).

Exemple 2.4 (Tirage avec remise)**.** *Si on dispose d'une urne contenant n boules numérotées de 1 à n (donc distinguables), alors on modélise le tirage de r boules avec remise par la probabilité uniforme sur l'univers $\Omega = \{1, \dots, n\}^r$, dont le cardinal vaut n^r. En conséquence, lors d'un tirage avec remise de deux cartes dans un jeu de 32 cartes, la probabilité que les cartes soient de la même couleur*[6] *vaut $(2 \times 16^2)/32^2 = 1/2$, tandis que la probabilité d'obtenir 2 as vaut $4^2/32^2 = 1/64$.*

4. Nous suivons ici l'usage anglo-saxon.
5. Sans doute pour se distinguer de « navigable », « irrigable », « infatigable », etc., « distinguables » s'écrit avec un « u ».
6. Par *couleur* on entend ici de manière non standard *rouge* ou *noir* et pas *pique*, *carreau*, *trèfle*, *cœur*.

Probabilités - Préparation à l'agrégation interne de mathématiques (11/116) - Copyright © Djalil Chafaï et Pierre-André Zitt. 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Exemple 2.5 (Tirage sans remise). *Si on dispose d'une urne contenant n boules numérotées de 1 à n (donc distinguables), alors on modélise le tirage de r boules sans remise par la probabilité uniforme sur l'univers $\Omega = \{T \subset \{1,\dots,n\} : \mathrm{card}(T) = r\}$, dont le cardinal vaut $\binom{n}{r} = \frac{n!}{r!(n-r)!}$. En conséquence, lors d'un tirage sans remise de deux cartes dans un jeu de 32 cartes, la probabilité de tirer deux cartes de même couleur vaut $2\binom{16}{2}/\binom{32}{2} = 15/31$, tandis que la probabilité d'obtenir 2 as vaut $\binom{4}{2}/\binom{32}{2} = 3/(32\times 31)$. Comme on peut s'y attendre, ces deux probabilités sont plus petites que celles avec remise.*

Sondage simple : deux modélisations, une probabilité. On considère l'expérience suivante : dans une urne avec 2 boules vertes et 3 boules jaunes, on tire deux boules sans remise. On cherche la probabilité de l'événement « obtenir une boule verte et une jaune ». Athanase a lu les exemples précédents et décide de modéliser l'expérience par des combinaisons : il considère Ω_A l'ensemble des combinaisons de 2 boules parmi 5 ($\mathrm{Card}(\Omega_A) = \binom{5}{2} = 10$), muni de l'équiprobabilité $\mathbb{P}_A$. Il y a $\binom{2}{1}\binom{3}{1} = 6$ combinaisons favorables, donc

$$\mathbb{P}_A[E] = \frac{6}{10} = \frac{3}{5}.$$

Bérénice opte pour l'univers Ω_B des arrangements de deux boules parmi les 5 : $\mathrm{Card}(\Omega_B) = 20$, muni de l'équiprobabilité $\mathbb{P}_B$. Il y a 6 arrangements du type (boule verte, boule jaune) et 6 du type (boule jaune, boule verte), donc :

$$\mathbb{P}_B[E] = \frac{12}{20} = \frac{3}{5}.$$

Athanase et Bérénice ont choisi des univers et des probabilités différentes, donc des codages différents de E, mais ils obtiennent la même probabilité.

Pile ou face : nombre de piles. Pour $0 \le k \le n$, quelle est la probabilité d'obtenir k fois pile en n parties de pile ou face avec une pièce de monnaie équilibrée ? L'univers est $\Omega = \{(a_1,\dots,a_n) \in \{0,1\}^n\}$ de cardinal 2^n, où 0 code face et 1 code pile. L'événement d'intérêt est $A = \{(a_1,\dots,a_n) : a_1 + \dots + a_n = k\}$, de cardinal $\binom{n}{k}$. La pièce étant équilibrée, on choisit le modèle d'équiprobabilité, et donc $\mathbb{P}[A] = \mathrm{card}(A)/\mathrm{card}(\Omega) = \binom{n}{k}2^{-n}$. Nous verrons qu'il s'agit d'un cas particulier de la loi binomiale (taille n et paramètre $1/2$).

Il faut faire attention au choix des résultats possibles, qui doivent être équiprobables. Considérons par exemple le résultat du jet simultané de deux dés indistinguables. En ordonnant les résultats des deux dés, on peut considérer l'univers $\Omega_1 = \{(i,j) : 1 \le i \le j \le 6\}$, qui décrit bien tout les résultats possibles de l'expérience. Cependant la probabilité uniforme sur Ω_1 donne trop de poids aux « doubles ».
Le bon modèle est la probabilité uniforme sur $\{(i,j) : 1 \le i, j \le 6\}$, qui force à distinguer les dés, et donc à reformuler les événements où ils ne sont pas distingués.
Si l'on cherche par exemple la probabilité d'avoir un double 1, le premier modèle donne $1/21$, le second $1/36$; pour la probabilté d'obtenir un 1 et un 2, le premier modèle donne encore $1/21$, la bonne réponse étant $2/36 = 1/18$.

Sondage simple : modèle. Considérons une urne contenant $N = N_1 + N_2$ boules dont N_1 sont blanches et N_2 rouges. On effectue un tirage sans remise de $n \le N$ boules dans l'urne. Il y a $\binom{N}{n}$ tirages possibles. Adoptons le modèle de la probabilité uniforme sur l'ensemble de ces possibilités, c'est-à-dire sur l'univers Ω des sous-ensembles de $\{1,\dots,N\}$ à n éléments. Pour tout $0 \le k \le n$, le nombre de tirages avec k boules blanches est $\binom{N_1}{k}\binom{N_2}{n-k}$, et la probabilité de tirer k boules blanches vaut donc

$$\frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}.$$

Cette formule définit la **loi hypergéométrique** sur les sous-populations de taille n d'une population de taille N à deux types (voir le théorème 3.14). Ceci montre au passage l'identité de Vandermonde :

$$\binom{N}{n} = \sum_{k=0}^{n} \binom{N_1}{k}\binom{N_2}{n-k}.$$

Probabilités - Préparation à l'agrégation interne de mathématiques (12/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Supposons par exemple que l'on tire une main de Poker (5 cartes parmi 52). Pour calculer la probabilité d'obtenir une paire d'As, on sépare les 52 cartes en $N_1 = 4$ as et $N_2 = 48$ autres cartes :

$$\mathbb{P}\left[\text{« une paire d'As »}\right] = \frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}}.$$

Ce modèle peut se généraliser à plus de « couleurs », on parle de loi hypergéométrique multitype. Sur l'exemple du Poker, si l'on cherche la probabilité d'obtenir une paire d'As et un paire de rois, on obtient, en séparant les cartes en $N_1 = 4$ as, $N_2 = 4$ rois et $N_3 = 44$ autres cartes, l'expression :

$$\mathbb{P}\left[\text{« paire d'As et paire de rois »}\right] = \frac{\binom{4}{2}\binom{4}{2}\binom{44}{1}}{\binom{52}{5}}.$$

Autre exemple : dans une commode contenant N vêtements dont N_1 pantalons, N_2 chemises et N_3 vestes, on tire uniformément 3 vêtements. Quelle est la probabilité d'obtenir un costume complet ?

Remarque 2.6 (Équiprobabilité)**.** *Même si l'équiprobabilité est le choix le plus courant quand Ω est fini, il n'est naturellement pas le seul possible ! Ainsi, pour modéliser une pièce biaisée, qui tombe sur pile avec probabilité p, on choisit $\Omega = \{0,1\}$, $\mathscr{F} = \mathscr{P}(\Omega)$ et $\mathbb{P}$ qui donne les poids p et $1-p$ aux atomes $\{1\}$ et $\{0\}$.*

Remarque 2.7 (Philosophie)**.** *Que signifie au juste « la pièce a une chance sur deux de tomber sur pile » ? La mécanique classique affirme que le mouvement de la pièce est déterministe : si on connaît parfaitement les conditions du lancer il n'y a pas d'aléatoire. Les probabilités pallient donc ce manque d'information, et peuvent être interprétées de manières variées : classique, fréquentiste, logique, bayésienne, « propensionniste », etc. Même si nous connaissions les conditions du lancer avec une grande précision, la sensibilité aux erreurs des équations de la mécanique classique pourraient rendre le résultat du lancer imprévisible. Les probabilités pallient donc également cet excès de complexité.*

2.3 Quelques exemples remarquables

Problème des anniversaires

Calculons la probabilité p_n pour que dans une classe de n élèves, au moins deux d'entre eux soient nés le même jour. On suppose pour simplifier que les naissances sont uniformes sur les jours de l'année, et on ne tient pas compte des années bissextiles. On modélise cette expérience par la probabilité uniforme sur l'univers $\Omega = \{1,\dots,d\}^n$ où $d = 365$, dont le cardinal est d^n. Cela correspond à faire n tirages avec remise dans une urne contenant d boules numérotés de 1 à d. Si A est l'événement « deux élèves au moins sont nés le même jour » alors A^c correspond à n tirages sans remise ordonnés (arrangements !) et donc (pour $n \le d$ car $p_n = 1$ sinon)

$$p_n = \mathbb{P}[A] = 1 - \mathbb{P}\left[A^c\right] = 1 - \frac{d(d-1)\cdots(d-n+1)}{d^n} = 1 - \prod_{k=1}^{n-1}\left(1 - \frac{k}{d}\right).$$

La suite $(p_n)_{n\ge1}$ croît de manière sigmoïde, avec un phénomène de seuil (mode de la pente) vers $20-30$:

$$p_{20} \approx 0{,}41, \quad p_{30} \approx 0{,}71, \quad p_{40} \approx 0{,}89, \quad p_{50} \approx 0{,}97.$$

Ceci s'explique par le fait que le nombre de couple d'élèves est quadratique en n. Par contraste, si B est l'événement « un élève au moins est né le même jour que l'enseignant » alors on a

$$\mathbb{P}[B] = 1 - \mathbb{P}\left[B^c\right] = 1 - \left(\frac{d-1}{d}\right)^n = 1 - \left(1 - \frac{1}{d}\right)^n,$$

formule qui ne fait par apparaître de phénomène de seuil. Cette fois-ci, le nombre de couples (élève, enseignant) est linéaire en n.

Problème du chevalier de Méré

Si l'on jette 4 fois un dé à six faces, la probabilité d'obtenir un 6 vaut $1-(5/6)^4 \approx 0{,}52 > 1/2$. Si l'on jette 24 fois deux dés à six faces, la probabilité d'obtenir un double six vaut $1-(35/36)^{24} \approx 0{,}49 < 1/2$. Le

Probabilités - Préparation à l'agrégation interne de mathématiques (13/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

chevalier de Méré était un noble de la cour de Louis XIV, qui trouvait ces résultats contre-intuitifs[7], pensant que les deux probabilités s'obtenaient en multipliant la probabilité sur un lancer par le nombre de lancers (il obtenait donc 2 chances sur 3 pour les deux cas) : on retrouve le problème de formalisation évoqué plus haut. La modélisation correcte dans le premier cas est l'équiprobabilité sur l'univers $\{1,\dots,6\}^4$ et dans le second cas, l'équiprobabilité sur l'univers $(\{1,\dots,6\}^2)^{24} = \{(i,j) : 1 \le i,j \le 6\}^{24}$.

Permutations aléatoires

L'ensemble $\mathscr{S}_n$ des permutations de $\{1,\dots,n\}$ muni de la composition $\circ$ constitue ce qu'on appelle le *groupe symétrique*. Il s'agit d'un groupe fini non abélien de cardinal $n!$. La loi uniforme μ sur $\mathscr{S}_n$ affecte la probabilité $1/n!$ à chaque atome de $\mathscr{S}_n$: elle modélise par exemple l'état d'un paquet de n cartes mélangées. La loi uniforme est la seule loi sur $\mathscr{S}_n$ qui soit invariante par toute translation du groupe (à droite ou à gauche). La condition est évidemment nécessaire, et sa suffisance s'établit en observant que si μ est invariante par toute translation (disons à gauche) alors pour tous $\sigma,\sigma' \in \mathscr{S}_n$,

$$\mu(\sigma \circ \sigma') = \mu(\sigma'),$$

et donc μ affecte la même probabilité à tous les atomes de $\mathscr{S}_n$ (en posant $\sigma^{-1} = \sigma'$), elle est uniforme.

2.4 Expériences répétées

Nous avons vu plus haut que pour modéliser la réalisation consécutive de deux expériences indépendantes, représentées respectivement par $(\Omega_1,\mathscr{F}_1)$ et $(\Omega_2,\mathscr{F}_2)$, on pouvait introduire l'univers produit $\Omega = \Omega_1 \times \Omega_2$ et le munir de la tribu produit $\mathscr{F}_1 \otimes \mathscr{F}_2$. Si on dispose pour chacune expérience d'une loi de probabilité, on définit naturellement la :

Définition 2.8 (Probabilité produit)**.** *Si $(\Omega_1,\mathscr{F}_1,\mathbb{P}_1)$ et $(\Omega_2,\mathscr{F}_2,\mathbb{P}_2)$ sont deux espaces probabilisés, il existe une unique mesure de probabilité notée $\mathbb{P}_1 \otimes \mathbb{P}_2$ sur le produit $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2)$, appelée (mesure de)* ***probabilité produit****, qui vérifie*

$$(\mathbb{P}_1 \otimes \mathbb{P}_2)(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)$$

pour tout événement produit $A_1 \times A_2 \in \mathscr{F}_1 \times \mathscr{F}_2$. De même, si $(\Omega,\mathscr{F},\mathbb{P})$ est un espace probabilisé et que l'on munit $\Omega^{\mathbb{N}}$ de la tribu des cylindres $\mathscr{F}^{\otimes\infty}$, il existe une unique mesure de probabilité $\mathbb{P}^{\otimes\infty}$ sur $(\Omega^{\mathbb{N}},\mathscr{F}^{\otimes\infty})$ qui vérifie

$$\mathbb{P}^{\otimes\infty}(A_0 \times A_1 \times A_2 \times \cdots) = \mathbb{P}[A_0]\,\mathbb{P}[A_1]\,\mathbb{P}[A_2]\cdots$$

pour tout cylindre $A_0 \times A_1 \times A_2 \cdots \in \mathscr{F}^{\otimes\infty}$ (il s'agit à droite d'un produit fini car $\mathbb{P}[A_n] = 1$ à partir d'un certain rang sur n).

L'existence et l'unicité sont admises.

Remarque 2.9 (Produit et équiprobabilité)**.** *Si $(\Omega_1,\mathscr{F}_1,\mathbb{P}_1)$ et $(\Omega_2,\mathscr{F}_2,\mathbb{P}_2)$ vérifient l'équiprobabilité alors $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ vérifie aussi l'équiprobabilité. Cette propriété était déjà cachée dans les considérations sur les tirages ordonnés sans remise de r boules parmi n : l'ensemble des r-uplets est le produit $\{1,2,\dots n\}^r$. C'est également le cadre utilisé pour étudier le problème des anniversaires ou celui du chevalier de Méré (exemples de la section 2.3).*

Remarque 2.10 (Produit et indépendance)**.** *Plaçons-nous sur $\Omega^{\mathbb{N}}$ muni de la tribu cylindrique et de la probabilité produit. Soit m et n deux entiers, A_1 une partie de Ω^m et B_1 une partie de Ω^n. Alors, pour*

$$A = A_1 \times \Omega^n \times \Omega \times \Omega \times \cdots,$$
$$B = \Omega^m \times B_1 \times \Omega \times \Omega \times \cdots,$$

on peut facilement montrer que $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$. On réinterprétera cette égalité plus tard en disant que si A et B dépendent de lancers distincts, alors ils sont indépendants.

7. Blaise Pascal fait part de ce problème à Pierre de Fermat dans l'une de ses lettres. C'est à propos de ce même chevalier de Méré et de Pascal que Siméon Denis Poisson déclare, en introduction de son traité *Recherches sur la probabilité des jugements en matière criminelle et en matière civile* : « Un problème relatif aux jeux de hasard, proposé à un austère janséniste par un homme du monde, a été l'origine du calcul des probabilités ». Poisson fait ici allusion à un autre problème suggéré par Méré et discuté par Pascal et Fermat, le problème « des points », ou de la partie interrompue : comment répartir équitablement les mises dans un jeu de hasard si l'on doit s'arrêter avant la fin ?

Probabilités - Préparation à l'agrégation interne de mathématiques (14/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Si l'ensemble Ω est infini, et en particulier dans le cas d'une expérience répétée indéfiniment, les propriétés suivantes sont utiles.

Théorème 2.11 (Suites d'événements, « continuité » des probabilités)**.** *Pour toute suite* (A_n) *d'événements sur* $(\Omega, \mathscr{F}, \mathbb{P})$,

1. *si* (A_n) *est croissante (pour l'inclusion) alors* $\mathbb{P}[A_n] \nearrow \mathbb{P}[\bigcup_n A_n]$ *;*
2. *si* (A_n) *est décroissante (pour l'inclusion) alors* $\mathbb{P}[A_n] \searrow \mathbb{P}[\bigcap_n A_n]$ *;*
3. $\mathbb{P}[\bigcup_n A_n] \leq \sum_n \mathbb{P}[A_n]$ *;*
4. *si* $\mathbb{P}[A_n] = 0$ *pour tout* n *alors* $\mathbb{P}[\bigcup_n A_n] = 0$ *;*
5. *si* $\mathbb{P}[A_n] = 1$ *pour tout* n *alors* $\mathbb{P}[\bigcap_n A_n] = 1$.

Les deux premières propriétés traduisent une sorte de continuité de l'application $\mathbb{P}$ vis-à-vis des convergences monotones.

Démonstration. Pour la première propriété, les $B_n = A_n \setminus A_{n-1}$ sont deux à deux disjoints, d'où $\mathbb{P}[\bigcup_n A_n] = \mathbb{P}[\bigcup_n B_n] = \sum_n \mathbb{P}[B_n] = \lim_n \sum_{m \leq n} \mathbb{P}[B_m] = \lim_n \mathbb{P}[A_n]$. La deuxième propriété s'en déduit par passage au complémentaire. On vérifie la troisième propriété d'abord pour les unions finies, puis on passe à la limite en utilisant le premier point. Les deux dernières propriétés s'ensuivent. □

Pile ou face : suites d'événements et limites. Illustrons quelques-unes de ces propriétés dans le cas du jeu de pile ou face infini. On considère $\Omega_0 = \{0, 1\}$ muni de la tribu grossière $\mathscr{F}_0$ et de la loi de Bernoulli $\mathbb{P}_0$ qui donne le poids $1/2$ aux singletons $\{0\}$ et $\{1\}$. On pose $\Omega = \Omega_0^{\mathbb{N}}$ que l'on munit de la tribu cylindrique $\mathscr{F}_0^{\otimes \mathbb{N}}$ et de la probabilité produit $(\mathbb{P}_0)^{\otimes \mathbb{N}}$.

Si $A_n =$ « les n premiers lancers donnent pile », la suite A_n est décroissante pour l'inclusion. On en déduit que :

$$\mathbb{P}\left[\bigcap_n A_n\right] = \mathbb{P}\left[\text{« tous les lancers donnent pile »}\right] = \lim \mathbb{P}[A_n] = \lim(1/2^n) = 0.$$

Par conséquent, la mesure produit $\mathbb{P}^{\otimes \mathbb{N}}$ donne un poids nul au singleton $\{(1, 1, \ldots 1, \ldots)\} \in \{0, 1\}^{\mathbb{N}}$.

Plus généralement, fixons k et posons $B_n =$ « il y a au plus k faces sur les n premiers lancers ». La suite B_n décroît pour l'inclusion. Si n est de la forme $2km$, B_n est inclus dans l'intersection de m événements

$$\text{« au plus } k \text{ faces sur les lancers 1 à } 2k \text{ »} \cap \text{« au plus } k \text{ faces sur les lancers } (2k+1) \text{ à } 4k \text{ »}$$
$$\cdots \cap \text{« au moins } k \text{ faces sur les lancers } 2k(m-1)+1 \text{ à } 2km \text{ ».}$$

Chacun des m événements a une probabilité $1/2$ (par symétrie) ; grâce à la remarque 2.10 on en déduit

$$\mathbb{P}[B_{2km}] \leq 1/2^m.$$

Par monotonie on tire $\mathbb{P}[B_n] \searrow 0$. La propriété 2 du théorème implique alors :

$$\mathbb{P}\left[\bigcap B_n\right] = \mathbb{P}\left[\text{« il y a au plus } k \text{ faces en tout »}\right] = 0,$$

autrement dit

$$\mathbb{P}\left[\text{« il y a au moins } k+1 \text{ faces »}\right] = 1.$$

Enfin, si l'on appelle ce dernier événement C_k, on déduit de la propriété 5 :

$$\mathbb{P}\left[\text{« il y a une infinité de faces »}\right] = \mathbb{P}\left[\bigcap_k C_k\right] = 1.$$

On verra plus loin une autre preuve de ces résultats.

2.5 Conditionnement

Intuitivement, une mesure de probabilité permet de quantifier le manque d'information en affectant un poids aux issues possibles. Si l'on dispose d'information supplémentaire, les poids changent. Plaçons nous

Probabilités - Préparation à l'agrégation interne de mathématiques (15/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

d'abord dans un modèle équiprobable fini $(\Omega, \mathscr{P}(\Omega), \mathbb{P})$, par exemple le jet d'un dé équilibré. La probabilité (« inconditionnelle ») d'un événement A (par exemple « obtenir 4 ou plus ») est donnée par

$$\mathbb{P}[A] = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} = \frac{1}{2}.$$

Si on dispose de l'information supplémentaire qu'un événement B a eu lieu (par exemple $B =$ « le résultat est pair » $= \{2,4,6\}$), les « cas possibles » changent (Ω est « remplacé » par B) et les cas favorables sont restreints (seuls restent les cas qui appartiennent à A et à B). On obtient donc la **probabilité conditionnelle** :

$$\mathbb{P}[A|B] = \frac{\mathrm{Card}\, A \cap B}{\mathrm{Card}\, B} = \frac{\mathrm{Card}\, A / \mathrm{Card}\, \Omega}{\mathrm{Card}(A \cap B) / \mathrm{Card}\, \Omega} = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Cette dernière formule ne fait pas appel au choix spécifique de l'équiprobabilité ; on l'utilise comme définition générale.

Définition 2.12 (Probabilité conditionnelle)**.** *Si $A, B \in \mathscr{F}$ avec $\mathbb{P}[B] > 0$ alors la **probabilité conditionnelle de A sachant B** est la quantité (parfois notée $\mathbb{P}_B(A)$) suivante :*

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

Théorème 2.13 (Propriétés importantes)**.**
$\mathbb{P}[\cdot|B]$ est une probabilité. *Si $\mathbb{P}[B] > 0$ alors l'application*

$$\begin{aligned} \mathbb{P}[\cdot|B] : \mathscr{F} &\to [0,1], \\ A &\mapsto \mathbb{P}[A|B] \end{aligned}$$

est une mesure de probabilité sur $(\Omega, \mathscr{F})$ appelée probabilité conditionnelle. *Elle vérifie donc* toutes les propriétés des probabilités *énoncées dans le théorème 2.2 ; en particulier $\mathbb{P}[\Omega|B] = 1$, $\mathbb{P}[A^c|B] = 1 - \mathbb{P}[A|B]$, et pour toute suite (A_n) d'événements disjoints,*

$$\mathbb{P}\left[\bigcup_n A_n \middle| B\right] = \sum_n \mathbb{P}[A_n|B].$$

Probabilités totales, 2e version. *Si $0 < \mathbb{P}[B] < 1$ alors*

$$\mathbb{P}[A] = \mathbb{P}[A|B]\,\mathbb{P}[B] + \mathbb{P}\left[A \middle| B^c\right]\mathbb{P}\left[B^c\right]$$

Plus généralement, si $\Omega = \bigcup_n B_n$ est une partition de Ω avec $\mathbb{P}[B_n] > 0$ alors

$$\mathbb{P}[A] = \sum_n \mathbb{P}[A|B_n]\,\mathbb{P}[B_n].$$

Si $A \cap B = \varnothing$, $\mathbb{P}[A|B] = 0$. La probabilité conditionnelle $\mathbb{P}[\cdot|B]$ est donc en quelque sorte « portée par B ». Plus précisément, on peut montrer que $\mathbb{P}[\cdot|B]$ est une mesure de probabilité sur $(B, \mathscr{F}_B)$ où $\mathscr{F}_B = \{C \cap B : B \in \mathscr{F}\}$ est la tribu trace de $\mathscr{F}$ sur B. Notons que $\mathbb{P}[\cdot|\Omega] = \mathbb{P}$.

Si A est fixé, l'application $\mathbb{P}[A|\cdot] : B \mapsto \mathbb{P}[A|B]$ n'est *pas* une probabilité. L'égalité $\mathbb{P}[A|B^c] = 1 - \mathbb{P}[A|B]$ n'a par exemple aucune raison d'être vraie en général.

En pratique, le conditionnement est souvent utilisé quand l'expérience étudiée est en plusieurs étapes.

Exemple 2.14 (Test de dépistage de maladie)**.** *Des laboratoires pharmaceutiques ont mis au point un test médical pour dépister une maladie. Les experts pensent qu'une personne sur mille est malade dans la population. De plus, des expériences ont montré que le test déclare positifs* 99% *des malades qu'on lui soumet, et qu'il déclare malades* 2% *des personnes saines qu'on lui soumet. On choisit une personne au hasard (première étape : elle peut être saine ou malade), et on la soumet au test (deuxième étape : le test peut être positif ou négatif). Si on définit les événements*

$$A = \text{« le test médical est positif »} \quad \text{et} \quad B = \text{« la personne est malade »}$$

alors les données se traduisent par $\mathbb{P}[B] = 1/1000$, $\mathbb{P}[A|B] = 99/100$ et $\mathbb{P}[A|B^c] = 2/100$. La probablité (« totale ») de l'événement A se calcule alors en écrivant :

$$\begin{aligned} \mathbb{P}[A] &= \mathbb{P}[A|B]\,\mathbb{P}[B] + \mathbb{P}\left[A|B^c\right]\mathbb{P}\left[B^c\right] \\ &= (99/100)(1/1000) + (2/100)(999/1000) \\ &\approx 0.02. \end{aligned}$$

On a conditionné par le résultat de la première étape pour calculer la probabilité de l'événement complet.

Probabilités - Préparation à l'agrégation interne de mathématiques (16/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Si la première étape est vue comme une « cause » et la seconde comme une « conséquence », on peut essayer de retrouver la cause en partant de la conséquence : c'est l'objet de la formule de Bayes, ou formule de « probabilités des causes », qui permet d'inverser l'ordre d'un conditionnement.

Théorème 2.15 (Formule de Bayes)**.** *Si* $\mathbb{P}[A] > 0$ *et* $0 < \mathbb{P}[B] < 1$ *alors*

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]} = \frac{\mathbb{P}[A|B]\,\mathbb{P}[B]}{\mathbb{P}[A \cap B] + \mathbb{P}[A \cap B^c]} = \frac{\mathbb{P}[A|B]\,\mathbb{P}[B]}{\mathbb{P}[A|B]\,\mathbb{P}[B] + \mathbb{P}[A|B^c]\,\mathbb{P}[B^c]}.$$

La formule de Bayes permet de calculer $\mathbb{P}[B|A]$ à partir des données $\mathbb{P}[A|B], \mathbb{P}[A|B^c], \mathbb{P}[B]$. Notons la formule suivante parfois utile en pratique pour les calculs numériques :

$$\mathbb{P}[B|A] = \frac{1}{1 + \frac{\mathbb{P}[A|B^c]}{\mathbb{P}[A|B]} \frac{\mathbb{P}[B^c]}{\mathbb{P}[B]}}.$$

Exemple 2.16 (Test de dépistage, suite)**.** *La question naturelle dans le cadre du test de dépistage modélisé ci-dessus est de connaître la probabilité d'être malade sachant que le test est positif. La formule donne :*

$$\mathbb{P}[B|A] = \frac{1}{1 + \frac{\mathbb{P}[A|B^c]}{\mathbb{P}[A|B]} \frac{\mathbb{P}[B^c]}{\mathbb{P}[B]}} = \frac{1}{1 + \frac{2 \times 999}{99}} \approx \frac{1}{20} = 0,05.$$

Le test n'est vraiment pas efficace de ce point de vue ! Le paradoxe vient du fait que la maladie est si rare qu'il y a plus de « faux positifs » (2% *des* 99.9% *de personnes saines, soit environ* 2% *de la population) que de vrais positifs (*99% *des* 0.1% *de personnes malades, soit environ* 0.1% *de la population). Cet exemple est l'occasion de rappeler que du point de vue statistique, un test comporte deux types d'erreur (faux positifs et faux négatifs) qui ne jouent pas un rôle symétrique du point de vue du risque modélisé.*

2.6 Indépendance

Intuitivement, deux expériences sont indépendantes lorsqu'elles ne sont pas reliées par une relation causale. Par exemple, le jet de deux dés à six faces équilibrés, est modélisé par l'univers $\Omega \times \Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$. L'absence de relation causale entre les deux dés suggère de considérer la mesure de probabilité uniforme sur $(\Omega \times \Omega, \mathscr{P}(\Omega \times \Omega))$, qui se trouve être la mesure de probabilité produit des mesures de probabilités uniformes. En particulier, si $A \times \Omega$ (respectivement $\Omega \times B$) est un événement qui ne concerne que le résultat du premier (respectivement second) jet de dé, alors [8]

$$\begin{aligned}
\mathbb{P}[(A \times \Omega) \cap (\Omega \times B)] &= \frac{\mathrm{Card}((A \times \Omega) \cap (\Omega \times B))}{\mathrm{Card}(\Omega \times \Omega)} \\
&= \frac{\mathrm{Card}(A)\mathrm{Card}(B)}{\mathrm{Card}(\Omega)\mathrm{Card}(\Omega)} \\
&= \mathbb{P}[A]\,\mathbb{P}[B] \\
&= \mathbb{P}[A \times \Omega]\,\mathbb{P}[\Omega \times B]
\end{aligned}$$

(par abus de notation $\mathbb{P}$ est utilisé à la fois pour la mesure de probabilité uniforme sur Ω et sur $\Omega \times \Omega$).

Plus généralement, en revenant à l'interprétation intuitive des probabilités conditionnelles, lorsque deux événements A et B ne sont pas reliés causalement, l'information « B s'est produit » ne devrait pas modifier notre estimation de la probabilité de A, et on devrait avoir : $\mathbb{P}[A|B] = \mathbb{P}[A]$, ou encore

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B].$$

Cette formule remarquable conduit à la définition générale suivante de l'indépendance, bien au-delà du cas de l'équiprobabilité.

Définition 2.17 (Indépendance de deux événements)**.** *Soit* $(\Omega, \mathscr{F}, \mathbb{P})$ *un espace probabilisé. On dit que les* ***deux événements*** $A, B \in \mathscr{F}$ *sont* ***indépendants*** *lorsque*

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B].$$

Exemple 2.18 (Espaces produits)**.** *Remarquons une fois de plus que le choix de la probabilité produit pour représenter une suite d'expériences (remarque 2.10) est intrinsèquement lié à l'indépendance entre les expériences.*

8. Sur le plan cartésien $\Omega \times \Omega$, l'événement $A \times \Omega$ est une bande verticale et $\Omega \times B$ une bande horizontale.

Probabilités - Préparation à l'agrégation interne de mathématiques (17/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Exemple 2.19. *Considérons le lancer d'un dé équilibré, modélisé par* $\Omega = \{1,2,3,4,5,6\}$ *muni de la tribu grossière et de la mesure de probabilité uniforme. Si* $A =$ *« le résultat est* ≤ 4 *» et* $B =$ *« le résultat est pair » alors* A *et* B *sont indépendants : en utilisant les cardinaux :* $\mathbb{P}[A \cap B] = \frac{2}{6} = \frac{4}{6}\frac{3}{6} = \mathbb{P}[A]\mathbb{P}[B]$.

Remarque 2.20. *Si* A *et* B *sont indépendants alors* A^c *et* B^c *le sont car*

$$\mathbb{P}\left[A^c \cap B^c\right] = 1 - \mathbb{P}[A \cup B] = 1 - \mathbb{P}[A] - \mathbb{P}[B] + \mathbb{P}[A]\mathbb{P}[B] = (1 - \mathbb{P}[A])(1 - \mathbb{P}[B]).$$

L'indépendance n'est *pas transitive* ! Si A est indépendant de B, et B est indépendant de C, alors A n'est pas nécessairement indépendant de C. Contre-exemple : $C = A$ si $\mathbb{P}[A] \in]0,1[$, car si A est indépendant de lui-même, alors $\mathbb{P}[A] = \mathbb{P}[A \cap A] = \mathbb{P}[A]^2$, donc $\mathbb{P}[A] \in \{0,1\}$.

L'indépendance n'est *pas stable par union* ! Si A et B sont indépendants de C, $A \cup B$ n'est pas nécessairement indépendant de C.
Contre-exemple : pour deux tirages de pile ou face on considère

— $A =$ « le premier lancer donne face » » $= \{(0,0),(0,1)\}$
— $B =$ « le second lancer donne pile » $= \{(1,1),(0,1)\}$
— $C =$ « les deux premiers lancers donnent le même résultat » $= \{(0,0),(1,1)\}$.

Alors

$$A \cap C = \{(0,0)\},\ B \cap C = \{(1,1)\},\ A \cup B = \{(0,0),(0,1),(1,1)\},\ (A \cup B) \cap C = \{(0,0),(1,1)\}$$

et ainsi

$$\mathbb{P}[A \cap C] = \mathbb{P}[B \cap C] = \frac{1}{4} = \frac{1}{2}\frac{1}{2} = \mathbb{P}[A]\mathbb{P}[C] = \mathbb{P}[B]\mathbb{P}[C],$$

donc A et B sont bien tous deux indépendants de C. En revanche

$$\mathbb{P}[(A \cup B) \cap C] = \frac{1}{2} \neq \frac{3}{8} = \mathbb{P}[A \cup B]\mathbb{P}[C].$$

Pour obtenir l'indépendance de $A \cup B$ et de C une notion plus forte est donc nécessaire.

Définition 2.21 (Indépendance d'une famille d'événements)**.** *Si* $(A_i)_{i \in I}$ *est une famille d'événements, on dit qu'ils sont* ***indépendants***[9] *lorsque pour toute partie* ***finie*** $J \subset I$,

$$\mathbb{P}\left[\bigcap_{j \in J} A_j\right] = \prod_{j \in J} \mathbb{P}[A_j].$$

De même, si $(\mathscr{F}_i)_{i \in I}$ *est une* ***famille de tribus*** *sur* Ω, *on dit qu'elles sont indépendantes lorsque* $(A_i)_{i \in I}$ *sont indépendants dès que* $A_i \in \mathscr{F}_i$ *pour tout* $i \in I$.

Remarque 2.22 (Lien avec l'indépendance deux à deux)**.** *L'indépendance d'une famille implique clairement l'indépendance des événements pris deux à deux. La réciproque est fausse en général. Les événements* A, B *et* C *de l'exemple précédent sont par exemple indépendants deux à deux, mais pas mutuellement indépendants.*

Théorème 2.23 (Indépendance)**.** *Si* $(A_i)_{i \in I}$ *est une famille d'évenements indépendants, alors :*

1. *les* $(A_i^c)_{i \in I}$ *sont indépendants ;*
2. *si* J *et* K *sont deux sous-ensembles de* I *disjoints, et si l'on note* $\mathscr{F}_J$ *(resp.* $\mathscr{F}_K$*) la tribu engendrée par les* $(A_i)_{i \in J}$ *(resp. les* $(A_i)_{i \in K}$*), alors les tribus* $\mathscr{F}_J$ *et* $\mathscr{F}_K$ *sont indépendantes.*

Démonstration. Pour la première propriété, on raisonne par récurrence à partir de

$$\begin{aligned}\mathbb{P}\left[A^c \cap B^c\right] &= 1 - \mathbb{P}[A \cup B] = 1 - \mathbb{P}[A] - \mathbb{P}[B] + \mathbb{P}[A \cap B] \\ &= 1 - \mathbb{P}[A] - \mathbb{P}[B] + \mathbb{P}[A]\mathbb{P}[B] = (1 - \mathbb{P}[A])(1 - \mathbb{P}[B]) \\ &= \mathbb{P}\left[A^c\right]\mathbb{P}\left[B^c\right].\end{aligned}$$

La preuve de la deuxième propriété est admise. □

9. Dans la définition on écrit souvent « mutuellement indépendants », mais à l'usage on parle simplement d'indépendance.

Probabilités - Préparation à l'agrégation interne de mathématiques (18/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Pile ou face : indépendance. La deuxième propriété écrite en termes de tribus peut sembler complexe. Sur l'exemple du jeu de pile ou face répété, elle permet de justifier des propriétés très intuitives. Pour $A_i =$ « le i^{e} lancer donne pile », les $(A_i)_{i\in\mathbb{N}}$ sont indépendants. Si A est par exemple l'événement « les 10 lancers pairs tombent tous sur pile » et B est défini par « sur les lancers 1, 3 et 7, il y a au moins deux piles », alors A et B sont indépendants : en effet, en posant J (resp. K) l'ensemble des entiers pairs (resp. impairs), A est dans $\mathscr{F}_J$ et B dans $\mathscr{F}_K$.

De façon encore plus élémentaire, cette propriété implique que si A, B, C sont indépendants, alors $A\cup B$ est indépendant de C.

Ne pas confondre « A et B **incompatibles** » avec « A et B **indépendants** ». La première notion est purement ensembliste et s'interprète comme « A et B ne peuvent pas arriver simultanément ». La seconde nécessite une mesure de probabilité, intuitivement elle signifie que « savoir si A a eu lieu ne donne aucune information sur B ». Notons que si A et B sont à la fois indépendants et incompatibles alors $\mathbb{P}[A]\mathbb{P}[B] = \mathbb{P}[A\cap B] = \mathbb{P}[\varnothing] = 0$ et donc $\mathbb{P}[A] = 0$ ou $\mathbb{P}[B] = 0$. Notons enfin que si A et B sont indépendants alors A^c et B^c le sont, tandis que si A et B sont incompatibles, alors A^c et B^c ne le sont que s'ils forment une partition de Ω.

Lemme 2.24 (Borel-Cantelli)**.** *Soit (A_n) une suite d'événements dans $(\Omega,\mathscr{F},\mathbb{P})$.*

1. *Cantelli : si $\sum_n \mathbb{P}[A_n] < \infty$ alors $\mathbb{P}\left[\overline{\lim} A_n\right] = 0$;*
2. *Borel (loi du zéro-un) : si les (A_n) sont **indépendants** alors*

$$\mathbb{P}\left[\overline{\lim} A_n\right] = \begin{cases} 0 & ssi \quad \sum_n \mathbb{P}[A_n] < \infty \\ 1 & ssi \quad \sum_n \mathbb{P}[A_n] = \infty. \end{cases}$$

La seconde partie du lemme contient une réciproque à la première partie.

Démonstration. Pour la partie Cantelli : la suite (B_n) définie par $B_n = \cup_{m\geq n} A_m$ est décroissante. Par conséquent, si $\sum_n \mathbb{P}[A_n] < \infty$ alors

$$\mathbb{P}\left[\overline{\lim} A_n\right] = \mathbb{P}[\cap_n B_n] = \lim_n \mathbb{P}[B_n] \leq \lim_n \sum_{m\geq n} \mathbb{P}[A_m] = 0.$$

Pour la partie Borel : la première partie réduit le problème à établir que si $\sum_n \mathbb{P}[A_n] = \infty$ alors $\mathbb{P}\left[\overline{\lim} A_n\right] = 1$, ou encore $\mathbb{P}\left[\underline{\lim} A_n^c\right] = 0$. Si l'on pose $B_n = \cap_{m\geq n} A_m^c$ alors (B_n) est croissante pour l'inclusion, ce qui entraîne $\mathbb{P}\left[\underline{\lim} A_n^c\right] = \lim_n \mathbb{P}[B_n]$ par le théorème 2.11. À présent, l'indépendance des (A_n^c), l'inégalité $1 - x \leq e^{-x}$ valable pour $x \in \mathbb{R}$, et $\sum_n \mathbb{P}[A_n] = \infty$, donnent pour tout n :

$$0 \leq \mathbb{P}[B_n] = \prod_{m\geq n} \mathbb{P}\left[A_n^c\right] \leq \prod_{m\geq n} e^{-\mathbb{P}[A_m]} = \exp\left(-\sum_{m\geq n} \mathbb{P}[A_m]\right) = 0. \qquad \square$$

L'intérêt de ce lemme peut ne pas apparaître au premier abord. On verra plus loin que la première partie permet par exemple d'établir une version forte de la loi des grands nombres (Théorème 5.5). Donnons dès à présent une illustration directe de la première partie du lemme.

Exemple 2.25 (Application de la première partie du lemme de Borel-Cantelli)**.** *On modélise l'évolution d'une population hermaphrodite. Pour simplifier, on suppose que les individus sont éternels. À l'instant 0, la population compte 2 individus. Pour tout $n \geq 1$, à l'instant n, la population compte $n+2$ individus, et un couple d'individus choisi au hasard fait un enfant, ce qui augmente la population de 1 individu. Soit A_n l'événement « les deux premiers individus font un enfant à l'instant n ». On a $\mathbb{P}[A_1] = 1$ mais $\mathbb{P}[A_n] \sim c/n^2$ quand $n \to \infty$, et donc $\sum_n \mathbb{P}[A_n] < \infty$. Grâce à la première partie du lemme de Borel-Cantelli, presque sûrement, le couple formé par les deux premiers individus ne fait plus d'enfant à partir d'un certain temps.*

Pour illustrer la seconde partie du lemme de Borel-Cantelli, nous revenons au jeu de pile ou face.

Pile ou face : singes dactylographes. Pour illustrer la seconde partie du lemme, revenons à l'exemple du pile ou face infini. Si l'on définit $A_n =$ « le n^{e} lancer donne face », les A_n sont indépendants, et

$$\sum \mathbb{P}[A_n] = \sum_n (1/2) = \infty,$$

Probabilités - Préparation à l'agrégation interne de mathématiques (19/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

donc la deuxième partie du lemme de Borel–Cantelli implique

$$\mathbb{P}\left[\overline{\lim} A_n\right] = 1.$$

Or $\overline{\lim} A_n =$ « une infinité de A_n se réalisent », autrement dit

$$\mathbb{P}\left[\text{« il y a une infinité de faces »}\right] = 1.$$

Cet argument peut s'adapter pour montrer plus généralement que n'importe quelle séquence finie de symboles se répète une infinité de fois (et en particulier apparaît au moins une fois) : la suite de 0 et de 1 contient donc presque-sûrement le codage en binaire des œuvres complètes de Shakespeare [10].

La seconde partie du lemme affirme qu'un certain événement ($\overline{\lim} A_n$) ne peut être que de probabilité 0 ou 1. Ce théorème admet la généralisation suivante.

Théorème 2.26 (Loi du zéro-un de Kolmogorov)**.** *Soit* $(\Omega, \mathscr{F}, \mathbb{P})$ *un espace probabilisé et* $(\mathscr{F}_n)_{n\geq 1}$ *une suite de tribus sur* Ω*, incluses dans* $\mathscr{F}$ *et* ***indépendantes****. Soit* $\mathscr{G}_n$ *la tribu engendrée par* $\cup_{m\geq n}\mathscr{F}_m$*. Alors* $\mathbb{P}[A] \in \{0,1\}$ *pour tout* A *dans la* ***tribu terminale*** $\mathscr{G}_\infty = \cap_n \mathscr{G}_n$*.*

En quelque sorte, on a $\mathscr{G}_\infty = \overline{\lim}\,\mathscr{F}_n$. Notons que si $A_n \in \mathscr{F}_n$ pour tout n alors $\underline{\lim}_n A_n$ et $\overline{\lim}_n A_n$ sont dans la tribu terminale et donc sont de probabilité 0 ou 1, ce qui montre que le second lemme de Borel-Cantelli découle de la loi du zéro-un de Kolmogorov.

Démonstration. Les tribus $\mathscr{F}_n$ et $\mathscr{G}_{n+1}$ sont indépendantes, et donc la tribu engendrée par $\cup_n \mathscr{F}_n$ est indépendante de la tribu $\mathscr{G}_\infty$. Comme $\mathscr{G}_\infty \subset \cup_n \mathscr{F}_n$ on en déduit que tout événement de $\mathscr{G}_\infty$ est indépendant de lui même, et est donc de probabilité 0 ou 1. □

2.7 Équiprobabilité continue : la mesure de Lebesgue

Toute cette section est *stricto sensu* hors programme ; elle permet cependant de mettre en perspective les notions de variables continues (qui elles sont au programme !).

Les axiomes définissant les mesures positives (positivité, Σ-additivité) paraissent convenir très bien pour parler de surface ou de volume : la surface d'un ensemble en deux morceaux devrait bien être la somme des surfaces des morceaux. Cette intuition peut être rendue rigoureuse et on a le résultat suivant.

Théorème 2.27 (Mesure de Lebesgue)**.** *Soit* d *un entier et* $\mathscr{B}(\mathbb{R}^d)$ *la tribu borélienne sur* $\mathbb{R}^d$*. Il existe une unique mesure positive* λ *sur* $\mathscr{B}(\mathbb{R}^d)$*, appelée mesure de Lebesgue, telle que, pour tout pavé* $E = [a_1, b_1] \times [a_2, b_2] \cdots [a_d, b_d]$*,*

$$\lambda(E) = \prod_{i=1}^{d} (b_i - a_i).$$

10. Ce théorème est souvent expliqué en disant qu'un singe éternel qui taperait éternellement sur une machine à écrire finirait par reproduire les œuvres complètes de Shakespeare. L'image est parfois attribuée à Borel, qui l'utilise dans sa note *La mécanique statistique et l'irréversibilité* publiée au Journal de Physique en 1913. « On a souvent cherché à donner une idée de l'extrême rareté des cas exceptionnels, rareté qui dépasse tout ce que notre imagination peut concevoir ; voici une comparaison qui me parait particulièrement frappante. Concevons qu'on ait dressé un million de singes à frapper au hasard sur les touches d'une machine à écrire et que, sous la surveillance de contremaîtres illettrés, ces singes dactylographes travaillent avec ardeur dix heures par jour avec un million de machines à écrire de types variés. Les contremaîtres illettrés rassembleraient les feuilles noircies et les relieraient en volumes. Et, au bout d'un an, ces volumes se trouveraient renfermer la copie exacte des livres de toute nature et de toutes langues conservés dans les plus riches bibliothèques du monde. Telle est la probabilité pour qu'il se produise, pendant un instant très court, dans un espace de quelque étendue, un écart notable de ce que la Mécanique statistique considère comme le phénomène le plus probable. Supposer que cet écart ainsi produit subsistera pendant quelques secondes revient à admettre que, pendant plusieurs années, notre armée de singes dactylographes, travaillant toujours dans les mêmes conditions, fournira chaque jour la copie exacte de tous les imprimés, livres et journaux, qui paraîtront la semaine suivante sur toute la surface du globe. Il est plus simple de dire que ces écarts improbables sont purement impossibles. »

Notons que Borel ne cherche pas ici à illustrer le théorème, mais bien à souligner que dans son application pratique (ici à la mécanique statistique), si l'on se fixe un temps d'attente raisonnable, l'événement rare ne se produira pas !

Probabilités - Préparation à l'agrégation interne de mathématiques (20/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Le carré unité figuré en clair est l'ensemble Ω ; les événements d'intérêt sont figurés en foncé. À gauche, l'événement « Athanase arrive avant Bérénice » est codé par le triangle $x < y$ d'aire 1/2. À droite, l'événement « A. et B. ne s'attendent pas plus de 10 min. », d'aire 11/36. Comme l'aire totale du carré vaut 1, ce sont aussi les probabilités des événements correspondants.

FIGURE 2.2 – Rendez-vous d'Athanase et Bérénice

Si Ω est un sous-ensemble raisonnable [11] de $\mathbb{R}^d$, posons $\mathscr{F}$ la collection des $\{A \cap \Omega; A \in \mathscr{B}(\mathbb{R}^d)\}$. Alors $\mathscr{F}$ est une tribu [12] et l'application

$$\begin{aligned} \mathbb{P} : \mathscr{F} &\to [0,1] \\ A &\mapsto \frac{\lambda(A)}{\lambda(\Omega)} \end{aligned}$$

est une probabilité sur $(\Omega, \mathscr{F})$: la normalisation est évidente et la Σ-additivité provient de celle de λ. Cette probabilité représente le choix d'un point uniformément dans l'ensemble Ω ; la probabilité que ce point tombe dans une région A est *proportionnelle au volume de* A (en dimension 2, à l'aire de A ; en dimension 1 à sa longueur).

Fléchettes : le modèle. On joue aux fléchettes sur une cible ronde modélisée par le disque unité dans le plan : $\Omega = D(0,1) := \{(x,y) | x^2 + y^2 \leq 1.\}$.

On suppose que le point d'impact est réparti uniformément sur la cible : la probabilité de tomber dans une région est proportionnelle à son aire.

Si on ne marque des points que quand la fléchette tombe à une distance 1/2 du centre, on a par exemple :

$$\mathbb{P}\left[\text{« marquer des points »}\right] = \frac{\lambda(D(0,1/2))}{\lambda(D(0,1))} = \frac{\pi(1/2)^2}{\pi 1^2} = \frac{1}{4}.$$

Exemple 2.28 (Corrosion de carrosserie)**.** *Pour étudier la corrosion d'une carrosserie, on peut chercher à modéliser le point d'impact d'une goutte de pluie. Ce point sera* a priori *uniforme sur la surface exposée.*

De la même manière que les calculs dans le cas uniforme discret se ramènent à du dénombrement, le cas uniforme continu se réduit à des calculs de longueurs, d'aires ou de volumes.

Dans l'exemple du jeu de fléchettes, cherchons la probabilité de toucher exactement le centre $O = (0,0)$. Pour tout $r > 0$, $\{O\}$ est inclus dans le disque $D(O,r)$, donc $\mathbb{P}[\{O\}] \leq r^2$. Par conséquent la probabilité de toucher le centre est *nulle*. Ceci est valable pour n'importe quel autre point de la cible.

De façon encore plus surprenante, si A est une partie dénombrable de Ω (par exemple l'ensemble des points du disque à coordonnées rationnelles), la Σ-additivité entraîne que

$$\mathbb{P}[A] = \sum_{x \in A} \mathbb{P}[\{x\}] = 0.$$

Exemple 2.29 (Rendez-vous)**.** *Athanase et Bérénice se donnent rendez-vous entre* 12 *et* 13*h. Le moment d'arrivée de chacun est uniforme sur cet intervalle de temps que l'on modélise par l'intervalle réel* $[0,1]$.

11. C'est-à-dire que $\Omega \in \mathscr{B}(\mathbb{R}^d)$ et $\lambda(\Omega) > 0$.
12. C'est la *tribu trace* déjà croisée plus haut à propos du conditionnement.

Probabilités - Préparation à l'agrégation interne de mathématiques (21/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

La probabilité qu'Athanase arrive avant Bérénice doit valoir $1/2$ *par symétrie ; vérifions ce fait. Le moment d'arrivée d'Athanase, comme celui de Bérénice, est modélisé par la probabilité uniforme sur l'intervalle* $[0,1]$. *De façon complètement similaire au cas discret, comme les arrivées des deux individus sont (supposées) indépendantes l'une de l'autre, l'expérience complète se modélise par la loi produit sur l'espace produit, qui n'est autre que la loi uniforme sur le carré* $[0,1]\times[0,1]$ *: l'abscisse correspond au moment d'arrivée d'Athanase et l'ordonnée à celui de Bérénice.*

L'événement « A. arrive avant B. » est alors le triangle $\{(x,y)\in[0,1]^2, x<y\}$, *d'aire* $1/2$. *On retrouve bien le résultat annoncé.*

On peut alors répondre à des questions plus complexes. Calculons par exemple la probabilité que le temps passé par le premier arrivé à attendre son ami(e) ne dépasse pas 10 *minutes. Cet événement correspond à la région du carré* $E=\left\{(x,y)\in[0,1]^2, \left|x-y\right|\leq\frac{1}{6}\right\}$ *représentée à droite dans la figure 2.2. Son aire vaut* $11/36$, *qui est aussi la probabilité cherchée.*

Probabilités - Préparation à l'agrégation interne de mathématiques (22/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Chapitre 3

Variables aléatoires et intégration

3.1 Variables aléatoires réelles

Dans le chapitre précédent, nous nous sommes intéressés *via* la notion d'événement à des conséquences *qualitatives* d'une expérience aléatoire : le dé a-t-il (oui ou non) donné un résultat pair ? la fléchette est-elle (oui ou non) dans la bonne zone ? Ma soupe du soir était-elle liquide ou congelée ? Pour étudier des conséquences *quantitatives*, il faut introduire la notion de **variable aléatoire**, qui fournira un outil pour poser des questions plus complexes : quel est le résultat du dé ? À quelle distance du centre la fléchette est-elle tombée ? Quelle est la température de ma soupe ?

