

Probabilités et processus stochastiques

Springer

Paris

Berlin

Heidelberg

New York

Hong Kong

Londres

Milan

Tokyo

Yves Caumel

Probabilités et processus stochastiques



Springer

Yves Caumel
ISAE-ENSICA
1, place Émile Blouin
31056 Toulouse Cedex

ISBN 978-2-8178-0162-9 Springer Paris Berlin Heidelberg New York

© Springer-Verlag France, 2011

Imprimé en France

Springer-Verlag France est membre du groupe Springer Science + Business Media

Cet ouvrage est soumis au copyright. Tous droits réservés, notamment la reproduction et la représentation, la traduction, la réimpression, l'exposé, la reproduction des illustrations et des tableaux, la transmission par voie d'enregistrement sonore ou visuel, la reproduction par microfilm ou tout autre moyen ainsi que la conservation des banques de données. La loi française sur le copyright du 9 septembre 1965 dans la version en vigueur n'autorise une reproduction intégrale ou partielle que dans certains cas, et en principe moyennant les paiements des droits. Toute représentation, reproduction, contrefaçon ou conservation dans une banque de données par quelque procédé que ce soit est sanctionnée par la loi pénale sur le copyright.

L'utilisation dans cet ouvrage de désignations, dénominations commerciales, marques de fabrique, etc., même sans spécification ne signifie pas que ces termes soient libres de la législation sur les marques de fabrique et la protection des marques et qu'ils puissent être utilisés par chacun.

La maison d'édition décline toute responsabilité quant à l'exactitude des indications de dosage et des modes d'emplois. Dans chaque cas il incombe à l'utilisateur de vérifier les informations données par comparaison à la littérature existante.

Maquette de couverture : Jean-François Montmarché
Détail du tableau : © Bloc Images



Collection
Statistique et probabilités appliquées
dirigée par Yadolah Dodge

Professeur Honoraire
Université de Neuchâtel
Suisse
yadolah.dodge@unine.ch

Comité éditorial :

Aurore Delaigle

Département de mathématiques
et de statistique
Université de Melbourne
Victoria 3010
Australie

Christian Genest

Département de mathématiques
et de statistique
Université McGill
Montréal H3A 2K6
Canada

Marc Hallin

Université libre de Bruxelles
Campus de la Plaine
CP 210
1050 Bruxelles
Belgique

Ludovic Lebart

Télécom-ParisTech
46, rue Barrault
75634 Paris Cedex 13
France

Christian Mazza

Département de mathématiques
Université de Fribourg
Chemin du Musée 23
CH-1700 Fribourg
Suisse

Stephan Morgenthaler

École Polytechnique Fédérale
de Lausanne
Département de Mathématiques
1015 Lausanne
Suisse

Louis-Paul Rivest

Département de mathématiques
et de statistique
Université Laval
Québec G1V 0A6
Canada

Gilbert Saporta

Conservatoire national
des arts et métiers
292, rue Saint-Martin
75141 Paris Cedex 3
France

Dans la même collection :

- *Statistique. La théorie et ses applications*
Michel Lejeune, avril 2004
- *Optimisation appliquée*
Yadolah Dodge, octobre 2004
- *Le choix bayésien. Principes et pratique*
Christian P. Robert, novembre 2005
- *Régression. Théorie et applications*
Pierre-André Cornillon, Éric Matzner-Løber, janvier 2007
- *Le raisonnement bayésien. Modélisation et inférence*
Éric Parent, Jacques Bernier, juillet 2007
- *Premiers pas en simulation*
Yadolah Dodge, Giuseppe Melfi, juin 2008
- *Génétique statistique*
Stephan Morgenthaler, juillet 2008
- *Maîtriser l'aléatoire. Exercices résolus de probabilités et statistique, 2^e édition*
Eva Cantoni, Philippe Huber, Elvezio Ronchetti, septembre 2009
- *Pratique du calcul bayésien*
Jean-Jacques Boreux, Éric Parent, décembre 2009
- *Statistique. La théorie et ses applications, 2^e édition*
Michel Lejeune, septembre 2010
- *Le logiciel R*
Pierre Lafaye de Micheaux, Rémy Drouilhet, Benoît Liquet, novembre 2010

Présentation

La théorie des probabilités résulte d'un long processus de modélisation des phénomènes aléatoires, inauguré au XVII^e siècle par l'étude des jeux de hasard, pour aboutir aujourd'hui à la théorisation de phénomènes aussi complexes que les processus de diffusion en physique ou l'évolution des marchés financiers. En 1931, le mathématicien soviétique Andreï Kolmogorov, prenant appui sur la toute récente théorie de la mesure et de l'intégration développée par Borel et Lebesgue, donna au corpus dénommé jusqu'alors Calcul des probabilités une structure axiomatique-déductive, propre aux mathématiques contemporaines, assurant la théorie des probabilités sur des bases solides.

Ce livre s'adresse aux étudiants en mathématiques de niveaux L2, L3 et M1, ainsi qu'aux élèves ingénieurs et plus généralement aux étudiants et aux chercheurs sollicités par des problèmes de modélisation de phénomènes et de systèmes aléatoires. Il devrait leur permettre de maîtriser les concepts et les méthodes probabilistes afin de les appliquer à des domaines aussi variés que le traitement du signal, l'automatique, la théorie des réseaux, la recherche opérationnelle, les sciences biologiques et économiques.

Les prérequis constitués des connaissances mathématiques dispensées durant les deux premières années d'un cursus universitaire du type mathématique, ou dans le cadre des classes préparatoires aux écoles d'ingénieurs et de commerce, pourront être complétés par les notions basiques d'analyse fonctionnelle : mesure et intégration, espaces de Hilbert.

J'ai rédigé ce cours dans le souci permanent d'éviter la pesante et souvent inefficace linéarité de l'exposé déductif, que les limites horaires d'un cours de mathématiques en école d'ingénieurs rendent impraticables. À de rares exceptions près, seules les démonstrations peu techniques et relativement courtes ont été données ; en revanche certaines démonstrations abordables et ayant un intérêt pédagogique sont proposées sous forme d'exercices.

L'ouvrage est composé de onze chapitres : les six premiers constituent un exposé de la théorie des probabilités ; les quatre suivants sont consacrés aux processus aléatoires classiques et à leurs applications, quant au dernier chapitre c'est un ensemble de problèmes suivis de leurs corrigés. Le premier chapitre concerne les probabilités définies sur des ensembles au plus dénombrables ; les deux chapitres suivants mettent en place les notions théoriques centrales s'articulant sur les concepts de variable et de vecteur aléatoire. Dans le chapitre 4

sont rassemblées les méthodes permettant d'effectuer des calculs de lois, dont la problématique générale consiste à déterminer la loi d'une variable ou d'un vecteur aléatoire fonction d'un vecteur aléatoire X de loi connue. Le chapitre 5 définit les différents modes de convergence de la théorie des probabilités et expose les théorèmes limites classiques qui s'y rattachent. Le chapitre 6 traite du conditionnement et de ses applications à la fiabilité et à la théorie de l'information. Les chapitres 7 et 8 sont consacrés à l'étude des chaînes de Markov discrètes ainsi qu'aux processus de Poisson et à leur extension aux processus de renouvellement. Le chapitre 9 est consacré aux chaînes de Markov à temps continu, à leurs applications aux processus de naissance et de mort ainsi qu'à la théorie des files d'attente. Dans le chapitre 10 on expose les bases de la théorie des processus du second ordre et du processus brownien.

Afin de montrer, s'il en était besoin, combien sont vivantes les théories mathématiques et les pensées qui y sont mises en œuvre, j'ai parsemé ce cours de notules historiques portant sur le lent développement de la théorie probabiliste, ainsi que de remarques de nature épistémologique concernant la notion de hasard. De brèves biographies tentent de donner une épaisseur humaine à ces créateurs souvent méconnus que sont les mathématiciens, si joliment baptisés par le philosophe Gaston Bachelard « prophètes de l'abstrait ».

Quiconque a un jour abordé la théorie des probabilités sait combien sa compréhension nécessite la pratique d'une intuition et d'un savoir-faire spécifiques : seul un apprentissage actif où les exercices joueront un rôle prépondérant, se révélera productif. Pour répondre à cette exigence, cent cinquante exercices et problèmes environ, dont la majorité sont corrigés, couvrent la totalité du cours et sont orientés, pour beaucoup d'entre eux, vers les applications. Leur niveau de difficulté est gradué de la façon suivante :

- ♠ : application directe du cours ;
- ♠♠ : nécessite une bonne compréhension du cours ;
- ♠♠♠ : résolution exigeant un savoir-faire certain.

J'adresse mes plus vifs remerciements à Claudie Chabriac pour son aide précieuse tout au long de la gestation de cet ouvrage ainsi qu'à Emmanuel Lochin pour son aide technique et son soutien moral. Je remercie chaleureusement mes collègues et particulièrement Ludovic d'Estampes et Florence Nicol pour leurs lectures critiques, ainsi que Rémi Diana et Thomas Ferrandiz pour leur aide lors de la retranscription du manuscrit. Que tous trouvent ici l'expression de mon amicale et chaleureuse gratitude.

Yves Caumel

Sommaire

Présentation	vii
1 Probabilités sur les ensembles finis	1
1.1 Notions de système aléatoire et d'espace de probabilité	1
1.2 Indépendance d'événements	6
1.3 Événements équiprobables dans le cas où Ω est fini	7
1.4 Corrigés des exercices	13
2 Variables aléatoires	19
2.1 Notion de variable aléatoire	19
2.2 Variables aléatoires discrètes	21
2.3 Variables aléatoires absolument continues	29
2.4 Variables gaussiennes	34
2.5 Corrigés des exercices	39
3 Vecteurs aléatoires	45
3.1 Un exemple	45
3.2 Description des vecteurs aléatoires	46
3.2.1 Généralités	46
3.2.2 Vecteurs discrets	49
3.2.3 Vecteurs absolument continus	52
3.3 Indépendance et corrélation	55
3.4 Fonctions génératrice et caractéristique	57
3.4.1 Fonction génératrice	57
3.4.2 Fonction caractéristique	58
3.5 Le vecteur gaussien	60
3.6 Corrigés des exercices	66
4 Calcul de lois	71
4.1 Approche du problème sur deux exemples	71
4.2 Méthodes classiques	73
4.2.1 Méthode du changement de variables	74
4.2.2 Méthode d'identification fonctionnelle	75

4.2.3	Variables aléatoires définies comme maximum ou minimum d'un ensemble de v.a.	78
4.2.4	Lois de variables mixtes	80
4.3	Méthodes de simulation des lois de probabilité	81
4.4	Quelques densités classiques	83
4.5	Corrigés des exercices	85
5	Convergences et limites des suites aléatoires	93
5.1	Convergence en probabilité	93
5.2	Convergence en loi	95
5.3	Théorèmes aux limites	99
5.4	Convergence presque sûre	104
5.5	Convergence en moyenne quadratique	105
5.6	Corrigés des exercices	109
6	Probabilités, lois et espérances conditionnelles	113
6.1	Conditionnement d'événements	113
6.2	Lois conditionnelles	117
6.2.1	Loi d'une variable discrète conditionnée par une variable discrète	117
6.2.2	Loi d'une variable absolument continue conditionnée par une variable discrète	119
6.2.3	Loi d'une variable discrète conditionnée par une variable absolument continue	120
6.2.4	Loi d'une variable continue conditionnée par une variable continue	120
6.2.5	Loi d'une variable continue conditionnée par un événement quelconque	122
6.3	Espérance conditionnelle	126
6.4	Variance conditionnelle	132
6.5	Loi et espérance conditionnelle des vecteurs gaussiens	134
6.6	Notions d'entropie et d'information	135
6.7	Retour sur la notion de hasard	141
6.8	Corrigés des exercices	143
7	Chaînes de Markov discrètes	149
7.1	Introduction aux processus aléatoires	149
7.2	Définition et caractérisation des chaînes de Markov discrètes	152
7.3	Classification des états	157
7.3.1	Relation de communication entre états	157
7.3.2	États récurrents et transitoires	158
7.3.3	États périodiques	164
7.4	Distributions stationnaires et distributions limites	165
7.5	Compléments	170

7.5.1	Estimation de la matrice de transition P et de la loi stationnaire π d'une chaîne finie irréductible	170
7.5.2	Probabilité d'absorption par les classes récurrentes dans le cas d'une chaîne finie	171
7.5.3	Temps moyens de séjour dans les états transitoires . . .	171
7.6	Corrigés des exercices	173
8	Processus de Poisson et de renouvellement	179
8.1	Processus de Poisson homogène	179
8.2	Processus de Poisson spatiaux	188
8.3	Processus de Poisson homogènes composés	190
8.4	Processus de renouvellement	191
8.4.1	Description et propriétés	191
8.4.2	Comportement asymptotique d'un processus de renouvellement	192
8.5	Corrigés des exercices	194
9	Chaînes de Markov à temps continu et files d'attente	203
9.1	Chaînes de Markov à temps continu	203
9.1.1	Définition et propriétés	203
9.1.2	Étude du régime stationnaire	210
9.2	Processus de naissance et de mort	212
9.2.1	Définition et propriétés	212
9.2.2	Étude du régime stationnaire des processus de naissance et de mort	213
9.3	Files d'attente	214
9.3.1	Définition et exemples	214
9.3.2	Étude du régime stationnaire des files d'attente	216
9.3.3	Réseaux de Jackson	222
9.4	Corrigés des exercices	229
10	Processus du second ordre	235
10.1	Généralités	235
10.2	Processus stationnaires du second ordre	237
10.3	Propriétés spectrales des processus SSL	238
10.4	Processus gaussiens stationnaires	240
10.5	Processus ergodiques	241
10.6	Processus de Wiener ou brownien et bruit blanc	242
10.7	Histoire de la modélisation du mouvement brownien	251
10.8	Corrigés des exercices	256
11	Problèmes	261
11.1	Énoncés des problèmes	261
11.2	Corrigés des problèmes	273

Annexes	293
Index	300

Chapitre 1

Probabilités sur les ensembles finis

Le hasard est-il une propriété intrinsèque de la réalité physique ou bien se manifeste-t-il comme une conséquence des limitations théoriques et expérimentales internes aux sciences qui le mettent en œuvre ? C'est cette deuxième hypothèse que défendait le probabiliste Émile Borel lorsqu'il affirmait au début du siècle dernier que « *le hasard n'est que le nom donné à notre ignorance et n'existerait pas pour un être omniscient* » ; mais, afin de modérer son propos, il ajoutait « *quelles que soient la nature et la rapidité des progrès scientifiques, il y aura toujours une place pour lui* ».

La théorie des probabilités ne nous apprend rien sur la nature profonde du hasard, ce n'est pas là son objectif : elle a été conçue pour être opératoire. Née de la curiosité des hommes pour les phénomènes aléatoires d'origine ludique ou naturelle, la théorie des probabilités est, depuis les années 1930, solidement adossée à la théorie de la mesure et de l'intégration, et permet de modéliser les phénomènes et les systèmes aléatoires à des fins pratiques de description et de prévision dans tous les domaines de l'activité humaine.

1.1 Notions de système aléatoire et d'espace de probabilité

À une action donnée, un **système aléatoire** donne une réponse imprévisible appelée **épreuve** ou **événement aléatoire élémentaire**. Cependant, dans le cadre de la théorie des probabilités, l'incertitude à laquelle est confronté l'observateur est tempérée par la connaissance de l'ensemble des événements aléatoires susceptibles d'être générés par le système.

Quelques exemples :

Système aléatoire	Action	Réponse aléatoire
Un dé	Un seul lancer	Face du dé obtenue
Une machine	Fonctionnement	Durée de vie
Une cible D , un projectile	Tir du projectile	Impact dans D
Jeu de 32 cartes	Tirer une carte	Une carte parmi les 32

Définition 1.1 *L'ensemble des événements élémentaires d'un système aléatoire est dénommé **espace fondamental**, ou **espace des observables**, et noté Ω . Les exemples ci-dessus illustrent la diversité de nature de l'espace fondamental, qui peut être un ensemble fini, infini dénombrable ou non dénombrable.*

À un événement quelconque est associée une partie de Ω constituée d'événements élémentaires pertinents pour l'observateur.

Exemple 1.1 Dans le cas de l'expérience associée au jet d'un dé équilibré, considérons par exemple les événements : « obtenir un chiffre pair » décrit par le sous-ensemble $\{2, 4, 6\}$ ou bien « obtenir un chiffre strictement supérieur à 4 » décrit par $\{5, 6\}$.

Exemple 1.2 S'il s'agit d'une expérience de tir sur une cible circulaire de rayon R , un événement consistera en l'atteinte d'un sous-ensemble donné de la cible. Par exemple, l'un des quadrants ou bien l'intérieur d'un cercle concentrique de rayon $r < R$.

La modélisation d'un système aléatoire nécessite que soit défini un ensemble \mathcal{T} constitué d'événements inclus dans l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω , suffisamment riche pour être stable par rapport aux opérations d'intersection, d'union et de complémentation qui lui confèrent une structure de tribu. Ces opérations ensemblistes nécessaires à l'expression des événements les plus divers sont respectivement équivalentes aux opérations logiques de disjonction, de conjonction et de négation.

Tableau de correspondance

Théorie des probabilités	Théorie des ensembles
négation de A (notée : \bar{A})	A^c = complémentaire de A
A et B	$A \cap B$ = intersection de A et B
A ou B	$A \cup B$ = union de A et B
A ou B avec A, B incompatibles	$A \cup B$ avec $A \cap B = \emptyset$ (notée $A \uplus B$ ou $A + B$)
événement certain	ensemble Ω
événement impossible	sous-ensemble vide \emptyset
réalisation de B entraîne celle de A	$B \subset A$

Définition 1.2 L'ensemble \mathcal{T} des événements définis sur l'espace fondamental Ω , est une **tribu**. Rappelons les axiomes définissant une tribu :

- (T1) l'ensemble vide \emptyset appartient à \mathcal{T} ;
- (T2) \mathcal{T} est stable par **union dénombrable** ;
- (T3) \mathcal{T} est stable par **complémentation**.

Si Ω est un ensemble fini ou infini dénombrable, la tribu \mathcal{T} est l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω ; si Ω est l'un des ensembles \mathbb{R} ou \mathbb{R}^k , ou une de leurs parties, la tribu \mathcal{T} est la tribu borélienne définie sur cet ensemble. On rappelle que la tribu borélienne, est la plus petite tribu que l'on peut définir sur \mathbb{R} contenant tous les sous-ensembles associés aux événements susceptibles d'être observés : ensembles finis ou infinis dénombrables, intervalles ouverts ou fermés, et leurs unions et intersections, etc.

Exemple 1.3 Supposons qu'une urne contienne 1 boule blanche et n boules noires ; l'épreuve consiste à tirer une boule après l'autre jusqu'à l'obtention de la boule blanche. Soit B l'événement « apparition d'une boule blanche », et N l'événement « apparition d'une boule noire ». L'espace fondamental associé à ce modèle aléatoire est : $\Omega = \{B, (N, B), \dots, \underbrace{(N, \dots, N, B)}_{n \text{ fois}}\}$.

À ce stade de la construction de la théorie des probabilités, nous disposons d'un ensemble d'événements élémentaires Ω , muni d'une tribu \mathcal{T} d'événements ; reste à définir une mesure positive sur (Ω, \mathcal{T}) , qui permettra d'évaluer la probabilité de réalisation d'un événement aléatoire quelconque appartenant à \mathcal{T} . Tout au long de ce cours, le lecteur pourra vérifier que la théorie de la mesure et de l'intégration à laquelle il pourra se référer constitue le cadre mathématique naturel de la théorie des probabilités.

Définition 1.3 On appelle **mesure de probabilité ou probabilité définie sur** (Ω, \mathcal{T}) , toute application $P : \mathcal{T} \rightarrow [0, 1]$ vérifiant les axiomes :

$$(P_1) \quad P(\Omega) = 1$$

(P₂) *Sigma-additivité* : pour tout sous-ensemble dénombrable d'événements disjoints $(A_i)_{i \in I}$ de \mathcal{T} , dont l'union est désignée par $\biguplus A_i$:

$$P(\biguplus A_i) = \sum_{i \in I} P(A_i).$$

Le triplet (Ω, \mathcal{T}, P) est un **espace de probabilité ou espace probabilisé**.

Les événements seront désignés par les premières lettres de l'alphabet éventuellement indexées : A, B, C, \dots

Voici rassemblées dans un seul théorème les propriétés de la mesure de probabilité, dont la plupart ont un fort contenu intuitif qui en facilite l'interprétation.

Théorème 1.1 (*Propriétés de la mesure de probabilité*)

(a) $\forall A \in \mathcal{F}, P(\bar{A}) = 1 - P(A)$ où \bar{A} est le complémentaire de A dans Ω .

(b) $\forall A, B \in \mathcal{F}, (A \subset B \Rightarrow P(A) \leq P(B))$.

(c) $\forall A, B \in \mathcal{F}, P(A) + P(B) = P(A \cap B) + P(A \cup B)$.

(d) *Convergence croissante monotone* : soit une suite croissante $(A_n)_n$ d'événements, alors :

$$P\left(\bigcup_n A_n\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

Résultat analogue pour la décroissance monotone : soit une suite décroissante $(A_n)_n$, alors :

$$P\left(\bigcap_n A_n\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

(e) *Pour toute famille finie ou dénombrable d'événements $(A_i)_i$:*

$$P\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} P(A_i).$$

(f) *Identité de Poincaré* : pour toute suite finie d'événements $(A_i)_{i=1, \dots, n}$:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) - \\ &\quad \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n). \end{aligned}$$

Preuve

(a) $A \uplus \bar{A} = \Omega \Rightarrow P(A) + P(\bar{A}) = 1$.

(b) $A \subset B \iff \exists B', B = A \uplus B' \Rightarrow P(B) = P(A) + P(B')$ donc :

$$P(B) \geq P(A).$$

(c) $A \cup B = A \uplus (B \setminus A)$ où $B \setminus A = B \cap \bar{A}$.

$$P(A \cup B) = P(A) + P(B \setminus A) = P(A) + P(B) - P(A \cap B).$$

(d) On applique (c) à la suite $(B_n)_n$ définie par $B_n = A_n - A_{n-1}$.

(e) Conséquence de (c).

(f) Par récurrence à partir de (c). □

Théorème 1.2 (des probabilités totales)

Si $(B_i)_{i \in I}$ est un système exhaustif ou complet d'événements, défini par $\biguplus_{i \in I} B_i = \Omega$, alors pour tout $A \in \mathcal{T}$,

$$P(A) = \sum_{i \in I} P(A \cap B_i).$$

Définition 1.4 Un événement de probabilité nulle est dit **négligeable**; une propriété vraie sauf sur un événement négligeable est qualifiée de **presque sûre**.

Intermède biographique : Andrei Kolmogorov (1903-1987)

Créateur de la théorie axiomatique des probabilités, Kolmogorov, de nationalité russe, fut l'un des grands mathématiciens du siècle dernier. En 1929, il proposa une axiomatisation de la théorie des probabilités fondée sur la théorie de la mesure dont l'exposé définitif *Théorie générale de la mesure et des probabilités* parut en 1933. Ses travaux s'orientèrent ensuite vers l'étude des processus stochastiques généraux et des processus stationnaires; on lui doit aussi de nombreux travaux en topologie, en mécanique classique et en théorie des automates finis.

Exemple 1.4 Quel nombre minimal de personnes faut-il dans une assemblée pour que deux d'entre elles au moins aient le même anniversaire, avec une probabilité $\frac{1}{2}$? (On suppose que les dates de naissance sont équiprobables sur les 365 jours et on ne tient pas compte des années bissextiles).

La probabilité p_n pour qu'il n'y en ait aucune parmi n personnes qui soit née le même jour que n'importe quelle autre, est :

$$p_n = \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \dots \left(1 - \frac{n-1}{365}\right).$$

D'où la probabilité $1 - p_n$ pour qu'il y en ait au moins deux qui soient nées le même jour. Si une assemblée comporte au moins vingt-trois personnes, la probabilité de réaliser cet événement est au moins égale à $\frac{1}{2}$. Si elle comporte quarante personnes, on a 90% de chances de réaliser cet événement.

1.1 ♠ Quel doit être le nombre minimal de personnes pour que la probabilité que deux d'entre elles au moins soient nées le même mois, soit égale à 0,9? (On considère que tous les mois ont la même longueur.)

1.2 ♠♠ Application de l'identité de Poincaré

Un préposé répartit au hasard n lettres, destinées à n personnes, à raison d'une lettre par boîte. Calculer la probabilité pour que chaque lettre arrive à son destinataire, ainsi que la probabilité pour qu'il n'y en ait aucune qui arrive à son destinataire; que vaut-elle quand n est très grand?

Intermède biographique : Émile Borel (1871-1956)

Né à Saint-Affrique en Aveyron, É. Borel fit des études de mathématiques à l'École normale supérieure où il devint professeur. Nommé peu après à la Sorbonne, à la chaire de la théorie des fonctions puis à celle de la théorie des probabilités, il quitta son poste pour entrer en politique sous l'impulsion de son ami Paul Painlevé; d'abord député dans son département d'origine en 1924, il devint deux ans après ministre de la Marine. Emprisonné par le régime de Vichy, auquel il s'opposa farouchement, il fonda dès sa libération un réseau de résistance dans sa région natale.

Borel développa les bases de la théorie de la mesure, sur lesquelles H. Lebesgue put édifier sa théorie de l'intégration. Il participa au long développement de la théorie des probabilités et de ses ramifications tout en se passionnant pour la pédagogie et la vulgarisation scientifique; en théorie des jeux, ses travaux annoncent ceux de J. von Neumann (1903-1957).

Approche empirique du concept de probabilité

(1) On répète n fois une épreuve consistant à tirer un des trois événements $\{A, B, C\}$, dont on ignore les probabilités; la fréquence empirique $\frac{n_A}{n}$, où n_A est le nombre d'apparitions de A pour n tirages, est une bonne approximation de la probabilité *a priori* $P(A)$, si n est suffisamment grand; même constat pour les fréquences empiriques $\frac{n_B}{n}$ et $\frac{n_C}{n}$. Les fréquences empiriques $\frac{n_A}{n}$, $\frac{n_B}{n}$, $\frac{n_C}{n}$ vérifient l'axiome : $\frac{n_A}{n} + \frac{n_B}{n} + \frac{n_C}{n} = 1$.

(2) Le jeu de « pile ou face » est un bon modèle pour illustrer l'approche empirique. Considérons les événements : $A = (\text{obtenir Pile en un seul jet})$, et $\bar{A} = (\text{obtenir Face en un seul jet})$. Si la pièce utilisée est parfaitement équilibrée du point de vue dynamique, $P(A) = P(\bar{A}) = \frac{1}{2}$; on constate sur un ensemble de n jets que $\frac{n_A}{n}$ et $\frac{n_{\bar{A}}}{n}$ convergent vers $\frac{1}{2}$, lorsque n s'accroît.

1.2 Indépendance d'événements

Deux événements sont indépendants si l'information apportée par la réalisation de l'un d'eux ne donne aucune information sur la réalisation de l'autre événement. Dans une première approche, on dira que deux événements sont indépendants si la probabilité de leur réalisation conjointe est égale au produit de leurs probabilités respectives.

Définition 1.5 Deux événements A, B d'un espace probabilisé (Ω, \mathcal{F}, P) sont **indépendants** si et seulement si :

$$P(A \cap B) = P(A).P(B).$$

Exemple 1.5 Soit E un jeu de trente-deux cartes ; A l'événement « cartes supérieures au 10 » ; B l'événement « la carte est un roi ». A et B sont dépendants car : $P(A \cap B) = \frac{4}{32}$ et $P(A)P(B) = \frac{16}{32} \cdot \frac{4}{32}$.

Théorème 1.3 Si A et B sont indépendants, alors les événements :

$$\begin{array}{l} \overline{A} \text{ et } \overline{B} \\ A \text{ et } \overline{B} \\ \overline{A} \text{ et } B \end{array} \text{ sont indépendants.}$$

On a un résultat analogue pour deux événements dépendants.

1.3 ♠ Démontrer le théorème.

Définition 1.6 Une famille d'événements $(A_i)_{i \in I}$ est **mutuellement indépendante** ou **indépendants** dans leur ensemble si pour tout sous-ensemble fini J inclus dans I :

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

Il est donc évident que des événements mutuellement indépendants sont indépendants deux à deux.

Dans la suite du cours, l'indépendance des événements sera supposée être mutuelle.

1.3 Événements équiprobables dans le cas où Ω est fini

L'analyse combinatoire, branche des mathématiques qui traite les problèmes de dénombrements sur des ensembles finis, est l'outil de calcul privilégié pour déterminer les probabilités de ce type.

Théorème 1.4 Une probabilité P sur un ensemble Ω fini ou dénombrable est complètement caractérisée par l'ensemble des valeurs $P(\omega_i)$ qu'elle prend sur les événements simples ω_i constitutifs de Ω . Pour tout événement A de \mathcal{T} :

$$P(A) = \sum_{\omega_j \in A} P(\omega_j).$$

Définition 1.7 Soit (Ω, \mathcal{T}, P) un espace de probabilité fini de taille n , dont les événements élémentaires sont équiprobables, c'est-à-dire si :

$$\forall i = 1, \dots, n \quad P(\omega_i) = \frac{1}{n}.$$

Si A est un événement de la tribu \mathcal{F} , sa probabilité est définie par :

$$P(A) = \frac{\text{nombre de réalisations } \omega_i \text{ favorables à } A}{\text{nombre } n \text{ de réalisations possibles}}.$$

Exemple 1.6 Soit Ω un jeu de trente-deux cartes, A l'événement « un carreau inférieur ou égal à 10 » ;

$$P(A) = \frac{4}{32} = \frac{1}{8}.$$

Définition 1.8 Le tirage d'un sous-ensemble de k éléments dans un ensemble qui en comporte n , distincts entre eux, s'effectue selon deux modalités :

(1) avec ou sans remise ;

(2) en tenant compte ou non de l'ordre du sous-ensemble.

Il existe trois types de tirages classiques :

- la **combinaison** est un sous-ensemble obtenu **sans remise et en ne tenant pas compte de l'ordre** ; une combinaison s'identifie donc à une partie ;

- la **combinaison avec répétition** est un sous-ensemble obtenu **avec remise et en ne tenant pas compte de l'ordre** de l'échantillon ;

- l' **arrangement** est un sous-ensemble obtenu **sans remise et en tenant compte de l'ordre** de l'échantillon.

Définition 1.9 Étant donné un ensemble de n éléments, on définit :

(1) Le **nombre** A_n^k **d'arrangements** de k éléments :

$$A_n^k = \frac{n!}{(n-k)!}.$$

(2) Le **nombre** C_n^k , noté encore $\binom{n}{k}$, **de combinaisons** de k éléments parmi n :

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n.(n-1) \dots (n-k+1)}{k!}.$$

Sachant qu'il y a $k!$ arrangements distincts d'une partie de k éléments, et dans ce cas on parle aussi de permutations de k éléments, on en déduit :

$$A_n^k = k! C_n^k.$$

(3) Le **nombre** K_n^k **de combinaisons à répétitions** de k éléments :

$$K_n^k = \binom{n+k-1}{k}.$$

1.4 ♠♠ (1) Démontrer que le nombre de partitions en m parties A_1, \dots, A_m d'un ensemble de n éléments, telles que pour tout $i=1, \dots, m$ $|A_i| = k_i$, est égal à $\frac{n!}{k_1! \dots k_m!}$, dit **coefficient multinomial** et noté :

$$\binom{n}{k_1, \dots, k_m}.$$

(2) Établir les **propriétés importantes** suivantes des coefficients binomiaux grâce à des raisonnements de forme ensembliste :

$$\binom{n}{k} = \binom{n}{n-k}; \quad \sum_{k=0}^n \binom{n}{k} = 2^n; \quad \binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$

D'une grande utilité pratique, la formule de Stirling doit être mémoriser.

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \left(1 + O\left(\frac{1}{n}\right)\right) \text{ quand } n \rightarrow +\infty.$$

Intermède historique : l'origine de la notion de probabilité

Les jeux de hasard ont été pratiqués dans les plus anciennes civilisations d'Asie et d'Orient. Ainsi jouait-on à l'astragale qui est l'ancêtre du jeu de dé, en Irak et en Égypte, 3 000 ans avant J-C ; l'astragale est le nom d'un os en forme de cube, et *alzhar* signifie *dé* en arabe. En Grèce puis à Rome, les jeux de dés étaient largement répandus et connus sous le nom d'*aleae*, auquel on doit le mot *aléatoire*. Ce n'est pourtant qu'à partir du XV^e et du XVI^e siècle, que l'on s'intéressa à la régularité de certaines séquences apparaissant dans les jeux de hasard ; les premiers dénombrements et calculs de probabilité d'occurrences de telles séquences sont dus aux savants mathématiciens de la Renaissance italienne. L'un des plus célèbres d'entre eux, Gerolamo Cardano (1501-1576), à la fois médecin, astrologue et mathématicien, configuration intellectuelle fréquente à cette époque, est surtout connu pour avoir donné la solution générale des équations du troisième degré, et conçu le mécanisme qui porte en français le nom de *cardan*. Il écrivit en 1526 un traité intitulé *Liber de ludo aleae* (Livre des jeux de hasard), dans lequel il faisait usage de règles empiriques qui s'apparentent à la propriété de sigma-additivité des mesures.

De nombreux obstacles parsemèrent le chemin de la modélisation des phénomènes aléatoires, parmi lesquels nous mentionnerons les plus importants :

- la longue descendance de l'idéalisme platonicien selon lequel les phénomènes terrestres, contrairement aux phénomènes célestes, ne peuvent être régis par des lois d'ordre et d'harmonie ; conception de l'univers qui fut battue en brèche par Galilée, dont les observations permirent de conclure à l'homogénéité des mondes terrestre et céleste ;

- la croyance nourrie par les religions monothéistes que tout est réglé par la providence divine ;

- l'incompatibilité avec l'hypothèse du hasard du principe de causalité, qui dominait toute la physique classique ;
- la difficulté de développer des méthodes combinatoires.

1.5 ♠ À propos du jeu de bridge

(1) Combien de mains différentes un joueur peut-il recevoir au jeu de bridge (sans tenir compte de l'ordre, bien sûr) ? (Le jeu de bridge se joue avec cinquante-deux cartes, quatre joueurs, chacun des quatre joueurs recevant treize cartes)

(2) Probabilité qu'un seul des joueurs ait les treize cœurs ? Probabilité pour que chacun des quatre joueurs ait treize cartes d'une même couleur ?

1.6 ♠ Dénombrer les configurations possibles de n objets identiques dans r boîtes ordonnées. Le calcul de ce type de dénombrements a été essentiel au développement de la physique statistique.

1.7 ♠ Problème du double-six

À l'époque de Pascal, riche des premières tentatives de modélisation du hasard, le chevalier de Méré, personnage bien en cour et ami du philosophe, curieux des sciences et joueur impénitent, demanda à ce dernier, lequel des deux événements A_1 ou A_2 , était le plus fréquent :

$A_1 = \{\text{obtenir au moins un six en 4 lancers d'un dé}\} ;$

$A_2 = \{\text{obtenir au moins un double-six en 24 lancers de deux dés}\}.$

Qu'en pensez-vous ?

1.8 ♠♠ On considère une population de taille n dont on extrait un échantillon de taille p , puis après remise, un deuxième échantillon de taille q . Supposant $p + q < n$, quelle est la probabilité que les deux échantillons n'aient aucun élément commun ? Probabilité pour qu'il n'y ait qu'un élément commun ?

Application numérique : $n = 100$ et $p = q = 10$.

1.9 ♠ Un groupe de $2n$ personnes s'assoit autour d'une table :

- (1) De combien de façons peuvent-elles le faire ?
- (2) S'il y a autant de femmes que d'hommes, quelle est la probabilité de respecter l'alternance homme, femme, homme... autour de la table ?

1.10 ♠ Problème concernant le jeu de *passé-dix*, posé par le duc de Toscane à Galilée

Pourquoi un total de 10, obtenu en lançant trois dés, est-il plus probable qu'un total de 9, alors que ces deux nombres se partagent en autant de façons additives à l'aide des triplets de nombres compris entre 1 et 6.

1.11 ♠♠ Quelle est la probabilité pour que les nombres obtenus à la suite de n jets d'un dé soient dans un ordre non strictement croissant ?

Intermède : systèmes générateurs de hasard

Jusqu'au milieu du XIX^e siècle, le principe déterministe était avec le principe de causalité l'un des paradigmes dominants des sciences de la nature. La célèbre formulation qu'en donna Laplace peut se résumer ainsi : *la connaissance des conditions initiales d'un système physique et des lois qui en régissent son évolution doit permettre d'en prédire l'évolution future ou d'en restituer l'évolution passée*. Le mathématicien Henri Poincaré (1854-1912) s'attaquant au très difficile « problème à n corps », qui consiste à résoudre les équations régissant la dynamique de n planètes liées deux à deux par les forces d'attraction gravitationnelle, allait porter un coup d'arrêt à l'ambitieux programme laplacien. En effet, il mit en évidence des solutions très sensibles aux conditions initiales évoluant vers des trajectoires imprévisibles aujourd'hui qualifiées de chaotiques. Un tel système non linéaire est donc déterministe par la nature même des équations qui régissent son évolution et très sensible aux valeurs des conditions initiales, comme l'est par exemple le pendule excité pour certaines valeurs du rapport (*fréquence propre/fréquence d'excitation*). Le dé est un exemple simple de cette classe de systèmes, puisqu'une infime différence entre les valeurs des paramètres de position et de vitesse initiales du dé correspondant à plusieurs lancers suffit pour que soient générées des faces différentes. C'est ce caractère chaotique qui fait du dé le générateur aléatoire le plus anciennement pratiqué. De nos jours, il revient aux théories des systèmes non linéaires et du chaos dynamique de modéliser et d'étudier de tels phénomènes, présents dans de nombreuses sciences (physique, biologie...). Le lecteur intéressé par la dynamique du dé et des intrications subtiles des six bassins attracteurs correspondant aux six faces du dé devrait trouver son bonheur dans le livre de Strzalko J. et al. *Dynamics of gambling : origins of randomness in mechanical systems*, Springer, 2005.

L'apparence aléatoire peut être trompeuse !