Définition 3.1 (Variable aléatoire réelle)**.** *Si* $(\Omega, \mathscr{F}, \mathbb{P})$ *est un espace probabilisé, on appelle* ***variable aléatoire réelle*** *(abrégé v.a.r.) toute application* $X : \Omega \to \mathbb{R}$ *telle que*

$$\{X \in I\} := \{\omega \in \Omega : X(\omega) \in I\} = X^{-1}(I) \in \mathscr{F}$$

pour tout intervalle $I \subset \mathbb{R}$*. On dit que* X *est une v.a.r.* ***discrète*** *lorsque* $X(\Omega)$ *est fini ou infini dénombrable, typiquement* $X(\Omega) = \mathbb{N}$*.*

Pile ou face : variable aléatoire « nombre de piles ». On joue n fois à pile ou face, ce qu'on modélise par l'espace probabilisé $(\Omega, \mathscr{F}, \mathbb{P})$ où $\Omega = \{0,1\}^n$, $\mathscr{F}$ est la tribu grossière et $\mathbb{P}$ l'équiprobabilité.

On s'intéresse au « nombre de piles » : cette quantité *aléatoire* qui *varie* en fonction de l'événement élémentaire qui se réalise est codée par la fonction $X : \Omega \to \mathbb{R}$ définie pour tout $\omega = (a_1, \cdots, a_n) \in \{0,1\}^n i$ par

$$X(\omega) = \sum_{i=1}^{n} a_i.$$

Cette fonction est bien une variable aléatoire, puisque $X^{-1}(I)$ est une partie de Ω et que $\mathscr{F} = \mathscr{P}(\Omega)$. Comme l'image directe $X(\Omega)$ est l'ensemble fini $\{0, 1, \ldots, n\}$, X est discrète. On verra plus loin que X suit une loi binomiale.

Dans les cas où $\mathscr{F}$ n'est pas la tribu grossière, il faut en principe vérifier la condition de mesurabilité [1].

Théorème 3.2 (Caractérisation)**.** *Une application* $X : (\Omega, \mathscr{F}) \to \mathbb{R}$ *est une v.a.r. si et seulement si* $\{X \leq x\} = X^{-1}(]-\infty, x]) \in \mathscr{F}$ *pour tout* $x \in \mathbb{R}$*.*

Démonstration. Découle des axiomes des tribus car tout intervalle de $\mathbb{R}$ s'obtient en utilisant un nombre au plus dénombrable d'intervalles de la forme $]-\infty, x]$ et symboles $\cap, \cup, ^c$. Par exemple $]a, b] =]-\infty, b] \cap]-\infty, a]^c$ et $[a, b] = \cap_{n \in \mathbb{N}^*}]a - 1/n, b]$. □

Fléchettes : abscisse. On reprend l'exemple du tir de fléchettes sur la cible ronde : $\Omega = \{(x, y) \in \mathbb{R}^2 ; x^2 + y^2 \leq 1\}$, $\mathscr{F}$ est la tribu borélienne et $\mathbb{P}$ la mesure de Lebesgue normalisée.

1. C'est-à-dire qu'il faut vérifier que l'ensemble $\{X \in I\}$ fait partie de $\mathscr{F}$, l'ensemble des partie de Ω auxquelles $\mathbb{P}$ attribue une mesure.

Probabilités - Préparation à l'agrégation interne de mathématiques (23/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

L'abscisse du point d'impact est représentée par la fonction

$$X : \begin{cases} \Omega & \to \mathbb{R} \\ \omega = (x, y) & \mapsto x. \end{cases}$$

On peut vérifier la propriété de mesurabilité et montrer que X est bien une variable aléatoire.

Exemple 3.3. *Si $(\Omega, \mathscr{F}, \mathbb{P})$ est un espace de probabilité et $A \in \mathscr{F}$ alors $\mathbf{1}_A$ est une variable aléatoire discrète booléenne (i.e. prenant les valeurs 0 ou 1). On dit qu'il s'agit d'une variable aléatoire de Bernoulli de paramètre $\mathbb{P}[\mathbf{1}_A = 1] = \mathbb{P}[A]$. Ainsi, les variables aléatoires (quantitatives) permettent aussi de parler des événements (qualitatifs). Plus généralement, si $x_1, \dots, x_n \in \mathbb{R}$ et $A_1, \dots, A_n \in \mathscr{F}$ alors $x_1 \mathbf{1}_{A_1} + \cdots + x_n \mathbf{1}_{A_n}$ est une v.a.r. discrète.*

On rappelle que $\mathbb{R}$ est muni de la tribu borélienne, engendrée par les intervalles. On dit que $f : \mathbb{R} \to \mathbb{R}$ est **borélienne** si pour tout intervalle I, l'image réciproque $f^{-1}(I)$ appartient à la tribu borélienne $\mathscr{B}(\mathbb{R})$. On admet que toute fonction continue est borélienne.

Théorème 3.4 (Stabilité)**.** *Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace probabilisé. L'ensemble $L_0(\Omega, \mathscr{F})$ des variables aléatoires sur Ω est une $\mathbb{R}$-algèbre commutative : pour toutes variables X et Y, tous réels λ et μ, les fonctions $(\lambda X + \mu Y)$ et XY sont des variables aléatoires ; elles sont discrètes si X et Y le sont.*

De plus, si $f : \mathbb{R} \to \mathbb{R}$ est borélienne, alors $f(X) := f \circ X : \omega \mapsto f(X(\omega))$ est une v.a.r. ; elle est discrète si X l'est.

$$\begin{tikzcd} (\Omega, \mathscr{F}) \arrow[r, "X"] \arrow[dr, "f \circ X"'] & (\mathbb{R}, \mathscr{B}(\mathbb{R})) \arrow[d, "f"] \\ & (\mathbb{R}, \mathscr{B}(\mathbb{R})) \end{tikzcd}$$

La notion de fonction borélienne n'est pas explicitement au programme, mais le fait que $f(X)$ est une variable aléatoire si f est raisonnable l'est.

Remarque 3.5. *Le produit ou la somme de variables non-discrètes peut être discret ; de même $f(X)$ peut être discrète même si X ne l'est pas.*

Ce théorème est admis.

Fléchettes : variable aléatoire. Pour le jeu de fléchettes, la distance au centre $R : \omega \mapsto \sqrt{X^2(\omega) + Y^2(\omega)}$ est une variable aléatoire. En effet X en est une, donc X^2 aussi (la fonction $x \mapsto x^2$ est continue donc borélienne). Par somme $X^2 + Y^2$ est une variable aléatoire ; on conclut en composant par la fonction continue $x \mapsto \sqrt{x}$.

3.2 Fonction de répartition et loi

Fonction de répartition

On a vu précédemment que la même expérience pouvait se modéliser par des univers Ω différents, mais que deux modélisations différentes devaient donner la même probabilité aux événements d'intérêt. De même, ce qui importe pour l'étude de propriétés quantitatives n'est *pas* tant *la variable aléatoire en elle-même* mais bien les *probabilités que cette variable tombe dans telle ou telle région* de $\mathbb{R}$. Une première façon de coder mathématiquement ce comportement est la notion de **fonction de répartition**.

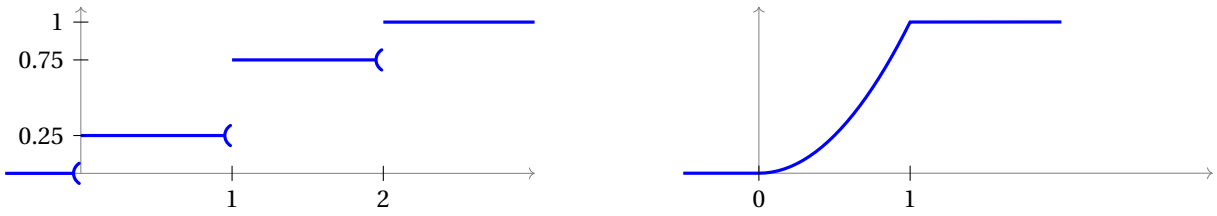
Définition 3.6 (Fonction de répartition)**.** *La fonction de répartition d'une v.a.r. X est la fonction $F_X : \mathbb{R} \to [0, 1]$ définie pour tout $x \in \mathbb{R}$ par :*

$$F_X(x) := \mathbb{P}[X \leq x] = \mathbb{P}[\{\omega : X(\omega) \leq x\}].$$

Théorème 3.7 (Propriété des fonction de répartition)**.** *Si X est une v.a.r. alors*

1. *F_X est croissante et continue à droite ;*
2. *$\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$.*

Probabilités - Préparation à l'agrégation interne de mathématiques (24/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



À gauche : la fonction de répartition du nombre de piles sur 2 lancers d'une pièce équilibrée. La médiane est unique et vaut 1 ; tous les réels dans $[0,1[$ sont des quantiles d'ordre 0.25.
À droite : la fonction de répartition de la distance au centre pour une fléchette lancée uniformément sur le disque unité. Cette fonction induit une bijection de $]0,1[$ sur $]0,1[$ donc tous les quantiles sont uniques.

FIGURE 3.1 – Deux fonctions de répartition

Démonstration. La fonction F_X est croissante car si $x \le y$ alors $\{X \le x\} \subset \{X \le y\}$. Les propriétés suivantes s'obtiennent en utilisant le théorème 2.11, successivement avec...

— l'intersection décroissante $\bigcap_n \{X \le x_n\} = \{X \le x\}$ pour toute suite $(x_n) \searrow x$;
— l'union croissante $\bigcup_n \{X \le x_n\} = \mathbb{R}$ pour toute suite $(x_n) \nearrow \infty$;
— l'intersection décroissante $\bigcap_n \{X \le x_n\} = \varnothing$ pour toute suite $(x_n) \searrow -\infty$. □

Remarque 3.8. *On peut montrer que ces deux propriétés caractérisent les fonctions de répartition : si une fonction F vérifie ces propriétés, alors on peut construire un espace probabilisé $(\Omega, \mathscr{F}, \mathbb{P})$ et une v.a.r. X sur Ω telle que $F_X = F$.*

Fléchettes : fonction de répartition. Soit R la distance au centre pour une fléchette lancée uniformément sur le disque unité. Pour $r < 0$, $\mathbb{P}[R \le r] = 0$ et pour $r \ge 1$, $\mathbb{P}[R \le r] = 1$. La région intéressante est l'intervalle $r \in [0,1]$: pour un tel r, $\{R \le r\} = \{\omega = (x,y) : x^2 + y^2 \le r\} = D(O,r)$ est d'aire πr^2, donc $\mathbb{P}[R \le r] = \frac{\pi r^2}{\pi 1^2} = r^2$. Cette fonction est représentée dans la figure 3.1.

Théorème 3.9 (Fonction de répartition et probabilités d'intervalles)**.** *Pour tous réels $x < y$:*

$$\mathbb{P}\left[x < X \le y\right] = F_X(y) - F_X(x),$$
$$\mathbb{P}[X < x] = F_X(x^-) := \lim_{\substack{z \to x \\ z < x}} F_X(z),$$
$$\mathbb{P}[X = x] = F(x) - F_X(x^-).$$

Démonstration. Pour la première on écrit l'union disjointe $\{X \le y\} = \{X \le x\} \cup \{x < X \le y\}$. Pour la seconde, on réutilise le théorème 2.11 avec l'union croissante $\{X < x\} = \cup_n \{X \le x_n\}$ pour toute suite $(x_n) \nearrow x$ avec $x_n < x$. □

Définition 3.10 (Médiane, quartiles, quantiles)**.** *Soit X est une v.a.r. de fonction de répartition F_X. Pour tout réel $\alpha \in]0,1[$, si $F_X(x^-) \le \alpha$ et $F_X(x) \ge \alpha$ on dit que x est un **quantile d'ordre α** de X. Ceci est vérifié en particulier si $F_X(x) = \alpha$. Pour $\alpha = \frac{1}{2}$ on parle de **médiane**. Les quantiles d'ordre 1/4, 1/2, 3/4 sont appelés **quartiles**, ceux d'ordre 1/10, 2/10, etc., **déciles**.*

Remarque 3.11. *La notion de quantile permet de décrire grossièrement la répartition des valeurs de la v.a.r. X ; en particulier la médiane est une valeur « centrale », et l'écart entre les quartiles donne une idée de l'« étalement » de la variable. En pratique on peut par exemple étudier le salaire médian, l'âge médian, etc. ; regarder le salaire des 10% les plus riches revient à chercher le 9e décile.*

Le plus célèbre des quantiles est le quantile d'ordre $1 - 0.05/2 = 0.975$ de la loi gaussienne standard $\mathcal{N}(0,1)$; il vaut environ 1.96. Ce quantile intervient dans l'expression des bornes de l'intervalle de confiance sur la moyenne obtenu avec le théorème central limite comme on le verra plus loin.

Remarquons que l'équation $F(x) = \alpha$ peut avoir 0, 1 ou une infinité de solutions. Dans le premier cas, le quantile d'ordre α est unique, c'est graphiquement l'ascisse du point où le graphe de F « saute » le niveau α. Dans le dernier cas le quantile n'est a priori *pas unique, même si l'on fixe parfois une règle pour choisir une solution particulière. Pour une variable discrète, F est constante par morceaux et $F(x) = \alpha$ a toujours 0 ou une infinité de solutions.*

Probabilités - Préparation à l'agrégation interne de mathématiques (25/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Loi d'une variable aléatoire

Le « comportement » d'une variable aléatoire, que l'on a étudié au-dessus *via* la fonction de répartition, peut aussi être représenté par sa loi.

Définition 3.12 (Loi)**.** *La **loi** d'une v.a.r. X est la mesure de probabilité $\mathbb{P}_X$ sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ définie pour tout intervalle $I \subset \mathbb{R}$ par :*

$$\mathbb{P}_X(I) = \mathbb{P}[X \in I] = \mathbb{P}[\{\omega, X(\omega) \in I\}].$$

Si μ est une mesure de probabilité sur $\mathbb{R}$, on note $X \sim \mu$ et on dit que « X suit la loi μ » lorsque la v.a.r. X a pour loi μ. Le terme « loi » est synonyme de « mesure de probabilité ».

La loi de X et sa fonction de répartition contiennent les mêmes informations :

Théorème 3.13 (Caractérisation de la loi par la fonction de répartition)**.** *Si X et Y sont deux v.a.r. alors $\mathbb{P}_X = \mathbb{P}_Y$ si et seulement si $F_X = F_Y$.*

Démonstration. Reprendre la preuve du théorème 3.2. □

Sondage simple : deux modélisations, une loi. Athanase et Bérénice reprennent leur expérience de tirage sans remise de deux boules dans une urne contenant 2 boules vertes et 3 boules jaunes. On note X le nombre de boules vertes tirées. Athanase garde son choix d'univers et considère Ω_A l'ensemble des combinaisons de deux boules parmi 5 ($\mathrm{Card}(\Omega_A) = 10$). Sa variable $X = X_A$ est une fonction de Ω_A dans $\mathbb{R}$. Il en déduit les probabilités :

x	0	1	2
$\mathbb{P}[X_A = x]$	3/10	3/5	1/10

Bérénice opte pour l'univers Ω_B des arrangements de deux boules parmi les 5 : $\mathrm{Card}(\Omega_B) = 20$. Pour sa variable $X_B : \Omega_B \to \mathbb{R}$, elle obtient :

x	0	1	2
$\mathbb{P}[X_B = x]$	6/20	$6/20 + 6/20 = 3/5$	$2/20 = 1/10$

Athanase et Bérénice obtiennent donc des variables aléatoires définies sur des espaces différents, mais ces deux variables ont *la même loi.*

Théorème 3.14 (Lois discrètes)**.** *La loi d'une v.a.r. discrète, à valeurs dans un ensemble au plus dénombrable E, est caractérisée par la donnée de $\mathbb{P}[X = x]$ pour tout $x \in E$, car pour tout intervalle $I \subset \mathbb{R}$ on a, en raison du fait que $I \cap E$ est au plus dénombrable,*

$$\mathbb{P}[X \in I] = \mathbb{P}[X \in I \cap E] = \sum_{x \in I \cap E} \mathbb{P}[X = x].$$

*La fonction $x \mapsto \mathbb{P}[X = x]$ n'a pas de nom classique en français ; l'appelation à la mode anglaise est **fonction de masse** (*probability mass function*).*

Pile ou face : quelques lois discrètes classiques. Le jeu de pile ou face permet de définir plusieurs lois classiques. On fixe ici un $p \in]0,1[$ et on considère des lancers répétés d'une pièce qui tombe sur pile avec probabilité p. On représente toujours pile par 1 et face par 0.

Loi de Bernoulli : On note X le résultat du premier lancer. L'ensemble des valeurs possibles est $E = \{0, 1\}$ et

$$\mathbb{P}[X = 1] = 1 - \mathbb{P}[X = 0] = p.$$

Loi binomiale : On note S la somme des résultats des n premiers lancers. On dit que S suit la loi binomiale de paramètres n et p, et l'on note $S \sim \mathrm{Binom}(n, p)$; S prend ses valeurs dans $E = \{0, 1, \ldots, n\}$ et pour tout $k \in E$:

$$\mathbb{P}[X = k] = \binom{k}{n} p^k (1-p)^{n-k}.$$

Loi géométrique : On note T le numéro du premier lancer qui donne pile. On dit que T suit la loi géométrique de paramètre p, et l'on note $T \sim \mathrm{Geom}(p)$; T prend ses valeurs dans $E = \mathbb{N}^*$, et pour

Probabilités - Préparation à l'agrégation interne de mathématiques (26/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

tout $k \geq 1$:

$$\mathbb{P}[X = k] = (1-p)^{k-1} p.$$

On verra plus loin que dans certains cas, la loi binomiale est bien approchée par la **loi de Poisson** $\mathrm{Poi}(\lambda)$, caractérisée pour un paramètre $\lambda > 0$ par $E = \mathbb{N}$ et pour tout $k \in \mathbb{N}$,

$$\mathbb{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Sondage simple : loi hypergéométrique et approximation binomiale. Nous avons déjà vu précédemment sous l'angle combinatoire la **loi hypergéométrique** : si dans une population de $N = N_1 + N_2$ individus dont N_1 sont de type 1 et N_2 de type 2, on effectue un sondage sans remise sur $n \leq N$ individus, alors le nombre X d'individus de type 1, parmi les n individus tirés, suit la loi hypergéométrique $\mathrm{HyperGeom}(N_1, N_2, n)$ sur $E = \{0, 1, \ldots, n\}$ donnée pour tout $0 \leq k \leq n$ par

$$\mathbb{P}[X = k] = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}.$$

Si la taille de la population tend vers l'infini mais que la proportion d'individus de type 1 converge vers p, la loi hypergéométrique converge vers une loi binomiale. Plus précisément, on considère deux suites $N_1(m)$ et $N_2(m)$ qui tendent vers l'infini et telles que

$$\frac{N_1(m)}{N_1(m) + N_2(m)} \xrightarrow[m \to \infty]{} p.$$

On fixe un entier n, et on considère X_m une variable de loi $\mathrm{HyperGeom}(N_1, N_2, n)$. Alors, pour tout $0 \leq k \leq n$, la formule de Stirling[2] entraîne

$$\begin{aligned}\mathbb{P}[X_m = k] &= \frac{\binom{N_1(m)}{k}\binom{N_2(m)}{n-k}}{\binom{N_1(m)+N_2(m)}{n}} \\ &\xrightarrow[m \to \infty]{} \binom{n}{k} p^k (1-p)^{n-k}.\end{aligned}$$

Intuitivement, si la population est très grande devant le nombre de tirages, le tirage sans remise (hypergéométrique) est très proche du tirage avec remise (binomial). En quelque sorte, le résultat du lancer d'une pièce de monnaie correspond à un tirage dans une grande population à deux types constituée par toutes les possibilités physiques du lancer. Cela permet de concevoir le jeu de pile ou face de paramètre p quelconque comme une approximation de tirages équiprobables dans une grande population à deux types. Plus généralement, les modèles non équiprobables peuvent toujours être réduits à des modèles équiprobables, intrinsèquement combinatoires, éventuellement via un passage à la limite sur des paramètres.

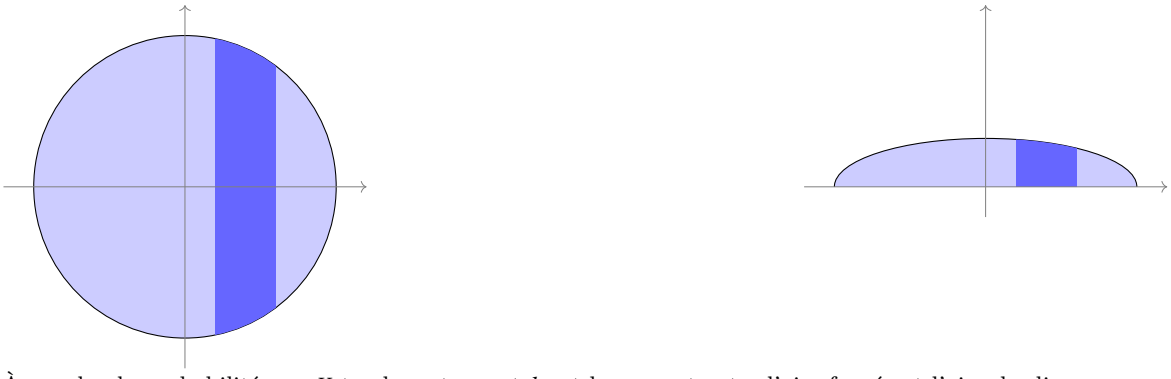
Exemple 3.15 (Lois uniformes)**.** *On dit qu'une v.a.r. X suit la loi uniforme sur l'ensemble fini $\{1, 2, \ldots, n\}$ lorsque $\mathbb{P}[X = k] = 1/n$ pour tout $1 \leq k \leq n$. Dans ce cas, F_X est constante sur les morceaux $]-\infty, 0[, [0, 1[, \ldots, [n-1, n[, [n, \infty[$ et y prend les valeurs $0, 1/n, \ldots, (n-1)/n$, 1. On dit qu'une v.a.r. X suit la loi uniforme sur $[0, 1]$ lorsque $\mathbb{P}_X(I) = |I \cap [0, 1]|$ pour tout intervalle $I \subset \mathbb{R}$. Dans ce cas, pour tout $x \in \mathbb{R}$,*

$$F_X(x) = \begin{cases} 0 & si\ x \leq 0 \\ x & si\ x \in [0, 1] \\ 1 & si\ x \geq 1. \end{cases}$$

Exercice 3.16 (Algorithme de simulation de Fisher-Yates ou de Knuth)**.** *Si $U_1, \ldots, U_n$ sont des v.a. indépendantes avec U_k de loi uniforme sur $\{1, \ldots, k\}$ pour tout $1 \leq k \leq n$, montrer que le produit de transpositions aléatoires $\sigma_n = (1, U_1) \cdots (n, U_n)$ dans le groupe symétrique $\mathscr{S}_n$ suit la loi uniforme sur $\mathscr{S}_n$. Indication : montrer, par récurrence, que $\mathbb{P}[\sigma_n = \sigma]$ ne dépend pas de $\sigma \in \mathscr{S}_n$. Montrer qu'il en est de même du produit inversé $(n, U_n) \cdots (1, U_1)$. Préciser le lien avec le pseudo-code informatique suivant :*

2. $n! \sim \sqrt{2\pi n}(n/e)^n$ ou plus précisément $n! = \sqrt{2\pi n}(n/e)^n(1 + \mathscr{O}_{n\to\infty}(1/n))$.

Probabilités - Préparation à l'agrégation interne de mathématiques (27/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



À gauche, la probabilité que X tombe entre a et b est le rapport entre l'aire foncée et l'aire du disque.
Par symétrie, ce rapport est le même si on supprime la partie inférieure de la cible. On peut ensuite changer l'échelle en y pour obtenir une aire totale de 1, toujours sans changer les rapports d'aire. On obtient la courbe de droite, représentative de la fonction $f_X : x \mapsto \mathbf{1}_{[-1,1]}(x)\frac{2}{\pi}\sqrt{1-x^2}$, qui est donc une densité pour X.

Figure 3.2 – Fléchettes : loi de l'abscisse

```
for k from length(v) downto 2 do swap(v[k],v[ceil(k*rand)])
```

Définition 3.17 (Lois à densité)**.** *On dit qu'une fonction continue par morceaux $f : \mathbb{R} \to \mathbb{R}$ est une* ***densité*** *(de probabilité) lorsqu'elle vérifie*

$$f \geq 0 \quad \text{et} \quad \int_{-\infty}^{+\infty} f(x)\,dx = 1.$$

Si X est une v.a.r., on dit que la loi de X admet la densité f lorsque pour tout intervalle $I \subset \mathbb{R}$,

$$\mathbb{P}[X \in I] = \int_I f(x)\,dx.$$

On dit alors que X a une ***loi à densité****. Dans ce cas, F_X est la primitive de f valant 1 en $+\infty$.*

Nous renvoyons à l'annexe G pour quelques rappels sur les résultats de théorie de l'intégration au programme.

Fléchettes : densités. La distance au centre R dans le jeu de fléchettes est une variable à densité : la fonction

$$f_R : r \mapsto \begin{cases} 2r & \text{si } r \in [0,1], \\ 0 & \text{sinon,} \end{cases}$$

obtenue en dérivant F_R est une densité de R. Notons que la fonction g_R définie par $g_R(r) = f_R(r)$ pour $r \neq 1$ et $g_R(1) = 0$ est également une densité pour R : il n'y a donc pas unicité.

Si l'on note X l'abscisse du point d'impact, le raisonnement indiqué sous la figure 3.2 montre que X admet la densité

$$f_X : x \mapsto \frac{2}{\pi}\mathbf{1}_{[-1,1]}(x)\sqrt{1-x^2}.$$

On dit que X suit la « loi du demi-cercle ».

Voici quelques exemples de lois à densité.

Loi uniforme sur $[a,b]$: $f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$ avec $a < b$; c'est l'analogue continu de l'équiprobabilité.

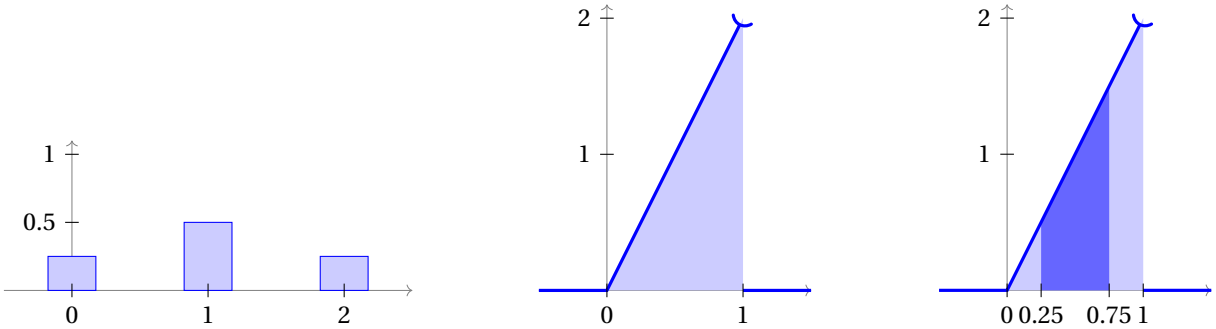
Loi exponentielle : $f(x) = \lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$ avec $\lambda > 0$; elle apparaît pour modéliser des durées de vie, des temps d'attente (*cf* l'annexe E).

Loi normale (ou gaussienne) $f(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{(x-m)^2}{2\sigma^2})$ avec $m \in \mathbb{R}$ et $\sigma > 0$. Elle est fondamentale en raison de son apparition dans le théorème limite central ; nous lui consacrerons une section plus loin.

Loi de Cauchy : pour $a > 0$, $f(x) = \frac{a}{\pi(a^2+x^2)}$. C'est un exemple simple de loi à « queue lourde » : f décroît lentement à l'infini, et X prend souvent de grandes valeurs. On verra plus loin que loi de Cauchy ne *vérifie pas la loi des grands nombres.*

De nombreux autres exemples apparaissent naturellement en statistiques : lois de Pareto, loi du χ^2, loi Beta, loi Gamma... Un bon exercice consiste à calculer lorsque cela est possible la fonction de répartition.

Probabilités - Préparation à l'agrégation interne de mathématiques (28/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



À gauche, le diagramme en bâtons réprésentant la fonction de masse de la variable « nombre de piles sur deux lancers ». La probabilité d'obtenir chaque valeur est proportionnelle à la hauteur du bâton correspondant.
Au milieu, une densité de la variable R, distance au centre pour un jeu de fléchettes. L'aire sous la courbe vaut $\int f(x)dx = 1$.
À droite, pour la même variable R, la probabilité de tomber entre deux valeurs est égale à l'aire sous la courbe entre ces deux valeurs : $\mathbb{P}[0.25 \le R \le 0.75] = \int_{0.25}^{0.75} 2r dr = 1/2$.

FIGURE 3.3 – Diagrammes en bâtons et densités

Simulation par la fonction de répartition

De nombreux problèmes en probabilités et statistiques sont trop complexes pour être traités analytiquement ; on a alors souvent recours à des simulations numériques. Pour les mener à bien il est utile de pouvoir simuler des tirages de variables aléatoires suivant des lois diverses. Supposons que l'on ait accès à un générateur de loi uniforme (fonction communément appelée `rand`). Nous allons voir que si l'on sait « inverser » la fonction de répartition d'une variable X, alors on sait simuler X.

Les fonctions de répartition ne sont en général pas bijectives (voir par exemple la figure 3.1). On peut toutefois « presque » les inverser.

Définition 3.18 (Inverse généralisé)**.** *Soit F une fonction croissante, continue à droite, de limites* 0 *et* 1 *en $-\infty$ et $+\infty$. La fonction G définie par*

$$\begin{aligned} G : [0,1] &\to \overline{\mathbb{R}} \\ y &\mapsto G(y) := \inf\{x \in \mathbb{R} : F(x) \ge y\} \end{aligned}$$

est appelée ***inverse généralisé*** *de F. Elle est croissante et vérifie*

$$\begin{aligned} & G(]0,1[) \subset \mathbb{R}, \\ \forall (x,y) \in \mathbb{R} \times [0,1], \quad & F(G(y)) \ge y \quad \textit{et} \quad G(F(x)) \le x, \\ \forall (x,y) \in \mathbb{R} \times [0,1], \quad & (G(y) \le x) \Leftrightarrow (y \le F(x)). \end{aligned}$$

Démonstration. La définition de G entraîne immédiatement sa croissance ; l'hypothèse sur les limites garantit que $G(]0,1[)$ est inclus dans $\mathbb{R}$. La définition entraîne également :

— si $F(x) \ge y$, alors $G(y) \le x$,

— si $x > G(y)$, il existe $x' \in [G(y), x]$ tel que $F(x') \ge y$.

Le premier point appliqué à $y = F(x)$ montre que $G(F(x)) \le x$. Comme F est continue à droite, le deuxième point entraîne $F(G(y)) \ge y$.

Pour montrer le sens direct de l'équivalence, il suffit de composer $G(y) \le x$ par la fonction croissante F et d'utiliser $y \le F(G(y))$. La réciproque vient de même en utilisant la croissance de G. □

Théorème 3.19 (Méthode d'inversion)**.** *Soit X une v.a.r. de fonction de répartition F_X et U une variable de loi uniforme sur $]0,1[$. Si l'on note G l'inverse généralisé de F_X, alors la variable $X' = G(U)$ a même loi que X.*

Si G est explicite on peut donc simuler X à partir de U.

Démonstration. Soit x' un réel. Utilisons l'équivalence indiquée à la fin de la définition de G :

$$\begin{aligned} \{\omega : X'(\omega) \le x'\} &= \{\omega : G(U(\omega)) \le x'\} \\ &= \{\omega : U(\omega) \le F(x')\} \\ &= U^{-1}\big(]-\infty, F(x')]\big). \end{aligned}$$

Probabilités - Préparation à l'agrégation interne de mathématiques (29/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Comme U est une variable aléatoire, cet ensemble est dans $\mathscr{F}$, et on peut calculer sa probabilité :

$$\mathbb{P}\left[X' \leq x'\right] = \mathbb{P}\left[U \leq F(x')\right] = F_U(F(x')).$$

La fonction de répartition de la loi uniforme U est connue ; en particulier $F_U(y) = y$ pour $y \in [0,1]$. Par conséquent

$$\mathbb{P}\left[X' \leq x'\right] = F(x')$$

et X' a la même fonction de répartition — et donc la même loi — que X. □

Remarque 3.20 (Cas continu)**.** *On peut vérifier que si F est continue et strictement croissante, elle établit une bijection de $\mathbb{R}$ sur $]0,1[$ d'inverse (la restriction de) G.*

Voyons maintenant quelques applications de ce théorème.

Exemple 3.21 (Simulation d'une loi discrète finie)**.** *Soit U une v.a.r. uniforme sur $[0,1]$. Pour tout $p \in [0,1]$, la v.a.r. $\mathbf{1}_{\{U \leq p\}}$ suit la loi de Bernoulli de paramètre p. Plus généralement, soit $p_1, \ldots, p_n \in [0,1]$ avec $p_1 + \cdots + p_n = 1$. Posons $a_0 = 0$ et $a_i = p_1 + \cdots + p_i$ pour tout $1 \leq i \leq n$. La v.a.r. X à valeurs dans $\{1, \ldots, n\}$ qui vaut par définition i sur l'événement $\{U \in [a_{i-1}, a_i]\}$ vérifie forcément $\mathbb{P}[X = i] = p_i$ pour tout $1 \leq i \leq n$.*

Exemple 3.22 (Simulation de la loi uniforme sur $[a,b]$)**.** *Si U suit la loi uniforme sur $[0,1]$ alors pour tout $a < b$, la v.a.r. $(b-a)U + a$ suit la loi uniforme sur $[a,b]$.*

Exemple 3.23 (Simulation de la loi exponentielle)**.** *Si U suit la loi uniforme sur $[0,1]$ alors $-\ln(1-U)/\lambda$ suit la loi exponentielle de paramètre λ. Comme $1-U$ et U ont même loi, on peut utiliser alternativement $-\ln(U)/\lambda$. Sur un ordinateur, la fonction $\ln$ dilate la discrétisation de U prés de 0 et la précision sera mauvaise dans cette zone. Cependant, cette dilatation est modérée (en $\ln$), et la simulation reste correcte.*

Exemple 3.24 (Simulation de la loi de Cauchy)**.** *Si U suit la loi uniforme sur $[0,1]$ alors $\tan(\pi U - \pi/2) = -\frac{1}{\tan(\pi U)}$ suit la loi de Cauchy. Sur un ordinateur, la fonction $1/\tan(\pi\cdot)$ dilate la discrétisation de U près de 0 et de 1. Cette dilatation est plus forte que dans l'exemple précédent (en $1/x$ en 0). Les grandes valeurs de la loi de Cauchy seront mal simulées alors qu'elles ont une forte influence sur la moyenne.*

3.3 Espérance — définition et propriétés générales

Espérance dans le cas fini

Définition 3.25. *Soit X une v.a. discrète à valeurs dans un ensemble E **fini**. L'**espérance** de la variable X est la quantité :*

$$\mathbb{E}[X] := \sum_{x \in E} x \cdot \mathbb{P}[X = x].$$

Si X représente un gain, $\mathbb{E}[X]$ est ce qu'on peut « raisonnablement espérer » gagner en une expérience ; c'est la *moyenne* des valeurs de X, *pondérée* par les poids $\mathbb{P}[X = x]$.

On peut montrer (et c'est un bon exercice) que cette espérance est *linéaire* : si X et Y sont deux variables prenant un nombre fini de valeurs, si λ et μ sont deux réels, alors $\lambda X + \mu Y$ prend un nombre fini de valeurs et :

$$\mathbb{E}\left[\lambda X + \mu Y\right] = \lambda \mathbb{E}[X] + \mu \mathbb{E}[Y].$$

Pile ou face : espérance de la loi binomiale. On joue à pile ou face, la probabilité d'obtenir pile étant $p \in]0,1[$. Calculons l'espérance des variables X et S définies plus haut. La variable de Bernoulli X prend ses valeurs dans $\{0,1\}$, son espérance vaut

$$\mathbb{E}[X] = \sum_{x \in \{0,1\}} x\mathbb{P}[X = x] = 0 \cdot (1-p) + 1 \cdot p = p.$$

La variable binomiale S vaut k avec probabilité $\binom{n}{k}(1-p)^{n-k}p^k$; son espérance vaut

$$\mathbb{E}[S] = \sum_{k=0}^{n} k \cdot \binom{n}{k}(1-p)^{n-k}p^k.$$

Pour calculer cette somme on peut par exemple utiliser, pour $k \geq 1$, la formule[3] combinatoire $k\binom{n}{k} = n\binom{n-1}{k-1}$

Probabilités - Préparation à l'agrégation interne de mathématiques (30/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

et écrire :

$$\begin{aligned}
\mathbb{E}[S] &= \sum_{k=1}^{n} k\binom{n}{k}(1-p)^{n-k}p^k \\
&= n\sum_{k=1}^{n}\binom{n-1}{k-1}(1-p)^{(n-1)-(k-1)}p^{k-1}p \\
&= np\sum_{k'=0}^{n-1}\binom{n-1}{k'}(1-p)^{(n-1)-k'}p^{k'} \\
&= np(1-p+p)^{n-1} \\
&= np.
\end{aligned}$$

où l'on a fait le changement d'indice $k' = k-1$ avant d'utiliser à l'avant-dernière ligne la formule du binôme.

Nous verrons plus loin d'autres façons de calculer cette espérance.

Définition générale

La bonne façon de généraliser la notion d'espérance à des variables aléatoires quelconques n'est *a priori* pas claire : si E est infini dénombrable il peut y avoir des problèmes de convergence, et pour E infini indénombrable la somme n'a plus de sens.

On peut alors chercher à étendre la notion en gardant la propriété de linéarité.

Théorème 3.26 (Espérance des variables positives – Admis)**.** *Soit $L_+(\Omega,\mathscr{F},\mathbb{P})$ l'ensemble des variables aléatoires définies sur $(\Omega,\mathscr{F},\mathbb{P})$ et à valeurs dans $[0,\infty]$. Il existe une unique application $\mathbb{E}: L_+(\Omega,\mathscr{F},\mathbb{P}) \mapsto [0,\infty]$ avec les propriétés suivantes (convention $0\times\infty = 0$) :*

1. *$\mathbb{E}[\mathbf{1}_A] = \mathbb{P}[A]$ pour tout $A \in \mathscr{F}$, et en particulier $\mathbb{E}[\mathbf{1}_\Omega] = 1$;*
2. *$\mathbb{E}$ est linéaire : $\mathbb{E}\left[\lambda X + \mu Y\right] = \lambda\mathbb{E}[X] + \mu\mathbb{E}[Y]$ pour tous $X, Y \in L_+(\Omega,\mathscr{F},\mathbb{P})$ et $\lambda,\mu \in \mathbb{R}_+$;*
3. *$\mathbb{E}[\lim_{n\to\infty} X_n] = \lim_{n\to\infty}\mathbb{E}[X_n]$ pour toute suite (X_n) croissante de $L_+(\Omega,\mathscr{F},\mathbb{P})$.*

La troisième propriété est connue sous le nom de *convergence monotone*.

Preuve du cas discret. Si X prend un nombre fini de valeurs, elle s'écrit $X = \sum_{k=1}^{n} x_k \mathbf{1}_{A_k}$ avec $A_k = \{\omega : X(\omega) = x_k\} \in \mathscr{F}$. Grâce aux deux premières propriétés on retrouve bien la formule $\mathbb{E}[X] = \sum_{k=1}^{n} x_k\mathbb{P}[A_k]$. Si X est une variable aléatoire discrète positive, alors il existe une suite $(X_n)_{n\geq 1}$ croissante de v.a.r. du type précédent telles que $X = \lim_{n\to\infty} X_n$ d'où $\mathbb{E}[X] = \sum_{x\in X(\Omega)} x\mathbb{P}[X = x]$ par convergence monotone pour les séries. On admet le résultat au-delà des v.a.r. discrètes. □

Exercice 3.27. *Montrer que si $X \geq 0$ et $\mathbb{E}[X] < \infty$ alors $\mathbb{P}[X < \infty] = 1$.*

Généralisons maintenant la notion d'espérance à des variables de signe quelconque. Pour cela, si $X : (\Omega,\mathscr{F},\mathbb{P}) \to \mathbb{R}$ est une v.a.r., on la décompose en différences de variables positives en posant $X_+ = \max(X,0)$ et $X_- = \max(-X,0)$; on a alors :

$$|X| = X_+ + X_- \quad \text{et} \quad X = X_+ - X_-.$$

Définition 3.28 (Espérance des variables quelconques)**.** *On dit qu'une v.a.r. $X : (\Omega,\mathscr{F},\mathbb{P}) \to \mathbb{R}$ est **intégrable** lorsque $\mathbb{E}|X| = \mathbb{E}[X_+] + \mathbb{E}[X_-] < \infty$. On définit alors l'**espérance** de X en posant*

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-].$$

On note $L^1(\Omega,\mathscr{F},\mathbb{P})$ l'ensemble des v.a.r. intégrables, et plus généralement $L^p(\Omega,\mathscr{F},\mathbb{P})$ l'ensemble des v.a.r. X telles que $|X|^p$ est intégrable.

L'espérance possède les propriétés fondamentales et immédiates suivantes :

i) linéarité : si $X, Y \in L^1(\Omega,\mathscr{F},\mathbb{P})$ et $\lambda,\mu \in \mathbb{R}$ alors $\lambda X + \mu Y \in L^1(\Omega,\mathscr{F},\mathbb{P})$ et

$$\mathbb{E}\left[\lambda X + \mu Y\right] = \lambda\mathbb{E}[X] + \mu\mathbb{E}[Y] \ ;$$

3. Pour justifier cette formule, on peut compter le nombre de façons de choisir parmi n personnes un groupe de k personnes et un chef de groupe. Pour cela on peut d'abord choisir le groupe ($\binom{n}{k}$ choix) puis choisir un chef dans le groupe (k choix), ou on peut choisir d'abord le chef (n choix) puis compléter le groupe avec $k-1$ autres personnes ($\binom{n-1}{k-1}$ choix).

Probabilités - Préparation à l'agrégation interne de mathématiques (31/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

ii) positivité : si $X \in L_+(\Omega, \mathscr{F}, \mathbb{P})$ alors $\mathbb{E}[X] \geq 0$;

iii) croissance : si $X, Y \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ avec $X \leq Y$ alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

L'inégalité triangulaire donne $|\mathbb{E}[X_+] - \mathbb{E}[X_-]| \leq \mathbb{E}[X_+] + \mathbb{E}[X_-]$, donc

$$|\mathbb{E}[X]| \leq \mathbb{E}[|X|].$$

Notons que $\mathbb{E}$ est une forme linéaire sur l'espace vectoriel $L^1(\Omega, \mathscr{F}, \mathbb{P})$. Si X est une v.a.r. constante et égale à un réel c alors X est intégrable et $\mathbb{E}[X] = c$.

Définition 3.29 (Moments)**.** *Si $X \in L^p(\Omega, \mathscr{F}, \mathbb{P})$ avec $p \in \mathbb{N}^*$, on dit que $\mathbb{E}[X^p]$ est le* ***moment***[4] *d'ordre p de X. En particulier, $\mathbb{E}[X]$ est le « premier moment » de X.*

Si X est bornée, c'est-à-dire $\mathbb{P}[|X| \leq r] = 1$ pour un réel $r \in \mathbb{R}_+$, alors $X \in L^p(\Omega, \mathscr{F}, \mathbb{P})$ pour tout $p \geq 1$ et X possède dans ce cas des moments de tout ordre.

Exercice 3.30 (Espérance des indicatrices)**.** *Le principe d'inclusion-exclusion (théorème 2.3) découle via les propriétés de l'espérance de l'identité suivante :*

$$1 - \mathbf{1}_{\cup_{1 \leq i \leq r} A_i} = \prod_{1 \leq i \leq r} \mathbf{1}_{A_i^c} = \prod_{1 \leq i \leq r} (1 - \mathbf{1}_{A_i}) = \sum_{k=1}^{r} \sum_{1 \leq i_1 < \cdots < i_k \leq r} (-1)^k \mathbf{1}_{A_{i_1} \cap \cdots \cap A_{i_k}}.$$

Les inégalités de Boole-Bonferonni raffinent le principe d'inclusion-exclusion :

$$\mathbb{P}[\cup_{1 \leq i \leq r} A_i] - \sum_{k=1}^{m} (-1)^{k+1} S_k \quad \textit{est} \quad \begin{cases} \geq 0 & \textit{si } m \textit{ impair} \\ \leq 0 & \textit{si } m \textit{ pair} \\ = 0 & \textit{si } m = r \textit{ (inclusion-exclusion !)} \end{cases}$$

En prenant $m = 1$ à gauche et $m = 2$ à droite on obtient le cas particulier (pour $r \geq 2$)

$$\underbrace{\mathbb{P}[A_1] + \cdots + \mathbb{P}[A_r]}_{S_1} \leq \mathbb{P}[\cup_{1 \leq i \leq r} A_i] \leq \underbrace{\mathbb{P}[A_1] + \cdots + \mathbb{P}[A_r]}_{S_1} - \underbrace{\sum_{i<j} \mathbb{P}\left[A_i \cap A_j\right]}_{S_2}.$$

Pour établir les inégalités de Boole-Bonferroni, on commence par observer que si

$$x_0 \leq \cdots \leq x_k \geq x_{k+1} \geq \cdots \geq x_r$$

sont des nombres réels vérifiant $\sum_{k=0}^{r} (-1)^k x_k = 0$ alors $\sum_{k=0}^{m} (-1)^k x_k \geq 0$ pour les m pairs et ≤ 0 pour les m impairs (ceci mérite une démonstration, qui est omise). Appliquée à la suite des coefficients binomiaux $x_0 = \binom{r}{0}, \ldots, x_r = \binom{r}{r}$, cette observation donne

$$\sum_{k=0}^{m} (-1)^k \binom{r}{k} \quad \textit{est} \quad \begin{cases} \geq 0 & \textit{si } m \textit{ impair} \\ \leq 0 & \textit{si } m \textit{ pair} \\ = 0 & \textit{si } m = r \textit{ (formule du binôme !).} \end{cases}$$

À présent, si $r(\omega)$ désigne le nombre d'indices $j \in \{1, \ldots, r\}$ tels que $\omega \in A_j$ alors

$$\sum_{1 \leq i_1 < \cdots < i_k \leq r} \mathbf{1}_{A_{i_1}}(\omega) \cdots \mathbf{1}_{A_{i_k}}(\omega) = \binom{r(\omega)}{k}.$$

Le résultat désiré découle ensuite de la linéarité et de la positivité de l'espérance. On trouvera une application des inégalités de Boole-Bonferroni en fiabilité dans le livre de Delmas et Jourdain [DJ07].

4. Ce terme vient du vocabulaire physique — le moment (appelé usuellement en français « quantité de mouvement ») d'un système étant une moyenne des vitesses de ses éléments, pondérée par leurs masses.

Probabilités - Préparation à l'agrégation interne de mathématiques (32/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

3.4 Espérance des variables aléatoires discrètes

L'écriture d'une variable discrète positive X comme limite monotone de sommes finies d'indicatrices montre que la formule reste valable pour des variables discrètes positives :

$$\mathbb{E}[X] = \sum_{x \in E} x \cdot \mathbb{P}[X = x].$$

Nous renvoyons à l'annexe G pour des rappels sur le sens de la somme. Le théorème du transfert permet plus généralement de calculer l'espérance de fonctions de v.a.r. discrètes.

Théorème 3.31 (Espérance et formule du transfert pour les v.a.r. discrètes)**.** *Si X est une v.a.r. discrète à valeurs dans un ensemble au plus dénombrable E alors $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ si et seulement si la somme $\sum_{x \in E} |x| \mathbb{P}[X = x]$ est finie, et on a alors*

$$\mathbb{E}[X] = \sum_{x \in E} x \mathbb{P}[X = x].$$

Plus généralement, pour toute fonction $\varphi : \mathbb{R} \to \mathbb{R}$, on a $\varphi(X) \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ si et seulement si la somme $\sum_{x \in E} \left|\varphi(x)\right| \mathbb{P}[X = x]$ est finie, et on a alors la ***formule du transfert***

$$\mathbb{E}\left[\varphi(X)\right] = \sum_{x \in E} \varphi(x) \mathbb{P}[X = x].$$

Remarque 3.32 (Terminologie)**.** *Le terme de* transfert *vient du formalisme de théorie de la mesure : l'espérance d'une variable $X : \Omega \to E$ s'y interprète comme une intégration sur l'espace de départ Ω, par rapport à la mesure (de probabilité) $\mathbb{P}$. Le « transfert » consiste à calculer cette espérance comme une somme sur l'espace d'arrivée E.*

Démonstration. Quitte à numéroter les éléments de E, on peut supposer que $E = \mathbb{N}$. En écrivant $\varphi = \varphi_+ - \varphi_-$ on se ramène au cas où $\varphi \geq 0$. On a par convergence monotone

$$\mathbb{E}\left[\varphi(X)\right] = \lim_{n \to \infty} \mathbb{E}\left[\varphi(X) \mathbf{1}_{\{X \leq n\}}\right]$$

et le résultat découle alors du fait que $\mathbb{E}\left[\varphi(X) \mathbf{1}_{\{X \leq n\}}\right] = \sum_{k=0}^{n} \varphi(k) \mathbb{P}[X = k]$. □

Exercice 3.33 (Transfert)**.** *Retrouver la formule du transfert pour les v.a.r. discrètes en partant de la formule $\mathbb{E}\left[\varphi(X)\right] = \sum_z z \mathbb{P}\left[\varphi(X) = z\right]$. De même, retrouver la linéarité de l'espérance pour les v.a.r. discrètes en partant de la formule $\mathbb{E}[X + Y] = \sum_z z \mathbb{P}[X + Y = z]$.*

Théorème 3.34. *Si X est une v.a.r. à valeurs dans $\mathbb{N}$ alors*

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{P}[X \geq n].$$

Démonstration. Partons de la formule de l'espérance :

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \mathbb{P}[X = k] = \sum_{k=1}^{\infty} k \mathbb{P}[X = k].$$

Écrivons k comme somme[5] de k fois l'entier 1 :

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \sum_{j=1}^{k} \mathbb{P}[X = k] = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \mathbf{1}_{j \leq k} \mathbb{P}[X = k].$$

Comme tous les termes sont positifs, on peut intervertir les sommes (c'est un cas particulier du théorème de Fubini–Tonelli) :

$$\mathbb{E}[X] = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \mathbf{1}_{j \leq k} \mathbb{P}[X = k] = \sum_{j=1}^{\infty} \mathbb{P}\left[X \geq j\right].$$ □

5. C'est donc une transformation d'Abel.

Probabilités - Préparation à l'agrégation interne de mathématiques (33/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Pile ou face : espérance de la loi géométrique. Pour trouver l'espérance du temps d'attente T du premier pile, on doit calculer la somme :

$$\mathbb{E}[T] = \sum_{k=1}^{\infty} k(1-p)^{k-1} p.$$

La méthode la plus courante est peut-être de factoriser par p, de voir la somme restante comme une fonction de p et d'en chercher une primitive en intégrant sous le signe somme. Le théorème précédent fournit une réponse rapide, puisqu'il n'y a qu'à sommer des séries géométriques : en posant $q = 1-p$,

$$\mathbb{E}[T] = \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} q^{k-1} p = \sum_{j=1}^{\infty} q^{j-1} = \frac{1}{1-q} = \frac{1}{p}.$$

Par exemple, le nombre moyen de lancers à faire pour obtenir un 6 sur un dé équilibré est $1/(1/6) = 6$.

Exercice 3.35 (Probabilités et convexité : inégalité de Jensen)**.** *Soit $\varphi : I \to \mathbb{R}$ une fonction convexe sur un intervalle $I \subset \mathbb{R}$ et X une v.a.r. discrète à valeurs dans I telle que X et $\varphi(X)$ sont intégrables. Alors*

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

et de plus, l'égalité est atteinte si $\mathbb{P}[X = c] = 1$ pour une constante c, et c'est le seul cas d'égalité possible lorsque est φ strictement convexe. Indication : se ramener au cas où X est discrète finie, utiliser le théorème du transfert et la définition de la convexité. Cas particuliers importants : $\varphi(x) = |x|$, $\varphi(x) = x^2$, $\varphi(x) = |x|^p$ avec $p \geq 2$, $\varphi(x) = e^x$, $\varphi(x) = x\ln(x)$ (si $X \geq 0$).

Exercice 3.36 (Application de l'inégalité de Jensen à l'entropie de Boltzmann-Shannon)**.** *L'entropie $H(X)$ de la loi d'une v.a. discrète finie X prenant les valeurs $x_1, \ldots, x_n$ avec les probabilités $p_1, \ldots, p_n$ est*

$$H(X) = -\sum_{k=1}^{n} p_k \ln(p_k),$$

avec la convention $0\ln(0) = 0$. Calculée avec un logarithme en base 2, elle représente le nombre moyen de bits par symboles nécessaires au codage dans un canal parfait avec un alphabet de n symboles[6]*. On définit de même l'entropie d'une v.a. discrète infinie, sous réserve que la série converge. En utilisant l'inégalité de Jensen, montrer que l'entropie relative définie pour toutes lois discrètes P et Q sur $\mathbb{N}$ par*

$$\mathrm{Ent}(Q|P) = \sum_n \frac{Q_n}{P_n} \ln \frac{Q_n}{P_n} P_n = \sum_n Q_n \ln \frac{Q_n}{P_n}$$

vérifie $\mathrm{Ent}(Q|P) \geq 0$ avec égalité si et seulement si $P = Q$. En déduire également que parmi les lois sur $\mathbb{N}$ de même moyenne fixée, la loi géométrique est l'unique loi qui maximise l'entropie, et calculer la valeur du maximum. Montrer également que parmi les lois de même support fini, la loi uniforme est l'unique loi qui maximise l'entropie, et calculer la valeur du maximum.

3.5 Espérance des variables aléatoires à densité

Le théorème du transfert permet de calculer l'espérance de v.a.r. à densité et plus généralement l'espérance de fonctions de v.a.r. à densité. Il montre en particulier que $\mathbb{E}\left[\varphi(X)\right]$ ne dépend que de la loi de X via sa densité. Attention : si X est à densité, alors $\varphi(X)$ n'est pas forcément à densité.

Théorème 3.37 (Espérance et formule du transfert pour les v.a.r. à densité — Admis)**.** *Si X est une v.a.r. continue de densité f alors $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ si et seulement si la fonction $x \mapsto |x| f(x)$ est intégrable sur $\mathbb{R}$. On a alors*

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x)\, dx.$$

Plus généralement, pour toute $\varphi : \mathbb{R} \to \mathbb{R}$ borélienne, on a $\varphi(X) \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ si et seulement si la fonction $x \mapsto |\varphi(x)| f(x)$ est intégrable sur $\mathbb{R}$. On a alors la formule du transfert :

$$\mathbb{E}\left[\varphi(X)\right] = \int_{-\infty}^{+\infty} \varphi(x) f(x)\, dx.$$

6. Le codage peut être réalisé par le code de Huffman, présent partout en informatique, y compris dans les images au format JPEG et dans les fichiers musicaux au format MP3.

Probabilités - Préparation à l'agrégation interne de mathématiques (34/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Eléments de démonstration. Soit I un intervalle de $\mathbb{R}$, et soit $\varphi = \mathbf{1}_I$. Par définition de $\mathbb{E}$ et f,

$$\mathbb{E}\left[\varphi(X)\right] = \mathbb{E}\left[\mathbf{1}_I(X)\right] = \mathbb{P}\left[X \in I\right] = \int_I f(x)\,dx = \int_{-\infty}^{+\infty} \varphi(x) f(x)\,dx :$$

la formule du transfert est donc valable pour les fonctions φ indicatrices d'intervalles.

On admet qu'elle reste valable lorsque I est un borélien de $\mathbb{R}$. Par linéarité, elle reste valable pour toute fonction φ étagée (i.e. constante sur un nombre fini de boréliens). Pour établir le résultat pour toute fonction φ borélienne, on se ramène tout d'abord au cas où $\varphi \geq 0$ en utilisant la décomposition $\varphi = \varphi_+ - \varphi_-$, puis on considère une suite croissante $(\varphi_n)_{n\geq 1}$ de fonctions positives étagées (constantes sur un nombre fini de boréliens) convergeant vers φ (existence admise) et on obtient par convergence monotone $\mathbb{E}\left[\varphi(X)\right] = \lim_{n\to\infty} \mathbb{E}\left[\varphi_n(X)\right]$ et $\lim_{n\to\infty} \mathbb{E}\left[\varphi_n(X)\right] = \int_{-\infty}^{+\infty} \varphi(x) f(x)\,dx$. □

Fléchettes : espérance. La distance moyenne au centre dans le jeu de fléchettes vaut :

$$\mathbb{E}\left[R\right] = \int r f_R(r) dr = \int_0^1 r \cdot 2r dr = \frac{2}{3}.$$

Posons $S = R^2$. On peut calculer $\mathbb{E}\left[S\right]$ de deux manières. La première est de trouver la loi de S, en cherchant sa fonction de répartition : pour $s \in [0,1]$, on trouve facilement $\mathbb{P}\left[S \leq s\right] = \mathbb{P}\left[R \leq \sqrt{s}\right] = s$. Ainsi S suit la loi uniforme sur $[0,1]$, d'espérance $1/2$.

La deuxième méthode est d'appliquer le théorème du transfert avec la fonction $\varphi : x \mapsto x^2$:

$$\mathbb{E}\left[S\right] = \int_{\mathbb{R}} r^2 f_R(r) dr = \int_0^1 2r^3 dr = 1/2.$$

Exercice 3.38 (Moyenne)**.** *En utilisant le théorème du transfert, retrouver les formules pour la moyenne de la loi uniforme, exponentielle, et normale. La loi de Cauchy possède-t-elle une moyenne ? Une médiane ? Idem pour les lois de Pareto et de Student.*

Remarque 3.39 (Queues lourdes)**.** *Les lois sans espérance comme la loi de Cauchy ne sont pas des objets exotiques réservés aux contre-exemples : de nombreux phénomènes naturels donnent des échantillons répartis en lois de puissance comme la loi de Cauchy, de Pareto, de Student. Notons que si X n'a pas de variance alors X^2 n'a pas d'espérance.*

Exercice 3.40 (Calcul des moments)**.** *En utilisant l'intégration par parties, montrer que les moments de la loi exponentielle de paramètre λ sont donnés pour tout $n \geq 1$ par*

$$m_n = \int_0^\infty x^n \lambda e^{-\lambda x}\,dx = \frac{n!}{\lambda^n}$$

(on peut établir la formule de récurrence $m_{n+1} = \lambda^{-1}(n+1)m_n$). Montrer que les moments d'ordre impairs de la loi normale standard $\mathcal{N}(0,1)$ sont nuls tandis que les moments pairs sont donnés par

$$\int_{-\infty}^{+\infty} x^{2n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}\,dx = \prod_{k=1}^{n} (2k-1) = \frac{(2n)!}{2n(2n-2)\cdots 2} = \frac{(2n)!}{2^n n!}.$$

Montrer que les moments d'ordre impairs de la loi du demi-cercle de densité

$$x \mapsto \frac{2}{\pi}\sqrt{1-x^2}\mathbf{1}_{[-1,1]}(x)$$

sont nuls tandis que les moments d'ordre pair sont les nombres de Catalan $\frac{1}{n+1}\binom{2n}{n}$.

Exercice 3.41. *Soit X une v.a.r. positive et intégrable, de densité f et de fonction de répartition F. Montrer en utilisant le théorème de convergence dominée que*

$$\lim_{r\to\infty} r\mathbb{P}\left[X > r\right] = \lim_{r\to\infty} \mathbb{E}\left[r\mathbf{1}_{[r,\infty[}(X)\right] = 0.$$

En déduire par intégration par parties basée sur $-(1-F)' = f$ sur un intervalle $[0,r]$ l'analogue suivant du théorème 3.34 :

$$\mathbb{E}\left[X\right] = \lim_{r\to\infty} \int_0^r x f(x)\,dx = \int_0^\infty \mathbb{P}\left[X > x\right]\,dx.$$

Probabilités - Préparation à l'agrégation interne de mathématiques (35/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Plus généralement, soit X une v.a.r. pas forcément positive, telle que $|X|^p$ est intégrable pour un réel $p \geq 1$. Montrer au moyen du théorème de Fubini-Tonelli que

$$\mathbb{E}\left[|X|^p\right] = p\int_0^\infty t^{p-1}\mathbb{P}\left[|X| > t\right] dt.$$

Cette identité est importante : elle relie moments et queue de distribution. Nous reviendrons sur ce lien à propos des inégalités de Markov et Bienaymé–Tchebychev.

Exercice 3.42 (Inégalité de Jensen et entropie)**.** *L'inégalité de Jensen de l'exercice 3.35 reste valable au delà du cadre discret : si $\varphi : I \to \mathbb{R}$ est une fonction convexe sur un intervalle $I \subset \mathbb{R}$ et si X est une v.a.r. à valeurs dans I telle que X et $\varphi(X)$ sont intégrables alors $\varphi(\mathbb{E}[X]) \leq \mathbb{E}\left[\varphi(X)\right]$. De plus, l'égalité est atteinte si $\mathbb{P}(X = c) = 1$ pour une constante c, et il s'agit de plus du seul cas d'égalité possible lorsque φ est strictement convexe. On définit l'entropie de Boltzmann-Shannon d'une v.a.r. X de densité f par*

$$H(X) = -\int f(x)\ln(f(x))\, dx$$

lorsque l'intégrale existe. En théorie de l'information, cette quantité intervient dans le calcul de la capacité des canaux de communication, voire également l'exercice 3.36 sur l'analogue discret. En utilisant l'inégalité de Jensen, montrer que l'entropie relative, définie pour deux lois P et Q de densités $f > 0$ et $g > 0$ par

$$\mathrm{Ent}(Q|P) = \int \frac{g(x)}{f(x)} \ln \frac{g(x)}{f(x)} f(x)\, dx = \int g(x) \ln \frac{g(x)}{f(x)}(x)\, dx$$

vérifie $\mathrm{Ent}(Q|P) \geq 0$ avec égalité si et seulement si $P = Q$. En déduire que parmi les lois à densité sur un intervalle $[a, b]$ fixé (respectivement sur $\mathbb{R}_+$ à moyenne fixée, respectivement sur $\mathbb{R}$ centrée à variance fixée) la loi uniforme (respectivement la loi exponentielle, respectivement la loi gaussienne) est l'unique loi qui maximise l'entropie.

3.6 Variance

L'espérance d'une variable est un « indicateur de centralité », qui donne une idée de la grandeur de la variable. Des variables très différentes peuvent avoir même espérance [7]. On a vu précédemment que la notion de quantile permettait de caractériser l'« étalement » d'une variable ; malheureusement les quantiles ne sont pas toujours simples à manipuler [8].

La **variance** d'une v.a.r. est une manière pratique de quantifier l'étalement. Elle est définie sur l'espace $L^2(\Omega, \mathscr{F}, \mathbb{P}) = \{X : (\Omega, \mathscr{F}, \mathbb{R}) \to \mathbb{R}, \mathbb{E}\left[|X|^2\right] < \infty\}$ des v.a.r. de carré intégrable.

Théorème 3.43 (Carré intégrable)**.** *L'ensemble $L^2(\Omega, \mathscr{F}, \mathbb{P})$ est un espace vectoriel et si $X, Y \in L^2(\Omega, \mathscr{F}, \mathbb{P})$ alors $XY \in L^1(\Omega, \mathscr{F}, \mathbb{P})$. En particulier, $L^2(\Omega, \mathscr{F}, \mathbb{P}) \subset L^1(\Omega, \mathscr{F}, \mathbb{P})$.*

Démonstration. On a $(X+Y)^2 \leq 2(X^2 + Y^2)$ et donc $L^2(\Omega, \mathscr{F}, \mathbb{P})$ est un espace vectoriel. De plus, $XY = \frac{1}{2}((X+Y)^2 - X^2 - Y^2)$ et donc $XY \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ si $X, Y \in L^2(\Omega, \mathscr{F}, \mathbb{P})$. □

Comme $\mathbb{E}\left[X^2\right] \geq 0$, l'application $(X, Y) \mapsto \mathbb{E}[XY]$ est bilinéaire, symétrique et positive [9] sur $L^2(\Omega, \mathscr{F}, \mathbb{P})$ et en particulier, on dispose de l'inégalité de Cauchy-Schwarz :

$$|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]}.$$

Définition 3.44 (Variance)**.** *La **variance** de $X \in L^2(\Omega, \mathscr{F}, \mathbb{P})$ est le nombre réel* positif

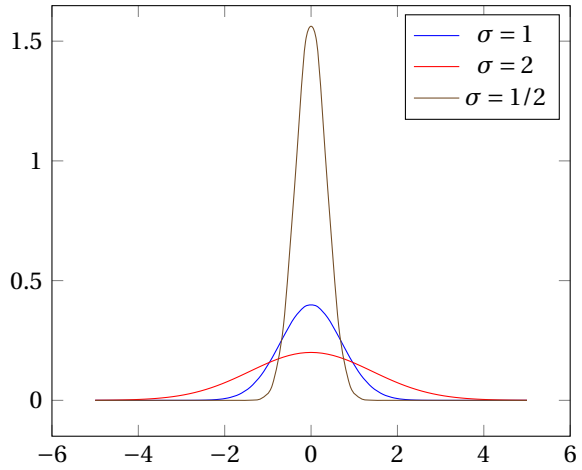
$$\sigma^2(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

7. D'un point de vue plus statistique, une note moyenne de 10 peut être obtenue en mettant 10 à tous les élèves, ou en mettant 20 à la moitié de la classe et 0 à l'autre ; on peut avoir une température moyenne agréable en mettant « les pieds dans la glace et la tête dans le four »...

8. Par exemple, connaître les trois quartiles de deux variables X et Y ne suffit pas pour déterminer les quartiles de $X + Y$.

9. Cette application est « presque » un produit scalaire : si $\mathbb{E}\left[X^2\right] = 0$, alors $X = 0$ avec probabilité 1 (c'est-à-dire qu'il existe $A \in \mathscr{F}$ tel que $\mathbb{P}[A] = 1$ et pour tout $\omega \in A$, $X(\omega) = 0$). Pour obtenir un véritable produit scalaire il faut identifier les variables aléatoires quand elles coïncident avec probabilité 1.

Probabilités - Préparation à l'agrégation interne de mathématiques (36/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



Ci-contre, densités

$$f_\sigma(x) = (2\pi\sigma)^{-1/2}\exp(-((x-m)/\sigma)^2)$$

des lois normales $\mathcal{N}(m,\sigma^2)$, pour une moyenne $m=0$ et trois valeurs de l'écart-type σ. Plus σ est grand, plus la densité est étalée ; plus σ est petit plus la loi se concentre, en terme de masse, autour de sa moyenne.

FIGURE 3.4 – Variance des lois normales

La variance de X représente la moyenne du carré des écarts à la moyenne. Pour obtenir un nombre de même dimension que X, on définit l'**écart-type** de X par

$$\sigma(X) = \sqrt{\sigma^2(X)}.$$

Comme $\sigma^2(X) = \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right]$, on obtient la formule de König

$$\sigma^2(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.$$

On a $\sigma^2(X) = 0$ si et seulement si $\mathbb{P}[X = \mathbb{E}[X]] = 1$ (c'est-à-dire si X est presque sûrement constante).

Fléchettes : variance. La distance au centre R pour le jeu de fléchettes est bornée par 1, elle est donc dans L^2. La variance vaut $\sigma^2(R) = \mathbb{E}\left[R^2\right] - \mathbb{E}[R]^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$.

Pile ou face : variance. La variance d'une loi de Bernoulli de paramètre p vaut $p(1-p)$. Elle est maximale quand $p = 1/2$: dans ce cas, l'espérance vaut $1/2$, la variance $1/4$ et l'écart-type $1/2$. On peut remarquer que dans ce cas très particulier l'écart entre X et sa moyenne vaut toujours exactement $1/2$: il est donc rassurant que ce soit également la valeur de l'écart-type...

Si p est très proche de 0, $\mathbb{P}[X=0] = 1-p$ est proche de 1 donc X est souvent proche de sa moyenne ; la variance tend vers 0, on dit que la variable est très « concentrée » autour de sa moyenne.

Exercice 3.45 (Moindres carrés)**.** *Montrer que si X est une v.a.r. de carré intégrable alors*

$$\sigma^2(X) = \min_{m\in\mathbb{R}} \mathbb{E}\left[(X-m)^2\right],$$

le minimum étant atteint en $m = \mathbb{E}[X]$. Indication : développer en m et utiliser la linéarité de l'espérance. Ainsi $\sigma(X) = \sqrt{\sigma^2(X)}$ est la distance des moindres carrés (i.e. L^2) du point X de L^2 au sous-espace vectoriel de $\mathrm{L}^2(\Omega,\mathscr{F},\mathbb{P})$ formé par les v.a.r. constantes. En d'autres termes, $\mathbb{E}[X]$ est la constante la plus proche de X au sens L^2, et la distance vaut $\sigma(X)$.

Exercice 3.46 (Moyenne et variance des lois usuelles)**.** *Établir que la loi normale $\mathcal{N}(m,\sigma)$ de densité $f_{m,\sigma}(x) = (2\pi\sigma)^{-1/2}e^{-((x-m)/\sigma)^2)}$ a pour moyenne m et pour écart-type σ. Plus généralement, recalculer la moyenne et la variance pour les lois usuelles (table 3.1).*

Probabilités - Préparation à l'agrégation interne de mathématiques (37/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

3.7 Inégalités de Markov et de Bienaymé-Tchebychev

Comme $\mathbb{R}$ est réunion dénombrable de compacts, si X est une v.a.r. alors pour tout $\varepsilon > 0$ il existe un compact $K \subset \mathbb{R}$ tel que $\mathbb{P}[X \notin K] \leq \varepsilon$, et on dit que la loi de X est *tendue*. Les inégalités de Markov et de Bienaymé-Tchebychev ci-dessous permettent de mieux quantifier la propriété de tension lorsque X possède des moments finis.

Théorème 3.47 (Inégalité de Markov)**.** *Si X est intégrables à valeurs positives, alors pour tout $r > 0$,*

$$\mathbb{P}[X \geq r] \leq \frac{\mathbb{E}[X]}{r}.$$

L'inégalité de Markov n'a pas d'intérêt quand $r \leq \mathbb{E}[X]$ (la borne est alors ≥ 1). L'inégalité de Markov affirme qu'une v.a.r. $X \geq 0$ intégrable est toujours concentrée autour de 0 : si par exemple $\mathbb{E}[X] = 1$, alors X dépasse 100 avec une probabilité inférieure ou égale à 1%.

Démonstration. Découle de la croissance de l'espérance utilisée avec $r\mathbf{1}_{\{X \geq r\}} \leq X$. Notons que si X est à densité, la variable de gauche est discrète et celle de droite à densité, ce qui justifie de développer une théorie générale pour l'espérance. □

La seconde inégalité permet de traduire quantitativement l'interprétation de la variance comme indicateur de l'étalement d'une variable.

Théorème 3.48 (Bienaymé-Tchebychev)**.** *Si $X \in L^2(\Omega, \mathscr{F}, \mathbb{P})$ alors pour tout $r > 0$,*

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq r] \leq \frac{\sigma^2(X)}{r^2}.$$

La probabilité de faire un écart de r à la moyenne est donc d'autant plus petite que r est grand, et d'autant plus petite que le variance est petite.

On dit parfois que $[\mathbb{E}[X] \pm r]$ est un **intervalle de fluctuation** pour X, de niveau $1 - \frac{\sigma^2(X)}{r^2}$. On parle plus souvent d'**inégalité de déviation**.

Démonstration. L'inégalité de Markov (théorème 3.47) appliquée à la v.a.r. positive $(X - \mathbb{E}[X])^2$ donne

$$\begin{aligned}\mathbb{P}[|X - \mathbb{E}[X]| \geq r] &= \mathbb{P}\left[(X - \mathbb{E}[X])^2 \geq r^2\right] \\ &\leq \frac{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}{r^2} = \frac{\sigma^2(X)}{r^2}.\end{aligned}$$ □

L'inégalité n'a pas d'intérêt lorsque $r \leq \sigma(X)$ car la borne est alors plus grande que 1.

Remarque 3.49 (Utilité de l'inégalité de Bienaymé-Tchebychev)**.** *L'inégalité est très générale : elle s'applique dès que la variable a un moment d'ordre 2. Elle est donc très utile pour établir des résultats théoriques, comme on le verra plus loin à propos de la loi des grands nombres.*

Sur des exemples concrets simples, par contraste, elle peut donner des résultats très loin de la valeur optimale. Si X est par exemple le résultat d'un lancer de dé ($\mathbb{E}[X] = 7/2$, $\sigma^2(X) = 35/12$), le caractère discret de X et l'inégalité de Bienaymé–Tchebychev donnent

$$\mathbb{P}[X \notin \{2,3,4,5\}] = \mathbb{P}[|X - \mathbb{E}[X]| \geq r] \leq \frac{35}{12r^2}$$

pour tout $r \in]3/2, 5/2]$. La borne la meilleure est obtenue en faisant tendre r vers $5/2$, ce qui donne

$$\mathbb{P}[X \notin \{2,3,4,5\}] \leq 7/15 \approx 0.46,$$

alors que la valeur exacte est $1/3$. Si l'on veut par la même méthode majorer $\mathbb{P}[X \notin \{3,4\}]$, l'inégalité ne donne rien (le meilleur $r = 3/2$ étant plus petit que l'écart-type).

Plus généralement, soit $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ une fonction mesurable croissante telle que $\varphi(r) > 0$ pour tout $r > 0$. Si $\varphi(|X - \mathbb{E}[X]|)$ est intégrable alors pour tout $r > 0$

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq r] \leq \frac{\mathbb{E}\left[\varphi(|X - \mathbb{E}[X]|)\right]}{\varphi(r)}.$$

Pour $\varphi(r) = r^2$ on retrouve l'inégalité de Bienaymé-Tchebychev. Ce résultat est le plus souvent appliqué avec $\varphi(r) = r^p$ pour $p \geq 1$, ou avec $\varphi(r) = \exp(r)$. La « morale » est la suivante : plus X est intégrable, plus elle est concentrée autour de sa moyenne.

Probabilités - Préparation à l'agrégation interne de mathématiques (38/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Exercice 3.50 (Inégalité de Paley-Zygmund)**.** *Soit X une v.a.r. positive de carré intégrable. Montrer que pour tout réel $\theta \in [0,1]$, on a l'inégalité de déviation :*

$$\mathbb{P}[X > \theta\mathbb{E}[X]] \geq (1-\theta)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}\left[X^2\right]} = (1-\theta)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X]^2 + \sigma^2(X)}.$$

Indication : utiliser $X \leq \theta\mathbb{E}[X] + X\mathbf{1}_{\{X>\theta\mathbb{E}[X]\}}$ puis l'inégalité de Cauchy-Schwarz.

3.8 Fonction génératrice

Définition et propriétés générales

Nous avons déjà vu trois façons de caractériser la loi d'une variable aléatoire :

— par sa fonction de masse (dans le cas discret) ou sa densité (dans le cas continu) ;

— par la définition, c'est-à-dire par la donnée pour tout intervalle I de la quantité

$$\mathbb{P}[X \in I] = \mathbb{E}[\mathbf{1}_I(X)] \ ;$$

— par sa fonction de répartition, c'est-à-dire par la donnée pour tout x de la quantité

$$F_X(x) = \mathbb{P}[X \leq x] = \mathbb{E}\left[\mathbf{1}_{]-\infty,x]}(X)\right];$$

Plus généralement, on peut caractériser la loi de X par la donnée des $\{\mathbb{E}\left[f(X)\right] : f \in \mathfrak{F}\}$ dès lors que la classe de fonctions $\mathfrak{F}$ est assez riche. On peut en particulier faire les choix suivants :

— indicatrices d'intervalles de la forme $]-\infty, x]$, ou d'intervalles quelconques (cas précédents) ;

— fonctions continues bornées $\mathbb{R} \to \mathbb{R}$ (permettent d'approcher les indicatrices) ;

— fonctions mesurables positives $\mathbb{R} \to \mathbb{R}$ (contient les indicatrices d'intervalles !) ;

— fonctions trigonométriques de la forme $x \mapsto e^{itx}$ avec $t \in \mathbb{R}$ (transformée de Fourier [10]) ;

— fonctions de la forme $x \mapsto s^x$ avec $s \in [0,1]$ (fonction génératrice, si X discrète) ;

— fonctions de la forme $x \mapsto e^{-tx}$ avec $t \geq 0$ (transformée de Laplace, si $X \geq 0$).

Un de ces choix est explicitement au programme, celui des *fonctions génératrices*, utilisé pour traiter des variables *discrètes à valeurs dans* $\mathbb{N}$. Pour les v.a.r. et lois continues, on utilise plutôt les transformées de Laplace ou les fonctions caractéristiques (hors programme).

Définition 3.51 (Fonction génératrice)**.** *Si P est une loi de probabilité sur $\mathbb{N}$ alors sa* ***fonction génératrice****, notée $g_P : [0,1] \to \mathbb{R}$, est définie pour tout $s \in]-1,1]$ par*

$$g_P(s) = \sum_{n=0}^{\infty} s^n P(\{n\}).$$

Si X est une v.a.r. de loi P sur $\mathbb{N}$ alors on note $g_X = g_P$:

$$g_X(s) = \sum_{n=0}^{\infty} s^n \mathbb{P}[X = n] = \mathbb{E}\left[s^X\right].$$

Remarque 3.52 (Ensemble de définition)**.** *L'ensemble de définition de la fonction génératrice peut varier suivant les auteurs et le cadre. Comme g_X est une série entière et $g_X(1) = 1 < \infty$, on peut définir g_X sur le disque unité complexe fermé $\{z \in \mathbb{C}, |z| \leq 1\}$.*

Remarque 3.53 (Liens avec les transformée de Fourier et de Laplace)**.** *La transformée de Laplace se déduit de la fonction génératrice : pour $t \geq 0$, $\mathbb{E}\left[\exp(-tX)\right] = g_X(e^{-t})$. Par la remarque précédente, on peut étudier la fonction génératrice sur le cercle unité de $\mathbb{C}$ et retrouver la fonction caractéristique : $\theta \in \mathbb{R}_+ \mapsto \mathbb{E}\left[\exp(i\theta X)\right] = g_X(e^{i\theta})$, c'est-à-dire $g_X|_{\{z\in\mathbb{C}:|z|=1\}}$.*

Théorème 3.54 (Propriétés des fonctions génératrices)**.** *Soit X une des v.a. à valeurs dans $\mathbb{N}$. Sa fonction génératrice g_X vérifie les propriétés suivantes.*

i) croissance : *g_X est croissante sur l'intervalle $[0,1]$, avec $g_X(0) = \mathbb{P}[X = 0]$ et $g_X(1) = 1$;*

10. En théorie des probabilités, on parle plutôt de **fonction caractéristique** (de la loi).

Probabilités - Préparation à l'agrégation interne de mathématiques (39/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

ii) fonction « génératrice » : *g_X est $\mathscr{C}^\infty$ sur $]-1,1[$ et ses dérivées « engendrent les probabilités » : $g_X^{(n)}(0) = (n!)\mathbb{P}[X=n]$ pour tout $n \in \mathbb{N}$;*

iii) caractérisation de la loi : *Si Y est une autre v.a. à valeurs dans $\mathbb{N}$, X et Y ont même loi si et seulement si $g_X = g_Y$;*

iv) moments : *si $X(X-1)\cdots(X-k+1)$ est intégrable ($k \in \mathbb{N}^*$) alors*

$$\mathbb{E}[X(X-1)\cdots(X-k+1)] = \lim_{s \nearrow 1} g_X^{(k)}(s).$$

L'expression $\mathbb{E}[X(X-1)\cdots(X-k+1)]$ est appelée **moment factoriel** d'ordre k de X.

Démonstration. Le 1. est immédiat. Le 2. découle du fait que le rayon de convergence de la série entière $\sum_{n=0}^{\infty} z^n \mathbb{P}[X=n]$ est supérieur ou égal à 1. Pour le 3., si $g_X = g_Y$ alors

$$\mathbb{P}[X=n] = g_X^{(n)}(0) = g_Y^{(n)}(0) = \mathbb{P}[Y=n]$$

pour tout $n \in \mathbb{N}$ et la réciproque est évidente. Le 4. s'obtient en calculant $g_X^{(n)}(s)$ pour $s \in]0,1[$ (dérivation sous le signe somme) puis en utilisant le théorème de convergence monotone (faire tendre s vers 1). □

L'un des principaux intérêts des fonctions génératrices sera mis en lumière plus loin dans la section 4.5, voyons-en déjà une utilisation calculatoire.

Exemple 3.55 (Loi géométrique)**.** *La loi géométrique* $\mathrm{Geom}(p)$ *de paramètre p est donnée pour tout $k \geq 1$ par $\mathbb{P}[T=k] = q^{k-1}p$, où $q := (1-p)$. Sa fonction génératrice s'obtient par une simple somme géométrique :*

$$g_T(s) = \sum_{k \geq 1} s^k q^{k-1} p = \frac{sp}{1-sq}.$$

On dérive cette fonction deux fois :

$$g_T'(s) = \frac{p}{(1-sq)^2} \quad \text{et} \quad g_T''(s) = \frac{2pq}{(1-sq)^3}.$$

On en déduit $\mathbb{E}[T] = g_T'(1) = \frac{1}{p}$ et $\mathbb{E}[T(T-1)] = g_T''(1) = \frac{2q}{p^2}$, d'où

$$\begin{aligned}\sigma^2(T) &= \mathbb{E}\left[T^2\right] - \mathbb{E}[T]^2 = \mathbb{E}[T(T-1)] + \mathbb{E}[T](1-\mathbb{E}[T]) \\ &= \frac{2q}{p^2} - \frac{q}{p^2} = \frac{q}{p^2}.\end{aligned}$$

On peut ainsi calculer les moments de T en ne faisant que des dérivations, évitant les calculs de sommes.

La fonction génératrice permet aussi d'étudier la convergence de suites de variables aléatoires :

Théorème 3.56 (Fonctions génératrices et convergence en loi)**.** *Soit X une v.a.r. et (X_n) une suite de variables aléatoires, toutes à valeurs dans $\mathbb{N}$, de fonctions génératrices g et g_n. Si pour tout $s \in [0,1[$,*

$$g_n(s) \xrightarrow[n\to\infty]{} g(s),$$

alors pour tout $k \in \mathbb{N}$,

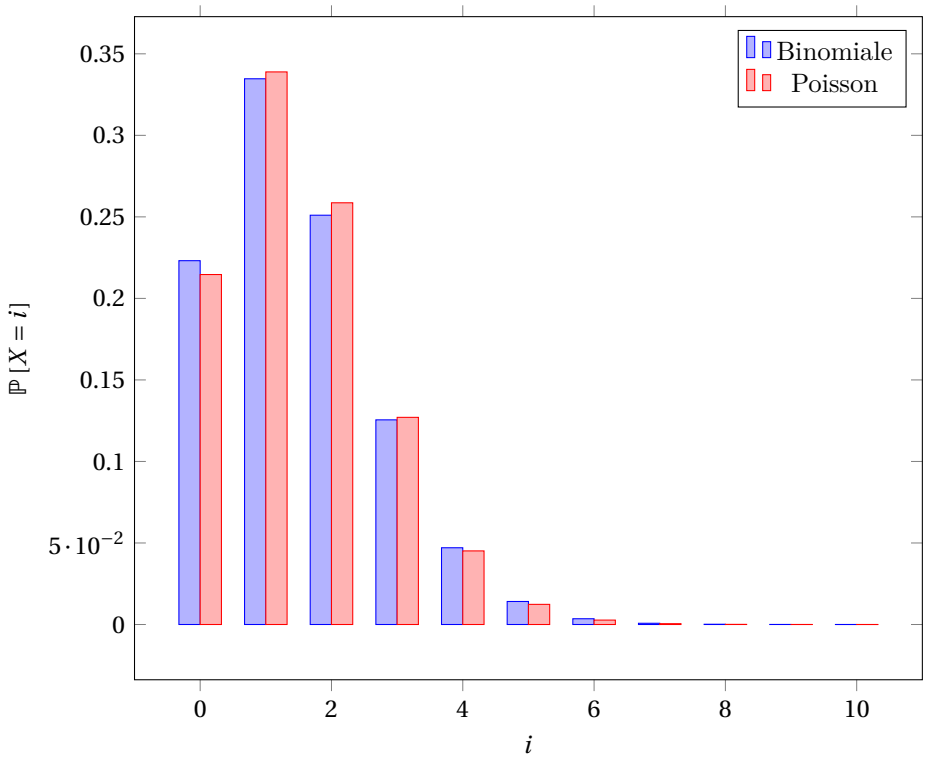
$$\mathbb{P}[X_n = k] \xrightarrow[n\to\infty]{} \mathbb{P}[X=k].$$

On dit que (X_n) ***converge en loi*** *vers X.*

Démonstration. Ce théorème peut se déduire de résultats plus généraux. Pour le montrer de façon élémentaire, on peut procéder par récurrence sur k. Pour initialiser à $k=0$ il suffit de prendre $s=0$. Notons $a_{n,j} = \mathbb{P}\left[X_n = j\right]$ et $a_j = \mathbb{P}\left[X=j\right]$; supposons que l'on a $a_{n,j} \xrightarrow[n\to\infty]{} a_j$, pour tout $j \leq k-1$. On décompose les deux sommes :

$$\begin{aligned}g_n(s) &= \sum_{j=0}^{k-1} a_{n,j} s^j + a_{n,k} s^k + s^k \sum_{j=1}^{\infty} a_{n,k+j} s^j \\ g(s) &= \sum_{j=0}^{k-1} a_j s^j + a_k s^k + s^k \sum_{j=1}^{\infty} a_{k+j} s^j.\end{aligned}$$

Probabilités - Préparation à l'agrégation interne de mathématiques (40/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



Les diagrammes en bâtons de la loi binomiale $\mathrm{Binom}(n, p)$
et de la loi de Poisson $\mathrm{Poi}(np)$ pour $p = 1/20$ et $n = 30$.

FIGURE 3.5 – Approximation de la loi binomiale par la loi de Poisson.

En soustrayant la deuxième ligne de la première, on a pour tout $s \in]0,1[$,

$$a_{n,k} - a_k = \frac{g_n(s) - g(s)}{s^k} - \frac{\sum_{j=0}^{k-1}(a_{n,j} - a_j)s^j}{s^k} - \sum_{j=1}^{\infty}(a_{n,k+j} - a_{k+j})s^j.$$

On prend les valeurs absolues ; dans le dernier terme on majore sauvagement $|a_{n,k} - a_k|$ par 2 (ce sont des probabilités !) :

$$|a_{n,k} - a_k| \leq \left|\frac{g_n(s) - g(s)}{s^k}\right| + \left|\frac{\sum_{j=0}^{k-1}(a_{n,j} - a_j)s^j}{s^k}\right| + 2\sum_{j=1}^{\infty} s^j.$$

Toujours à s fixé, on fait tendre n vers l'infini : à droite, le premier terme disparaît (c'est l'hypothèse) et le second aussi (par hypothèse de récurrence) :

$$\limsup_n |a_{n,k} - a_k| \leq \frac{2s}{1-s}.$$

Ceci étant valable pour tout $s \in]0,1[$, il suffit maintenant de faire tendre s vers 0 pour obtenir :

$$\mathbb{P}[X_n = k] = a_{n,k} \xrightarrow[n\to\infty]{} a_k = \mathbb{P}[X = k]. \qquad \square$$

Application : l'approximation binomiale/Poisson

Le nombre de succès S_n lors de la répétition de n expériences indépendantes, qui réussissent chacune avec probabilité p, suit une loi binomiale de paramètres n et p. L'espérance de S_n vaut np.

Supposons que n soit grand, p petit mais que np soit « de l'ordre de 1 » : donnons quelques exemples.

— Le nombre de parties gagnées à la roulette sur une série de 100 en misant sur un chiffre : $p = 1/38$, $n = 100$ et $np \approx 2.7$.

— Pendant une pluie d'intensité modérée, il tombe $6mm$ par heure et par mètre carré de surface, soit approximativement $0.1L$ par minute par mètre carré. Il y a 10^4 centimètres carrés dans un mètre

Probabilités - Préparation à l'agrégation interne de mathématiques (41/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

carré, et approximativement 10^6 gouttes d'eau dans un litre. En notant S_n le nombre de gouttes tombant en une minute sur une petite surface d'un centimètre carré pendant une minute de pluie, S_n suit une loi binomiale de paramètre $n = 10^5$ et $p = 10^{-4}$, et $np \approx 10$.

Pour donner un sens mathématique précis on considère une limite où $p = p_n$ dépend de n, et où np_n converge vers une limite finie.

Théorème 3.57 (Approximation binomiale/Poisson)**.** *Soit X une variable aléatoire de loi de Poisson* $\mathrm{Poi}(\lambda)$*. Soit p_n une suite de réels tels que $\frac{p_n}{n} \to \lambda$, et pour tout n, soit S_n une variable de loi binomiale* $\mathrm{Binom}(n, p_n)$*. Alors pour tout entier k,*

$$\mathbb{P}[S_n = k] \xrightarrow[n\to\infty]{} \exp(-\lambda)\frac{\lambda^k}{k!} = \mathbb{P}[X = k].$$

Démonstration. En écrivant le coefficient binomial comme

$$\binom{n}{k} = \frac{1}{k!}n(n-1)\cdots(n-k+1) \sim \frac{n^k}{k!},$$

et en raisonnant par équivalents on peut établir le résultat ; les détails sont laissés en exercice.

Grâce au résultat établi plus haut sur les fonctions génératrices, on peut procéder autrement. La fonction génératrice de S_n a été calculée précédemment, elle vaut

$$g_{S_n}(s) = (q_n + p_n s)^n = (1 - p_n(s-1))^n = \exp\big(n\log(1 - p_n(s-1))\big).$$

Quand n tend vers l'infini, $n\log(1 - p_n(s-1)) \sim -\lambda(s-1)$, donc pour tout $s \in]-1, 1[$,

$$g_{S_n}(s) \xrightarrow[n\to\infty]{} \exp(-\lambda(1-s)) = g_X(s).$$

où $g_X(s)$ est la fonction génératrice d'une loi de Poisson. Il suffit alors d'appliquer le théorème 3.56. □

Probabilités - Préparation à l'agrégation interne de mathématiques (42/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Loi	**Support**	**Fonc. de masse/densité**	**Moyenne**	**Variance**	**Modèle**
Bernoulli	$\{0,1\}$	$\mathbb{P}[X=1]=p$	p	$p(1-p)$	Pile ou face
Rademacher	$\{-1,1\}$	$\mathbb{P}[X=1]=p$	$2p-1$	$p(1-p)$	Pile ou face
Binomiale	$\{0,1,\dots,n\}$	$\mathbb{P}[X=k]=\binom{k}{n}p^k(1-p)^{n-k}$	np	$np(1-p)$	Nombre de succès
Poisson	$\mathbb{N}$	$\mathbb{P}[X=k]=e^{-\lambda}\frac{\lambda^k}{k!}$	λ	λ	Limite de lois binomiales
Hypergéom.	$\{0,1,\dots,n\}$	$\mathbb{P}[X=k]=\frac{\binom{k}{N_1}\binom{n-k}{N_2}}{\binom{n}{N}}$	$n\frac{N_1}{N}$	$\frac{nN_1N_2(N-n)}{N^2(N-1)}$	Sondage simple
Géom.	$\mathbb{N}^*$	$\mathbb{P}[X=k]=(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{p}{(1-p)^2}$	Temps d'attente d'un succès
Géom.	$\mathbb{N}$	$\mathbb{P}[X=k]=(1-p)^k p$	$\frac{1-p}{p}$	$\frac{p}{(1-p)^2}$	Temps d'attente d'un succès
Uniforme	$\{1,\dots,n\}$	$\mathbb{P}[X=k]=\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	Équiprobabilité (discrète)
Uniforme	$[a,b]\subset\mathbb{R}$	$x\mapsto\frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$	$\frac{b-a}{2}$	$\frac{(b-a)^2}{12}$	Équiprobabilité (continue)
Expo.	$\mathbb{R}_+$	$x\mapsto\lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Temps d'attente (continu)
Normale	$\mathbb{R}$	$x\mapsto\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-m)^2/(2\sigma^2)}$	m	σ^2	Théorème central limite
Student	$\mathbb{R}$	$x\mapsto\frac{\Gamma(\frac{1}{2}(a+1))}{\sqrt{a\pi}\Gamma(\frac{1}{2}a)}\left(1+\frac{x^2}{a}\right)^{-\frac{1}{2}(a+1)}$	0 $(a>1)$	$\frac{a}{a-2}$ $(a>2)$	Tests statistiques
χ^2	$\mathbb{R}_+$	$x\mapsto\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$	n	$2n$	Tests statistiques
Cauchy	$\mathbb{R}$	$x\mapsto\frac{1}{\pi(1+x^2)}$	non	non	Loi à queue lourde
Pareto	$[1,\infty[$	$x\mapsto\frac{n}{x^{n+1}}\mathbf{1}_{[1,\infty[}(x)$	$\frac{n}{n-1}$ $(n>1)$	$\frac{n}{(n-1)^2(n-2)}$ $(n>2)$	

La loi de Cauchy est une loi de Student avec $a=1$. La loi exponentielle est une loi Gamma avec $a=1$.

Table 3.1 – Quelques lois sur $\mathbb{N}$ ou $\mathbb{R}$.

Probabilités - Préparation à l'agrégation interne de mathématiques (43/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Probabilités - Préparation à l'agrégation interne de mathématiques (44/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Chapitre 4

Vecteurs aléatoires

4.1 Définition, loi d'un vecteur aléatoire

Athanase veut déguster une pizza. Il prend au hasard 4 couverts dans son tiroir, qui contient 2 fourchettes, 3 couteaux et 4 cuillères. Il note X le nombre de fourchettes et Y le nombre de couteaux.

S'il compte manger avec les mains, il n'a besoin que d'un couteau, et seule la variable Y l'intéresse : on a vu plus haut qu'elle suit une loi $\mathsf{HyperGeom}(2,7;4)$. S'il veut jouer à catapulter des boulettes de pain à la fourchette, il n'étudiera que X, qui suit la loi $\mathsf{HyperGeom}(3,6;4)$. Cependant, s'il veut mettre une belle table pour séduire Bérénice, il veut connaître la probabilité qu'il y ait deux fourchettes et deux couteaux : il doit donc étudier *simultanément* X et Y.

Définition 4.1 (Vecteur aléatoire)**.** *Un* ***vecteur aléatoire*** $X = (X_1,\dots,X_d)$ *de* $\mathbb{R}^d$ *est une suite* $X_1,\dots,$ X_d *de variables aléatoires réelles définies sur un même espace* $(\Omega,\mathscr{F},\mathbb{P})$*. La loi de* X *est la mesure de probabilité* $\mathbb{P}_X$ *sur* $\mathbb{R}^d$ *définie pour tout produit d'intervalles* $I_1 \times \cdots \times I_d$ *par*

$$\mathbb{P}_X(I_1 \times \cdots \times I_d) = \mathbb{P}\left[X_1 \in I_1,\dots,X_d \in I_d\right] = \mathbb{P}\left[X \in I_1 \times \cdots \times I_d\right].$$

Les lois des v.a.r. $X_1,\dots,$ X_d *sont les* lois marginales *du vecteur aléatoire* X*.*

Le vecteur aléatoire X de $\mathbb{R}^d$ est discret lorsque $X(\Omega)$ est au plus dénombrable, et sa loi est alors entièrement déterminée par la donnée pour tout $(x_1,\dots,x_d) \in X(\Omega)$ de

$$\mathbb{P}\left[X_1 = x_1,\dots,X_d = x_d\right].$$

Les lois marginales sont des v.a.r. discrètes et leur loi s'obtient en sommant par rapport à toutes les autres variables. Par exemple, la loi de la v.a.r. X_1 est donnée par

$$\mathbb{P}\left[X_1 = x_1\right] = \sum_{x_2 \in X_2(\Omega),\dots,x_d \in X_d(\Omega)} \mathbb{P}\left[X_1 = x_1,\dots,X_d = x_d\right].$$

Exemple 4.2 (Tableau de contingence)**.** *Reprenons l'exemple des couverts d'Athanase. La loi jointe de* X *et* Y *peut être représentée par le tableau suivant (les probabilités sont exprimées en* $1/126^e$*) :*

$X \backslash Y$	0	1	2	3	loi de X
0	1	12	18	4	35
1	8	36	24	2	70
2	6	12	3	0	21
loi de Y	15	60	45	6	

À l'intersection de la ligne i *et de la colonne* j *on écrit (en* $1/126^e$*) la probabilité* $\mathbb{P}\left[X = i, Y = j\right]$*. Sur les « marges » du tableau, la somme en ligne donne la fonction de masse de* X*, et la somme en colonne celle de* Y*.*

Exemple 4.3 (Loi hypergéométrique multitypes)**.** *Le cas général de l'expérience des couverts d'Athanase est la* ***loi hypergéométrique multitype****. On fixe un nombre de couleurs* $d \geq 2$ *et on considère une urne contenant* N_1 *boules de la première couleurs,* N_2 *de la deuxième,…,* N_d *de la dernière couleur.*

Probabilités - Préparation à l'agrégation interne de mathématiques (45/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

On tire n boules dans cette urne et on note X_i le nombre de boules de couleur i tirées. Pour tout n-uplet $(n_1,\dots,n_d)\in\mathbb{N}^d$ tel que $\sum n_i = n$ et $n_i \le N_i$ pour tout i, on a

$$\mathbb{P}\left[X_1 = n_1,\dots,X_d = n_d\right] = \frac{\binom{N_1}{n_1}\cdots\binom{N_d}{n_d}}{\binom{N}{n}}.$$

Comme dans l'exemple des couverts, les marginales suivent des lois hypergéométriques classiques : $X_i \sim \mathrm{HyperGeom}(N_i, N-N_i, n)$.

Lorsque les nombres $N_1,\dots,N_d$ tendent vers l'infini de sorte que les proportions $(N_i/N,\dots,N_d/N)$ convergent vers un vecteur $(p_1,p_2,\dots,p_d)$, la loi hypergéométrique converge vers la ***loi multinomiale*** *de taille n et de paramètre $(p_1,\dots,p_d)$, généralisation de la binomiale, définie par :*

$$\mathbb{P}\left[X_1 = n_1,\dots,X_d = n_d\right] = \frac{n!}{n_1!\cdots n_d!}p_1^{n_1}\cdots p_d^{n_d}.$$

On dit qu'une fonction $f:\mathbb{R}^d\to\mathbb{R}$ est une *densité* de probabilité lorsque

$$f\ge 0 \quad \text{et} \quad \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(x_1,\dots,x_d)\,dx_1\cdots dx_d = 1.$$

On dit que le vecteur aléatoire X admet pour densité f lorsque pour tout pavé $I_1\times\cdots\times I_d$,

$$\mathbb{P}\left[X\in I_1\times\cdots\times I_d\right] = \int_{I_1}\cdots\int_{I_d} f(x_1,\dots,x_d)\,dx_1\cdots dx_d.$$

Comme en dimension 1, la densité d'un vecteur n'est pas unique car on peut légèrement la modifier. Les lois marginales sont également à densité et leur densité s'obtient en intégrant f par rapport à toutes les autres variables (prendre $I_j=\mathbb{R}$ pour $j\ne i$). Par exemple, la densité de X_1 est

$$x_1\in\mathbb{R}\mapsto\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(x_1,\dots,x_d)\,dx_2\dots dx_d.$$

Le calcul peut être mené par intégrations successives grâce au théorème de Fubini-Tonelli.

Le théorème de Fubini-Tonelli est fondamental et simple : ne pas en avoir peur !

Fléchettes : couple abscisse/ordonnée. Notons X l'abscisse et Y l'ordonnée dans le jeu de fléchettes. Si $I\times J$ est un pavé de $\mathbb{R}^2$,

$$\mathbb{P}\left[(X,Y)\in I\times J\right] = \frac{\text{aire de } (I\times J)\cap\Omega}{\text{aire de } \Omega}.$$

L'aire de $I\times J$ est égale à $\iint \mathbf{1}_{I\times J}(x,y)dxdy$; on admet que l'analogue est vrai pour la région $(I\times J)\cap\Omega$, de sorte que

$$\mathbb{P}\left[(X,Y)\in I\times J\right] = \frac{1}{\pi}\iint \mathbf{1}_{(I\times J)\cap\Omega}(x,y)dxdy = \iint_{I\times J}\mathbf{1}_{\Omega}(x,y)dxdy.$$

La fonction $(x,y)\mapsto\frac{1}{\pi}\mathbf{1}_\Omega(x,y)$ est donc une densité pour le vecteur (X,Y).

Retrouvons la loi du demi-cercle par intégration : X admet la densité

$$f(x) = \int_{\mathbb{R}}\frac{1}{\pi}\mathbf{1}_\Omega(x,y)dy = \frac{1}{\pi}\int_{\mathbb{R}}\mathbf{1}_{x^2+y^2\le 1}dy = \frac{1}{\pi}\int_{\mathbb{R}}\mathbf{1}_{y^2\le 1-x^2}dy.$$

Si $|x|>1$ l'indicatrice est toujours nulle et $f(x)$ vaut 0. Si $|x|\le 1$ on a

$$f(x) = \frac{1}{\pi}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}dy = \frac{2}{\pi}\sqrt{1-x^2}.$$

On retrouve bien la loi du demi-cercle.

Théorème 4.4 (Théorème du transfert pour les vecteurs aléatoires – Admis)**.** *Soit $X=(X_1,\dots,X_d)$ un vecteur aléatoire de $\mathbb{R}^d$ et $\varphi:\mathbb{R}^d\to\mathbb{R}$ une fonction borélienne.*

Probabilités - Préparation à l'agrégation interne de mathématiques (46/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

1. *Si X est discret et si la série $\sum_{x\in X(\Omega)} \left|\varphi(x)\right| \mathbb{P}[X=x]$ converge alors*

$$\mathbb{E}\left[\varphi(X_1,\dots,X_d)\right] = \sum_{(x_1,\dots,x_d)\in\Omega(X)} \varphi(x_1,\dots,x_d)\mathbb{P}[X_1=x_1,\dots,X_d=x_d].$$

2. *Si X a pour densité f et si $x\in\mathbb{R}^d \mapsto |\varphi(x)|f(x)$ est intégrable alors*

$$\mathbb{E}\left[\varphi(X_1,\dots,X_d)\right] = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} \varphi(x_1,\dots,x_d)\, f(x_1,\dots,x_d)\, dx_1\cdots dx_d.$$

Démonstration. On procède comme pour les variables aléatoires. □

Exercice 4.5 (Linéarité de l'espérance)**.** *Soit (X,Y) un vecteur aléatoire de $\mathbb{R}^2$ de densité f. En utilisant le théorème du transfert pour les vecteurs aléatoires, retrouver la propriété de linéarité de l'espérance : $\mathbb{E}[X+Y]=\mathbb{E}[X]+\mathbb{E}[Y]$.*

4.2 Indépendance de variables aléatoires

Définition 4.6 (Indépendance)**.** *Si $X=(X_1,\dots,X_d)$ est un vecteur aléatoire de $\mathbb{R}^d$ alors on dit que les v.a.r. $X_1,\dots,X_d$ sont* indépendantes *lorsque pour tous intervalles $I_1,\dots,I_d$ de $\mathbb{R}$ les événements $\{X_1\in I_d\},\dots,\{X_d\in I_d\}$ sont indépendants.*

Si par exemple A et B sont deux événements alors les v.a.r. booléennes $\mathbf{1}_A$ et $\mathbf{1}_B$ sont indépendantes si et seulement si A et B sont indépendants. Une variable aléatoire constante est proportionnelle à $\mathbf{1}_\Omega$ et est donc indépendante de toutes les autres v.a.r.