En effet, considérons sur un cercle de diamètre unitaire la suite des points $(M_n)_n$ d'abscisses curvilignes $a.n$, a étant un nombre transcendant (par exemple : $a = \frac{\pi}{10}$, $n \in \mathbb{N}$). Un observateur ignorant le système de génération des points et n'ayant accès qu'à un petit arc de cercle délimité par une petite fenêtre, de largeur égale à $\frac{1}{1.000}$ de la circonférence, verra les points se répartir au hasard sur l'arc de cercle observé. L'illusion du hasard sera tenace et il lui faudra un peu de travail pour réduire l'apparence aléatoire à une explication déterministe du phénomène. Au fait, pourquoi avoir choisi un nombre transcendant ?

1.12 ♠♠ Propriété de raréfaction des nombres premiers

On tire de façon équiprobable un nombre dans $\{1, \dots, n\}$.

Soient l'événement $A_p = (\text{le nombre tiré est divisible par } p)$

et $\phi(n)$ = le nombre d'entiers inférieurs à n et premiers avec n (on rappelle que k et n sont premiers entre eux si $\text{PGCD}(k,n)=1$).

- (1) Si p divise n , déterminer $P(A_p)$.
 (2) Prouver que $\phi(n) = n \prod_{i=1}^m (1 - \frac{1}{p_i})$ où $\{p_1, \dots, p_m\}$ est l'ensemble des premiers $\leq n$ et qui divisent n .

À la fin du siècle dernier, les mathématiciens français et suisse Hadamard et de La Vallée-Poussin, démontrèrent indépendamment l'un de l'autre le très beau résultat déjà pressenti par Euler, connu sous le nom d'approximation du logarithme intégral, à l'origine de l'important théorème de raréfaction croissante des nombres premiers :

$$\text{si } n \text{ est grand, } \phi(n) \simeq \int_2^n \frac{1}{\log_n(t)} dt \simeq \frac{n}{\log(n)}.$$

Proximité des fréquences calculées et des fréquences observées :

Nombre n d'entiers considérés	10^2	10^3	10^4	10^5	10^7
$\phi(n)$	25	168	1 229	9 592	664 579
Nombre $p(n)$ des premiers $\leq n$	21	159	1 213	9 555	664 241
Erreur relative $\frac{\phi(n)-p(n)}{p(n)}$	16%	5,3%	1,3%	0,3%	0,05%

Grâce à une démonstration analogue à celle de l'exercice précédent, on démontre un très beau résultat de la théorie des nombres premiers : la probabilité $P(\text{deux entiers tirés au hasard dans } 1, 2, \dots, n \text{ soient premiers})$ converge vers $\frac{6}{\pi^2}$ quand n devient très grand.

Les problèmes 1 et 2 se rapportent à ce chapitre.

1.4 Corrigés des exercices

Exercice 1.1

Soit $\phi(n)$ la probabilité qu'aucun couple de personnes ne soit né le même mois dans un groupe de n personnes. On a, pour $n \leq 12$,

$$\phi(n) = \left(1 - \frac{1}{12}\right) \left(1 - \frac{2}{12}\right) \cdots \left(1 - \frac{n-1}{12}\right).$$

L'entier n tel que $\phi(n) \leq 0,1$ soit le plus proche de 0,1 est $n = 8$.

Exercice 1.2

On numérote les lettres et les boîtes des destinataires : à chaque lettre, on associe la boîte choisie par le facteur.

(1) Il y a $n!$ distributions possibles pour les lettres et une seule correspond au cas où chaque lettre est à sa place, d'où la probabilité $\frac{1}{n!}$ que chaque lettre atteigne son destinataire.

(2) Soit A_i l'événement « la lettre i est à sa place ». On cherche $P(E)$ où

$$E = \overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_n} = \overline{A_1 \cup A_2 \cup \cdots \cup A_n},$$

donc $P(E) = 1 - P(A_1 \cup A_2 \cup \cdots \cup A_n)$; on détermine $P(A_1 \cup A_2 \cup \cdots \cup A_n)$ à l'aide de la formule de Poincaré, que l'on établit par récurrence sur n :

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \left(\sum_{1 \leq i_1, \dots, i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \right).$$

$P(A_{i_1} \cap \cdots \cap A_{i_k}) = \frac{(n-k)!}{n!}$, car pour $A_{i_1} \cap \cdots \cap A_{i_k}$, les lettres i_1, \dots, i_k sont nécessairement dans leur boîte et les $n-k$ lettres restantes sont réparties au hasard parmi les $n-k$ boîtes restantes, soit $(n-k)!$ choix possibles.

Or il y a $C_n^k = \frac{n!}{k!(n-k)!}$ façons de choisir les $\{i_1, \dots, i_k\}$ (k éléments parmi n) et donc

$$\sum_{1 \leq i_1, \dots, i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}) = \frac{1}{k!}.$$

Finalement,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \frac{1}{k!} \text{ et } P(E) = 1 - P\left(\bigcup_{i=1}^n A_i\right) = 1 - \sum_{k=1}^n (-1)^{k+1} \frac{1}{k!}.$$

La probabilité pour qu'aucune lettre n'atteigne son destinataire converge vers $\frac{1}{e}$ quand n tend vers l'infini.

Exercice 1.3

$$\begin{aligned}
P(\bar{A})P(\bar{B}) &= (1 - P(A))(1 - P(B)) = 1 - P(A) - P(B) + P(A)P(B) \\
&= 1 - P(A) - P(B) + P(A \cap B) = 1 - [P(A) + P(B) - P(A \cap B)] \\
&= 1 - P(A \cup B) = P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B}).
\end{aligned}$$

Démonstration analogue pour les deux autres formules.

Exercice 1.4

(1) Nombre de partitions : (nombre de choix de k_1 éléments parmi n) \times (nombre de choix de k_2 éléments parmi $n - k_1$ éléments) $\times \dots$, soit :

$$\frac{n!}{k_1!(n - k_1)!} \times \frac{(n - k_1)!}{k_2!(n - k_1 - k_2)!} \times \dots \times \frac{(n - k_1 - \dots - k_{m_1})!}{k_m!0!} = \frac{n!}{k_1!k_2! \dots k_m!}.$$

(2) Il revient au même de choisir k éléments que d'en éliminer $(n - k)$, d'où l'égalité attendue.

$$\sum_{k=0}^n \binom{n}{k} \text{ est le nombre total de parties de l'ensemble, soit } 2^n.$$

Pour établir la dernière propriété, on considère un élément particulier ; pour dénombrer les parties à k éléments, on considère deux cas : soit cet élément en fait partie et il faut alors choisir $k - 1$ éléments parmi $n - 1$, soit il n'en fait pas partie et il faut alors choisir k éléments parmi $n - 1$.

Exercice 1.5

$$(1) \binom{52}{13}.$$

$$(2) p_1 = \frac{4 \times \binom{39}{13} \times \binom{26}{13}}{\binom{52}{13} \times \binom{39}{13} \times \binom{26}{13}} = \frac{4}{\binom{52}{13}}; p_2 = \frac{4!}{\binom{52}{13} \times \binom{39}{13} \times \binom{26}{13}}.$$

Exercice 1.6

On a n objets « o » et r boîtes que l'on peut délimiter par $r - 1$ cloisons « c », ce qui fait en tout $n + r - 1$ entités (cloisons ou objets). Une configuration est parfaitement déterminée par un $(n + r - 1)$ -uplet constitué de $r - 1$ « c » et de n « o ». Il y a autant de tels $(n + r - 1)$ -uplets que de façons de placer les $r - 1$ « c », soit $\binom{n+r-1}{r-1}$.

Exercice 1.7

Expérience 1 : lancer d'un dé quatre fois de suite.

$$\Omega_1 = \{(x_1, x_2, x_3, x_4) ; 1 \leq x_i \leq 6\}$$

A_1 : « obtenir au moins un 6 ». Tous les résultats ont même chance de se réaliser donc

$$P(A_1) = \frac{\text{card}(A_1)}{\text{card}(\Omega_1)}.$$

Il est plus pratique de déterminer $P(\overline{A_1})$ où $\overline{A_1}$ est l'événement « pas de 6 ».

$$P(\overline{A_1}) = 1 - P(A_1) \text{ avec } \overline{A_1} = \{(x_1, x_2, x_3, x_4) ; 1 \leq x_i \leq 5\}.$$

$$P(\overline{A_1}) = \frac{\text{card}(\overline{A_1})}{\text{card}(\Omega)} \text{ avec } \text{card}(\overline{A_1}) = 5^4 \text{ et } \text{card}(\Omega) = 6^4 \text{ d'où } P(\overline{A_1}) = \left(\frac{5}{6}\right)^4 \text{ et}$$

$$P(A_1) = 1 - \left(\frac{5}{6}\right)^4 \approx 0,518.$$

Expérience 2 : lancer de deux dés vingt-quatre fois de suite.

$$\Omega_2 = \{((x_1, y_1), \dots, (x_{24}, y_{24})) ; 1 \leq x_i, y_i \leq 6\},$$

A_2 : « faire au moins un double 6 ». Tous les résultats ont même chance de se réaliser donc

$$P(A_2) = \frac{\text{card}(A_2)}{\text{card}(\Omega_2)}.$$

Il est plus pratique de déterminer $P(\overline{A_2})$ où $\overline{A_2}$ est l'événement « pas de double 6 ».

$$P(\overline{A_2}) = 1 - P(A_2) \text{ avec } \overline{A_2} = \{((x_1, y_1), \dots, (x_{24}, y_{24})) ; (x_i, y_i) \neq (6, 6)\}.$$

On a trente-cinq choix pour chaque (x_i, y_i) donc $\text{card}(\overline{A_2}) = (35)^{24}$.

$$\text{Finalement } P(\overline{A_2}) = \frac{\text{card}(\overline{A_2})}{\text{card}(\Omega_2)} = \left(\frac{35}{36}\right)^{24} \text{ et}$$

$$P(A_2) = 1 - P(\overline{A_2}) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0,491.$$

Conclusion : il est plus probable de faire au moins un six en quatre lancers qu'un double six en vingt-quatre lancers de deux dés.

Exercice 1.8

Un résultat est ici un couple $(\{x_1, \dots, x_p\}, \{y_1, \dots, y_q\})$, $\{x_1, \dots, x_p\}$ et $\{y_1, \dots, y_q\}$ étant deux échantillons de tailles respectives p et q de la population de taille n . Pour le premier échantillon, il y a C_n^p choix possibles, et pour le deuxième, il y en a C_n^q , car on peut réutiliser les éléments du premier. Ainsi,

$$\text{card}(\Omega) = C_n^p \times C_n^q.$$

(1) Soit A_1 = « les 2 échantillons n'ont pas d'éléments en commun ».

Pour le premier échantillon, on a toujours C_n^p choix possibles ; pour le deuxième, il faut choisir q éléments parmi $n - p$ (pour ne pas reprendre les éléments du premier) donc il reste C_{n-p}^q choix (ce qui exige $p + q \leq n$). Ainsi

$$\text{card}(A_1) = C_n^p \times C_{n-p}^q \text{ et } P(A_1) = \frac{\text{card}(A_1)}{\text{card}(\Omega)} = \frac{C_{n-p}^q}{C_n^q},$$

$$\text{d'où } P(A_1) = \frac{(n-p)!(n-q)!}{n!(n-p-q)!}.$$

(2) $A_2 = \ll \text{les deux échantillons ont 1 élément en commun} \gg$.

On choisit l'élément commun aux deux échantillons : n choix possibles.
On choisit les $p-1$ autres éléments du premier échantillon parmi $n-1$: C_{n-1}^{p-1} choix possibles.

On choisit enfin les $q-1$ autres éléments du deuxième échantillon parmi les $n-p$ qui ne sont pas dans le premier : C_{n-p}^{q-1} choix possibles.

$$\begin{aligned} \text{Finalement, } \text{card}(A_2) &= n \times C_{n-1}^{p-1} \times C_{n-p}^{q-1} \\ &= \frac{n(n-1)!}{(p-1)!(n-p)!} \frac{(n-p)!}{(q-1)!(n-p-q+1)!} \end{aligned}$$

et comme $\text{card}(\Omega) = \frac{n!}{p!(n-p)!} \times \frac{n!}{q!(n-q)!}$, on a :

$$P(A_2) = \frac{n!}{(p-1)!(q-1)!(n-p-q+1)!} \times \frac{p!(n-p)!q!(n-q)!}{n!n!}$$

d'où, en simplifiant, $P(A_2) = \frac{pq(n-p)!(n-q)!}{n!(n-p-q+1)!}$.

Exercice 1.9

On obtient la même configuration par rotation (chaque personne a même voisin de droite et même voisin de gauche). Ainsi, s'il y a $2n$ places, il y a $2n$ configurations identiques.

(1) Première méthode : on numérote les individus (par exemple, selon leur ordre d'arrivée). À chaque individu, on associe une chaise et une seule (celle qu'il choisit) : il y a alors $(2n)!$ possibilités mais chaque configuration est obtenue $2n$ fois.

Nombre de configurations différentes : $\frac{(2n)!}{2n} = (2n-1)!$.

Deuxième méthode : pour éviter les configurations identiques, on fixe la première personne sur la chaise qu'il a choisie et pour les autres on a $(2n-1)!$ possibilités.

(2) Cas où l'on respecte l'alternance.

Première méthode : le premier a $2n$ choix, puis les places des hommes et des femmes sont imposées : n choix possibles pour ceux de l'autre sexe, d'où $n!$ possibilités et $n-1$ possibles pour ceux du même sexe, d'où $(n-1)!$ possibilités. Finalement, $2n \times (n!) \times (n-1)!$ possibilités, dont $2n$ qui donnent la même configuration (par rotation).

Nombre de configurations respectant l'alternance : $n!(n-1)!$

Deuxième méthode : une fois le premier placé, il reste $n!$ possibilités pour les personnes de l'autre sexe et $(n-1)!$ pour celles du même sexe. D'où $n!(n-1)!$ configurations directement.

Exercice 1.10

L'expérience est ici un lancer de trois dés. On a donc

$$\Omega = \{(x_1, x_2, x_3) ; 1 \leq x_i \leq 6\} \text{ et } \text{card}(\Omega) = 6^3 = 216.$$

On considère les deux événements A (respectivement B) « faire un total de 9 » (respectivement de 10).

Chacun des triplets ayant la même chance de se réaliser, on a $P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$,
(respectivement $P(B) = \frac{\text{card}(B)}{\text{card}(\Omega)}$).

Nombre d'éléments de A :

- avec 1, 2, 6, on a 6 triplets ;
- avec 1, 3, 5, on a 6 triplets ;
- avec 1, 4, 4, on a 3 triplets ;
- avec 2, 2, 5, on a 3 triplets ;
- avec 2, 3, 4, on a 6 triplets ;
- avec 3, 3, 3, on a 1 triplet.

Ainsi, on a vingt-cinq triplets, obtenus à partir de six suites croissantes de 3 entiers de $\llbracket 1, 6 \rrbracket$.

Nombre d'éléments de B :

- avec 1, 3, 6, on a 6 triplets ;
- avec 1, 4, 5, on a 6 triplets ;
- avec 2, 2, 6, on a 3 triplets ;
- avec 2, 3, 5, on a 6 triplets ;
- avec 2, 4, 4, on a 3 triplets ;
- avec 3, 3, 4, on a 3 triplets.

Ainsi, on a vingt-sept triplets, obtenus à partir de six suites croissantes de trois entiers de $\llbracket 1, 6 \rrbracket$.

$$P(A) = \frac{25}{216} \text{ et } P(B) = \frac{27}{216}.$$

Exercice 1.11

On a donc n_1 fois la face 1, n_2 fois la face 2, ..., n_6 fois la face 6. Il faut donc trouver le nombre de 6-uples d'entiers (n_1, \dots, n_6) tels que $n_1 + \dots + n_6 = n$.

Il est plus simple d'aligner les n objets, puis de placer les cloisons des tiroirs : avec k tiroirs, on a $k - 1$ cloisons à placer. (Voir l'exercice 1.6.)

On a ainsi $\binom{n+k-1}{k-1} = \binom{n+k-1}{n}$ configurations différentes. Dans le cas du lancer de n dés, le nombre de résultats donnant une suite croissante est donc $\binom{n+5}{n}$ et la probabilité cherchée est $\frac{\binom{n+5}{n}}{6^n}$.

Exercice 1.12

(1) p divise n donc $P(A_p) = \frac{1}{p}$.

(2) Pour tout i , la probabilité que n ne soit pas divisible par p_i est $\left(1 - \frac{1}{p_i}\right)$ et la probabilité pour que n ne soit pas divisible par p_1, p_2, \dots, p_m est $\prod_{i=1}^m \left(1 - \frac{1}{p_i}\right)$.

Finalement :

$$\frac{\phi(n)}{n} = \prod_{i=1}^m \left(1 - \frac{1}{p_i}\right).$$

Chapitre 2

Variables aléatoires

Notion centrale de la théorie des probabilités, les variables aléatoires permettent de modéliser les phénomènes et les systèmes aléatoires.

2.1 Notion de variable aléatoire

Remarquons tout d'abord qu'en dépit de sa dénomination, une variable aléatoire n'est pas une variable au sens habituel. C'est une application d'un espace probabilisé (Ω, \mathcal{T}, P) dans un espace mesurable (E, \mathcal{E}) , où \mathcal{E} désigne une tribu définie sur E . Une variable aléatoire, que l'on désignera toujours par la majuscule d'une des dernières lettres de l'alphabet, permet de munir (E, \mathcal{E}) d'une probabilité appelée probabilité image. La seule condition exigée pour réaliser cet objectif est de supposer que la variable aléatoire soit mesurable.

Définition 2.1 *Toute application mesurable X d'un espace probabilisé (Ω, \mathcal{T}, P) dans un espace (E, \mathcal{E}) définit une variable aléatoire. X vérifie donc la propriété de mesurabilité :*

$$\forall B \in \mathcal{E}, \quad X^{-1}(B) \in \mathcal{T}.$$

On construit sur (E, \mathcal{E}) une probabilité dite probabilité image de la probabilité P par la variable aléatoire X , grâce au théorème :

Théorème 2.1 *La mesure P_X définie sur (E, \mathcal{E}) par*

$$\forall B \in \mathcal{E}, \quad P_X(B) = P(X^{-1}(B)),$$

est la probabilité image de P par X définie sur (E, \mathcal{E}) .

Preuve Vérifions les deux axiomes définissant une probabilité :

$$(a) \quad P_X(E) = P(X^{-1}(E)) = P(\Omega) = 1 ;$$

$$(b) P_X(\biguplus B_i) = P(X^{-1}(\biguplus B_i)) = \sum_i P(X^{-1}(B_i)) = \sum_i P_X(B_i) \text{ } (\Sigma\text{-additivité}) \quad \square$$

Si l'espace E est égal à \mathbb{R} , on le munit muni de sa tribu borélienne et les v.a. définies sur E sont dites réelles.

Exemple 2.1 Construction de la variable aléatoire de Bernoulli

Soient :

- l'espace des observables : $\Omega = \{\omega_1, \omega_2\}$;
- la tribu définie sur Ω : $\mathcal{T} = P(\Omega)$;
- la probabilité P définie sur Ω : $P(\omega_1) = p$, $P(\omega_2) = 1 - p$ où $p \in]0, 1[$.

La variable aléatoire de Bernoulli X de paramètre p est définie par :

$$X(\omega_1) = 1, X(\omega_2) = 0.$$

$$P_X(1) = P(X^{-1}(1)) = P(\omega_1) = p \text{ et } P_X(0) = P(X^{-1}(0)) = P(\omega_2) = 1 - p.$$

Cette v.a. qui permet de modéliser les tirages à deux issues possibles, illustrés par le modèle « pile ou face », est à la base de la construction des v.a. classiques telles que les v.a. binomiale ou géométrique.

Noter que : $\forall x \in \{0, 1\} : P_X(x) = p^x(1 - p)^{1-x}$.

On utilisera désormais l'abréviation v.a. à la place de variable aléatoire.

Étant donné les espaces de probabilité $(\Omega_i, \mathcal{T}_i, P_i)_{i=1, \dots, n}$, il s'agit de construire un espace de probabilité sur l'espace produit $\Omega = \Omega_1 \times \Omega_2 \dots \Omega_n$. On définit sur Ω la tribu \mathcal{T} égale à la tribu engendrée par les pavés rectangulaires $B_1 \times B_2 \times \dots \times B_n$ où B_i parcourt la tribu \mathcal{T}_i : on rappelle que la tribu engendrée par un sous-ensemble $\Omega' \subset \Omega$ est définie par l'intersection de toutes les tribus définies sur Ω contenant Ω' . D'autre part, la tribu engendrée par les ouverts de \mathbb{R} est la tribu borélienne.

Théorème 2.2 *Sur (Ω, \mathcal{T}) il existe une probabilité unique P , dite **probabilité produit** définie par :*

$$\forall (B_1, B_2, \dots, B_n) \in \mathcal{T}_1 \times \mathcal{T}_2 \dots \times \mathcal{T}_n :$$

$$P(B_1 \times B_2 \times \dots \times B_n) = P_1(B_1) \dots P_n(B_n).$$

Construction de la v.a. binomiale

Un groupe de n étudiants est censé suivre un cours ; tout étudiant a la probabilité p d'être présent au cours indépendamment du comportement des autres étudiants. À chaque étudiant i on associe l'ensemble Ω_i décrivant

ses états possibles {absence, présence}, muni de la tribu $\mathcal{T}_i = \mathcal{P}(\Omega_i)$ et de la probabilité P , définie par : $P_i(\text{présence}) = p$ et $P_i(\text{absence}) = 1 - p$.

L'espace produit $\Omega = \prod_{i=1}^n \Omega_i$ décrit les états possibles de l'ensemble des étudiants. On le munit de la probabilité produit P définie par le théorème précédent.

Considérons la v.a. S définie par :

$$S : \omega = (\omega_1, \dots, \omega_n) \in \Omega \longrightarrow S(\omega) = \sum_i \omega_i \in \{0, 1, \dots, n\};$$

$S(\omega)$ représente le nombre de présents.

La probabilité image P_S est définie sur $\{0, 1, \dots, n\}$:

$\forall k \in \{0, 1, \dots, n\}$:

$$P_S(k) = P(S^{-1}(k)) = P(\{\omega \text{ tels que } S(\omega) = k\}).$$

L'ensemble $\{\omega \text{ tel que } S(\omega) = k\}$ étant constitué de C_n^k configurations incompatibles, chacune ayant la probabilité $p^k (1-p)^{n-k}$, on en déduit :

$$P_S(k) = C_n^k p^k (1-p)^{n-k}.$$

Définition 2.2 S est la v.a. **binomiale** de paramètres (n, p) de loi P_S notée $\mathcal{B}(n; p)$.

2.2 Variables aléatoires discrètes

Définition 2.3 Une v.a. réelle X définie sur (Ω, \mathcal{T}, P) est discrète si $X(\Omega)$ est un ensemble au plus dénombrable de \mathbb{R} .

$$X(\Omega) = \{x_i\}_{i \in \mathbb{I}} \subseteq \mathbb{Z};$$

$$\forall i \quad P_X(x_i) = P(X^{-1}(x_i)) \stackrel{\text{noté}}{=} p_i.$$

On pourra écrire : $P_X = \sum_i p_i \delta_{x_i}$ où δ_{x_i} est la mesure de support $\{x_i\}$.

Rappelons que pour tout borélien B , $\delta_{x_i}(B)$ est égale à 1 si $x_i \in B$ et à 0 sinon.

Définition 2.4 La représentation graphique ou **diagramme d'une loi discrète** est le graphe de la fonction :

$$x \in \{x_i\}_{i \in \mathbb{I}} \longrightarrow P_X(x).$$

Définition 2.5 La **fonction de répartition** F_X de la variable aléatoire quelconque X , est la fonction :

$$F_X : x \in \mathbb{R} \rightarrow F_X(x) = P_X([-\infty, x]) = P(\{\omega \in \Omega / X(\omega) \leq x\}).$$

$$\forall a \text{ et } b \text{ tels que } a \leq b : P_X([a, b]) = F_X(b) - F_X(a).$$

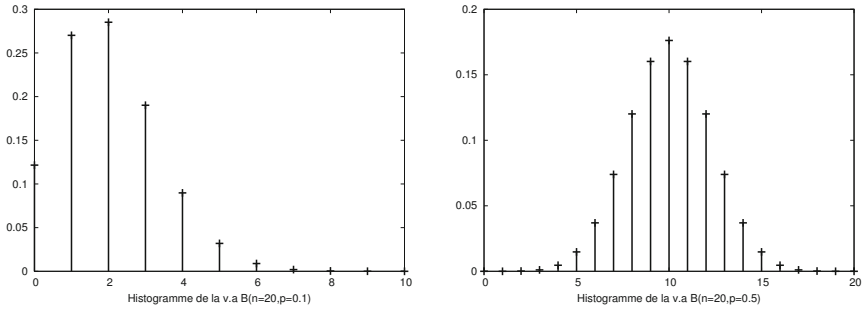


Figure 2.1 – Diagramme des lois binomiales $\mathcal{B}(n = 20; p = 0.1)$ et $\mathcal{B}(n = 20; p = 0.5)$.

Exemple 2.2 Si X est une v.a. binomiale $\mathcal{B}(n; p)$:

$$F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} C_n^k p^k (1-p)^{n-k},$$

où $\lfloor x \rfloor$ est le plus grand entier inférieur ou égal à x .

Théorème 2.3 (Propriétés de la fonction de répartition)

(1) F_X est croissante et continue à droite ; si X est discrète à valeurs x_i , F_X est discontinue en ces points.

(2) $\lim_{x \rightarrow +\infty} F_X(x) = 1$; $\lim_{x \rightarrow -\infty} F_X(x) = 0$.

Preuve (1) $x_1 < x_2$ alors $F_X(x_2) = F_X(x_1) + P_X(\lfloor x_1, x_2 \rfloor) \geq F_X(x_1)$.

Si l'on pose $]-\infty, x] = \cap_{n \in \mathbb{N}}]-\infty, x + \frac{1}{n}]$ il est facile de prouver la continuité à droite de F_X .

(2) $\lim_{x \rightarrow +\infty} F_X(x) = \lim_{x \rightarrow +\infty} P_X(\lfloor -\infty, x \rfloor) = P_X(\lim_{x \rightarrow +\infty} \lfloor -\infty, x \rfloor) = P_X(\mathbb{R}) = 1$ par continuité monotone de P_X . Raisonnement analogue pour la deuxième égalité. \square

Définition 2.6 (Espérance mathématique d'une v.a. quelconque)

L'espérance mathématique d'une v.a. X définie sur l'espace probabilisé (Ω, \mathcal{F}, P) est définie par :

$$\int_{X(\Omega)} x dP_X(x) = \int_{\Omega} X(\omega) dP(\omega) \text{ notée } E(X)$$

où \int doit être interprété comme symbole d'une intégrale classique dans le cas d'une v.a. absolument continue (voir : section suivante), ou celui d'une série dans le cas d'une v.a. discrète. Dans ce dernier cas, l'espérance mathématique

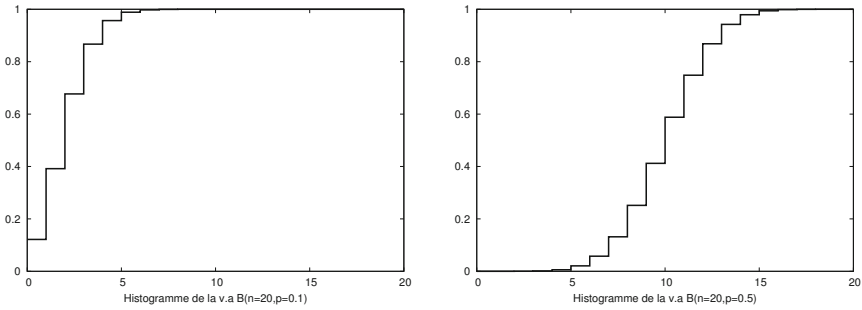


Figure 2.2 – Graphes des fonctions de répartition de lois binomiales : $\mathcal{B}(n = 20; p = 0, 1)$; $\mathcal{B}(n = 20; p = 0, 5)$.

est la moyenne de l'ensemble des valeurs possibles de X , pondérées par leurs probabilités respectives.

Définition 2.7 *L'espérance mathématique d'une v.a. discrète X de loi $P_X(x_i) = p_i$ ($i \in \mathbb{I} \subset \mathbb{Z}$) existe si $\sum_i |x_i| p_i < +\infty$; elle est définie par :*

$$\sum_{i \in \mathbb{I}} p_i x_i =_{(\text{noté})} E(X).$$

Analogie mécanique : l'espérance mathématique est le centre de gravité d'un ensemble de masses $(p_i)_i$ situées aux abscisses $(x_i)_i$ de la droite réelle.

2.1 ♠ Si X est une v.a. à valeurs entières positives, démontrer que :

$$E(X) = \sum_{n \geq 0} P(X > n).$$

Remarque 2.1 On dit que la v.a. X domine la v.a. Y si $P(X > Y) > \frac{1}{2}$; cependant la domination de X sur Y n'entraîne pas $E(X) > E(Y)$. Cette affirmation, qui peut paraître paradoxale, est illustrée par l'exercice suivant.

2.2 ♠ Étranges dés !

Deux dés équilibrés A et B portent respectivement les faces $\{3, 4, 10, 11, 12, 18\}$ et $\{2, 7, 8, 9, 16, 17\}$ (voir MANSUY R., *Des dés diaboliques* Quadrature, numéro 74). Soient les v.a. X_A et X_B associées aux faces obtenues.

(1) Montrer que $P(X_A > X_B) > \frac{1}{2}$, alors que $E(X_A) < E(X_B)$.

(2) Soit un troisième dé C dont l'ensemble des faces est $\{1, 5, 6, 13, 14, 15\}$.

Montrer que B domine C et que, contrairement à l'intuition de transitivité de la relation de domination, C domine A . Ce paradoxe est à rapprocher du paradoxe

de Condorcet. Il est facile d'imaginer un jeu, où face à un adversaire qui aura choisi l'un des trois dés, il vous suffira de choisir parmi les deux autres celui qui domine le dé choisi par votre adversaire pour l'emporter avec une probabilité supérieure à 0,5.

Théorème 2.4 (*Propriétés de l'espérance d'une v.a. discrète*) *L'espérance mathématique est un opérateur E défini sur l'espace vectoriel des variables aléatoires intégrables, qui vérifie les propriétés :*

(1) *pour toute v.a. X égale à c presque sûrement :*

$$E(X) = c.$$

(2) *X positive implique $E(X)$ positive.*

(3) $\forall B = \{x_i\}_i$ *sous-ensemble dénombrable de \mathbb{R} :*

$$E(\mathbf{1}_B) = P_X(B).$$

(4) *propriété de transfert : si h est une fonction définie et à valeurs dans \mathbb{R} telle que $\sum_i |h(x_i)|P_X(x_i)$ soit fini, alors :*

$$E(h(X)) = \sum_i h(x_i)P_X(x_i).$$

Preuve

(1) est triviale.

(2) les valeurs de X étant positives, leur moyenne pondérée l'est aussi.

(3) $E(\mathbf{1}_B) = \sum_i \mathbf{1}_B(x_i)P_X(x_i) = P_X(B).$

□

Les propriétés de linéarité de l'espérance seront définies au théorème 3.3 du chapitre suivant.

Définition 2.8 *Si $\sum_i p_i |x_i|^n < +\infty$, alors le **moment d'ordre n** de la v.a. X existe, et est défini par :*

$$\sum_{i \in \mathbf{I}} p_i x_i^n =_{(\text{noté})} E(X^n).$$

La notion de variance est l'indicateur classique de dispersion des valeurs d'une v.a., autour de son espérance mathématique :

Définition 2.9 La **variance** $V(X)$ d'une v.a. réelle X existe et est finie si $E(X^2)$ est fini. Elle est définie comme moyenne des écarts quadratiques de X autour de son espérance $E(X)$:

$$E((X - E(X))^2) \stackrel{\text{noté}}{=} V(X).$$

Analogie mécanique : $V(X)$ est comparable au moment d'inertie d'une distribution de masses sur la droite réelle par rapport à son centre de gravité $E(X)$. En pratique, on utilise l'expression équivalente et plus opératoire :

$$V(X) = E(X^2) - (E(X))^2.$$

La variance est toujours positive ou nulle car $E(X^2) \geq (E(X))^2$.

Définition 2.10 L'écart-type σ_X de la v.a. X est défini par :

$$\sqrt{V(X)} = \sigma_X.$$

Théorème 2.5 (Propriétés de la variance) Si la v.a. X est telle que $V(X)$ existe :

$$\begin{aligned} \forall a \in \mathbb{R} : \quad & V(a + X) = V(X) ; \\ & V(aX) = a^2 V(X) \text{ et } \sigma_{aX} = |a| \sigma_X. \end{aligned}$$

Preuve Démonstrations faciles laissées au lecteur. □

Fonction génératrice

Définition 2.11 À toute v.a. X à valeurs dans \mathbb{N} , on associe sa **fonction génératrice** $g_X(z)$ définie par la série entière :

$$g_X(z) = E(z^X) = \sum_{n \in \mathbb{N}} P(X = n) z^n, \quad \forall z \in D(0, 1) \subset \mathbb{C},$$

$g_X(z)$ est la transformée en Z (utilisée par les automaticiens) de la suite $\{P(X = n)\}_{n \in \mathbb{N}}$, continue dans son disque de convergence $D(0, 1)$, et holomorphe dans son intérieur $\overset{\circ}{D}(0, 1)$.

La fonction génératrice d'une v.a. entière positive caractérise complètement sa loi. D'un point de vue pratique, son utilisation facilite souvent le calcul de l'espérance et de la variance d'une v.a.

Théorème 2.6 La fonction génératrice g_X d'une v.a. entière X caractérise sa loi :

$$\forall n \in \mathbb{N}, \quad P_X(n) = \frac{1}{n!} \frac{d^n g_X(z)}{dz^n} \Big|_{z=0}.$$

Théorème 2.7 Pour toute v.a. X entière telle que $E(X^i)$ existe $\forall i \in \{1, 2, \dots, k\}$:

$$E(X) = \frac{dg_X(z)}{dz} \Big|_{z=1}; \quad E(X^2) = \frac{d^2 g_X(z)}{dz^2} \Big|_{z=1} + E(X);$$

$$\forall i \in \{1, 2, \dots, k\} : E(X(X-1)\dots(X-(k-1))) = \frac{d^i g_X(z)}{dz^i} \Big|_{z=1}.$$

Définition 2.12 Une v.a. X est dite de **Poisson** d'intensité λ (strictement positive), si sa loi est définie par :

$$(\forall k \in \mathbb{N}) \quad P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!};$$

$$E(X) = V(X) = \lambda.$$

On note :

$$X \sim \mathcal{P}(\lambda).$$

Les variables de Poisson décrivent le nombre aléatoire d'événements tels que les pannes d'un système ou les arrivées de clients dans un service.

Exemple 2.3 Si X est de loi de Poisson de paramètre λ , alors :

$g_X(z) = \sum_{n \geq 0} e^{-\lambda} z^n \frac{\lambda^n}{n!} = e^{\lambda(z-1)}$. D'où l'on déduit grâce au théorème précédent les résultats connus :

$$E(X) = \lambda e^{\lambda(z-1)} \Big|_{z=1} = \lambda \quad \text{et} \quad P_X(n) = \frac{1}{n!} \frac{d^n g_X(z)}{dz^n} \Big|_{z=0} = \frac{\lambda^n e^{-\lambda}}{n!}.$$

2.3 ♠ (1) Vérifier que la loi de Poisson est bien une loi de probabilité; démontrer que sa variance est égale à λ .

(2) Calculer $E\left(\frac{1}{1+X}\right)$.

2.4 ♠ Démontrer, à l'aide de deux méthodes, que l'espérance d'une v.a. $\mathcal{B}(n, p)$ est égale à np et que sa variance est égale à $np(1-p)$.

Définition 2.13 *Le modèle décrit par une variable géométrique est constitué par le dispositif suivant : on procède à des épreuves indépendantes ayant chacune deux issues : A qui apparaît avec la probabilité p et \bar{A} apparaissant avec la probabilité $(1 - p)$. On appelle **variable géométrique sur \mathbb{N}^*** le nombre X d'épreuves nécessaires à la première apparition de A. L'événement $(X = k)$ équivaut à l'événement : \bar{A} apparaît consécutivement $(k - 1)$ fois, puis apparaît A, ces événements consécutifs étant indépendants :*

$$P(X = k) = P(\bar{A}) \dots P(\bar{A})P(A) = (1 - p)^{k-1}p.$$

Sa loi est définie par :

$$\forall k \in \mathbb{N}^* \quad P(X = k) = p(1 - p)^{k-1}.$$

$$E(X) = \frac{1}{p} \quad V(X) = \frac{1 - p}{p^2}.$$

On note : $X \sim \mathcal{G}_{\mathbb{N}^*}(p)$.

Loi de la v.a. **géométrique sur \mathbb{N}** est définie par :

$$\forall k \in \mathbb{N}, P(X = k) = p(1 - p)^k; E(X) = \frac{1 - p}{p} \text{ et } V(X) = \frac{1 - p}{p^2}.$$

Exemple 2.4 Procédons à des lancers d'une pièce de monnaie : la v.a. X associée au rang du premier lancer donnant pile est de loi géométrique :

$$P(X = k) = (1 - p)^{k-1}p,$$

où $P(\text{un lancer donne pile}) = p$.

2.5 ♠♠ Un radar explore une région de l'espace contenant n objets ; lors d'un balayage, chaque objet est capté avec la probabilité p .

(1) Déterminer les probabilités pour qu'un objet donné soit capté au terme de 2, 3, ..., b balayages ; en déduire la probabilité pour que k objets donnés parmi les n , soient captés au terme de b balayage.

(2) Quel est le nombre b^* de balayages pour que la probabilité de capter tous les objets soit égale à une probabilité fixée π^* ?

(3) Combien de balayages doit-on effectuer pour que le nombre moyen d'objets découverts soit supérieur ou égal à m ?