Exercice 4.7 (De la loi de Poisson à la loi binomiale)**.** *Soit (X,Y) un couple de v.a. indépendantes de loi de Poisson de paramètres respectifs λ et μ. On suppose que la somme vaut n, et on cherche la distribution de X. Plus précisément, on fixe n et pour tout $k\in\{0,1,\dots n\}$ on cherche $\mathbb{P}[X=k|X+Y=n]$. En utilisant la définition de la probabilité conditionnelle,*

$$\mathbb{P}[X=k|X+Y=n] = \frac{\mathbb{P}[X=k \text{ et } X+Y=n]}{\mathbb{P}[X+Y=n]}.$$

La somme $X+Y$ suit une loi de Poisson de paramètre $\lambda+\mu$. Au numérateur, on réécrit l'événement :

$$\{X=k \text{ et } X+Y=n\} = \{X=k \text{ et } Y=n-k\}.$$

Comme X et Y sont indépendantes, les événements $X=k$ et $Y=n-k$ sont indépendants. On en déduit

$$\begin{aligned}\mathbb{P}[X=k|X+Y=n] &= \frac{\mathbb{P}[X=k]\,\mathbb{P}[Y=n-k]}{\mathbb{P}[X+Y=n]}\\ &= \frac{e^{-\lambda}\lambda^k/(k!)e^{-\mu}\mu^{n-k}/(n-k)!}{e^{-\lambda-\mu}(\lambda+\mu)^n/(n!)}\\ &= \binom{n}{k}p^k(1-p)^{n-k},\end{aligned}$$

où $p=\lambda/(\lambda+\mu)$. On retrouve l'expression d'une loi binomiale : on dit que pour tout n, la loi conditionnelle de X sachant $\{X+Y=n\}$ est la loi $\mathrm{Binom}(n,p)$.

Théorème 4.8 (Espérance et indépendance – Admis)**.** *Si $X=(X_1,\dots,X_d)$ un vecteur aléatoire de $\mathbb{R}^d$ alors $X_1,\dots,X_d$ sont indépendantes si et seulement si pour toutes fonctions boréliennes positives $\varphi_1,\dots,\varphi_d:\mathbb{R}\to\mathbb{R}_+$, on a, dans $\mathbb{R}_+\cup\{+\infty\}$,*

$$\mathbb{E}\left[\prod_{i=1}^d \varphi_i(X_i)\right] = \prod_{i=1}^d \mathbb{E}\left[\varphi_i(X_i)\right].$$

De plus, si $X_1,\dots,X_d$ sont indépendantes alors pour toutes fonctions $\varphi_1,\dots,\varphi_d:\mathbb{R}\to\mathbb{R}$ boréliennes vérifiant $\varphi_1(X_1),\dots,\varphi_d(X_d)\in L^1(\Omega,\mathscr{F},\mathbb{P})$, on a

$$\prod_{i=1}^d \varphi_i(X_i)\in L^1(\Omega,\mathscr{F},\mathbb{P}) \quad \text{et} \quad \mathbb{E}\left[\prod_{i=1}^d \varphi_i(X_i)\right] = \prod_{i=1}^d \mathbb{E}\left[\varphi_i(X_i)\right].$$

En particulier $\mathbb{E}[X_1\cdots X_d]=\mathbb{E}[X_1]\cdots\mathbb{E}[X_d]$ si $X_1,\dots,X_d$ sont dans $L^1(\Omega,\mathscr{F},\mathbb{P})$.

Probabilités - Préparation à l'agrégation interne de mathématiques (47/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Démonstration. Si la première propriété est vraie, alors son application à des fonctions de la forme $\varphi_i = \mathbf{1}_{I_i}$ où I_i est un intervalle de $\mathbb{R}$ fournit l'indépendance de X_1,..., X_d. Réciproquement, on procède par approximation, linéarité de l'espérance, et convergence monotone à partir de ces fonctions élémentaires. Pour établir la seconde propriété, on utilise la $\varphi_i = (\varphi_i)_+ - (\varphi_i)_-$, la linéarité de l'espérance, et la première propriété. □

Théorème 4.9 (Indépendance et structure produit des densités – Admis)**.**

1. *Si les v.a.r. X_1,..., X_d sont indépendantes de densités $f_1, \dots, f_d$ alors le vecteur $X = (X_1, \dots, X_d)$ admet la densité $x \mapsto (f_1 \otimes \cdots \otimes f_d)(x) = f_1(x_1) \cdots f_d(x_d)$*
2. *Les composantes X_1,..., X_d d'un vecteur aléatoire X de $\mathbb{R}^d$ de densité f de marginales $f_1, \dots, f_d$ sont indépendantes ssi X admet aussi $f_1 \otimes \cdots \otimes f_d$ comme densité.*

Démonstration. Découle de la définition de l'indépendance et du théorème de Fubini-Tonelli. □

Exemple 4.10 (Loi de Cauchy)**.** *Soient X et Y des v.a.r. indépendantes de loi normale $\mathcal{N}(0,1)$. On pose $C = X/Y$. Pour trouver la loi de C, on peut utiliser le théorème du transfert. Soit φ l'indicatrice d'un segment $[a,b]$: essayons de mettre $\mathbb{E}\left[\phi(C)\right]$ sous la forme $\int_a^b f(u)du = \int f(u)\varphi(u)du$ pour une densité de probabilité f à déterminer. Comme X et Y sont indépendantes, le couple (X,Y) admet la densité $(1/(2\pi))\exp(-(x^2+y^2)/2)$, donc*

$$\begin{aligned}\mathbb{E}\left[\varphi(C)\right] = \mathbb{E}\left[\varphi(X/Y)\right] &= \frac{1}{2\pi}\iint_{\mathbb{R}^2} \varphi(x/y)e^{-\frac{x^2+y^2}{2}}dxdy \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty}\int_0^{\infty} \varphi(x/y)e^{-\frac{x^2+y^2}{2}}dxdy.\end{aligned}$$

On fait le changement de variables $(u,v) = (x/y, y)$. C'est un difféomorphisme de $\mathbb{R}\times]0,\infty[$. Son inverse est $(x,y) = (uv, v)$, de matrice jacobienne $\begin{pmatrix} v & u \\ 0 & 1 \end{pmatrix}$. Par changement de variables on a donc :

$$\mathbb{E}\left[\varphi(C)\right] = \frac{1}{\pi}\int_0^{\infty}\int_{-\infty}^{\infty} \varphi(u)e^{-\frac{u^2v^2+v^2}{2}}vdudv.$$

En intégrant sur v (toujours le théorème de Fubini), on obtient :

$$\mathbb{E}\left[\varphi(C)\right] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{1+u^2}\varphi(u)du.$$

On a donc bien écrit $\mathbb{P}\left[C \in [a,b]\right] = \mathbb{E}\left[\varphi(C)\right] = \int_a^b f(u)du$ où $f(u) = \frac{1}{\pi(1+u^2)}$ est la densité d'une loi de Cauchy. Par conséquent C suit une loi de Cauchy. On en déduit que si C suit une loi de Cauchy, c'est également le cas pour $1/C$.

Remarque 4.11 (Simulation de la loi normale)**.** *Pour simuler une loi normale, la méthode générale d'inversion vue dans le théorème 3.19 n'est pas commode car la fonction de répartition n'a pas d'expression explicite. Plusieurs algorithmes alternatifs existent, parmi lesquels l'algorithme polaire de Box-Muller. Soit (X,Y) un vecteur aléatoire de $\mathbb{R}^2$ de coordonnées polaires (r,θ). Alors X et Y sont indépendantes de loi $\mathcal{N}(0,1)$ si et seulement si r et θ sont indépendantes avec $r^2 \sim \mathrm{Exp}(1/2) = \mathrm{Gamma}(1,1/2) = \chi^2(2)$ et $\theta \sim \mathrm{Unif}([0,2\pi])$. En effet :*

$$\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}dxdy = re^{-\frac{r^2}{2}}\mathbf{1}_{\mathbb{R}_+}(r)\frac{1}{2\pi}\mathbf{1}_{[0,2\pi]}(\theta)drd\theta.$$

On obtient d'emblée deux réalisations indépendantes de $\mathcal{N}(0,1)$ c'est-à-dire une réalisation de $\mathcal{N}(0,I_2)$.

Pour des raisons de performance et de précision, certains logiciels utilisent plutôt une méthode de discrétisation-rejet (algorithme du Ziggurat de Marsaglia).

Théorème 4.12 (Indépendance et convolution pour le cas à densité)**.** *Si X,Y sont deux variables aléatoires indépendantes de densités f et g alors $X+Y$ admet la densité*

$$z \in \mathbb{R} \mapsto (f * g)(z) = \int_{-\infty}^{+\infty} f(z-y)g(y)\,dy.$$

Probabilités - Préparation à l'agrégation interne de mathématiques (48/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Démonstration. Comme X et Y sont indépendantes, le couple (X,Y) admet pour densité la fonction produit $(x,y) \mapsto f(x)g(y)$. Le théorème du transfert pour le couple (X,Y) donne alors pour tout $t \in \mathbb{R}$,

$$\mathbb{P}[X+Y \leq t] = \int\int \mathbf{1}_{\{(x,y):x+y\leq t\}} f(x)g(y)\,dxdy.$$

En effectuant le changement de variable $(x,y) \mapsto (z,y)$ avec $z = x+y$ il vient

$$\mathbb{P}[X+Y \leq t] = \int\int \mathbf{1}_{\{z\leq t\}} f(z-y)g(y)\,dydz.$$

Le théorème de Fubini-Tonelli donne à présent

$$\mathbb{P}[X+Y \leq t] = \int_{-\infty}^{t}\left(\int_{-\infty}^{+\infty} f(z-y)g(y)\,dy\right)dz. \qquad \square$$

Corollaire 4.13 (Pas d'égalité pour les couples denses)**.** *Si (X_1,X_2) est un couple de v.a. indépendantes à densité, alors $\mathbb{P}[X_1 = X_2] = 0$: avec probabilité 1, l'une des variables est strictement inférieure à l'autre.*

Plus généralement, si $(X_1,\dots,X_n)$ est un vecteur de v.a. indépendantes à densité, alors avec probabilité 1, il existe une permutation π de $\{1,\dots,n\}$ telle que $X_{\pi(1)} < X_{\pi(2)} < \cdots < X_{\pi(n)}$.

Démonstration. Les variables X_1 et $-X_2$ sont indépendantes et à densité, le théorème implique donc que $X_1 - X_2$ est à densité. Par conséquent $\mathbb{P}[X_1 - X_2 = 0] = 0$. Pour la généralisation, on majore la probabilité que deux variables soient égales par la somme sur tous les couples (i,j) de $\mathbb{P}\left[X_i = X_j\right]$. $\square$

Exemple 4.14 (Somme de lois de Cauchy)**.** *. Si X et Y sont indépendantes et suivent des loi de Cauchy de paramètres a et b (*i.e. $f_X(x) = \frac{1}{\pi}\frac{a}{a^2+x^2}$*), $X+Y$ admet la densité*

$$f_{X+Y}(z) = \frac{ab}{\pi^2}\int_{\mathbb{R}} \frac{1}{a^2+(z-y)^2}\cdot\frac{1}{b^2+y^2}\,dy.$$

On peut montrer[1] *que cette intégrale vaut*

$$f_{X+Y}(z) = \frac{1}{\pi}\frac{(a+b)}{(a+b)^2+z^2}.$$

Par conséquent, $X+Y$ suit la loi de Cauchy de paramètre $a+b$.

Remarque 4.15 (Statistique d'ordre et vecteur des rangs)**.** *Soient $X_1,\dots, X_n$ des v.a.r. indépendantes et de même loi admettant une densité. L'hypothèse de densité entraîne* via *le corollaire 4.13 qu'avec probabilité 1, il existe une unique permutation (aléatoire) π_X, élément du groupe symétrique $\mathscr{S}_n$, telle que $X_{\pi_X(1)} < \cdots < X_{\pi_X(n)}$. On parle de* statistique d'ordre de l'échantillon $X_1,\dots, X_n$. *En particulier*

$$X_{\pi_X(1)} = \min(X_1,\dots,X_n) \quad \textit{et} \quad X_{\pi_X(n)} = \max(X_1,\dots,X_n).$$

On dit que le vecteur $(\pi_X^{-1}(1),\dots,\pi_X^{-1}(n))$ est le vecteur des rangs de l'échantillon $X_1,\dots, X_n$. *La loi du vecteur X est échangeable, c'est-à-dire que les vecteurs aléatoires X et $X_\sigma = (X_{\sigma(1)},\dots,X_{\sigma(n)})$ ont la même loi quelque soit $\sigma \in \mathscr{S}_n$, d'où*

$$\mathbb{P}[\pi_X = \sigma] = \mathbb{P}\left[X_{\sigma(1)} < \cdots < X_{\sigma(n)}\right] = \mathbb{P}[X_1 < \cdots < X_n].$$

Le membre de droite ne dépend pas du choix de $\sigma \in \mathscr{S}_n$ et vaut donc $1/\mathrm{card}(S_n) = 1/n!$. Ainsi, σ suit la loi uniforme sur $\mathscr{S}_n$. Alternativement, il est également possible d'observer que $\pi_{X_\sigma} = \pi_X \circ \sigma$ quelque soit $\sigma \in \mathscr{S}_n$, donc que la loi de π_X est invariante par toute translation, et donc que π_X suit la loi uniforme sur $\mathscr{S}_n$ (exemple de la section 2.3). Il se trouve que π_X et la statistique d'ordre X_{π_X} sont indépendantes, car pour tout borélien $A \subset \{(x_1,\dots,x_n) \in \mathbb{R}^n : x_1 < \cdots < x_n\}$ et toute permutation $\sigma \in \mathscr{S}_n$,

$$\mathbb{P}\left[\pi_X = \sigma, X_{\pi_X} \in A\right] = \mathbb{P}\left[X_{\sigma(1)} < \cdots < X_{\sigma(n)}, X_\sigma \in A\right] = \frac{1}{n!}\mathbb{P}\left[X \in \{A_{\sigma'} : \sigma' \in \mathscr{S}_n\}\right].$$

1. Avec les méthodes au programme, on peut utiliser une décomposition en éléments simples. La méthode des résidus (hors-programme) fournit la réponse plus rapidement. La méthode la plus simple pour établir le résultat n'est pas de passer par la densité, mais d'utiliser les fonctions caractéristiques (hors-programme).

Probabilités - Préparation à l'agrégation interne de mathématiques (49/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

4.3 Moyenne et matrice de covariance

Couples de variables, covariance

Un vecteur aléatoire de taille 2 est appelé **couple** de variables alétoires.

Définition 4.16 (Covariance). *La **covariance** de $X, Y \in L^2(\Omega, \mathscr{F}, \mathbb{P})$ est définie par*

$$\mathrm{Cov}(X,Y) := \mathbb{E}\left[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\right] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

Cette définition a bien un sens : l'inégalité de Cauchy-Schwarz montre que la v.a.r. $(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])$ est intégrable et que l'on a

$$\mathbb{E}\left[|(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])|\right] \le \sigma(X)\sigma(Y).$$

La covariance est symétrique : $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$. Elle est reliée à la variance par $\mathrm{Cov}(X,X) = \sigma^2(X)$, et on a la formule bilinéaire

$$\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y) + 2\,\mathrm{Cov}(X,Y).$$

Si X, Y sont indépendantes alors $\mathrm{Cov}(X,Y) = 0$ (on dit que X, Y sont non corrélées ou décorrélées) et

$$\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y).$$

La réciproque est fausse : $\mathrm{Cov}(X,Y) = 0$ n'implique pas que X et Y sont indépendantes (on verra plus loin une réciproque dans un cas gaussien). Contre exemple : $\mathrm{Cov}(U,U^2) = \mathbb{E}\left[U^3\right] - \mathbb{E}[U]\,\mathbb{E}\left[U^2\right] = 0$ si U est uniforme sur $[-1,1]$.

Plus généralement, considérons une suite X_1, ..., X_n de variables dans $L^1(\Omega, \mathscr{F}, \mathbb{P})$. Par linéarité de l'espérance (*nul besoin d'indépendance*), on a

$$\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]$$

Si X_1, ..., X_n sont dans $L^2(\Omega, \mathscr{F}, \mathbb{P})$ alors

$$\sigma^2(X_1 + \cdots + X_n) = \sigma^2(X_1) + \cdots + \sigma^2(X_n) + 2\sum_{1\le k<l\le n} \mathrm{Cov}(X_k, X_l).$$

Si maintenant X_1,..., X_n sont *non-corrélées* (en particulier si elles sont *indépendantes*), les termes de covariance disparaissent et l'on a :

$$\sigma^2(X_1 + \cdots + X_n) = \sigma^2(X_1) + \cdots + \sigma^2(X_n).$$

Mais la variance est *quadratique*, et non pas linéaire : $\sigma^2(aX+b) = a^2\sigma^2(X)$. Voyons comment appliquer ces propriétés aux lois binomiale et hypergéométrique.

Pile ou face : Calcul des moments. Lorsque l'on tire n boules avec remise dans une urne contenant N_1 boules ocre et N_2 boules indigo, si l'on note $X_k = \mathbf{1}_{\text{la } k^{\text{e}} \text{ boule est ocre}}$, alors les $(X_k)_{k=1,\dots,n}$ sont i.i.d. de loi de Bernoulli de paramètre $p = N_1/(N_1+N_2)$. Le nombre S_n de boules ocre tirées suit la loi binomiale $\mathrm{Binom}(n,p)$. On retrouve l'espérance de la loi binomiale, calculée « à la main » précédemment : $\mathbb{E}[S_n] = \sum_{k=1}^n \mathbb{E}[X_k] = np$. Comme les $(X_k)_{k=1,\dots,n}$ sont indépendantes,

$$\sigma^2(S_n) = \sum_{k=1}^n \sigma^2(X_k) = np(1-p).$$

Sondage simple : Calcul des moments. On reprend l'exemple ci-dessus en considérant des tirages *sans remise*. On retrouve le cadre du sondage simple. Avec les mêmes notations, les variables $(X_k)_{k=1,\dots,n}$ sont toujours de loi de Bernoulli de paramètre $p = N_1/(N_1+N_2)$ mais elles ne sont plus indépendantes. Ceci ne change pas le calcul de l'espérance : on retrouve $\mathbb{E}[S_n] = np$. Pour la variance, on applique la formule

$$\sigma^2(S_n) = \sum_{k=1}^n \sigma^2(X_k) + 2\sum_{k<l} \mathrm{Cov}(X_k, X_l).$$

Pour $k<l$, $\mathrm{Cov}(X_k, X_l) = \mathbb{E}[X_k X_l] - p^2$. On voit ensuite que $X_k X_l$ est une variable de Bernoulli. En revenant

Probabilités - Préparation à l'agrégation interne de mathématiques (50/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

à l'intérprétation combinatoire, il y a $N!/(N-n)!$ arrangement possibles de n boules parmi N, d'où

$$\begin{aligned}\mathbb{E}[X_k X_l] &= \mathbb{P}[X_k X_l = 1] = \mathbb{P}[\text{les boules } k \text{ et } l \text{ sont ocre}] \\ &= \frac{\frac{N_1(N_1-1)(N-2)!}{(N-n)!}}{\frac{N!}{(N-n)!}} \\ &= \frac{N_1(N_1-1)}{N(N-1)}.\end{aligned}$$

On en déduit facilement $\mathrm{Cov}(X_k, X_l) = -p(1-p)\frac{1}{N-1}$. Finalement

$$\sigma^2(S_n) = np(1-p) - 2\frac{n}{n-1}2\frac{p(1-p)}{N-1} = np(1-p)\frac{N-n}{N-1}.$$

Fléchettes : Décorrélation n'implique pas indépendance. Dans le jeu de fléchettes notons X l'abscisse et Y l'ordonnée. Ces deux variables sont de moyenne nulle, donc $\mathrm{Cov}(X,Y) = \mathbb{E}[XY]$. Cette espérance est donnée par la formule du transfert, et l'intégrale résultante peut se calculer en appliquant le théorème de Fubini :

$$\begin{aligned}\mathbb{E}[XY] &= \iint xy\frac{1}{\pi}\mathbf{1}_{x^2+y^2\le 1}\,dx\,dy \\ &= \frac{1}{\pi}\int_{-1}^{1} x\left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y\,dy\right)dx \\ &= 0.\end{aligned}$$

Les deux variables X et Y sont donc décorrélées.

En revanche, elle ne sont pas indépendantes. Intuitivement, connaître l'abscisse du point d'impact donne une information sur l'ordonnée ; en prenant par exemple $I = J = [\sqrt{2}/2, 1]$, on a

$$\mathbb{P}[(X,Y) \in I \times J] = 0 \neq \mathbb{P}[X \in I]\,\mathbb{P}[Y \in J].$$

Définition 4.17 (Coefficient de corrélation de Pearson)**.** *Si $X, Y \in L^2(\Omega, \mathscr{F}, \mathbb{P})$ avec $\sigma^2(X) > 0$ et $\sigma^2(Y) > 0$ alors on appelle **corrélation** de X et Y la quantité*

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\sigma^2(X)}\sqrt{\sigma^2(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

Le coefficient de corrélation mesure en quelque sorte la dépendance linéaire. En effet, l'inégalité de Cauchy-Schwarz et ses cas d'égalité permet d'établir les propriétés suivantes :

1. $-1 \le \rho(X,Y) \le 1$;
2. $\rho(X,Y) = 1$ si et seulement si $\mathbb{P}[X = aY + b] = 1$ pour des réels $a > 0$ et b ;
3. $\rho(X,Y) = -1$ si et seulement si $\mathbb{P}[X = aY + b] = 1$ pour des réels $a < 0$ et b.

Le cas vectoriel

Définition 4.18 (Vecteur moyenne et matrice de covariance)**.** *Soit X un vecteur colonne aléatoire de $\mathbb{R}^d$. Si ses composantes sont intégrables alors on définit son **vecteur moyenne***

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])^\top.$$

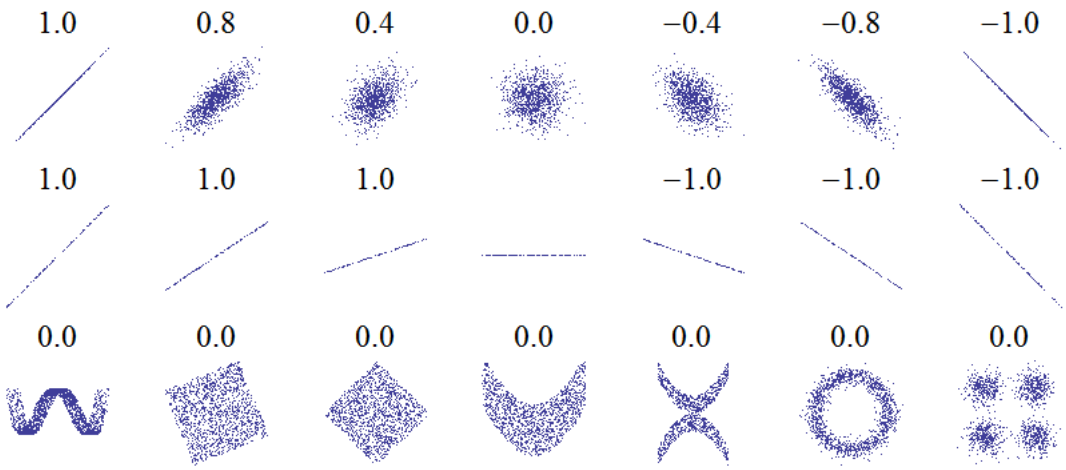
*Si les composantes $X_1, \ldots, X_n$ de X sont de carré intégrable, alors on appelle **matrice de covariance** $\Sigma(X)$ la matrice symétrique $d \times d$ donnée pour tous $i, j \in \{1, \ldots, d\}$ par*

$$\Sigma(X)_{i,j} = \mathrm{Cov}(X_i, X_j).$$

On dit que X est **centré** si $\mathbb{E}[X] = 0$; il est dit **réduit** si $\Sigma(X)$ est la matrice identité. La diagonale de $\Sigma(X)$ est constituée des variances des composantes de X. Si les composantes de X sont indépendantes alors $\Sigma(X)$ est diagonale. La réciproque est fausse en général, mais vraie pour les vecteurs gaussiens, comme nous le verrons dans la section 4.4.

Les deux « résumés » que sont le vecteur moyenne et la matrice de covariance se comportent bien vis-à-vis des transformations affines.

Probabilités - Préparation à l'agrégation interne de mathématiques (51/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



Au dessus de chaque nuage de points, on a indiqué le coefficient de corrélation ρ correspondant.

— La première ligne montre la situation « idéale » : la corrélation forte correspond à des droites, la corrélation nulle à un nuage étalé et la corrélation varie « continûment » avec le nuage.

— La deuxième ligne montre que le signe de la corrélation n'est pas toujours simple à interpréter, puisqu'il varie de -1 à 1 alors que le nuage ne change presque pas.

— Enfin la troisième ligne met en garde contre l'utilisation de la corrélation comme seul indicateur de la dépendance : dans tous les nuages, la corrélation est nulle, pourtant l'abscisse et l'ordonnée sont dans chaque dessin liés par un lien fort, mais non-linéaire.

Figure 4.1 – Corrélation et dépendance (source : Wikipédia).

Théorème 4.19 (Transformations affines)**.** *Soit X un vecteur (colonne) aléatoire de $\mathbb{R}^d$ à composantes de carré intégrable. Si A est une matrice $n \times d$ et b un vecteur de $\mathbb{R}^n$, alors le vecteur aléatoire $AX + b$ de $\mathbb{R}^n$ a pour moyenne et covariance :*

$$\mathbb{E}[AX+b] = A\mathbb{E}[X] + b$$
$$\Sigma(AX+b) = A\Sigma(X)A^\top.$$

La *linéarité* est essentielle : pour une transformation φ générale, $\mathbb{E}\left[\varphi(X)\right] \neq \varphi(\mathbb{E}[X])$.

Démonstration. Par linéarité de l'espérance, on a $\mathbb{E}[AX+b] = A\mathbb{E}[X] + b$. Pour la covariance, on se ramène au cas où $b = 0$ et $\mathbb{E}[X] = 0$ pour lequel on a

$$\Sigma(AX) = \mathbb{E}\left[(AX)(AX)^\top\right] = \mathbb{E}\left[AXX^\top A^\top\right] = A\mathbb{E}\left[XX^\top\right]A^\top = A\Sigma(X)A^\top. \qquad \square$$

Exercice 4.20 (Linéarité)**.** *Montrer que si X est un vecteur (colonne) aléatoire de $\mathbb{R}^d$ à composantes de carré intégrable, centrées, alors pour toute matrice $A \in \mathcal{M}_n(\mathbb{R})$, $\mathbb{E}(X^\top AX) = \mathrm{Trace}(A\Sigma(X))$.*

Rappelons qu'une matrice symétrique $A \in \mathcal{M}_d(\mathbb{R})$ est dite **semi-définie positive** si pour tout $v \in \mathbb{R}^d$, $v^\top Av \geq 0$, et qu'elle est **définie positive** si l'inégalité est stricte pour les vecteurs non-nuls[2].

Théorème 4.21 (Structure des matrices de covariance)**.** *La matrice de covariance d'un vecteur aléatoire est toujours symétrique et semi-définie positive.*

Démonstration. Par le résultat précédent, X et $\mathbb{E}[X]$ ont même covariance et on peut supposer que X est centré : on a donc $\Sigma(X) = \mathbb{E}\left[XX^\top\right]$. Si v est un vecteur colonne de $\mathbb{R}^d$, alors la matrice $vv^\top$ est symétrique, semi-définie positive, de rang 1. Ses valeurs propres sont 0 et $v^\top v = \|v\|_2^2$. Pour tout ω, la matrice aléatoire $(XX^\top)(\omega)$ est symétrique, semi-définie positive, de rang 1. Par conséquent, son espérance Σ est également symétrique. Le fait que Σ soit semi-définie positive découle de la linéarité de l'espérance, car si u est un vecteur colonne de $\mathbb{R}^d$,

$$u^\top \Sigma u = u^\top \mathbb{E}\left[XX^\top\right]u = \mathbb{E}\left[u^\top XX^\top u\right] \geq 0.$$

En revanche, Σ peut être de rang quelconque entre 1 et d, bien que $XX^\top$ soit de rang 1 (il faut concevoir l'espérance comme une combinaison convexe infinie). $\square$

2. On dit parfois simplement « positive » pour « semi-définie positive », au risque de confondre avec la condition $a_{ij} \geq 0$.

Probabilités - Préparation à l'agrégation interne de mathématiques (52/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Remarque 4.22 (Cône)**.** *L'ensemble des matrices $d \times d$ symétriques semi-définies positives est un cône convexe fermé : si A et B en sont deux éléments, alors pour tous réels $\lambda, \mu \geq 0$, $\lambda A + \mu B$ l'est également. La frontière de ce cône est constituée par les éléments du cône qui ne sont pas de plein rang. L'intérieur du cône est le cône convexe ouvert des matrices symétriques définies positives de dimension $d \times d$.*

Nous allons voir que le théorème précedent admet une réciproque. Pour cela nous avons besoin du résultat d'algèbre linéaire suivant.

Théorème 4.23 (Racines carrées matricielles)**.** *Toute matrice symétrique semi-définie positive Σ de dimension $d \times d$ s'écrit $\Sigma = AA^\top$ où A est une matrice de dimension $d \times d$. Une telle matrice A, appelée* ***racine carrée*** *de Σ, n'est pas unique en général.*

Démonstration. Le théorème spectral fournit une matrice diagonale D et une matrice orthogonale P telles que $\Sigma = PDP^\top$. De plus, $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_d)$ où $\{\lambda_1, \ldots, \lambda_d\} \subset \mathbb{R}_+^d$ est le spectre de Σ. Ainsi, $\Sigma = AA^\top$ où $A = P\mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d})$. Un autre choix possible est $A = P\mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d})P^\top$, qui fournit une racine carrée symétrique et semi-définie positive. Dans les deux cas, les matrices Σ et A ont le même rang. Alternativement, la décomposition de Cholesky fournit une matrice triangulaire inférieure A à diagonale positive ou nulle qui vérifie $AA^\top = \Sigma$. Une telle matrice se calcule par un algorithme récursif simple et explicite. En effet, l'équation $AA^\top = \Sigma$ est équivalente au système d'équations suivant : pour tous $1 \leq i \leq j \leq d$

$$A_{i,i} A_{j,i} = \Sigma_{i,j} - \sum_{k=1}^{i-1} A_{i,k} A_{j,k}.$$

La diagonale de Σ est positive ou nulle, strictement positive lorsque Σ est inversible. Dans ce dernier cas, il existe une unique matrice A triangulaire inférieure à diagonale strictement positive telle que $AA^\top = \Sigma$, et la décomposition de Cholesky constitue alors un cas particulier de la décomposition LU des matrices inversibles. □

Théorème 4.24 (Généricité des matrices de covariance)**.** *Toute matrice $\Sigma \in \mathcal{M}_d(\mathbb{R})$ symétrique semi-définie positive est la matrice de covariance d'un vecteur aléatoire de $\mathbb{R}^d$.*

Démonstration. Soit A une racine carrée matricielle de Σ et X un vecteur aléatoire de $\mathbb{R}^d$ dont les composantes sont indépendantes centrées et réduites. La matrice de covariance de X est I_d. Le vecteur aléatoire AX est centré, de matrice de covariance $AI_dA^\top = \Sigma$. □

Le produit de Schur-Hadamard $A \circ B$ de deux matrices A et B de dimension $d \times d$ est le produit terme à terme défini par $(A \circ B)_{i,j} = A_{i,j} B_{i,j}$ pour tout $1 \leq i, j \leq d$. Les considérations précédentes fournissent une preuve probabiliste du résultat d'algèbre linéaire suivant :

Corollaire 4.25 (Schur)**.** *Si A et B sont deux matrices symétriques semi-définies positives de même dimension, alors $A \circ B$ est symétrique semi-définie positive.*

Démonstration. Soient X et Y deux vecteurs aléatoires indépendants et centrés de $\mathbb{R}^d$ tels que $\Sigma(X) = A$ et $\Sigma(Y) = B$. Le vecteur aléatoire Z de $\mathbb{R}^d$ défini par $Z_i = X_i Y_i$ pour tout $1 \leq i \leq d$ est centré, et sa matrice de covariance est donnée par $A \circ B$ car X et Y sont indépendants et centrés. Ainsi, la matrice symétrique $A \circ B$ est semi-définie positive en tant que matrice de covariance d'un vecteur aléatoire ! □

4.4 Loi normale multivariée, vecteurs gaussiens

La loi normale unidimensionnelle se généralise en dimension d de la manière suivante.

Définition 4.26 (Loi normale ou loi de Gauss ou loi gaussienne)**.** *Soit $m \in \mathbb{R}^d$ et Σ une matrice symétrique $d \times d$ définie positive. On dit que le vecteur aléatoire $X = (X_1, \ldots, X_d)$ suit la* ***loi normale*** *de moyenne m et de matrice de covariance Σ, et on note $X \sim \mathcal{N}(m, \Sigma)$ lorsque X admet la densité*

$$x \in \mathbb{R}^d \mapsto \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}\langle \Sigma^{-1} x, x \rangle\right).$$

La loi est qualifiée de ***standard*** *lorsque $m = 0$ et $\Sigma = I_d$.*

Remarque 4.27. *Si $X \sim \mathcal{N}(m, \Sigma)$, le vecteur m est bien l'espérance de X, et la matrice Σ est la matrice de covariance $\Sigma(X)$ définie à la section précédente : $\Sigma_{i,j} = \mathrm{Cov}(X_i, X_j)$.*

Probabilités - Préparation à l'agrégation interne de mathématiques (53/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Dans le cas multidimensionnel, dans la notation $\mathcal{N}(m,\Sigma)$ le Σ désigne toujours la matrice de covariance. Dans le cas univarié, le deuxième paramètre peut être soit la variance, soit l'écart-type, suivant les auteurs.

Théorème 4.28 (Corrélation et indépendance dans le cas gaussien)**.** *Si $X_1,\dots,X_d$ sont des variables gaussiennes unidimensionnelles indépendantes, avec $X_i \sim \mathcal{N}(m_i,\sigma_i^2)$, le vecteur $(X_1,\dots,X_d)^\top$ suit la loi normale $\mathcal{N}(m,\mathrm{Diag}(\sigma_1^2,\dots,\sigma_d^2))$.*

Réciproquement, si $X=(X_1,\dots,X_d)$ suit la loi normale $\mathcal{N}(m,\Sigma)$ et si Σ est diagonale, alors la densité de X est produit, et les marginales (X_i) sont indépendantes.

Exercice 4.29 (Somme de gaussiennes indépendantes)**.** *En utilisant le produit de convolution, montrer que si X_1 et X_2 sont deux v.a.r. indépendantes de loi normales $\mathcal{N}(m_1,\sigma_1^2)$ et $\mathcal{N}(m_2,\sigma_2^2)$ alors $X+Y$ suit la loi normale $\mathcal{N}(m_1+m_2,\sigma_1^2+\sigma_2^2)$.*

Exercice 4.30 (Fonctions affines et lois normales)**.** *Soit Σ une matrice $d\times d$ symétrique dont toutes les valeurs propres sont strictement positive. Soit A une matrice telle que $\Sigma = AA^\top$ (obtenue par exemple* via *le théorème spectral ou* via *la décomposition de Cholesky). Établir au moyen du théorème du transfert et d'un changement de variable que si $Z\sim\mathcal{N}(0,I_d)$ alors $AZ+m\sim\mathcal{N}(m,\Sigma)$. En déduire que si $X\sim\mathcal{N}(m,\Sigma)$ et $v\in\mathbb{R}^d$ alors $\langle X,v\rangle\sim\mathcal{N}(\langle m,v\rangle,\langle v,\Sigma v\rangle)$ et en particulier $X_i\sim\mathcal{N}(m_i,\Sigma_{i,i})$ pour tout $1\le i\le d$. Ainsi, les lois marginales d'un vecteur aléatoire de loi normale sont toutes de loi normale.*

Exercice 4.31 (Vecteurs aléatoires à lois marginales gaussiennes et vecteurs gaussiens)**.** *Montrer que si X et Y sont deux v.a.r. indépendantes avec X de loi de Rademacher de paramètre $1/2$ et Y de loi normale $\mathcal{N}(0,1)$ alors la v.a.r. XY suit la loi normale $\mathcal{N}(0,1)$ tandis que la v.a.r. $XY+Y$ ne suit pas la loi normale. En déduire que le vecteur aléatoire (XY,Y) de $\mathbb{R}^2$ a des lois marginales normales mais ne suit pas une loi normale sur $\mathbb{R}^2$. Les fonctions caractéristiques (hors programme) permettent d'établir un résultat positif dans cet esprit en allant au delà des simples lois marginales : si Z est un vecteur aléatoire de $\mathbb{R}^d$ tel que $v\cdot Z = v_1Z_1+\dots+v_dZ_d$ suit une loi normale sur $\mathbb{R}$ pour tout $v\in\mathbb{R}^d$ alors Z est suit une loi normale sur $\mathbb{R}^d$.*

4.5 Fonctions génératrices et variables indépendantes

La puissance des fonctions génératrices (et de leurs analogues continus : transformée de Laplace et fonctions caractéristiques) tient entre autres au résultat suivant.

Théorème 4.32 (Fonction génératrice et somme de variables indépendantes)**.** *Si X et Y sont deux v.a.r.* indépendantes *à valeurs dans $\mathbb{N}$, de fonctions génératrices g_X et g_Y, alors pour tout $s\in]-1,1]$,*

$$g_{X+Y}(s) = g_X(s)g_Y(s).$$

Autrement dit, l'opération « somme de deux variables indépendantes », qui se traduit sur les fonctions de masse ou les densités par une opération de *convolution* souvent difficile à calculer, s'exprime très simplement sur les fonctions génératrices par un *produit*.

Pile ou face : retour sur les moments. Si $(X_1,\dots X_n)$ sont des variables i.i.d. de loi de Bernoulli de paramètre p, la somme $S_n=\sum X_i$ suit la loi binomiale $\mathrm{Binom}(n,p)$. Comme la fonction génératrice de la loi de Bernoulli est $g_X(s)=\mathbb{E}\left[s^X\right]=q+ps$, il vient

$$g_S(s) = (q+ps)^n.$$

On en déduit, pour $n\ge 2$,

$$g_S'(s) = np(q+ps)^{n-1} \qquad \text{et} \qquad g_S''(s) = n(n-1)p^2(q+ps)^{n-2},$$

ce qui permet de retrouver

$$\mathbb{E}[S] = g_S'(1) = np \qquad \text{et} \qquad \sigma^2(X) = g_S''(1)+g_S'(1)(1-g_S'(1)) = npq.$$

Exemple 4.33. *Soit X_1, X_2 les résultats de deux dés équilibrés indépendants. La somme S a pour fonction génératrice :*

$$g_S(s) = g_{X_1}(s)g_{X_2}(s) = \frac{1}{36}\left(\sum_{k=1}^{6} x^k\right)^2 = \frac{1}{36}\sum_{l=2}^{12} a_l x^l$$

Probabilités - Préparation à l'agrégation interne de mathématiques (54/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

où $a_l = \mathbb{P}[S = l] = l - 1$ *pour* $l = 2, 3, \ldots, 7$ *et* $a_l = 13 - l$ *pour* $l = 8, 9, \ldots, 12$. *La somme* S *ne suit donc pas une loi uniforme : obtenir* 7 *est par exemple bien plus probable qu'obtenir* 2 *ou* 12.

Peut-on truquer les deux dés de façon à obtenir une somme S *uniforme ? Supposons que l'on puisse le faire : il existerait deux familles* (a_k) *et* (b_k) *telles que*

$$\left(\sum_{k=1}^{6} a_k x^k\right)\left(\sum_{k=1}^{6} b_k x^k\right) = \sum_{l=2}^{12} x^l.$$

En simplifiant par x^2 *on obtient :*

$$\left(\sum_{k=0}^{5} a_{k+1} x^k\right)\left(\sum_{k=0}^{5} b_{k+1} x^k\right) = \sum_{l=0}^{10} x^l.$$

Les racines du polynôme de droite sont les racines 11e *non-triviales de l'unité :* $\{\exp(2ik\pi/11); 1 \le k \le 10\}$. *Aucune n'est réelle. À gauche, chacun des polynômes est de degré impair à coefficients réels et possède donc au moins une racine réelle : c'est contradictoire.*

Exemple 4.34 (Loi de Poisson)**.** *Si* X *suit une loi de Poisson de paramètre* λ, *sa fonction génératrice est donnée par :*

$$g_X(s) = \sum_{k \ge 0} e^{-\lambda} \frac{\lambda^k}{k!} s^k = \exp(-\lambda(1 - s)).$$

On en déduit immédiatement que la somme de deux variables de Poisson indépendantes de paramètres λ *et* μ *est encore distribuée suivant une loi de Poisson, de paramètre* $\lambda + \mu$.

4.6 Retour sur l'approximation binomiale/Poisson

Dans cette section on revient sur l'approximation de la loi binomiale par une loi de Poisson lorsque n est grand et p petit, en montrant que l'on peut contrôler l'erreur expicitement. Cette section peut être omise en première lecture.

Pour donner un sens précis aux approximations de lois, une approche est d'introduire des *distances entre lois de probabilité*. Dans le cas discret, on utilise souvent la distance en variation totale.

Définition 4.35 (Distance en variation totale)**.** *Si* X, X' *sont des v.a. discrètes à valeurs dans* $\mathbb{N}$, *alors la distance* l^1 *entre les suites* $(\mathbb{P}[X = k])$ *et* $(\mathbb{P}[X' = k])$ *est appelée* ***distance en variation totale****, notée :*

$$\begin{aligned} d_V(X, X') &:= \sum_k \left|\mathbb{P}[X = k] - \mathbb{P}[X' = k]\right| \\ &= 2 - 2\sum_k \min(\mathbb{P}[X = k], \mathbb{P}[X' = k]). \end{aligned}$$

La seconde expression provient de l'égalité $|x - x'| = x + x' - 2\min(x, x')$.

Le théorème suivant donne un sens rigoureux à la fameuse phrase « convergence de la binomiale vers la loi de Poisson quand n est grand, p petit et np raisonnable ».

Théorème 4.36 (Inégalité de poissonisation de Le Cam)**.** *Soit* $S_n = X_1 + \cdots + X_n$ *où* $X_1, \ldots, X_n$ *sont des v.a. indépendantes avec* X_i *de loi de Bernoulli de paramètre* $p_i = \mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0]$ *pour tout* $1 \le i \le n$. *Soit* T_n *une variable de loi de Poisson* $\mathrm{Poi}(p_1 + \ldots + p_n)$. *Alors* $\mathbb{E}[T_n] = \mathbb{E}[S_n]$ *et*

$$d_V(S_n, T_n) \le 2\sum_{k=1}^{n} p_k^2.$$

En particulier, si tous les p_i *sont égaux à* p, *alors* S_n *suit une loi binomiale et l'on a :*

$$\sum_{k=0}^{\infty} \left|\mathbb{P}[S_n = k] - e^{-np}\frac{(np)^k}{k!}\right| \le 2np^2.$$

Ainsi, si $p = p_n$ dépend de n et vérifie $np_n \to \lambda$ (par exemple $p_n = \lambda/n$) alors l'erreur d'approximation entre binomiale et Poisson, quantifiée par la distance en variation totale, est majorée par $2np_n^2 = \mathcal{O}(1/n)$.

Probabilités - Préparation à l'agrégation interne de mathématiques (55/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Preuve du théorème 4.36. Si X est une variable de Bernoulli de paramètre p et T une variable de loi de Poisson de paramètre p, la seconde expression de la variation totale donne

$$d_V(X,T) = 2 - 2\min(1-p, e^{-p}) - 2\min(p, pe^{-p}) = 2 - 2(1-p) - 2pe^{-p} = 2p(1-e^{-p}) \leq 2p^2.$$

Le théorème s'en déduit en écrivant T_n comme somme de n variables de Poisson de paramètres p_k, et en appliquant le lemme suivant :

Lemme 4.37. *Soit (X,Y) et (X',Y') deux couples de variables indépendantes. Alors :*

$$d_V(X+Y, X'+Y') \leq d_V(X,X') + d_V(Y,Y').$$

Pour établir le lemme, on note a_j, b_j, a'_j et b'_j les probabilités d'être égal à j pour les variables aléatoires X, Y, X' et Y'. Ensuite, pour tout k, on écrit

$$\begin{aligned}\left|\mathbb{P}[X+Y=k] - \mathbb{P}\left[X'+Y'=k\right]\right| &= \left|\sum_{(i,j);i+j=k}\left(a_i b_j - a'_i b'_j\right)\right| \\ &\leq \sum_{(i,j);i+j=k}\left(a_i\left|b_j - b'_j\right| + \left|a_i - a'_i\right|b'_j\right).\end{aligned}$$

En sommant sur k, on reconstruit tous les couples $(i,j) \in \mathbb{N}^2$:

$$\begin{aligned}\sum_k \left|\mathbb{P}[X+Y=k] - \mathbb{P}\left[X'+Y'=k\right]\right| &\leq (\sum_i a_i)\sum_j \left|b_j - b'_j\right| + \sum_i \left|a_i - a'_i\right| \sum_j b'_j \\ &\leq d_V(X,X') + d_V(Y,Y').\end{aligned}$$

□

Probabilités - Préparation à l'agrégation interne de mathématiques (56/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Chapitre 5

Théorèmes limites

5.1 Loi faible des grands nombres

Pile ou face : point de vue statistique fréquentiste. On dispose d'une pièce dont on ne sait pas si elle est équilibrée ou non ; elle tombe sur pile avec une probabilité p (inconnue). On lance (n fois ou une infinité de fois) la pièce, et on note X_k le résultat (dans $\{0,1\}$) du k^{e} lancer. L'univers et la tribu sont fixés, mais à chaque $p \in [0,1]$ on peut associer une probabilité $\mathbb{P}_p$ décrivant l'expérience : sous $\mathbb{P}_p$, les variables X_k sont indépendantes, et $\mathbb{P}_p[X_k = 1] = p$.

On cherche alors à *estimer* le paramètre inconnu p au vu d'une réalisation de l'expérience. L'idée naturelle est de choisir comme estimateur de p la fréquence empirique des piles :

$$\frac{S_n}{n} = \frac{1}{n}\sum_{k=1}^{n} X_i.$$

Une première justification de ce choix est qu'en moyenne, il donne la bonne réponse [1] :

$$\mathbb{E}_p\left[\frac{S_n}{n}\right] = \frac{1}{n}\sum_{k=1}^{n} p = p.$$

La loi des grands nombres donne une autre justification : la quantité aléatoire $\frac{S_n}{n}$ approche bien p quand n grandit [2]. Comme $\frac{S_n}{n}$ est formellement une fonction (de Ω dans $\mathbb{R}$), il faut préciser en quel sens cette convergence a lieu.

Théorème 5.1 (Loi faible des grands nombres)**.** *Soient* $(X_n)_{n\geq 1}$ *des v.a.r.* indépendantes*, de* même loi *et* intégrables *; on note* m *leur espérance commune. Alors pour tout* $\varepsilon > 0$*,*

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \varepsilon\right] = 0.$$

On dit aussi que la suite $(n^{-1}(X_1 + \cdots + X_n))_{n\geq 1}$ *converge* ***en probabilité*** *vers* m*.*

Le terme « loi » dans l'expression « loi faible » doit être compris comme dans l'expression « loi de la nature ». La loi faible doit son nom au type de la convergence, car la convergence en probabilité est plus faible que la convergence presque sûre fournie par la « loi forte ».

Preuve partielle. On se place sous l'hypothèse plus forte que les X_i sont de carré intégrable. En utilisant l'inégalité de Bienaymé-Tchebychev (théorème 3.48) pour la v.a.r. $S_n = X_1 + \cdots + X_n - nm$, et le fait que $\sigma^2(S_n) = n\sigma^2(X_1)$ (grâce à l'indépendance), il vient

$$\begin{aligned}\mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \varepsilon\right) &= \mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq n\varepsilon) \\ &\leq \frac{\sigma^2(S_n)}{n^2\varepsilon^2} = \frac{n\sigma^2(X_1)}{n^2}\varepsilon^2 = \mathcal{O}_{n\to\infty}\left(\frac{1}{n}\right).\end{aligned}$$

1. On dit que l'estimateur S_n/n est *sans biais*.
2. On parle de *consistance* de l'estimateur.

Probabilités - Préparation à l'agrégation interne de mathématiques (57/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Le cas général est admis : l'idée de la preuve est d'approcher les X_i par des variables Y_i de carré intégrable, par exemple en tronquant à un niveau K :

$$Y_i = \begin{cases} X_i & \text{si } |X_i| \leq K, \\ -K & \text{si } X_i \leq -K, \\ K & \text{si } X_i \geq K. \end{cases}$$

Il faut ensuite montrer que la somme normalisée des Y_i ne s'éloigne pas trop de celle des X_i. □

Remarque 5.2 (Isotropie et hypothèses de la loi faible)**.** *La loi faible des grands nombre reste valable au delà du cas des variables aléatoires indépendantes équidistribuées. En effet, il suffit par exemple que $\sigma^2(S_n) = o(n^2)$, ce qui signifie que $\frac{S_n}{n} \to 0$ dans L^2 quand $n \to \infty$. C'est le cas par exemple lorsque le vecteur aléatoire $(X_1, \ldots, X_n)$ est isotrope, c'est-à-dire que sa matrice de covariance est I_n, c'est-à-dire que $X_1, \ldots, X_n$ sont non-corrélées et de même variance :*

$$\sigma^2(S_n) = \sum_{1 \leq i,j \leq n} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n} \sigma^2(X_i) = \mathcal{O}(n) = o(n^2).$$

Exemple 5.3 (Hypothèses, (contre-)exemple des lois de Cauchy)**.** *Si la loi des X_n est à support compact alors elles possèdent des moments de tout ordre, et en particulier ces v.a.r. sont intégrables (ceci comprend le cas Bernoulli).*

A contrario, *si $X_1, \ldots, X_n$ sont des v.a.r. indépendantes de loi de Cauchy de paramètre a, leur moyenne empirique $\frac{1}{n}(X_1 + \cdots + X_n)$ suit également la loi de Cauchy (il suffit d'utiliser le résultat de l'exemple 4.14 sur la somme de deux lois de Cauchy, de faire une récurrence, et de vérifier qu'un multiple d'une variable de Cauchy reste de Cauchy). La moyenne empirique reste aléatoire, et la convergence n'a pas lieu : la loi des grands nombres* n'est pas vérifiée *pour cette loi à queue lourde. Cette propriété concerne en réalité toute une classe de lois (par exemple lois de Pareto ou Student).*

Exemple 5.4 (Polynômes de Bernstein et théorème de Weierstrass)**.** *Un théorème célèbre de Weierstrass affirme que pour tout intervalle compact $[a,b] \subset \mathbb{R}$, la restriction des fonctions polynômes $\mathbb{R}[X]$ à $[a,b]$ est dense dans l'ensemble des fonctions continues $\mathscr{C}([a,b],\mathbb{R})$, pour la topologie de la norme uniforme $\|\cdot\|_\infty$. On peut l'établir en utilisant la même idée que dans la preuve de la loi faible des grands nombres.*

On commence grâce à un argument de translation et dilatation par se ramener au cas où $[a,b] = [0,1]$. Fixons f dans $\mathscr{C}([0,1],\mathbb{R})$. On va montrer que les polynômes de Bernstein $(P_n)_{n \geq 1}$ définis par

$$P_n(X) = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) X^k (1-X)^{n-k}$$

convergent uniformément vers f sur $[0,1]$. Fixons d'abord $x \in [0,1]$, et donnons-nous une variable $S_n^x = S_n$ une variable de loi binomiale $\mathrm{Binom}(n,x)$. Par la formule du transfert appliquée à $g : x \mapsto f(x/n)$,

$$\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \mathbb{E}\left[g(S_n)\right] = \sum_{k=0}^{n} \mathbb{P}\left[S_n = k\right] g(k) = P_n(x).$$

Par conséquent, pour tout $x \in [0,1]$ et $n \in \mathbb{N}^*$,

$$f(x) - P_n(x) = \mathbb{E}\left[f(x) - f\left(\frac{S_n}{n}\right)\right].$$

Fixons un $\varepsilon > 0$ arbitrairement petit. Comme f est continue sur l'intervalle compact $[0,1]$, elle est uniformément continue d'après le théorème de Heine : il existe donc $\eta > 0$ tel que $|f(x) - f(y)| \leq \varepsilon$ pour tous $x, y \in [0,1]$ tels que $|x - y| \leq \eta$.

Considérons l'événement $A_n = \left\{\left|\frac{1}{n}S_n^x - x\right| \leq \eta\right\}$. Comme $\mathbb{E}\left[\frac{1}{n}S_n^x\right] = x$ et $\sigma^2(\frac{1}{n}S_n^x) = \frac{1}{n}x(1-x)$, on a par l'inégalité de Bienaymé–Tchebychev :

$$\mathbb{P}\left[A_n^c\right] \leq \frac{x(1-x)}{n\eta^2} \leq \frac{1}{4n\eta^2}.$$

Probabilités - Préparation à l'agrégation interne de mathématiques (58/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Il existe donc un entier N, dépendant de ε mais pas de x, tel que pour tout $n \geq N$, $\mathbb{P}\left[A_n^c\right] \leq \varepsilon$, d'où

$$\begin{aligned}
\left|f(x) - P_n(x)\right| &\leq \mathbb{E}\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right] \\
&= \mathbb{E}\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right| \mathbf{1}_{A_n}\right] + \mathbb{E}\left(\left|f(x) - f\left(\frac{S_n}{n}\right)\right| \mathbf{1}_{A_n^c}\right) \\
&\leq \mathbb{E}(\varepsilon) + \mathbb{E}\left(2\|f\|_\infty \mathbf{1}_{A_n^c}\right) \\
&\leq (1 + 2\|f\|_\infty)\varepsilon.
\end{aligned}$$

Cette borne est uniforme en $x \in [0,1]$, et le théorème est établi.

Remarque : le théorème de Weierstrass permet d'établir, en utilisant la caractérisation de la loi par les fonctions tests continues et bornées, que si X et Y sont deux v.a.r. bornées avec une suite de moments identique, c'est-à-dire que $\mathbb{E}[X^n] = \mathbb{E}[Y^n]$ pour tout $n \geq 0$, alors X et Y ont même loi.

5.2 Loi forte des grands nombres

Théorème 5.5 (Loi forte des grands nombres)**.** *Si $(X_n)_{n\geq 1}$ est une suite de v.a.r. indépendantes et de même loi possédant une espérance m alors*

$$\mathbb{P}\left[\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = m\right] = 1.$$

On dit que la suite $(n^{-1}(X_1 + \cdots + X_n))_{n\geq 1}$ converge ***presque sûrement*** *vers m.*

La preuve de ce résultat est hors-programme. Notons toutefois qu'en plus de son intérêt intrinsèque, elle fournit une application naturelle du lemme de Borel-Cantelli.

La preuve fait appel à trois idées. Les deux premières sont les mêmes que pour la loi faible : une troncature, pour se ramener à des variables bornées, et une inégalité de concentration (du type Bienaymé–Tchebychev) pour contrôler les écarts à la moyenne. Pour établir le résultat presque-sûr, l'idée nouvelle est de faire appel au lemme de Borel-Cantelli. La preuve complète peut se trouver par exemple dans le livre de Paul S. Toulouse [Tou99] ou de Feller [Fel68]. Nous donnons ci-dessous deux preuves rapides lorsque les variables sont bornées (par une constante ou dans L^4) ce qui nous dispense de troncature. Il est également possible d'établir que pour une suite $(X_n)_{n\geq 1}$ de v.a.r. indépendantes et de même loi, la condition d'intégrabilité $\mathbb{E}[|X_1|] < \infty$ (i.e. X_1 possède une espérance) est nécessaire et suffisante pour que la loi forte des grands nombres ait lieu.

Exercice 5.6 (Mesurabilité)**.** *Montrer que l'ensemble $\left\{\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = m\right\} \subset \Omega$ est bien un événement (un élément de $\mathscr{F}$), en la réécrivant avec des unions et intersections dénombrables d'éléments de $\mathscr{F}$.*

Pour la preuve du théorème, quitte à remplacer les X_i par $X_i - \mathbb{E}[X_i]$, on peut supposer que $m = 0$. On pose $S_n = X_1 + \cdots + X_n$. Pour montrer le théorème il suffit d'établir :

$$\mathbb{P}\left[\overline{\lim_n}\{|S_n| > n\varepsilon\}\right] = 0.$$

En effet, ceci implique, pour une suite $\varepsilon_k \searrow 0$ arbitraire, et pour tout k,

$$\mathbb{P}\left[\left(\overline{\lim_n} \frac{|S_n|}{n}\right) > \varepsilon_k\right] = \mathbb{P}\left[\overline{\lim_n}\{|S_n| > n\varepsilon\}\right] = 0.$$

En prenant l'union dénombrable en k, le théorème 2.11 donne $\mathbb{P}\left[\overline{\lim}_n \frac{S_n}{n} > 0\right] = 0$, d'où l'on déduit $\mathbb{P}[\lim S_n/n = 0] = 1$.

Preuve du théorème 5.5 pour des variables bornées. On suppose qu'il existe une constante $C > 0$ telle que $\mathbb{P}[|X_i| \leq C] = 1$ (ne dépend pas de i car les variables ont même loi). La première partie du lemme 2.24 de Borel-Cantelli permet de se ramener à établir que

$$\sum_{n\geq 1} \mathbb{P}[|S_n| \geq \varepsilon n] < \infty$$

Probabilités - Préparation à l'agrégation interne de mathématiques (59/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

pour tout $\varepsilon > 0$ fixé. Or pour tout $r > 0$ et tout entier $n \geq 1$, l'inégalité de Markov donne

$$\mathbb{P}[S_n \geq n\varepsilon] \leq \mathbb{P}\left[e^{rS_n} \geq rn\varepsilon\right] \leq e^{-rn\varepsilon}\mathbb{E}\left[e^{rS_n}\right] = e^{rn\varepsilon}\mathbb{E}\left[e^{rX_1}\right]^n, \tag{5.1}$$

où l'égalité finale provient du fait que les variables $(X_n)_{n\geq 1}$ sont indépendantes et de même loi. À présent, comme $m = \mathbb{E}[X_1] = 0$ et $\mathbb{P}[|X_1| \leq C] = 1$, il vient, en utilisant l'inégalité élémentaire $e^t - t \leq e^{2t^2}$ si $t \in [0, 1/2]$ et $e^t - t \leq e^t \leq e^{2t^2}$ si $t \geq 1/2$,

$$\mathbb{E}\left[e^{rX_1}\right] \leq e^{rC} - rC \leq e^{2r^2C^2}.$$

Le meilleur choix pour le paramètre r est $r = \varepsilon/(4C^2)$; en reportant dans (5.1) il vient :

$$\mathbb{P}[S_n \geq n\varepsilon] \leq e^{rn\varepsilon - 2nr^2C^2} \leq e^{-n\varepsilon^2/(8C^2)}.$$

En combinant ceci à la même inégalité pour les variables $(-X_1)_{n\geq 1}$ on obtient enfin

$$\mathbb{P}[|S_n| \geq n\varepsilon] \leq 2e^{-n\varepsilon^2/(8C^2)}.$$

Le membre de droite est bien le terme général d'une série convergente, comme désiré. □

Cas des variables bornées dans L^4. On suppose que les variables $(X_n)_{n\geq 1}$ sont bornées dans L^4, c'est à dire que $\tau^4 = \mathbb{E}\left[X_1^4\right] < \infty$. On a alors

$$\mathbb{E}\left[S_n^4\right] = n\tau^4 + 3n(n-1)\sigma^4 = \mathcal{O}(n^2).$$

Ainsi, pour tout $\varepsilon > 0$, par l'inégalité de Markov,

$$\sum_n \mathbb{P}(|S_n| > n\varepsilon) = \sum_n \mathbb{P}\left[|S_n|^4 > n^4\varepsilon^4\right] \leq \sum_n \frac{\mathbb{E}\left[S_n^4\right]}{n^4\varepsilon^4} < \infty$$

et le résultat découle à présent de la première partie du lemme 2.24 de Borel-Cantelli. Notons que si $\mathbb{P}[|X_1| \leq C] = 1$ alors la suite $(X_n)_{n\geq 1}$ est bornée dans L^4, et la seconde preuve est donc plus puissante. D'autre part, elle reste valable même si les variables ne sont pas de même loi, pourvu qu'elles soient indépendantes et bornées dans L^4. □

Exercice 5.7 (De la loi forte à la loi faible)**.** *La loi faible (théorème 5.1) découle de la loi forte (théorème 5.5) car $\mathbb{P}\left[n^{-1}|S_n| > \varepsilon\right] \leq \mathbb{P}[A_n]$ avec $A_n = \{\sup_{k\geq n} k^{-1}|S_k| > \varepsilon\}$ et comme la suite $(A_n)_{n\geq 1}$ est croissante, on a $\lim_{n\to\infty}\mathbb{P}[A_n] = \mathbb{P}[\cap_n A_n] = 1$ quand $m = 0$.*

Exemple 5.8 (Marche aléatoire simple sur $\mathbb{Z}$)**.** *La marche aléatoire simple peut modéliser la position d'une particule dans un fluide, et sa version continue, le mouvement Brownien, a été introduite en physique par Einstein et Langevin, et en finance mathématique par Bachelier. Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi de Rademacher de paramètre $p \in [0,1]$, c'est-à-dire que $\mathbb{P}[X_n = 1] = 1 - \mathbb{P}[X_n = -1] = p$ pour tout $n \geq 1$, modélisant les incréments dus au choc avec les particules du fluide. La marche aléatoire simple sur $\mathbb{Z}$ est la suite $(S_n)_{n\geq 1}$ où $S_n = X_1 + \cdots + X_n$. La loi forte des grands nombres entraîne que si $p \neq 1/2$ alors $(S_n)_{n\geq 1}$ diverge vers $\pm\infty$ avec probabilité 1 quand $n \to \infty$.*

Exemple 5.9 (Inégalité de Jensen et loi des grands nombres)**.** *L'inégalité de Jensen affirme que si X est une v.a.r. intégrable et $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction convexe telle que $\varphi(X)$ est intégrable, alors*

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

En particulier, si $\varphi(x) = x^2$ on retrouve $\mathbb{E}[X]^2 \leq \mathbb{E}\left[X^2\right]$, et si $\varphi(x) = |x|$ on retrouve $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

Pour déduire l'inégalité de la loi des grands nombres, on introduit une suite $(X_k)_{1\leq k\leq n}$ de variables indépendantes et de même loi que X. Par définition de la convexité, on a l'inégalité suivante :

$$\varphi\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \leq \frac{1}{n}\sum_{i=1}^n \varphi(X_i).$$

Comme les X_i sont intégrables, indépendantes et de même loi, on peut leur appliquer la loi des grands nombres : $\frac{1}{n}\sum_{i=1}^n X_i$ converge presque-sûrement vers $\mathbb{E}[X]$. Comme φ est convexe, elle est continue, donc $\varphi\left(\frac{1}{n}\sum X_i\right)$ converge presque-sûrement vers $\varphi(\mathbb{E}[X])$. Par une nouvelle application de la loi des grands nombres, cette fois à la suite $(\varphi(X_i))$, le terme de droite tend p.s. vers $\mathbb{E}\left[\varphi(X)\right]$, ce qui achève la preuve.

Probabilités - Préparation à l'agrégation interne de mathématiques (60/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Remarque 5.10 (Convergence monotone ou théorème de Fubini-Tonelli)**.** *Le lecteur familier avec l'intégrale de Lebesgue connaît bien le théorème de convergence monotone : si* $(X_n)_{n\geq 1}$ *est une suite croissante de v.a.r. à valeurs dans* $[0,\infty]$ *alors*

$$\lim_n \mathbb{E}[X_n] = \mathbb{E}\left[\lim_n X_n\right].$$

D'autre part, si X *est une v.a.r. sur* $[0,\infty]$ *vérifiant* $\mathbb{E}[X] < \infty$ *alors* $\mathbb{P}[X < \infty] = 1$*. La première partie du lemme de Borel-Cantelli en découle car*

$$\sum_n \mathbb{P}[A_n] = \sum_n \mathbb{E}\left[\mathbf{1}_{A_n}\right] = \mathbb{E}\left(\sum_n \mathbf{1}_{A_n}\right) = \mathbb{E}\left(\mathbf{1}_{\overline{\lim}_n A}\right).$$

On peut également voir ce résultat comme une conséquence du théorème de Fubini-Tonelli plutôt que comme une application du théorème de convergence monotone. Une autre conséquence du théorème de convergence monotone (ou du théorème de Fubini-Tonelli) est que

$$\sum_n \mathbb{E}[|Y_n|] < \infty \quad \Rightarrow \quad \mathbb{P}\left[\lim_n Y_n = 0\right] = 1.$$

En effet, on a $\mathbb{E}\left[\sum_n |Y_n|\right] = \sum_n \mathbb{E}[|Y_n|] < \infty$ *et donc* $\sum_n |Y_n|$ *est une v.a.r. sur* $[0,\infty]$ *d'espérance finie, et donc finie avec probabilité* 1*, ce qui implique que* $|Y_n|$ *tend vers* 0 *avec probabilité* 1*. Cette observation suggère une preuve alternative du théorème 5.5 :*

$$\mathbb{E}\left(\sum_n \left(\frac{S_n}{n}\right)^4\right) = \sum_n \mathbb{E}\left(\left(\frac{S_n}{n}\right)^4\right) < \infty \quad \textit{d'où} \quad \mathbb{P}\left(\lim_n \frac{S_n}{n} = 0\right) = 1.$$

Remarque 5.11 (Suite)**.** *Soit* $(X_n)_{n\geq 1}$ *une suite de v.a.r. indépendantes de carré intégrable. Si* $\sum_n \mathrm{Var}(X_n) < \infty$ *alors en vertu du théorème de convergence monotone ou du théorème de Fubini-Tonelli positif on a* $\mathbb{P}[\lim_n X_n - \mathbb{E}[X_n] = 0] = 1$*. En particulier, si* $(X_n)_{n\geq 1}$ *sont des v.a.r. i.i.d. centrées de carré intégrable alors*

$$\mathbb{P}\left[\lim_n n^{-1} X_n = 0\right] = 1.$$

Attention, il ne s'agit pas de la loi des grands nombres, qui concerne la quantité $n^{-1}(X_1 + \cdots + X_n)$*.*

Exemple 5.12 (Covariance empirique)**.** *Soit* X *un vecteur colonne aléatoire de* $\mathbb{R}^m$ *centré et de matrice de covariance* Σ*. Soient* $X_1, \ldots, X_n$ *des vecteurs colonne aléatoires de* $\mathbb{R}^m$ *de même loi que* X*. La* matrice de covariance empirique Σ_n *est définie par*

$$\Sigma_n = \frac{1}{n}(X_1 X_1^\top + \cdots + X_n X_n^\top) = \frac{1}{n}\mathbb{X}^\top \mathbb{X}$$

où $\mathbb{X}$ *est la matrice aléatoire* $m \times n$ *dont les lignes sont* $X_1^\top, \ldots, X_n^\top$*. La matrice aléatoire* Σ_n *est symétrique semi-définie positive car combinaison convexe de telles matrices. On a*

$$\mathbb{E}[\Sigma_n] = \Sigma$$

et la loi forte des grands nombres indique que Σ_n *converge entrée par entrée vers* Σ*, avec probabilité* 1*. Comme l'ensemble des matrices inversible est ouvert, si* Σ *est inversible, alors avec probabilité* 1*, la matrice* Σ_n *est inversible pour* n *assez grand.*

5.3 Théorème limite central

La première version de ce théorème semble être dûe à Abraham de Moivre. Il a été développé plus tard par Laplace, par Tchebychev puis par Lyapunov notamment. Georg Pólya publie en 1920 un article intitulé « Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem », où il qualifie ce « théorème limite » (Grenzwertsatz) de « zentral » (primordial, principal) en raison de son caractère universel. Si la traduction anglaise semble fixée (« central limit theorem » souvent abrégé en CLT), l'usage francophone, partant de la traduction littérale « théorème limite central » (TLC), attribue souvent la centralité à la limite en parlant de « théorème de la limite centrale », allant parfois même jusqu'à l'audacieux « théorème central limite » (TCL).

Probabilités - Préparation à l'agrégation interne de mathématiques (61/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Énoncé du théorème

On cherche à estimer l'espérance m d'une variable aléatoire X au vu d'un échantillon $X_1,\dots, X_n$ de répétitions indépendantes de X. La loi des grands nombres montre que la moyenne empirique

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}$$

est un estimateur consistant et sans biais de m. Ce résultat ne suffit pas en pratique : on a besoin de contrôler l'erreur faite en estimant m par $\overline{X}_n$.

Cette erreur $E_n = \overline{X}_n - m$ est de moyenne nulle. On peut aisément calculer sa variance :

$$\sigma^2(E_n) = \sigma^2(X)/n.$$

L'écart-type de l'erreur est donc *proportionnel à* $1/\sqrt{n}$.

Pour étudier E_n plus finement, on la normalise en la divisant par son écart-type et l'on pose

$$Z_n = \frac{E_n}{\sigma(E_n)} = \frac{\sqrt{n}}{\sigma(X)}\left(\frac{S_n}{n} - m\right).$$

L'erreur normalisée Z_n est alors, par définition, « centrée » (de moyenne nulle) et « réduite » (d'écart-type 1). Si (X_n) est « raisonnable », cette erreur normalisée converge :

Théorème 5.13 (Théorème limite central)**.** *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi, de variance non nulle et finie σ^2 et de moyenne m. Alors pour tout intervalle I de $\mathbb{R}$, l'erreur normalisée $Z_n = (\frac{S_n}{n} - m)/(\sigma/\sqrt{n})$ vérifie*

$$\mathbb{P}\left[Z_n \in I\right] \xrightarrow[n\to\infty]{} \int_I \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \ ;$$

le terme de droite n'est autre que $\mathbb{P}\left[Z \in I\right]$ pour $Z \sim \mathcal{N}(0,1)$. On dit que Z_n ***converge en loi*** *vers Z.*

Remarque 5.14 (Universalité)**.** *Ce résultat ne dépend de la loi des X_i qu'à travers leur moyenne et leur variance (qui déterminent la manière de normaliser l'erreur) : la loi limite est universelle.*

Notons aussi que si les X_i suivent une loi normale $\mathcal{N}(0,1)$, l'erreur normalisée Z_n suit exactement *la loi normale $\mathcal{N}(0,1)$ — en ce sens, le théorème central limite ressemble à un théorème de point fixe.*

Bien que la preuve du TLC soit hors-programme, nous en donnerons une plus bas.

Illustrations et applications

Soit $\alpha \in]0, 1/2[$ un réel — α s'interprétera comme un risque d'erreur, les choix typiques étant $\alpha = 0.05$ et $\alpha = 0.01$. Si Z suit une loi normale $\mathcal{N}(0,1)$, et si l'on note q_α le quantile d'ordre $1-\alpha/2$ de la loi normale, alors

$$\mathbb{P}\left[-q_\alpha \leq Z \leq q_\alpha\right] = 1 - \alpha.$$

Il y a donc une probabilité $1-\alpha$ que la variable aléatoire Z tombe dans l'intervalle (déterministe) $[\pm q_\alpha]$ — on dit parfois que $I_F(\alpha) := [-q_\alpha, q_\alpha]$ est un **intervalle de fluctuation** au niveau $1-\alpha$ de la variable Z.

Le théorème limite central implique

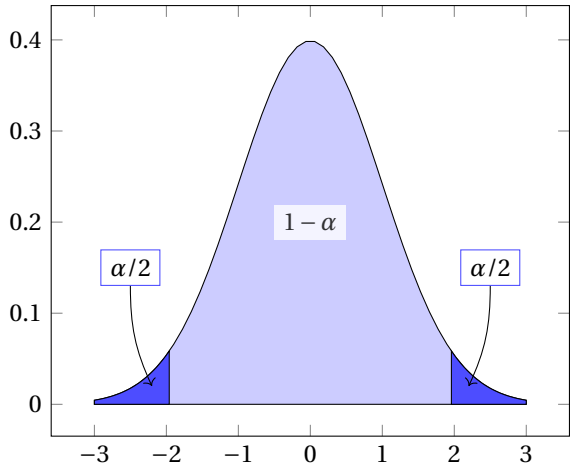
$$\mathbb{P}\left[Z_n \in I_\alpha\right] \xrightarrow[n\to\infty]{} 1 - \alpha.$$

Pile ou face : fluctuations. On jette n fois une pièce équilibrée et on note S_n le nombre de piles. La loi (forte) des grands nombres dit que si n tend vers l'infini, S_n/n tend vers $1/2$. Pour préciser cela utilisons le théorème limite central : si l'on admet que n est assez grand pour que $\mathbb{P}\left[Z_n \in I\right]$ soit proche de sa limite, on a

$$\begin{aligned} 1 - \alpha &= \mathbb{P}\left[Z \in I_\alpha\right] \approx \mathbb{P}\left[Z_n \in I_F(\alpha)\right] \\ &= \mathbb{P}\left[S_n \in \left[mn - q_\alpha \frac{\sigma}{\sqrt{n}}, mn + q_\alpha \frac{\sigma}{\sqrt{n}}\right]\right]. \end{aligned}$$

Comme $m = 1/2$ et $\sigma = 1/4$, on trouve pour $n = 1000$ et $\alpha = 5\%$: $\mathbb{P}\left[S_{1000} \in [500 \pm 15.49]\right] \approx 0.95$.

Probabilités - Préparation à l'agrégation interne de mathématiques (62/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



Densité de la loi normale centrée réduite. L'aire totale sous la courbe vaut 1. Pour obtenir une région centrée $[-q_\alpha, q_\alpha]$ d'aire $1-\alpha$ il faut choisir q_α comme le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale.

FIGURE 5.1 – Quantile et fluctuation de la loi normale standard

Dans le cas précédent, on a supposé l'expérience parfaitement connue (on sait que la pièce est équilibrée) et on a cherché à calculer les probabilités de divers événements — c'est le point de vue *probabiliste*.

L'utilisation la plus courante du théorème correspond au point de vue *statistique*. Nous avons vu précédemment que la loi des grands nombres donnait une justification de l'utilisation de la moyenne empirique S_n/n comme estimateur de l'espérance (inconnue) d'une variable X. Le TLC peut alors donner une idée de l'erreur d'estimation. Ce renversement de point de vue consiste à écrire

$$\begin{aligned} Z_n(\omega) \in I_F(\alpha) &\Leftrightarrow |Z_n(\omega)| \leq q_\alpha \\ &\Leftrightarrow \left| \frac{S_n(\omega)/n - m}{\sigma/\sqrt{n}} \right| \leq q_\alpha \\ &\Leftrightarrow \left| \frac{S_n(\omega)}{n} - m \right| \leq q_\alpha \frac{\sigma}{\sqrt{n}} \\ &\Leftrightarrow m \in \left[\frac{S_n(\omega)}{n} \pm q_\alpha \frac{\sigma}{\sqrt{n}} \right]. \end{aligned}$$

La première ligne correspond à l'appartenance d'une quantité aléatoire à un intervalle déterministe ; dans la dernière on regarde l'appartenance du réel m (déterministe) à un intervalle aléatoire.

Cet intervalle aléatoire

$$I_C(\alpha) = \left[\overline{X}_n - \frac{\sigma q_\alpha}{\sqrt{n}}, \overline{X}_n + \frac{\sigma q_\alpha}{\sqrt{n}} \right]$$

est appelé **intervalle de confiance**[3] pour l'estimation de m au niveau $1-\alpha$, et l'on a

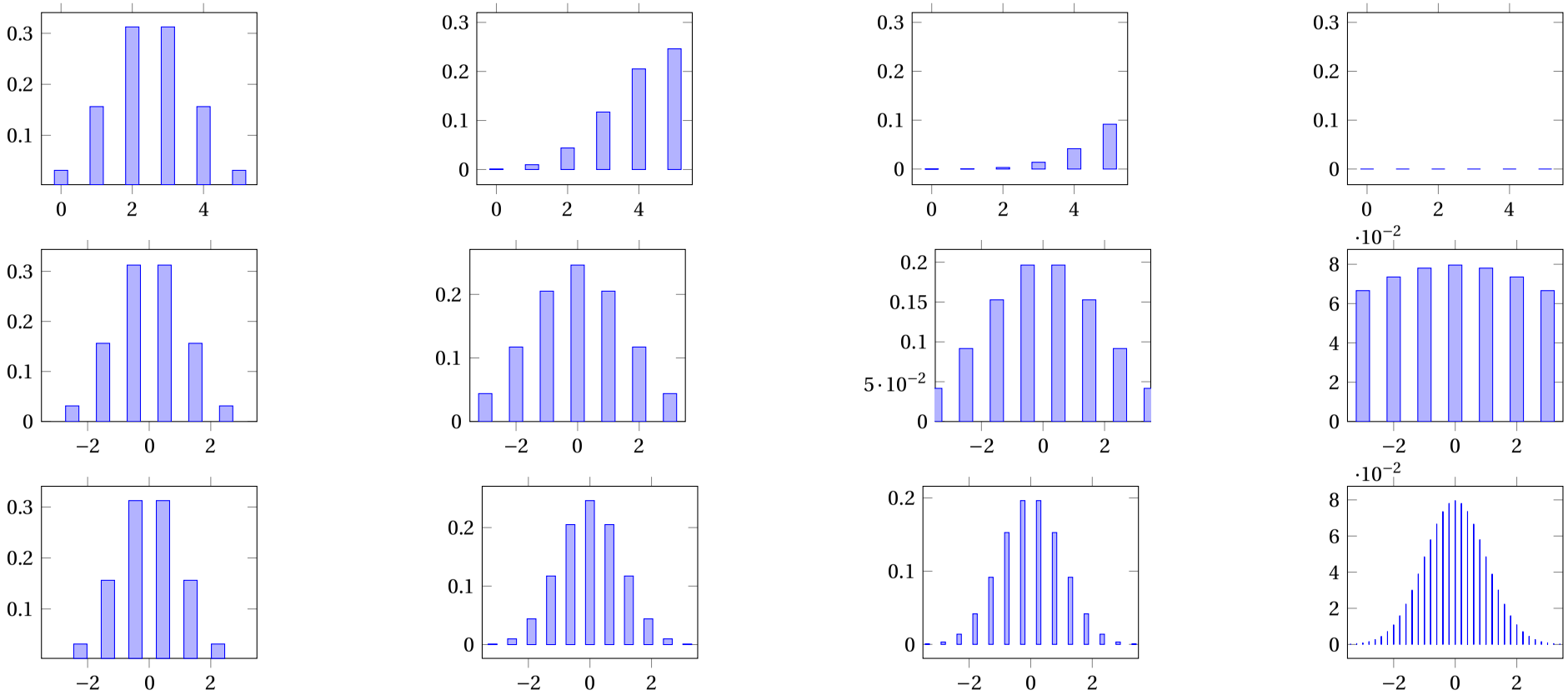
$$\mathbb{P}\left[m \in I_C(\alpha)\right] = \mathbb{P}\left[Z_n \in I_F(\alpha)\right] \xrightarrow[n\to\infty]{} 1-\alpha.$$

L'intervalle $I_C(\alpha)$ est de largeur $2\sigma q_\alpha n^{-1/2}$, qui tend vers 0 si $n \to \infty$, et vers ∞ si $\alpha \to 0$. Cependant, l'intervalle de confiance $I_{n,\alpha}$ n'est calculable que si σ est connu. Comme ce n'est en général pas le cas, il faut remplacer σ par un estimateur, et travailler un peu plus pour montrer que l'intervalle de confiance reste justifié[4]. Dans le cas particulier de variables de Bernoulli on peut cependant se passer de cette dernière étape.

3. On parle plus exactement d'intervalle de confiance asymptotique : l'intervalle contient le paramètre avec une probabilité qui tend vers $1-\alpha$.

4. Le résultat permettant cette justification est le lemme de Slutsky, il est présenté dans les compléments, p. 100. Dans le cas de variables de Bernoulli, on « remplace » donc σ par son estimateur $\sqrt{\overline{X_n}(1-\overline{X}_n)}$ pour obtenir l'intervalle dit de Wald : $I_C^{\text{Wald}}(\alpha) = \left[\overline{X}_n \pm \frac{q_\alpha \sqrt{\overline{X}_n(1-\overline{X}_n)}}{\sqrt{n}} \right]$. Cet intervalle est plus petit (et donc meilleur) que I'_C .

Probabilités - Préparation à l'agrégation interne de mathématiques (63/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



Première ligne : les diagrammes en bâtons de la loi binomiale pour $p = 1/2$ et $n = \ldots$; les échelles en x et y sont fixées. La masse s'échappe vers la droite.
Deuxième ligne : la loi de la variable centrée $S_n - np$, l'échelle en x est fixée. Le diagramme s'aplatit et on ne voit qu'un fraction de plus en plus petite de la masse.
Troisième ligne : la loi de la variable centrée et réduite $(S_n - np)/\sqrt{np(1-p)}$.

Figure 5.2 – Diagramme en bâtons de la binomiale

Probabilités - Préparation à l'agrégation interne de mathématiques (64/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Pile ou face : intervalle de confiance. Les variables (X_i) suivent une loi de Bernoulli de paramètre p inconnu, pour lequel on cherche un intervalle de confiance. L'écart-type $\sigma = \sqrt{p(1-p)}$ est inconnu, mais comme $p \in [0,1]$, il ne dépasse pas $1/2$. Par conséquent,

$$I'_C(\alpha) := \left[\overline{X}_n - \frac{q_\alpha}{2\sqrt{n}}, \overline{X}_n + \frac{q_\alpha}{2\sqrt{n}}\right] \supset I_C(\alpha),$$

ce qui donne finalement

$$\mathbb{P}\left[m \in I'_C(\alpha)\right] \geq \mathbb{P}\left[m \in I_C(\alpha)\right] = 1 - \alpha.$$

Le fameux intervalle de confiance à « 95% » s'obtient avec $\alpha = 0.05$ et $q_{0.05} = 1.96$.

Exercice 5.15 (Du théorème limite central à la loi faible des grands nombres)**.** *Il est possible de déduire la loi faible des grands nombres du théorème de la limite centrale. Notons $a_n = \sigma/\sqrt{n}$. Le TLC affirme que Z_n converge en loi vers Z, il s'agit d'en déduire que $a_n Z_n$ converge en probabilité vers* 0.

Pour cela on peut utiliser deux résultats rappelés dans l'annexe H.1. Comme a_n tend vers 0*, le lemme de Slutsky montre que $a_n Z_n$ converge en loi vers $0 \cdot Z = 0$. Comme cette variable limite est constante, la convergence a en fait lieu en probabilité (*cf. *remarque H.5).*

On donnera une application du TCL au problème de la ruine du joueur dans l'annexe D.

Preuve du théorème limite central

La preuve du TLC est hors-programme.

La preuve basée sur les fonctions caractéristiques est rapide mais utilise des outils hors-programme ; nous allons présenter une preuve utilisant une inégalité de couplage.

En reprenant les notations de l'énoncé du TLC, il s'agit d'établir la convergence de $\mathbb{P}[Z_n \in I]$ vers $\mathbb{P}[Z \in I]$, c'est-à-dire de montrer :

$$\text{Pour toute } f \text{ indicatrice d'intervalle,} \quad \mathbb{E}\left[f(Z_n)\right] \to \mathbb{E}\left[f(Z)\right].$$

Nous utiliserons le résultat suivant :

Lemme 5.16 (Inégalité de couplage de Lindeberg)**.** *Soient X_1, Y_1,..., X_n, Y_n des v.a.r. indépendantes telles que $\mathbb{E}\left[|X_k|^3\right] < \infty$ et $Y_k \sim \mathcal{N}(\mathbb{E}[X_k], \sigma^2(X_k))$ pour tout $1 \leq k \leq n$. Alors pour toute $f \in \mathscr{C}^3(\mathbb{R},\mathbb{R})$ avec f, f', f'', f''' bornées, en posant $\tau_k^3 = \mathbb{E}\left[|X_k - \mathbb{E}[X_k]|^3\right]$,*

$$\left|\mathbb{E}\left[f(X_1 + \cdots + X_n)\right] - \mathbb{E}\left[f(Y_1 + \cdots + Y_n)\right]\right| \leq \frac{\tau_1^3 + \cdots + \tau_n^3}{2} \left\|f'''\right\|_\infty.$$

Démonstration. Quitte à translater f on peut se placer dans le cas où $\mathbb{E}[X_k] = 0$ pour tout $1 \leq k \leq n$. Pour comparer $\mathbb{E}\left[f(X_1 + \cdots + X_n)\right]$ à $\mathbb{E}\left[f(Y_1 + \cdots + Y_n)\right]$, l'idée est de remplacer, de proche en proche, les X_i par les Y_i. Fixons $n \geq 1$ et posons, pour tout $1 \leq k \leq n$:

$$Z_k = X_1 + \cdots + X_{k-1} + Y_{k+1} + \cdots + Y_n.$$

Avec ces notations, on a la somme télescopique

$$f(X_1 + \cdots + X_n) - f(Y_1 + \cdots + Y_n) = \sum_{k=1}^{n} (f(Z_k + X_k) - f(Z_k + Y_k)).$$

La formule de Taylor-Lagrange appliquée à f à l'ordre 2 en Z_k donne

$$f(Z_k + X_k) = f(Z_k) + f'(Z_k)X_k + f''(Z_k)\frac{X_k^2}{2!} + f'''(A_k)\frac{X_k^3}{3!},$$

où A_k est un réel entre Z_k et $Z_k + X_k$. De même on a

$$f(Z_k + Y_k) = f(Z_k) + f'(Z_k)Y_k + f''(Z_k)\frac{Y_k^2}{2!} + f'''(B_k)\frac{Y_k^3}{3!}.$$

Probabilités - Préparation à l'agrégation interne de mathématiques (65/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

On prend l'espérance de la différence de ces expressions :

$$\mathbb{E}\left[f(Z_k+X_k)-f(Z_k+Y_k)\right]=\mathbb{E}\left[f(Z_k)(X_k-Y_k)\right]+\frac{1}{2}\mathbb{E}\left[f''(Z_k)(X_k^2-Y_k^2)\right]+\frac{1}{3!}\mathbb{E}\left[f'''(A_k)X_k^3+f'''(B_k)Y_k^3\right].$$

On utilise maintenant le fait que (X_k,Y_k) et Z_k sont *indépendantes*. Comme les deux premiers moments de X_k et Y_k coïncident, il vient

$$\left|\mathbb{E}\left[f(Z_k+X_k)-f(Z_k+Y_k)\right]\right|\leq\frac{\|f'''\|_\infty}{3!}\mathbb{E}\left[|X_k|^3+|Y_k|^3\right].$$

Comme $Y_k=\sigma(X_k)G_k$ avec $G_k\sim\mathcal{N}(0,1)$ et comme $\mathbb{E}\left[|G_k|^3\right]=4/\sqrt{2\pi}\leq 2$, on obtient

$$\mathbb{E}\left[|Y_k|^3\right]=\mathbb{E}\left[|X_k|^2\right]^{3/2}\mathbb{E}\left[|G_k|^3\right]\leq 2\mathbb{E}\left[|X_k|^3\right].$$

Il ne reste plus qu'à sommer les majorations pour conclure. □

Preuve du TLC (théorème 5.13). On supposera dans cette preuve que $\mathbb{E}\left[|X_1-m|^3\right]=\tau^3<\infty$, ce qui couvre le cas Bernoulli. On montre le résultat en deux temps.

Convergence pour les fonctions régulières. Posons $X_i'=(X_i-m)/(\sigma\sqrt{n})$: on a

$$\mathbb{E}\left[X_i'\right]=0,\qquad\sigma^2(X_i')=1/n,\qquad\sum_{i=1}^n X_i'=Z_n.$$

Soit Y_i une suite de variables indépendantes de loi gaussienne $\mathcal{N}(0,1/n)$. Alors les (X_i',Y_i) vérifient les hypothèses du lemme 5.16, avec pour tout i, $\tau_i^3=\mathbb{E}\left[\left|X_i'\right|^3\right]=\frac{\tau^3}{\sigma^3n^{3/2}}$. Comme de plus $Z=\sum Y_i$ suit la loi normale $\mathcal{N}(0,1)$, on en déduit :

$$\left|\mathbb{E}\left[f(Z_n)-f(Z)\right]\right|\leq\frac{\tau^3}{\sigma^3}\|f'''\|_\infty\cdot\frac{1}{\sqrt{n}},$$

d'où la convergence de $\mathbb{E}\left[f(Z_n)\right]$ vers $\mathbb{E}\left[f(Z)\right]$ pour toute fonction $f\in\mathscr{C}^3(\mathbb{R},\mathbb{R})$ telle que f,f',f'',f''' sont bornées.

Approximation des indicatrices. Soit I un intervalle de $\mathbb{R}$; on se limitera au cas $I=]-\infty,a]$, les autres étant similaires. Pour tout ε, on peut classiquement construire des fonctions « plateau » f_ε et g_ε de classe $\mathscr{C}^3$ telles que

$$\mathbf{1}_{]-\infty,a-\varepsilon]}\leq f_\varepsilon\leq\mathbf{1}_I\leq g_\varepsilon\leq\mathbf{1}_{]-\infty,a+\varepsilon]},$$

De plus, par compacité, les dérivées de f_ε et g_ε jusqu'à l'ordre 3 sont bornées. Concentrons-nous sur les inégalités de droite. En appliquant $\mathbf{1}_I\leq g_\varepsilon$ à la variable Z_n et en prenant l'espérance on trouve

$$\mathbb{E}\left[\mathbf{1}_I(Z_n)\right]\leq\mathbb{E}\left[g_\varepsilon(Z_n)\right].$$

On prend la limite supérieure : le terme de droite converge grâce à la première étape.

$$\limsup\mathbb{E}\left[\mathbf{1}_I(Z_n)\right]\leq\mathbb{E}\left[g_\varepsilon(Z)\right]\leq\mathbb{E}\left[\mathbf{1}_{]-\infty,a+\varepsilon]}(Z)\right]\leq\mathbb{P}\left[Z\leq a+\varepsilon\right].$$

Le premier terme ne dépend pas de ε, et la fonction de répartition de la gaussienne est continue, donc

$$\limsup\mathbb{E}\left[\mathbf{1}_I(Z_n)\right]\leq\mathbb{E}\left[\mathbf{1}_I(Z)\right].$$

Le même raisonnement fournit $\mathbb{E}\left[\mathbf{1}_I(Z)\right]\leq\liminf\mathbb{E}\left[\mathbf{1}_I(Z_n)\right]$. Finalement on a bien établi, pour tout intervalle I,

$$\mathbb{P}\left[Z_n\in I\right]\to\mathbb{P}\left[Z\in I\right].$$ □

5.4 Approximation de la loi binomiale par la loi normale

Soit $(X_n)_{n\geq1}$ une suite de v.a.r. indépendantes et de même loi, de moyenne m et de variance non nulle et finie σ^2. On pose $S_n=X_1+\cdots+X_n$ pour tout $n\geq1$. Le théorème 5.13 de la limite centrale indique que pour tout $t\in\mathbb{R}$,

$$\lim_{n\to\infty}\mathbb{P}\left[Z_n\leq t\right]=\mathbb{P}\left[Z\leq t\right].$$

Probabilités - Préparation à l'agrégation interne de mathématiques (66/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Le théorème de Berry-Esseen précise cette proximité en affirmant que pour tout $t \in \mathbb{R}$ et $n \geq 1$,

$$\sup_{t\in\mathbb{R}} |\mathbb{P}[Z_n \leq t] - \mathbb{P}[Z \leq t]| \leq \frac{\tau^3}{\sqrt{n}\sigma^3}$$

où $\tau^3 = \mathbb{E}\left[|X_1 - \mathbb{E}[X_1]|^3\right]$. Lorsque $(X_n)_{n\geq 1}$ sont de Bernoulli de paramètre $p \in]0,1[$, on peut expliciter la borne de droite : pour $q = 1-p$, $\sigma^2 = pq$ et $\tau^3 = pq(1-2pq)$ ce qui donne

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}\left[S_n \leq np + t\sqrt{npq}\right] - \mathbb{P}[Z \leq t]\right| \leq \frac{1-2pq}{\sqrt{npq}}.$$

Cette approximation de la loi binomiale par la loi normale est bonne quand $(1-2p(1-p))/\sqrt{np(1-p)}$ est petit. À n fixé, cette borne est minimale pour $p = 1/2$ mais explose quand p se rapproche de 0 ou de 1. Une preuve du théorème de Berry-Esseen se trouve dans le livre de Feller [Fel68]. Notons que l'on peut adapter [5] notre preuve du théorème central limite pour obtenir une borne en $\mathcal{O}(n^{-1/8})$.

Il est possible de quantifier la proximité de la loi binomiale à la loi normale en utilisant la *densité* plutôt que la fonction de répartition : c'est l'approche historique, illustrée par le théorème suivant.

Théorème 5.17 (de Moivre et Laplace)**.** *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi de Bernoulli de moyenne $p \in]0,1[$ et d'écart-type $\sigma = \sqrt{p(1-p)}$. Alors pour tout $n \geq 1$, la variable $S_n = X_1 + \cdots + X_n$, de loi* $\mathrm{Binom}(n,p)$ *d'espérance $m_n = np$ et d'écart-type $\sigma_n = \sqrt{n}\sigma$ et sa fonction de masse, convenablement normalisée, converge vers la densité gaussienne $\mathcal{N}(0,\sigma^2)$. Plus précisément, si n et k tendent simultanément vers l'infini de telle façon que $\lim_{n,k\to\infty}(k-m_n)/\sigma_n = x$ alors*

$$\sigma_n \mathbb{P}[S_n = k] \xrightarrow[k,n\to+\infty]{} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Démonstration. Procédons par étapes.

Échelles de convergence. On pose $q = 1-p$ et $l = n-k$. La relation $(k-m_n)/\sigma_n \to x$ entraîne en particulier les convergences suivantes :

$$\frac{k}{n} \xrightarrow[k,n\to+\infty]{} p, \qquad \frac{l}{n} \xrightarrow[k,n\to+\infty]{} q, \qquad \frac{l-nq}{\sigma_n} \xrightarrow[k,n\to+\infty]{} -x.$$

Décomposition. Par définition de la loi binomiale, puis en utilisant la formule de Stirling, on obtient :

$$\begin{aligned}
\sqrt{n}\mathbb{P}[S_n = k] &= \sqrt{n}\binom{n}{k} p^k q^l = \sqrt{n}\frac{n!}{k!l!} p^k q^l \\
&\sim \sqrt{n}\frac{\sqrt{2\pi n}(n/e)^n}{\sqrt{2\pi k}(k/e)^k\sqrt{2\pi l}(l/e)^l} p^k q^l \\
&\sim \frac{1}{\sqrt{2\pi}}\frac{n}{\sqrt{kl}}\left(\frac{pn}{k}\right)^k\left(\frac{qn}{l}\right)^l \\
&\sim \frac{1}{\sqrt{2\pi}\sigma}\exp\left(k\ln\left(\frac{pn}{k}\right) + l\ln\left(\frac{qn}{l}\right)\right),
\end{aligned}$$

où l'on a utilisé les convergences de k/n vers p et l/n vers q.

Terme exponentiel. On cherche la limite du terme dans l'exponentielle. En posant $u_k = \frac{k-np}{k}$, le premier terme dans l'exponentielle devient :

$$k\ln\left(\frac{pn}{k}\right) = k\ln(1-u_k).$$

Or la relation $(k-m_n)/\sigma_n \to x$ entraîne $u_k \sim \frac{x\sigma}{p\sqrt{n}}$, donc

$$\begin{aligned}
k\ln\left(\frac{pn}{k}\right) &= k\left(-u_k - \frac{1}{2}u_k^2 + o(u_k^2)\right) \\
&= -ku_k - \frac{1}{2}ku_k^2 + o(ku_k^2).
\end{aligned}$$

5. Il s'agit de remarquer que f_ε peut être choisie de manière à ce que $\|f_\varepsilon'''\|_\infty \leq \varepsilon^{-3}$, puis de choisir ε dépendant de n, de façon que $\varepsilon^4 = \mathcal{O}(n^{-1/2})$.

Probabilités - Préparation à l'agrégation interne de mathématiques (67/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Comme $ku_k^2 \sim \frac{x^2\sigma^2}{p^2}\frac{k}{n} \sim x^2 q$, il vient :

$$k\ln\left(\frac{pn}{k}\right) = np - k - q\frac{x^2}{2} + o(1).$$

De même

$$l\ln\left(\frac{nq}{l}\right) = nq - l - p\frac{x^2}{2} + o(1).$$

On somme ces deux expressions ; comme $p+q=1$ et $k+l=n$, le terme dans l'exponentielle converge vers $-x^2/2$. Finalement on a bien :

$$(\sigma\sqrt{n})\mathbb{P}\left[S_n = k\right] \xrightarrow[k,n\to+\infty]{} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right),$$

comme annoncé. □

Ici la convergence est obtenue « ponctuellement » en x ; en raffinant l'étude asymptotique via une amélioration de la formule de Stirling, on peut montrer le résultat plus fort suivant :

Théorème 5.18 (de Moivre et Laplace, uniforme)**.** *Si $S_n \sim \mathrm{Binom}(n,p)$ avec $0<p<1$ et $q=1-p$, de moyenne $m_n = np$ et d'écart-type $\sigma_n = \sqrt{npq}$, alors pour tous $-\infty < a < b < +\infty$ on a la convergence uniforme suivante :*

$$\lim_{n\to\infty}\sqrt{n}\sup_{k\in I_n(a,b)}\left|\mathbb{P}\left[S_n = k\right] - \frac{1}{\sqrt{2\pi}\sigma_n}\exp\left(-\frac{(k-m_n)^2}{2\sigma_n^2}\right)\right| = 0$$

où

$$I_n(a,b) = \left\{0 \le k \le n : \frac{k-m_n}{\sigma_n} \in [a,b]\right\}.$$

Nous renvoyons au livre de Dacunha-Castelle et Duflo [DCD94] pour une preuve.

Le théorème 5.18 fournit par intégration le TLC dans le cas Bernoulli :

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{S_n - m_n}{\sigma_n} \in [a,b]\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx.$$

Probabilités - Préparation à l'agrégation interne de mathématiques (68/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Compléments

Probabilités - Préparation à l'agrégation interne de mathématiques (69/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Probabilités - Préparation à l'agrégation interne de mathématiques (70/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Annexe **A**

Biais par la taille

A.1 Un cas simple

Problème A.1 (Biais par la taille)**.** *On cherche à tirer uniformément un foyer parmi les foyers français. Pour cela on prend un français au hasard dans la population et on choisit son foyer. La procédure est-elle correcte ?*

Pour clarifier la situation, supposons que l'on s'intéresse uniquement à la taille des foyers. Pour tout entier $k \geq 1$, soit p_k la fréquence des foyers de taille k dans la population française, de sorte que $\sum_{k\geq 1} p_k = 1$. Notons $m := \sum_k k p_k$ la taille moyenne des foyers.

Prenons donc un français au hasard dans la population et notons T la taille du foyer auquel il appartient. Si la procédure d'échantillonage proposée était correcte, on devrait avoir $\mathbb{P}[T = k] = p_k$. Nous allons voir qu'en réalité, $\mathbb{P}[T = k] = \frac{k}{m} p_k$: les grands foyers sont sur-représentés et les petits, sous-représentés. Ce phénomène est appelé *biais par la taille*[1]. Il s'agit sans doute du biais d'échantillonnage le plus connu. Ce biais est d'autant plus important que la taille du foyer diffère de la taille moyenne m.

Solution. Notons N_k le nombre de foyers de taille k dans la population française : il y a donc $F = \sum_k N_k$ foyers et $N = \sum_k k N_k$ individus en France, la proportion de foyers de taille k étant $p_k = N_k / F$. On vérifie de plus aisément l'égalité $N = mF$.

On modélise le choix d'un français au hasard par le tirage d'un entier selon la loi uniforme sur l'intervalle $[\![1, N]\!]$. Il y a N_k foyers de taille k qui comptent au total kN_k individus. Par définition de la loi uniforme (formule « cas favorables sur cas totaux ») la probabilité que ce français appartiennent à un foyer de taille k est donc

$$\frac{kN_k}{N} = \frac{kp_kF}{N} = \frac{k}{m} p_k,$$

ce qui est la formule annoncée. □

A.2 Estimation dans une file d'attente

Problème A.2 (Temps d'attente à un guichet)**.** *Le gestionnaire d'une administration décide d'évaluer le temps que les usagers passent au guichet. Supposons pour simplifier qu'il y a un unique guichet. Le gestionnaire vient à une heure fixée, attend que la personne actuellement au guichet ait fini son opération et lui demande combien de temps elle a attendu.*

Cette procédure donne-t-elle une bonne estimation du temps moyen d'attente des usagers ?

Comme précédemment, la procédure n'est pas correcte : le temps d'attente observé est biaisé, il y a plus de chances que le gestionnaire arrive pendant qu'une requête particulièrement longue est en train d'être traitée.

Modélisons la situation en temps discret. Le guichet ouvre le matin au temps $n = 0$. Les usagers arrivent ; le k^{e} usager reste au guichet pendant X_k secondes. Pour rester dans un cadre simple on suppose les choses suivantes :

— les temps de service X_k sont des variables aléatoires à valeurs entières ;

— les X_k sont indépendantes, de même loi : $\mathbb{P}[X_k = i] = p_i$, et d'espérance commune $m = \sum k p_k < \infty$.

1. *In English :* $\sum_{k\geq 1} \frac{k}{m} p_k \delta_k$ *is the size-biased law constructed from the initial law* $\sum_{k\geq 1} p_k \delta_k$.

Probabilités - Préparation à l'agrégation interne de mathématiques (71/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

On note $V_k = 1$ si le temps k correspond au début du service d'un usager, $V_k = 0$ sinon. Si par exemple $X_1 = 2$, $X_2 = 3$ et $X_3 = 1$, on aura $V_0 = V_2 = V_5 = V_6 = 1$ et $V_1 = V_3 = V_4 = 0$.

Le gestionnaire arrive à un temps n que l'on va supposer grand. La suite $S_k = \sum_{j=1}^{k} X_k$ est strictement croissante, il existe donc un unique K (aléatoire) tel que $S_{K-1} \leq n < S_K$. Ce que le gestionnaire observe est le temps de service du K^{e} usager, à savoir X_K.

Cherchons la loi de ce temps X_K. Pour tout i, on décompose l'événement $X_K = i$ en une union disjointe suivant les valeurs de K :

$$\begin{aligned}\{X_K = i\} = \bigcup_k \{X_K = i, K = k\} &= \bigcup_k \left\{ X_k = i, \sum_{j=1}^{k-1} X_j \leq n < \sum_{j=1}^{k-1} X_j + X_k \right\} \\ &= \bigcup_k \left\{ X_k = i, \sum_{j=1}^{k-1} X_j \leq n < \sum_{j=1}^{k-1} X_j + i \right\}.\end{aligned}$$

On prend maintenant les probabilités des deux côtés ; à droite, on utilise le fait que l'union est disjointe et l'indépendance des X_k pour écrire :

$$\mathbb{P}[X_K = i] = \sum_k \mathbb{P}[X_k = i]\,\mathbb{P}[S_{k-1} \leq n < S_{k-1} + i] = p_i \sum_k \mathbb{P}[n - i < S_{k-1} \leq n].$$

Enfin, en écrivant la probabilité comme espérance d'indicatrice, on peut faire rentrer la somme dans l'espérance et écrire :

$$\begin{aligned}\mathbb{P}[X_K = i] &= p_i \mathbb{E}\left[\sum_k \mathbf{1}_{n-i<S_{k-1}\leq n}\right] \\ &= p_i \mathbb{E}[\text{« nombre de renouvellements entre } n-i+1 \text{ et } n \text{ »}] \\ &= p_i \sum_{n-i+1}^{n} \mathbb{E}[V_j]. \qquad \text{(A.1)}\end{aligned}$$

Pour conclure, il nous faut calculer $\mathbb{E}[V_j]$, ou au moins donner sa limite quand n tend vers l'infini. Le résultat général (admis) est le suivant.

Théorème A.3 (Renouvellement)**.** *Si la distribution des X_k est « non-réticulée », c'est-à-dire si il n'existe aucun $a \geq 2$ tel que X_k soit presque-sûrement un multiple de a, alors $\mathbb{E}[V_n] \to 1/m$.*

Corollaire A.4. *Quand n tend vers l'infini, le temps observé par le gestionnaire converge en loi vers la loi de X_1 biaisée par sa taille :*

$$\mathbb{P}[X_K = i] \xrightarrow[n\to\infty]{} \frac{i p_i}{m}.$$

Démonstration. Le corollaire est une conséquence immédiate du théorème et de la formule (A.1) puisque la somme dans le terme de droite contient un nombre fini (i) de termes, qui tendent tous vers $1/m$.

La preuve du théorème nous entraînerait trop loin. Notons tout de même que dans le cas particulier où la loi des X_i est géométrique de paramètre $1/m$, le résultat est simple. En effet, l'expérience consiste alors à lancer à chaque instant une pièce : si elle tombe sur pile le service s'arrête et on passe à l'usager suivant, si elle tombe sur face le service continue. Les variables V_n sont donc dans ce cas indépendantes, et suivent toutes une loi de Bernoulli de paramètre $1/m$, d'espérance $1/m$. □

Ce biais par la taille apparaît dans de nombreuses situations pratiques. Pour citer un exemple, on s'intéresse en foresterie à l'estimation du nombre d'arbres dans une parcelle, à leur taille, leur rayon, etc. ; les méthodes pratiques de comptage (que ce soit par parcours au sol ou télédétection) induisent souvent un biais, qui dans les cas les plus simples est un biais par la taille ou par la surface.