Ce problème est analogue à celui, ô combien lancinant pour l'auteur d'un ouvrage scientifique, posé par la correction typographique d'épreuves écrites. Il se formule ainsi : chaque lecture correctrice d'un texte qui comporte n erreurs, permet de détecter et de corriger chacune d'elle avec la probabilité p ; les lectures étant successives et indépendantes entre elles, il s'agit de déterminer le

nombre de lectures que l'on doit effectuer pour détecter 95 % des n erreurs. (Une page contient en moyenne une ou deux erreurs, le livre contient 330 pages et $p = \frac{1}{2}$)

2.6 ♠♠ Un sauteur en hauteur tente de franchir successivement les hauteurs $1, 2, \dots, N$. Les résultats des sauts sont indépendants entre eux, et la probabilité de succès pour un saut de hauteur N est égale à $\frac{1}{N}$; d'autre part, le sauteur est éliminé dès son premier échec.

Déterminer la loi de la hauteur X du dernier saut réussi, ainsi que $E(X)$ et $V(X)$ en utilisant la fonction génératrice.

2.7 ♠♠♠ Une entreprise stocke en début d'année n unités d'un produit donné. Un produit vendu dans l'année rapporte un bénéfice b , tandis qu'un produit non vendu et restant en stock occasionne une perte a .

(1) Exprimer la v.a. $R_n(X)$ représentant le revenu annuel correspondant à une demande de X produits et un stock de n . En déduire l'expression du revenu annuel moyen $E(R_n)$.

(2) Déterminer la valeur n^* du stock permettant d'optimiser le revenu annuel moyen (on pourra raisonner à partir de la différence $E(R_{n+1}) - E(R_n)$).

(3) Dans le commerce vestimentaire, il est courant que les magasins revendent les vêtements trois fois leur prix d'achat ($b = 2a$). En faisant l'hypothèse que le nombre d'acheteurs d'un produit donné suit une loi géométrique ($P(X = x) = p(1 - p)^x$) de paramètre $p = 0, 1$, calculer la valeur de n^* .

Intermède historique : mathématisation du hasard au XVII^e siècle.

En 1654, Blaise Pascal (1623-1662) philosophe et mathématicien, soumit à son contemporain Pierre de Fermat (1601-1665), magistrat exerçant à Toulouse et mathématicien, les *problèmes des partis et du double-six* (voir : exercice 7, chapitre 1) que lui avait proposés son ami le chevalier de Méré, gentilhomme mondain et joueur réputé. Le problème des partis se formule ainsi : *comment partager le gain d'un tournoi interrompu avant sa fin, entre deux joueurs de force égale*. Les deux mathématiciens le résolurent par des méthodes différentes et enrichirent de nouveaux concepts le tout récent *Calcul des chances*, ancêtre de la théorie des probabilités.

Par ailleurs, dans la correspondance épistolaire entre les deux savants figure le fameux pari de Pascal, dans lequel (une fois n'est pas coutume dans l'œuvre du philosophe !) Dieu n'est pas seul maître du jeu, mais un être contingent comme vous et moi, et dont la probabilité d'existence est nettement inférieure à l'unité. Rappelons la formulation qu'en donna Pascal, apologiste zélé de la religion chrétienne : « *même si Dieu a une probabilité infime p d'exister, l'homme a intérêt à croire en lui, car l'espérance de gain spirituel obtenu par cette adhésion est égale à l'infini* ». En effet, cette espérance s'écrit aujourd'hui : « $p \cdot \infty + (1 - p) \cdot 0 = \infty$ ». Avec cette conclusion optimiste, où était mise en

œuvre la notion d'espérance mathématique, Pascal pouvait affirmer, que même si l'hypothèse de l'existence de Dieu était de faible probabilité, l'homme en retirerait un gain infini lors de sa résurrection. Les problèmes posés par Pascal et les solutions qu'il en donna, le confortèrent dans l'intuition qu'il établissait une nouvelle branche des mathématiques qu'il nomma : *géométrie du hasard*.

Durant la même période, le savant hollandais Christiaan Huygens (1629-1695) fit paraître en 1657 un traité de mathématique ayant pour titre *De ratiociniis in aleae ludo* (*Du calcul des jeux de dé*) considéré comme une synthèse des résultats du calcul des probabilités de son époque, dans lequel sont notamment définies les notions d'échantillonnage avec ou sans remise et d'espérance mathématique.

2.3 Variables aléatoires absolument continues

Omniprésentes dans la plupart des domaines scientifiques et industriels, ces variables aléatoires permettent de modéliser les répartitions aléatoires continues. L'usage tolère la dénomination de variables continues.

Définition 2.14 Une v.a. X est dite **absolument continue** ou **continue**, s'il existe une fonction f_X presque sûrement positive dite **densité de probabilité** (ou **densité**) de X d'intégrale sur \mathbb{R} égale à l'unité et telle que :

$$\text{pour tout borélien } B \text{ de } \mathbb{R}, \quad P_X(B) = \int_B f_X(x) dx.$$

Le résultat précédent s'exprime sous une forme plus intuitive :

$$\forall x \in \mathbb{R}, \quad P(X \in [x, x + \Delta x]) = f_X(x) \Delta x + o(\Delta x).$$

Rappel : une fonction vérifie presque sûrement une propriété, si elle la vérifie partout sauf éventuellement sur un ensemble de mesure nulle.

Dans le cadre de la théorie de la mesure, une mesure positive μ portée par \mathbb{R} est qualifiée d'absolument continue par rapport à la mesure de Lebesgue m , si pour tout borélien B de mesure de Lebesgue nulle, $\mu(B)=0$. Sous cette hypothèse, le théorème de Radon-Nykodim affirme qu'il existe nécessairement une fonction positive f appelée densité de μ , telle que pour tout borélien B , $\mu(B) = \int_B f(x) dm(x)$. C'est en ce sens qu'une v.a. est dite absolument continue, au sens où sa loi de probabilité P_X est une mesure absolument continue par rapport à m .

Définition 2.15 La **fonction de répartition** F_X d'une v.a. X absolument continue est définie par :

$$\forall x \in \mathbb{R}, \quad F_X(x) = \int_{-\infty}^x f_X(u) du = P_X([-\infty, x]).$$

Théorème 2.8 Soit P_X une loi absolument continue, définie par sa densité f_X .

(1) $\forall a, b$ réels tels que $a < b$:

$$P_X([a, b]) = \int_a^b f_X(x) dx.$$

(2) $\forall a \in \mathbb{R}, P_X(a) = 0$. (Interprétation : une variable absolument continue ne charge aucun point d'une probabilité strictement positive.)

(3) La fonction de répartition F_X est continue sur \mathbb{R} et $f_X = F_X'$ sauf éventuellement sur un sous-ensemble de mesure de Lebesgue nulle.

Preuve (1) $P_X([a, b]) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx$;

(2) $P_X(a) = \int_{\{a\}} f_X(x) dx = 0$, car $\{a\}$ est de mesure de Lebesgue nulle. \square

Définition 2.16 Le **support** D d'une v.a. est l'adhérence de l'ensemble des points sur lequel la densité est strictement positive ; on fera suivre l'expression d'une densité par la fonction indicatrice $\mathbf{1}_D$ de son support D .

Deux variables aléatoires distinctes peuvent avoir la même loi. C'est le cas de deux variables normales centrées X et sa symétrique $-X$, toutes deux de même densité ; ce résultat est vrai pour les v.a. X de lois symétriques par rapport à l'origine.

Définition 2.17 La **médiane** x_M d'une loi de probabilité d'une v.a. X continue est définie par :

$$P(X < x_M) = P(X > x_M) = \frac{1}{2}.$$

Définition 2.18 X est une **variable uniforme** de support $[a, b]$ si sa densité est définie par :

$$f_X(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) ;$$

$$E(X) = \frac{a+b}{2} ; V(X) = \frac{(b-a)^2}{12} ; x_M = E(X).$$

Pour tout borélien B de $[a, b]$, $P(X \in B) = \frac{\mu(B)}{(b-a)}$, où μ est la mesure de Lebesgue.

Définition 2.19 X est une **variable exponentielle de paramètre λ** , notée $\mathcal{E}(\lambda)$ si sa densité est définie par :

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

$$E(X) = \frac{1}{\lambda} ; V(X) = \frac{1}{\lambda^2}.$$

Les variables exponentielles modélisent les temps d'attente à l'entrée d'un service ou les durées de vie de systèmes à usure non cumulative.

Construction de la variable exponentielle comme limite de la variable géométrique

Considérons un système mis en marche au temps 0, pouvant tomber en panne sur tout intervalle de temps d'étendue h (petit par rapport à 1) avec la probabilité λh ; on suppose en outre que les états de la machine dans des intervalles de temps disjoints sont indépendants.

La probabilité P (la machine tombe pour la première fois en panne dans l'intervalle de temps $[nh, (n+1)h]$) est donc égale à $(1 - \lambda h)^n \lambda h$, qui est la loi géométrique de paramètre λh . Si l'on pose $t = nh$, la probabilité précédente s'écrit $(1 - \lambda \frac{t}{n})^n \lambda \frac{t}{n}$. Pour une valeur t fixée, faisons tendre le pas h vers 0 et n vers $+\infty$: $(1 - \lambda \frac{t}{n})^n \lambda \frac{t}{n}$ converge alors vers $\exp(-\lambda t) \lambda dt$: $P(\text{la machine tombe pour la première fois en panne dans } [t, t+dt]) \simeq f(t) dt$.

Conclusion : la durée de vie de la machine est de loi exponentielle définie par sa densité $f(t) = \lambda \exp(-\lambda t)$.

La variable exponentielle est donc l'analogue continu de la variable géométrique; elles vérifient toutes deux la propriété d'absence de mémoire qui les caractérise, étudiée au chapitre 6.

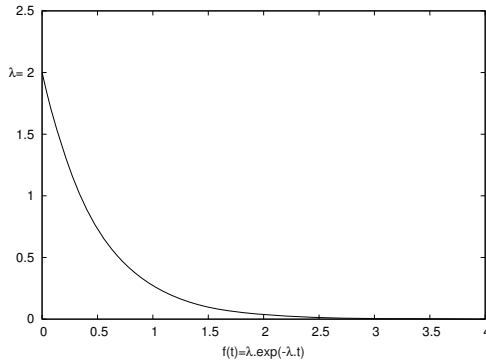


Figure 2.3 – Densité de la loi exponentielle.

2.8 ♠♠ Considérons une v.a. X exponentielle de paramètre λ . Soit Y la partie entière de X notée $\lfloor X \rfloor$.

- (1) Démontrer que Y est de loi géométrique;
- (2) Calculer $E(Y)$, $V(Y)$ ainsi que la loi de $X - Y$.

[Indication : poser $A = \ll X - Y \leq u \gg$, $B_k = \ll Y = k \gg$, et utiliser le théorème 1.2 des probabilités totales.]

Définition 2.20 L'espérance $E(X)$ d'une v.a. X de densité f_X est définie, si $\int_{-\infty}^{+\infty} |x| f_X(x) dx < +\infty$, par :

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx.$$

Définition 2.21 Le moment d'ordre n $E(X^n)$ de la v.a. X est défini, si $\int |x|^n \cdot f_X(x) dx$ existe

$$E(X^n) = \int x^n \cdot f_X(x) dx.$$

La propriété de transfert déjà énoncée pour les v.a. discrètes (voir théorème 2.4), permet de calculer l'espérance d'une fonction $\varphi(X)$ d'une v.a. X sans devoir en déterminer sa loi :

Théorème 2.9 Étant donné une v.a. X de densité f_X et une fonction φ telle que $\int_{-\infty}^{+\infty} |\varphi(x)| f_X(x) dx < +\infty$ alors : $E(\varphi(X)) = \int_{-\infty}^{+\infty} \varphi(x) f_X(x) dx$.

Définition 2.22 La v.a. de Cauchy est de densité :

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{\pi(1+x^2)}.$$

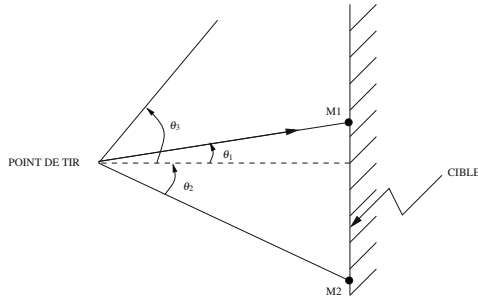
Remarque 2.2 Pour tout entier positif n , $E(X^n) = \int_{-\infty}^{+\infty} x^n \frac{1}{\pi(1+x^2)} dx$ n'est pas définie : en conséquence, la v.a. de Cauchy n'a ni espérance ni variance.

2.9 ♠ Construction géométrique de la variable de Cauchy

Soit X une v.a. de densité f_X et h une fonction croissante définie sur le support de la densité et à valeurs dans \mathbb{R} : exprimer $F_{h(X)}$ en fonction de F_X et en déduire la densité de $h(X)$. Appliquer ce résultat à la construction de la loi de Cauchy : X est une v.a. uniforme sur $]-\frac{\pi}{2}, \frac{\pi}{2}[$ et $h(x) = tg(x)$.

Supposons une cible constituée par la droite verticale située à une distance l du point de tir placé en 0. En ce point un tireur « fou » tire au hasard, au sens où son angle Θ de tir varie uniformément dans $]-\frac{\pi}{2}, \frac{\pi}{2}[$.

Si grande que soit l'ordonnée y de l'impact, la probabilité que l'impact atteigne la cible au-delà de y n'est pas négligeable : en effet, $P(Y > y) = P(\frac{\pi}{2} > \Theta > \theta) = \frac{1}{2}(1 - \frac{2}{\pi}) \cdot \theta$, où $\theta = \arctg(y)$. Cette propension qu'a la v.a. de Cauchy à prendre de très grandes valeurs avec des probabilités non négligeables a pour conséquence la non-existence de son espérance et donc d'aucun de ses moments.



Théorème 2.10 (*Inégalités classiques*)

(1) **Inégalité de Tchebychev** (la plus fréquemment utilisée) : pour toute v.a. X d'espérance m et de variance σ^2 finie :

$$\forall a > 0, P(|X - m| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Si $a = n\sigma$, $P(|X - m| \geq n\sigma) \leq \frac{1}{n^2}$.

(2) **Règle des trois sigmas** (conséquence de l'inégalité de Tchebychev) :

pour toute v.a. X de moyenne m et de variance σ^2 :

$$P(X \in [m - 3\sigma ; m + 3\sigma]) \geq 0,9.$$

Interprétation : au moins 90% de la masse de la distribution de probabilité est condensée sur un segment centré en sa moyenne et d'étendue égale à six écarts types.

(3) **Inégalité de Markov** : pour toute v.a. X positive et toute constante a strictement positive :

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

(4) **Inégalité de Jensen** : X une v.a. quelconque et h une fonction convexe sur l'ensemble des valeurs de X , alors si $E(X)$ et $E(h(X))$ sont finies, on a :

$$h(E(X)) \leq E(h(X)).$$

Preuve (1) Posons $U = X - E(X)$; l'inégalité s'écrit :

$$\begin{aligned} P(|U| \geq a) &\leq \frac{E(U^2)}{a^2} \iff a^2 \left(\int_{-\infty}^{-a} dP_U(u) + \int_a^{+\infty} dP_U(u) \right) \leq E(U^2) \\ &= \int_{-\infty}^{-a} u^2 dP_U(u) + \int_{-a}^a u^2 dP_U(u) + \int_a^{+\infty} u^2 dP_U(u). \end{aligned}$$

$$(3) P(X \geq a) = \int_a^{+\infty} dP_X(x) \leq \frac{1}{a} \int_a^{+\infty} x dP_X(x) \leq \frac{E(X)}{a}.$$

(4) Preuve dans le cas discret : $h(E(X)) = h(\sum p_i x_i) \leq \sum p_i h(x_i) = E(h(X))$.

□

2.4 Variables gaussiennes

Dans la pratique, la variable gaussienne est la plus fréquemment utilisée, d'où son autre dénomination de variable *normale* donnée par les probabilistes du XIX^e siècle. L'un des grands théorèmes de la théorie des probabilités, le théorème de la limite centrale, fonde rigoureusement ce constat (voir théorème 5.9).

Définition 2.23 X est une **variable gaussienne ou normale** d'espérance m et de variance σ^2 , si sa loi est définie par la densité :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \mathbf{1}_{\mathbb{R}}(x).$$

On écrit : $X \sim \mathcal{N}(m, \sigma^2)$.

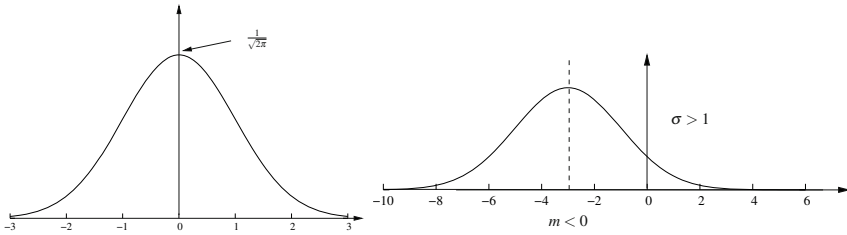
Si $m = 0$, X est dite *centrée* ; si $\sigma = 1$, X est dite *réduite*.

Remarque 2.3 Si X^* est une v.a. $\mathcal{N}(0, 1)$, alors :

$$P(-1,96 < X^* < 1,96) \simeq 0,95 \text{ et } P(-2,57 < X^* < 2,57) \simeq 0,99.$$

Remarque 2.4 La suite des densités $f_{X_n}(x) = \frac{n}{\sqrt{2\pi}} e^{-\frac{n^2 x^2}{2}}$ des v.a. X_n centrées de variances $\frac{1}{n^2}$ converge au sens des distributions vers la distribution de Dirac δ_0 .

- 2.10 ♠** (1) Soit la v.a. X de densité f_X et $Y = aX + b$. Densité de Y ?
 (2) Application : X est de loi $\mathcal{N}(0, 1)$ et $Y = \sigma X + m$.

Figure 2.4 – Graphes de la densité $\mathcal{N}(0, 1)$ et de la densité $\mathcal{N}(-3, 4)$

2.11 ♠ X est une v.a. de densité f_X ; calculer la densité, l'espérance et la variance de $|X|$. Considérer le cas où X est de loi $\mathcal{N}(0, \sigma^2)$.

2.12 ♠♠ Slalom

Un skieur affronte un slalom de vingt portes ; s'il ne tombe pas, son temps de parcours (en secondes) X suit une loi normale $\mathcal{N}(m = 50, \sigma^2 = 4)$. À chaque porte il peut tomber avec la probabilité $\frac{1}{20}$, ce qui lui fait perdre 4 secondes. On admet que les chutes sont indépendantes entre elles.

Quelle probabilité a-t-il de décrocher la Flèche de platine (attribuée lorsque le temps de descente est inférieur à 52) ?

[Indication : exprimer le temps de parcours sous forme d'une somme de v.a.]

Intermède biographique : Carl Friedrich Gauss (1777-1855)

Gauss, mathématicien de génie, né à Brunswick (Allemagne) de parents ouvriers, a produit une œuvre scientifique monumentale et sans doute inégalée. Enfant prodige, aux dons rapidement remarqués et mis en valeur par ses professeurs, il obtint ses premiers résultats à l'âge de dix-huit ans à l'université de Göttingen, et publia en 1801 un ouvrage fondamental sur l'arithmétique intitulé *Disquisitiones arithmeticae*. Son œuvre concerne la plupart des domaines mathématiques et de la physique, notamment : la géométrie différentielle, la géodésie, les géométries non euclidiennes, l'algèbre, la théorie des nombres, l'électromagnétisme, l'optique. Mathématicien solitaire, Gauss a eu peu d'élèves mais sans doute les plus brillants, tels que : Dirichlet, Dedekind et Riemann. Il apparaît comme étant l'un des premiers mathématiciens modernes annonçant la conception formaliste des mathématiques, lorsqu'il affirmait que : « *Les mathématiciens font abstraction de la nature des objets mathématiques et de la signification de leurs relations, qu'ils se contentent d'énumérer et de relier rigoureusement.* »

Calcul des probabilités gaussiennes

X étant de loi normale $\mathcal{N}(m, \sigma^2)$, donner une approximation à quatre décimales de la probabilité $P(a < X < b)$, pour tous réels a et b tels que $a < b$. Les tables

classiques de la loi normale (voir Annexes à la fin de l'ouvrage) donnent les valeurs $F(x)$ de la fonction de répartition de la variable gaussienne centrée et réduite X^* , pour des valeurs x comprises entre 0 et 2,99, car $P(|X^*| > 3) \simeq 0,001$.

Sachant que $X = m + \sigma X^*$:

$$P(a < X < b) = P\left(\frac{a-m}{\sigma} < X^* < \frac{b-m}{\sigma}\right) = F\left(\frac{b-m}{\sigma}\right) - F\left(\frac{a-m}{\sigma}\right).$$

1° cas : a et b sont supérieurs à m .

$$P(a < X < b) = F\left(\frac{b-m}{\sigma}\right) - F\left(\frac{a-m}{\sigma}\right).$$

Exemple 1 : $m = 2$, $\sigma = 3$, $a = 4$, $b = 6$: $F(\frac{b-m}{\sigma}) = F(\frac{4}{3}) \simeq F(1,33) = 0,9082$.

$F(\frac{a-m}{\sigma}) = F(\frac{2}{3}) \simeq F(0,66) = 0,7454$, d'où $P(a < X < b) = 0,1628$.

2° cas : a est inférieur à m , b est supérieur à m .

$$P(a < X < b) = F\left(\frac{b-m}{\sigma}\right) + F\left(\frac{m-a}{\sigma}\right) - 1.$$

$F(\frac{a-m}{\sigma}) = 1 - F(\frac{m-a}{\sigma})$ par symétrie de F par rapport à 0.

Exemple 2 : $a = 1$, $m = 2$, $\sigma = 3$, $b = 6$.

$$\begin{aligned} P(a < X < b) &= F\left(\frac{4}{3}\right) + F\left(\frac{1}{3}\right) - 1 \simeq F(1,33) + F(0,33) - 1 \\ &= 0,9082 + 0,6293 - 1 = 0,5375. \end{aligned}$$

3° cas : a et b sont inférieurs à m .

$$P(a < X < b) = F\left(\frac{m-a}{\sigma}\right) - F\left(\frac{m-b}{\sigma}\right).$$

Exemple 3 : $a = -2$, $b = -1$, $m = 2$, $\sigma = 3$.

$P(a < X < b) = F(\frac{4}{3}) - F(\frac{3}{3}) \simeq F(1,33) - F(1) = 0,908 - 0,8413 = 0,067$.

On démontrera dans son cadre théorique naturel au chapitre 5, le résultat important d'approximation de la loi binomiale par la loi gaussienne : si X_n est une v. a binomiale $\mathcal{B}(n; p)$ alors la loi de $\frac{X_n}{n}$ est approchée par la loi normale $\mathcal{N}(p, \frac{p(1-p)}{n})$ pour des valeurs de n supérieures à 30.

2.13 ♠ Où il est établi rigoureusement qu'en ces temps incertains on ne devrait jamais manquer d'air, fût-il pollué !

Dans une pièce fermée de volume V et contenant N molécules d'air (N étant un multiple du nombre d'Avogadro), on considère un volume v petit par rapport à V contenant un nombre aléatoire X de molécules d'oxygène, autour du nez d'un quidam immobile.

(1) Utiliser la remarque précédente pour démontrer que la loi de X est proche d'une loi normale dont on déterminera les paramètres.

(2) Si on considère que les signes d'asphyxie se manifestent dès que X varie autour de sa moyenne au-delà de 1% de sa valeur, déterminer la probabilité de cet événement. En déduire la conclusion (rassurante) annoncée. (Considérer l'approximation précédente.)

Définition 2.24 On appelle **fonction d'erreur**, notée $\text{erf}(x)$, la fonction définie par :

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \quad \forall x \geq 0.$$

La fonction d'erreur complémentaire, notée $\text{erfc}(x)$:

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-\frac{t^2}{2}} dt \quad \forall x \geq 0.$$

Application : $P(|X| \geq x) = \text{erfc}(\frac{x}{\sqrt{2}})$ et $P(|X| < x) = \text{erf}(\frac{x}{\sqrt{2}})$.

Loi de Pareto

Cette loi fut découverte par le mathématicien italien V. Pareto (1848-1923) qui tentait de modéliser en 1880 la répartition de la richesse des ménages italiens en fonction de leurs revenus. Il fit un constat, aujourd'hui vérifié et transposable à de nombreux domaines, que 20% des ménages possédaient 80% de la richesse nationale. La loi dite du 80/20, qui, de manière plus réaliste, pourrait aujourd'hui s'appeler 85/15, s'applique à de nombreux domaines de l'économie sociale et industrielle. Quelques exemples d'application de la loi :

- 15% de la population mondiale détient 85% des richesses ;
- 20% des moyens humains et financiers permettent d'atteindre 80% des objectifs ;
- 20% des noms de famille d'un pays sont portés par 80% de la population ;
- 20% des ventes de grandes surfaces fournissent 80% des bénéfices.

2.14 ♠ (1) La v.a. de Pareto X est de densité :

$$f_X(x) = \frac{K(x_m, \alpha)}{x^{\alpha+1}} \mathbf{1}_{[x_m, +\infty[}(x) \text{ où } x_m \text{ et } \alpha > 0.$$

Tracer les graphes de f_X pour $(\alpha = 1, x_m = 2)$ et $(\alpha = 2, x_m = 3)$. Déterminer la constante $K(x_m, \alpha)$.

(2) À quelles conditions l'espérance et la variance sont-elles définies ? Les calculer.

(3) Au vu des graphes des densités, utiliser ce modèle pour illustrer les exemples présentés.

Remarquable construction de la loi normale due à Gauss

Gauss postule que la densité f qui régit les erreurs x entâchant l'observation d'une quantité donnée θ vérifie les propriétés suivantes :

(P_1) les erreurs sont aussi souvent positives que négatives, d'où l'hypothèse de parité de f ;

(P_2) les grandes erreurs sont improbables : $\lim_{|x| \rightarrow +\infty} f(x) = 0$;

(P_3) f vérifie : $f \geq 0$ et $\int_{-\infty}^{+\infty} f(x) dx = 1$.

Sachant que les observations sont indépendantes entre elles, la probabilité d'effectuer des erreurs successives comprises entre x_1 et $x_1 + dx_1$, x_2 et $x_2 + dx_2$, ... , x_n et $x_n + dx_n$ vaut $\prod_{i=1}^n f(x_i) dx_i$.

Problème : les x_i étant connus, quelle est la valeur de θ la plus vraisemblable, c'est-à-dire celle qui maximise la probabilité d'observation des (x_i) ou, ce qui revient au même, des erreurs $(e_i)_i$ où $e_i = x_i - \theta$.

Gauss recherche donc la fonction f telle que $L(\theta) = \prod_{i=1}^n f(x_i - \theta)$ atteigne son maximum en $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, et qui maximise la somme $\sum_{i=1}^n (x_i - \hat{\theta})^2$. En langage contemporain, il s'agit de déterminer f telle que l'estimation de θ par la méthode des moindres carrés équivale à l'estimation de θ par la méthode du maximum de vraisemblance.

Soit $\ln(L(\theta)) = \sum_i \ln(f(x_i - \theta))$.

Pour tout n , pour tout $(x_i)_{i=1, \dots, n}$, (*) $\frac{\partial}{\partial \theta} \ln(L(\theta))$ doit être égal à 0. Posons $h = \ln(f)$. (*) équivaut à (**) $\sum_i h'(x_i - \bar{x}_i) = 0$.

Gauss pose $x_2 = x_3 = \dots = x_n = x_1 - ny$, d'où $\bar{x} = x_1 - (n-1)y$.

(**) équivaut à $h((n-1)y) + (n-1)h'(-y)$ pour tout n et pour tout y .

Or f est paire : $h'(-y) = -h'(y)$ et $h'((n-1)y) = (n-1)h'(y)$, soit

$$\frac{h'((n-1)y)}{(n-1)y} = \frac{h'(y)}{y} = a \text{ (constante),}$$

d'où : $\ln f(y) = h(y) = a \frac{y^2}{2} + b$ et par les propriétés $(P_i)_i$: f est la densité gaussienne centrée.

Les problèmes 3 et 4 se rapportent à ce chapitre.

2.5 Corrigés des exercices

Exercice 2.1

$$\sum_{i \geq 0} P([X > i]) = \sum_{i \geq 0} \sum_{k=i+1}^{+\infty} P([X = k]) = \sum_{k=1}^{+\infty} \sum_{i=0}^{k-1} P([X = k]) = \sum_{k=1}^{+\infty} kP([X = k]) = \mathbb{E}(X).$$

Exercice 2.2

- (1) $P(X_A > X_B) = \frac{1}{36} \times (\text{nombre de } (a_i, b_j)_{1 \leq i, j \leq 6} \text{ tels que } a_i > b_j) = \frac{20}{36} > \frac{1}{2}$.
 Or $E(X_A) = \frac{58}{6} < E(X_B) = \frac{59}{36}$.
 (2) La deuxième question relève d'un traitement analogue.

Exercice 2.3

$$(1) E(X) = \sum_{n=1}^{+\infty} n e^{-\lambda} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} \sum_{n=1}^{+\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda.$$

$$\begin{aligned} E(X^2) &= \sum_{n=1}^{+\infty} n^2 e^{-\lambda} \frac{\lambda^n}{n!} = \sum_{n=2}^{+\infty} n(n-1) e^{-\lambda} \frac{\lambda^n}{n!} + \sum_{n=1}^{+\infty} n e^{-\lambda} \frac{\lambda^n}{n!} \\ &= \lambda^2 e^{-\lambda} \sum_{n=2}^{+\infty} \frac{\lambda^{n-2}}{(n-2)!} + \lambda = \lambda^2 + \lambda. \end{aligned}$$

$$(2) E\left(\frac{1}{1+X}\right) = \sum_{k=0}^{+\infty} \frac{1}{k+1} P(X=k) = \frac{e^{-\lambda}}{\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^{k+1}}{(k+1)!} = \frac{e^{-\lambda}}{\lambda} (e^\lambda - 1) = \frac{1-e^{-\lambda}}{\lambda}.$$

Exercice 2.4

$E(X) = E(\sum_i B_i) = \sum_i E(B_i) = np$ et $V(X) = \sum_i V(B_i) = np(1-p)$ par indépendance des B_i . On peut utiliser aussi la fonction génératrice.

Exercice 2.5

- (1) $P(\text{objet donné soit capté par 2 balayages}) = 1 - (1-p)^2$. De même,

$$P(\text{objet donné soit capté par } b \text{ balayages}) = 1 - (1-p)^b$$

et

$$P(k \text{ objets soient captés par } b \text{ balayages}) = C_n^k (1 - (1-p)^b)^k (1-p)^{b(n-k)}.$$

- (2) $P(\text{tous les objets soient captés par } b \text{ balayages}) = (1 - (1-p)^b)^n$,
 d'où $\pi^* = (1 - (1-p)^{b^*})^n$, soit $(1-p)^{b^*} = 1 - (\pi^*)^{1/n}$ et finalement $b^* = \frac{\ln(1 - (\pi^*)^{1/n})}{\ln(1-p)}$.

(3) On a vu en (2) que le nombre X d'objets captés en b balayages est de loi $\mathcal{B}(n; 1 - (1-p)^b)$; donc $E(X) = n(1 - (1-p)^b)$ et $E(X) \geq m$ si et seulement si $b \geq \frac{\ln(1 - (\frac{m}{n}))}{\ln(1-p)}$.

Exercice 2.6

Soit A_k l'événement « le sauteur réussit la hauteur k ». On a, par hypothèse, $P(A_k) = \frac{1}{k}$.

Si X désigne la hauteur du dernier saut réussi, on a $X(\Omega) = \mathbb{N}^*$ (car $P(A_1) = 1$) et, pour $n \in \mathbb{N}^*$,

$$[X = n] = A_1 \cap \cdots \cap A_n \cap \overline{A_{n+1}}.$$

Les sauts étant supposés indépendants, les événements A_k le sont aussi et les probabilités se multiplient, soit :

$$P([X = n]) = \prod_{k=1}^n P(A_k) \times P(\overline{A_{n+1}}) = \frac{1}{n!} \times \left(1 - \frac{1}{n+1}\right).$$

Ainsi, $P([X = n]) = \frac{1}{n!} - \frac{1}{(n+1)!}$ pour tout $n \in \mathbb{N}^*$.

$$\begin{aligned} G_X(z) &= \mathbb{E}(z^X) = \sum_{n=1}^{+\infty} z^n \left(\frac{1}{n!} - \frac{1}{(n+1)!} \right) = \sum_{n=1}^{+\infty} \frac{z^n}{n!} - \sum_{n=1}^{+\infty} \frac{z^n}{(n+1)!} \\ &= e^z - 1 - \frac{1}{z} (e^z - 1 - z) = e^z \left(1 - \frac{1}{z} \right) + \frac{1}{z}. \end{aligned}$$

$$G'_X(z) = e^z \left(1 - \frac{1}{z} + \frac{1}{z^2} \right) - \frac{1}{z^2} \text{ et } G''_X(z) = e^z \left(1 - \frac{1}{z} + \frac{1}{z^2} + \frac{1}{z^2} - \frac{2}{z^3} \right) + \frac{2}{z^3},$$

donc $G'_X(1) = e - 1$ et $G''_X(1) + G'_X(1) - (G'_X(1))^2 = 2 + (e - 1) - (e - 1)^2 = 3e - e^2$, soit :

$$\mathbb{E}(X) = e - 1 \text{ et } V(X) = 3e - e^2.$$

Exercice 2.7

(1) Si $X < n$, $R_n = Xb - (n - X)a$ et si $X \geq n$, $R_n = nb$.

Revenu moyen annuel $E(R_n)$, correspondant à un stock de taille n :

$$\begin{aligned}
 E(R_n) &= \sum_{x=0}^{n-1} (xb - (n-x)a)P(X=x) + \sum_{x=n}^{+\infty} nbP(X=x) \\
 &= (a+b) \sum_{x=0}^{n-1} xP(X=x) - na \sum_{x=0}^{n-1} P(X=x) + nb \sum_{x=n}^{+\infty} P(X=x) \\
 &= (a+b) \sum_{x=0}^{n-1} xP(X=x) - na \sum_{x=0}^{n-1} P(X=x) + nb \left[1 - \sum_{k=0}^{n-1} P(X=k) \right] \\
 &= (a+b) \sum_{x=0}^{n-1} xP(X=x) - n(a+b) \sum_{x=0}^{n-1} P(X=x) + nb \\
 &= nb + (a+b) \sum_{x=0}^{n-1} (x-n)P(X=x).
 \end{aligned}$$

$$(2) \quad E(R_{n+1}) = (n+1)b + (a+b) \sum_{x=0}^n (x-n-1)P(X=x).$$

D'où

$$E(R_{n+1}) - E(R_n) = b - (a+b) \sum_{x=0}^n P(X=x) = b - (a+b)P(X < n+1).$$

Le revenu annuel augmente tant que $E(R_{n+1}) - E(R_n) \geq 0$, la valeur n^* optimisant le stock sera donc obtenue pour la première valeur n telle que $E(R_{n+1}) - E(R_n) < 0$, c'est-à-dire :

$$b - (a+b)P(X < n+1) < 0,$$

et donc on cherche la première valeur n telle que : $P(X < n+1) > \frac{b}{a+b}$.

(3) On a maintenant $b = 2a$ et $P(X=x) = pq^x$ avec $p = 0,1$ et $q = 1-p = 0,9$.

$P(X < n+1) = \sum_{x=0}^n P(X=x) = p \left(\frac{1-q^{n+1}}{1-q} \right) = 1 - q^{n+1} < \frac{2}{3}$ si et seulement si $q^{n+1} > \frac{1}{3}$, soit $n+1 > -\frac{\ln(3)}{\ln(0,9)}$ ou encore $n > 9,4$; d'où $n^* = 10$.

Exercice 2.8

(1) $P([Y = k]) = P([k \leq X < k+1]) = F_X(k+1) - F_X(k) = (1 - e^{-\lambda(k+1)}) - (1 - e^{-\lambda k})$, soit $P([Y = k]) = e^{-\lambda k}(1 - e^{-\lambda})$ pour tout $k \in \mathbb{N}$. Donc Y suit la loi géométrique $\mathcal{G}_0(1 - e^{-\lambda})$.

$$E(X) = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2} \quad \text{et} \quad E(Y) = \frac{e^{-\lambda}}{1-e^{-\lambda}} = \frac{1}{e^{\lambda}-1}.$$

(2) Pour tout $u \in [0, 1[$,

$$P([X - Y \leq u]) = P\left(\bigcup_{k \in \mathbb{N}} [X - Y \leq u] \cap [Y = k]\right) = \sum_{k \geq 0} P([k \leq X \leq k+u]).$$

$P([k \leq X \leq k+u]) = F_X(k+u) - F_X(k) = e^{-\lambda k} - e^{-\lambda(k+u)} = e^{-\lambda k}(1 - e^{-\lambda u})$, et $\sum_{k \geq 0} e^{-\lambda k} = \frac{1}{1-e^{-\lambda}}$:

$$F_{X-Y}(u) = \frac{1-e^{-\lambda u}}{1-e^{-\lambda}} \text{ pour } u \in [0, 1[\text{ et } f_{X-Y}(u) = \frac{\lambda e^{-\lambda u}}{1-e^{-\lambda}} \mathbf{1}_{[0,1[}(u).$$

Exercice 2.9

(1) $P([h(X) \leq y]) = P([X \leq h^{-1}(y)])$ d'où

$$f_Y(y) = \frac{d}{dy} F_X(h^{-1}(y)) = \frac{d}{dy} (h^{-1}(y)) f_X(h^{-1}(y)).$$

$$\text{Si } f_X(x) = \frac{1}{\pi} \mathbf{1}_{]-\frac{\pi}{2}, \frac{\pi}{2}[}(x), h^{-1}(y) = \arctan(y), f_Y(y) = \frac{1}{\pi(1+y^2)} \mathbf{1}_{]-\infty, +\infty[}(y).$$

Exercice 2.10

(1) $Y = aX + b$; $F_Y(y) = P([Y \leq y]) = P([aX + b \leq y]) = P([aX \leq y - b])$.

Si $a > 0$, $F_Y(y) = P\left([X \leq \frac{y-b}{a}]\right) = F_X\left(\frac{y-b}{a}\right)$ et $f_Y(y) = F'_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$.

Si $a < 0$, $F_Y(y) = P\left([X \geq \frac{y-b}{a}]\right) = 1 - P\left([X < \frac{y-b}{a}]\right) = 1 - F_X\left(\frac{y-b}{a}\right)$, car F_X est continue et $P\left([X = \frac{y-b}{a}]\right) = 0$. On a alors $f_Y(y) = F'_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right)$.

$$\text{Conclusion : } f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

(2) Si X^* est de loi $\mathcal{N}(0, 1)$, $f_{X^*}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$,

$$f_{\sigma X^* + m}(y) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right).$$

Exercice 2.11

$$F_T(t) = P([T \leq t]) = P([|X| \leq t]).$$

Si $t \leq 0$, $F_T(t) = 0$.

Si $t > 0$, $F_T(t) = P([-t \leq X \leq t]) = F_X(t) - F_X(-t)$:

$$f_T(t) = F'_T(t) = f_X(t) + f_X(-t).$$

Finalement, $f_T(t) = [f_X(t) + f_X(-t)] \mathbf{1}_{]0, +\infty[}(t)$.

Exercice 2.12

Si T est le temps du parcours, on a $T = X + 4 \times N$, où X suit (approximativement) la loi normale $\mathcal{N}(50, 4)$, et N la loi binomiale $\mathcal{B}(20, \frac{1}{20})$ (il y a vingt portes, chacune ayant la probabilité $p = \frac{1}{20}$ d'être accrochée. On cherche

$P([T < 52])$. On a :

$$\begin{aligned} P([T < 52]) &= P([X + 4N < 52]) = P\left(\bigcup_{n=0}^{20} [X + 4n < 52] \cap [N = n]\right) \\ &= P\left(\bigcup_{n=0}^{20} [X + 4n < 52] \cap [N = n]\right) = P\left(\bigcup_{n=0}^{20} [X < 52 - 4n] \cap [N = n]\right) \\ &= \sum_{n=0}^{20} P([X < 52 - 4n] \cap [N = n]). \end{aligned}$$

Comme X et N sont indépendantes, on a

$$P([T < 52]) = \sum_{n=0}^{20} P([X < 52 - 4n])P([N = n]).$$

X suit la loi $\mathcal{N}(50, 4)$ donc $\frac{X-50}{2}$ suit la loi $\mathcal{N}(0, 1)$, de fonction de répartition Φ , donnée par les tables pour $x > 0$, (avec, pour $x < 0$, $\Phi(x) = 1 - \Phi(-x)$). Ainsi :

$$P([X < 52 - 4n]) = P\left(\left[\frac{X - 50}{2} < \frac{52 - 4n - 50}{2}\right]\right) = \Phi(1 - 2n)$$

et $P([N = n]) = C_{20}^n p^n (1-p)^{20-n} = C_{20}^n \left(\frac{1}{20}\right)^n \left(\frac{19}{20}\right)^{20-n}$ et donc :

$$P([T < 52]) = \sum_{n=0}^{20} \Phi(1 - 2n) C_{20}^n \frac{19^{20-n}}{20^{20}}.$$

Pour $n=0$, $\Phi(1) \approx 0,841$, pour $n=1$, $\Phi(-1) \approx 0,159$, pour $n=2$, $\Phi(-3) \approx 0,001$ et pour $n=3$, $\Phi(-5)$ est négligeable. On peut donc négliger les valeurs de $n=3$ à $n=20$, ce qui allège agréablement les calculs, puisqu'il ne reste que trois termes à calculer.

$$P([T < 52]) \approx \Phi(1) \left(\frac{19}{20}\right)^{20} + 20\Phi(-1) \frac{19^{19}}{20^{20}} + 190\Phi(-3) \frac{19^{18}}{20^{20}} \approx 0,37.$$

Exercice 2.13

(1) $X \sim \mathcal{B}(N, \frac{v}{V})$ qu'on approche par la loi normale

$\mathcal{N}\left(\frac{Nv}{V}, \sigma^2 = \frac{v}{V} \left(1 - \frac{v}{V}\right) N\right)$ où $\frac{N}{V}$ est le nombre de molécules par litre.

(2) $\frac{\sigma_X}{E(X)} \approx \frac{1}{\sqrt{N \cdot v/V}}$. Calculons $P\left(\frac{|X - E(X)|}{E(X)} < 0,01\right)$:

$\frac{N}{V} \times v$ est de l'ordre de 10^{23} et $\sigma \approx \sqrt{10^{23}} \approx 3,1 \cdot 10^{11}$, d'où il vient :
 $\left(\frac{X - 10^{23}}{3,1 \cdot 10^{11}}\right) \sim \mathcal{N}(0, 1)$ et la probabilité cherchée vaut $P([-3, 1 \cdot 10^{11} < Y < 3, 1 \cdot 10^{11}]) =$

$1 - \alpha$, avec α de l'ordre de 10^{-100} , d'où, à l'échelle du temps humain, l'impossibilité de s'asphyxier.

Exercice 2.14

- (1) $\int_{\mathbb{R}} f_X(x) dx = \frac{K(x_m, \alpha)}{\alpha x_m^\alpha} = 1$ donc $K(x_m, \alpha) = \alpha x_m^\alpha$.
- (2) $\mathbb{E}(X^n)$ fini si et seulement si $\alpha + 1 - n > 1$, c'est-à-dire $\alpha > n$. D'où :
- $$\mathbb{E}(X^n) = \frac{\alpha x_m^\alpha}{\alpha - n}.$$

Chapitre 3

Vecteurs aléatoires

Les vecteurs aléatoires permettent de modéliser des phénomènes et des systèmes dont la description nécessite plusieurs v.a.

3.1 Un exemple

On lance un projectile ponctuel (idéalisée concevable à l'époque du laser et de la haute résolution) sur une cible circulaire D de rayon 1 centrée en 0, selon une probabilité uniformément répartie sur la cible.

Pour tout borélien B du disque D , $P((X,Y) \in B) = \frac{\text{Aire}(B)}{\pi}$, où (X,Y) sont les coordonnées du point d'impact sur la cible.

La probabilité $P_{X,Y}$ du vecteur aléatoire (X,Y) est définie par la densité constante sur le disque :

$$f_{X,Y}(x,y) = \frac{1}{\pi} \mathbf{1}_D(x,y).$$

Donc, par analogie avec la définition de la densité d'une v.a. absolument continue, on a :

$$P((X,Y) \in B) = \int_B f_{X,Y}(x,y) dx dy = \text{Aire}(B)/\pi.$$

Calculons la probabilité que $X \in [a,b] \subset [-1,1]$.

Soit le borélien B , défini comme intersection de la bande verticale de support $[a,b]$ avec D .

$$\begin{aligned} P(X \in [a,b]) &= P((X,Y) \in B) \\ &= \int_B f_{X,Y}(x,y) dx dy \\ &= \frac{1}{\pi} \int_a^b \left(\int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} dy \right) dx = \int_a^b \frac{2}{\pi} \sqrt{1-x^2} dx. \end{aligned}$$

L'égalité précédente étant valide pour tout intervalle $[a, b] \subset [-1, 1]$, on en déduit que X a pour densité :

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

Une démonstration analogue permettrait d'établir : $f_Y(y) = \frac{2}{\pi} \sqrt{1-y^2} \mathbf{1}_{[-1,1]}(y)$.

3.2 Description des vecteurs aléatoires

3.2.1 Généralités

Dans ce qui suit, on désignera par β_k la tribu des boréliens de \mathbb{R}^k .

Définition 3.1 *Un vecteur aléatoire réel de dimension k est un vecteur dont les composantes sont des v.a. réelles ; il est décrit par une application mesurable $X = (X_1, \dots, X_k)$ définie sur un espace probabilisé (Ω, \mathcal{T}, P) à valeurs dans (\mathbb{R}^k, β_k) . Pour tout événement B appartenant à la tribu β_k , $X^{-1}(B) \in \mathcal{T}$.*

Définition 3.2 *La loi de probabilité P_X dite loi conjointe de X est définie sur (\mathbb{R}^k, β_k) par :*

$$\forall B \in \beta_k, P_X(B) = P((X_1, \dots, X_k) \in B).$$

La fonction de répartition F_X est définie par :

$$\forall (x_1, x_2, \dots, x_k) \in \mathbb{R}^k \quad F_X(x_1, \dots, x_k) = P(X_1 \leq x_1, \dots, X_k \leq x_k).$$

La connaissance des lois marginales des $(X_i)_i$ ne suffit évidemment pas pour déterminer la loi conjointe du vecteur (X_1, X_2, \dots, X_k) et plus généralement celle d'une v.a. $h(X_1, \dots, X_k)$ fonction des composantes $(X_i)_i$.

Définition 3.3 *Indépendance des composantes d'un vecteur aléatoire.*

Les v.a. X_1, X_2, \dots, X_k sont mutuellement indépendantes, si pour tout k -uplet de boréliens (B_1, B_2, \dots, B_k) de \mathbb{R}^k :

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_k) = P(X_1 \in B_1)P(X_2 \in B_2) \dots P(X_n \in B_k).$$

Interprétation : les événements décrits par les v.a. $(X_i)_{i=1, \dots, k}$ sont indépendants entre eux ; autrement dit, la réalisation d'un sous-ensemble quelconque d'entre eux, n'a aucune influence sur la réalisation des autres.

Théorème 3.1 *Si X est un vecteur aléatoire de dimension n de loi P_X et si φ est une application réelle définie sur \mathbb{R}^k telle que $E(|\varphi(X)|) < +\infty$, alors :*

$$E(\varphi(X)) = \int_{\mathbb{R}^k} \varphi(x) dP_X(x).$$

Application : soient le vecteur $X = (X_1, X_2)$ de loi $P_{(X_1, X_2)}$, et φ définie par $\varphi(u, v) = u + v$.

$$\begin{aligned} E(X_1 + X_2) &= \int_{\mathbb{R}^2} (x_1 + x_2) dP_{(X_1, X_2)}(x_1, x_2), \\ &= \int_{\mathbb{R}^2} x_1 dP_{(X_1, X_2)}(x_1, x_2) + \int_{\mathbb{R}^2} x_2 dP_{(X_1, X_2)}(x_1, x_2), \\ &= \int_{\mathbb{R}} x_1 dP_{X_1}(x_1) + \int_{\mathbb{R}} x_2 dP_{X_2}(x_2) = E(X_1) + E(X_2). \end{aligned}$$

Définition 3.4 *L'espérance d'un vecteur aléatoire $X = (X_1, \dots, X_k)$ est définie par le vecteur :*

$$(E(X_1), E(X_2), \dots, E(X_k)) \stackrel{\text{noté}}{=} E(X).$$

Théorème 3.2 *(Caractérisation de l'indépendance des composantes d'un vecteur aléatoire.)*

X_1, X_2, \dots, X_k sont indépendantes si et seulement si pour tout n -uplet $(\varphi_1, \dots, \varphi_k)$ de fonctions bornées de \mathbb{R} dans \mathbb{R} :

$$E(\varphi_1(X_1) \dots \varphi_k(X_k)) = E(\varphi_1(X_1)) \dots E(\varphi_k(X_k)).$$

Théorème 3.3 *(Linéarité de l'espérance mathématique)*

(a) Si X et Y sont des vecteurs aléatoires de même dimension :

$$E(X + Y) = E(X) + E(Y).$$

(b) $\forall a \in \mathbb{R}, E(aX) = aE(X)$.

Définition 3.5 À tout couple de v.a. (X, Y) admettant des moments d'ordre deux, on associe la **covariance** de ses composantes définie par :

$$E(XY) - E(X)E(Y) \stackrel{\text{noté}}{=} \text{Cov}(X, Y).$$

Théorème 3.4 Étant donné un couple de v.a. (X, Y) admettant des moments d'ordre deux, on a :

(a) $V(X + Y) = V(X) + V(Y) + 2(E(XY) - E(X)E(Y))$;

(b) si X, Y sont indépendantes : $\text{Cov}(X, Y) = 0$ et $V(X + Y) = V(X) + V(Y)$.

Preuve

- (a) $V(X+Y) = E((X+Y)^2) - (E(X)+E(Y))^2$ qu'il suffit de développer.
 (b) Appliquons le théorème 3.1 dans le cas où $\varphi(x,y) = xy$:

$$E(XY) = \int_{\mathbb{R}^2} xy dP_{(X,Y)}(x,y).$$

Puis, par indépendance de X et Y ,

$$E(XY) = \int_{\mathbb{R}^2} xy dP_X(x) dP_Y(y) = \int_{\mathbb{R}} x dP_X(x) \int_{\mathbb{R}} y dP_Y(y) = E(X)E(Y)$$

d'où la proposition (b), grâce au développement de $V(X+Y)$. \square

3.1 ♠ Optimisation d'un multisenseur

On estime un paramètre p inconnu à l'aide de deux systèmes de mesure dont les estimations sont données par les v.a. X_1 et X_2 :

$$X_1 = p + U_1 \text{ et } X_2 = p + U_2,$$

où les v.a. U_i représentent des erreurs indépendantes centrées et de variances σ_i^2 . Étant donné l'estimateur de p , noté \hat{p} , de la forme $\hat{p} = aX_1 + bX_2$; déterminer les nombres a et b de manière que $E(\hat{p}) = p$ (\hat{p} est dit sans biais) et que $V(\hat{p}^*)$ soit minimal.

Définition 3.6 Indépendance des vecteurs aléatoires

Les vecteurs aléatoires X_1, X_2, \dots, X_k , de dimensions respectives n_1, \dots, n_k sont indépendants, si pour tout k -uplet (B_1, B_2, \dots, B_k) de boréliens de $\mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_k}$:

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_k \in B_k) = P(X_1 \in B_1)P(X_2 \in B_2) \dots P(X_k \in B_k).$$

L'indépendance de variables ou de vecteurs aléatoires ne se conserve pas nécessairement par transformation fonctionnelle, comme l'illustre l'exemple suivant.

Exemple 3.1 Considérons : $U = X$ et $V = (\text{sign}(X)Y)$, où X et Y sont des v.a. indépendantes de même loi : $\frac{1}{3}\delta_{-1} + \frac{2}{3}\delta_1$.

Démontrons que les v.a. U et V sont dépendantes.

$$\begin{aligned} P(U = 1, V = 1) &= P(X = 1 \text{ et } \text{sign}(X)Y = 1) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9} ; \\ P(U = 1) &= \frac{2}{3}. \end{aligned}$$

$$\begin{aligned} P(V = 1) &= P((\text{sign}(X) = 1 \text{ et } Y = 1) \text{ ou } (\text{sign}(X) = -1 \text{ et } Y = -1)) \\ &= P((\text{sign}(X) = 1 \text{ et } Y = 1)) + P(\text{sign}(X) = -1 \text{ et } Y = -1)) \\ &= \frac{2}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3} = \frac{5}{9} \end{aligned}$$

On en déduit $P(U = 1)P(V = 1) = \frac{10}{27} \neq P(U = 1, V = 1)$ et donc la dépendance de U et V .

3.2.2 Vecteurs discrets

Un vecteur aléatoire X discret de dimension k prend ses valeurs dans un ensemble au plus dénombrable E inclus dans \mathbb{R}^k .

Définition 3.7 La loi P_X , dite loi conjointe du vecteur X , est définie par :

$$\forall x \in E, \quad P_X(x_1, \dots, x_k) \stackrel{\text{déf}}{=} P(X_1 = x_1, \dots, X_k = x_k).$$

Théorème 3.5 La loi P_{X_i} de X_i , dite **loi marginale** de X_i est définie par :

$$P_{X_i}(x_i) = P(X_i = x_i) = \sum_{x_1 \in E_1} \dots \sum_{x_{i-1} \in E_{i-1}} \sum_{x_{i+1} \in E_{i+1}} \dots \sum_{x_k \in E_k} P_X(x_1, x_2, \dots, x_k).$$

Preuve

$$\begin{aligned} P(X_i = x_i) &= P(X_1 \in E_1, \dots, X_i = x_i, \dots, X_k \in E_k) \\ &= \sum_{x_1 \in E_1} \dots \sum_{x_{i-1} \in E_{i-1}} \sum_{x_{i+1} \in E_{i+1}} \dots \sum_{x_k \in E_k} P_X(x_1, x_2, \dots, x_k). \end{aligned}$$

□

3.2 ♠♠ On tire des boules l'une à la suite de l'autre, sans remise, dans une urne contenant 2 boules blanches et N boules noires, jusqu'à ce que sorte la deuxième boule blanche; soient les v.a. X et Y , égales aux numéros de tirage des deux blanches.

(1) Loi et espérance de X ? [*rappel* : $\sum_{k=1}^N k^2 = \frac{N(N+1)(2N+1)}{6}$].

(2) Par un raisonnement simple, déterminer la loi de (X, Y) . Dessiner le support du vecteur aléatoire (X, Y) . Calculer la loi et l'espérance de Y .

On pourra généraliser l'étude précédente au cas de N noires et de B blanches.

Exemple 3.2 Étude du vecteur multinomial

Le vecteur multinomial est l'extension de dimension k de la variable binomiale. Soit l'espace fondamental $\Omega = \{A_1, A_2, \dots, A_k\}$, où chaque événement A_i a la probabilité de réalisation p_i , sachant que $\sum p_i = 1$. On procède à n tirages avec remise dans Ω : le vecteur aléatoire (X_1, \dots, X_k) compte les réalisations respectives (x_1, x_2, \dots, x_k) des événements A_1, \dots, A_k . D'un point de vue pratique, ce modèle probabiliste décrit un sondage avec remise par échantillons de taille n , dans une population partitionnée en k sous-ensembles, décrivant par exemple les classes socio-professionnelles d'un pays ou d'une région. Déterminons : $P(X_1 = x_1, \dots, X_k = x_k)$.

Calculons le nombre d'échantillons réalisant l'événement (x_1 fois A_1 , x_2 fois A_2, \dots, x_k fois A_k). Compte tenu de l'indépendance des tirages, on démontre que le nombre de tels échantillons est égal à :

$$C_n^{x_1} \cdot C_{n-x_1}^{x_2} \cdots C_{n-(x_1+\dots+x_{k-1})}^{x_k} = \frac{n!}{x_1! x_2! \dots x_k!}.$$

Or la probabilité de tirer un échantillon de la forme ci-dessus est égale à $p_1^{x_1} \dots p_k^{x_k}$.

$$\text{D'où : } P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}.$$

On vérifie facilement que X_i est binomiale de loi $\mathcal{B}(n, p_i)$ et que

$$P(X_1 = x_1, \dots, X_k = x_k) \neq P(X_1 = x_1) \dots P(X_k = x_k),$$

pour tout (x_1, \dots, x_k) de somme n , ce qui équivaut à la dépendance des v.a. $(X_i)_{i=1, \dots, k}$.

Définition 3.8 *Le vecteur aléatoire $X = (X_1, \dots, X_k)$ est dit multinomial de paramètres $(p_i)_i$, si pour tout k -uplet (x_1, \dots, x_k) à valeurs dans $\{0, 1, \dots, n\}$ et tels que $\sum_{i=1}^k x_i = n$:*

$$P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k},$$

où les p_i sont des réels positifs de somme égale à un.

Définition 3.9 *Matrice de covariance d'un vecteur aléatoire X*

Covariances des couples de vecteurs (X_i, X_j) :

$$E(X_i X_j) - E(X_i)E(X_j) \stackrel{\text{noté}}{=} \text{Cov}(X_i, X_j).$$

Matrice de covariance du vecteur X , notée $\text{Cov}(X)$ ou Γ_X :

$$\Gamma_X = E(X \cdot {}^t X) - E(X) \cdot {}^t E(X).$$

$$\Gamma_X = \begin{pmatrix} \sigma_1^2 & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_k) \\ \vdots & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \text{Cov}(X_k, X_1) & \cdots & \cdots & \sigma_k^2 \end{pmatrix}.$$

Définition 3.10 *Les composantes X_i et X_j sont dites **décorrélées** si :*

$$\text{Cov}(X_i, X_j) = 0.$$

Théorème 3.6 (Propriétés de l'espérance et de la matrice de covariance des vecteurs aléatoires) Soit X un vecteur aléatoire de dimension k :

P1. La forme quadratique définie par ${}^t x \text{Cov}(X) x$ ($x \in \mathbb{R}^k$) est symétrique, semi-définie positive.

P2. Pour toute application linéaire T de \mathbb{R}^k sur lui-même :

$$E(T(X)) = T(E(X)) ;, \quad \text{Cov}(T(X)) = T \text{Cov}(X) {}^t T.$$

P3. Si X et Y sont des vecteurs de même dimension, a et b des réels quelconques, alors :

$$\text{Cov}(aX + bY) = a^2 \text{Cov}(X) + b^2 \text{Cov}(Y) + 2ab \text{Cov}(X, Y).$$

P4. $V(\sum X_i) = \sum V(X_i) + \sum_i \sum_j \text{cov}(X_i, X_j).$

Si les composantes X_i sont non corrélées deux à deux : $V(\sum_i X_i) = \sum_i V(X_i).$

Preuve

(P1) Symétrie : $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i).$

Semi-définie positivité : pour tout $u = (u_1, u_2, \dots, u_k)$

$$\begin{aligned} {}^t u \text{Cov}(X) u &= \sum_i \sum_j u_i u_j \text{cov}(X_i, X_j) \\ &= E\left(\left(\sum_i u_i (X_i - m_i)\right)\left(\sum_j u_j (X_j - m_j)\right)\right) = E(Y^2), \geq 0 \end{aligned}$$

où $Y = \sum_i u_i (X_i - m_i).$

(P2) Soit $T = (t_{ij})_{i,j}$ alors $TX = (\sum_j t_{ij} X_j)_i :$

$$E(TX) = \left(\sum_j t_{ij} E(X_j)\right)_i = TE(X)$$

$$\begin{aligned} \text{Cov}(TX) &= E(TX \cdot {}^t (TX)) - E(TX) {}^t E(TX) \\ &= E(TX \cdot {}^t X \cdot {}^t T) - TE(X) {}^t E(X) {}^t T = T \text{Cov}(X) {}^t T. \end{aligned}$$

(P3)

$$\begin{aligned} \text{(a)} \quad \text{Cov}(aX) &= a^2 \text{Cov}(X) \\ \text{(b)} \quad \text{Cov}(X + Y) &\stackrel{\text{déf}}{=} E((X + Y) {}^t (X + Y)) - E(X + Y) {}^t E(X + Y) \\ &= \text{Cov}(X) + \text{Cov}(Y) + \\ &\quad E(X {}^t Y) - E(X) {}^t E(Y) + E(Y {}^t X) - E(Y) {}^t E(X) \\ &= \text{Cov}(X) + \text{Cov}(Y) + 2\text{Cov}(X, Y). \end{aligned}$$

(P4) Évident. □

Exemple 3.3 Étude du vecteur multinomial (suite et fin).

X_i est de loi binomiale $\mathcal{B}(np_i, np_i(1-p_i))$, d'où $E(X) = {}^t(np_1, \dots, np_k)$.

Détermination de la matrice de covariance de X :

posons $U_j^{(i)} = \begin{cases} 1 & \text{si le } i\text{-ième tirage fait apparaître } A_j ; \\ 0 & \text{sinon.} \end{cases}$

Il est facile de vérifier que $(X_1, X_2, \dots, X_k) = \sum_{i=1}^n (U_1^{(i)}, \dots, U_k^{(i)})$ où les $U_j^{(i)}$ sont des v.a. indicatrices des événements A_j pour chacun des n tirages ; les vecteurs $U^{(i)} = (U_1^{(i)}, \dots, U_k^{(i)})$ sont indépendants.

D'autre part, $E(U_m^{(i)} U_r^{(i)}) = 0$ si $m \neq r$, d'où :

$$\text{Cov}(U_m^{(i)}, U_r^{(i)}) = -E(U_m^{(i)})E(U_r^{(i)}) = -p_m p_r. \quad (*)$$

On en déduit : $\text{cov}(X_m, X_r) = -np_m p_r$:

$$\text{Cov}(X) = n \begin{pmatrix} p_1(1-p_1) & \dots & -p_1 p_k \\ \vdots & \ddots & \vdots \\ -p_1 p_k & \dots & p_k(1-p_k) \end{pmatrix}.$$

3.2.3 Vecteurs absolument continus

Définition 3.11 La loi P_X du vecteur absolument continu $X = (X_1, X_2, \dots, X_k)$ est définie par sa **densité** $f_X(x_1, x_2, \dots, x_k)$:

$$\forall B \in \beta_k \quad P_X(B) = \int_B f_X(x_1, x_2, \dots, x_k) dx_1 dx_2 \dots dx_k.$$

La densité f_X est une fonction positive de \mathbb{R}^k dans \mathbb{R} , d'intégrale égale à l'unité.

Théorème 3.7 La **densité marginale** f_{X_i} de X_i est définie par :

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{k-1}} f_X(x_1, \dots, x_k) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_k.$$

Théorème 3.8 Soient X un vecteur aléatoire de dimension k , de densité f_X , et une fonction mesurable h de \mathbb{R}^k dans \mathbb{R} telle que $E(h(X))$ existe ;

$$E(h(X)) = \int_{\mathbb{R}^k} h(x) f_X(x) dx.$$

Théorème 3.9 (Dérivation de l'espérance mathématique d'une fonction paramétrée de v.a.)

Soient X un vecteur aléatoire de dimension k , de densité f_X , et g une fonction réelle définie sur $\mathbb{R}^k \times \mathbb{I}$ dérivable par rapport à t variant dans l'intervalle \mathbb{I} , et telle que pour tout $t \in \mathbb{I}$, la fonction $f_X(x)g(x,t)$ soit intégrable par rapport à x , et $\left| \frac{\partial g(x,t)}{\partial t} \right|$ soit bornée par une fonction intégrable par rapport à x , alors :

$$\frac{\partial E(g(X,t))}{\partial t} = E \left(\frac{\partial X(g(X,t))}{\partial t} \right).$$

(Ce résultat est une application directe du théorème classique de dérivation d'une intégrale paramétrée.)

Calcul de la probabilité d'événements de la forme $(X \leq Y)$ ou $(X+Y \leq t)$

$P_{(X,Y)}$ est la loi du couple (X,Y) :

$$(1) P(X \leq Y) = P_{(X,Y)}((x,y) \mid x \leq y) = \int \int_{x \leq y} dP_{(X,Y)}(x,y).$$

Si (X,Y) est de densité $f_{(X,Y)}$: $P(X \leq Y) = \int \int_{x \leq y} f_{(X,Y)}(x,y) dx dy$.

Application : (X,Y) désigne les durées de vie indépendantes de deux composants électroniques, de densités exponentielles $\mathcal{E}(\mu)$ et $\mathcal{E}(\nu)$. Supposons que les deux composants sont mis en service au même instant ; déterminons la probabilité pour que le composant de durée de vie X tombe en panne le premier.

$$\begin{aligned} P(X \leq Y) &= \int \int_{x \leq y} \mu e^{-\mu x} \nu e^{-\nu y} dx dy \\ &= \mu \nu \int_0^{+\infty} (e^{-\nu y} \int_0^y e^{-\mu x} dx) dy = \frac{\mu}{\mu + \nu}. \end{aligned}$$

Par symétrie : $P(Y \leq X) = 1 - P(X < Y) = 1 - \frac{\mu}{\mu + \nu} = \frac{\nu}{\mu + \nu}$.

$$\begin{aligned} (2) F_{X+Y}(t) &= P_{(X,Y)}((x,y) \mid x+y \leq t) \\ &= \int \int_{x+y \leq t} dP_{(X,Y)}(x,y) \\ &= \int \int_{x+y \leq t} f_{(X,Y)}(x,y) dx dy. \end{aligned}$$

Supposons toujours l'indépendance de X et Y :

$$\begin{aligned} \int \int_{x+y \leq t} f_Y(y) f_X(x) dx dy &= \int_{-\infty}^{+\infty} (f_Y(y) \int_{-\infty}^{t-y} f_X(x) dx) dy \\ &= \int_{-\infty}^{+\infty} f_Y(y) F_X(t-y) dy. \end{aligned}$$

Si F_X est dérivable sur son domaine de définition,

$$f_{X+Y}(t) = \frac{d}{dt}(F_{X+Y}(t)) = \frac{d}{dt} \left(\int_{-\infty}^{+\infty} f_Y(y) F_X(t-y) dy \right),$$

égal, par application du théorème de dérivation sous le signe somme, à :

$$\int_{-\infty}^{+\infty} f_Y(y) \frac{d}{dt} F_X(t-y) dy = \int_{-\infty}^{+\infty} f_Y(y) f_X(t-y) dy = f_X * f_Y(t).$$

On vient de démontrer le théorème important suivant, qui fera l'objet d'une démonstration différente dans le chapitre 4 (théorème 4.5).

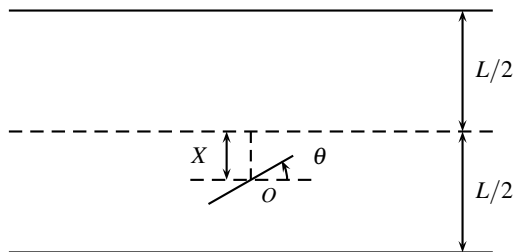
Théorème 3.10 *La densité f_{X+Y} de la somme de deux v.a. indépendantes X, Y est égale au produit de convolution $f_X * f_Y$ des densités des v.a.*

3.3 ♠♠ Problème de Buffon (1707-1788)

Le comte de Buffon (1707-1788), issu d'une famille de noblesse de robe, est plus connu pour ses travaux en sciences de la nature, que pour ceux qu'il a consacrés aux mathématiques. Pourtant il a laissé à la postérité un mémoire intitulé *Sur le jeu du franc-carreau*, dans lequel il calcule les probabilités qu'un écu jeté sur un parquet pavé de carrés égaux, ne coupe pas de pavé ou en coupe un ou plusieurs. La formulation moderne de ce problème est la suivante : on jette au hasard une aiguille de longueur l sur un parquet formé de lames horizontales de largeur L ($l < L$) et de longueur infinie.

Calculer la probabilité pour que l'aiguille soit à cheval sur deux lames.

[Indication : on paramètre l'aiguille par la distance X séparant son milieu O de la droite (en pointillé) partageant la lame de parquet en deux parties égales et par l'angle θ qu'elle fait avec la verticale.]



En inférer une méthode permettant d'obtenir une approximation (lente!) de π , mise en œuvre depuis une cinquantaine d'années au palais de la Découverte à Paris. Le problème dit de Buffon-Synge consiste à déterminer l'espérance du nombre d'intersections d'une corde souple de longueur ℓ jetée sur le plancher.

3.3 Indépendance et corrélation

Théorème 3.11 *Si le vecteur aléatoire $X = (X_1, X_2, \dots, X_k)$ est de densité f_X et les v.a. X_i de densités respectives f_{X_i} , alors :*

$$(\text{Indépendance des } X_i) \iff f_X(x_1, \dots, x_k) = \prod_{i=1}^k f_{X_i}(x_i) \quad \forall (x_1, \dots, x_k).$$

Preuve (\Leftarrow) $P_X(B_1 \times \dots \times B_k) = P(X_1 \in B_1, \dots, X_k \in B_k) = \int_{B_1 \times \dots \times B_k} f_X(x) dx$
 $= \int_{B_1 \times \dots \times B_k} f_{X_1}(x_1) \dots f_{X_k}(x_k) dx_1 \dots dx_k = \int_{B_1} f_{X_1}(x_1) dx_1 \dots \int_{B_k} f_{X_k}(x_k) dx_k.$
 $= P_{X_1}(B_1) \dots P_{X_k}(B_k).$ La réciproque est triviale. \square

3.4 ♠ On vise une cible de rayon R , centrée en 0 : supposons que les coordonnées (X, Y) du point d'impact sont de loi normale $\mathcal{N}(0, \sigma = 0,6R)$, X et Y étant indépendantes, donc de densité $f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{2\pi(0,6R)^2} e^{-\frac{(x^2+y^2)}{2(0,36)R^2}}.$

(1) Probabilité pour que l'impact soit sur la cible ? La calculer pour $R = 1$.

(2) Quel est le nombre minimal de tirs garantissant l'atteinte de la cible au moins une fois, avec 95% de certitude ?

Théorème 3.12 *Soit le vecteur $X = (X_i)_{i=1, \dots, k}$ de composantes indépendantes telles que $E(|X_i|) < +\infty$ pour tout i , alors :*

$$E(|X_1 X_2 \dots X_k|) < +\infty \text{ et } E(X_1 X_2 \dots X_k) = \prod_{i=1}^k E(X_i).$$

Preuve (dans le cas d'un vecteur continu) :

$$\begin{aligned} E(X_1 \dots X_k) &= \int_{\mathbb{R}^k} x_1 \dots x_k f_X(x_1, \dots, x_k) dx_1 \dots dx_k \\ &= \left(\int x_1 f_{X_1}(x_1) dx_1 \right) \dots \left(\int x_k f_{X_k}(x_k) dx_k \right) = \prod_{i=1}^k E(X_i). \end{aligned}$$

\square

Théorème 3.13 *(L'indépendance de deux v.a. entraîne leur décorrélation)*

$$(X_1, X_2 \text{ indépendantes}) \implies (\text{Cov}(X_1, X_2) = 0).$$

La réciproque $(\text{Cov}(X_1, X_2) = 0 \implies X_1, X_2 \text{ indépendantes})$ n'est vraie que si (X_1, X_2) est un vecteur gaussien ou un vecteur de Bernoulli. (On rappelle qu'un vecteur de Bernoulli (X_1, X_2) prend ses valeurs dans l'espace $\{0, 1\}^2$.)

Preuve X_1, X_2 indépendants $\Rightarrow E(X_1 X_2) - E(X_1)E(X_2) = 0$. \square

Exemple 3.4 Variables dépendantes et pourtant décorrélées

Soit Φ une v.a. uniforme sur $[0, 2\pi]$; posons $X = \cos(\Phi)$ et $Y = \sin(\Phi)$.

Le support de (X, Y) , qui est le cercle unité, n'est pas un produit cartésien, d'où la dépendance de X et Y , bien que leur covariance égale à $E(\frac{\sin 2\Phi}{2}) - E(\cos(\Phi)).E(\sin(\Phi))$ soit nulle.

Théorème 3.14 (*Condition nécessaire d'indépendance des composantes*)

$$(\text{Indépendance des } X_i) \implies (\text{support}(X_1, \dots, X_k) = \bigotimes_{i=1}^n \text{supports des } X_i).$$

Remarque 3.1 En pratique, on utilisera souvent ce théorème sous sa forme contraposée comme condition suffisante de dépendance : si le support de (X_1, \dots, X_k) est différent d'un pavé rectangle égal au produit cartésien des supports des X_i alors les composantes X_i sont dépendantes.

Définition 3.12 Le *coefficient de corrélation linéaire* de X_1 et X_2 , noté ρ_{X_1, X_2} , est égal à la covariance des v.a. normalisées $(\frac{X_1}{\sigma_{X_1}}, \frac{X_2}{\sigma_{X_2}})$:

$$\rho_{X_1, X_2} = \text{cov}\left(\frac{X_1}{\sigma_{X_1}}, \frac{X_2}{\sigma_{X_2}}\right) = \frac{\text{cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}}.$$

Théorème 3.15 (*Propriétés du coefficient de corrélation linéaire de deux v.a.*) (1) Pour tout couple de v.a. (X_1, X_2) : $|\rho_{X_1, X_2}| \leq 1$.

(2) $\rho_{X_1, X_2} = \pm 1 \iff$ Il existe une relation affine entre X_1 et X_2 : X_1 et X_2 sont dites colinéaires.

(3) $\rho_{X_1, X_2} = 0 \iff$ Aucune relation affine entre X_1 et X_2 : X_1 et X_2 sont dites décorrélées.

En pratique, ρ_{X_1, X_2} est utilisé comme un indice de dépendance affine de deux v.a.

Preuve

(1) Soit le vecteur centré (Y_1, Y_2) :

$$E((Y_2 - aY_1)^2) = a^2 E(Y_1^2) - 2a E(Y_1 Y_2) + E(Y_2^2) \geq 0, \forall a \in \mathbb{R}.$$

Le discriminant de l'équation de variable a est ≤ 0 , ce qui entraîne l'inégalité de Cauchy-Schwarz :

$$|E(Y_1 Y_2)| \leq \sqrt{E(Y_1^2) E(Y_2^2)}.$$

Appliquons cette inégalité aux v.a. centrées $Y_1 = (X_1 - E(X_1))$ et $Y_2 = (X_2 - E(X_2))$:

$$|\rho_{X_1, X_2}| = \left| \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}} \right| \leq 1.$$

(2) Remarquons que (*) est une égalité dans le seul cas où $E((Y_2 - aY_1)^2) = 0$, équivalent à $(Y_2 = aY_1$ presque sûrement).

Dans ce dernier cas, $\left| \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}} \right| = 1$ et $\frac{X_2 - E(X_2)}{\sigma_{X_2}} = \pm 1 \cdot \frac{X_1 - E(X_1)}{\sigma_{X_1}}$. \square

Remarque 3.2 Deux v.a. U et V , toutes deux fonctions du même vecteur de composantes indépendantes (X_1, \dots, X_k) ne sont pas nécessairement dépendantes.

C'est par exemple le cas où X_1 et X_2 sont des v.a. indépendantes uniformément distribuées sur $[0, 1]$; on établira (voir chapitre suivant) que :

$U = (-2 \ln X_1)^{\frac{1}{2}} \cos(2\pi X_2)$ et $V = (-2 \ln X_1)^{\frac{1}{2}} \sin(2\pi X_2)$ sont des variables gaussiennes $\mathcal{N}(0, 1)$ indépendantes utiles pour générer des v.a. gaussiennes.

3.5 ♠♠ Formules classiques des erreurs relatives

X et Y sont des v.a. réelles d'espérances et de variances connues ; soit $Z = f(X, Y)$ où $f \in \mathcal{C}^2(\mathbb{R}^2, \mathbb{R})$. Il s'agit de trouver une méthode d'approximation de $E(f(X, Y))$ et de $V(f(X, Y))$. On rappelle que $f(x, y)$ admet le développement à l'ordre deux au voisinage du point (x_0, y_0) , de la forme :

$$\begin{aligned} & f(x_0, y_0) + (x - x_0)f'_x(x_0, y_0) + (y - y_0)f'_y(x_0, y_0) \\ & + \frac{1}{2}(x - x_0)^2 f''_{xx}(x_0, y_0) + \frac{1}{2}(y - y_0)^2 f''_{yy}(x_0, y_0) + (x - x_0)(y - y_0)f''_{xy}(x_0, y_0). \end{aligned}$$

(1) Déterminer une approximation de $E(Z)$. L'appliquer au cas $Z = XY$.