Probabilités - Préparation à l'agrégation interne de mathématiques (72/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Annexe **B**

Jeu de pile ou face

B.1 Récapitulatif

On modélise une infinité de lancers au jeu de pile ou face avec une pièce équilibrée par l'espace probabilisé produit $(\{0,1\}^{\mathbb{N}^*}, \mathscr{F}, \mathbb{P})$ où $\mathscr{F}$ est la tribu engendrée par les cylindres et $\mathbb{P}$ la mesure de probabilité produit associée à la mesure de probabilité uniforme sur $\{0,1\}$. Ici pile est codé 1 et face est codé 0. Lorsque la pièce n'est pas équilibrée et donne face avec probabilité $p \in [0,1]$, on munit $\{0,1\}$ de la loi qui affecte la probabilité p à 1 et $1-p$ à 0. La suite des coordonnées dans cet espace produit constitue une suite de v.a.r. $(X_n)_{n\geq 1}$ indépendantes et de même loi de Bernoulli :

$$\mathbb{P}[X_n = 1] = 1 - \mathbb{P}[X_n = 0] = p \in [0,1].$$

On a déjà illustré précédemment sur cet exemple de nombreuses notions :

— les notions d'univers et de tribu, pages 5 et 7, et celle de variable aléatoire, page 23 ;

— les propriétés des suites d'événements et le lemme de Borel–Cantelli, pages 15 et 19

— la notion d'indépendance, page 18 ;

— trois lois classiques (loi de Bernoulli, loi binomiale et loi géométrique) et leurs premiers moments, pages 26, 33, 50 et 54 ;

— sur la loi binomiale, les notions d'intervalle de confiance et de fluctuation, pages 62, 63.

B.2 Temps d'attente des succès successifs

Si on compte les piles comme des succès et les faces comme des échecs, rappelons que la loi du temps d'attente du premier succès est la loi géométrique $\mathsf{Geom}(p)$, définie pour tout $k \geq 1$ par $\mathbb{P}[T = k] = (1-p)^{k-1}p$.

Pour tout $r \in \mathbb{N}^*$, le nombre de lancers T_r nécessaires pour obtenir r succès est défini par récurrence par $T_1 = T$ et $T_{r+1} = \inf\{n > T_r : X_n = 1\}$. Les v.a.r. $T_1, T_2 - T_1, T_3 - T_2, \ldots$ sont indépendantes et de même loi géométrique $\mathsf{Geom}(p)$. La loi de la variable aléatoire T_r est appelée loi de Pascal ou **loi binomiale-négative**[1]. Elle est donnée par le produit de convolution $(\mathsf{Geo}_{\mathbb{N}^*}(p))^{*r}$, qui se traduit par la formule suivante pour $k \geq r$:

$$\begin{aligned}\mathbb{P}[T_r = k] &= \sum_{\substack{k_1 \geq 1, \ldots, k_r \geq 1 \\ k_1 + \cdots + k_r = k}} (1-p)^{k_1 - 1} p \cdots (1-p)^{k_r - 1} p \\ &= (1-p)^{k-r} p^r \binom{k-1}{r-1}\end{aligned}$$

Les deux premiers moments se calculent facilement puisque T est somme de variables indépendantes de loi géométrique :

$$\mathbb{E}[T_r] = r\mathbb{E}[T] = \frac{r}{p} \quad \text{et} \quad \sigma^2(T_r) = r\sigma^2(T) = r\frac{1-p}{p^2}.$$

1. Le coefficient binomial $\binom{k-1}{r-1}$ peut se réécrire $\frac{(k-1)(k-2)\cdots r}{(k-r)!} = (-1)^r \frac{(-r)(-r-1)\cdots(1-k)}{(k-r)!}$ ce qu'on note parfois $\binom{-r}{k-r}$. On obtient ainsi une formule très proche de celle de la loi binomiale, mais où apparaît le coefficient binomial généralisé $\binom{-r}{k-r}$, ce qui explique le nom de la loi.

Probabilités - Préparation à l'agrégation interne de mathématiques (73/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Le processus de Bernoulli $(B_n)_{n\geq 0}$ est donné par $B_n = B_0 + S_n$ où B_0 est une variable aléatoire quelconque. Ses trajectoires sont constantes par morceaux, avec des sauts d'amplitude +1, et les temps de saut sont donnés par $(T_r)_{r\geq 1}$ (temps inter-sauts indépendantes et de même loi géométrique). Le processus de Bernoulli est le processus de comptage de tops espacés par des durées indépendantes de même loi géométriques. De ce point de vue, il constitue un analogue à temps discret du processus de Poisson simple. C'est pour ce processus que l'on a pu montrer complètement le phénomène de biais par la taille dans l'annexe précédente.

B.3 Fluctuations non asymptotiques

Rappelons que la loi forte des grands nombres et le théorème de la limite centrale s'écrivent ici

$$\mathbb{P}\left[\frac{S_n}{n} \to p\right] = 1 \quad \text{et} \quad \frac{\sqrt{n}}{\sqrt{p(1-p)}}\left(\frac{S_n}{n} - p\right) \xrightarrow[n\to+\infty]{\mathscr{L}} \mathscr{N}(0,1),$$

et que cela donne des intervalles de confiance asymptotiques pour p : l'intervalle dit de Wald, dont nous verrons plus bas (p. 100) la justification, et sa « simplification » vue page 63. Cet intervalle est assez mauvais en pratique : la convergence n'est pas très rapide et pour de petites valeurs de n ou pour des p proches de 0 ou 1, la « probabilité de couverture » (probabilité que le vrai paramètre p appartienne à l'intervalle de confiance) peut être bien inférieure à $1-\alpha$.

De nombreuses alternatives sont disponibles dans ce cas. Une possibilité est de s'appuyer sur la correspondance beta-binomiale : si $U_1, \dots, U_n$ sont des v.a. indépendantes et de loi uniforme sur $[0,1]$ et si $U_{(1)} \leq \cdots \leq U_{(n)}$ est leur réordonnement alors $U_{(k)} \sim \mathrm{Beta}(k, n-k+1)$ a pour densité $t \in [0,1] \mapsto (\int_0^1 s^{k-1}(1-s)^{n-k}\,ds)^{-1} t^{k-1}(1-t)^{n-k}$, et

$$\mathbb{P}[S_n \geq k] = \mathbb{P}\left[\mathbf{1}_{\{U_1\leq p\}} + \cdots + \mathbf{1}_{\{U_n \leq p\}} \geq k\right] = \mathbb{P}\left[U_{(k)} \leq p\right].$$

On peut grâce à cela définir un intervalle de confiance pour p. Pour le voir, fixons n et notons, pour tout $1 \leq k \leq n$, $p_\alpha(k)$ le quantile d'ordre α de la loi $\mathrm{Beta}(k, n-k+1)$. On voit facilement que $k \mapsto p_\alpha(k)$ est croissante ; on la « complète » en posant $p_\alpha(0) = 0$ et $p_\alpha(n+1) = 1$.

Théorème B.1 (Intervalle de Clopper-Pearson)**.** *On pose $i_\alpha(k) = p_{\alpha/2}(k)$ et $s_\alpha(k) = p_{1-\alpha/2}(k+1)$. L'intervalle $I_{CP}(S_n) = [i_\alpha(S_n), s_\alpha(S_n)]$ est un intervalle de confiance au niveau α : si S_n suit la loi binomiale* $\mathrm{Binom}(n,p)$,

$$\mathbb{P}\left[p \in I_{CP}(S_n)\right] \geq 1-\alpha.$$

Démonstration. S'il est différent de 1, ce que l'on peut supposer, le paramètre p est nécessairement dans un intervalle du type $[i_\alpha(k-1), i_\alpha(k)[$ pour un k entre 1 et $n+1$. Pour cette valeur de k et pour tout entier l, $p \leq i_\alpha(l)$ si et seulement si $k \leq l$. Par conséquent

$$\mathbb{P}\left[p \leq i_\alpha(S_n)\right] = \mathbb{P}\left[k \leq S_n\right].$$

Si $k = n+1$ cette probabilité est nulle, sinon on utilise la correspondance Beta-binomiale et la définition de k pour écrire :

$$\begin{aligned}\mathbb{P}\left[p \leq i_\alpha(S_n)\right] &= \mathbb{P}\left[U_{(k)} \leq p\right] \\ &\leq \mathbb{P}\left[U_{(k)} \leq i_\alpha(k)\right] \\ &= \alpha/2.\end{aligned}$$

On raisonne de même pour montrer $\mathbb{P}\left[p \geq s_\alpha(S_n)\right] \leq \alpha/2$, ce qui permet de conclure. □

B.4 Lien avec la loi uniforme

Chaque réalisation de la suite (X_n) permet de construire un nombre réel dans l'intervalle $[0,1]$ via son écriture en base 2. Cela correspond à la surjection $\varsigma : \{0,\dots,1\}^{\mathbb{N}^*} \to [0,1]$ définie pour tout $x \in \{0,\dots,2\}^{\mathbb{N}^*}$ par

$$\varsigma(x) = \sum_{n=1}^{\infty} 2^{-n} x_n = \underbrace{0, x_1 \cdots x_n \cdots}_{\text{en base 2}}.$$

Probabilités - Préparation à l'agrégation interne de mathématiques (74/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Lorsque $p = 1/2$ alors la variable aléatoire U donnée par

$$U = \varsigma(X) = \sum_{n=1}^{\infty} 2^{-n} X_n$$

suit la loi uniforme sur $[0,1]$. Si $a = a_1 r^{-1} + \cdots + a_n 2^{-n}$ est un nombre diadique, on a

$$\begin{aligned}\mathbb{P}\left[a < U < a + 2^{-n}\right] &= \mathbb{P}\left[X_1 = a_1, \ldots, X_n = a_n\right] \\ &= \mathbb{P}\left[X_1 = a_1\right] \cdots \mathbb{P}\left[X_n = a_n\right] = 2^{-n}.\end{aligned}$$

Réciproquement, ce calcul montre que les coefficients (X_n) de l'écriture en base 2 d'une variable aléatoire uniforme sur $[0,1]$ sont indépendants de loi de Bernoulli symétrique sur $\{0,1\}$. Les nombres de $[0,1]$ dont l'écriture en base 2 est constante à partir d'un certain rang « ne comptent pas » en quelque sorte, et ς est presque sûrement une injection et donc presque sûrement une bijection. On peut donc en déduire une méthode pour générer d'un seul coup n réalisations indépendantes de loi de Bernoulli symétrique sur $\{0,1\}$ à partir d'une réalisation de précision n en base 2 d'une loi uniforme sur $[0,1]$.

Probabilités - Préparation à l'agrégation interne de mathématiques (75/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Probabilités - Préparation à l'agrégation interne de mathématiques (76/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Annexe **C**

Collectionneur d'images

C.1 Définition et formules exactes

Un collectionneur cherche à réunir les r images différentes d'un album pour enfants. Pour cela on suppose qu'il achète les images une à une, et que chaque image achetée est tirée uniformément parmi les r possibles. Si l'on note T le nombre total (aléatoire) d'images à acheter pour avoir la collection complète, que peut-on dire de T ?

Ce problème, appelé en anglais *coupon collector problem*, est un modèle important à ranger dans la même boite à outils que le jeu de pile ou face, auquel il est intimement relié. Un grand nombre de situations concrètes sont modélisables par le collectionneur d'images ou une de ses variantes. Nous nous limitons ici à la variante la plus simple.

Repérons l'image tiré la k^{e} fois par une variable aléatoire X_k, prenant ses valeurs dans $\{1,2,\dots,r\}$. Par hypothèse les variables $(X_k)_{k\in\mathbb{N}}$ sont indépendantes et de même loi uniforme sur $\{1,2,\dots,r\}$. L'instant où la collection est complète est défini par

$$T = \min\{n \geq 1 : \{X_1,\dots,X_n\} = \{1,\dots,r\}\} = \min\{n \geq 1 : \mathrm{Card}\{X_1,\dots,X_n\} = r\}.$$

Théorème C.1 (Expression combinatoire de la loi)**.** *On a $T \geq r$ et pour tout $n \geq r$,*

$$\mathbb{P}[T = n] = \frac{r!}{r^n}\left\{\begin{matrix} n-1 \\ r-1 \end{matrix}\right\}$$

où la notation entre accolades est le nombre de Stirling de seconde espèce $(n-1, r-1)$, c'est-à-dire le nombre de partitions en $r-1$ blocs d'un ensemble de $n-1$ éléments.

Démonstration. On a $X_T \notin \{X_1,\dots,X_{T-1}\}$ car l'image qui termine la collection n'a forcément jamais été vue auparavant. Fixons $n \geq r$. Il y a r^n choix possibles décrivant les n premiers tirages. Pour construire un choix favorable à l'événement $\{T = n\}$, il faut choisir la dernière image (r choix), répartir les $n-1$ images restants sur les $r-1$ types d'images restants $\left\{\begin{matrix} n-1 \\ r-1 \end{matrix}\right\}$ choix pour savoir quelles sont les images identiques, et $(r-1)!$ ordres possibles sur les $r-1$ types d'images restants). Le résultat désiré en découle. □

La queue de distribution peut elle aussi être calculée explicitement.

Théorème C.2 (Queue de distribution)**.** *Pour tout $n \geq 1$,*

$$\mathbb{P}[T > n] = \sum_{k=1}^{r} (-1)^{k-1}\binom{r}{k}\left(1 - \frac{k}{r}\right)^n.$$

Démonstration. Notons $E_{n,i} = \{X_1 \neq i,\dots,X_n \neq i\}$ l'événement « l'image i n'est toujours pas dans la collection après le n^{e} tirage ». On a alors

$$\mathbb{P}[T > n] = \mathbb{P}\left[E_{n,1} \cup \dots \cup E_{n,r}\right].$$

Si $i_1,\dots,i_k \in \{1,\dots,r\}$ sont distincts alors, avec $R = \{1,\dots,r\} \setminus \{i_1,\dots,i_k\}$,

$$\mathbb{P}\left[E_{n,i_1} \cap \dots \cap E_{n,i_k}\right] = \mathbb{P}[X_1 \in R]\cdots\mathbb{P}[X_n \in R] = \left(\frac{r-k}{r}\right)^n = \left(1 - \frac{k}{r}\right)^n.$$

Le résultat désiré découle alors du principe d'inclusion-exclusion (théorème 2.3). □

Probabilités - Préparation à l'agrégation interne de mathématiques (77/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Le théorème C.1 est superbe mais difficile à exploiter : il est par exemple malaisé d'en déduire l'espérance de T. De même, le fait que les signes soit alternés dans la formule du théorème C.2 rend délicate l'étude directe du comportement de la queue de T en fonction de n et r.

C.2 Une décomposition et ses conséquences

Le résultat suivant est intuitif et va fournir facilement les premiers résultats sur T ; il montre déjà que que $\mathbb{P}[T<\infty]=1$.

Lemme C.3 (Décomposition)**.** *Il existe r variables aléatoires G_1, G_2,…G_r, indépendantes, telles que :*
- *G_i suit la loi géométrique de paramètre $\pi_i = \frac{r-i+1}{r}$,*
- *la variable T se décompose comme $T=\sum G_i$.*

Démonstration. L'idée principale est de découper le temps T en r périodes, suivant le nombre d'images différentes collectées. La première image obtenue est automatiquement incluse dans la collection : on pose $G_1 \equiv 1$ le temps d'attente de cette image. Pour obtenir une nouvelle image après cela, il faut attendre un temps $G_2 = \min\{n \geq 1 : X_{G_1+n} \neq X_{G_1}$. Plus généralement, pour tout $1<i\leq r$, on pose

$$G_i = \min\{n \geq 1 : X_{G_{i-1}+n} \notin \{X_1,\dots,X_{G_{i-1}}\}\}$$

le temps nécessaire pour passer d'une collection de $i-1$ images différentes à i images différentes. On a $\mathrm{Card}(\{X_1,\dots,X_{G_i}\}) = i$ pour tout $1\leq i\leq n$. La variable aléatoire G_i est le temps d'apparition du premier gain dans un jeu de pile ou face spécial, pour lequel la probabilité de gagner vaut $(r-i+1)/r$ (puisqu'il reste $r-i+1$ images à collecter) : G_i suit donc la loi $\mathrm{Geom}_{\mathbb{N}^*}((r-i+1)/r)$. Ceci témoigne du fait qu'il est de plus en plus difficile d'obtenir une image d'un nouveau type au fil de la collection. □

Corollaire C.4 (Moments)**.** *L'espérance de T est donnée par :*

$$\mathbb{E}[T] = r\sum_{i=1}^{r}\frac{1}{i} = rH_r$$

où H_r est le r^e nombre harmonique, compris entre $\ln(r)$ et $\ln(r)+1$, et qui vérifie $H_r = \ln(r)+\gamma+o_{r\to\infty}(1)$ où $\gamma\approx 0.577$ est la constante d'Euler. La variance vaut

$$\sigma^2(T) = r\sum_{i=1}^{r-1}\frac{r-i}{i^2} \leq \frac{\pi^2}{6}r^2.$$

Démonstration. La linéarité de l'espérance et le fait que G_i suit une loi géométrique de paramètre $\pi_i = (r-i+1)/r$ donne

$$\mathbb{E}[T] = \sum_{i=1}^{r}\mathbb{E}[G_i] = \sum_{i=1}^{r}\frac{1}{\pi_i} = \sum_{i=1}^{r}\frac{r}{r-i+1} = r\sum_{i=1}^{r}\frac{1}{i}$$

comme annoncé. Comme de plus les G_i sont indépendantes, de variances respectives $\sigma^2(G_i) = (1-\pi_i)/\pi_i^2 = r(i-1)/(r-i+1)^2$, on en déduit la formule de la variance. Les propriétés de H_r sont classiques. □

Application numérique. Pour $r=150$, il faut acheter en moyenne $150(\ln(150)+\gamma)\approx 838$ images pour compléter la collection !

Ces simples contrôles de moments permettent d'établir la convergence suivante.

Théorème C.5 (Convergence)**.**

$$\frac{T}{r\ln(r)} \xrightarrow[r\to\infty]{\mathbb{P}} 1.$$

Démonstration. Par l'inégalité de Markov et les formules pour les moments de T, il vient pour tout $u>0$,

$$\begin{aligned}\mathbb{P}\left[\left|\frac{T}{r\ln(r)}-1\right|>u\right] &\leq \frac{\mathbb{E}\left[|T-r\ln(r)|^2\right]}{r^2\ln(r)^2u^2}\\ &= \frac{\sigma^2(T)+(\mathbb{E}[T]-r\ln(r))^2}{r^2\ln(r)^2u^2}\\ &\leq (\pi^2/6+1)\left(\frac{1}{u^2\ln(r)^2}\right).\end{aligned}$$

Pour tout u fixé le terme de droite tend vers 0 quand $r\to\infty$, ce qui prouve la convergence annoncée. □

Probabilités - Préparation à l'agrégation interne de mathématiques (78/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

La borne logarithmique sur la vitesse de convergence en probabilité est trop faible pour en déduire une convergence presque sûre au moyen du lemme de Borel-Cantelli. En revanche, la majoration obtenue par l'inégalité de Markov fournit un premier intervalle de fluctuation pour T : pour tout $t > 0$, la borne précedente appliquée à $u = t/\ln(r)$ donne

$$\mathbb{P}\left[T \in [r\ln(r) - rt, r\ln(r) + rt]\right] \geq 1 - (\pi^2/6 + 1)\frac{1}{t^2}.$$

À présent, pour α et r fixés, il faut choisir t assez grand pour que le second membre soit égal à $1-\alpha$. L'intervalle de fluctuation, de largeur $2rt$, se dégrade quand α diminue (car t augmente).

Application numérique. Pour $\alpha = 0.05$ et $r = 150$ on trouve l'intervalle $[-340, 1843]$. La borne supérieure est très grande et la borne inférieure n'apporte rien ! À α (et donc t) fixé, elle ne devient intéressante que si r est assez grand pour garantir $r\ln(r) - rt > r$.

Remarque C.6. *En n'utilisant que des informations sur les deux premiers moments de T on a pu montrer un résultat limite et obtenir un intervalle de fluctuation. En pratique ce dernier n'est pas très bon pour des valeurs raisonnables de r ; pour avoir mieux il faut étudier plus précisément la loi de T.*

C.3 Étude fine et fluctuations

Commençons par une première inégalité de déviation en reprenant la preuve du théorème C.2.

Théorème C.7 (Déviation)**.** *Pour tout réel $t > 0$,*

$$\mathbb{P}\left[T > r\ln(r) + rt\right] \leq \frac{r}{r-1}e^{-t}.$$

Démonstration. Soit n la partie entière de $r\ln(r) + rt$. On réécrit $\{T > n\}$ comme union des $E_{n,i}$, comme dans la preuve du théorème C.2, et on majore la probabilité de l'union par une somme :

$$\mathbb{P}\left[T > n\right] = \mathbb{P}\left[\cup_{i=1}^{r} E_{n,i}\right] \leq \sum_{i=1}^{r} \mathbb{P}\left[E_{n,i}\right].$$

Comme $\mathbb{P}\left[E_{n,i}\right] = (1 - 1/r)^n$, il vient

$$\mathbb{P}\left[T > r\ln(r) + rt\right] \leq \mathbb{P}\left[T > n\right] \leq r\left(1 - \frac{1}{r}\right)^{r\ln(r)+rt-1} \leq \frac{r^2}{r-1}\exp\left((r\ln(r) + rt)\ln(1 - 1/r)\right).$$

Le résultat suit en majorant $\ln(1 - 1/r)$ par $(-1/r)$. □

Pour un α et un r fixé, on peut en déduire un intervalle de fluctuation de T, en choisissant $t_\alpha = \ln\left(r/(\alpha(r-1))\right)$, pour lequel $\mathbb{P}\left[T \in [r, r\ln(r) + t_\alpha r]\right] \geq 1 - \alpha$.

Application numérique. Supposons qu'il y ait $r = 150$ images à collectionner, et fixons $\alpha = 0.05$. On trouve $t_\alpha \approx 3.0$, et il y a 95% de chances que T soit entre 150 et... 1201.

On peut enfin pousser l'analyse plus loin en prenant en compte le caractère alterné de la somme définissant la queue de la distribution de T. On rappelle que la loi de Gumbel a pour fonction de répartition $t \in \mathbb{R} \mapsto e^{-e^{-t}}$.

Théorème C.8 (Fluctuations asymptotiques)**.** *On a*

$$\frac{T - r\ln(r)}{r} \xrightarrow[r\to\infty]{loi} \text{Gumbel}.$$

Démonstration. Il suffit d'établir que pour tout $t \in \mathbb{R}$ on a

$$\lim_{r\to\infty} \mathbb{P}(T > r\ln(r) + tr) = S(t) = 1 - e^{-e^{-t}}.$$

Fixons donc $t \in \mathbb{R}$ et supposons que r est assez grand pour que $r\ln(r) + tr > r$. Soit $n_{t,r}$ l'entier défini par $n_{t,r} = r\ln(r) + tr$ si $r\ln(r) + tr \in \mathbb{N}$ et $n_{t,r} = \lfloor r\ln(r) + tr \rfloor + 1$ sinon. Le théorème C.2 donne

$$\mathbb{P}(T > r\ln(r) + tr) = \sum_{k=1}^{r} (-1)^{k-1}\binom{r}{k}\left(1 - \frac{k}{r}\right)^{n_{t,r}}.$$

Probabilités - Préparation à l'agrégation interne de mathématiques (79/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Comme $\binom{r}{k} \leq r^k/k!$ et $1-u \leq e^{-u}$ pour tout $u \geq 0$, on a

$$\binom{r}{k}\left(1-\frac{k}{r}\right)^{n_{t,r}} \underset{r\to\infty}{\overset{\leq}{\longrightarrow}} \frac{e^{-tk}}{k!}.$$

Enfin, par convergence dominée, on obtient

$$\lim_{r\to\infty} \sum_{k=1}^{r} (-1)^{k-1}\binom{r}{k}\left(1-\frac{k}{r}\right)^{n_{t,r}} = \sum_{k=1}^{\infty} (-1)^{k-1}\frac{e^{-tk}}{k!} = S(t). \qquad \square$$

Le théorème C.8 fournit un nouvel intervalle de fluctuation (asymptotique) pour T : pour tout réel $t \geq 0$,

$$\lim_{r\to\infty} \mathbb{P}\left[T \in [r\ln(r) - rt, r\ln(r) + rt]\right] = e^{-e^{-t}} - e^{-e^{t}}.$$

Application numérique. Toujours pour $r = 150$ et $\alpha = 0.05$, on résout numériquement $e^{-e^{-t}} - e^{-e^{t}} = \alpha$ pour obtenir $t_\alpha \approx 2.97$ et l'intervalle de fluctuation $[306, 1198]$. La borne supérieure est donc très proche de la précédente ; on a en revanche gagné une borne inférieure non-triviale.

Probabilités - Préparation à l'agrégation interne de mathématiques (80/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Annexe **D**

Marche aléatoire simple et ruine du joueur

D.1 Ruine du joueur

Soit $0 < p < 1$ un réel fixé et soit $(\varepsilon_n)_{n\geq 1}$ une suite de v.a. indépendantes et de même loi telles que $\mathbb{P}[\varepsilon_n = 1] = 1 - \mathbb{P}[\varepsilon_n = -1] = p$ pour tout $n \geq 1$. Soit X_0 une v.a. sur $\mathbb{Z}$ indépendante de la suite $(\varepsilon_n)_{n\geq 1}$. La **marche aléatoire simple** sur $\mathbb{Z}$, de paramètre p, est la suite récurrente aléatoire $(X_n)_{n\geq 0}$ sur $\mathbb{Z}$ définie par relation récursive suivante :

$$X_{n+1} = X_n + \varepsilon_{n+1} = X_0 + \varepsilon_1 + \cdots + \varepsilon_{n+1}$$

pour tout $n \geq 0$. Pour tout $n \geq 1$, la v.a. $\beta_n = (\varepsilon_n + 1)/2$ suit une loi de Bernoulli de paramètre p car on a $\mathbb{P}\left[\beta_n = 1\right] = 1 - \mathbb{P}\left[\beta_n = 0\right] = p$. Ainsi, pour tout $n \geq 0$,

$$\frac{X_n - X_0 + n}{2} = \beta_1 + \cdots + \beta_n \sim \mathrm{Binom}(n, p).$$

Un joueur invétéré joue à la roulette de façon répétée en misant toujours 1 € : il gagne un euro à chaque « victoire » et en perd un à chaque « défaite ». On suppose que les parties sont indépendantes, et que la probabilité de victoire vaut p.

La fortune initiale du joueur est de x euros ; il quitte le jeu lorsqu'il atteint la fortune $a < x$ euros (il est ruiné) ou $b > x$ euros (son objectif). Dans la suite, on adopte les notations $\mathbb{P}_x[\cdot] := \mathbb{P}[\cdot | X_0 = x]$ et $\mathbb{E}_x[\cdot] := \mathbb{E}[\cdot | X_0 = x]$.

Théorème D.1 (Sortie ou ruine)**.** *Soient $a, b \in \mathbb{Z}$ avec $a < b$. Soit τ_a, τ_b et τ les v.a. à valeurs dans $\mathbb{N} \cup \{\infty\}$ définies par*

$$\tau_a = \inf\{n \geq 0 : X_n = a\}, \quad \tau_b = \inf\{n \geq 0 : X_n = b\}, \quad \textit{et} \quad \tau = \min(\tau_a, \tau_b).$$

Alors pour tout $a \leq x \leq b$ on a $\mathbb{E}_x(\tau) < \infty$ et en posant $\rho = \dfrac{1-p}{p}$ on a

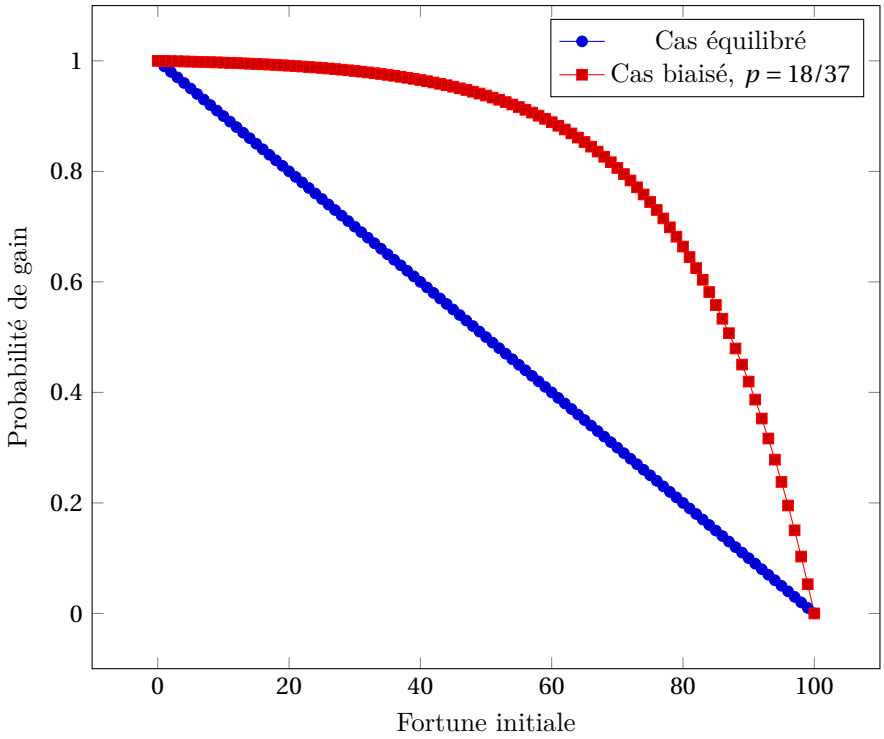
$$\mathbb{P}_x(X_\tau = a) = \begin{cases} \dfrac{\rho^b - \rho^x}{\rho^b - \rho^a} & \textit{si } p \neq \frac{1}{2}, \\ \dfrac{b-x}{b-a} & \textit{si } p = \frac{1}{2}. \end{cases} \quad \textit{et} \quad \mathbb{E}_x(\tau) = \begin{cases} \dfrac{x-a}{1-2p} - \dfrac{(b-a)}{1-2p}\dfrac{\rho^x - \rho^a}{\rho^b - \rho^a} & \textit{si } p \neq \frac{1}{2}, \\ (b-x)(x-a) & \textit{si } p = \frac{1}{2}. \end{cases}$$

Démonstration. Montrons tout d'abord que τ est d'espérance finie. Notons $L = b - a$ la longueur de l'intervalle. Remarquons que si le joueur remporte ses L premières parties, alors $\tau = \tau_a \leq L$. Plus généralement, si le joueur remporte les parties $n+1$, $n+2$,..., $n+L$, alors $\tau \leq n+L$: soit la marche avait déjà atteint $\{a, b\}$ avant le temps n, soit elle était encore entre a et b au temps n et les L gains successifs l'en font sortir.

Notons C_j l'événement « chanceux » $C_j = \left\{\varepsilon_{jL+1} = \varepsilon_{jL+2} = \cdots = \varepsilon_{jL} = 1\right\}$. Le raisonnement précédent montre que

$$\mathbb{P}_x[\tau > kL] \leq \mathbb{P}_x\left[\bigcap_{j=1}^{k} C_j^c\right].$$

Probabilités - Préparation à l'agrégation interne de mathématiques (81/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



Un joueur part avec $x \in \{1,\dots,99\}$ euros, joue un euro à chaque partie et part s'il est ruiné ou s'il atteint son objectif de 100 euros. On représente la probabilté de ruine en fonction de la fortune initiale dans deux cas : pour un jeu équilibré (pile ou face non-biaisé) et pour un jeu très légèrement déséquilibré (un pari « rouge » à la roulette française, de probabilité de gain $18/37 \approx 0.486$). Ce faible déséquilibre sur une partie conduit à une différence très nette dans les parties répétées : en partant de la somme moyenne $x = 50$ euros, le joueur a une chance sur deux d'être ruiné en jouant à pile ou face, et environ 94% de chances d'être ruiné à la roulette...

Figure D.1 – Ruine du joueur

Comme les pas (ε_i) sont supposés indépendants et que les C_j couvrent des périodes de temps disjointes, les événements C_j sont indépendants, donc

$$\mathbb{P}_x[\tau > kL] \leq \prod_{j=1}^{k} \mathbb{P}_x\left[C_j^c\right] = (1-p^L)^k.$$

Comme $1-p^L < 1$, on en déduit à l'aide du théorème 3.34 que $\mathbb{E}_x[\tau] < \infty$ (on peut en fait montrer que τ a des moments de tous ordres). En particulier $\mathbb{P}_x[\tau < \infty] = 1$.

Calculons maintenant la probabilité de ruine $r(x) = \mathbb{P}_x[X_\tau = a]$. On conditionne par le résultat de la première partie pour obtenir :

$$\begin{aligned} r(x) &= \mathbb{P}_x[X_\tau = a \mid X_1 = x+1]\, p + \mathbb{P}_x[X_\tau = a \mid X_1 = x-1]\,(1-p) \\ &= p r(x+1) + (1-p) r(x-1). \end{aligned}$$

L'ensemble des solutions de cette récurrence linéaire d'ordre deux est un espace vectoriel qui contient la solution constante 1. Si $p \neq 1/2$ alors ρ^x est aussi solution, linéairement indépendante de 1, et donc les solutions sont de la forme $A + B\rho^x$ avec A et B constantes. Les conditions aux bords $r(a) = 1$, $r(b) = 0$ fixent A et B, ce qui donne l'unique solution

$$r(x) = \frac{\rho^b - \rho^x}{\rho^b - \rho^a}.$$

Si $p = 1/2$ alors $\rho = 1$ et les deux solutions fondamentales précédentes sont confondues. Cependant, on observe que dans ce cas, x est également solution, linéairement indépendante de 1, et donc les solutions sont de la forme $A + Bx$ où A et B sont des constantes. Les conditions aux bords $r(a) = 1$ et $r(b) = 0$ fixent A et B, ce qui donne l'unique solution

$$r(x) = \frac{b-x}{b-a}.$$

Probabilités - Préparation à l'agrégation interne de mathématiques (82/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Calculons $R(x) = \mathbb{E}_x(\tau)$. En conditionnant selon X_1 on obtient [1] pour tout $a < x < b$ la récurrence linéaire

$$R(x) = pR(x+1) + (1-p)R(x-1) + 1.$$

La présence du second membre 1 fait rechercher des solutions particulières. Si $p \neq 1/2$ alors $x/(1-2p)$ est solution particulière, et les solutions de l'équation sont de la forme $R(x) = x/(1-2p) + A + B\rho^x$. Les conditions aux bords $R(a) = 0$ et $R(b) = 0$ donnent enfin

$$R(x) = \frac{x-a}{1-2p} - \frac{(b-a)}{1-2p}\frac{\rho^b - \rho^x}{\rho^b - \rho^a}.$$

Si $p = 1/2$ alors $-x^2$ est solution particulière, et les solutions sont de la forme $-x^2 + A + Bx$. Les conditions aux bords $R(a) = R(b) = 0$ donnent enfin

$$R(x) = (b-x)(x-a).$$

La même approche permet de calculer la fonction génératrice $\mathbb{E}_x(s^\tau \mid X_\tau = a)$. □

Remarque D.2. *Si $p = 1/2$, on a $\mathbb{P}_x(\tau_a < \infty) = 1$ et $\mathbb{P}_x(\tau_b < \infty) = 1$ pour tout $a \leq x \leq b$. Il est possible d'établir que si $p = 1/2$ alors avec probabilité 1, la suite aléatoire $(X_n)_{n\geq 0}$ visite presque sûrement chaque élément de $\mathbb{Z}$ une infinité de fois. En revanche, si $p \neq 1/2$ alors avec probabilité 1, la suite $(X_n)_{n\geq 0}$ ne visite qu'un nombre fini de fois chaque élément de $\mathbb{Z}$. On le voit bien dans les formules du théorème D.1 en faisant tendre a ou b vers l'infini.*

Remarque D.3 (Les théorèmes limites à la rescousse)**.** *Voici un autre argument pour établir que $\mathbb{P}_x(\tau < \infty) = 1$. Posons $m = 2p - 1$ et $\sigma^2 = 4p(1-p)$. Si $m \neq 0$ alors par la loi forte des grands nombres, p.s. $(X_n)_{n\geq 1}$ tend vers $+\infty$ si $m > 0$ et vers $-\infty$ si $m < 0$, et donc $\mathbb{P}_x(\tau < \infty) = 1$. Si $m = 0$ alors pour tout $n \geq 1$, en posant $I_n = \frac{1}{\sqrt{n}}\,]a,b[$, on a*

$$\mathbb{P}_x(\tau = \infty) \leq \mathbb{P}\,[a < X_n < b] = \mathbb{P}\left(\frac{X_n}{\sqrt{n}} \in I_n\right).$$

Or $(n^{-1/2}X_n)_{n\geq 1}$ converge en loi vers $\mathcal{N}(0,\sigma^2)$ par le théorème central limite. Mais I_n dépend de n. Cependant, comme $(I_n)_{n\geq 1}$ est décroissante,

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{X_n}{\sqrt{n}} \in I_n\right) \leq \inf_{m\geq 1} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{I_m} e^{-\frac{t^2}{2\sigma^2}}\, dt = 0.$$

D.2 Premier retour en zéro de la marche aléatoire

La loi du premier temps où une marche aléatoire partant de 0 y revient est totalement explicite.

Théorème D.4 (Premier retour en zéro et nombres de Catalan)**.** *Si $\tau = \inf\{n \geq 1 : X_n = 0\}$ alors pour tout $n \geq 0$,*

$$\mathbb{P}_0\,[\tau = 2n+2] = \frac{2}{n+1}\binom{2n}{n} p^{n+1}(1-p)^{n+1}.$$

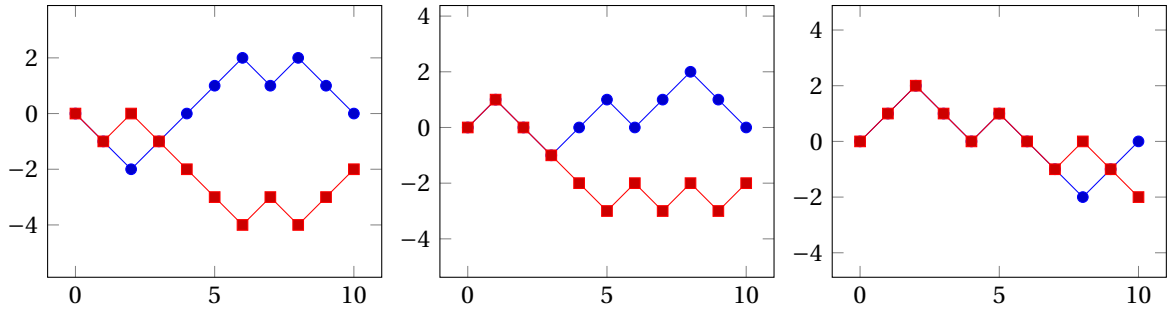
On reconnaît le n^{e} nombre de Catalan $\frac{1}{n+1}\binom{2n}{n}$. Ces nombres comptent, outre les chemins de la marche aléatoire simple, les mots de Dyck, les parenthésages, les triangulations d'un polygone, les partitions non croisées, les chemins sous-diagonaux dans le carré, les arbres planaires, etc. Les moments pairs de la loi du demi-cercle sont les nombres de Catalan. C'est l'occasion de souligner que la beauté de la combinatoire réside dans les bijections qu'elle révèle, entre des ensembles finis d'objets de natures *a priori* très différentes.

Preuve du théorème D.4. Sachant $\{X_0 = 0\}$, l'événement $\{\tau = 2n+2\}$ correspond à une trajectoire de longueur $2n+2$ partant de 0 et revenant à zéro en restant strictement positive ou strictement négative. Ces deux cas sont équiprobables, d'où le facteur 2 dans le résultat. Dans les deux cas, il y a eu forcément $n+1$ incréments $+1$ et $n+1$ incréments -1, d'où

$$\mathbb{P}_0(\tau = 2n+2) = 2C_n p^{n+1}(1-p)^{n+1}.$$

1. Cette méthode de conditionnement, déjà utilisée pour déterminer $r(x)$, est valable plus généralement pour toute chaîne de Markov.

Probabilités - Préparation à l'agrégation interne de mathématiques (83/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



Pour compter les « ponts positifs » (chemins de longueur n commençant et finissant en 0 et restants positifs) on dénombre l'ensemble complémentaire des ponts qui prennent au moins une valeur négative. À chacun de ces ponts (représentés en bleu), on associe un autre pont entre 0 et -2, représenté en rouge. Le chemin rouge suit le bleu jusqu'au premier -1, après quoi il en est l'image miroir (réfléchie par rapport à la droite $y=-1$). On obtient ainsi tous les ponts entre 0 et 2, ce qui permet de dénombrer les ponts positifs.

Figure D.2 – Compter les ponts

où C_n est le nombre de chemins de longueur $2n+2$ partant de zéro et revenant à zéro, et restant strictement positifs. Le premier incrément est forcément $+1$ et le dernier forcément -1 et C_n est égal au nombre de chemins de longueur $2n$ partant de zéro et revenant à zéro et restant positifs. Il y a n incréments $+1$ et n incréments -1. Considérons les chemins partant de zéro et revenant à zéro et contenant n incréments $+1$ et n incréments -1. Il y en a $\binom{2n}{n}$. Si un chemin de ce type n'est pas positif alors juste après la première position négative, modifions tous les incréments en permutant le signe des $+1$ et des -1 — cette transformation est illustrée dans la figure D.2. On obtient de la sorte un chemin avec $n-1$ incréments $+1$ et $n+1$ incréments -1, et il s'avère que tous les chemins partant de zéro avec $n-1$ incréments $+1$ et $n+1$ incréments -1 s'obtiennent de la sorte, et il y en a $\binom{2n}{n-1}$. Ainsi, $C_n = \binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n}$. Cette astuce est attribuée au mathématicien français Désiré André (1840–1917). □

De cette expression combinatoire, on peut déduire les résultats suivants.

Théorème D.5 (Récurrence et temps de retour pour la marche aléatoire)**.** *Soit τ le temps de premier retour en 0 d'une marche aléatoire de paramètre p.*

— *Si $p \neq 1/2$, alors $\mathbb{P}[\tau=\infty]>0$: avec probabilité positive, on ne revient jamais en zéro. On dit que la marche est transiente.*

— *Si $p=1/2$, alors $\mathbb{P}[\tau<\infty]=1$: on revient toujours en 0, la marche est dite récurrente. En revanche le temps moyen de retour est infini : $\mathbb{E}[\tau]=\infty$. On dit que la marche est récurrente nulle.*

Démonstration. On procède par étapes.

Une série entière. Comme la loi de τ est explicite, on peut mener à bien tous les calculs. Tout d'abord

$$\mathbb{P}[\tau<\infty] = \sum_{n=0}^{\infty} \mathbb{P}[\tau=2n+2] = 2pq \sum_{n=0}^{\infty} C_n (pq)^n = 2pqg(pq),$$

où g est la série entière[2] $g(s)=\sum_{n\geq 0} C_n s^n$. L'interprétation probabiliste ci-dessus montre que $g(1/4)$ converge, donc le rayon de convergence de g est supérieur où égal à $1/4$. Notons de plus que $2sg(s)$ appartient nécessairement à $[0,1]$, puisque c'est $\mathbb{P}[\tau<\infty]$ pour p tel que $pq=s$.

Une récurrence pour les nombres de Catalan. Les nombres de Catalan vérifient la relation de récurrence

$$C_{n+1} = \sum_{k=0}^{n} C_k C_{n-k}.$$

Ceci peut se voir sur l'interprétation combinatoire. C_{n+1} est le cardinal de l'ensemble $\mathscr{C}_{n+1}$ des chemins de longueur $2n+2$ qui restent positifs. On partitionne $\mathscr{C}_{n+1}$:

$$\mathscr{C}_{n+1} = \bigcup_{k=0}^{n} \{\text{chemins dont le premier retour en zéro est en } 2k+2\}.$$

2. En combinatoire, on dit que g est la fonction génératrice de la suite C_n, d'où la notation.

Probabilités - Préparation à l'agrégation interne de mathématiques (84/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Un chemin de $\mathscr{C}_{n+1}$ dont le premier zéro est en $2k+2$ est la concaténation d'un chemin de longueur $2k+2$ restant strictement positif, et d'un chemin de longueur $2n-2k$ restant positif ; d'où la formule annoncée.

Identification de g. Pour $s \geq 0$, on en déduit

$$\begin{aligned} g(s) &= C_0 + s \sum_{n=0}^{\infty} C_{n+1} s^n \\ &= 1 + s \sum_{0 \leq k \leq n} C_k s^k C_{n-k} s^{n-k} \\ &= 1 + s \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} C_k C_l s^k s^l \\ &= 1 + s g(s)^2, \end{aligned}$$

toutes les interversions étant autorisées puisque les termes sont positifs (toujours le théorème de Fubini) Pour $s \in]0, 1/4]$, on en déduit

$$g(s) = \frac{1 \pm \sqrt{1-4s}}{2s}.$$

Comme $2sg(s) \in [0,1]$, on a finalement $2sg(s) = 1 - \sqrt{1-4s}$. Par conséquent

$$\mathbb{P}[\tau < \infty] = 1 - \sqrt{1-4pq}.$$

Cette quantité vaut bien 1 pour $p = 1/2$, sinon elle est strictement inférieure à 1.

Espérance de τ. La variable τ est à valeurs dans $\mathbb{N} \cup \{\infty\}$. Si $p \neq 1/2$, $\mathbb{P}[\tau = \infty] > 0$ et l'espérance est automatiquement infinie. Si $p = 1/2$, $\mathbb{P}[\tau = \infty] = 0$, et

$$\mathbb{E}[\tau] = \sum_n (2n+2)\mathbb{P}[\tau = 2n+2] = \sum_{n=0}^{\infty} \binom{2n}{n} 4^{-n}.$$

Par la formule de Stirling, le terme général de cette série est équivalent à $1/\sqrt{\pi n}$, donc la série diverge et l'espérance de τ est infinie.

Note : les nombres de Catalan sont les moments pairs de la loi du demi-cercle (exercice 3.40). □

Probabilités - Préparation à l'agrégation interne de mathématiques (85/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Probabilités - Préparation à l'agrégation interne de mathématiques (86/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Annexe **E**

Lois exponentielles et durées de vie

E.1 Définition, premières propriétés

Rappelons que X suit une loi exponentielle de paramètre $\lambda > 0$ si sa fonction de répartition vaut : $F(x) = \mathbf{1}_{\mathbb{R}_+}(1 - e^{-\lambda x})$, ; de façon équivalente : $\mathbb{P}[X \geq x] = e^{-\lambda x}$ pour tout $x \geq 0$. La loi exponentielle admet la densité $\mathbf{1}_{\mathbb{R}_+} \lambda e^{-\lambda x}$.

Les lois exponentielles sont très importantes en modélisation stochastique. Cette importance peut s'expliquer par le fait qu'elles apparaissent naturellement comme analogue continu de la loi géométrique. La loi géométrique modélise l'attente d'un événement (premier succès dans un jeu de pile ou face répété) ; si on réalise les expériences de plus en plus vite, en diminuant la probabilité de succès, on obtient une loi exponentielle. Plus précisément on a le résultat suivant.

Théorème E.1 (Lois exponentielles et géométriques)**.**

Contraction de la loi géométrique. *Pour tout $n \geq 1$ soit $X_n \sim \mathrm{Geom}(p_n)$ avec $0 < p_n < 1$. Si np_n converge vers $\lambda > 0$, en posant X une variable de loi $\mathrm{Exp}(\lambda)$, on a pour tout $x \in \mathbb{R}$,*

$$\mathbb{P}\left[\frac{X_n}{n} \geq x\right] \xrightarrow{n\to\infty} e^{-\lambda x} = \mathbb{P}[X \geq x].$$

Discrétisation de la loi exponentielle. *Si $Y \sim \mathrm{Exp}(\lambda)$ et si $\lfloor Y \rfloor$ désigne la partie entière de Y, alors $\lfloor Y \rfloor$ et $Y - \lfloor Y \rfloor$ sont indépendantes et $1 + \lfloor Y \rfloor \sim \mathrm{Geom}(e^{-\lambda})$.*

Remarque E.2. *Le premier résultat établit,* via *le critère des fonctions de répartition, une* convergence en loi *— notion détaillée dans la définition H.2.*

Démonstration. Pour la première partie on écrit, pour tout $n \geq 1$ et tout $x \in \mathbb{R}$,

$$\mathbb{P}[X_n \geq nx] = \mathbb{P}[X_n \geq \lfloor nx \rfloor] = (1 - p_n)^{\lfloor nx \rfloor} \to e^{-\lambda x}.$$

La seconde partie découle de la formule $\mathbb{P}[Y > t + s \mid Y > s] = \mathbb{P}[Y > t] = e^{-\lambda t}$ pour tous $s, t \geq 0$, que l'on reverra plus bas et qui correspond à l'absence de mémoire des lois exponentielles. □

La pièce de monnaie, comme les cartes ou le tirage du loto, n'a pas de mémoire : le fait qu'elle soit tombée 5 fois sur face de suite n'augmente pas la probabilité d'apparition d'un pile. Cette propriété se généralise aux lois exponentielles et les caractérisent.

Théorème E.3 (Caractérisation des lois exponentielles par absence de mémoire)**.** *Pour toute v.a.r. X sur $\mathbb{R}_+$ telle que $\mathbb{P}[X > 0] > 0$, les propriétés suivantes sont équivalentes :*

1. *$\mathscr{L}(X)$ est une loi exponentielle ;*
2. *pour tout t, la « loi conditionnelle » $\mathscr{L}(X - t \mid X > t)$ est une exponentielle de paramètre λ, ce qui signifie que pour tout $\mathbb{P}[(X - t) \leq s | X > t] = 1 - e^{-\lambda s}$ pour tout $s \geq 0$.*

Démonstration. Si $X \sim \mathrm{Exp}(\lambda)$ alors $\mathscr{L}(X - t | X > t) = \mathscr{L}(X)$ car pour tous $s, t \geq 0$,

$$\mathbb{P}[X - t > s | X > t] = \frac{\mathbb{P}[X > t + s]}{\mathbb{P}[X > t]} = e^{-\lambda s}.$$

Probabilités - Préparation à l'agrégation interne de mathématiques (87/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Réciproquement, si l'on note $G(t) = \mathbb{P}[X > t]$, la seconde propriété implique, pour tous $s, t \geq 0$

$$G(t)G(s) = G(t+s).$$

Par continuité inférieure, $G(\varepsilon) = \mathbb{P}[X > \varepsilon] \to \mathbb{P}[X > 0] > 0$ lorsque $\varepsilon \to 0^+$. Ainsi, $G(\varepsilon) > 0$ pour $\varepsilon > 0$ assez petit. Si $t > 0$ alors $t \leq n\varepsilon$ avec $n \in \mathbb{N}$ et donc

$$G(t) = \mathbb{P}[X > t] \geq \mathbb{P}[X > n\varepsilon] = G(\varepsilon)^n.$$

Ainsi, $G(t) > 0$ pour tout $t \geq 0$. D'autre part, les solutions non identiquement nulles de l'équation fonctionnelle $G(t+s) = G(t)G(s)$ pour tous $s, t \in \mathbb{R}_+$ sont de la forme $G(t) = G(1)^t$ (considérer les $t \in \mathbb{Q}$ puis utiliser la décroissance de G lorsque $t \in \mathbb{R}_+$). □

Les variables exponentielles modélisent souvent des temps d'attente d'un « client » (usager à la poste, ordinateur personnel,...) qui souhaite utiliser un « serveur » (guichet ouvert, serveur d'un fournisseur d'accès,...). Si plusieurs serveurs fonctionnent en parallèle, le temps d'attente du client sera donné par le minimum d'un ensemble de variables. Si les temps sont exponentiels la situation est particulièrement simple.

Théorème E.4 (Minimum – Horloges en compétition)**.** *Si $E_1, \ldots, E_n$ sont des v.a.r. indépendantes de loi exponentielle de paramètres $\lambda_1, \ldots, \lambda_n$ alors*

$$M = \min(E_1, \ldots, E_n) \sim \mathrm{Exp}(\lambda_1 + \cdots + \lambda_n).$$

De plus, avec probabilité 1*, le minimum M est atteint pour un unique entier aléatoire K indépendant de M et de loi donnée pour tout $1 \leq k \leq n$ par*

$$\mathbb{P}[K = k] = \mathbb{P}[M = E_k] = \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}.$$

Démonstration. On a $\mathbb{P}[M \geq x] = \mathbb{P}[E_1 \geq x] \cdots \mathbb{P}[E_n \geq x] = e^{-(\lambda_1 + \cdots + \lambda_n)x}$ pour tout $x \geq 0$, et cela montre que M suit la loi exponentielle de paramètre $\lambda_1 + \ldots + \lambda_n$. Pour traiter K, nous allons déterminer directement la loi du couple[1] (M, K), ce qui fournira à nouveau la loi de M. Comme les v.a.r. sont indépendantes et de loi à densité, avec probabilité 1, l'entier aléatoire K est bien défini sur $\{1, \ldots, n\}$ (voir le corollaire 4.13). Pour tout $1 \leq k \leq n$ et $t \geq 0$ on a

$$\{M \geq t \text{ et } K = k\} = \{E_k \geq t \text{ et } E_{k'} > E_k \text{ pour tout } k' \neq k\}.$$

Par hypothèse sur les variables aléatoires $E_1, \ldots, E_n$ il vient

$$\mathbb{P}[M \geq t \text{ et } K = k] = \int_t^\infty \lambda_k e^{-s\lambda_k} \prod_{k' \neq k} \mathbb{P}[E_{k'} > s]\, ds = \left(1 - e^{-t(\lambda_1 + \cdots + \lambda_n)}\right) \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}.$$

On en déduit que les variables aléatoires M et K sont indépendantes, et de plus, les lois de M et K s'obtiennent en prenant $t = 0$ et en sommant en k respectivement. □

Si $E_1, \ldots, E_n$ sont indépendantes et de loi exponentielle de paramètre λ alors pour tout réel $t \geq 0$ la v.a.r. discrète $\mathbf{1}_{\{E_1 > t\}} + \cdots + \mathbf{1}_{\{E_n > t\}} = \mathrm{Card}\{1 \leq k \leq n : E_k > t\}$, c'est à dire le nombre de serveurs encore occupés au temps t, suit la loi binomiale $\mathrm{Binom}(n, e^{-\lambda t})$ car les indicatrices sont indépendantes de loi de Bernoulli.

Symétriquement on peut s'intéresser au premier temps où toutes les horloges ont sonné.

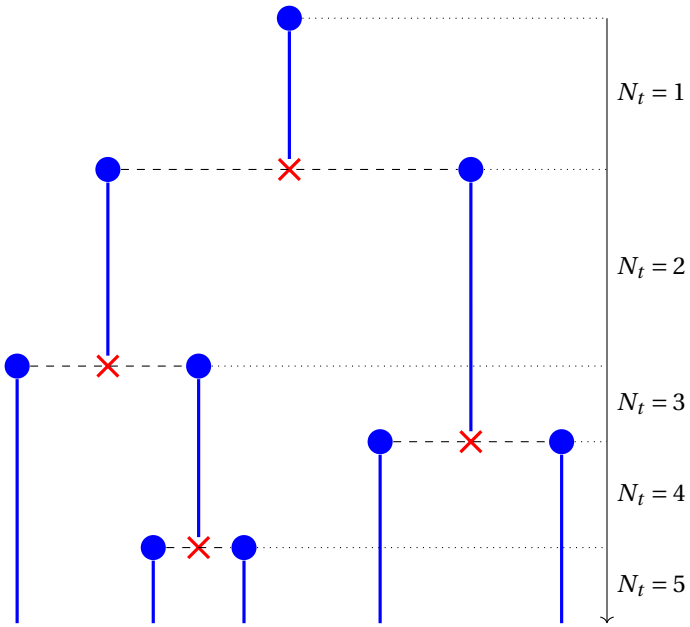
Théorème E.5 (Maximum)**.** *Si $E_1, \ldots, E_n$ sont indépendantes de loi $\mathrm{Exp}(\lambda)$ alors*

$$\mathscr{L}(\max(E_1, \ldots, E_n)) = \mathscr{L}(F_1 + \cdots + F_n)$$

où $F_1, \ldots, F_n$ sont indépendantes avec $F_k \sim \mathrm{Exp}(k\lambda)$ pour tout $1 \leq k \leq n$.

1. Notons que ce couple, dont la première composante est à densité et la seconde est discrète, ne rentre pas dans le cadre du programme.

Probabilités - Préparation à l'agrégation interne de mathématiques (88/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



À l'instant 0 (en haut du graphique) il n'y a qu'une seule particule. Elle vit un certain temps aléatoire de loi exponentielle (longueur du segment bleu) puis meurt (c'est la croix rouge) en donnant naissance à deux particules filles (les segments horizontaux indiquent cette filiation, leur longueur est arbitraire). Chacune de ces deux particules évolue de la même manière.

FIGURE E.1 – Le processus de Yule

Démonstration. Cette décomposition en somme est très proche de celle qui apparaît dans le problème du collectionneur d'images (lemme C.3). Donnons d'abord une justification intuitive du résultat en ce sens. Soit F_n le temps où la première horloge sonne (exponentiel de paramètre $n\lambda$ comme minimum de n variables exponentielles). Après ce temps, il faut attendre un temps F_{n-1} pour entendre une horloge sonner : il reste $n-1$ horloges et, par absence de mémoire[2], chacune suit une loi exponentielle, donc F_{n-1} suit bien la loi $\mathrm{Exp}(\lambda(n-1))$. On procède de même jusqu'à F_1, temps d'attente entre l'avant-dernière sonnerie d'horloge et la dernière.

Proposons maintenant une preuve plus analytique. Posons $S_n = F_1 + \cdots + F_n$. La densité de $M_n = \max(E_1, \ldots, E_n)$ est

$$f_n(x) = (\mathbb{P}[M_n \le x])' = ((1-e^{-\lambda x})^n \mathbf{1}_{\mathbb{R}_+}(x))' = n\lambda(1-e^{-\lambda x})^{n-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x).$$

Montrons par récurrence sur n que S_n a pour densité f_n. C'est vrai pour $n=1$. Si cela est vrai pour n, alors la densité de S_{n+1} est, en notant g_λ la densité de $\mathrm{Exp}(\lambda)$,

$$\begin{aligned} f_n * g_{(n+1)\lambda}(y) &= (n+1)\lambda \int_{-\infty}^{y} f_n(x) e^{-\lambda(n+1)(y-x)}\, dx \\ &= \lambda(n+1) n\lambda e^{-\lambda(n+1)y} \int_0^y (e^{\lambda x}-1)^{n-1} e^{\lambda x}\, dx \\ &= \lambda(n+1) e^{-\lambda(n+1)y} (e^{\lambda y}-1)^n \\ &= \lambda(n+1) e^{-\lambda y}(1-e^{-n\lambda y})^n = f_{n+1}(y). \end{aligned}$$

□

Illustrons une autre conséquence de cette décomposition sur un modèle de population. Considérons un arbre binaire infini représentant la descendance d'une cellule mère (voir la figure E.1). Supposons que chaque bout de branche a une longueur aléatoire qui représente la durée de vie avant division. On suppose que toutes ces longueurs sont des v.a.r. indépendantes de loi exponentielle de paramètre λ. À l'instant $t \ge 0$, l'arbre possède N_t branches, et $N_0 = 1$. Le processus $(N_t)_{t\ge 0}$ est appelé processus de Yule de paramètre λ.

Corollaire E.6 (Taille de la population du processus de Yule)**.** *Pour tout $t \ge 0$ la taille de la population N_t suit la loi géométrique* $\mathrm{Geom}(e^{-\lambda t})$. *En particulier,*

$$\mathbb{E}[N_t] = e^{\lambda t} \quad \textit{et} \quad \mathrm{Var}(N_t) = (1-e^{-\lambda t})e^{2\lambda t}.$$

2. On admet ici que l'absence de mémoire peut s'appliquer après le temps aléatoire F_n.

Probabilités - Préparation à l'agrégation interne de mathématiques (89/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Démonstration. Notons F_1 le temps d'attente de la première division, F_2 le temps à attendre après F_1 pour observer la deuxième division, etc. Par définition $N_t = \mathrm{Card}\{n \geq 1 : S_n \leq t\}$ où $S_n = F_1 + \cdots + F_n$. La propriété d'absence de mémoire des lois exponentielles implique que les $(F_n)_{n\geq 1}$ sont indépendantes et que $F_n \sim \mathrm{Exp}(n\lambda)$ pour tout $n \geq 1$ (puisque F_n peut être vu comme le minimum de n variables $\mathrm{Exp}(\lambda)$ indépendantes correspondant au temps de vie des n particules présentes au temps S_{n-1}). D'après le théorème E.5, la v.a.r. S_n a même loi que $\max(E_1, \ldots, E_n)$ où $E_1, \ldots, E_n$ sont des v.a.r. indépendantes de loi exponentielle de paramètre λ. Par conséquent, on a pour tout n :

$$\mathbb{P}[N_t - 1 \geq n] = \mathbb{P}[S_n \leq t] = \mathbb{P}[E_1 \leq t] \cdots \mathbb{P}[E_n \leq t] = (1 - e^{-\lambda t})^n. \qquad \square$$

Reprenons le vocabulaire « client-serveur » et supposons maintenant que les serveurs fonctionnent en série : le client doit passer un temps E_1 avec le premier serveur, E_2 avec le deuxième, etc. Le client enregistre sa progression (le nombre de serveurs vus) dans une variable N_t.

Théorème E.7 (Horloges en série, comptage, processus de Poisson)**.** *Soit $(E_n)_{n\geq 1}$ des v.a.r. de loi exponentielle de paramètre λ, et $T_n = \sum_{i=1}^n E_i$. Pour tout $t \geq 0$, soit N_t la v.a.r. de comptage*

$$N_t = \mathrm{Card}\{n \geq 1 : T_n \leq t\}.$$

Alors :

1. *$(T_1, \ldots, T_n)$ a pour densité $(t_1, \ldots, t_n) \mapsto \lambda^n e^{-\lambda t_n} \mathbf{1}_{\{0<t_1<\cdots<t_n\}}(t_1, \ldots, t_n)$;*
2. *T_n suit la loi Gamma de densité $t \mapsto \frac{t^{n-1}}{(n-1)!} \lambda^n e^{-\lambda t} \mathbf{1}_{\mathbb{R}_+}(t)$;*
3. *N_t suit la loi de Poisson $\mathrm{Poi}(\lambda t)$.*

Démonstration. La propriété 1 s'obtient par le changement de variable linéaire triangulaire

$$(s_1, s_2, \ldots, s_n) \mapsto (s_1, s_1 + s_2, \ldots, s_1 + \cdots + s_n)$$

à partir de la loi de $(T_1, T_2 - T_1, \ldots, T_n - T_{n-1})$ de densité

$$(s_1, \ldots, s_n) \mapsto \prod_{i=1}^n \lambda e^{-\lambda s_i} \mathbf{1}_{\mathbb{R}_+}(s_i) = \lambda^n e^{-\lambda(s_1 + \cdots + s_n)} \mathbf{1}_{\mathbb{R}_+^n}(s_1, \ldots, s_n).$$

On obtient le point 2 par récurrence sur n. Pour le point 3 on écrit $\{N_t = n\} = \{T_n \leq t < T_{n+1}\}$, d'où $\mathbb{P}[N_t = n] = \mathbb{P}[T_n \leq t] - \mathbb{P}[T_{n+1} \leq t]$, et on utilise la propriété 2. $\square$

E.2 Modélisation des durées de vie

Comme on l'a vu, les lois exponentielles sont utiles pour modéliser les durées dans les processus de renouvellement sans mémoire comme les files d'attente : caisse de supermarché, feu tricolore, etc. Elles sont également utilisées pour modéliser des phénomènes physiques comme la désintégration de particules radioactives. En revanche, la propriété d'absence de mémoire montre qu'elles sont inadaptées pour modéliser les durées de vie d'individus, de composants électriques, ... pour lesquels on observe un phénomène de vieillissement.

On modélise une durée de vie par une variable aléatoire réelle positive X. On peut penser à la durée de vie d'un organisme vivant, d'une entreprise, d'une situation donnée comme le chômage, etc. Supposons que X admet une densité $f : \mathbb{R}_+ \to \mathbb{R}_+$ telle que $f(t) > 0$ pour tout $t > 0$. La loi de X est caractérisée par la donnée de la densité f ou encore par la donnée de la fonction de répartition F définie par

$$F(t) = \mathbb{P}[X \leq t] = \int_0^t f(s)\, ds$$

pour tout $t \geq 0$. La **fonction de survie** est donnée par $S(t) = 1 - F(t) = \mathbb{P}[X > t]$ pour tout $t \geq 0$; elle représente la probabilité de mourir après l'instant t, i.e. de survivre à la période $[0, t]$. Il est clair que la fonction de survie caractérise la loi de X, car il en va de même de la fonction de répartition F. La fonction de survie vérifie :

— S prend ses valeurs dans l'intervalle $[0, 1]$;

— S est continue ;

— S est décroissante ;

Probabilités - Préparation à l'agrégation interne de mathématiques (90/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

— $S(0) = 1$ et $\lim_{t\to\infty} S(t) = 0$.

Remarque E.8. *La première et la troisième propriété ainsi que la limite en ∞ découlent immédiatement des propriétés correspondantes pour F. La continuité complète (et pas seulement à droite) et la valeur en 0 viennent de l'hypothèse que X est à densité sur $\mathbb{R}^+$.*

La **fonction de hasard** h est définie pour tout $t \geq 0$ par

$$h(t) = \frac{f(t)}{S(t)} = \frac{F'(t)}{1 - F(t)} = -\frac{S'(t)}{S(t)}.$$

La fonction de hasard est positive. Le terme *hasard* est un anglicisme qui signifie ici « danger » ou « risque ». En effet, la fonction de hasard s'interprète comme un taux de mortalité instantané, c'est à dire une probabilité de mort par unité de temps : pour tout $t \geq 0$,

$$\mathbb{P}\left[X \in [t, t+s] | X > t\right] = h(t) \times s + o_{s\to 0}(s),$$

ou encore :

$$h(t) = \frac{d}{ds}\Big|_{s=0} (\mathbb{P}\left[t < X < t+s \,|\, X > t\right]).$$

Pour le voir, on écrit

$$\begin{aligned}\frac{F(t+s)}{S(t)} &= \frac{\mathbb{P}\left[X < t+s\right]}{\mathbb{P}\left[X > t\right]} = \frac{\mathbb{P}\left[X < t\right] + \mathbb{P}\left[t < X < t+s\right]}{\mathbb{P}\left[X > t\right]} \\ &= 1 + \frac{\mathbb{P}\left[t < X < t+s\right]}{\mathbb{P}\left[X > t\right]} = 1 + \mathbb{P}\left[t < X < t+s \,|\, X > t\right],\end{aligned}$$

puis l'on dérive en s.

La fonction de hasard caractérise la loi, puisque l'on peut exprimer S en fonction de h par :

$$S(t) = \exp\left(-\int_0^t h(u)\, du\right).$$

On cherche souvent à modéliser les durées de vies via leur fonction de hasard h plutôt que via la densité f. Voici deux exemples concrets de fonctions de hasard :

Durée de vie humaine. La fonction h part d'une valeur positive et décroît violemment (ce pic initial correspond à la mortalité infantile), puis a un plat muni de deux bosses vers 18–22 ans (accidents deux roues) et 40 ans (accidents cardiaques), puis remonte de manière convexe (vieillesse). Les pics de la fonctions de hasard correspondent à une diminution de la durée de vie.

Durée du chômage. la fonction h a l'allure de la fonction $x \mapsto (1+x)\exp(-x)$, le pic mou correspondant au chômage de longue durée.

Durée de vie d'un composant. On a coutume de dire que la fonction de hasard typique d'un composant a une forme de baignoire : rodage (décroissance), exploitation (long plateau), usure (croissance).

On peut encore définir d'autres points de vue pour étudier la durée de vie. Une idée est de s'intéresser aux quantités *conditionnelles*. Pour tout $t \geq 0$, la **fonction de survie conditionnelle** est la probabilité de survivre encore pendant un temps s sachant qu'on a déjà vécu un temps t :

$$S(s \,|\, t) = \mathbb{P}\left[X > t+s \,|\, X > t\right].$$

Ces fonctions peuvent s'exprimer directement en fonction de h : pour $s, t > 0$,

$$S(s \,|\, t) = \frac{S(t+s)}{S(t)} = \exp\left(-\int_t^{t+s} h(u)\, du\right).$$

Fixons t. La fonction $s \mapsto S(s|t)$ vérifie toutes les hypothèses d'une fonction de survie ; on peut donc lui associer une loi (appelée « loi conditionnelle de $X - t$ sachant $X > t$ »), de fonction de répartition $F_{X-t|X>t} : s \mapsto 1 - S(s|t)$.

Supposons X intégrable. La **durée de vie moyenne restante** est définie pour chaque t comme l'espérance de la loi conditionnelle de X sachant $X \geq t$. Par le résultat indiqué dans l'exercice 3.41,

$$\begin{aligned} r(t) &= \mathbb{E}\left[X - t \,|\, X > t\right] = \mathbb{E}\left[X \,|\, X > t\right] - t \\ &:= \int_0^\infty S(s|t)\, ds. \end{aligned}$$

Probabilités - Préparation à l'agrégation interne de mathématiques (91/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

L'intégrale est finie puisqu'elle se réécrit

$$r(t) = \frac{1}{S(t)} \int_0^\infty S(s+t)\,ds = \frac{1}{S(t)} \int_t^\infty S(s)\,ds \leq \frac{1}{S(t)} \mathbb{E}[X] < \infty.$$

Remarque E.9. *Si X est une durée de vie humaine mesurée en années, $t + r(t)$ est l'*espérance de vie à t ans.

La fonction r caractérise la loi de X. En effet, il est clair que r est entièrement déterminée par la loi de X. Réciproquement, connaissant la fonction r, on a pour tout $t \geq 0$,

$$\frac{1}{r(t)} = \frac{S(t)}{\int_t^\infty S(s)\,ds}, \tag{E.1}$$

relation qu'il s'agit d'« inverser » pour trouver S en fonction de r. On reconnaît à droite (l'opposé de) la dérivée d'un logarithme : en intégrant entre 0 et t, on obtient

$$\int_0^t \frac{1}{r(s)}\,ds = \ln\Big(\int_0^\infty S(s)\Big) - \ln\Big(\int_t^\infty S(s)\,ds\Big).$$

L'intégrale $\int_0^\infty S(s)\,ds$ vaut $\mathbb{E}[X] = r(0)$, donc

$$\int_t^\infty S(s)\,ds = r(0) \exp\left(-\int_0^t \frac{1}{r(s)}\,ds\right).$$

En se rappelant de la formule (E.1), on en déduit l'expression de S en fonction de r :

$$S(t) = \frac{r(0)}{r(t)} \exp\left(-\int_0^t \frac{1}{r(s)}\,ds\right).$$

Ainsi, la durée de vie moyenne restante caractérise bien la loi.

On dispose d'une caractérisation de l'indépendance temporelle qui découle de l'absence de mémoire des lois exponentielles : pour tout $\lambda > 0$, il y a équivalence entre :

— X suit la loi exponentielle de moyenne $1/\lambda$;
— $f(t) = \lambda e^{-\lambda t}$ pour tout $t \geq 0$;
— $F(t) = 1 - e^{-\lambda t}$ pour tout $t \geq 0$;
— $S(t) = e^{-\lambda t}$ pour tout $t \geq 0$;
— $h(t) = \lambda$ pour tout $t \geq 0$ (constance de la fonction de hasard)
— $S(t \mid t_0) = e^{-\lambda t}$ pour tous $t_0, t \geq 0$ (absence de mémoire)
— $S(t + t_0) = S(t)S(t_0)$ pour tous $t_0, t \geq 0$;
— $r(t) = 1/\lambda$ pour tout $t \geq 0$.

Remarque E.10 (Fiabilité des systèmes)**.** *Les notions de fonction de survie et de fonction de hasard sont centrales dans l'étude quantitative de la fiabilité des systèmes. Pour un système simple, constitué de composants montés en série, la durée de vie du système est le minimum des durées de vie des composants ; si les composants sont en parallèle il faut considérer le maximum. Pour des systèmes plus complexes on doit définir de nouvelles notions : les fonctions de structure, les coupes minimales, les arbres de défaillance…*

Probabilités - Préparation à l'agrégation interne de mathématiques (92/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Annexe **F**

Extrêmes

F.1 Modélisation des phénomènes extrêmes

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi. Les résultats les plus classiques de la théorie des probabilités (loi des grands nombres, théorème limite central) concernent la somme $S_n = \sum_{i=1}^n X_i$. Le TLC peut en particulier s'interpréter comme une identification de la *bonne normalisation* de S_n : pour $a_n = 1/(\sqrt{n}\sigma)$ et $b_n = -\sqrt{n}\mu/\sigma$, $a_n S_n + b_n$ converge en loi (vers la loi centrée réduite $\mathcal{N}(0,1)$).

Dans de nombreuses applications ce n'est pas la somme des X_i qui importe mais leur plus grande valeur :

$$M_n = \max(X_1, \ldots, X_n).$$

Si X_i est une performance sportive, M_n est le record au temps n, si X_i est une hauteur d'eau, la valeur de M_n détermine si il y a ou non une crue... L'étude de ces grandes valeurs est appelée « théorie des valeurs extrêmes ».

Remarquons d'emblée que M_n est croissante [1], par conséquent elle converge presque-sûrement vers une variable M_∞ à valeurs dans $]-\infty, \infty]$. Cette valeur limite est en réalité déterministe.

Théorème F.1 (Limite de M_n)**.** *On a toujours*

$$\mathbb{P}\Big(\lim_{n\to\infty} M_n = x_F\Big) = 1 \quad \textit{où} \quad x_F = \sup\{x \in \mathbb{R} : F_{X_1}(x) < 1\} \in \mathbb{R} \cup \{\infty\}.$$

Le réel x_F est appelé « bord droit du support de X_1 ».

Démonstration. Notons déjà que pour tout $x \in \mathbb{R}$ et tout $n \in \mathbb{N}$ on a

$$F_{M_n}(x) = \mathbb{P}[M_n \leq x] = \mathbb{P}[X_1 \leq x] \cdots \mathbb{P}[X_n \leq x] = F_{X_1}^n(x).$$

Pour tout $x < x_F$, on a $F_{X_1}(x) < 1$ et donc

$$\mathbb{P}[M_n \leq x] = F_{X_1}^n(x) \xrightarrow[n\to+\infty]{} 0.$$

De plus, dans le cas où $x_F < \infty$, on a pour tout $x \geq x_F$, $F(x) = 1$, et donc

$$\mathbb{P}[M_n \leq x] = F_{X_1}^n(x) \to 1.$$

Ainsi, la suite $(M_n)_{n\geq 1}$ converge en probabilité vers x_F. Comme on a vu précédemment qu'elle convergeait presque sûrement (et donc en probabilité) vers M_∞, on en déduit que $M_\infty = x_F$. □

Dans la suite on va chercher à étudier le comportement de $x_F - M_n$ (si x_F est fini) ou à voir à quelle vitesse M_n tend vers l'infini (si $x_F = \infty$), ce qui reviendra à identifier de bons choix pour les suites a_n et b_n.

1. Ceci signifie que pour tout ω, la suite $M_n(\omega)$ est croissante.

Probabilités - Préparation à l'agrégation interne de mathématiques (93/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

F.2 Quatre exemples simples

Considérons d'abord le cas où les X_i sont des variables de Bernoulli indépendantes de paramètre p. Dans ce cas la loi de M_n est elle-même une loi de Bernoulli, de paramètre $1-(1-p)^n$. Cette loi converge vers une masse de Dirac en 1. On peut vérifier qu'il n'y a pas d'autre normalisation intéressante. Ceci correspond au fait qu'il n'y a pas d'« approche » du maximum : soit on l'a exactement atteint, soit ce n'est pas le cas, la suite $x_F - M_n$ est nulle à partir d'un certain rang. Ce phénomène se reproduit dès que $\mathbb{P}[X_1 = x_F] > 0$ (on dit que la loi de X_1 a un atome en x_F).

Trois autres situations sont particulièrement simples à étudier, pour X_1 suivant l'une des trois lois :

— uniforme (queue à droite nulle),

— exponentielle (queue à droite à décroissance exponentielle),

— de Cauchy (queue à droite à décroissante polynomiale).

Si X_1 suit la loi uniforme sur $[0,\theta]$, $x_F = \theta$ est fini. Le résultat suivant montre qu'en en certain sens, l'écart $(\theta - M_n)$ entre le maximum actuel et le maximum théorique se comporte comme $\frac{\theta}{n}E$ où E suit une loi exponentielle.

Théorème F.2 (Extrêmes de lois uniformes : loi de Weibull)**.** *Si X_1 suit la loi uniforme sur $[0,\theta]$ avec $\theta > 0$ alors*

$$\lim_{n\to\infty} F_{n(\theta^{-1}M_n - 1)}(x) = e^x \mathbf{1}_{\mathbb{R}_-}(x) + \mathbf{1}_{\mathbb{R}_+}(x).$$

pour tout $x \in \mathbb{R}$. La limite est la fonction de répartition de $-E$ où E suit une loi exponentielle (on dit parfois qu'il s'agit de la loi de Weibull des extrêmes, à ne pas confondre avec les lois de Weibull utilisées pour modéliser les durées de vie).

Démonstration. Pour tout $x \le 0$ on a $n^{-1}x + 1 \le 1$ et

$$\mathbb{P}\left[M_n \le \theta(n^{-1}x+1)\right] = (n^{-1}x+1)^n \to e^x. \qquad \square$$

Remarque F.3. *Cela donne la vitesse et la loi de fluctuation (non gaussienne) de l'estimateur $\widehat{\theta}_n = \max\{U_1,\dots,U_n\}$ de θ où $U_1,\dots,U_n$ sont i.i.d. de loi uniforme sur $[0,\theta]$ (il s'agit d'un modèle statistique non régulier). Le théorème* (F.2) *indique une fluctuation non normale pour l'estimateur M_n de θ.*

Pour les lois exponentielles et de Cauchy, $x_F = \infty$ et M_n tend donc presque sûrement vers l'infini ; la question est de savoir à quelle vitesse.

Théorème F.4 (Extrême de lois exponentielles : loi de Gumbel)**.** *Si X_1 est exponentielle de moyenne $1/\lambda$ alors*

$$\lim_{n\to\infty} F_{\lambda M_n - \ln(n)}(x) = e^{-e^{-x}}$$

pour tout $x \in \mathbb{R}$. La limite est la fonction de répartition d'une loi de Gumbel.

Intuitivement, (la loi de) M_n est donc proche de (la loi de) $\frac{1}{\lambda}(\ln(n) + G)$ où G suit une loi de Gumbel.

Application : on peut approcher $\mathbb{P}[M_n \le x]$ par $F(\lambda x - \ln(n))$ où F est la fonction de répartition de la loi de Gumbel. Par exemple pour $\lambda = 1/10$, $x = 50$, $n = 100$, on a

$$\mathbb{P}[M_{100} \ge 50] \approx e^{-e^{-(\lambda x - \ln(n))}} = 0{,}49023$$

tandis que le calcul exact donne

$$\mathbb{P}[M_{100} \ge 50] = 1 - (1 - e^{-\lambda x})^n = 0{,}49139.$$

Démonstration. Pour tout $x \in \mathbb{R}$ tel que $x + \ln(n) \ge 0$ (toujours vrai si n assez grand)

$$\mathbb{P}[\lambda M_n - \ln(n) \le x] = (1 - n^{-1}e^{-x})^n \to e^{-e^{-x}}. \qquad \square$$

Pour la loi de Cauchy, les grandes valeurs sont très probables, on s'attend donc à obtenir une vitesse plus rapide que $\ln(n)$.

Théorème F.5 (Extrêmes de lois de Cauchy : loi de Fréchet)**.** *Si X_1 suit la loi de Cauchy alors*

$$\lim_{n\to\infty} F_{\pi n^{-1} M_n}(x) = e^{-1/x}\mathbf{1}_{\mathbb{R}_+^*}(x)$$

pour tout $x \in \mathbb{R}$. La limite est la fonction de répartition de la loi de Fréchet.

Démonstration. Comme $\arctan(x) = \pi/2 - 1/x + \mathcal{O}_{x\to\infty}(1/x^2)$, pour tout $x \ge 0$,

$$\mathbb{P}\Big(M_n \le \frac{nx}{\pi}\Big) = \left(\int_{-\infty}^{nx\pi^{-1}} \frac{dy}{\pi(1+x^2)}\right)^n = \left(1 - \frac{1}{nx} + \mathcal{O}_{x\to\infty}(n^{-2})\right)^n \to e^{-1/x}. \qquad \square$$

Probabilités - Préparation à l'agrégation interne de mathématiques (94/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

F.3 Un résultat général

Nous avons vu dans la section précédente que lorsque X_1 suit une loi de Bernoulli, une loi uniforme, une loi exponentielle ou une loi de Cauchy, il existe une suite déterministe $(a_n, b_n)_{n\geq 1}$ avec $a_n > 0$ ainsi qu'une loi L sur $\mathbb{R}$ de fonction de répartition F tel que pour tout $x \in \mathbb{R}$,

$$\lim_{n\to\infty} F_{a_n M_n + b_n}(x) = F(x).$$

Le théorème suivant affirme que ces cas particuliers sont d'une certaine manière les seuls possibles.

Théorème F.6 (Extrêmes de Gnedenko-Fréchet-Fisher-Tippett)**.** *S'il existe une suite $(a_n, b_n)_{n\geq 1}$ et une loi L de fonction de répartition F telles que pour tout $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} F_{a_n M_n + b_n}(x) = F(x)$$

alors, à translation/dilatation près, L suit :

1. *soit une loi de Dirac ;*
2. *soit une loi de Weibull*[2] *avec $F(x) = e^{-(-x)^\alpha}\mathbf{1}_{\mathbb{R}_-}(x) + \mathbf{1}_{\mathbb{R}_+}(x)$ pour un paramètre $\alpha > 0$;*
3. *soit une loi de Gumbel avec $F(x) = e^{-e^{-x}}\mathbf{1}_{\mathbb{R}}(x)$;*
4. *soit une loi de Fréchet avec $F(x) = e^{-x^{-\alpha}}\mathbf{1}_{\mathbb{R}_+}(x)$ pour un paramètre $\alpha > 0$.*

Remarque F.7. *Il se peut qu'aucun choix de a_n, b_n ne permette d'obtenir une convergence vers une variable autre qu'un Dirac : on l'a vu pour la loi de Bernoulli, c'est également le cas si X_1 suit une loi géométrique ou une loi de Poisson.*

La théorie des extrêmes fournit des conditions nécessaires et suffisantes sur F_{X_1} pour l'appartenance aux bassins d'attraction des trois lois des extrêmes. Ces conditions portent sur la queue à droite. La loi de Weibull apparaît pour les lois dont la queue à droite est nulle (ex. loi uniforme), la loi de Gumbel apparaît pour les lois dont la queue à droite est exponentielle (ex. lois exponentielle et normale), et la loi de Fréchet apparaît pour les lois dont la queue à droite est polynomiale (lois de Cauchy, de Student, de Pareto).

Théorème F.8 (Lois des extrêmes)**.** *Les trois lois limites non-triviales sont **max-stables** : si X_1 suit une de ces lois, il existe a_n et b_n tels que $M_n \overset{d}{=} a_n X_1 + b_n$. Plus précisément,*

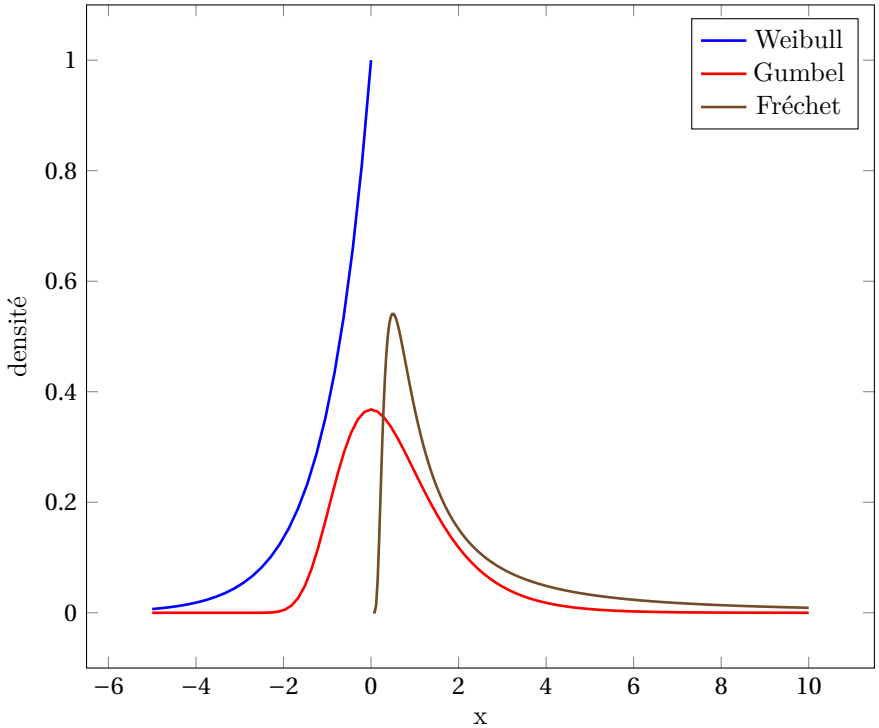
1. *si X est Weibull de paramètre α alors $M_n \overset{d}{=} n^{-1/\alpha} X_1$;*
2. *si X est Gumbel alors $M_n \overset{d}{=} X_1 + \ln(n)$;*
3. *si X est Fréchet de paramètre α alors $M_n \overset{d}{=} n^{1/\alpha} X_1$.*

De plus, pour tout $\alpha > 0$ et toute variable aléatoire X, il y a équivalence entre :

1. *$-X^{-1}$ suit une loi de Weibull de paramètre α ;*
2. *$\ln(X^\alpha)$ suit une loi de Gumbel ;*
3. *X suit une loi de Fréchet de paramètre α.*

2. Attention, la terminologie diffère de celle utilisée pour modéliser les durées de vie en fiabilité/survie.

Probabilités - Préparation à l'agrégation interne de mathématiques (95/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/



Probabilités - Préparation à l'agrégation interne de mathématiques (96/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Annexe **G**

Familles sommables et intégrales de Riemann

Le calcul des probabilités fait appel à de nombreux endroits à des calculs de somme et des calculs d'intégrales. Dans le cadre fini, on sait bien que l'on peut manipuler les sommes aisément : en modifiant l'ordre de sommation, en intervertissant plusieurs indices de sommation…

Pour généraliser la définition et les modes de calcul aux sommes infinies, on procède en *deux temps*.

1. On considère d'abord *le cas positif* : dans ce cadre, les sommes sont définies dans $[0,\infty]$ (la valeur ∞ est autorisée), et toutes les manipulations usuelles fonctionnent.
2. Pour une somme de termes *de signe quelconque*, on considère d'abord la somme des valeurs absolues. Si celle-ci a une valeur finie, on peut de nouveau faire les manipulations usuelles.

Ce même mode de définition en deux temps se retrouve pour la définition d'intégrale de Riemann de fonctions continues par morceaux ; nous l'avons également utilisée dans la définition 3.28 d'une variable aléatoire de signe quelconque. C'est, plus généralement, l'approche classique d'intégration en théorie de la mesure, qui regroupe et intègre tous ces résultats dans un même cadre théorique.

G.1 Familles sommables

Rappelons que si a_n est une suite à termes positifs, on peut toujours définir $S = \sum_n a_n = \lim_n (\sum_{k=1}^n a_k) \in [0,\infty]$, et on peut effectuer la somme dans n'importe quel ordre : pour toute permutation σ de $\mathbb{N}$,

$$\sum_n a_n = \sum_n a_{\sigma(n)}.$$

De même, si la série $\sum a_n$ est absolument convergente, c'est-à-dire si $\sum_n |a_n| < \infty$, alors elle converge et on peut effectuer la somme dans n'importe quel ordre :

$$\sum_n a_n = \sum_n a_{\sigma(n)}.$$

Si maintenant E est un ensemble dénombrable quelconque, et $(a_k)_{k\in E}$ une famille indexée par E, les propriétés précédentes permettent de donner un sens à l'expression $\sum_{k\in E} a_k$. On se donne une énumération quelconque $(k_0, k_1, \ldots, k_n, \ldots)$ de E, et on pose

1. $\sum_{k\in E} a_k = \sum_n a_{k_n}$, quand les a_k sont positifs, cette somme est dans $[0,\infty]$;
2. quand $\sum_E |a_k| < \infty$, on pose $\sum_E a_k = \sum_n a_{k_n}$. Cette somme est dans $]-\infty,\infty[$.

Les propriétés précédentes garantissent que la somme ainsi définie ne dépend pas de l'énumération choisie.

On peut également suivre l'approche en deux temps évoquée plus haut pour définir $\sum_{k\in E} a_k$ de la façon suivante :

1. Si les a_k sont positifs, on pose $\sum_{k\in E} a_k = \sup\left\{\sum_{k\in F} a_k ; F \subset E, F \text{ fini}\right\} \in [0,\infty]$.
2. Si les a_k sont quelconques, et si $\sum_{k\in E} |a_k| < \infty$, on pose $a_k^+ = \max(a_k, 0)$, $a_k^- = \max(-a_k, 0)$. Alors a^+ et a^- sont positives, de sommes finies, et l'on pose

$$\sum_{k\in E} a_k = \sum_{k\in E} a_k^+ - \sum_{k\in E} a_k^-.$$

La « philosophie » présentée dans l'introduction se traduit par le résultat suivant :

Probabilités - Préparation à l'agrégation interne de mathématiques (97/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Théorème G.1 (Fubini, cas discret). *Si $(a_{kl})_{k\in E, l\in F}$ est une famille de réels positifs, alors*

$$\sum_{k,l} a_{kl} = \sum_k \sum_l a_{kl} = \sum_l \sum_k a_{kl}.$$

La même formule est valable pour des a_{kl} de signe quelconque, si l'on suppose de plus que $\sum_{k,l} |a_{k,l}| < \infty$.

G.2 Intégration de Riemann

La notion d'intégration au programme est l'intégrale de Riemann. Sur un segment $[a,b]$, une condition suffisante de Riemann-intégrabilité est la continuité par morceaux : il existe une subdivision $(a_0, \dots, a_N)$ de $[a,b]$ telle que f est continue sur $]a_i, a_{i+1}[$ avec des limites à droite en a_i et à gauche en a_{i+1}. De même être C^k par morceaux signifie être C^k sur chacun des intervalles ouverts, et admettre, de même que ses k premières dérivées, des limites à gauche et à droite aux points a_i.

Ces notions s'étendent à un intervalle ouvert, borné ou non, en disant que f doit être continue par morceaux (respectivement C^k par morceaux) sur tout segment inclus dans I, ainsi f peut avoir un nombre dénombrable de points de discontinuité.

L'intégrale de Riemann d'une fonction continue par morceaux sur un intervalle I quelconque se définit alors de manière exactement similaire aux familles sommables, en suivant l'approche en deux temps :

1. Si f est positive, $\int_I f$ est par définition la borne supérieure des intégrales sur les segments inclus dans I.
2. Si f est de signe quelconque, et si $\int_I |f| < \infty$, on décompose f en différence de fonctions positives $f = f_+ - f_-$, et on pose $\int_I f = \int_I f_+ - \int_I f_-$.

Rappelons que si f est intégrable sur $\mathbb{R}$, son intégrale coïncide avec la limite des intégrales $\int_{-M}^{M} f(x)dx$, quand M tend vers l'infini.

On dispose là aussi du théorème de Fubini :

Théorème G.2 (Fubini, cas continu). *Si $f : I \times J \to \mathbb{R}$ est une fonction continue à valeurs positives, alors*

$$\iint_{I\times J} f(x,y)dxdy = \int_I \left(\int_J f(x,y)dy\right)dx = \int_J \left(\int_I f(x,y)dx\right)dy.$$

La même formule est valable pour une fonction de signe quelconque, si l'on suppose de plus que $\iint |f(x,y)|dxdy < \infty$.

G.3 Les deux théorèmes fondamentaux admis

Ces théorèmes sont faciles à établir et plus généraux quand on dispose de la théorie d'intégration de Lebesgue. On fixe un intervalle I.

Théorème G.3 (Convergence monotone). *Soit f_n une suite croissante de fonctions continues par morceaux convergeant simplement vers f continue par morceaux. Alors f est intégrable si et seulement si $\sup_n \int f_n < \infty$, auquel cas l'intégrale de f est la limite des intégrales des f_n.*

Théorème G.4 (Convergence dominée). *Soit f_n une suite de fonctions à valeurs complexes (continues par morceaux) convergeant simplement vers f continue par morceaux. Si $\sup_n |f_n| \le g$ avec g continue par morceaux intégrable, alors f est intégrable et $\int f_n$ converge vers $\int f$.*

Preuve partielle. Pour f_n à valeurs réelles on pose $g_n(x) = \inf_{m\ge n} f_m(x)$. La suite g_n est bien définie (les $f_m(x)$ sont minorés par $-g(x)$), et est croissante. Elle converge simplement vers f. Comme g_n est majorée par g, la suite $(\int g_n)$ est majorée donc par convergence monotone $\int f = \lim \int g_n \le \liminf \int f_n$. On fait de même dans l'autre sens. □

Probabilités - Préparation à l'agrégation interne de mathématiques (98/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Annexe **H**

Convergences pour des suites de variables aléatoires

Dans toute cette section, $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$, X, Y sont des v.a.r. définies sur un même espace de probabilités $(\Omega, \mathscr{A}, \mathbb{P})$, de lois respectives μ_n, ν_n, μ et ν, et de fonction de répartitions respectives F_n, G_n, F et G.

H.1 Convergences de variables et de lois

Une suite de variables aléatoires est — formellement — une suite de fonctions, sur un espace muni d'une mesure de probabilité $\mathbb{P}$. On peut définir plusieurs notions de convergence pour une telle suite.

Définition H.1 (Convergences)**.** *On dit que X_n converge **presque sûrement**, et on note $X_n \xrightarrow[n\to\infty]{\text{p.s.}} X$, lorsque*

$$\begin{aligned}\mathbb{P}\left[\lim_{n\to\infty} X_n = X\right] &:= \mathbb{P}\left[\left\{\omega\in\Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right]\\ &= 1.\end{aligned}$$

C'est la notion de convergence qui apparaît dans la loi forte *des grands nombres.*

*On dit que la convergence a lieu **en probabilité**, et on note $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$, quand*

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}\left[|X_n - X| \geq \varepsilon\right] = 0.$$

C'est la notion de convergence qui apparaît dans la loi faible *des grands nombres.*

*Enfin, on dit que X_n converge vers X **en moyenne** (en moyenne quadratique, en moyenne d'ordre p), si pour $p = 1$ ($p = 2$, p quelconque dans $[1,\infty]$) on a :*

$$X \in \mathrm{L}^p \quad \text{et} \quad \lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^p\right] = 0,$$

qui signifie que $X_n \to X$ dans l'espace de Banach L^p. Les cas les plus utiles sont $p = 1$ et $p = 2$ (et $p = \infty$?!).

Stabilités des convergences. Les trois convergences sont stables par combinaisons linéaires finies. Les deux premières sont stables par composition avec une fonction continue : si X_n converge (p.s. ou en probabilité) vers X et $f : \mathbb{R} \to \mathbb{R}$ est continue, alors $f(X_n)$ converge (p.s. ou en probabilité) vers $f(X)$.

Extension aux vecteurs aléatoires. Les trois types de convergence s'étendent naturellement aux vecteurs aléatoires. Pour la convergence en probabilité ou en moyenne, on peut choisir une distance ou une norme.

Définition H.2 (Convergence en loi.)**.** *On dit que X_n **converge en loi** vers X, et on note $X_n \xrightarrow[n\to\infty]{\text{loi}} X$, ou encore $X_n \xrightarrow[n\to\infty]{\text{loi}} \mu$ si, pour toute fonction f continue bornée,*

$$\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X)\right].$$

C'est la notion de convergence qui apparaît dans le théorème limite central.

La convergence en loi est équivalente à la convergence de $\mathbb{E}\left[f(X_n)\right]$ vers $\mathbb{E}\left[f(X)\right]$:

Probabilités - Préparation à l'agrégation interne de mathématiques (99/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

1. *pour toute* $f:\mathbb{R}\to\mathbb{R}$, *de classe* $\mathscr{C}^\infty$ *et à support compact ;*
2. *pour toute* f *de la forme* $f(\cdot)=e^{it\cdot}$, $t\in\mathbb{R}$ *(c'est la* convergence des fonctions caractéristiques*) ;*
3. *pour toute* f *de la forme* $\mathbf{1}_{]-\infty,x]}$, *où* x *est point de continuité de la fonction de répartition* F_X *de* X.

Dans des cas particuliers on dispose d'autres caractérisations :

— ***(sur*** $\mathbb{Z}$***)*** *on peut choisir les fonctions du type* $f=\mathbf{1}_{\{x\}}$, $x\in\mathbb{Z}$, *c'est-à-dire vérifier la* convergence des fonctions de masse *;*

— ***(sur*** $\mathbb{R}_+$***)*** *on peut utiliser la* transformée de Laplace *en prenant comme fonctions tests les* $f(\cdot)=e^{-t\cdot}$, $t\geq 0$ *;*

— ***(sur*** $\mathbb{N}$***)*** *on peut utiliser les* fonctions génératrices *en prenant* f *du type* $f(\cdot)=s^{\cdot}=e^{\ln(s)\cdot}$, *pour* $s\in]0,1[$.

L'équivalence de toutes ces définitions est admise.

Remarque H.3 (Des convergences différentes)**.** *Comme son nom l'indique, la convergence en loi ne fait intervenir les variables* (X_n) *et* X *qu'à travers leurs lois. En particulier, les* X_n *et* X *peuvent tous être définis sur des espaces différents et converger en loi. À l'inverse, la convergence en probabilité étudie la différence* X_n-X, *ce qui suppose que* X_n *et* X *soient définis sur le même espace.*

Notons toutefois que si X *est la variable constante* c, *on peut toujours donner un sens à* $X_n-X=X_n-c$. *Nous verrons dans la remarque H.5 que dans ce cas particulier, convergence en loi et en probabilité sont en fait équivalentes.*

Stabilité. La convergence en loi est stable par composition avec une fonction continue.

Extension aux vecteurs. On utilise un produit scalaire dans la transformée de Fourier (fonction caractéristique) : itX devient $i\langle t,X\rangle$.

Sommes de variables indépendantes. Les transformées de Fourier, de Laplace, et les fonctions génératrices sont particulièrement utiles pour établir la convergence en loi de sommes de variables aléatoires indépendantes : l'exponentielle transforme somme en produit, puis l'indépendance transforme espérance de produit en produit d'espérances.

Lemme H.4 (de Slutsky)**.** *Si* $X_n\xrightarrow{\text{loi}}X$ *et* $Y_n\xrightarrow{\text{loi}}Y$ *où* Y *est une constante, alors* $(X_n,Y_n)\xrightarrow{\text{loi}}(X,Y)$. *En particulier,* $X_nY_n\xrightarrow{\text{loi}}XY$ *;* $X_n+Y_n\xrightarrow{\text{loi}}X+Y$ *;* $X_n/Y_n\xrightarrow{\text{loi}}X/Y$ *si* $Y\neq 0$.

Démonstration. On montre la convergence de la fonction caractéristique du couple (X_n,Y_n) vers celle de (X,Y). Comme Y est constante, on a pour tous $(s,t)\in\mathbb{R}^2$ et tout $n\geq 1$,

$$\begin{aligned}\mathbb{E}\left[e^{itX_n+isY_n}\right]-\mathbb{E}\left[e^{itX+isY}\right]&=\mathbb{E}\left[e^{itX_n+isY_n}\right]-\mathbb{E}\left[e^{itX_n+isY}\right]+\mathbb{E}\left[e^{itX_n+isY}\right]-\mathbb{E}\left[e^{itX+isY}\right]\\&=\mathbb{E}\left[e^{itX_n}\right](\exp isY_n-\exp isY)+\mathbb{E}\left[isY\right](\exp itX_n-\exp itX).\end{aligned}$$

Le second terme converge vers 0 car X_n converge en loi vers X. On peut majorer le premier terme en module par $\mathbb{E}\left[\left|e^{isY_n}-e^{isY}\right|\right]$. On verra plus loin que comme Y est constante, Y_n converge en probabilité vers Y ; on conclut en invoquant l'uniforme continuité de $x\mapsto e^{isx}$. □

Le lemme de Slutsky permet notamment de remplacer une moyenne ou une variance par un estimateur empirique dans un théorème de convergence en loi du même type que le théorème central limite, lié par exemple à un estimateur, ce qui s'avère pratique pour fabriquer une région de confiance ou un test d'hypothèse statistique.

Pile ou face : Lemme de Slutsky et intervalle de Wald. Considérons X_n une suite de variables i.i.d. de Bernoulli de paramètre p, S_n leur somme et Z_n la variable normalisée $(S_n-np)/\sqrt{np(1-p)}$. Par la loi (faible) des grands nombres, S_n/n converge en probabilité vers p, donc $\sqrt{S_n/n(1-S_n/n)}/\sqrt{p(1-p)}$ converge en probabilité vers 1. Par le TLC, Z_n converge vers la loi normale standard $\mathcal{N}(0,1)$. Le lemme de Slutsky assure alors la convergence du couple :

$$\left(Z_n,\frac{\sqrt{S_n/n(1-S_n/n)}}{\sqrt{p(1-p)}}\right)\xrightarrow[n\to\infty]{\text{loi}}(Z,1).$$

Probabilités - Préparation à l'agrégation interne de mathématiques (100/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Posons $Z'_n = (S_n - np)/\sqrt{n(S_n/n)(1-S_n/n)}$. Comme Z'_n est le produit des deux variables du couple :

$$Z'_n = Z_n \times \frac{(S_n/n)(1-S_n/n)}{p(1-p)},$$

Z'_n converge en loi vers la loi normale $\mathcal{N}(0,1)$.

Soit alors q_α le quantile d'ordre $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$, et notons I_W l'intervalle (aléatoire) $[S_n/n \pm q_\alpha\sqrt{(S_n/n)(1-S_n/n)}/\sqrt{n}]$.

$$\begin{aligned}\mathbb{P}\left[p \in I_W\right] &= \mathbb{P}\left[\left|\frac{S_n}{n} - p\right| \le q_\alpha\sqrt{(S_n/n)(1-S_n/n)}/\sqrt{n}\right]\\ &= \mathbb{P}\left[\left|Z'_n\right| \le q_\alpha\right]\\ &\xrightarrow[n\to\infty]{} 1-\alpha,\end{aligned}$$

et l'intervalle de Wald est bien un intervalle de confiance asymptotique au niveau $1-\alpha$.

H.2 Relations entre les convergences.

L'inégalité de Hölder permet d'établir que la convergence en moyenne d'ordre $q \ge 1$ implique la convergence en moyenne d'ordre $1 \le p \le q$ car en notant $1/r = 1/p - 1/q$:

$$\mathbb{E}\left[|X_n - X|^p\right]^{1/p} \le \mathbb{E}\left[|X_n - X|^q\right]^{1/q}\mathbb{E}\left[1^r\right]^{1/r} = \mathbb{E}\left[|X_n - X|^q\right]^{1/q}.$$

L'inégalité de Markov permet d'établir que la convergence en moyenne d'ordre $p \ge 1$ entraîne la convergence en probabilité :

$$\mathbb{P}\left[|X_n - X| \ge \varepsilon\right] \le \frac{\mathbb{E}\left[|X_n - X|^p\right]}{\varepsilon^p}.$$

On passe de la convergence presque sûre à la convergence en probabilité en traduisant tout d'abord les quantificateurs $\forall$ et $\exists$ par $\cap$ et $\cup$ comme suit :

$$\{\lim_{n\to\infty} X_n = X\} = \bigcap_{k=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcap_{n\ge m}\{|X_n - X| \le 1/k\}.$$

L'union sur m étant croissante, on peut utiliser la propriété de continuité du théorème 2.11 pour écrire, pour tout k fixé,

$$\begin{aligned}1 &= \mathbb{P}\left[\lim_{n\to\infty} X_n = X\right]\\ &\le \mathbb{P}\left[\bigcup_{m=1}^{\infty}\bigcap_{n\ge m}\{|X_n - X| \le 1/k\}\right]\\ &= \lim_{m\to\infty}\mathbb{P}\left[\bigcap_{n\ge m}\{|X_n - X| \le 1/k\}\right].\end{aligned}$$

Comme $\bigcap_{n\ge m} A_n \subset A_m$, on obtient $1 \le \underline{\lim}_{m\to\infty}\mathbb{P}\left[|X_n - X| \le 1/k\right]$, ce qui permet de conclure

Si $X_n \xrightarrow{\mathbb{P}} X$ alors $X_n \xrightarrow{\text{loi}} X$. En effet, pour toute fonction f continue à support compact, la fonction f est uniformément continue par le théorème de Heine, donc pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que $|f(x) - f(y)| \le \varepsilon$ si $|x - y| \le \eta$. En découpant, on en déduit :

$$\begin{aligned}\left|\mathbb{E}\left[f(X_n)\right] - \mathbb{E}\left[f(X)\right]\right| &\le \mathbb{E}\left[|f(X_n) - f(X)|\right]\\ &= \mathbb{E}\left[|f(X_n) - f(X)|\mathbf{1}_{|X_n - X| < \eta}\right] + \mathbb{E}\left[|f(X_n) - f(X)|\mathbf{1}_{|X_n - X| \ge \eta}\right]\\ &\le \varepsilon + 2\left\|f\right\|_\infty\mathbb{P}\left[|X_n - X| \ge \eta\right].\end{aligned}$$

Comme X_n convergence en probabilité, on en déduit $\limsup_n \left|\mathbb{E}\left[f(X_n)\right] - \mathbb{E}\left[f(X)\right]\right| \le \varepsilon$, ce qui conclut puisque ε est arbitraire.