(2) Calculer $E(X^2 Y^2)$, puis $V(XY)$; dans le cas où X et Y sont indépendantes, donner une approximation de $\frac{\sigma_{XY}}{E(XY)}$, si $\frac{\sigma_X}{E(X)}$ et $\frac{\sigma_Y}{E(Y)}$ sont $\ll 1$, et retrouver l'expression classique utilisée par les physiciens, de l'erreur relative d'un produit de valeurs aléatoires.

3.4 Fonctions génératrice et caractéristique

Ces deux concepts sont d'une grande efficacité opératoire lorsqu'il s'agit de déterminer les lois de sommes ou de différences de variables aléatoires indépendantes, ou des limites de suites de variables ou de vecteurs aléatoires.

3.4.1 Fonction génératrice

Définition 3.13 La fonction génératrice d'un vecteur aléatoire (X_1, \dots, X_k) à valeurs entières positives est définie par :

$$\forall i, \quad \forall z_i \in D(0, 1), \quad g_{(X_1, \dots, X_k)}(z_1, \dots, z_k) = E(z_1^{X_1} z_2^{X_2} \dots z_k^{X_k}).$$

On démontre facilement des propriétés analogues à celles de la fonction génératrice d'une v.a. ; par exemple $E(X_i) = \frac{\partial g_{(X_1, \dots, X_k)}(z_1, \dots, z_k)}{\partial z_i} \Big|_{(z_1=z_2=\dots=z_k=1)}$.

3.4.2 Fonction caractéristique

La fonction caractéristique d'une variable ou d'un vecteur aléatoire quelconque X est la transformation de Fourier (à un coefficient multiplicatif près) de sa mesure de probabilité, qui existe toujours et bénéficie de ses puissantes propriétés. À l'instar de la fonction génératrice, elle caractérise complètement la loi de probabilité et en contient donc toute l'information.

Définition 3.14 La *fonction caractéristique* d'un vecteur aléatoire $X = (X_1, \dots, X_k)$ de loi P_X est définie par :

$$\forall t \in \mathbb{R}^k, \quad E(e^{i\langle X, t \rangle}) = \int_{\mathbb{R}^k} e^{i\langle x, t \rangle} dP_X(x_1, \dots, x_k) = (\text{notée}) \quad \Phi_X(t_1, \dots, t_k).$$

($\langle X, t \rangle$ désigne le produit scalaire $\sum_j t_j X_j$). Si X est une v.a. discrète de loi $P(X = x_j) = p_j$, $\forall j \in \mathbb{J}$, alors :

$$\forall t \in \mathbb{R}, \quad \Phi_X(t) = \sum_{j \in \mathbb{J}} e^{itx_j} p_j.$$

Si X est une v.a. de densité f_X alors :

$$\forall t \in \mathbb{R}, \quad \Phi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(x) dx.$$

Si X est un vecteur aléatoire de densité $f_X(x_1, \dots, x_k)$, sa fonction caractéristique est définie par :

$$\Phi_X(t_1, \dots, t_k) = \int_{\mathbb{R}^k} e^{i\sum_{j=1}^k t_j x_j} f_X(x_1, \dots, x_k) dx_1 \dots dx_k.$$

Exemple 3.5

(1) X_1 variable de Bernoulli, de loi de probabilité $P_{X_1} = p\delta_1 + (1-p)\delta_0$:

$$\Phi_{X_1}(t) = e^{i \cdot 1 \cdot t} p + e^{i \cdot 0 \cdot t} (1-p) = (1-p) + p e^{it}.$$

(2) X_2 variable $\mathcal{N}(0, 1)$: $\Phi_{X_2}(t) = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2} dx$ (intégrale complexe que l'on calcule sur un contour rectangle standard centré en 0), est égale à :

$$\Phi_{X_2}(t) = e^{-\frac{t^2}{2}}.$$

Comme son nom l'indique, la fonction caractéristique d'une variable ou d'un vecteur aléatoire caractérise sa loi.

Théorème 3.16 *Étant donné deux vecteurs aléatoires X et Y , de même dimension et de fonctions caractéristiques Φ_X et Φ_Y :*

$$(\Phi_X = \Phi_Y) \iff (X \text{ et } Y \text{ ont même loi}).$$

Preuve Conséquence de la bijectivité de la transformation de Fourier. □

Théorème 3.17 *La fonction caractéristique d'une somme de v.a. indépendantes X_1, X_2, \dots, X_k de lois quelconques est égale au produit des fonctions caractéristiques :*

$$\forall t \in \mathbb{R}, \quad \Phi_{(\sum_{j=1}^k X_j)}(t) = \prod_{j=1}^k \Phi_{X_j}(t).$$

Preuve $\Phi_{\sum_j X_j}(t) = E[e^{i \sum_j X_j t}] = \prod_j E[e^{i X_j t}] = \prod_{j=1}^k \Phi_{X_j}(t)$. □

3.6 ♠ Appliquer le théorème précédent pour montrer que la somme de n variables de Bernoulli $(B_j)_{j=1, \dots, n}$, indépendantes de paramètre p , est une v.a. binomiale $\mathcal{B}(n, p)$.

Théorème 3.18 *(Moments d'une v.a.)*

Si $E(|X|^n)$ existe et est fini, le moment d'ordre n est égal à :

$$E(X^n) = \frac{1}{i^n} \frac{d^n \Phi_X(t)}{dt^n} \Big|_{(t=0)}.$$

Preuve Le théorème de dérivation de l'espérance implique que $\frac{d\Phi_X(t)}{dt} = E\left(\frac{d}{dt}(e^{itX})\right) = iE(X e^{itX})$, d'où $\frac{1}{i} \frac{d\Phi_X(t)}{dt} \Big|_{t=0} = \frac{iE(X)}{i} = E(X)$.

Démonstration analogue pour le calcul de $E(X^n)$. □

3.7 ♠ Redémontrer grâce à la fonction caractéristique que :

$$X^* \text{ est } \mathcal{N}(0, 1) \iff X = \sigma X^* + m \text{ est } \mathcal{N}(m, \sigma^2).$$

Pour déterminer la loi d'une v.a. $Y = h(X_1, \dots, X_n)$, on verra dans le chapitre suivant consacré au calcul de lois qu'il peut être utile de déterminer la fonction caractéristique $\Phi_Y(t)$, puis d'en calculer l'inverse de Fourier, grâce au théorème ci-dessous, pour en déduire la loi P_Y .

Théorème 3.19 (Inversion de la fonction caractéristique)

(1) Soit la v.a. X de densité f_X et telle $\Phi_X \in L^1(\mathbb{R})$, alors :

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ixt} \Phi_X(t) dt \quad (p.p.).$$

(2) Soit le vecteur X de densité f_X tel que $\Phi_X \in L^1(\mathbb{R}^k)$, alors :

$$f_X(x_1, \dots, x_k) = \frac{1}{(2\pi)^k} \int_{\mathbb{R}^k} \Phi_X(t_1, \dots, t_k) e^{-i(t_1 x_1 + \dots + t_k x_k)} dt_1 \dots dt_k \quad (p.p.).$$

(3) Si X est à valeurs discrètes x_k :

$$P[X = x_k] = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T e^{-itx_k} \Phi_X(t) dt.$$

Exemple 3.6 Soit X , la v.a. de Bernoulli telle que $P(X = 0) = 1 - p$ et $P(X = 1) = p$: $\Phi_X(t) = (1 - p) + pe^{it}$.

Retrouvons, grâce au théorème 3.19, la valeur de $P(X = 0)$:

$$\begin{aligned} P(X = 0) &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T e^{-it \cdot 0} ((1 - p) + pe^{it}) dt \\ &= \lim_{T \rightarrow +\infty} \frac{1}{2T} [(1 - p)t + \frac{p}{i} e^{it}]_{-T}^T = \lim_{T \rightarrow +\infty} ((1 - p) + p \frac{\sin T}{T}) = 1 - p. \end{aligned}$$

3.5 Le vecteur gaussien

L'étude des vecteurs gaussiens dont on a déjà souligné l'importance sera complétée à la fin du chapitre 6.

Définition 3.15 Un vecteur aléatoire $X = (X_1, \dots, X_k)$ est **gaussien ou normal** si toute combinaison linéaire $\sum_{i=1}^k a_i X_i$ ($a_i \in \mathbb{R}$) de ses composantes est une variable gaussienne.

Théorème 3.20 Les composantes d'un vecteur gaussien sont des v.a. gaussiennes.

Preuve Pour tout i , on choisit $a_i = 1$ dans la définition 3.15 et pour tout $j \neq i$ $a_j = 0$: on conclut à la normalité de X_i grâce à la définition. \square

La définition précédente est fondée sur le théorème de Cramer-Wold qui affirme que la loi d'un vecteur aléatoire quelconque (X_1, X_2, \dots, X_n) est caractérisée par la connaissance des lois de toutes les combinaisons linéaires possibles de ses composantes.

Théorème 3.21 (Fonction caractéristique du vecteur gaussien)

La fonction caractéristique du vecteur gaussien X^* centré réduit de dimension k est définie par :

$$\Phi_{X^*}(u_1, u_2, \dots, u_k) = \exp\left(-\frac{1}{2} \sum_{j=1}^k u_j^2\right).$$

La fonction caractéristique du vecteur aléatoire gaussien $\mathcal{N}(m, \Gamma_X)$, est définie par :

$$\Phi_X(u_1, u_2, \dots, u_k) = \exp\left(i \sum_j u_j m_j - \frac{1}{2} u^t \Gamma u\right) \text{ où } u = (u_1, \dots, u_k) \in \mathbb{R}^k.$$

Théorème 3.22 Si les v.a. indépendantes $(X_i)_{i=1, \dots, k}$ sont gaussiennes, alors le vecteur $X = (X_1, X_2, \dots, X_k)$ est gaussien.

Preuve $\forall a_1, \dots, a_k \in \mathbb{R}$, les v.a. $a_1 X_1, \dots, a_k X_k$ sont gaussiennes indépendantes, donc leur somme est gaussienne et, d'après la définition 3.15 le vecteur (X_1, \dots, X_k) est gaussien. \square

Attention ! Sauf si elles sont indépendantes, la normalité de chacune des composantes n'implique pas celle du vecteur, comme le montre l'exemple suivant.

Exemple 3.7 Soit le vecteur (X, Y) de dimension deux, de densité $f_{X,Y}$ égale à $\frac{1}{\pi} e^{-\frac{x^2+y^2}{2}}$ sur le premier et le troisième quadrant déterminé par un repère orthonormé, et nulle sur les deux autres quadrants. Par construction le vecteur (X, Y) n'est pas gaussien, mais ses densités marginales le sont, respectivement égales à $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ et $\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$. Le produit des densités marginales n'étant pas égal à la densité conjointe, les composantes sont dépendantes.

Théorème 3.23 Il existe deux classes de vecteurs gaussiens selon que leur matrice de covariance Γ_X est ou n'est pas une matrice régulière.

(1) Si Γ_X est régulière, X admet une densité f_X de support \mathbb{R}^k :

$$f_X(x) = \frac{1}{(\sqrt{2\pi})^k \sqrt{|\det(\Gamma_X)|}} \exp\left(-\frac{1}{2} (x - m_X)^t \Gamma_X^{-1} (x - m_X)\right)$$

$$\forall i \quad f_{X_i}(x_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x_i - m_i)^2}{2\sigma_i^2}}.$$

(2) Si Γ_X est singulière, de rang r strictement inférieur à n , le vecteur aléatoire X de loi $\mathcal{N}(m_X, \Gamma_X)$ prend ses valeurs dans le sous-espace affine $m_X \oplus$

F_X support de la densité de X , où F_X est le sous-espace vectoriel de dimension r , engendré par les vecteurs lignes de Γ_X . Dans ce cas, le vecteur X est dit dégénéré.

Théorème 3.24 *Étant donné un vecteur gaussien :*

(Indépendance des composantes) \iff (Décorrélacion des composantes).

Preuve (Indépendance des X_i) $\iff \forall x \in \mathbb{R}^k, f_X(x) = \prod f_{X_i}(x_i)$

$\iff f_X(x)$ de la forme : $\text{cte.} \exp(-\frac{1}{2} \sum_i \frac{(x_i - m_i)^2}{\sigma_i^2})$

$\iff \Gamma_X$ diagonale $\iff (X_i)_i$ décorrélées. □

Exemple 3.8 Description du vecteur gaussien à deux composantes $X = (X_1, X_2)$ de vecteur espérance ${}^t m = (m_1, m_2)$ et de matrice de covariance régulière Γ_X :

$$\Gamma_X = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}.$$

$$f_X(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2} Q(x_1, x_2) \right\}.$$

où $Q(x_1, x_2)$ est la forme quadratique définie si $\rho \neq \pm 1$ par :

$${}^t(x - m) \Gamma_X^{-1} (x - m) = \frac{1}{(1 - \rho^2)} \left(\left(\frac{x_1 - m_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - m_2}{\sigma_2} \right)^2 - 2\rho \frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_1 \sigma_2} \right).$$

Si $\rho = +1$, alors Γ_X est une matrice singulière et la distribution de probabilité est portée par la droite d'équation : $y = \frac{\sigma_2}{\sigma_1}(x - m_1) + m_2$. Si $\rho = -1$, la distribution de probabilité est portée par la droite d'équation $y = -\frac{\sigma_2}{\sigma_1}(x - m_1) + m_2$.

3.8 ♠ Soit $X = (X_1, X_2)$ de matrice de covariance $\Gamma_X = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$ et $m_X = (1, 2)$.

Déterminer le coefficient de corrélation $\rho_{(X_1, X_2)}$ et la densité $f_{(X_1, X_2)}$.

Définition 3.16 Un vecteur gaussien X^* est **centré réduit**, si $E(X^*) = 0$ et si $\Gamma_{X^*} = \mathbf{I}_d$.

Les composantes X_i^* sont indépendantes et pour tout couple (i, j) :

$$V(X_i) = 1, \quad m_i = 0 \quad \text{et} \quad \text{Cov}(X_i, X_j) = 0.$$

La densité f_{X^*} est définie par :

$$\forall x \in \mathbb{R}^k, \quad f_{X^*}(x) = (2\pi)^{-\frac{k}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^k x_i^2\right).$$

Attention ! Contrairement aux v.a. (voir théorème 3.24), deux vecteurs X, Y normaux peuvent être décorrélés et être dépendants. Pour que leur décorrélation entraîne leur indépendance, ils doivent être normaux dans leur ensemble, au sens où le vecteur (X, Y) obtenu par concaténation des composantes de X et Y est normal.

Définition 3.17 *L'ellipse de dispersion ou d'équidensité d'un vecteur gaussien, associée à un réel positif α , est l'ensemble des points (x_1, \dots, x_n) tels que :*

$$f_X(x_1, \dots, x_n) = \alpha.$$

Considérons le cas d'un vecteur gaussien (X_1, X_2) : $f_X(x_1, x_2) = \alpha$ équivaut à l'équation $Q(x_1, x_2) = Cte$, laquelle décrit une ellipse centrée en (m_1, m_2) .

Les ellipses de dispersion sont analogues aux géodésiques des surfaces définies dans l'espace euclidien de dimension 3 (voir fig 3.1).

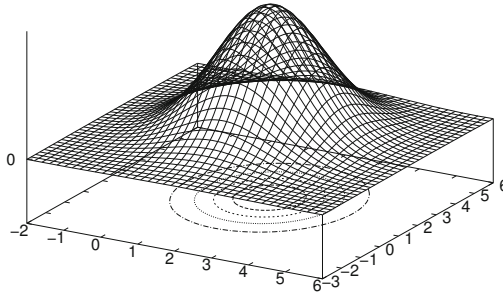


Figure 3.1 – Graphe d'une densité normale de dimension deux.

Théorème 3.25 (Transformation affine d'un vecteur gaussien)

Étant donné :

- un vecteur gaussien X de dimension k de loi $\mathcal{N}(m_X, \Gamma_X)$;
- une matrice A de dimension (n, k) ;
- un vecteur constant B de dimension n ;

$$Y = AX + B \text{ est de loi } \mathcal{N}(Am_X + B, A\Gamma_X^t A).$$

3.9 ♠♠ (1) Démontrer le théorème précédent.

(2) Application : soit $X = (X_1, X_2)$ le vecteur gaussien d'espérance nulle et de matrice de covariance $\Gamma_X = \begin{pmatrix} 1 & 3 \\ 3 & 4 \end{pmatrix}$ et $A = \begin{pmatrix} a & -1 \\ 2 & 1 \end{pmatrix}$.

Pour quelle valeur de a , le vecteur $Y = AX$ a-t-il des composantes indépendantes ?

Théorème 3.26 *Tout vecteur gaussien non dégénéré peut être transformé grâce à une transformation affine adéquate, en un vecteur gaussien non dégénéré à composantes indépendantes.*

Le théorème suivant montre que l'on peut toujours exprimer un vecteur gaussien quelconque X comme transformée affine du vecteur gaussien standard X^* .

Théorème 3.27 *Si X est un vecteur gaussien de dimension n et de loi $\mathcal{N}(m_X, \Gamma_X)$, alors il existe une matrice B de dimension $n \times k$, de rang k , telle que :*

$$X = m_X + BX^*,$$

où X^* est le vecteur gaussien centré réduit.

Preuve Fondée sur le résultat classique d'algèbre linéaire : toute matrice $n \times n$, symétrique et de rang k , peut s'exprimer sous la forme du produit d'une matrice $n \times k$, de rang k , par sa transposée. On peut déterminer la matrice B sous la forme triangulaire supérieure, à l'aide de l'algorithme de Cholesky. \square

Théorème 3.28 *Pour tout vecteur gaussien $\mathcal{N}(m_X, \Gamma_X)$, il existe une isométrie $X \rightarrow Y = AX + B$ où A est une matrice orthogonale telle que Y soit de loi $\mathcal{N}(0, \Gamma_Y)$ où Γ_Y est diagonale.*

3.10 ♠♠ Caractérisation géométrique des vecteurs gaussiens centrés

Soit $X = (X_1, X_2)$ un vecteur aléatoire dans \mathbb{R}^2 tel que X_1 et X_2 soient indépendantes de densités respectives f_1 et f_2 . Démontrer que si la densité f_X est invariante par rotation, les v.a. X_1 et X_2 sont gaussiennes. (On suppose f_1 et f_2 dérivables, et on construit deux équations différentielles fonction de f_1, f_1', f_2, f_2' , que l'on résoudra.)

Forme la plus générale d'une loi de probabilité d'un vecteur aléatoire ou d'une v.a.

La réponse est donnée par un théorème établissant que la probabilité P est égale à la somme d'une mesure discrète, d'une mesure absolument continue

et d'une mesure diffuse. Une mesure diffuse dans l'espace \mathbb{R}^k est une mesure de support non dénombrable dont la mesure de Lebesgue est nulle. D'un point de vue pratique, les probabilités définies par des mesures diffuses apparaissent rarement.

Considérons par exemple dans \mathbb{R}^2 la mesure discrète μ_1 définie par $\sum_{i \in \mathbb{N}} p_i \delta_{x_i}$ de masse π_1 , la mesure absolument continue μ_2 définie par $\mu_2(B) = \int_B f(x) dx$ pour tout borélien B de \mathbb{R}^2 , tel que $\mu_2(\mathbb{R}) = \pi_2$, et la mesure diffuse μ_3 associée à un vecteur normal dégénéré dont le support est une droite et dont la masse est égale à π_3 , telle que $\pi_1 + \pi_2 + \pi_3 = 1$. Alors $P = \mu_1 + \mu_2 + \mu_3$ définit une probabilité sur \mathbb{R}^2 .

Les problèmes 5 et 6 se rapportent à ce chapitre.

3.6 Corrigés des exercices

Exercice 3.1

$\mathbb{E}(\hat{p}) = a\mathbb{E}(X_1) + b\mathbb{E}(X_2) = ap + bp = (a+b)p$; si \hat{p} est sans biais, alors $a+b=1$ ($a, b > 0$).

$$V(\hat{p}) = a^2V(X_1) + b^2V(X_2) = a^2\sigma_1^2 + b^2\sigma_2^2.$$

Il s'agit donc de minimiser $\phi(a, b) = a^2\sigma_1^2 + b^2\sigma_2^2$ sous la contrainte $a+b=1$. On obtient $a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$, $b = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$. Si, par exemple, $\sigma_1^2 > \sigma_2^2$, alors $a < b$, résultat intuitivement prévisible.

Exercice 3.2

(1) X est le numéro de la première boule blanche apparue, donc $X(\Omega) = \{1, 2, \dots, N\}$.

$[X = k]$ est réalisé si la première boule blanche apparaît au rang k et la deuxième à l'un des rangs suivants : il y a donc $N+2-k$ choix équiprobables. Ainsi,

$$P([X = k]) = \frac{2(N+2-k)}{(N+1)(N+2)}.$$

On aurait pu montrer que : $P([X = k]) = \frac{N}{N+2} \cdot \frac{N-1}{N+1} \cdot \frac{N-2}{N} \cdots \frac{N-k+2}{N-k} \cdot \frac{2}{N-k-1}$

$$\begin{aligned} E(X) &= \sum_{k=1}^{N+1} kP([X = k]) = \frac{1}{C_{N+2}^2} \left[(N+2) \sum_{k=1}^{N+1} k - \sum_{k=1}^{N+1} k^2 \right] \\ &= \frac{1}{C_{N+2}^2} \left[(N+2) \frac{(N+1)(N+2)}{2} - \frac{(N+1)(N+2)(2N+3)}{6} \right] \\ &= N+2 - \frac{2N+3}{3} = \frac{3N+6-2N-3}{3} = \frac{N+3}{3}. \end{aligned}$$

(2) Y est le numéro de la deuxième boule blanche apparue, donc $Y(\Omega) = \{2, \dots, N+2\}$.

$[Y = k]$ est réalisé si la deuxième boule blanche apparaît au rang k et la première à l'un des rangs précédents : il y a donc $k-1$ choix équiprobables. Ainsi,

$$P([Y = k]) = \frac{2(k-1)}{(N+1)(N+2)}.$$

$P[(X, Y) = (i, j)] = \frac{2}{(N+1)(N+2)}$ si $1 \leq i < j \leq N+2$ et 0 sinon. On retrouve ainsi :

$$P([Y = j]) = \sum_{i=1}^{j-1} P([(X, Y) = (i, j)]) = \frac{2(j-1)}{(N+1)(N+2)}.$$

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{k=2}^{N+2} k P([Y = k]) = \frac{1}{C_{N+2}^2} \left[\sum_{k=2}^{N+2} k(k-1) \right] = \frac{1}{C_{N+2}^2} \left[\sum_{j=1}^{N+1} j(j+1) \right] \\ &= \frac{1}{C_{N+2}^2} \left[\frac{(N+1)(N+2)(2N+3)}{6} + \frac{(N+1)(N+2)}{2} \right] \\ &= \frac{2N+3}{3} + 1 = \frac{2N+3+3}{3} = \frac{2N+6}{3}. \end{aligned}$$

Finalement, $\mathbb{E}(Y) = \frac{2}{3}(N+3) = 2\mathbb{E}(X)$.

Résultat en accord avec l'intuition : en effet, si N est très grand, X et Y sont respectivement situés dans le voisinage de $\frac{N}{3}$ et $\frac{2N}{3}$.

Exercice 3.3

On exprime l'événement « l'aiguille coupe une parallèle » par l'inégalité :

$$\frac{X}{\cos \Theta} < \frac{\ell}{2} \text{ où } X \in [0, \frac{\ell}{2}] \text{ et } \Theta \in [0, \frac{\pi}{2}].$$

X et Θ sont indépendantes et de lois uniformes sur leurs supports. Il vient donc :

$$P([X < \frac{\ell}{2} \cos \Theta]) = \int \int_{x < \frac{\ell}{2} \cos \theta} f_X(x) f_\Theta(\theta) dx d\theta = \frac{4}{\pi L} \int_0^{\pi/2} \int_0^{\ell/2 \cos \theta} dx d\theta = \frac{2\ell}{\pi L}.$$

Après n jets, on compte le nombre N de fois où l'aiguille coupe une parallèle : on démontre, grâce à la loi des grands nombres (voir chapitre 5), que $\frac{N}{n}$ converge vers $\frac{2\ell}{\pi L}$ d'où une estimation de π égale à $\frac{2\ell n}{NL}$.

Exercice 3.4 (1)

$$\begin{aligned} P((X, Y) \in \text{cible}) &= \int_0^{2\pi} \int_0^R \frac{1}{2\pi \times 0,36R^2} e^{-\frac{r^2}{0,72 \times R^2}} dr d\theta \\ &= \frac{1}{0,36R^2} \int_0^R e^{-\frac{r^2}{0,72 \times R^2}} dr = \frac{0,6R}{0,36R^2} \int_0^{1/0,6} e^{-\frac{1}{2}u^2} du \\ &= \frac{1}{0,6R} P(0 \leq \mathcal{N}(0, 1) \leq 1,66) = \frac{0,452}{0,6R}, \end{aligned}$$

égal à 0,753 si $R = 1$.

(2) $P(\text{atteindre la cible au moins une fois en } n \text{ tirs})$
 $= 1 - P(\text{ne pas l'atteindre en } n \text{ tirs}) = 1 - (1 - 0,753)^n$; donc $1 - (0,247)^n = 0,95$.
 n est le premier entier immédiatement supérieur ou égal à $\frac{2,995}{1,396}$, soit $n = 3$.

Exercice 3.5

(1) Remplaçons (x, y) par le couple (X, Y) et (x_0, y_0) par $(\mathbb{E}(X), \mathbb{E}(Y))$.
 Sachant que $\mathbb{E}(X - \mathbb{E}(X)) = \mathbb{E}(Y - \mathbb{E}(Y)) = 0$:

$$\begin{aligned} \mathbb{E}(Z) = \mathbb{E}(f(X, Y)) &\approx f(\mathbb{E}(X), \mathbb{E}(Y)) + \frac{1}{2}V(X)f''_{xx}(\mathbb{E}(X), \mathbb{E}(Y)) + \frac{1}{2}V(Y)f''_{yy}(\mathbb{E}(X), \mathbb{E}(Y)) \\ &\quad + \text{Cov}(X, Y)f''_{xy}(\mathbb{E}(X), \mathbb{E}(Y)). \end{aligned}$$

On vérifie que $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) + \text{cov}(X, Y)$.

(2) $V(XY) = \mathbb{E}(X^2Y^2) - (\mathbb{E}(XY))^2 = \mathbb{E}(X^2)\mathbb{E}(Y^2) + \text{cov}(X^2, Y^2) - (\mathbb{E}(XY))^2$.

Si X et Y sont indépendantes, $V(XY) = \mathbb{E}(X^2)\mathbb{E}(Y^2) - (\mathbb{E}(XY))^2$.

$$\frac{V(XY)}{\mathbb{E}(XY)^2} = \frac{\mathbb{E}(X^2)\mathbb{E}(Y^2)}{\mathbb{E}(X)^2\mathbb{E}(Y)^2} - 1 = \left(\frac{V(X)}{\mathbb{E}(X)^2} + 1\right)\left(\frac{V(Y)}{\mathbb{E}(Y)^2} + 1\right) - 1 \approx \frac{V(X)}{\mathbb{E}(X)^2} + \frac{V(Y)}{\mathbb{E}(Y)^2}.$$

Donc : $\frac{V(XY)}{\mathbb{E}(XY)^2} \approx \frac{V(X)}{\mathbb{E}(X)^2} + \frac{V(Y)}{\mathbb{E}(Y)^2} \approx \left(\sqrt{\frac{V(X)}{\mathbb{E}(X)^2}} + \sqrt{\frac{V(Y)}{\mathbb{E}(Y)^2}}\right)^2$ d'où l'approximation classique de l'erreur relative d'un produit $X \cdot Y$:

$$\frac{\sigma_{X \cdot Y}}{\mathbb{E}(XY)} \approx \frac{\sigma_X}{\mathbb{E}(X)} + \frac{\sigma_Y}{\mathbb{E}(Y)}.$$

Exercice 3.6

Soit B une v.a. de loi de Bernoulli : $\Phi_B(t) = pe^{it} + 1 - p$.

$$\Phi_{\sum_{j=1}^n B_j}(t) = \prod_{j=1}^n \Phi_{B_j}(t) = (pe^{it} + 1 - p)^n = \sum_{k=0}^n e^{itk} \binom{n}{k} p^k (1-p)^{n-k} \text{ qui est la}$$

fonction caractéristique d'une v.a. de loi binomiale $\mathcal{B}(n, p)$.

Exercice 3.7

$$\Phi_X(t) = \mathbb{E}(e^{itX}) = \mathbb{E}(e^{it(\sigma X^* + m)}) = e^{itm} \mathbb{E}(e^{it\sigma X^*}) = e^{itm} \Phi_{X^*}(\sigma t) = e^{itm} e^{-\frac{\sigma^2 t^2}{2}}.$$

Exercice 3.8

$$\rho_{X_1, X_2} = \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}} = \frac{1}{\sqrt{6}} \cdot f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{5}} \exp\left(-\frac{1}{2}Q(x_1, x_2)\right) \text{ où}$$

$$Q(x_1, x_2) = (x_1 - 1, x_2 - 2)\Gamma_X^{-1}(x_1 - 1, x_2 - 2) \text{ et } \Gamma_X^{-1} = \frac{1}{5} \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}.$$

Exercice 3.9

(1) Y vérifie la caractérisation du vecteur gaussien.

$$\mathbb{E}(AX + B) = A\mathbb{E}(X) + B \text{ (linéarité de l'espérance).}$$

$$\text{Cov}(AX + B) = \text{Cov}(AX) = A \cdot \text{Cov}(X) \cdot {}^t A.$$

(2) (Y_1, Y_2) indépendantes $\iff A\Gamma_X {}^tA$ diagonale $\iff a = -1$.

Exercice 3.10

$f_{X_1, X_2}(x_1, x_2) = f_{X_1, X_2}(r(x_1, x_2)) = g(\rho)$ où $\rho = \sqrt{x_1^2 + x_2^2}$. De l'indépendance de X_1 et X_2 , on déduit :

$$g(\rho) = f_{X_1}(x_1)f_{X_2}(x_2) = f_1(x_1)f_2(x_2).$$

En dérivant par rapport à x_1 , on obtient : $f'_1(x_1)f_2(x_2) = g'(\rho) \times \frac{x_1}{\rho}$.

De même, en dérivant par rapport à x_2 , on obtient : $f_1(x_1)f'_2(x_2) = g'(\rho) \times \frac{x_2}{\rho}$.

On a donc, pour $x_1x_2 \neq 0$, $\frac{f'_1(x_1)}{x_1} \times f_2(x_2) = f_1(x_1) \times \frac{f'_2(x_2)}{x_2}$, soit

$$\frac{f'_1(x_1)}{x_1} \times \frac{1}{f_1(x_1)} = \frac{f'_2(x_2)}{x_2} \times \frac{1}{f_2(x_2)} = \lambda = \text{constante}.$$

D'où $\frac{f'_i(x_i)}{x_i} \times \frac{1}{f_i(x_i)} = \lambda$, soit $f'_i(x_i) = \lambda x_i f_i(x_i)$.

On est donc ramené à résoudre l'équation différentielle $y' - \lambda xy = 0$ qui donne, pour $y > 0$, $\ln y = \frac{\lambda}{2}x^2 + \mu$, soit $f_i(x) = C_i e^{\frac{\lambda}{2}x^2}$.

Pour des raisons de convergence, on a nécessairement $\lambda < 0$. On pose $\lambda = -\frac{1}{\sigma^2}$ et comme $\int f_i(x) dx = 1$, on a alors $C_i \int e^{-\frac{x^2}{2\sigma^2}} dx = C_i \sigma \sqrt{2\pi} = 1$, soit $C_i = \frac{1}{\sigma \sqrt{2\pi}}$. Finalement, $f_1(x) = f_2(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$: X_1 et X_2 suivent une même loi normale centrée.

Chapitre 4

Calcul de lois

Dans ce chapitre, on présente les méthodes classiques permettant de déterminer les lois de variables ou de vecteurs aléatoires obtenus comme somme, différence, produit et généralement comme fonctions de variables et vecteurs aléatoires de lois connues. Plus explicitement, étant donné un vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ de loi P_X , il s'agit de déterminer la loi de $h(X)$, où h est une fonction mesurable définie sur \mathbb{R}^n à valeurs dans \mathbb{R}^m .

4.1 Approche du problème sur deux exemples

(1) Soit X une v.a. normale centrée réduite $\mathcal{N}(0, 1)$; pour déterminer la loi de $Y = X^2$, on exprime F_Y en fonction de F_X :

$$\forall y > 0, F_Y(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq y) = F_X(\sqrt{y}) - F_X(-\sqrt{y}),$$

$$\text{d'où : } f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})).$$

$$\text{Si } f_X(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \text{ alors } f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} \mathbf{1}_{\mathbb{R}^+}(y),$$

$$E(Y) = E(X^2) = V(X) + (E(X))^2 = 1.$$

(2) Soient X_1 et X_2 des v.a. de Poisson indépendantes de paramètres λ_1

et λ_2 ; déterminons la loi de $X_1 + X_2$.

$$\begin{aligned}
 \forall k \in \mathbb{N}, P(X_1 + X_2 = k) &= \sum_{i=0}^k P(X_1 = i, X_2 = k - i) \\
 &= \sum_{i=0}^k P(X_1 = i) P(X_2 = k - i) = \sum_{i=0}^k \frac{e^{-\lambda_1} \lambda_1^i}{i!} \frac{e^{-\lambda_2} \lambda_2^{k-i}}{(k-i)!} \\
 &= e^{-(\lambda_1 + \lambda_2)} \sum_{i=0}^k \frac{1}{i!(k-i)!} \lambda_1^i \lambda_2^{k-i} \\
 &= \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} (\lambda_1 + \lambda_2)^k \sum_{i=0}^k \binom{k}{i} \frac{\lambda_1^i \lambda_2^{k-i}}{(\lambda_1 + \lambda_2)^k} \quad (*)
 \end{aligned}$$

Or $\frac{\lambda_1^i \lambda_2^{k-i}}{(\lambda_1 + \lambda_2)^k} = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^i \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{k-i}$, d'où grâce à la formule du binôme :

$$P(X_1 + X_2 = k) = \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} (\lambda_1 + \lambda_2)^k.$$

Conclusion : $X_1 + X_2$ est de loi de Poisson de paramètre $(\lambda_1 + \lambda_2)$.

Théorème 4.1 (*Stabilité des sommes finies de variables indépendantes binomiales et de Poisson*)

(1) Toute somme finie de v.a. de Poisson indépendantes de paramètres respectifs $(\lambda_i)_{i=1, \dots, n}$ est une v.a. de Poisson de paramètre $\sum_i \lambda_i$.

(2) Toute somme finie de v.a. binomiales indépendantes de paramètres respectifs (n_i, p) est une binomiale $\mathcal{B}(\sum_i n_i, p)$.

4.1 ♠ Démontrer la proposition (2) du théorème dans le cas de deux variables en utilisant la méthode de l'exemple introductif (2).

4.2 ♠ Deux machines (1) et (2) dont les fonctionnements sont indépendants, sont mises en marche au même instant $t = 0$; leurs durées de vie respectives sont décrites par les v.a. T_1 et T_2 entières positives. À chaque instant n , leurs probabilités respectives de panne sont p_1 et p_2 .

(1) Pour tout temps $n \in \mathbb{N}^*$, exprimer l'événement $(T_1 = n)$ en fonction des événements $((1) \text{ tombe en panne au temps } i)_{i=1, \dots, n}$.

(2) En déduire la loi de T_i ($i = 1, 2$) et de $T_1 - T_2$.

Calculer $P(T_2 < T_1)$, puis $P(T_1 < T_2)$ et $P(T_2 = T_1)$.

Intermède biographique : Siméon Denis Poisson (1781-1840)

Issu d'un milieu modeste, le très jeune Siméon commença des études en chirurgie puis en médecine ; sa phobie des blessures et du sang l'oblige à arrêter ses études. Entré à l'École Centrale (ancêtre des lycées) de Pithiviers, il excelle en mathématiques et en physique, prépare et réussit brillamment à dix-sept ans le concours d'entrée à l'École polytechnique, où son talent est remarqué par ses professeurs : Laplace, Monge et Lagrange. Dominant l'analyse de son temps, ses recherches sur les séries de Fourier annoncent celles de Dirichlet et de Riemann ; il laisse aussi de nombreux travaux en mécanique, en physique mathématique et en théorie des probabilités, notamment dans son livre, au titre joliment incongru : *Recherches sur la probabilité des jugements en matière criminelle et civile*.

4.2 Méthodes classiques

Le théorème de la probabilité image est le résultat fondamental du calcul des lois.

Théorème 4.2 (de la probabilité image) *Étant donné :*

- un vecteur aléatoire X défini sur (Ω, \mathcal{T}, P) à valeurs dans $(\mathbb{R}^n, \beta_n, P_X)$;
- une fonction mesurable h de $(\mathbb{R}^n, \beta_n, P_X)$ dans (\mathbb{R}^m, β_m) .

On définit la probabilité image $P_{h(X)}$ du vecteur aléatoire $h(X)$ sur (\mathbb{R}^m, β_m) :

$$\forall B \in \beta_m, P_{h(X)}(B) = P_X(h^{-1}(B)).$$

Preuve $P_{h(X)}(B) = P(h(X) \in B) = P(\{\omega \in \Omega \mid X(\omega) \in h^{-1}(B)\}) = P_X(h^{-1}(B))$.
(Ce résultat est une illustration du théorème de la mesure-image exposé dans le cadre de la théorie de la mesure et de l'intégration.) \square

Exemple 4.1 Application du théorème aux deux exemples introductifs

(1) $h(X) = X^2$, $B = [0, y]$ donc :

$$F_{X^2}(y) = P_X(h^{-1}([0, y])) = P_X([- \sqrt{y}, \sqrt{y}]) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

(2) $h(x_1, x_2) = x_1 + x_2$ et $B = \{k\}, k \in \mathbb{N}$:

$$\begin{aligned} P_{X_1+X_2}(k) &= P_{(X_1, X_2)}(h^{-1}(k)) = P_{(X_1, X_2)}((0, k), (1, k-1), \dots, (k, 0)) \\ &= P\left(\bigoplus_{i=0}^k (X_1 = i) \cap (X_2 = k-i)\right) \\ &= \sum_{i=0}^k P(X_1 = i) P(X_2 = k-i) \text{ (par } \Sigma\text{-additivité et indépendance).} \end{aligned}$$

4.2.1 Méthode du changement de variables

Nous l'appliquerons au cas des vecteurs dont la loi est définie par une densité. Rappelons qu'un changement de variables est défini par un difféomorphisme (fonction de classe C^1 , bijective et d'inverse C^1).

Théorème 4.3 *Si aux hypothèses générales du calcul de lois on adjoint l'hypothèse de difféomorphisme de h , alors la densité du vecteur $h(X)$ est définie par :*

$$f_{h(X)}(y) = f_X(h^{-1}(y)) |J(h^{-1}(y))|,$$

où $J(h^{-1}(y))$ est le jacobien de h^{-1} en y .

Preuve On recherche la densité g de $h(X)$ sous la forme $P_{h(X)}(B) = \int_B g(y) dy$.

$$\forall B, P_{h(X)}(B) = P_X(h^{-1}(B)) = \int_{h^{-1}(B)} f_X(x) dx = \int_B f_X(h^{-1}(y)) |J(h^{-1}(y))| dy.$$

D'où l'égalité presque partout de $g(y)$ et $f_X(h^{-1}(y)) |J(h^{-1}(y))|$.

Si f est continue sur son support, l'égalité précédente est alors vraie partout.

□

Remarque 4.1 Dans le cas où la fonction h est non bijective, on utilise la méthode de complémentation qui consiste à considérer h comme composante d'une fonction H bijective.

Application de la méthode de complémentation au calcul de la loi du rapport $\frac{X}{Y}$.

La fonction $h : \begin{pmatrix} X \\ Y \end{pmatrix} \longrightarrow \frac{X}{Y}$ n'étant pas bijective, considérons le difféomorphisme $H : \begin{pmatrix} X \\ Y \end{pmatrix} \longrightarrow \begin{pmatrix} \frac{X}{Y} \\ X \end{pmatrix}$ et appliquons le théorème précédent :

$$\begin{aligned} f_{(X/Y, X)}(u, v) &= f_{H(X, Y)}(u, v) = f_{X, Y}(H^{-1}(u, v)) |J(H^{-1}(u, v))| \\ &= f_{X, Y} \left(v, \frac{v}{u} \right) \frac{|v|}{u^2}, \end{aligned}$$

d'où la densité recherchée : $f_{\frac{X}{Y}}(u) = \int f_{X, Y} \left(v, \frac{v}{u} \right) \frac{|v|}{u^2} dv$.

4.3 ♠♠ Calculer $f_{\frac{X}{Y}}(u)$ dans le cas où X et Y sont des variables indépendantes de loi uniforme de support $]0, 1]$. [Indication : distinguer les deux cas ($0 < u < 1$) et ($1 \leq u$).] Que vaut $E \left(\frac{X}{Y} \right)$?

4.4 ♠♠ Soient X_1, X_2, \dots, X_n des v.a. indépendantes de loi normale $\mathcal{N}(m, \sigma^2)$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ et $U_i = X_i - \bar{X}$.

(1) Loi de \bar{X} ?

(2) On pose $X = {}^t(X_1, X_2, \dots, X_n)$ et $U = {}^t(U_1, U_2, \dots, U_n)$.

Écrire U sous la forme $A.X$; en déduire que U est gaussien. Calculer $E(U)$ et $V(U)$. Démontrer que $\text{Cov}(U_i, \bar{X}) = 0$ et en déduire que U est indépendant de X .

4.5 ♠ La variable de Cauchy comme rapport de variables normales

Soit un vecteur gaussien centré (X, Y) de variances σ_X^2 , σ_Y^2 et de covariance nulle. Démontrer que la v.a. $Z = \frac{X}{Y}$ est une v.a. de Cauchy de densité :

$$f_Z(z) = \frac{\sigma_x}{\sigma_y} \cdot \frac{1}{\pi(z^2 + (\frac{\sigma_x}{\sigma_y})^2)}.$$

4.2.2 Méthode d'identification fonctionnelle

Cette méthode efficace et élégante est fondée sur le théorème ci-dessous.

Théorème 4.4 Soient X un vecteur aléatoire de dimension k de loi P_X et $Y = h(X)$, où h est une fonction mesurable définie sur \mathbb{R}^k . Pour toute fonction φ définie sur \mathbb{R}^k , telle que $E(|\varphi(h(X))|)$ soit fini, alors $E(\varphi(h(X)))$ est égal à :

$$\int_{\mathbb{R}^k} \varphi(h(x)) dP_X(x) = \int_{\mathbb{R}^n} \varphi(y) dP_Y(y).$$

Exemple 4.2 Application au calcul de la loi d'une somme de deux v.a. absolument continues.

Soit le couple $X = (X_1, X_2)$ de loi P_X et $h(x_1, x_2) = x_1 + x_2$;

$E(\varphi(Y)) = E(\varphi(X_1 + X_2)) = \int_{\mathbb{R}^2} \varphi(x_1 + x_2) dP_X(x_1, x_2)$
égal d'après le théorème 4.4 à :

$$\int_{\mathbb{R}^2} \varphi(y) dP_X(x_1, y - x_1). \quad (*)$$

Si X_1 et X_2 sont indépendantes, alors $(*)$ est égale à
 $\int_{\mathbb{R}} \varphi(y) \int_{\mathbb{R}} dP_{X_1}(x_1) dP_{X_2}(y - x_1),$

et grâce au théorème précédent :

$$dP_Y(y) = \int_{\mathbb{R}} dP_{X_1}(x_1) dP_{X_2}(y - x_1).$$

Supposons X_1, X_2 de densités f_{X_1} et f_{X_2} :

$$dP_Y(y) = \int_{\mathbb{R}} f_{X_1}(x_1) f_{X_2}(y - x_1) dx_1 dy.$$

Conclusion : Y est une v.a. de densité $f_Y = f_{X_1} * f_{X_2}$.

On pourrait déterminer la densité de $(X_1 - X_2)$ en procédant de façon analogue, mais il est plus rapide de remarquer que $f_{-X_2}(x) = f_{X_2}(-x)$ et de conclure :

$$f_{X_1 - X_2}(u) = f_{X_1} * f_{-X_2}(u) = \int f_{X_1}(s) f_{X_2}(s - u) ds.$$

4.6 ♠♠ Déterminer à l'aide de la méthode d'identification fonctionnelle la densité du produit de deux v.a. X_1, X_2 connaissant la densité $f_{(X_1, X_2)}$ de support $D_{X_1} \times D_{X_2}$.

[solution : $f_{X_1, X_2}(t) = \int_{D_{X_2}} f_{(X_1, X_2)}(\frac{t}{x_2}, x_2) \frac{1}{|x_2|} \mathbf{1}_{D_{X_1}}(\frac{t}{x_2}) dx_2$]

On suppose que X_1 et X_2 sont indépendantes, de même loi uniforme sur $[0, 1]$. Déterminer la loi de $X_1 \cdot X_2$.

4.7 ♠ Soient X_1, X_2 v.a. indépendantes de densité f_1 et f_2 .
Montrer que : $P(X_i \leq X_j) = \int_{\mathbb{R}} F_{X_i}(x) \cdot f_{X_j}(x) \cdot dx$ où (i, j) distincts $\in \{1, 2\}$.
En déduire que $P(X_1 \leq X_2) = P(X_2 \leq X_1) = \frac{1}{2}$ si $f_1 = f_2$.

Théorème 4.5 La densité f_S de la somme S des v.a. indépendantes X_1, X_2, \dots, X_n de densités respectives f_{X_1}, \dots, f_{X_n} est égale au produit de convolution (commutatif et associatif) des densités :

$$f_S = f_{X_1} * f_{X_2} * \dots * f_{X_n}.$$

On a un résultat équivalent pour une somme de v.a. discrètes indépendantes :

$$P_S = P_{X_1} * \dots * P_{X_n}.$$

où $*$ désigne le produit de convolution des mesures.

Dans le cas de v.a. entières positives X, Y , le théorème suivant explicite la loi de $X+Y$.

Théorème 4.6 Soient X, Y des v.a. entières positives indépendantes de lois $P_X = \sum_{i \in \mathbb{N}} p_i \delta_i$ et $P_Y = \sum_{j \in \mathbb{N}} q_j \delta_j$. Alors la loi P_{X+Y} est définie par :

$$\forall k \in \mathbb{N}, P_{X+Y}(k) = P_X * P_Y(k) = \sum_{i=0}^k p_i \cdot q_{k-i}.$$

Preuve On rappelle le résultat d'analyse : $\delta_i * \delta_j = \delta_{i+j}$.
 $\forall k \in \mathbb{N}, P_X * P_Y(k) = (\sum_{i \in \mathbb{N}} p_i \delta_i) * (\sum_{j \in \mathbb{N}} q_j \delta_j)(k) = \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} p_i q_j \delta_{i+j}(k) = \sum_{i=0}^k p_i q_{k-i}$.
 \square

Théorème 4.7 (Stabilité du caractère normal par combinaison linéaire finie de v.a. normales indépendantes)

Soient n variables aléatoires normales indépendantes X_1, \dots, X_n de lois respectives $\mathcal{N}(m_i, \sigma_i^2)$. Pour tous réels a_1, \dots, a_n :

$$\sum_{i=1}^n a_i X_i \text{ est } \mathcal{N}(\sum_{i=1}^n a_i m_i ; \sum_{i=1}^n a_i^2 \sigma_i^2).$$

Preuve Par définition, $\Phi_{X_j}(t) = e^{im_j t - \frac{1}{2}\sigma_j^2 t^2}$;

donc $\Phi_{a_j X_j}(t) = \Phi_{X_j}(a_j t) = e^{ia_j m_j t - \frac{1}{2}\sigma_j^2 a_j^2 t^2}$. (X_j indépendantes) $\Rightarrow \Phi_{\sum_j a_j X_j}(t) = \prod_j e^{ia_j m_j t - \frac{1}{2}\sigma_j^2 a_j^2 t^2} = e^{i\sum_j a_j m_j t - \frac{1}{2}t^2(\sum_j \sigma_j^2 a_j^2)}$ qui est de la forme $e^{im - \frac{1}{2}t^2\sigma^2}$. La v.a. $\sum_j a_j X_j$ est donc une v.a. normale de moyenne $m = \sum_j a_j m_j$ et de variance $\sigma^2 = \sum_j a_j^2 \sigma_j^2$. \square

Remarque 4.2 Le calcul de l'intégrale de convolution lorsque les supports des densités f_{X_i} sont bornés n'est pas toujours facile et exige beaucoup de rigueur dans la conduite des calculs comme le montre l'exemple suivant.

Soient X_1 et X_2 , indépendantes, de même loi uniforme sur $[-\frac{L}{2}, \frac{L}{2}]$. Calculons la densité de $(X_1 - X_2)$.

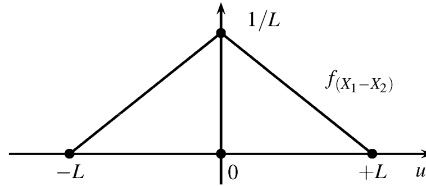
$$\begin{aligned} f_{X_1 - X_2}(u) &= f_{X_1} * f_{-X_2}(u) = \int f_{X_1}(s) f_{X_2}(s - u) ds \\ &= \int \frac{1}{L} \mathbf{1}_{[-\frac{L}{2}, \frac{L}{2}]}(s) \frac{1}{L} \mathbf{1}_{[-\frac{L}{2}, \frac{L}{2}]}(s - u) ds. \end{aligned}$$

Intégrale qui est non nulle si : $(-\frac{L}{2} \leq s \leq \frac{L}{2})$ et $(-\frac{L}{2} \leq s - u \leq \frac{L}{2})$
 $\iff \max(-\frac{L}{2}, u - \frac{L}{2}) \leq s \leq \min(\frac{L}{2}, u + \frac{L}{2})$.

On doit donc considérer deux cas selon la position de u :

- si $u \in [-L, 0]$ alors $f_{X_1 - X_2}(u) = \frac{u+L}{L^2}$;
- si $u \in [0, L]$ alors $f_{X_1 - X_2}(u) = \frac{L-u}{L^2}$.

Conclusion : $f_{X_1 - X_2}(u) = \frac{u+L}{L^2} \mathbf{1}_{[-L, 0]}(u) + \frac{L-u}{L^2} \mathbf{1}_{[0, L]}(u)$.



Si X_1 et X_2 sont indépendantes de même densité symétrique par rapport à l'axe des y , comme dans l'exemple précédent, alors :

$$f_{X_1 - X_2} = f_{X_1} * f_{-X_2} = f_{X_1} * f_{X_2} = f_{X_1 + X_2}.$$

Ainsi, si X_1 et X_2 sont de loi $\mathcal{N}(0, 1)$: $X_1 - X_2$ et $X_1 + X_2$ sont de même loi $\mathcal{N}(0, 2)$.

4.8 ♠ Deux personnes arrivent, indépendamment l'une de l'autre, selon une loi temporelle uniforme entre 8 et 9 heures, à un lieu de rendez-vous, et décident que la première arrivée ne devra pas attendre l'autre plus de 10 minutes.

(1) Déterminer les lois de $X_1 - X_2$ et de $|X_1 - X_2|$ où X_1 et X_2 sont les temps d'arrivée des personnes.

(2) Probabilité pour qu'elles se rencontrent ? (On pourra s'aider de la représentation géométrique de cet événement comme borélien de $[0, 1] \times [0, 1]$).

4.9 ♠♠ X et Y deux v.a. exponentielles indépendantes de paramètre 1.

Déterminer les lois de $(X, X+Y)$ puis de $X+Y$ grâce à la méthode du changement de variables ?

4.10 ♠♠ X et Y v.a. indépendantes de lois gamma de paramètres respectifs (λ, α) et (λ, β) : ainsi, $f_X(x) = \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)}$.

(1) Loi conjointe et lois marginales du vecteur $(X+Y, \frac{X}{X+Y})$?

(2) Loi de la somme de n v.a. exponentielles indépendantes de paramètre λ ? On rappelle que la loi exponentielle de paramètre (λ) est identique à la loi gamma de paramètres $(\lambda, 1)$.

4.11 ♠ Un ascenseur ne peut supporter une charge supérieure à 1 200 kg. Chaque utilisateur potentiel de l'ascenseur a un poids X_i aléatoire de loi normale $\mathcal{N}(m=80 ; \sigma^2=225)$.

Notation : $Y_n = \sum_{i=1}^n X_i$ est la v.a. associée aux poids de n individus.

Quel est le nombre maximal de personnes que l'on peut autoriser à monter dans l'ascenseur si l'on veut que le risque de surcharge et donc de non-démarrage soit égal à 10^{-5} ? [Indication : $P(\mathcal{N}(0, 1) > 4, 5) = 3 \cdot 10^{-6}$]

4.12 ♠ Soient les v.a. indépendantes : X normale centrée réduite et B la v.a. de Bernoulli définie par $P(B=1) = P(B=-1) = 0,5$: posons $Y = BX$.

(1) Montrer que Y est normale centrée réduite.

(2) X et Y sont-elles dépendantes ? Décorrélées ?

$P(X+Y=0) = \frac{1}{2}$, ce qui implique que $(X+Y)$ n'est pas de loi absolument continue ; il s'agit d'un exemple frappant d'une somme de deux v.a. dépendantes normales qui n'est pas de loi normale.

4.2.3 Variables aléatoires définies comme maximum ou minimum d'un ensemble de v.a.

Les v.a. de ce type sont mises en œuvre dans les problèmes de fiabilité.

(A) **Durée de vie d'un système en série**

Rappelons que la durée de vie d'un système est sa durée de fonctionnement jusqu'à la première panne. Étant donné n composants C_1, C_2, \dots, C_n en série, de durées de vie décrites par des v.a. indépendantes X_i , quelle est la loi de la durée de vie $Y = \min_{i=1, \dots, n}(X_i)$ du système ?

La méthode consiste à expliciter la fonction de répartition de Y à l'aide des fonctions de répartition des X_i .

$$\begin{aligned} F_Y(y) &= P\left(\min_i(X_i) \leq y\right) = 1 - P(\min_i(X_i) > y) \\ &= 1 - \prod_{i=1}^n P(X_i > y) = 1 - \prod_{i=1}^n (1 - F_{X_i}(y)). \end{aligned}$$

Si les v.a. X_i sont de même loi exponentielle $\mathcal{E}(\lambda)$, la v.a. $\min(X_i)$ est de densité exponentielle $\mathcal{E}(n\lambda)$:

$$f_{\min_i(X_i)}(y) = n\lambda e^{-n\lambda y} \mathbf{1}_{\mathbb{R}_+}(y).$$

On en conclut que la durée de vie d'un système en série de n composants, chacun ayant une durée de vie de loi exponentielle $\mathcal{E}(\lambda)$, est aussi de loi exponentielle $\mathcal{E}(n\lambda)$. La durée de vie moyenne du système en série égale à $E(Y) = \frac{1}{n\lambda}$ est n fois plus petite que la durée de vie $E(X) = \frac{1}{\lambda}$ de chacun de ses composants, résultat qui confirme la fragilité d'un système en série.

(B) Durée de vie d'un système en parallèle

Soient les n composants en parallèle : X_i est la durée de vie du composant C_i et $Z = \max_i(X_i)$ la durée de vie du système.

$$F_Z(z) = \prod_{i=1}^n F_{X_i}(z).$$

Si les v.a. X_i sont de même loi :

$$f_Z(z) = n f_X(z) (F_X(z))^{n-1}.$$

4.13 ♠♠ (1) Déterminer la densité f_Z de Z dans le cas où les X_i sont de lois exponentielles $\mathcal{E}(\lambda_i)_{i=1, \dots, n}$. Démontrer par récurrence que :

$$E(Z) = \sum_i \frac{1}{\lambda_i} - \sum_{i < j} \frac{1}{\lambda_i + \lambda_j} + \sum_{i < j < k} \frac{1}{\lambda_i + \lambda_j + \lambda_k} \dots + (-1)^{n+1} \frac{1}{\lambda_1 + \lambda_2 + \dots + \lambda_n}.$$

(2) Montrer que, si on triple de façon identique un circuit de commande sur un avion par exemple, on ne triple pas sa durée de vie moyenne, mais on l'augmente dans une proportion à déterminer.

4.2.4 Loïs de variables mixtes

Certaines v.a. sont définies comme fonction de v.a. discrètes et de v.a. absolument continues : nous allons illustrer cette situation dans le cas d'une somme et d'un produit.

Soient les v.a. indépendantes X_1 de densité f_{X_1} et X_2 de loi discrète $P_{X_2} = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}$ avec $x_n \in \mathbb{R}$.

Loi de $Y = X_1 + X_2$

Pour tout réel y , l'événement $(X_1 + X_2 \leq y)$ se décompose, grâce au théorème des probabilités totales, en l'union :

$$\bigcup_{n \in \mathbb{N}} ((X_2 = x_n) \cap (Y \leq y)) = \bigsqcup_n ((X_2 = x_n) \cap (X_1 \leq y - x_n)).$$

L'indépendance de X_1 et X_2 entraîne :

$$F_Y(y) = \sum_n P(X_2 = x_n) P(X_1 \leq y - x_n) = \sum_n p_n F_{X_1}(y - x_n)$$

et par dérivation :

$$f_{X_1+X_2}(y) = \sum_n p_n f_{X_1}(y - x_n)$$

Application : $P(X_2 = -1) = p$ et $P(X_2 = 1) = 1 - p$; $X_1 \sim N(0, \sigma^2)$;
d'où :

$$f_Y(y) = p \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y+1)^2}{2\sigma^2}} + (1-p) \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2\sigma^2}}$$

Loi de $Z = X_1 X_2$

On supposera d'abord que : $\forall n, x_n \neq 0$

$$(\forall z \in \mathbb{R}) (X_1 X_2 \leq z) = \bigsqcup_n ((X_2 = x_n) \cap (X_1 X_2 \leq z)) ;$$

$$P(Z \leq z) = \sum_n P(X_2 = n) P(x_n X_1 \leq z)$$

.

$$\begin{aligned} F_Z(z) &= \sum_{x_n > 0} p_n P(X_1 \leq \frac{z}{x_n}) + \sum_{x_n < 0} p_n P(X_1 > \frac{z}{x_n}) \\ &= \sum_{x_n > 0} p_n F_{X_1}(\frac{z}{x_n}) + \sum_{x_n < 0} p_n (1 - F_{X_1}(\frac{z}{x_n})), \end{aligned}$$

puis par dérivation :

$$f_Z(z) = \sum_n p_n \frac{1}{|x_n|} f_{X_1}(\frac{z}{x_n}).$$

Remarque 4.3 En pratique, on devrait toujours s'assurer au terme du calcul d'une loi quelconque, qu'elle vérifie l'axiome $P(\Omega) = 1$. Bien que cette vérification ne soit qu'une condition nécessaire de l'exactitude de l'expression de la loi à laquelle on aboutit, elle permet de conclure avec une quasi certitude. Cette méthode joue un rôle analogue à la méthode dite de la *preuve par 9*, pratiquée naguère dans les écoles, dont la vérification était une condition nécessaire de la justesse du résultat d'une multiplication.

4.3 Méthodes de simulation des lois de probabilité

La simulation de phénomènes aléatoires grâce aux variables aléatoires qui les décrivent est d'une grande utilité pratique : calcul d'intégrales par les méthodes de Monte-Carlo, simulation du fonctionnement de réseaux (informatiques, routiers, aériens) et de l'évolution de systèmes physiques. Aux deux méthodes présentées dans ce paragraphe, il faut ajouter la méthode de rejet qu'on n'expose pas ici.

(4.3.1) Génération de la v.a. de loi uniforme sur $[0, 1]$

Les algorithmes multiplicatifs à congruence linéaire permettent de générer des suites d'entiers $(y_n)_n$ ayant l'apparence d'une distribution aléatoire sur un intervalle $[0, m-1]$, à l'aide de relations de récurrence de la forme : $y_{n+1} = a.y_n + b \text{ [modulo } m]$, où les entiers donnés a et m sont premiers entre eux et vérifient $m \gg a$. Par exemple, la fonction RANDOM du logiciel de simulation stochastique SCILAB utilise les valeurs $a = 843\,314\,861$, $b = 453\,816\,693$, $m = 2^{31}$ pour générer des suites de nombres *pseudo-aléatoires* de la forme $(x_n = \frac{y_n}{m})_{n \in \mathbb{N}}$, qui simulent des tirages uniformément répartis sur $[0, 1]$.

(4.3.2) Génération de v.a. de loi quelconque

On a choisi de présenter trois méthodes ayant un intérêt pédagogique.

(a) Méthode de la fonction de répartition

Elle est fondée sur le théorème :

Théorème 4.8 Pour toute v.a. X , de fonction de répartition F_X :

X et $F_X^{-1}(U)$ sont de même loi.

(U étant la v.a. uniforme sur $]0, 1[$.)

F_X^{-1} est la fonction quantile : $\forall u, F_X^{-1}(u) = \inf \{t \text{ réels ; } F(t) \geq u\}$.

Cette méthode est peu utilisée, car la fonction $F_X(U)$ est souvent coûteuse à simuler lorsqu'elle contient des fonctions transcendantes (par exemple e^{-x^2}); d'autre part, elle donne des résultats peu satisfaisants pour des valeurs de U voisines des points où la dérivée de F_X est très petite.

Exemple 4.3 Simulation de la v.a. exponentielle X de paramètre λ :

$$F_X^{-1}(U) = -\frac{1}{\lambda} \ln(U).$$

4.14 ♠ Simulation d'une v.a. géométrique

Si X est de loi $\mathcal{E}(\lambda)$ en déduire que :

$$\forall n, P(n-1 < X \leq n) = (e^{-\lambda})^{n-1} (1 - e^{-\lambda}).$$

En déduire que la v.a. égale à la partie entière (X) est de loi géométrique de paramètre $p = 1 - e^{-\lambda}$.

(b) Simulation de v.a. à valeurs dans un ensemble fini

La méthode est fondée sur le théorème :

Théorème 4.9 La v.a. X de loi $P = \sum_{i=1}^{i=n} p_i \delta_{x_i}$, est simulée par la v.a. :

$$X = x_1 \mathbf{1}_{\{U < p_1\}} + \sum_{i=1}^{n-2} x_{i+1} \mathbf{1}_{\{p_1 + \dots + p_i \leq U \leq p_1 + \dots + p_{i+1}\}} \\ + x_n \mathbf{1}_{\{p_1 + \dots + p_{n-2} + p_{n-1} \leq U \leq 1\}}.$$

4.15 ♠♠ (1) Démontrer la proposition précédente.

(2) Quelle loi la v.a. $\sum_{i=1}^{i=n} \mathbf{1}_{\{U_i < p_i\}}$ simule-t-elle ? [réponse : v.a. binomiale]

(c) Simulation de v.a. normales par la méthode de Box-Muller

4.16 ♠♠ Méthode polaire (dite de Box-Muller) de simulation de la loi normale $\mathcal{N}(0, 1)$

Soient X_1 et X_2 , indépendantes, de loi uniforme sur $[0, 1]$.

En déduire que $X = (-2 \ln(X_1))^{\frac{1}{2}} \cos(2\pi X_2)$ et $Y = (-2 \ln(X_1))^{\frac{1}{2}} \sin(2\pi X_2)$ sont indépendantes de loi $\mathcal{N}(0, 1)$.

4.4 Quelques densités classiques

Soit un vecteur aléatoire (X_1, X_2) de densité f_{X_1, X_2} .

(1) X_1 et X_2 sont dépendantes

$$f_{X_1+X_2}(u) = \frac{1}{2} \int f_{(X_1, X_2)}\left(\frac{u+v}{2}, \frac{u-v}{2}\right) dv ;$$

$$f_{X_1-X_2}(u) = \frac{1}{2} \int f_{(X_1, X_2)}\left(\frac{u+v}{2}, \frac{u-v}{2}\right) du ;$$

$$f_{X_1, X_2}(u) = \int_{\mathbb{R}} f_{(X_1, X_2)}\left(\frac{u}{v}, v\right) \frac{1}{|v|} dv ;$$

$$f_{\frac{X_1}{X_2}}(u) = \int_{\mathbb{R}} f_{(X_1, X_2)}\left(v, \frac{v}{u}\right) \frac{v}{u^2} dv.$$

(2) X_1 et X_2 sont indépendantes

$$f_{X_1+X_2}(u) = \int_{\mathbb{R}} f_{X_1}(u-v) f_{X_2}(v) dv ;$$

$$f_{X_1-X_2}(u) = \int_{\mathbb{R}} f_{X_1}(v) f_{X_2}(v-u) dv ;$$

$$f_{X_1 X_2}(u) = \int_{\mathbb{R}} f_{X_1}\left(\frac{u}{v}\right) f_{X_2}(v) \frac{1}{|v|} dv ;$$

$$f_{\frac{X_1}{X_2}}(u) = \int_{\mathbb{R}} f_{X_1}\left(v\right) f_{X_2}\left(\frac{v}{u}\right) \frac{v}{u^2} dv.$$

Présentation de trois nouvelles v.a. et de leurs lois, construites à partir de v.a. gaussiennes centrées réduites, d'usage fréquent en théorie statistique.

Définition 4.1 (1) Étant donné n v.a. gaussiennes $(X_i)_{i=1,2,\dots,n}$ indépendantes $\mathcal{N}(0,1)$, la v.a. du **chi-deux** à n degrés de liberté est définie par :

$$\sum_{i=1}^n X_i^2 \underset{\text{(noté)}}{=} \chi^{(2)}(n).$$

(2) Étant donné $n+p$ v.a. gaussiennes $(X_i)_{i=1,2,\dots,n+p}$ indépendantes de loi $\mathcal{N}(0,1)$, et les v.a. du chi-deux $\chi^{(2)}(n) = \sum_{i=1}^n X_i^2$ et $\chi^{(2)}(p) = \sum_{i=n+1}^{n+p} X_i^2$, indépendantes par construction, la v.a. de **Fisher-Snedecor** $F(n; p)$ de paramètres (n, p) :

$$\frac{\frac{\chi^{(2)}(n)}{n}}{\frac{\chi^{(2)}(p)}{p}} \underset{\text{(noté)}}{=} F(n; p).$$

(3) La v.a. de **Student** T_n à n degrés de liberté est définie par :

$$\frac{U}{\sqrt{\frac{\chi^{(2)}(n)}{n}}} \stackrel{(noté)}{=} T_n.$$

U est une v.a. normale $\mathcal{N}(0, 1)$ indépendante de la v.a. $\chi^{(2)}(n)$.

Théorème 4.10 (de Fisher)

Si X_1, \dots, X_n sont des v.a. indépendantes $\mathcal{N}(m, \sigma^2)$, alors :

$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ et $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ sont indépendantes,

et \bar{X}_n est $\mathcal{N}(m, \frac{\sigma^2}{n})$ et $\frac{(n-1)S^2}{\sigma^2}$ est une v.a. $\chi^{(2)}(n-1)$.

Théorème 4.11 (Caractéristiques des v.a. $\chi^{(2)}(n)$, de Fisher-Snedecor et de Student)

(1) Densité de la v.a. du chi-deux à n d.d.l. $\chi^{(2)}(n)$:

$$\forall x \geq 0 \quad f_{\chi^{(2)}(n)}(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \exp(-\frac{x}{2}) x^{\frac{n}{2}-1},$$

$$E(\chi^{(2)}(n)) = n ; V(\chi^{(2)}(n)) = 2n.$$

(2) Densité de la v.a. de Fisher-Snedecor $F(n, p)$ de paramètres (n, p) :

$$\forall x \geq 0, \quad f_{F(n; p)}(x) = \frac{n^{\frac{n}{2}} p^{\frac{p}{2}} x^{\frac{n}{2}-1}}{\beta(\frac{n}{2}, \frac{p}{2}) (p + nx)^{\frac{(n+p)}{2}}},$$

$$E(F(n, p)) = \frac{p}{p-2} \text{ si } p > 2.$$

(3) Densité de la v.a. de Student T_n à n d.d.l. :

$$\forall x \in \mathbb{R} \quad f_{T_n}(x) = \frac{1}{\sqrt{n} B(\frac{1}{2}, \frac{n}{2}) (1 + \frac{x^2}{n})^{\frac{(n+1)}{2}}}.$$

$$E(T_n) = 0 ; V(T_n) = \frac{n}{n-2}.$$

Preuve (1) La fonction de répartition $F_{\chi^{(2)}(n)}(u)$ est définie par :

$$(2\pi)^{-\frac{n}{2}} \int \int_{r^2 \leq u} \exp(-\frac{1}{2} \sum_{i=1}^n y_i^2) dy_1 \dots dy_n \text{ où l'on a posé } r^2 = \sum_{i=1}^n y_i^2.$$

Si Ω est l'angle solide associé aux coordonnées polaires, on a :

$$F_{\chi^{(2)}(n)}(u) = (2\pi)^{-\frac{n}{2}} \int d\Omega \int_0^{\sqrt{u}} e^{(-\frac{r^2}{2})} r^{n-1} dr ; \int d\Omega = \frac{2}{\Gamma(\frac{n}{2})} \pi^{\frac{n}{2}}.$$

On pose $t = r^2$ dans l'intégrale ci-dessus pour aboutir au résultat. □

Les problèmes 7, 8 et 9 se rapportent à ce chapitre.

4.5 Corrigés des exercices

Exercice 4.1

$$\forall k \in \{0, 1, \dots, n_1 + n_2\}$$

$$\begin{aligned} P([X_1 + X_2 = k]) &= \sum_{i=0}^k P([X_1 = i])P([X_2 = k - i]) \\ &= \sum_{i=0}^k \frac{n_1!}{i!(n_1 - i)!} p^i (1 - p)^{n_1 - i} \frac{(n - n_1)!}{(k - i)!(n - n_1 - k + i)!} p^{k - i} (1 - p)^{n - n_1 - k + i} \\ &= p^k (1 - p)^{n - k} \sum_{i=0}^k \binom{n_1}{i} \binom{n - n_1}{k - i} = p^k (1 - p)^{n - k} \binom{n}{k}, \text{ où } n = n_1 + n_2. \end{aligned}$$

Exercice 4.2

$$(1) P([T_i = k]) = (1 - p_i)^{k-1} p_i \text{ (loi géométrique de paramètre } p_i).$$

$$\begin{aligned} P([T_1 - T_2 = k]) &= P\left(\bigcup_{i=1}^{+\infty} ([T_1 = k + i] \cap [T_2 = i])\right) = \sum_{i=1}^{+\infty} P([T_1 = k + i])P([T_2 = i]) \\ &= \frac{p_1 p_2 (1 - p_1)^k}{p_1 + p_2 - p_1 p_2}. \end{aligned}$$

$$(2) P([T_2 < T_1]) = P([T_1 - T_2 > 0]) = \sum_{k \geq 1} P([T_1 - T_2 = k]) = \frac{p_2(1 - p_1)}{p_1 + p_2 - p_1 p_2} \quad (*)$$

$$P([T_2 < T_1]) = \frac{p_1(1 - p_2)}{p_1 + p_2 - p_1 p_2} \text{ (il suffit d'échanger } p_1 \text{ et } p_2 \text{ dans } (*)) \text{ d'où}$$

$$P([T_2 = T_1]) = 1 - P([T_1 < T_2]) - P([T_2 < T_1]) = \frac{p_1 p_2}{p_1 + p_2 - p_1 p_2}.$$

Exercice 4.3

Méthode du changement de variable :

$$f_{X/Y}(u) = \int f_{X,Y}\left(v, \frac{v}{u}\right) \frac{v}{u^2} dv = \frac{1}{u^2} \int \mathbf{1}_{]0,1[}(v) \mathbf{1}_{]0,1[}\left(\frac{v}{u}\right) v dv.$$

Deux cas :

- $0 < u < 1$ et $0 < v < 1$: $f_{X/Y}(u) = \frac{1}{u^2} \int_0^u v dv = \frac{1}{2} \mathbf{1}_{]0,1[}(u)$;
- $1 \leq u$ et $0 < v < 1$: $f_{X/Y}(u) = \frac{1}{2u^2} \mathbf{1}_{]1,+\infty[}(u)$.

$$f_{X/Y}(u) = \frac{1}{2} \mathbf{1}_{]0,1[}(u) + \frac{1}{2u^2} \mathbf{1}_{]1,+\infty[}(u).$$

Au voisinage de $+\infty$, $u f_{X/Y}(u) \sim \frac{1}{2u}$, n'est pas intégrable : X/Y a donc une espérance infinie.

Exercice 4.4

(1) \bar{X} est normale (combinaison linéaire de v.a. normales).

$$\mathbb{E}(\bar{X}) = m; V(\bar{X}) = \frac{\sigma^2}{n}.$$

(2) Pour tout i , U_i est l'image par transformation linéaire du vecteur (X_1, X_2, \dots, X_n) : elle est donc normale d'espérance $\mathbb{E}(U_i) = \mathbb{E}(X_i - \bar{X}) = \mathbb{E}(X_i) - \mathbb{E}(\bar{X}) = 0$.

$$\begin{aligned} \text{cov}(U_i, U_j) &= \mathbb{E}(((X_i - m) - (\bar{X} - m)) \cdot ((X_j - m) - (\bar{X} - m))) \\ &= -2\mathbb{E}((X_i - m)(\bar{X} - m)) + V(\bar{X}) = -\frac{\sigma^2}{n}. \end{aligned}$$

Par l'indépendance des X_i .

$$\begin{aligned} V(U_i) &= \left(1 - \frac{1}{n}\right)^2 V(X_i) + \frac{1}{n^2} (n-1)V(X_j) \\ &= \sigma^2 \left(1 - \frac{2}{n} + \frac{1}{n^2} + \frac{1}{n} - \frac{1}{n^2}\right) = \sigma^2 \left(1 - \frac{1}{n}\right). \end{aligned}$$

On a vu que U est de la forme $A.X$: donc U est gaussien.

$$\begin{aligned} \text{cov}(U_i, \bar{X}) &= \mathbb{E}(((X_i - m) - (\bar{X} - m)) \cdot (\bar{X} - m)) \\ &= \text{cov}(X_i, \bar{X}) - V(\bar{X}) = 0. \end{aligned}$$

Comme, pour tout i , (U_i, \bar{X}) est gaussien, la décorrélation de U_i et \bar{X} implique leur indépendance.

Exercice 4.6

Pour toute application φ continue bornée définie sur \mathbb{R}^2 , on obtient par application du théorème 4.4 :

$$E(\varphi(X_1 X_2)) = \int_{\mathbb{R}^2} \varphi(x_1 x_2) f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 = \int_{\mathbb{R}} \varphi(y) dP_{X_1 \cdot X_2}(y).$$

Afin d'expliciter $P_{X_1 \cdot X_2}$, on développe la première intégrale après avoir posé $y = x_1 \cdot x_2$:

$$\int_{D_Y} \left(\int_{D_{X_2}} \varphi(y) f_{X_1} \left(\frac{y}{x_2} \right) f_{X_2}(x_2) \frac{1}{|x_2|} dx_2 \right) dy,$$

d'où, par application du théorème d'identification fonctionnelle :

$$dP_Y(y) = \left(\int_{(x_2 \neq 0, \frac{y}{x_2} \in D_{X_1})} f_{X_1} \left(\frac{y}{x_2} \right) f_{X_2}(x_2) \frac{1}{|x_2|} dx_2 \right) dy.$$

Supposons X_1, X_2 indépendantes de loi $\mathcal{U}([0, 1])$:

$$f_{X_1 X_2}(y) = \int_{(0 \leq \frac{y}{x_2} \leq 1) \text{ et } (0 \leq x_2 \leq 1)} \frac{1}{x_2} dx_2 ;$$

le domaine d'intégration est donc défini par $]y, +\infty[\cap [0, 1]$, soit $]y, 1]$; d'où :

$$f_{X_1 X_2}(y) = \int_y^1 \frac{1}{x_2} dx_2 = -\ln y.$$

Exercice 4.7

$$f_{X_2 - X_1}(x) = \int f_{X_2}(s) \cdot f_{X_1}(s - x) ds$$

$$\begin{aligned} P(X_1 \leq X_2) &= P(X_1 - X_2 \geq 0) = \int_0^{+\infty} f_{X_2 - X_1} \cdot dx = \int_0^{+\infty} \left(\int f_{X_2}(s) \cdot f_{X_1}(s - x) ds \right) dx \\ &\stackrel{(u=s-x)}{=} \int f_{X_2}(s) \left(\int_s^{-\infty} -f_{X_1}(u) du \right) ds = \int f_{X_2}(s) \cdot F_{X_1}(s) ds \end{aligned}$$

Par symétrie, $P(X_2 \leq X_1) = \int f_{X_1}(s) \cdot F_{X_2}(s) ds$; si $f_1 = f_2$, alors : $P(X_2 \leq X_1) = P(X_1 \leq X_2) = \frac{1}{2}$.