On a finalement montré les relations suivantes entre les convergences.

$$\begin{array}{ccccc} & & \boxed{\text{CV en moyenne } \mathrm{L}^p} & & \\ & & \Downarrow & & \\ & & \boxed{\text{CV en moyenne } \mathrm{L}^1} & & \\ & & \Downarrow & & \\ \boxed{\text{CV presque sûre}} & \Rightarrow & \boxed{\text{CV en probabilité}} & \Rightarrow & \boxed{\text{CV en loi}}\end{array}$$

Probabilités - Préparation à l'agrégation interne de mathématiques (101/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Remarque H.5 (Implications inverses)**.** *On peut dans certains cas « remonter les flèches ». En particulier si X est constante (et égale à c) alors la convergence en loi entraîne la convergence en probabilité : si $f_{c,\varepsilon} : \mathbb{R} \to \mathbb{R}_+$ est continue bornée avec $f(c) = 1$ et $\mathbf{1}_{]c-\varepsilon,c+\varepsilon[} \geq f_{c,\varepsilon}$ alors*

$$\mathbb{P}\left[|X_n - X| < \varepsilon\right] = \mathbb{E}\left[\mathbf{1}_{]c-\varepsilon,c+\varepsilon[}(X_n)\right] \geq \mathbb{E}\left[f_{c,\varepsilon}(X_n)\right] \to \mathbb{E}\left[f_{c,\varepsilon}(c)\right] = 1.$$

Remonter vers la convergence en moyenne revient souvent à intervertir des limites ; c'est l'objet de résultats fondamentaux de théorie de la mesure et de l'intégration que nous évoquons plus bas.

Contrexemples.

1. *Suite convergeant en probabilité mais pas presque sûrement.* On peut considérer une suite « tournante » sur $[0,1]$. Posons $X_{m+k} = \mathbf{1}_{[k,k+1[/m}(U)$ pour $0 \leq k < m$ et $m \geq 1$, où U est uniforme sur $[0,1]$. On a alors $X_n \xrightarrow{\mathbb{P}} 0$ mais $(X_n)_{n\geq 1}$ ne converge pas p.s. car « tous les ω voient passer des 0 et des 1 une infinité de fois ».
2. *Suite convergeant en loi mais pas en probabilité.* Soit X_n uniforme sur $[0,1]$ pour tout $n \geq 1$, et X uniforme sur $[0,1]$ et indépendante de $(X_n)_{n\geq 1}$. On a $X_n \xrightarrow{\text{loi}} X$ car $\mathscr{L}(X_n) = \mathscr{L}(X)$ pour tout $n \geq 1$. Mais si $\varepsilon > 0$ alors $\mathbb{P}\left[|X_n - X| \geq \varepsilon\right]$ est non nul et ne dépend pas de n, et donc $(X_n)_{n\geq 1}$ ne converge pas en probabilité vers X.
3. *Suite convergeant en probabilité mais pas en moyenne.* On pose $X_n = n\mathbf{1}_{[0,1/n]}(U)$ où U est une v.a.r. uniforme sur $[0,1]$. On a $\mathbb{P}\left[|X_n - 0| \geq \varepsilon\right] = \mathbb{P}\left[U \leq 1/n\right] = 1/n \to 0$ pour tout $\varepsilon > 0$, donc $X_n \xrightarrow{\mathbb{P}} 0$. Cependant, $\mathbb{E}\left[|X_n - 0|\right] = n\int_0^1 \mathbf{1}_{[0,1/n]}(u)\,du = 1 \not\to 0$ et donc $(X_n)_{n\geq 1}$ ne converge pas en moyenne vers 0.

H.3 Passer à la limite dans une espérance

Théorème H.6 (Convergence monotone)**.** *Si $(X_n)_{n\geq 1}$ est positive et croissante alors X_n converge presque-sûrement dans $\mathbb{R}_+ \cup \{+\infty\}$, et*

$$\lim_{n\to\infty} \mathbb{E}\left[X_n\right] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right].$$

En particulier, $\mathbb{E}\left[\lim_{n\to\infty} X_n\right] < \infty$ ssi $\lim_{n\to\infty} \mathbb{E}(X_n) < \infty$.

On utilise souvent le cas particulier avec le fait suivant : si $X \geq 0$ d'espérance finie, alors $\mathbb{P}\left[X < \infty\right] = 1$. On en déduit ainsi une preuve très sympathique du premier lemme de Borel-Cantelli :

$$\sum_n \mathbb{P}\left[A_n\right] = \sum_n \mathbb{E}\left[\mathbf{1}_{A_n}\right] = \mathbb{E}\left[\sum_n \mathbf{1}_{A_n}\right] \quad \text{et} \quad \left\{\sum_n \mathbf{1}_{A_n} = \infty\right\} = \overline{\lim_n}\, A_n.$$

La même méthode permet de réétablir très rapidement une loi forte des grands nombres pour des variables aléatoires indépendantes et bornées dans L^4 (pas forcément de même loi), comme celle vue dans la section 5.2, page 60. En effet, si $S_n := (X_1 + \cdots + X_n)$ et $\sup_n \mathbb{E}\left[X_n^4\right] < \infty$ alors $\mathbb{E}\left[S_n^4\right] = \mathscr{O}(n^2)$ (c'est le même développement de la puissance 4 que dans la preuve vue précédemment), d'où $\mathbb{E}\left[\sum_n (S_n/n)^4\right] = \sum_n \mathbb{E}\left[S_n^4\right] < \infty$. Par conséquent $\sum_n (S_n/n)^4$ est presque sûrement finie, donc avec probabilité 1, $|S_n(\omega)/n| \to 0$ comme (racine quatrième du) terme général d'une série convergente !

Lemme H.7 (Lemme de Fatou)**.** *Si $(X_n)_{n\geq 1}$ est positive alors*

$$\underline{\lim}_{n\to\infty}\, \mathbb{E}(X_n) \geq \mathbb{E}(\underline{\lim}_{n\to\infty}\, X_n).$$

Ce lemme est utile pour les suites qui sont ni monotones ni bornées. Pour se souvenir de du sens de l'inégalité, on peut retenir [1] que tout se passe comme si $\underline{\lim}_n$ était concave.

Théorème H.8 (Convergence dominée)**.** *Si $X_n \xrightarrow{\text{p.s.}} X$ et $\sup_n |X_n| \leq Y$ pour une v.a. Y telle que $\mathbb{E}\left[Y\right] < \infty$, alors on peut « passer à la limite dans l'espérance » :*

$$\lim_{n\to\infty} \mathbb{E}\left[X_n\right] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right] = \mathbb{E}\left[X\right].$$

Remarque H.9 (Modes de convergence)**.** *Si $\sup_n |X_n|$ est intégrable alors la convergence presque sûre entraîne la convergence en moyenne : X est intégrable par convergence dominée, et on peut réappliquer le théorème à $X_n - X$, dominée par $Y + |X|$, pour montrer la convergence en moyenne.*

Ce théorème permet également de déduire la convergence en loi de la convergence presque sûre sans passer par la convergence en probabilité.

1. Cette astuce suppose de se rappeler du sens de l'inégalité de Jensen...

Probabilités - Préparation à l'agrégation interne de mathématiques (102/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Enfin, on peut « remonter » de la convergence en probabilité à la convergence en moyenne sous une hypothèse plus faible que celle du théorème de convergence dominée ; c'est l'objet de la définition suivante.

Définition H.10 (Intégrabilité uniforme.)**.** *Pour toute famille de variables aléatoires $(X_i)_{i\in I}$ finie ou infinie, dénombrable ou non, les propriétés suivantes sont équivalentes :*

1. $\lim_{R\to+\infty}\sup_{i\in I}\mathbb{E}(|X_i|\mathbf{1}_{\{|X_i|\geq R\}})=0$;
2. [2] $\forall\varepsilon>0,\ \exists\delta>0,\ \forall A\in\mathscr{A},\ \mathbb{P}(A)\leq\delta\Rightarrow\sup_{i\in I}\mathbb{E}(|X_i|\mathbf{1}_A)\leq\varepsilon$;
3. [3] *il existe une fonction $\varphi:\mathbb{R}_+\to\mathbb{R}_+$ positive croissante et convexe telle que*

$$\lim_{x\to+\infty}\frac{\varphi(x)}{x}=+\infty\quad\textit{et}\quad\sup_{i\in I}\mathbb{E}(\varphi(|X_i|))<\infty.$$

Une famille qui vérifie ces propriétés est dite ***uniformément intégrable*** *(UI).*

Une famille UI ne peut être constituée que de v.a.r. intégrables. Une famille finie de v.a.r. est UI ssi les v.a.r. qui la constituent sont toutes intégrables. En vertu du troisième critère, une famille bornée dans L^p avec $p>1$ est toujours UI, car il suffit de prendre $\varphi(x)=|x|^p$. Enfin, une famille dominée dans L^1 (au sens où $\sup_{i\in I}|X_i|\leq Y$ avec $\mathbb{E}[Y]<\infty$) est toujours UI.

Grâce au troisième critère, domination dans L^1 entraîne bornitude dans « L^{1+} » ! En particulier, si une v.a.r. X est intégrable alors $\varphi(X)$ est intégrable pour une fonction φ sur-linéaire qui dépend toutefois de X. Pour comprendre ce phénomène, on peut penser aux séries de Riemann : si $\sum_n(1/n^\alpha)<\infty$ alors $\sum_n(1/n^{\alpha-\varepsilon})<\infty$ pour $\varepsilon>0$ assez petit, car la condition de convergence des séries de Riemann est « ouverte » plutôt que « fermée ».

L'uniforme intégrabilité est la condition la plus faible pour remonter de la convergence en probabilité à une convergence en moyenne, comme le montre le théorème suivant.

Théorème H.11 (Uniforme intégrabilité et interversion limite/espérance)**.** *Soit (X_n) une suite de variables intégrables, qui converge en probabilité vers une variable X intégrable.*

Alors X_n converge dans L^1 vers X si et seulement si les X_n sont uniformément intégrables.

Démonstration. En considérant $Y_n=X_n-X$ on se ramène au cas où X est nulle. On note $G(R)=\sup_n\mathbb{E}\left[|X_n|\mathbf{1}_{|X_n|\geq R}\right]$: l'uniforme intégrabilité signifie que $G(R)$ tend vers 0 en l'infini.

Si X_n converge dans L^1 vers 0, $\mathbb{E}[|X_n|]$ tend vers 0. Pour tout $\epsilon>0$, il existe I_ϵ fini tel que $\mathbb{E}[|X_n|]\leq\epsilon$ si $n\notin I_\epsilon$, donc

$$G(R)\leq\max_{n\in I_\epsilon}\mathbb{E}\left[|X_n|\mathbf{1}_{X_n}>R\right]+\epsilon.$$

On prend la limite supérieure en R : à droite le max porte sur un ensemble fini et chaque terme tend vers 0 par convergence dominée. Par conséquent

$$\limsup G(R)\leq\epsilon,$$

pour tout ϵ positif, et finalement $G(R)$ tend vers 0.

Réciproquement supposons que les X_n sont UI et montrons que $\mathbb{E}[|X_n|]$ tend vers 0. L'idée est de faire un double découpage : les petites valeurs contribuent peu à l'espérance, les valeurs moyennes sont contrôlées par la convergence en probabilité, et les grandes valeurs par l'uniforme intégrabilité. Plus précisément pour tout couple (ϵ,R) on a :

$$\begin{aligned}\mathbb{E}[|X_n|]&=\mathbb{E}\left[|X_n|\mathbf{1}_{[0,\epsilon]}(|X_n|)\right]+\mathbb{E}\left[|X_n|\mathbf{1}_{[\epsilon,R]}(|X_n|)\right]+\mathbb{E}\left[|X_n|\mathbf{1}_{[R,\infty[}(|X_n|)\right]\\&\leq\epsilon+R\mathbb{P}[|X_n|\geq\epsilon]+G(R).\end{aligned}$$

On prend la limite supérieure en n — le deuxième terme disparaît puisque X_n converge en probabilité :

$$\limsup_n\mathbb{E}[|X_n|]\leq\epsilon+G(R).$$

On fait tendre ϵ vers 0 et R vers l'infini pour conclure. □

Sur le plan topologique, une famille est UI ssi elle est faiblement relativement compacte dans L^1.

2. Critère « $\varepsilon-\delta$ ».
3. Critère de Charles-Jean Étienne Gustave Nicolas de la Vallée Poussin. Ceci n'est pas un canular.

Probabilités - Préparation à l'agrégation interne de mathématiques (103/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Probabilités - Préparation à l'agrégation interne de mathématiques (104/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Annexe I

Fonctions caractéristiques et vecteurs gaussiens

I.1 Fonction caractéristique

La loi des variables et vecteur aléatoires ainsi que leur indépendance peuvent être étudiées au moyen de transformées, chacune correspondant à une classe de fonctions test particulière (ces classes sont liées par changement de variable). Pour les variables et vecteurs discrets, c'est la fonction génératrice qu'on a coutume d'utiliser pour sa simplicité. Pour les variables aléatoires positives, c'est plutôt la transformée de Laplace qui est utilisée. Plus généralement, pour des vecteurs aléatoires quelconques, ont utilise la fonction caractéristique, ou transformée de Fourier. La fonction caractéristique des vecteurs aléatoires gaussiens possède des propriétés remarquables. Soulignons que ces transformées sont avant tout liées aux lois, plutôt qu'aux variables ou vecteurs qui suivent ces lois.

Définition I.1 (Fonction caractéristique ou transformée de Fourier)**.** *La fonction caractéristique d'une v.a.r. est la fonction $\Phi_X : \mathbb{R} \to \{z \in \mathbb{C} : |z| \leq 1\}$ définie pour tout $t \in \mathbb{R}$ par $\Phi_X(t) = \mathbb{E}\left[e^{itX}\right]$. Plus généralement, la fonction caractéristique d'un vecteur aléatoire X de $\mathbb{R}^d$ est la fonction $\Phi_X : \mathbb{R}^d \to \{z \in \mathbb{C} : |z| \leq 1\}$ définie pour tout $t \in \mathbb{R}^d$ par*

$$\Phi_X(t) = \mathbb{E}\left[e^{i\langle t, X\rangle}\right].$$

La fonction caractéristique est liée à la fonction génératrice : $g_X(e^{it}) = \Phi_X(t)$. Si X est une v.a.r. telle que X^k est intégrable pour tout $0 \leq k \leq n$ alors Φ_X est n fois dérivable en 0 et $\Phi_X^{(k)}(0) = i^k \mathbb{E}\left[X^k\right]$ pour tout $1 \leq k \leq n$. La transformée de Laplace définie par

$$t \in \mathbb{R}^n \mapsto \mathbb{E}\left[e^{\langle t, X\rangle}\right] \in \mathbb{R}_+ \cup \{\infty\}$$

n'a pas l'avantage d'être partout finie comme Φ_X. Le calcul effectif de Φ_X peut être mené grâce à la formule du transfert, en utilisant au besoin $e^{i\theta} = \cos(\theta) + i\sin(\theta)$.

Théorème I.2 (Caractérisation de la loi – Admis)**.** *Deux vecteurs aléatoires de $\mathbb{R}^d$ ont même loi si et seulement si ils ont même fonction caractéristique.*

Si X est un vecteur aléatoire de $\mathbb{R}^d$ et si $t \in \mathbb{R}$ et $s \in \mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ alors $t\langle s, X\rangle = \langle ts, X\rangle$ est une v.a.r. et $\langle s, X\rangle s$ est la projection de X sur la droite $\mathbb{R}s$.

Corollaire I.3 (Cramér-Wold – Caractérisation par projections)**.** *La loi d'un vecteur aléatoire X de $\mathbb{R}^d$ est caractérisée par les l'ensemble des lois de $\langle s, X\rangle$ pour tout $s \in \mathbb{S}^{d-1}$.*

Corollaire I.4 (Caractérisation de l'indépendance)**.** *Deux vecteurs aléatoires X et Y de $\mathbb{R}^d$ et $\mathbb{R}^{d'}$ sont indépendants si et seulement si pour tous $s \in \mathbb{R}^d$ et $t \in \mathbb{R}^d$*

$$\Phi_{(X,Y)}(s,t) = \Phi_X(s)\Phi_Y(t).$$

En particulier, si X et Y sont indépendants et $d = d'$ alors pour tout $t \in \mathbb{R}^d$,

$$\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t).$$

Soit X une v.a.r. telle que X^n est intégrable pour tout $n \in \mathbb{N}$. On dit que la loi de X est caractérisée par ses moments lorsque pour toute v.a.r. Y, si on a $\mathbb{E}[Y^n] = \mathbb{E}[X^n]$ pour tout $n \in \mathbb{N}$ alors la v.a.r. Y a la même loi que X. Le résultat suivant entraîne que la loi normale standard $\mathcal{N}(0,1)$ et la loi exponentielle sont caractérisées par leurs moments.

Probabilités - Préparation à l'agrégation interne de mathématiques (105/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Théorème I.5 (Théorème des moments de Stieltjes)**.** *Soit X une v.a.r. avec X^n intégrable pour tout $n \in \mathbb{N}$. Posons $m_n = \mathbb{E}[X^n]$. Les propositions suivantes sont équivalentes :*

1. *Φ_X est analytique sur un voisinage de 0 ;*
2. *Φ_X est analytique sur $\mathbb{R}$;*
3. *$\overline{\lim}_{n\to\infty} \left(\frac{1}{n!}|m_n|\right)^{\frac{1}{n}} < \infty$.*

Si ces conditions sont vérifiées alors la loi de X est caractérisée par ses moments. En particulier, une loi à support compact est caractérisée par ses moments.

La formule de Stirling donne $(1/n!)^{1/n} = \mathcal{O}_{n\to\infty}(1/n)$. Par conséquent, si $\overline{\lim}_{n\to\infty} \frac{1}{n}|m_n|^{1/n} < \infty$, alors la loi de X est caractérisée par ses moments.

Démonstration. Pour tout $n \in \mathbb{N}$, on a $\mathbb{E}[|X|^n] < \infty$ et donc Φ_X est n fois dérivable sur $\mathbb{R}$. De plus, $\Phi_X^{(n)}$ est continue sur $\mathbb{R}$ et pour tout $t \in \mathbb{R}$,

$$\Phi_X^{(n)}(t) = \mathbb{E}\left[(iX)^n e^{itX}\right].$$

En particulier, $\Phi_X^{(n)}(0) = i^n m_n$, et la série de Taylor de Φ_X en 0 est déterminée par la suite $(m_n)_{n\geq 1}$. Comme le rayon de convergence r d'une série entière $\sum_n a_n z^n$ est donné par la formule de Hadamard $r^{-1} = \overline{\lim}_n |a_n|^{\frac{1}{n}}$, on obtient l'équivalence de 1 et 3 (prendre $a_n = i^n m_n/n!$). D'autre part, comme pour tout $n \in \mathbb{N}$ et tous $s, t \in \mathbb{R}$,

$$\left| e^{isx}\left(e^{itx} - 1 - \frac{itx}{1!} - \cdots - \frac{(itx)^{n-1}}{(n-1)!}\right)\right| \leq \frac{|tx|^n}{n!},$$

on a pour tout $n \in \mathbb{N}$ pair et tous $s, t \in \mathbb{R}$,

$$\left|\Phi_X(s+t) - \Phi_X(s) - \frac{t}{1!}\Phi_X'(s) - \cdots - \frac{t^{n-1}}{(n-1)!}\Phi_X^{(n-1)}(s)\right| \leq m_n \frac{|t|^n}{n!},$$

qui montre que $3 \Rightarrow 2$. Comme $2 \Rightarrow 1$, on a bien équivalence de 1-2-3. Si X est bornée, alors $\sup_n |m_n|^{\frac{1}{n}} < \infty$ et donc 3 a lieu en utilisant la formule de Stirling. Si 3 a lieu alors les arguments précédents donnent un $r > 0$ tel que Φ_X est développable en série entière en tout $x \in \mathbb{R}$ avec un rayon de convergence $\geq r$. De proche en proche, on obtient que Φ_X est caractérisée par ses dérivées en zéro. □

I.2 Application aux vecteurs gaussiens

Théorème I.6 (Vecteurs et lois gaussiennes)**.** *Si X est un vecteur aléatoire de $\mathbb{R}^d$ de moyenne m et de matrice de covariance Σ alors les propriétés suivantes sont équivalentes :*

1. *Toute combinaison linéaire des composantes de X suit une loi normale sur $\mathbb{R}$;*
2. *La fonction caractéristique de X est donnée pour tout $t \in \mathbb{R}^d$ par*

$$\Phi_X(t) = \mathbb{E}\left(e^{i\langle t, X\rangle}\right) = \exp\left(i\langle t, m\rangle - \frac{1}{2}\langle t, \Sigma t\rangle\right);$$

3. *$\mathcal{L}(X) = \mathcal{L}(m + AZ)$, où A est une matrice de dimension $d \times d$ vérifiant $AA^\top = \Sigma$ et Z est un vecteur aléatoire de $\mathbb{R}^d$ à composantes indépendantes et de loi $\mathcal{N}(0,1)$.*

On dit alors que X est un *vecteur gaussien*. Sa loi est caractérisée par son vecteur moyenne m et sa matrice de covariance Σ. Elle est notée $\mathcal{N}(m, \Sigma)$. On dit que c'est une loi gaussienne sur $\mathbb{R}^d$. La loi $\mathcal{N}(0, I_d)$ de Z est appelée *loi gaussienne standard*.

Démonstration. Une combinaison linéaire des composantes de X s'écrit $\langle u, X\rangle = u^\top X$ où u est un vecteur colonne déterministe de $\mathbb{R}^d$. L'équivalence 1) $\Leftrightarrow$ 2) découle donc de l'expression de la fonction caractéristique des lois gaussiennes sur $\mathbb{R}$, et du fait que la fonction caractéristique caractérise la loi. Cette dernière propriété montre également que la loi de X est caractérisée par m et Σ. L'équivalence 1) $\Leftrightarrow$ 3) provient du théorème 4.23 sur la racine carrée matricielle, associé au théorème 4.19 sur la transformation linéaire. □

Exemple I.7 (Nécessaire mais pas suffisant)**.** *Les composantes d'un vecteur gaussien sont gaussiennes, mais la réciproque est fausse. En effet, soit $X = (Y, \varepsilon Y)$ un vecteur aléatoire de $\mathbb{R}^2$ où Y et ε sont indépendantes avec $Y \sim \mathcal{N}(0,1)$ sur $\mathbb{R}$ et ε de loi de Rademacher symétrique : $\mathbb{P}[\varepsilon = \pm 1] = 1/2$. Les composantes Y et εY de X sont gaussiennes mais la combinaison linéaire $Y + \varepsilon Y$ ne l'est pas car $\mathbb{P}[Y + \varepsilon Y = 0] = \mathbb{P}[\varepsilon = -1] = 1/2$. De plus, $\mathrm{Cov}(Y, \varepsilon Y) = \mathbb{E}\left[Y^2\right]\mathbb{E}[\varepsilon] = 0$ mais Y et εY ne sont pas indépendantes.*

Probabilités - Préparation à l'agrégation interne de mathématiques (106/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Théorème I.8 (Existence de densité)**.** *La loi gaussienne $\mathcal{N}(m,\Sigma)$ sur $\mathbb{R}^d$ admet une densité de probabilité si et seulement si Σ est inversible donnée dans ce cas par*

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det\Sigma}} \exp\Big(-\frac{1}{2}\langle (x-m), \Sigma^{-1}(x-m)\rangle\Big).$$

Démonstration. Soit A une racine carrée matricielle de Σ de même rang $p \le d$ que Σ. Soit Z un vecteur gaussien standard Z de loi $\mathcal{N}(0,\mathrm{I}_d)$, de sorte que $m + AZ \sim \mathcal{N}(m,\Sigma)$. La loi $\mathcal{N}(m,\Sigma)$ est portée par le sous-espace affine $E = \{Az + m \text{ avec } z \in \mathbb{R}^d\}$ de dimension p. Si $p < d$, alors $E \neq \mathbb{R}^d$ et $\mathcal{N}(m,\Sigma)$ n'a pas de densité. De plus, $E = \mathbb{R}^d$ si et seulement si $p = d$, c'est-à-dire si et seulement si Σ est inversible. On peut choisir les composantes de Z indépendantes et de même loi $\mathcal{N}(0,1)$ sur $\mathbb{R}$. La loi de Z admet alors la densité de probabilité f_Z donnée, pour tout $z \in \mathbb{R}^d$, par

$$f_Z(z) = \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z_k^2\right) = \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}\|z\|_2^2\right).$$

Si $X \sim \mathcal{N}(m,\Sigma)$, alors pour toute indicatrice de pavé $h : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbb{E}\left[h(X)\right] = \mathbb{E}\left[h(AZ+m)\right] = \int_{\mathbb{R}^d} h(Az+m) f_Z(z)\, dz.$$

Si Σ est inversible, le changement de variable affine $x = Az + m$ est un difféomorphisme de $\mathbb{R}^d$ dans lui-même, de jacobien non nul égal à $\det(A^{-1})$. La décomposition $\Sigma = AA^\top$ entraîne que $|\det(A)| = \sqrt{\det(\Sigma)}$ et $\Sigma^{-1} = (AA^\top)^{-1} = (A^{-1})^\top A^{-1}$. On en déduit que

$$\mathbb{E}[h(X)] = \frac{1}{(2\pi)^{d/2}\sqrt{\det\Sigma}} \int_{\mathbb{R}^d} h(x) \exp\Big(-\frac{1}{2}(x-m)^\top \Sigma^{-1}(x-m)\Big)\, dx$$

d'où la formule annoncée pour la densité f. □

Théorème I.9 (Indépendance des composantes)**.** *Pour tout vecteur gaussien X de $\mathbb{R}^d$, les trois propriétés suivantes sont équivalentes :*

1. *les composantes $X_1, \ldots, X_d$ sont mutuellement indépendantes ;*
2. *les composantes $X_1, \ldots, X_d$ sont deux à deux indépendantes ;*
3. *la matrice de covariance Σ de X est diagonale.*

En particulier, un vecteur aléatoire gaussien est gaussien standard si et seulement si ses composantes sont indépendantes et de même loi normale centrée réduite $\mathcal{N}(0,1)$ sur $\mathbb{R}$.

Démonstration. Les implications 1) ⇒ 2) et 2) ⇒ 3) découlent des définitions. Vérifions que 3) ⇒ 1). Si on a $\Sigma = \mathrm{Diag}(\sigma_1^2, \ldots, \sigma_d^2)$, alors pour tout $t \in \mathbb{R}^d$,

$$\Phi_X(t) = \exp\Big(i\langle t, m\rangle - \frac{1}{2}\langle t, \Sigma t\rangle\Big) = \prod_{k=1}^{d} \exp\Big(i t_k m_k - \frac{1}{2}\sigma_k^2 t_k^2\Big) = \prod_{k=1}^{d} \Phi_{X_k}(t_k).$$ □

Les lois gaussiennes sont stables par transformation affine. En effet, si $X \sim \mathcal{N}(m,\Sigma)$ et $A \in \mathcal{M}_{p,d}(\mathbb{R})$ et $b \in \mathbb{R}^p$ alors $AX + b \sim \mathcal{N}(Am + b, A\Sigma A^\top)$. En particulier, $\mathcal{N}(0,\sigma^2\mathrm{I}_d)$ est invariante par rotation et symétries car si $X \sim \mathcal{N}(0,\sigma^2\mathrm{I}_d)$ et si A est $d \times d$ orthogonale alors AX a la même loi que X. Le théorème de Cochran raffine l'étude de $\mathcal{N}(0,\sigma^2\mathrm{I}_d)$.

Théorème I.10 (Cochran)**.** *Soit X un vecteur colonne aléatoire de $\mathbb{R}^n$ de loi $\mathcal{N}(m,\sigma^2\mathrm{I}_n)$ et $\mathbb{R}^n = E_1 \oplus \cdots \oplus E_p$ une décomposition de $\mathbb{R}^n$ en somme directe de p sous-espaces vectoriels orthogonaux de dimensions $d_1, \ldots, d_p$ avec $d_1 + \cdots + d_p = n$. Soit $\mathbf{P}_k$ la matrice du projecteur orthogonal sur E_k et $Y_k = \mathbf{P}_k X$ la projection orthogonale de X sur E_k.*

1. *Les projections $(Y_1, \ldots, Y_p)$ sont des vecteurs gaussiens indépendants et*

$$Y_k \sim \mathcal{N}(\mathbf{P}_k m, \sigma^2 \mathbf{P}_k).$$

2. *Les variables aléatoires $\|Y_1 - \mathbf{P}_1 m\|_2^2, \ldots, \|Y_p - \mathbf{P}_p m\|_2^2$ sont indépendantes et*

$$\sigma^{-2}\|Y_k - \mathbf{P}_k m\|_2^2 \sim \chi^2(d_k).$$

Probabilités - Préparation à l'agrégation interne de mathématiques (107/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Démonstration. On se ramène d'abord au cas où $m = 0$ par translation. Le vecteur aléatoire $Y = (Y_1, \dots, Y_p)^\top$ de $\mathbb{R}^{np}$ s'écrit $Y = AX$ où A est la matrice de dimension $np \times n$

$$A = \begin{pmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_p \end{pmatrix}.$$

Il en découle que Y suit la loi $\mathcal{N}(0, \sigma^2 AA^\top)$. Pour tout $1 \le i \le p$, on a $\mathbf{P}_i = \mathbf{P}_i^\top = \mathbf{P}_i^2$. De plus, $\mathbf{P}_i \mathbf{P}_j = 0$ si $1 \le i \ne j \le p$ car $E_i \perp E_j$. Par conséquent, $AA^\top = \mathrm{Diag}(\mathbf{P}_1, \dots, \mathbf{P}_p)$ est diagonale par blocs. On peut déduire du théorème I.9 que $Y_1, \dots, Y_p$ sont des vecteurs gaussiens indépendants avec $Y_k \sim \mathcal{N}(0, \sigma^2 \mathbf{P}_k)$ pour tout $1 \le k \le p$. En particulier, les variables aléatoires $\|Y_1\|_2^2, \dots, \|Y_p\|_2^2$ sont indépendantes. Il reste à déterminer leur loi. Pour tout $1 \le k \le p$, soit $B_k = \{e_{k,1}, \dots, e_{k,d_k}\}$ une base orthonormée de E_k. La réunion $B_1 \cup \dots \cup B_p$ constitue une base orthonormée de $\mathbb{R}^n$. Le vecteur X s'écrit dans cette base $X = Y_1 + \dots + Y_p$ avec $Y_k = a_{k,1} e_{k,1} + \dots + a_{k,d_k} e_{k,d_k}$ où $a_{k,i} = \langle X, e_{k,i} \rangle$. L'invariance par transformation orthogonale de la loi $\mathcal{N}(0, \sigma^2 \mathrm{I}_n)$ implique que les variables aléatoires $a_{k,i}$ sont indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$. Il en découle que pour tout $1 \le k \le p$,

$$\sigma^{-2} \|Y_k\|^2 = \sigma^{-2}(a_{k,1}^2 + \dots + a_{k,d_k}^2) \sim \chi^2(d_k). \qquad \square$$

I.3 Applications en statistiques

Commençons par montrer une version multivariée du théorème limite central, en admettant le résultat suivant.

Théorème I.11 (Paul Lévy – Admis)**.** *Si $(X_n)_{n \ge 1}$ et X sont des vecteurs aléatoires de $\mathbb{R}^d$ et si X admet une densité continue alors les propriétés suivantes sont équivalentes :*

1. $\lim_{n \to \infty} \mathbb{E}\left[f(X_n)\right] = \mathbb{E}\left[f(X)\right]$ *pour toute fonction continue et bornée $f : \mathbb{R}^d \to \mathbb{R}$;*
2. $\lim_{n \to \infty} \mathbb{E}\left[f(X_n)\right] = \mathbb{E}\left[f(X)\right]$ *pour toute indicatrice f de pavé ou de boule ;*
3. $\lim_{n \to \infty} \Phi_{X_n}(t) = \Phi_X(t)$ *pour tout $t \in \mathbb{R}^d$.*

Corollaire I.12 (Théorème de la limite centrale multivarié)**.** *Soit $(X_n)_{n \ge 1}$ une suite de vecteurs aléatoires de $\mathbb{R}^d$ indépendants et de même loi, dont les composantes sont de carré intégrable. Alors, en notant m et Σ le vecteur moyenne et la matrice de covariance de X_1, on a, avec $X \sim \mathcal{N}(0, \Sigma)$, pour tout pavé ou boule B de $\mathbb{R}^d$,*

$$\lim_{n \to \infty} \mathbb{P}\left(\sqrt{n}\left(\frac{X_1 + \dots + X_n}{n} - m\right) \in B\right) = \mathbb{P}\left[X \in B\right].$$

Démonstration. Le théorème de Paul Lévy ramène le problème à la convergence ponctuelle des fonctions caractéristiques vers celle de la loi gaussienne $\mathcal{N}(0, \Sigma)$. Quitte à remplacer les X_k par $X_k - m$, on peut supposer que $m = 0$. Comme $X_1, \dots, X_n$ sont des vecteurs aléatoires indépendants et de même loi, on a pour tout $t \in \mathbb{R}^d$

$$\Phi_{\frac{X_1 + \dots + X_n}{\sqrt{n}}}(t) = \mathbb{E}\left(\exp\left(i\langle \frac{t}{\sqrt{n}}, X_1 \rangle + \dots + i\langle \frac{t}{\sqrt{n}}, X_n \rangle\right)\right) = \left(\Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Pour tout $t \in \mathbb{R}^n$, la v.a.r. $\langle t, X_1 \rangle$ a pour moyenne 0 et pour variance $\langle t, \Sigma t \rangle$. Une formule de Taylor à l'ordre 2 en 0 pour $\Phi_{\langle t, X_1 \rangle}$ donne $\Phi_{X_1}(t) = 1 + \frac{1}{2}\langle t, \Sigma t \rangle + o_{t \to 0}(\|t\|_2^2)$ d'où

$$\Phi_{\frac{X_1 + \dots + X_n}{\sqrt{n}}}(t) = \left(1 + \frac{1}{2n}\langle t, \Sigma t \rangle + \|t\|_2^2 o_{n \to \infty}\left(\frac{1}{n}\right)\right)^n \to \exp\left(\frac{1}{2}\langle t, \Sigma t \rangle\right) = \Phi_{\mathcal{N}(0,\Sigma)}(t).$$

$\square$

Le théorème de Cochran permet une étude fine des échantillons gaussiens ce qui justifie de nombreuses applications statistiques.

Corollaire I.13 (Échantillons gaussiens)**.** *Soient $X_1, \dots, X_n$ des v.a.r. de loi normale $\mathcal{N}(m, \sigma^2)$ avec $\sigma^2 > 0$, de* moyenne empirique *et* variance empirique *définies par*

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \qquad \text{et} \qquad S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \overline{X}_n)^2.$$

Probabilités - Préparation à l'agrégation interne de mathématiques (108/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Alors les variables aléatoires $\overline{X}_n$ et S_n^2 sont indépendantes avec

$$\overline{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \qquad \text{et} \qquad \frac{(n-1)}{\sigma^2} S_n^2 \sim \chi^2(n-1).$$

De plus, la moyenne empirique studentisée T_n *vérifie*

$$T_n = \sqrt{n}\left(\frac{\overline{X}_n - m}{S_n}\right) \sim t(n-1).$$

Démonstration. Soit 1_n le vecteur de $\mathbb{R}^n$ dont toutes les coordonnées sont égales à 1. La matrice de la projection orthogonale sur $E_1 = \mathbb{R}1_n$ est donnée par

$$\mathbf{P}_1 = \frac{1_n 1_n^\top}{\|1_n\|^2} = \frac{1}{n} 1_n 1_n^\top.$$

Le sous-espace $E_2 = E_1^\perp$ est de dimension $n-1$ et la matrice de la projection orthogonale sur E_2 est $\mathbf{P}_2 = \mathrm{I}_n - \mathbf{P}_1$. On a $Y_1 = \mathbf{P}_1 X = \overline{X}_n 1_n$ et $Y_2 = \mathbf{P}_2 X = (X_1 - \overline{X}_n, \dots, X_n - \overline{X}_n)^\top$, ce qui entraîne $\|Y_2\|^2 = (n-1)S_n^2$. Le théorème de Cochran permet de conclure. □

Le théorème de Cochran permet également de justifier le célèbre test d'adéquation du χ^2. Bien qu'il ne s'agisse pas d'une véritable distance, on appelle distance du χ^2 entre deux lois de probabilité p et q sur un ensemble fini $\{1, \dots, k\}$ le nombre réel positif

$$D(p,q) = \sum_{i=1}^{k} \frac{(p_i - q_i)^2}{p_i}.$$

Cette quantité asymétrique en p et q vaut $+\infty$ si l'un des p_i est nul.

Théorème I.14 (Test d'adéquation du χ^2)**.** *Soit $p = (p_1, \dots, p_k)$ une loi sur $\{1, \dots, k\}$ et X_1,..., X_n des v.a.r. sur $\{1, \dots, k\}$ indépendantes et de loi $q = (q_1, \dots, q_k)$. On définit les effectifs théoriques $n_1, \dots, n_k$, empiriques $N_1, \dots, N_k$, et la loi $\widehat{p} = (\widehat{p}_1, \dots, \widehat{p}_k)$ par*

$$n_i = np_i \quad \text{et} \quad N_i = \mathbf{1}_{\{X_1 = i\}} + \dots + \mathbf{1}_{\{X_n = i\}} \quad \text{et} \quad \widehat{p}_i = \frac{N_i}{n}$$

Supposons que $p_1 > 0, \dots, p_k > 0$. Considérons la distance du χ^2 normalisée

$$D_n = nD(p, \widehat{p}) = n \sum_{i=1}^{k} \frac{(p_i - \widehat{p}_i)^2}{p_i} = \sum_{i=1}^{k} \frac{(n_i - N_i)^2}{n_i}.$$

Si $p = q$ alors pour tout $t \in \mathbb{R}$,

$$\lim_{n \to \infty} F_{D_n}(t) = F_{\chi^2(k-1)}(t)$$

Si $p \neq q$ alors avec probabilité 1,

$$\lim_{n \to \infty} D_n = +\infty.$$

Démonstration. Supposons que $p \neq q$. La loi forte des grands nombres du théorème 5.5 entraîne qu'avec probabilité 1, on a $\lim_{n \to \infty} N_i / n \to q_i$ pour tout $1 \leq i \leq k$ et donc

$$\lim_{n \to \infty} \frac{D_n}{n} = \sum_{i=1}^{k} \frac{(p_i - q_i)^2}{p_i} = D(p,q)$$

et comme $D(p,q) > 0$ car $p \neq q$, on obtient bien $\lim_{n \to \infty} D_n = +\infty$. Supposons à présent au contraire que $p = q$. Pour $1 \leq j \leq n$, soit V_j le vecteur aléatoire de $\mathbb{R}^k$ donné par

$$(V_j)_i = \frac{1}{\sqrt{p_i}} (\mathbf{1}_{\{X_t = i\}} - p_i).$$

Les vecteurs $V_1, \dots, V_n$ sont indépendantes et de même loi, et cette loi est centrée et de matrice de covariance $\Sigma = I_k - \sqrt{p}\sqrt{p}^\top$ avec $\sqrt{p}^\top = (\sqrt{p_1}, \dots, \sqrt{p_k})$. Le théorème I.12 de la limite centrale multivarié entraîne alors que pour tout pavé ou boule B de $\mathbb{R}^k$,

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{1}{\sqrt{n}}(V_1 + \dots + V_n) \in B\right) = \mathbb{P}\left[Z \in B\right]$$

Probabilités - Préparation à l'agrégation interne de mathématiques (109/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

où $Z \sim \mathcal{N}(0,\Sigma)$. Soit $\mathrm{Vect}(\sqrt{p})$ le sous-espace de $\mathbb{R}^k$ engendré par $\sqrt{p}$ et soit $H = \sqrt{p}\sqrt{p}^\top$ la matrice de projection orthogonale sur $\mathrm{Vect}(\sqrt{p})$. La matrice de projection orthogonale sur $\mathrm{Vect}(\sqrt{p})^\perp$ est $I_k - H = \Sigma$. Cette matrice est de rang $k-1$ car H est de rang 1, et le théorème I.10 de Cochran donne $\|Z\|_2^2 \sim \chi^2(k-1)$. Il ne reste plus qu'à observer que

$$D_n = \left\| \frac{1}{\sqrt{n}}(V_1 + \cdots + V_n) \right\|_2^2.$$

□

Dans la pratique, on connaît p mais pas q, et on souhaite décider au vu de $X_1,\ldots, X_n$ si $p = q$ ou non. Cette décision est prise au moyen d'un test d'adéquation asymptotique. Plus précisément, on fixe un niveau de confiance $\alpha \in (0,1)$ comme par exemple $\alpha = 0,05$, puis on détermine le quantile a_α d'ordre $1-\alpha$ de la loi du chi-deux $\chi^2(k-1)$, ce qui donne la région d'acceptation du test $\mathscr{A}_\alpha = [0, a_\alpha]$. La règle de décision est la suivante :

si $D_n \in \mathscr{A}_\alpha$ alors on accepte l'hypothèse $p = q$ et sinon on la rejette.

La probabilité de rejeter à tort tend vers α quand $n \to \infty$ (erreur de première espèce).
La probabilité d'accepter à tort tend vers 0 quand $n \to \infty$ (erreur de seconde espèce).
À $X_1,\ldots, X_n$ fixés, plus α est petit, moins on rejette à tort mais plus on accepte à tort.

Probabilités - Préparation à l'agrégation interne de mathématiques (110/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Annexe J

Combinatoire, loi de Poisson, et partitions

En combinatoire, le n^{e} nombre de Bell B_n compte le nombre de partitions d'un ensemble à n éléments. On a $B_0 = 1$, $B_1 = 1$, $B_2 = 2$, et $(B_n)_{n\geq 0}$ vérifie la récurrence

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$$

qui se démontre de la manière suivante : pour choisir une partition de $\{1,\dots,n+1\}$ on choisit le nombre k d'éléments qui n'appartiennent pas au même bloc que 1, puis ces k éléments parmi n, puis on partitionne ces k éléments avec les B_k possibilités. La formule de récurrence se réécrit de la manière suivante :

$$\frac{B_{n+1}}{n!} = \sum_{k_1+k_2=n} \frac{1}{k_1!}\frac{B_{k_2}}{k_2!}$$

ce qui donne l'identité des séries entières suivante[1] :

$$\sum_{n=0}^{\infty} \frac{B_{n+1}}{n!} z^n = \sum_{k_1=0}^{\infty} \frac{1}{k_1!} z^{k_1} \sum_{k_2=0}^{\infty} \frac{B_{k_2}}{k_2!} z^{k_2} = e^z \sum_{n=0}^{\infty} \frac{B_n}{n!} z^n,$$

qui s'écrit $G'(z) = e^z G(z)$ où $G(z) = \sum_{n=0}^{\infty} \frac{B_n}{n!} z^n$, ce qui donne

$$G(z) = e^{e^z - 1}.$$

On reconnaît la fonction génératrice de la loi de Poisson de paramètre 1, prise au point e^z. Il en découle que B_n est le moment d'ordre n de la loi de Poisson de paramètre 1, c'est-à-dire

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}$$

(formule dite de Dobinski). Notons par ailleurs que si $\left\{ {n \atop k} \right\}$ désigne le nombre de partitions à k blocs d'un ensemble à n éléments (nombre de Stirling de seconde espèce) alors

$$B_n = \sum_{k=1}^{n} \left\{ {n \atop k} \right\}.$$

On dispose de la formule de récurrence

$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k \left\{ {n-1 \atop k} \right\} \quad \text{avec conditions au bord} \quad \left\{ {n \atop 1} \right\} = 1 \quad \text{et} \quad \left\{ {n \atop n} \right\} = 1$$

car pour choisir une partition de $\{1,\dots,n+1\}$ ayant k blocs il faut et il suffit soit de choisir une partition de $\{1,\dots,n\}$ ayant $k-1$ blocs et de la compléter avec le bloc singleton $\{n+1\}$, soit d'ajouter l'élément $n+1$

1. Pour justifier la convergence, on peut établir la majoration $B_n \leq n^n$ en remarquant que toute fonction de $\{1,\dots,n\}$ dans lui-même induit une partition de l'ensemble de départ : $\{1,\dots,n\} = \bigcup_i f^{-1}(\{i\})$; on en déduit une minoration du rayon de convergence par $1/e > 0$.

Probabilités - Préparation à l'agrégation interne de mathématiques (111/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

à l'un des k blocs d'une partition partition de $\{1,\dots,n\}$ ayant k blocs. Si X est une variable aléatoire de loi de Poisson de paramètre λ alors

$$\mathbb{E}\left[X^n\right] = \sum_{k=1}^{n} \left\{ {n \atop k} \right\} \lambda^k \quad \text{en particulier} \quad \mathbb{E}\left[X^n\right] = B_n \quad \text{si } \lambda = 1.$$

On dispose également de la formule explicite suivante :

$$\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} j^n$$

qui peut s'obtenir grâce au principe d'inclusion-exclusion en remarquant que le nombre de Stirling de seconde espèce est égal au nombre de surjections de $\{1,\dots,n\}$ dans $\{1,\dots,k\}$.

Intéressons-nous à la simulation de la loi uniforme sur l'ensemble Π_n des partitions de l'ensemble $\{1,\dots,n\}$, de cardinal B_n. Cette loi affecte le même poids $1/B_n$ à chaque élément de Π_n.

Théorème J.1 (Algorithme de Stam)**.** *Soit K un entier aléatoire valant k avec probabilité $k^n/(k!eB_n)$ pour tout $k \geq 0$. Sachant K, soient $C_1,\dots,C_n$ des variables aléatoires i.i.d. de loi uniforme sur $\{1,\dots,K\}$. Soit P la partition aléatoire de $\{1,\dots,n\}$ obtenue en décidant que i,j sont dans le même bloc si et seulement si $C_i = C_j$. Alors P suit la loi uniforme sur Π_n.*

La loi de K est bien définie grâce à la formule de Dobinski. Il est commode d'interpréter $C_1,\dots,C_n$ comme des couleurs, les blocs de P regroupant donc les éléments par couleur.

Démonstration. Il suffit d'observer que pour tout $p \in \Pi_n$, en notant b son nombre de blocs,

$$\mathbb{P}(P = p) = \sum_{k=b}^{\infty} \mathbb{P}(P = p | K = k)\mathbb{P}(K = k) = \sum_{k=b}^{\infty} \frac{k(k-1)\cdots(k-b+1)}{k^n} \frac{k^n}{k!eB_n} = \frac{1}{B_n}.$$

□

Probabilités - Préparation à l'agrégation interne de mathématiques (112/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Bibliographie

[BC07] B. Bercu et D. Chafaï – *Modélisation stochastique et simulation*, Dunod/SMAI, 2007.

[BL07] P. Barbe et M. Ledoux – *Probabilité*, EDP Sciences, 2007.

[CGCDM11] M. Cottrell, V. Genon-Catalot, C. Duhamel et T. Meyre – *Exercices de probabilités*, Cassini, 2011.

[DCD94] D. Dacunha-Castelle et M. Duflo – *Probabilités et statistique. problèmes à temps fixe*, Masson, 1994.

[DJ07] J.-F. Delmas et B. Jourdain – *Modèles aléatoires*, Mathematiques et Applications 57, Springer, 2007.

[Fel68] W. Feller – *An introduction to probability theory and its applications*, Wiley, 1968.

[FFF12] D. Foata, J. Franchi et A. Fuchs – *Calcul des probabilités – cours, exercices et problèmes corrigés*, Dunod, 2012.

[Tou99] P.-S. Toulouse – *Thèmes de probabilité et statistique*, Dunod, 1999.

Table des figures

2.1 Codage d'un multiensemble par une coloration. 11
2.2 Rendez-vous d'Athanase et Bérénice 21

3.1 Deux fonctions de répartition 25
3.2 Fléchettes : loi de l'abscisse 28
3.3 Diagrammes en bâtons et densités 29
3.4 Variance des lois normales 37
3.5 Approximation de la loi binomiale par la loi de Poisson. 41

4.1 Corrélation et dépendance (source : Wikipédia). 52

5.1 Quantile et fluctuation de la loi normale standard 63
5.2 Diagramme en bâtons de la binomiale 64

D.1 Ruine du joueur 82

Probabilités - Préparation à l'agrégation interne de mathématiques (113/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

D.2 Compter les ponts 84

E.1 Le processus de Yule 89

Liste des tableaux

1.1 Approche ensembliste des probabilités. On écrit *événement* ou *évènement*. 6

3.1 Quelques lois sur $\mathbb{N}$ ou $\mathbb{R}$. 43

Probabilités - Préparation à l'agrégation interne de mathématiques (114/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

Index

arrangement, 11

borélienne, 24

centré, 51
combinaison, 11
converge en loi, 40, 62, 99
corrélation, 51
couple, 50
covariance, 50

densité, 28
discrète, 23
distance en variation totale, 55
durée de vie moyenne restante, 91
déciles, 25
définie positive, 52

en moyenne, 99
en probabilité, 57, 99
épreuves, 5
équiprobabilité, 10
espace probabilisé, 9
espérance, 30, 31
événements, 9
événements élémentaires, 5

fonction caractéristique, 39
fonction de hasard, 91
fonction de masse, 26
fonction de répartition, 24
fonction de survie, 90
fonction de survie conditionnelle, 91
fonction génératrice, 39
formule du transfert, 33

incompatibles, 6
indépendants, 17, 18
intervalle de confiance, 63
intervalle de fluctuation, 38, 62, 79
intégrable, 31
inverse généralisé, 29
inégalité de déviation, 38

loi, 26
loi binomiale-négative, 73
loi de Poisson, 27
loi de probabilité, 9
loi hypergéométrique, 12, 27
loi hypergéométrique multitype, 45
loi multinomiale, 46
loi normale, 53
loi à densité, 28

marche aléatoire simple, 81
matrice de covariance, 51
max-stables, 95
mesure de probabilité, 9
moment, 32
moment factoriel, 40
médiane, 25

pile ou face
- espérance de la loi binomiale, 30
- espérance de la loi géométrique, 33
- Calcul des moments, 50
- fluctuations, 62
- intervalle de confiance, 63
- indépendance, 18
- quelques lois discrètes classiques, 26
- retour sur les moments, 54
- nombre de piles, 12
- singes dactylographes, 19
- point de vue statistique fréquentiste, 57
- suites d'événements et limites, 15
- tribus, 7
- univers, 5
- variable aléatoire « nombre de piles », 23
- variance, 37
- Lemme de Slutsky et intervalle de Wald, 100

presque sûrement, 59, 99
probabilité conditionnelle, 16
probabilité conditionnelle de A sachant B, 16
probabilité produit, 14

quantile d'ordre α, 25
quartiles, 25

racine carrée, 53
réduit, 51

Probabilités - Préparation à l'agrégation interne de mathématiques (115/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/

semi-définie positive, 52
standard, 53

tribu, 6
tribu borélienne, 7
tribu cylindrique, 7
tribu engendrée, 7
tribu grossière, 7
tribu terminale, 20

uniformément intégrable, 103
univers, 5

variable aléatoire, 23
variable aléatoire réelle, 23
variance, 36
vecteur aléatoire, 45
vecteur moyenne, 51

écart-type, 37

Probabilités - Préparation à l'agrégation interne de mathématiques (116/116) - Copyright © Djalil Chafaï et Pierre-André Zitt, 2014 - ISBN 978-2-9541710-2-9 - EAN 9782954171029 - http://djalil.chafai.net/