Exercice 4.8

Deux méthodes :

(1) Méthode de convolution : $X_1 - X_2$ prend ses valeurs dans $[-1, 1]$

$$f_{X_1 - X_2}(t) = \int f_{X_1}(s) f_{X_2}(s - t) ds = \int_0^1 \mathbf{1}_{[0, 1]}(s - t) ds.$$

- si $t \in [-1, 0]$, alors $\int_0^1 \mathbf{1}_{[0, 1]}(s - t) ds = 1 + t$;
- si $t \in [0, 1]$, alors $\int_0^1 \mathbf{1}_{[0, 1]}(s - t) ds = 1 - t$. D'où :

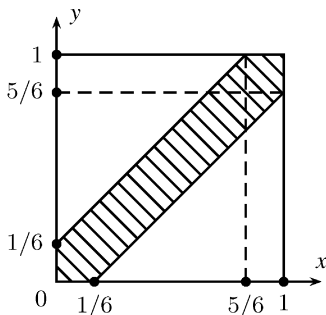
$$f_{X_1 - X_2}(t) = (1 + t) \mathbf{1}_{[-1, 0]}(t) + (1 - t) \mathbf{1}_{[0, 1]}(t).$$

Or, pour toute v.a. réelle X , $f_{|X|}(x) = (f_X(x) + f_X(-x)) \mathbf{1}_{[0, +\infty[}(x)$, d'où :

$$f_{|X_1 - X_2|}(t) = 2(1 - t) \mathbf{1}_{[0, 1]}(t).$$

$$P\left(|X_1 - X_2| \leq \frac{1}{6}\right) = 2 \int_0^{1/6} (1 - t) dt = \frac{11}{36}.$$

(2) Méthode géométrique : on représente l'espace des états, couples des temps d'arrivée de chaque personne, par le carré $[0, 1] \times [0, 1]$. L'événement $|X_1 - X_2| \leq \frac{1}{6}$ correspond à la partie hachurée autour de la diagonale, de surface $\frac{11}{36}$.

**Exercice 4.9**

X et Y étant indépendantes, on a :

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x)f_Y(y) = e^{-x}\mathbf{1}_{]0,+\infty[}(x)e^{-y}\mathbf{1}_{]0,+\infty[}(y) \\ &= e^{-(x+y)}\mathbf{1}_{]0,+\infty[}(x)\mathbf{1}_{]0,+\infty[}(y). \end{aligned}$$

$(U,V) = (X, X+Y)$ avec comme support de (X,Y) : $\{(x,y) \in \mathbb{R}^2 ; x > 0, y > 0\}$. De $u = x$ et $v = x + y$, on déduit $x = u$ et $y = v - u$. On en déduit le support de (U,V) :

$$D = \{(u,v) \in \mathbb{R}^2 ; u > 0, v - u > 0\} = \{(u,v) \in \mathbb{R}^2 ; u > 0, v > u\}.$$

$h : (x,y) \mapsto (u,v) = (x, x+y)$ bijection de \mathbb{R}^2 sur lui-même.

$$h^{-1} : (u,v) \mapsto (x,y) = (u, v-u) \quad J_{h^{-1}}(u,v) = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1.$$

$$\begin{aligned} f_{h(X,Y)}(u,v) &= f_{X,Y}(h^{-1}(u,v)) \times |J_{h^{-1}}(u,v)| \\ &= f_{X,Y}(u, v-u) = f_X(u)f_Y(v-u) = e^{-v}\mathbf{1}_{]0,+\infty[}(u)\mathbf{1}_{]0,+\infty[}(v-u). \end{aligned}$$

$$f_{U,V}(u,v) = e^{-v}\mathbf{1}_D(u,v) \text{ où } D = \{(u,v) \in \mathbb{R}^2 ; 0 < u < v\}.$$

$$\begin{aligned} f_V(v) &= \int f_{U,V}(u,v) du = \int e^{-v}\mathbf{1}_{]0,+\infty[}(v)\mathbf{1}_{]0,v[}(u) du \\ &= e^{-v}\mathbf{1}_{]0,+\infty[}(v) \int_0^v du = ve^{-v}\mathbf{1}_{]0,+\infty[}(v). \end{aligned}$$

Ainsi, $V = X + Y$ suit la loi gamma $\gamma(1,2)$: $f_V(v) = ve^{-v}\mathbf{1}_{]0,+\infty[}(v)$.
On aurait pu calculer la loi de $X + Y$ à l'aide de la méthode classique fondée sur le résultat suivant : la fonction caractéristique d'une somme de plusieurs v.a. indépendantes est égale au produit des fonctions caractéristiques des v.a.

Exercice 4.10

$P_X = \gamma(\lambda, \alpha)$ et $P_Y = \gamma(\lambda, \beta)$ avec X et Y indépendantes d'où

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda(x+y)} x^{\alpha-1} y^{\beta-1} \mathbf{1}_{]0,+\infty[}(x) \mathbf{1}_{]0,+\infty[}(y).$$

(1) $h : (x,y) \mapsto (u,v) = \left(x+y, \frac{x}{x+y}\right)$ définie sur $\mathbb{R}^2 \setminus \{(x,-x), x \in \mathbb{R}\}$ et $h^{-1} : (u,v) \mapsto (x,y) = (uv, u(1-v))$ pour $u \neq 0$.

$$f_{U,V}(u,v) = f_{X,Y}(h^{-1}(u,v)) \times |J_{h^{-1}}(u,v)|$$

$$\text{avec } J_{h^{-1}}(u,v) = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -uv - u + uv = -u \text{ et}$$

$$\begin{aligned} f_{X,Y}(h^{-1}(u,v)) &= f_{X,Y}(uv, u(1-v)) \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda u} (uv)^{\alpha-1} (u(1-v))^{\beta-1} \mathbf{1}_{]0,+\infty[}(uv) \mathbf{1}_{]0,+\infty[}(u(1-v)) \end{aligned}$$

$$\text{avec } \begin{cases} uv > 0 \\ u(1-v) > 0 \end{cases} \text{ si et seulement si } \begin{cases} uv > 0 \\ u^2 v(1-v) > 0 \\ u(1-v) > 0 \end{cases}, \text{ soit } \begin{cases} v \in]0, 1[\\ u > 0 \end{cases} \text{ et}$$

$$f_{U,V}(u,v) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda u} v^{\alpha-1} (1-v)^{\beta-1} u^{\alpha+\beta-1} \mathbf{1}_{]0,+\infty[}(u) \mathbf{1}_{]0,1[}(v)$$

$$f_U(u) = \int f_{U,V}(u,v) dv = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda u} u^{\alpha+\beta-1} \mathbf{1}_{]0,+\infty[}(u) \int_0^1 v^{\alpha-1} (1-v)^{\beta-1} dv$$

$$f_V(v) = \int f_{U,V}(u,v) du = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} v^{\alpha-1} (1-v)^{\beta-1} \mathbf{1}_{]0,1[}(v) \int_0^{+\infty} u^{\alpha+\beta-1} e^{-\lambda u} du.$$

$$\text{Or } \int_0^{+\infty} u^{\alpha+\beta-1} e^{-\lambda u} du \stackrel{(t=\lambda u)}{=} \int_0^{+\infty} t^{\alpha+\beta-1} e^{-t} dt \times \frac{1}{\lambda^{\alpha+\beta}} = \frac{\Gamma(\alpha+\beta)}{\lambda^{\alpha+\beta}}, \text{ donc}$$

$$f_V(v) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} v^{\alpha-1} (1-v)^{\beta-1} \mathbf{1}_{]0,1[}(v).$$

Ainsi, V suit la loi bêta de paramètres α et β . On déduit de $\int f_V(v) dv = 1$:

$$\int_0^1 v^{\alpha-1} (1-v)^{\beta-1} dv = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

$$\text{On a alors } f_U(u) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda u} u^{\alpha+\beta-1} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \mathbf{1}_{]0,+\infty[}(u), \text{ soit :}$$

$$f_U(u) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} u^{\alpha+\beta-1} e^{-\lambda u} \mathbf{1}_{]0,+\infty[}(u).$$

Ainsi, U suit la loi gamma $\gamma(\lambda, \alpha + \beta)$.

(2) Si $P_{X_1} = \mathcal{E}(\lambda) = \gamma(\lambda, 1) = P_{X_2}$ on a vu que $P_{X_1+X_2} = \gamma(\lambda, 2)$ lorsque X_1 et X_2 sont indépendantes.

Par récurrence, supposons que $Z_n = X_1 + \dots + X_n$ suit la loi $\gamma(\lambda, n)$; alors $Z_{n+1} = Z_n + X_{n+1}$ avec $P_{Z_n} = \gamma(\lambda, n)$, $P_{X_{n+1}} = \gamma(\lambda, 1)$, et Z_n et X_{n+1} indépendantes donc $Z_n + X_{n+1}$ suit la loi $\gamma(\lambda, n+1)$.

Ainsi, si les X_i sont indépendantes de loi $\mathcal{E}(\lambda)$, alors $X_1 + \dots + X_n$ suit la loi $\gamma(\lambda, n)$.

Plus généralement, on peut faire le même raisonnement avec la loi gamma : si les X_i suivent la loi gamma $\gamma(\lambda, \alpha)$ et sont indépendantes, alors $X_1 + \dots + X_n$ suit la loi gamma $\gamma(\lambda, n\alpha)$.

Exercice 4.11

$Y_n \sim \mathcal{N}(80n; 225n)$.

$P(Y_n > 1.000) = P\left(\frac{Y_n - 80n}{15\sqrt{n}} > \frac{1.000 - 80n}{15\sqrt{n}}\right)$ avec $\frac{Y_n - 80n}{15\sqrt{n}}$ v.a. de loi $\mathcal{N}(0; 1)$.

On déduit de l'indication que $4,5 \approx \frac{1.000 - 80n}{15\sqrt{n}}$; donc, en posant $x = \sqrt{n}$, on a l'équation :

$$80x^2 + (15 \times 4.5)x - 1.000 = 0.$$

dont l'unique solution positive est $x \approx 3,1$ et donc $n = 10$.

Exercice 4.12

$$\begin{aligned} (1) \quad F_Y(z) &= P([Y \leq z]) = P([XB \leq z]) \\ &= P(([XB \leq z] \cap [B = -1]) \cup ([XB \leq z] \cap [B = 1])) \\ &= P(([-X \leq z] \cap [B = -1]) \cup ([X \leq z] \cap [B = 1])) \\ &= \frac{1}{2} (P([X \geq -z]) + P([X \leq z])) \\ &= \frac{1}{2} (1 - F_X(-z) + F_X(z)) \end{aligned}$$

$$f_Y(z) = \frac{1}{2\sqrt{2\pi}} \left[e^{-\frac{(-z)^2}{2}} + e^{-\frac{z^2}{2}} \right] = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

donc Y suit aussi la loi $\mathcal{N}(0, 1)$.

(2) X et Y sont dépendantes par construction et pourtant décorrélées :

$$\text{Cov}(X, Y) = \mathbb{E}(X.BX) - \mathbb{E}(X).\mathbb{E}(BX) = \mathbb{E}(B) (\mathbb{E}(X^2) - \mathbb{E}(X)^2) = 0.$$

Exercice 4.13

$$(1) F_Z(z) = \prod_{i=1}^n F_{X_i}(z), \text{ d'où :}$$

$$f_Z(z) = \sum_{i=1}^n f_{X_i}(z) F_{X_1}(z) F_{X_2}(z) \cdots F_{X_{i-1}}(z) F_{X_{i+1}}(z) \cdots F_{X_n}(z)$$

(2) $\mathbb{E}(X) = \frac{1}{\lambda}$, $\mathbb{E}(\max(X_1, X_2)) = \frac{1}{\lambda} + \frac{1}{\lambda} - \frac{1}{2\lambda} = \frac{3}{2\lambda}$: donc, si on double le circuit, le rapport des durées de vie est $(\frac{3}{2\lambda}) / \frac{1}{\lambda} = \frac{3}{2}$, d'où un gain égal à 50.

$\mathbb{E}(\max(X_1, X_2, X_3)) = \frac{3}{\lambda} - \frac{3}{2\lambda} + \frac{1}{3\lambda} = \frac{11}{6\lambda}$ d'où, un gain égal à $\frac{11}{6} - 1 = \frac{5}{6}$, soit 83 %.

Exercice 4.15

(1) $[X = x_1] = [U < p_1] : P([X = x_1]) = P([U < p_1]) = p_1$.
 $[X = x_2] = [p_1 < U < p_1 + p_2] : P([X = x_2]) = P([p_1 < U < p_1 + p_2]) = p_2$.
 Plus généralement, $[X = x_n] = [p_1 + \cdots + p_{n-1} < U < p_1 + \cdots + p_n] :$

$$P([X = x_n]) = p_n.$$

(2) Même méthode.

Exercice 4.16

$$U = e^{-\frac{(X^2+Y^2)}{2}} \text{ et } V = \frac{1}{2\pi} \arctan\left(\frac{Y}{X}\right)$$

$$J = e^{-\frac{x^2+y^2}{2}} \frac{1}{2\pi} \frac{x^2}{x^2+y^2} \begin{vmatrix} -x & -y \\ -\frac{y}{x^2} & \frac{1}{x} \end{vmatrix} = -\frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

donc $dudv = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy$, et la densité de (X, Y) est bien $\frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$ sur \mathbb{R}^2 .

Chapitre 5

Convergences et limites des suites aléatoires

En théorie des probabilités, comme en analyse, il existe plusieurs types de convergences ; pour un problème donné, on choisit le type de convergence le plus adapté. Nous étudions successivement la convergence en loi et la convergence en probabilité, qui sont mises en œuvre dans les deux théorèmes limites fondamentaux que sont la loi des grands nombres et le théorème de la limite centrale. On introduit ensuite les notions de convergence presque sûre et de convergence en moyenne quadratique, cette dernière étant nécessaire à l'étude des processus du second ordre associés aux signaux aléatoires d'énergie finie qui seront étudiés au chapitre 10.

5.1 Convergence en probabilité

Considérons un tirage avec remise de n variables de Bernoulli $(B_i)_{i=1,2,\dots,n}$, indépendantes de même loi :

$$\forall i, \quad P_{B_i} = p.\delta_1 + (1-p).\delta_0.$$

On rappelle que $S_n = \sum_{i=1}^n B_i$ est une v.a. binomiale $\mathcal{B}(n, p)$.

S_n dénombre les v.a. de Bernoulli ayant pris la valeur 1 au cours du tirage ; on s'attend donc à ce que la fréquence observée de l'événement $(B = 1)$ définie par la v.a. $\frac{S_n}{n}$ converge vers p lorsque n tend vers l'infini, résultat confirmé par le théorème de Bernoulli.

Théorème 5.1 (de Bernoulli)

Étant donné une v.a. binomiale S_n de loi $\mathcal{B}(n, p)$, la suite des fréquences

empiriques $(\frac{S_n}{n})_{n \in \mathbb{N}}$ converge vers la v.a. constante égale à p , au sens suivant :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P \left[\left| \frac{S_n}{n} - p \right| < \varepsilon \right] = 1.$$

Ce théorème important en statistique justifie l'approximation d'une probabilité p inconnue d'un événement donné, comme limite de la suite des fréquences empiriques de sa réalisation ; il s'agit d'un premier exemple de la notion d'estimation, concept central de la théorie statistique. En effet, pour estimer la probabilité inconnue p , il suffira de procéder à n épreuves de Bernoulli $(B_i)_{i=1, \dots, n}$, dont on calculera la somme S_n puis la moyenne $\frac{S_n}{n}$ pour n suffisamment grand, par exemple $n \geq 30$.

Simulation de la convergence de $\frac{S_n}{n}$

Un générateur aléatoire, par exemple une pièce de monnaie équilibrée, permet de simuler une suite de n tirages indépendants de « faces » et de « piles » ; au i -ème tirage, on associe la v.a. B_i égale à 0 si on obtient « face », et 1 si on obtient « pile ». On a donc : $P(B_i = 0) = P(B_i = 1) = \frac{1}{2}$.

On observe que la v.a. $\frac{S_n}{n}$ fluctue autour de sa valeur moyenne 0,5, avec une variance décroissante égale à $V\left(\frac{S_n}{n}\right) = \frac{1}{n^2} V(\sum_i B_i) = \frac{1}{n} V(B_i) = \frac{1}{4n}$. Nous verrons plus loin que le théorème central limite, qui met en œuvre la convergence en loi, permet de démontrer que la loi des amplitudes des fluctuations de $\frac{S_n}{n}$ autour de sa moyenne converge vers loi gaussienne.

5.1 ♠ Démontrer le théorème de Bernoulli, en utilisant l'inégalité de Tchebychev (théorème 2.9), que nous rappelons : si Y est une v.a. d'espérance m , de variance σ^2 , alors $\forall \varepsilon > 0, P(|Y - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$.

Au début du XIX^e siècle, Poisson étendit le théorème de Bernoulli au cas où les v.a. indépendantes $(B_i)_{i=1, \dots, n}$ ne sont pas de loi identique : $\forall i = 1, \dots, n, P(B_i = 1) = p_i$. Il démontra que la loi des grands nombres subsistait : $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P \left[\left| \frac{S_n}{n} - \bar{p}_n \right| < \varepsilon \right] = 1$, où $\bar{p}_n = \frac{\sum_{i=1}^n p_i}{n}$.

Intermède biographique : de Jacques Bernoulli (1654-1705) à Carl Friedrich Gauss (1777-1855)

Premier du nom d'une illustre lignée de mathématiciens d'origine hollandaise, Jacques Bernoulli exploita, comme de nombreux mathématiciens avant lui, les séquences obtenues au jeu de pile ou face, et mit ainsi en évidence l'identité asymptotique entre la probabilité *a priori* d'un événement et la limite de sa fréquence d'apparition, résultat qui anticipait la très importante Loi

des grands nombres, figurant dans son livre *Ars conjectandi* (*L'art des conjectures*). Il s'agissait de la première tentative de fondement empirique du concept de probabilité.

En 1718, dans son ouvrage *Doctrine of chances* dans lequel sont définies les notions d'indépendance d'événements, de fonction génératrice, de probabilités conditionnelles et de probabilités composées, le mathématicien français huguenot Abraham de Moivre (1667-1754), exilé à Londres, prolongea la loi des grands nombres par la première formulation du théorème de la limite centrale, auquel Laplace (1749-1827) puis Gauss devaient donner une forme plus générale. C'est ainsi qu'apparut la première loi de probabilité définie par une densité, connue à présent comme loi de Gauss ou loi normale, ainsi qualifiée à cause de l'usage fréquent qui en est fait. Laplace y voyait une preuve supplémentaire de l'existence de Dieu et de son infinie bonté, puisqu'il avait doté l'humanité d'une loi universelle permettant de maîtriser le hasard.

La convergence en probabilité fut mise en œuvre pour la première fois par J. Bernoulli.

Définition 5.1 Une suite de v.a. $(X_n)_n$ **converge en probabilité** vers la v.a. X si :

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1.$$

(On notera : $X_n \rightarrow_{(P)} X$.)

5.2 Convergence en loi

Définition et caractérisations de la convergence en loi, dite aussi convergence étroite

Définition 5.2 Une suite de variables ou de vecteurs aléatoires $(X_n)_n$ de dimension k définis sur (Ω, \mathcal{T}, P) **converge en loi** vers la variable ou le vecteur aléatoire X quand n tend vers $+\infty$ (on notera : $X_n \rightarrow_{(L)} X$), si l'une des deux propositions équivalentes suivantes est vérifiée :

(1) Pour toute fonction bornée et continue h de \mathbb{R}^k dans \mathbb{R} :

$$E(h(X_n))_{(n \rightarrow +\infty)} \rightarrow E(h(X)).$$

(2) La limite ponctuelle de la fonction de répartition F_{X_n} quand n tend vers l'infini est F_X , en tout point où F_X est continue.

(2.a) Si $(X_n)_n$ est une suite discrète et X une v.a. discrète :

$$(X_n \rightarrow_{(L)} X) \iff (\forall x_k, \lim_{n \rightarrow +\infty} P_{X_n}(x_k) = P_X(x_k)).$$

(2.b) Si les lois des X_n et de X sont définies par des densités f_{X_n} et f_X , la caractérisation précédente (2) équivaut à la convergence ponctuelle de la suite $(f_{X_n})_n$ vers f_X .

Théorème 5.2 La convergence en probabilité d'une suite de v.a. entraîne sa convergence en loi. La réciproque est vraie dans le seul cas où la limite de la suite est constante (p.s.).

Les fonctions caractéristiques et génératrices sont à la base des théorèmes de convergence particulièrement efficaces et d'un usage fréquent en pratique.

Théorème 5.3 (de Lévy) Soient une suite de variables ou de vecteurs aléatoires $(X_n)_n$ et la suite de leurs fonctions caractéristiques $(\Phi_{X_n})_n$:

$$(X_n \rightarrow_{(L)} X) \iff (\forall x, \lim_{n \rightarrow +\infty} \Phi_{X_n}(x) = \Phi_X(x) \text{ et } \Phi_X \text{ continue en } 0).$$

Théorème 5.4 Soient une suite $(X_n)_n$ de v.a. à valeurs entières positives, g_{X_n} et g_X les fonctions génératrices de X_n et de X :

$$(X_n \rightarrow_{(L)} X) \iff (\forall s, |s| \leq 1, \lim_{n \rightarrow +\infty} g_{X_n}(s) = g_X(s)).$$

Applications des deux théorèmes précédents :

(1) La suite de v.a. $(X_n)_n$ de lois respectives $\mathcal{B}\left(n; \frac{\lambda}{n}\right)$ (on suppose n tel que $\frac{\lambda}{n} < 1$) converge en loi vers la v.a. de Poisson $\mathcal{P}(\lambda)$. Prouvons ce résultat :

$$g_{X_n}(s) = \sum_{k=0}^n C_n^k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} s^k = \left(1 - \frac{\lambda}{n} + \frac{\lambda s}{n}\right)^n,$$

converge vers $\exp(\lambda(s-1))$ fonction génératrice de la loi de Poisson $\mathcal{P}(\lambda)$.

Ce résultat est aussi vrai pour une suite de v.a. $\mathcal{B}(n; p_n)$, si $\lim_{n \rightarrow +\infty} n.p_n = \lambda$.

(2) Démonstration du théorème de Moivre-Laplace ci-dessous (théorème 5.5), comme illustration du théorème de Lévy.

Remarque 5.1 L'approximation de Poisson de lois binomiales est valide dès que $n > 30$ et $np < 5$; l'avantage numérique de travailler avec une v.a. de Poisson, plutôt qu'avec une v.a. binomiale, est très appréciable quand n est grand. Montrons-le sur un exemple : soit un lot de cent ampoules, chacune a la probabilité $1/200$ d'être défectueuse. Calculons la probabilité pour que dix ampoules soient défectueuses :

$$\lambda = \frac{1}{2} \text{ et } P(X = 10) = e^{-\frac{1}{2}} \times \left[\frac{1}{2}\right]^{10} \times \frac{1}{10!},$$

qui est moins lourd à calculer que $\frac{100!}{90!10!} \frac{1}{200^{10}} (1 - \frac{1}{200})^{90}$!

Intermède biographique : Paul Lévy (1886-1971)

Mathématicien français, ingénieur du Corps des Mines, Paul Lévy enseigna durant une quarantaine d'années à l'École des mines de Paris et à l'École polytechnique. Après des travaux en analyse fonctionnelle, il s'orienta vers la théorie des probabilités qu'il enrichit de nombreux résultats : l'utilisation de la fonction caractéristique comme outil puissant, une nouvelle démonstration du théorème de la limite centrale, le développement de la théorie des fonctions aléatoires à accroissements indépendants et surtout l'approfondissement de la théorie du processus brownien.

5.2 ♠ Soient X_1, X_2, \dots, X_n des v.a. indépendantes de loi uniforme sur $[-1, 1]$ et $T_n = \frac{1}{n} \sum_{i=1}^n (X_i)^{-1}$.

(1) Calculer la fonction caractéristique de $(X_i)^{-1}$, au voisinage de $t = 0$.
[Indication : $\int_0^{+\infty} \frac{1 - \cos y}{y^2} dy = \frac{\pi}{2}$.]

(2) En déduire la loi de $\frac{2T_n}{\pi}$ lorsque $n \rightarrow +\infty$.

(3) Montrer que $P(|\sum_{i=1}^n X_i^{-1}| > \frac{n\pi}{2}) \rightarrow \frac{1}{2}, n \rightarrow +\infty$.

5.3 ♠♠ Soient une suite de v.a. uniformes indépendantes $(U_n)_n$ de support $[0, 1]$, indépendante d'une suite $(X_n)_n$ de v.a. indépendantes $\mathcal{B}(n, p)$. Vers quelle v.a. converge en loi la suite $(nY_n)_{n \in \mathbb{N}}$ si $Y_n = \min_{(0 \leq i \leq X_n)} U_i$? [Cet exercice nécessite la connaissance de la formule de recomposition du théorème 6.3].

Théorème 5.5 (de Moivre-Laplace)

Étant donnée la suite de v.a. $(X_n)_{n \in \mathbb{N}}$ de loi $\mathcal{B}(n, p)$; quand n est grand, la loi de $\frac{X_n}{n}$ est approchée par la loi $\mathcal{N}(p, \frac{(1-p)p}{n})$.

Preuve Appliquons le théorème de Lévy : pour démontrer que $Y_n = \frac{X_n - np}{\sqrt{npq}}$ converge en loi vers la v.a. $\mathcal{N}(0, 1)$, il suffit de démontrer que Φ_{Y_n} converge simplement vers $\Phi_{\mathcal{N}(0,1)}$.

$$\Phi_{Y_n}(t) = e^{-\left[it \frac{np}{\sqrt{npq}} \right]} \Phi_{X_n} \left[\frac{t}{\sqrt{npq}} \right] \text{ où : } \Phi_{X_n}(u) = [\Phi_X(u)]^n ;$$

$$\Phi_{Y_n}(t) = e^{-\left[it \frac{np}{\sqrt{npq}} \right]} \left[pe^{i \frac{t}{\sqrt{npq}}} + (1-p) \right]^n ;$$

$$\lim_{n \rightarrow \infty} \Phi_{Y_n}(t) = e^{-\frac{t^2}{2}}, \text{ fonction caractéristique de la v.a. } \mathcal{N}(0, 1).$$

En pratique, le théorème de De Moivre est utile pour calculer des probabilités d'événements de loi binomiale $\mathcal{B}(n, p)$ où n est supérieur à 30. Les

tabulations de la loi binomiale n'existant pas pour de grandes valeurs de n , on utilise alors l'approximation suivante très satisfaisante dès que $n > 30$, np et $n(1-p) > 5$

$$P(X_n = k) = F_{X_n}(k+1) - F_{X_n}(k) \simeq F(k+1) - F(k) ;$$

où F est la fonction de répartition de la loi normale $\mathcal{N}(p, \frac{p(1-p)}{n})$. \square

Deux remarques importantes

(1) Une suite de v.a. discrètes peut converger en loi vers une v.a. absolument continue.

Illustration : la suite $(X_n)_n$ de variables discrètes uniformes de lois $P(X_n = \frac{k}{n}) = \frac{1}{n+1}$ pour $k = 0, 1, \dots, n$, converge en loi vers la v.a. X uniforme sur $[0, 1]$.

$\Phi_{X_n}(t) = \sum_{k=0}^n \frac{1}{n+1} e^{it \frac{k}{n}} = \frac{1 - e^{it \frac{(n+1)}{n}}}{(n+1)(1 - e^{it \frac{1}{n}})}$ converge vers $\frac{e^{it} - 1}{it}$, continue en 0, quand n tend vers $+\infty$: d'où le résultat par application du théorème de Lévy, puisque $\Phi_X(t) = \frac{e^{it} - 1}{it}$.

(2) Inversement, une suite de v.a. admettant des densités peut converger vers une v.a. discrète. C'est le cas d'une suite de v.a. $(X_n)_n$ dont les densités $(f_n)_n$ forment une suite d'unités approchées, définie en analyse fonctionnelle comme suite de fonctions positives d'intégrales égales à l'unité et telles que pour tout voisinage V de 0 si petit soit-il $\lim_{n \rightarrow +\infty} \int_{C_V} f_n(x) dx$ est égale à 0, converge en loi vers la v.a. constante 0.

Illustration : la suite d'unités approchées égales aux densités gaussiennes $f_{X_n}(x) = \frac{n}{\sqrt{2\pi}} \exp(-\frac{n^2 x^2}{2})$ de moyenne nulle et de variance égale à $\frac{1}{n^2}$ converge en loi vers la probabilité discrète δ_0 .

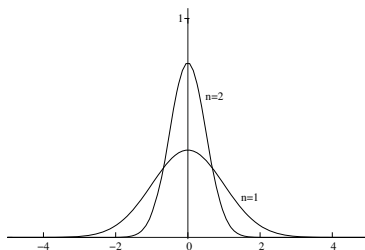


Figure 5.1 – Densités gaussiennes.

Théorème 5.6 Soient $(X_n)_n$ une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^k et f une fonction continue de \mathbb{R}^k dans \mathbb{R} . Si $(X_n)_n$ converge en loi (respectivement en probabilité) vers X , alors $(f(X_n))_n$ converge en loi (respectivement en probabilité) vers $f(X)$.

5.3 Théorèmes aux limites

La loi faible des grands nombres est une généralisation du théorème de Bernoulli (5.1) à la base de la méthode d'estimation de la moyenne en statistique inférentielle.

Théorème 5.7 (Loi faible des grands nombres)

Soit une suite $(X_n)_n$ de v.a. indépendantes, de même loi que X de classe $L^2(\mathbb{R})$ (i. e. : $E(|X^2|)$ fini), d'espérance m et de variance σ^2 , alors :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{(P)} m \quad \text{quand } n \rightarrow +\infty.$$

Preuve $V[\bar{X}_n] = \frac{\sigma^2}{n}$ et $E[\bar{X}_n] = m$, donc grâce au théorème de Tchebychev, si n tend vers $+\infty$, $P[|\bar{X}_n - m| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2}$ tend vers 0, $\forall \varepsilon > 0$. Conclusion : $\bar{X}_n \xrightarrow{(P)} m$. \square

Le théorème suivant ne suppose plus que les lois des X_i soient identiques.

Théorème 5.8 (Loi faible générale des grands nombres)

Soient les v.a. X_1, X_2, \dots, X_n , indépendantes, non nécessairement de même loi, de moyennes m_1, m_2, \dots, m_n et de variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ finies.

$$\left(\frac{1}{n} \sum_{i=1}^n m_i \xrightarrow{(n \rightarrow \infty)} m \text{ et } \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \xrightarrow{(n \rightarrow \infty)} 0 \right) \text{ implique :}$$

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{(n \rightarrow \infty)} m \quad \text{quand } n \rightarrow +\infty.$$

5.4 ♠ Démontrer le théorème précédent en appliquant l'inégalité de Tchebychev.

Les méthodes de Monte-Carlo

Ces méthodes qui portent le nom d'une des capitales européennes des jeux de hasard et d'argent, aujourd'hui bien désuète, sont à la base d'algorithmes de simulation statistique très puissants qui furent utilisés pour la première fois, à la fin de la Seconde Guerre mondiale par des chercheurs et des militaires américains confrontés aux problèmes posés par la mise au point de l'arme nucléaire. Il s'agissait de faciliter les calculs nécessaires à l'étude des réactions nucléaires, en vue de la fabrication de la première bombe atomique. Les

mathématiciens américains N. Metropolis (1925-1999) et S. Ulam (1909-1984) publièrent en 1949 le premier article de synthèse et produisirent en 1953 le premier algorithme de la méthode de Monte-Carlo. Aujourd'hui, ces méthodes sont utilisées dans les télécommunications (dimensionnement des centraux), en ingénierie financière, en physique nucléaire ou dans des méthodes de résolution d'équations aux dérivées partielles, et sont d'autant plus efficaces par rapport aux méthodes numériques que le nombre de variables mises en jeu dans le problème traité est grand. À l'instar des méthodes numériques, elles permettent de trouver une solution approchée à des problèmes n'ayant pas de solution analytique exacte, en utilisant toute la puissance des calculateurs vectoriels et parallèles. Pour une application élémentaire de la méthode, concernant l'estimation d'une intégrale simple, on pourra résoudre le problème 11. La méthode standard de Monte-Carlo est une application de la loi des grands nombres : étant donné un vecteur aléatoire $X = (X_1, X_2, \dots, X_k)$ de densité f_X , et une fonction réelle g définie sur le support de X , il s'agit d'obtenir une approximation de l'intégrale, $E(g(X)) = \int \dots \int_{\mathbb{R}^k} g(X) \cdot f_X(x) \cdot dx$, grâce à une simulation de la v.a. X , lorsque cette intégrale n'est pas analytiquement calculable et que les méthodes numériques ne sont pas satisfaisantes. Soit $X^{(i)} = (X_1^{(i)}, X_2^{(i)}, \dots, X_k^{(i)})_{i=1..n}$ l'échantillon obtenu par n simulations du vecteur X , et $Y^{(i)} = g(X^{(i)})$; la loi des grands nombres permet de conclure :

$$\lim_{n \rightarrow +\infty} \sum_{i=1}^n \frac{Y^{(i)}}{n} = E(g(X)).$$

Nous abordons maintenant l'un des théorèmes fondamentaux de la théorie des probabilités, dû aux travaux de Laplace puis de Gauss. Ce théorème fonde le caractère gaussien de nombreux phénomènes apparaissant dans des domaines aussi divers que la physique, la biologie ou l'économie.

Théorème 5.9 (Théorème de la limite centrale ou de Laplace-Gauss)

Soit la suite $(X_n)_n$ de v.a. indépendantes, de même loi, de moyenne m et de variance σ^2 , alors :

$$\frac{\bar{X}_n - m}{(\sigma/\sqrt{n})} \xrightarrow{(L)} \mathcal{N}(0, 1) \text{ quand } n \rightarrow +\infty.$$

Pratiquement, si n est au moins de l'ordre de la trentaine, on a les résultats asymptotiques équivalents :

$$(1) \quad \frac{\bar{X}_n - m}{(\sigma/\sqrt{n})} \sim \mathcal{N}(0, 1) ;$$

$$(2) \quad \bar{X}_n \sim \mathcal{N}\left(m; \frac{\sigma^2}{n}\right).$$

Interprétation : la moyenne de quantités aléatoires de même loi, de moyenne m et de variance σ^2 , mutuellement indépendantes, fluctue autour de

m , selon une loi gaussienne d'écart type $\frac{\sigma}{\sqrt{n}}$.

$$(3) \sum_{i=1}^n X_i \sim \mathcal{N}(nm; n\sigma^2).$$

L'équivalence asymptotique (3) ne supporte pas, contrairement aux autres, le passage à la limite quand n tend vers $+\infty$, car $V\left(\sum_{i=1}^n X_i\right) = n\sigma^2$ converge vers $+\infty$.

Preuve Utilisons le théorème de Lévy.

$$\Phi_{\bar{X}_n}(u) = \left(\Phi_{\frac{X}{n}}(u)\right)^n = \left(\Phi_X\left(\frac{u}{n}\right)\right)^n.$$

Par le développement limité à l'ordre deux, il vient :

$$\Phi_X\left(\frac{u}{n}\right) = \Phi_X(0) + \frac{u}{n}\Phi_X'(0) + \frac{u^2}{2n^2}\Phi_X''(0) + o\left(\frac{u^2}{n^2}\right).$$

Or, $\Phi_X(0) = 1$, $\Phi_X'(0) = im$ et $\Phi_X''(0) = -(m^2 + \sigma^2)$.

Calculons $\ln(\Phi_{\bar{X}_n}(u))$, en considérant que $\ln(1 + \varepsilon) \approx \varepsilon - \frac{\varepsilon^2}{2}$ si ε proche de 0 : pour n grand, $\ln(\Phi_{\bar{X}_n}(u)) = ium + \frac{u^2}{2n}(-\sigma^2) + o\left(\frac{1}{n^2}\right)$.

D'où $\Phi_{\bar{X}_n}(u) \approx e^{ium - \frac{u^2\sigma^2}{2n}}$ qui est la fonction caractéristique de la v.a. $\mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$, d'où la conclusion attendue, grâce au théorème de Lévy. \square

(Ce théorème remarquable fait de la variable gaussienne un attracteur des sommes de variables indépendantes de même loi ; d'un point de vue fonctionnel, il résulte des propriétés du produit de convolution de mesures identiques positives de masse unitaires.)

Le théorème de la limite centrale, qu'on pourra désigner par l'abréviation TLC ou TCL, se généralise aux suites de vecteurs aléatoires :

Théorème 5.10 *Soit la suite de vecteurs $(X_n)_n$ de dimension d , indépendants, de même loi de probabilité, d'espérance $m = (m_1, \dots, m_d)$ et de matrice de covariance Γ :*

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^n X_k - n.m \right) \xrightarrow{(L)} \text{vecteur } \mathcal{N}(0, \Gamma) \text{ quand } n \rightarrow +\infty.$$

$\mathcal{N}(0, \Gamma)$ désigne le vecteur gaussien d'espérance nulle et de matrice de covariance Γ .

Mise en évidence de la rapidité de la convergence en loi vers la loi normale, sous les hypothèses du théorème de la limite centrale.

Soit la suite de v.a. indépendantes $(X_n)_n$, toutes uniformément distribuées sur $[0, 1]$, on sait que $E(X_n) = \frac{1}{2}$, $V(X_n) = \frac{1}{12}$; posons $S_n = \sum_{k=1}^n X_k$.

On démontre que $f_{S_2}(x) = (1 - |x - 1|) \mathbf{1}_{[0,2]}(x)$ et plus généralement :

$$f_{S_n}(x) = \frac{1}{(n-1)!} \left(x^{n-1} \mathbf{1}_{[0,n]}(x) - C_n^1 (x-1)^{n-1} \mathbf{1}_{[1,n]}(x) + \dots \right. \\ \left. (-1)^k C_n^k (x-k)^{n-1} \mathbf{1}_{[k,n]}(x) \right).$$

f_{S_n} est donc un polynôme de degré $(n-1)$ sur chaque intervalle $]k, k+1]$, donc de classe \mathcal{C}^{n-2} (en chaque point k , il y a ajustement des polynômes jusqu'à l'ordre $(n-2)$ de dérivation).

La suite $(Z_n)_n$ définie par $Z_n = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$ converge en loi vers la v.a. gaussienne centrée-réduite.

On constate que dès que n est supérieur à sept, l'approximation de Z_n par la v.a. gaussienne $\mathcal{N}(0, 1)$ est quasi parfaite : la résolution des écrans ne permet pas de distinguer leurs densités respectives ! C'est dire la rapidité de la convergence en loi vers la loi normale mise en œuvre par le théorème central limite.

5.5 ♠♠ Inégalité des grandes déviations

(1) Soit X une v.a. positive ayant une espérance $E(X)$; démontrer l'inégalité de Markov :

$$\forall \lambda > 0 \quad E(X) \geq \lambda P(X \geq \lambda).$$

(2) Si X est la v.a. de Bernoulli $P(X = 1) = P(X = -1) = 0,5$, démontrer que pour tout α positif :

$$E(e^{\alpha X}) \leq e^{\frac{\alpha^2}{2}}.$$

(3) Soit la suite $(X_n)_n$ de v.a. indépendantes de même loi que la v.a. de Bernoulli X et $S_n = \sum_{i=1}^n X_i$. Démontrer que pour tout α positif :

$$P(S_n > \lambda) \leq e^{-\alpha \lambda} E(e^{\alpha S_n}) \leq e^{\frac{n\alpha^2}{2} - \alpha \lambda}.$$

En déduire l'inégalité des grandes déviations :

$$P(S_n > \lambda) \leq e^{-\frac{\lambda^2}{2n}},$$

qui est bien meilleure que l'inégalité de Tchebychev.

5.6 ♠ Dans un programme de calcul, on décide d'utiliser k chiffres significatifs après la virgule et d'arrondir tous les résultats à $\frac{1}{2} 10^{-k}$ près. On suppose que

l'on effectue 10^6 opérations successives, et que les erreurs commises pour chacune sont indépendantes, de loi uniforme sur $[-a, a]$, où $a = \frac{1}{2}10^{-k}$. On note S l'erreur commise sur le résultat final, et on veut calculer $P(|S| \leq 10^3 a)$. On note X_i l'erreur commise sur la i -ème opération (donc $S = \sum_{i=1}^{10^6} X_i$); en considérant 10^6 comme « grand », démontrer que $\sqrt{3}S/10^3 a$ suit approximativement une loi gaussienne $\mathcal{N}(0, 1)$ et conclure.

Des théorèmes apparentés au T.C.L dont les hypothèses sont moins restrictives en s'affranchissant par exemple de l'hypothèse d'identité des lois des v.a. X_i , mettent en évidence, plus encore que le T.C.L., le caractère fortement attractif de la v.a. gaussienne comme limite de sommes de v.a. indépendantes de lois quelconques. Tel est le cas du théorème suivant :

Théorème 5.11 (de Lindeberg-Feller)

Soit $(X_n)_n$ une suite de v.a. mutuellement indépendantes non nécessairement de même loi, d'espérances $(m_n)_n$ et de variances $(\sigma_n^2)_n$. Soit F_n la fonction de répartition de $(X_n - m_n)$ et $\theta_n^2 = \sum_{i=1}^n \sigma_i^2$.

Si, $\forall \varepsilon > 0 \lim_{n \rightarrow +\infty} \left(\frac{1}{\theta_n^2} \sum_{i=1}^n \int_{|x| > \varepsilon \theta_n} x^2 dF_i(x) \right) = 0$ alors :

$$\frac{\sum_{i=1}^n (X_i - m_i)}{\theta_n} \xrightarrow{(L)} N(0, 1), \text{ si } n \rightarrow +\infty.$$

Interprétation : si les $\frac{X_n - m_n}{\theta_n}$ sont uniformément petits, leur somme converge vers une v.a. normale $\mathcal{N}(0, 1)$. Ainsi, une v.a. résultante additivement d'un grand nombre de causes indépendantes, peut être considérée comme gaussienne. Plus encore que le TCL, le théorème de Lindeberg-Feller, dont les hypothèses sont plus réalistes, justifie l'omniprésence du caractère gaussien de nombreux phénomènes.

5.7 ♠♠ Soit une suite U_1, U_2, \dots, U_n de v.a. indépendantes de même loi uniforme sur $[0, 1]$. Étudier la convergence en loi des suites :

- (1) $Y_n = (\prod_{i=1}^n U_i)^{\frac{\alpha}{n}}$ où $\alpha > 0$;
- (2) $Z_n = e^{\alpha \sqrt{n}} (U_1 U_2 \dots U_n)^{\frac{\alpha}{\sqrt{n}}}$.

Intermède biographique : Pierre-Simon de Laplace (1749-1827)

Fils d'un fermier normand, Laplace commença sa formation dans un collège bénédictin local et poursuivit ses études à Caen puis à Paris, où le philosophe et mathématicien Jean Lerond d'Alembert (1717-1783) reconnaissant son génie, le nomma professeur à l'École militaire. Il devint peu après professeur dans la toute récente École polytechnique et participa à la réforme des poids et

mesures. Considéré avec Fourier (1768-1830) comme l'un des fondateurs de la physique mathématique, l'essentiel de son œuvre concerne le développement et les applications de l'analyse à la mécanique céleste et aux probabilités. Laplace rassemble et synthétise dans sa *Théorie analytique des probabilités* parue en 1812 les nombreux travaux de ses prédécesseurs : Pascal, Bernoulli, de Moivre et Bayes. Il affine leurs résultats et accroît leur théorisation en introduisant de nouveaux concepts et de nouvelles méthodes, comme l'indépendance d'événements, la règle des probabilités composées, la notion de loi limite et la méthode des moindres carrés si précieuse pour l'approximation et le traitement des données expérimentales. Prolongeant les travaux de Bernoulli, il fut le premier à étudier les suites de v.a. et proposa la notion de loi limite, ainsi qu'une théorie des erreurs qui sera approfondie ensuite par Gauss.

5.4 Convergence presque sûre

Bien qu'il s'agisse de la notion de convergence la plus naturelle, analogue à la convergence simple des suites de fonctions, son intérêt pratique est moindre que celui des autres types de convergence.

Définition 5.3 $(X_n)_n$ *converge presque sûrement vers* X si :

$$P\left(\left\{\omega \text{ tels que } \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\right\}\right) = 1 \quad (\text{On note : } X_n \xrightarrow{(p.s.)} X).$$

Théorème 5.12 *La convergence presque sûre d'une suite de v.a. entraîne sa convergence en probabilité et donc sa convergence en loi.*

Théorème 5.13 (Loi forte des grands nombres)

Si $(X_n)_n$ est une suite de v.a. indépendantes, de même loi que X telle que $E(|X|)$ soit finie, alors :

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{(p.s.)} E(X).$$

La loi forte des grands nombres s'applique aussi aux vecteurs aléatoires si l'on remplace dans son énoncé $E(|X|)$ par $E(\|X\|)$ où $\|X\|$ est la norme euclidienne du vecteur X .

Théorème 5.14 (Loi forte des grands nombres dans le cas de lois différentes)

Soit $(X_n)_n$ une suite de v.a. mutuellement indépendantes, de lois distinctes, d'espérances $(\mu_n)_n$ et de variances $(\sigma_n^2)_n$:

$$\left(\sum_{i=1}^{+\infty} \frac{\sigma_i^2}{i^2} \text{ convergente} \right) \implies \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{(p.s.)} 0\right).$$

5.8 ♠ Soit X de loi gamma $\gamma(p, \theta)$, de densité :

$$f_X(x) = \frac{1}{\Gamma(p)} e^{-\theta x} \theta^p x^{p-1} \mathbf{1}_{\mathbb{R}_+}(x).$$

(1) Utiliser l'expression de la fonction caractéristique d'une somme de v.a. indépendantes pour calculer la loi de $S_n = \sum_i X_i$, X_i de même loi que X .

(2) Loi asymptotique de S_n quand n grand ?

(3) Soit $Y_n = \frac{1}{n} \sum_{i=1}^n X_i^2$. Calculer la limite presque sûre de Y_n , quand n tend vers $+\infty$.

Pour sacrifier ne serait-ce qu'une fois dans ce cours à la vulgate mathématique de l'économie libérale, voici un exercice de modélisation financière !

5.9 ♠♠ Modèle simplifié d'évolution du prix d'une action en Bourse

On modélise le prix V_n d'une action, à l'instant $n.u$, où u est une unité de temps fixée (par exemple, un jour), par l'équation de récurrence :

$$V_n = (1 + r + X_n v) V_{n-1}$$

où r est le taux d'actualisation supposé constant, v la volatilité et $(X_n)_{n=1,2,\dots}$ une suite de v.a. indépendantes équidistribuées de loi $\frac{1}{2}(\delta_1 + \delta_{-1})$. La volatilité v est supposée vérifier : $0 < v < 1 + r$, $V_0 > 0$. (La volatilité est la variabilité à court terme des taux d'intérêt.)

(1) Montrer que l'espérance de $\ln(1 + r + X_n v)$ vaut $\frac{1}{2} \ln((1+r)^2 - v^2)$; en déduire que la suite $(\frac{1}{n} \ln(V_n))_n$ converge presque sûrement vers m . Traiter les deux cas : $m < 0$ et $m > 0$.

(2) Montrer que la variance σ^2 de $\ln(1 + r + X_n v)$ est égale à $\frac{1}{4} \ln\left(\left(\frac{1+r+v}{1+r-v}\right)\right)^2$; appliquer le TCL pour déterminer la loi de $\lim_{n \rightarrow +\infty} V_n$.

5.5 Convergence en moyenne quadratique

Ce type de convergence est adapté à l'étude des signaux et des processus aléatoires d'ordre deux, qui font l'objet du chapitre 10.

Définition 5.4 Une suite de v.a. $(X_n)_n$ d'ordre deux **converge en moyenne quadratique**, vers la v.a. X d'ordre deux, si :

$$E((X_n - X)^2) \xrightarrow{n \rightarrow +\infty} 0. \text{ (On note : } X_n \xrightarrow{(m.q.)} X \text{).}$$

Théorème 5.15 (Convergence en moyenne quadratique vers une constante)

Soit la suite $(X_n)_n$ de variables d'ordre deux :

$$\left(\lim_{(n \rightarrow +\infty)} E(X_n) = m \text{ et } \lim_{(n \rightarrow +\infty)} V(X_n) = 0 \right) \implies (X_n \xrightarrow{(m.q.)} m).$$

5.10 ♠ Prouver le théorème.

Le théorème suivant est une version moins restrictive de la loi des grands nombres, ne nécessitant que l'hypothèse de décorrélation des X_n .

Théorème 5.16 *Soit la suite de v.a. $(X_n)_n$ décorrélées deux à deux, de même moyenne m , telle que $\exists M > 0, \forall n \in \mathbb{N}, \sigma^2(X_n) < M$, alors :*

$$S_n/n \xrightarrow{(m.q.)} m.$$

Preuve

$$\begin{aligned} E \left[\left(\frac{S_n}{n} - m \right)^2 \right] &= \frac{1}{n^2} E \left[\left(\sum_{i=1}^n (X_i - m) \right)^2 \right] \\ &= \frac{1}{n^2} E \left[\sum_{i=1}^n (X_i - m)^2 + 2 \sum_{p \neq q} (X_p - m)(X_q - m) \right] \leq \frac{nM}{n^2}. \end{aligned}$$

D'où la conclusion quand n tend vers $+\infty$. \square

Théorème 5.17 *La convergence en moyenne quadratique d'une suite de v.a. $(X_n)_n$ entraîne sa convergence en probabilité et donc sa convergence en loi.*

Preuve L'inégalité de Jensen (voir théorème 8, chapitre 2) entraîne :

$$(E(X_n - X))^2 \leq E((X_n - X)^2).$$

Or, $X_n \xrightarrow{(m.q.)} X$, donc $E(|X_n - X|) \rightarrow 0$ si $n \rightarrow \infty$, et sachant que :

$$|E(X_n - X)| \leq E(|X_n - X|),$$

$\lim_{n \rightarrow +\infty} E(X_n) = E(X)$. D'autre part :

$$V(X_n - X) = E((X_n - X)^2) - (E(X_n - X))^2,$$

donc $V(X_n - X) \rightarrow 0$ si $X_n \xrightarrow{(m.q.)} X$.

L'inégalité de Tchebychev $P(|X_n - X| > \varepsilon) \leq \frac{V(X_n - X)}{\varepsilon^2}$ permet de conclure.

\square

5.11 ♠♠ (1) Soit la suite de vecteurs aléatoires indépendants $Z_n = (X_n, Y_n)$, de même loi normale dont les composantes sont centrées réduites et de coefficient de corrélation ρ ; on admet que $V(R_n) = \frac{cte}{n}$.

(1) Étudier les convergences quadratique et presque sûre de la suite $(R_n)_n$ définie par : $R_n = \frac{1}{n} \sum_{i=1}^n X_i Y_i$.

(2) Soit une suite $(X_n)_n$ de v.a. indépendantes de même loi de Cauchy et $S_n = \sum_{i=1}^n X_i$. Les suites $(U_n = \frac{S_n}{n})_n$ et $(V_n = \frac{S_n}{\sqrt{n}})_n$ convergent-elles, et en quel sens ?

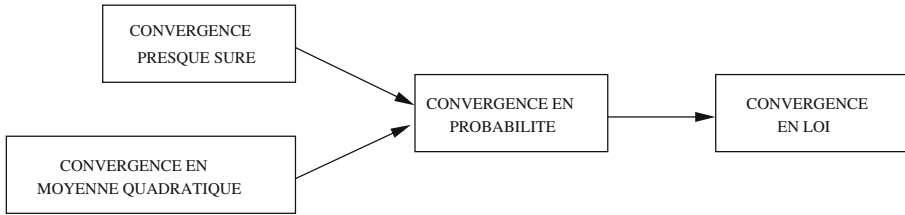


Figure 5.2 – Implications entre les divers types de convergence.

A propos du hasard en physique

(A) Le hasard en physique statistique

Dès le début du XVIII^e siècle, des physiciens observant au microscope l'agitation permanente et erratique de particules en suspension dans un liquide chauffé, pressentirent que de nombreux phénomènes d'apparence déterministe résultaient de la combinaison de micro phénomènes de nature aléatoire qui leur étaient sous-jacents : tel était le cas, par exemple, de la pression d'un gaz, conséquence de l'effet statistique moyen d'une multitude de chocs aléatoires d'amplitudes infimes à l'échelle atomique. Pour expliquer la pression et la viscosité des gaz par le seul mouvement moléculaire, le physicien anglais James-Clerk Maxwell (1831-1879) acquiesça au calcul des probabilités, postula l'équipartition aléatoire des positions et des vitesses des molécules autour de leurs moyennes, et formula ainsi l'hypothèse fondatrice de la théorie cinétique des gaz. C'est dans ce contexte qu'il appliqua avec succès la théorie des probabilités à la modélisation du comportement statistique moléculaire.

Ludwig Boltzmann (1844-1906), thermodynamicien allemand, enrichit la physique statistique d'une hypothèse qui s'est révélée être l'une des plus profondes sur la nature des grands systèmes physiques, en postulant qu'un gaz évolue toujours vers sa situation la plus probable. Dans son œuvre d'une grande originalité, le hasard a le statut d'un principe explicatif, au même titre que l'omniprésent principe de causalité qui dominait encore la physique de son époque. L'œuvre de Boltzmann, restée longtemps incomprise, fut prolongée et enrichie par J.W Gibbs (1839-1903) à des systèmes mécaniques très généraux, dont l'impact est aujourd'hui considérable, notamment en physique statistique quantique. À la fin du XIX^e siècle coexistaient deux théories apparemment irréconciliables : la mécanique newtonienne de nature déterministe, dont les lois vérifient le principe de réversibilité, et la physique statistique dont l'irréversibilité des lois est d'origine aléatoire. En physique statistique, la théorie des probabilités pallie notre incapacité à connaître tous les paramètres descriptifs d'un système constitué d'un grand nombre de composants.

(B) Le hasard en mécanique quantique

Contrairement au rôle qu'il joue en physique statistique, le hasard semble être une caractéristique intrinsèque du monde quantique. En effet, les probabilités n'y expriment pas des erreurs de mesure dues aux limitations de nos moyens d'investigation et ne relèvent pas de la méconnaissance des processus sous-jacents.

Dans les années 1930, deux conceptions de la réalité s'affrontaient au sein de la communauté des physiciens : la conception réaliste pour qui la réalité existe indépendamment de la connaissance que nous en donnent les théories physiques et la conception pragmatiste qui renonce à atteindre une réalité première. Pour Einstein (1879-1955), figure de proue de l'école réaliste, toute théorie physique est une représentation déterministe et aussi complète que possible de la réalité observée ; pour lui, l'intrusion du hasard en physique quantique n'est qu'un pis-aller, résultant de la méconnaissance de variables cachées inaccessibles à cause de la limitation provisoire des modèles théoriques et des moyens d'observation et d'expérimentation du moment. En opposition aux réalistes, l'école de Copenhague, dont Niels Bohr (1885-1962) fut le chef de file, pensait qu'il fallait s'en tenir à une conception instrumentaliste de la science fondée sur la seule exigence de capacités descriptives et prédictives fondées sur les modèles déterministes et probabilistes. Ces débats toujours d'actualité n'ont pas permis de trancher sur la nature du hasard : est-il une propriété de la réalité physique et des relations qui y sont mises en œuvre ou bien résulte-t-il des limites de nos moyens théoriques et pratiques d'investigation de la réalité ?

Les problèmes 10, 11 et 12 se rapportent à ce chapitre.

5.6 Corrigés des exercices

Exercice 5.1

Soit $Y_n = \frac{S_n}{n}$; $\mathbb{E}(Y_n) = p$, $V(Y_n) = \frac{1}{n^2}p(1-p)n = \frac{p(1-p)}{n}$. Appliquons l'inégalité classique de Tchebychev :

$$\text{pour tout } \varepsilon > 0, P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2}, \text{ donc } \lim_{n \rightarrow +\infty} P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) = 0.$$

On en déduit la convergence en probabilité de Y_n vers p .

Exercice 5.2

$$\begin{aligned} (1) \quad \phi_{1/X_j}(t) &= \frac{1}{2} \int_0^1 \left(e^{it/x} + e^{-it/x} \right) dx = \int_0^1 \cos\left(\frac{t}{x}\right) dx \\ &= |t| \int_{|t|}^{+\infty} \frac{\cos y}{y^2} dy = 1 - |t| \int_{|t|}^{+\infty} \frac{1 - \cos y}{y^2} dy, \end{aligned}$$

où on a posé : $x = \frac{|t|}{y}$. Donc, pour $t \rightarrow 0$, compte tenu de l'indication, $\phi_{1/X_j}(t) = 1 - \frac{\pi}{2}|t| + o(t)$.

(2) On en déduit : $\phi_{T_n}(t) = \left(1 - \frac{\pi}{2} \frac{|t|}{n} + o\left(\frac{1}{n}\right)\right)^n$ converge, quand n tend vers $+\infty$, vers $e^{-\frac{\pi}{2}|t|}$ qui est la fonction caractéristique d'une variable de Cauchy centrée.

$$(3) \quad \lim_{n \rightarrow +\infty} P\left(\left[\frac{2|T_n|}{\pi} > 1\right]\right) = \frac{2}{\pi} \int_0^{+\infty} \frac{du}{1+u^2} = \frac{1}{2}.$$

Exercice 5.3

$$\begin{aligned} P([Y_n \leq y]) &= \sum_{k=0}^n P([Y_n \leq y] | [X_n = k]) P([X_n = k]) \\ &= \sum_{k=0}^n P\left(\left[\min_{0 \leq i \leq k} (U_i) \leq k\right]\right) P([X_n = k]). \end{aligned}$$

$$F_{\min_{0 \leq i \leq k} (U_i)}(y) = 1 - P\left(\left[\min_{0 \leq i \leq k} (U_i) > y\right]\right) = 1 - (1 - F_U(y))^{k+1} \text{ d'où}$$

$$F_{Y_n}(y) = \sum_{k=0}^n (1 - (1-y)^{k+1}) \binom{n}{k} p^k (1-p)^{n-k} = 1 - (p(1-y) + 1-p)^n (1-y).$$

Sachant que $F_{aX}(x) = F_X\left(\frac{x}{a}\right)$ pour $a > 0$, on en déduit :

$$F_{\min_{0 \leq i \leq X_n} U_i}(y) = 1 - \left(p\left(1 - \frac{y}{n}\right) + 1 - p\right)^n \left(1 - \frac{y}{n}\right) \xrightarrow{n \rightarrow +\infty} 1 - e^{-py},$$

qui est la fonction de répartition d'une v.a. de loi exponentielle $\mathcal{E}(p)$.

Exercice 5.4

$$\mathbb{E}\left(\frac{S_n}{n}\right) = \frac{\sum_{i=1}^n \mu_i}{n}; \quad V\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2.$$

Par l'inégalité de Tchebychev : pour tout $\varepsilon > 0$, $P(|\bar{X}_n - \mathbb{E}(\bar{X}_n)| > \varepsilon) \leq \frac{1}{\varepsilon^2} \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2$ et, par passage à la limite quand $n \rightarrow +\infty$, $P(|\bar{X}_n - m| > \varepsilon) \rightarrow 0$. D'où la convergence en probabilité de \bar{X}_n vers m : il s'agit d'une généralisation de la loi faible des grands nombres.

Exercice 5.5

$$(1) \mathbb{E}(X) = \int_0^{+\infty} t dP_X(t) \geq \int_\lambda^{+\infty} t dP_X(t) \geq \lambda \int_\lambda^{+\infty} dP_X(t) = \lambda P([X \geq \lambda]).$$

$$(2) \mathbb{E}(e^{\alpha X}) = \frac{1}{2}(e^\alpha + e^{-\alpha}) = \text{ch}(\alpha) \text{ donc } \mathbb{E}(e^{\alpha X}) \leq e^{\alpha^2/2}.$$

(3) D'après (1) : $P([S_n > \lambda]) = P([e^{\alpha S_n} > e^{\alpha \lambda}]) \leq e^{-\alpha \lambda} \mathbb{E}(e^{\alpha S_n})$ et d'après l'indépendance des X_i , on a alors $\mathbb{E}(e^{\alpha S_n}) = \mathbb{E}(e^{\alpha X})^n \leq e^{n\alpha^2/2}$, d'où l'on déduit :

$$P([S_n > \lambda]) \leq e^{-\alpha \lambda} \mathbb{E}(e^{\alpha S_n}) \leq e^{\frac{n\alpha^2}{2} - \alpha \lambda}.$$

Pour $\alpha = \frac{\lambda}{n}$, on parvient à la majoration : $P([S_n > \lambda]) \leq e^{-\frac{\lambda^2}{2n}}$.

Exercice 5.6

X_i erreur associée à la i -ème opération, $S = \sum_{i=1}^{10^6} X_i$. Les X_i étant indépendantes, de même loi, d'après le théorème central limite :

$$\frac{S - E(S)}{\sigma_S} \text{ est de loi } \mathcal{N}(0; 1).$$

Or $\mathbb{E}(S) = 10^6 \mathbb{E}(X_i)$, avec $\mathbb{E}(X_i) = 0$, car X_i , erreur due à l'arrondi à $\frac{1}{2} 10^{-k}$ près, suit la loi uniforme sur $]-\frac{1}{2} 10^{-k}, \frac{1}{2} 10^{-k}[$, i.e. $]-a, a[$.

$V(S) = 10^6 V(X_i)$ (variables indépendantes), et $V(X_i) = \mathbb{E}(X_i^2) = \frac{1}{2a} \int_{-a}^a x^2 dx = \frac{a^2}{3}$, donc $\sqrt{V(S)} = 10^3 \frac{a}{\sqrt{3}}$ et $\frac{\sqrt{3}S}{10^3 a} \approx \mathcal{N}(0, 1)$. On a alors :

$$\begin{aligned} P(|S| \leq 10^3 a) &= P\left(\left[\frac{|S|}{10^3 a} \leq 1\right]\right) = P\left(\left[\frac{|S\sqrt{3}|}{10^3 a} \leq \sqrt{3}\right]\right) \\ &= \Phi(\sqrt{3}) - \Phi(-\sqrt{3}) = 2\Phi(\sqrt{3}) - 1. \end{aligned}$$

Or $\Phi(\sqrt{3}) \approx \Phi(1,732) \approx 0,958$, donc $P(|S| \leq 10^3 a) \approx 0,916$.

Exercice 5.8

- (1) La fonction caractéristique de S_n correspond à la loi $\gamma(p, n\theta)$.
 (2) Quand n devient grand, S_n est asymptotiquement proche (grâce au Théorème Central Limite) de la variable normale $\mathcal{N}\left(\frac{np}{\theta}, \frac{np}{\theta^2}\right)$.
 (3) D'après la loi forte des grands nombres, $Y_n = \frac{1}{n} \sum_{i=1}^n X_i^2$ converge presque sûrement vers $\mathbb{E}(X^2)$.

Exercice 5.10

Montrons que $\lim_{n \rightarrow +\infty} \mathbb{E}((X_n - m)^2) = 0$.

$$\mathbb{E}(X_n^2 + m^2 - 2mX_n) = V(X_n) + \mathbb{E}(X_n)^2 + m^2 - 2m\mathbb{E}(X_n) \xrightarrow{n \rightarrow +\infty} 0 + m^2 + m^2 - 2m^2 = 0.$$

Exercice 5.11

- (1) $\mathbb{E}(X_n Y_n) = \text{cov}(X_n, Y_n) + \mathbb{E}(X_n)\mathbb{E}(Y_n) = \rho$.

La loi forte des grands nombres (théorème 5.14) implique : $\frac{1}{n} \sum_{i=1}^n X_i Y_i = R_n$ converge presque sûrement vers $\mathbb{E}(X_n Y_n) = \rho$. La suite $(R_n)_n$ converge en moyenne quadratique vers ρ d'après le théorème 5.15.

(2) La suite $(U_n)_n$ n'est pas intégrable : la loi des grands nombres et le théorème central limite ne peuvent s'appliquer. Mais,

$$\phi_{U_n}(t) = \left(\phi_X \left(\frac{t}{n} \right) \right)^n = e^{-|t|} \text{ pour tout } t \in \mathbb{R},$$

donc (U_n) converge vers une variable de Cauchy identique. Pour tout $t \in \mathbb{R}$, $\phi_{V_n}(t) = e^{-\sqrt{n}|t|}$; $\phi_{V_n} \xrightarrow{n \rightarrow +\infty} \mathbf{1}_{\{0\}}$ qui est discontinue en 0 et ne peut pas être égale à une fonction caractéristique : $(V_n)_n$ ne converge pas en loi, et donc ne converge selon aucun autre mode de convergence (voir schéma des implications entre convergences en fin de chapitre).

Chapitre 6

Probabilités, lois et espérances conditionnelles

Un événement aléatoire A conditionne un événement B , si la réalisation de B dépend de celle de A . Nombreux sont les phénomènes physiques et les systèmes mis en œuvre dans les sciences de l'ingénieur, dont les états sont conditionnés. Dans ce chapitre, on construit les concepts et on expose les méthodes qui permettent de modéliser et de résoudre les problèmes de conditionnement les plus généraux.

6.1 Conditionnement d'événements

Soit un ensemble Ω de cardinal n , tel que $\Omega = \Omega_1 \cup \Omega_2$. Si l'on tire au hasard un élément dans Ω_2 , quelle est la probabilité p pour qu'il appartienne aussi à Ω_1 ?

Posons : $n_i = |\Omega_i|$ et $n_{1,2} = |\Omega_1 \cap \Omega_2|$.

$$p = \frac{|\Omega_1 \cap \Omega_2|}{|\Omega_2|} = \frac{n_{1,2}}{n_2} = \frac{n_{1,2}/n}{n_2/n} = \frac{P(\text{élément} \in \Omega_1 \cap \Omega_2)}{P(\text{élément} \in \Omega_2)}.$$

Cette formulation ensembliste se généralise au cas d'événements quelconques, selon la définition suivante :

Définition 6.1 *Étant donné deux événements A et B appartenant à la tribu \mathcal{T} définie sur un espace d'événements Ω , la **probabilité conditionnelle de B sachant A** (ou de B conditionné par A), où A est de probabilité non nulle, est définie par :*

$$P(B \text{ sachant } A) = \frac{P(A \cap B)}{P(A)} =_{(\text{noté})} P(B|A) \text{ ou } P^A(B).$$

Théorème 6.1 Deux événements quelconques A et B de probabilités non nulles vérifient :

$$P(A \cap B) = P(B|A).P(A) = P(A|B).P(B).$$

Définition 6.2 Deux événements A et B sont indépendants si $P^A(B) = P(B)$ ou, de façon équivalente, si $P^B(A) = P(A)$ ou encore $P(A \cap B) = P(A)P(B)$.

Exemple 6.1 Soient $\Omega = \{1, 2, \dots, 20\}$ et les événements $A = (\text{être un nombre premier})$ et $B = (\text{être un nombre de la forme } 3k + 1 \text{ où } k \in \mathbb{N})$. Démontrons que B conditionne A :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{|\Omega_B \cap \Omega_A|}{|\Omega_B|} = \frac{4}{7}, \text{ et}$$

$$P(A) = \frac{|\Omega_A|}{|\Omega|} = \frac{9}{20} \text{ différent de } P(A|B), \text{ donc } B \text{ conditionne } A.$$

6.1 ♠ Faisons l'hypothèse (optimiste) que deux politiciens mentent parfois, indépendamment l'un de l'autre, l'un disant la vérité trois fois sur quatre et l'autre quatre fois sur cinq. S'ils énoncent la même affirmation, quelle est la probabilité pour qu'elle soit vraie? [On posera : $E = \ll$ les deux politiciens énoncent la même assertion \gg et, pour $i = 1, 2$, $A_i : \ll$ le i -ème politicien ne ment pas \gg .]

Théorème 6.2 (Probabilité conditionnelle)

Étant donné un espace de probabilité (Ω, \mathcal{T}, P) , pour tout événement A de probabilité non nulle, la mesure $\frac{P(\cdot \cap A)}{P(A)} = (\text{notée}) P(\cdot|A)$ ou $P^A(\cdot)$ définit sur (Ω, \mathcal{T}, P) une probabilité, dite conditionnelle par rapport à l'événement A . Pour tout événement $B \in \mathcal{T}$:

$$P(B|A) = \frac{P(B \cap A)}{P(A)}.$$

Preuve $P(\Omega|A) = \frac{P(\Omega \cap A)}{P(A)} = 1$ et $P(\biguplus_{i=1}^n B_i|A) = \frac{P(\biguplus_{i=1}^n B_i \cap A)}{P(A)} = \sum_i \frac{P(B_i \cap A)}{P(A)} = \sum_i P(B_i|A) :$

$\frac{P(\cdot \cap A)}{P(A)}$ définit bien une loi de probabilité. □

6.2 ♠ Propriété caractéristique des systèmes « sans mémoire », décrits par des v.a. géométrique et exponentielle

Soit X une v.a. géométrique de loi $P(X = k) = (1 - p)^{k-1} p$ ($\forall k \in \mathbb{N}^*$).

(1) Démontrer que $P[X > m + n | X > n] = P[X > m]$. Interpréter ce résultat si X est la durée de vie d'un système.

(2) Cette propriété caractérise aussi les v.a. exponentielles. Est-ce surprenant ?

Attention ! L'indépendance de deux événements n'implique pas leur indépendance conditionnelle par rapport à un troisième : $P(A \cap B) = P(A)P(B)$ n'entraîne pas que $P(A \cap B | C) = P(A | C)P(B | C)$ pour un événement quelconque C .

Théorème 6.3 (Formules bayésiennes) Soient (Ω, \mathcal{T}, P) un espace probabilisé et une partition $(B_i)_i$ de Ω , dit système complet de Ω où chaque B_i est de probabilité non nulle.

$$(1) \text{ Formule de recomposition : } \forall A \in \mathcal{T}, P(A) = \sum_i^n P(A|B_i)P(B_i).$$

$$(2) \text{ Formule de Bayes : } P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}.$$

Preuve

$$(1) A = A \cap (\bigcup_i B_i) = \bigcup_i (A \cap B_i) \Rightarrow P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i).$$

$$(2) \sum_j P(A|B_j)P(B_j) = P(A) \text{ et } P(A|B_i)P(B_i) = P(B_i|A)P(A). \quad \square$$

Interprétation de la formule de Bayes dans le cadre d'un modèle causal.

Si un événement A résulte de l'une des n causes possibles $(B_i)_i$ incompatibles entre elles, alors connaissant l'ensemble des $P(A|B_i)$ et des $P(B_i)$, la formule de Bayes permet de déterminer la probabilité $P(B_i|A)$ que B_i soit la cause de A .

Exemple 6.2 Fiabilité d'un système de détection

Un système de détection répond par une alerte à la détection d'une panne intervenant sur une machine donnée. Soient A l'événement « l'alerte est donnée », et B l'événement « le système est en panne », vérifiant les probabilités conditionnelles suivantes : $P(A|B) = 0,99$ et $P(A|\bar{B}) = 0,002$.

Le problème que se pose le service de sécurité est le suivant : si l'alerte est donnée, quelle est la probabilité pour qu'elle corresponde à une panne, sachant qu'une panne survient avec une probabilité égale à 0,01 ?

Il s'agit de calculer $P(B|A)$ égal à :

$$\frac{P(A|B)P(B)}{P(B)P(A|B) + P(\bar{B})P(A|\bar{B})} = \frac{0,99 \times 0,01}{0,01 \times 0,99 + 0,99 \times 0,002} \simeq 0,83.$$

6.3 ♠ Le taux de réussite d'un examen donné est de 60 % pour les candidats issus de l'établissement A et de 80 % pour ceux issus de B . D'autre part, 55 % des candidats proviennent de A et 45 % de B .

Quel est le taux de réussite à cet examen ? Quelle est la probabilité pour qu'un candidat éliminé provienne de A ?

6.4 ♠ Test du sida

On estime que 130 000 personnes sont atteintes du sida en France, peuplée de 58 millions d'habitants ; le test de dépistage a une fiabilité de 0,998. Soit 0,001 la probabilité pour que le test soit positif à tort. Quelle est la probabilité qu'une personne n'ait pas le sida, alors que son test est positif ?

Pour clore cette introduction, on soumet deux paradoxes à la sagacité du lecteur.

6.5 ♠♠ Le paradoxe des trois pièces

Soient trois pièces dont l'une a deux faces rouges, la deuxième une face rouge et une face blanche, la troisième deux faces blanches. Je tire au hasard l'une des pièces qui vous présente l'une de ses faces qui est rouge. Quelle est la probabilité pour que la face masquée soit rouge ?

[On désignera respectivement par rr , bb et rb les trois pièces à deux faces rouges, deux faces blanches et deux faces de couleurs distinctes. Il s'agira d'exprimer $P(rr \mid \text{face visible rouge})$.]

Ce paradoxe se présente aussi sous une forme plus naturelle : vous sonnez au domicile d'une famille ayant deux enfants ; la porte s'ouvre et un garçon vous accueille. Quelle est la probabilité que l'autre enfant soit un garçon ?

6.6 ♠♠ Le paradoxe du prisonnier

L'exposé que l'on donne de ce paradoxe, s'inspire du problème dit de *Monty Hall*, qui fut un jeu télévisé fameux présenté naguère aux États-Unis. Supposons qu'un prisonnier soit placé devant trois portes fermées 1, 2 et 3 ; l'une d'elles lui permet de sortir définitivement de prison. Le directeur, qui connaît la configuration des portes, lui propose d'en choisir une au hasard et ouvre, après le choix du prisonnier, une porte qui n'est (évidemment) pas celle qui conduit à la liberté.

Le prisonnier augmente-t'il ses chances en modifiant son choix initial ?

Théorème 6.4 (*Conditionnements successifs*)

Soient les événements A_1, A_2, \dots, A_n vérifiant $P(A_1 \cap \dots \cap A_{n-1}) > 0$, alors :

$$P(\bigcap_{i=1}^n A_i) = P(A_n \mid A_1 \cap \dots \cap A_{n-1}) P(A_{n-1} \mid A_1 \cap \dots \cap A_{n-2}) \dots P(A_1).$$

Preuve (par récurrence) La propriété est évidente pour $n = 2$ et supposée vraie pour $n - 1$.

Démontrons la pour tout n : $P(\bigcap_{i=1}^n A_i) = P(A_n \mid A_1 \cap \dots \cap A_{n-1}) P(A_1 \cap \dots \cap A_{n-1}) = P(A_n \mid A_1 \cap \dots \cap A_{n-1}) P(A_{n-1} \mid A_1 \cap \dots \cap A_{n-2}) \dots P(A_1)$. \square

6.2 Lois conditionnelles

Dans ce paragraphe, il s'agit d'expliciter la probabilité conditionnelle dans chacun des quatre cas, caractérisé par le caractère discret ou continu des variables conditionnantes et des variables conditionnées.

6.2.1 Loi d'une variable discrète conditionnée par une variable discrète

Théorème 6.5 *Étant donné un vecteur aléatoire (X, Y) défini sur (Ω, \mathcal{T}, P) à valeurs dans le sous-ensemble au plus dénombrable $D \subset \mathbb{R} \times \mathbb{R}$, de loi conjointe $P_{X,Y}$ connue. La loi conditionnelle de Y sachant X est définie par :*

$$\forall (x_i, y_j) \in D \text{ tel que } P_X(x_i) \neq 0 :$$

$$P_Y(y_j|X = x_i) = P_{Y|X=x_i}(y_j) = \frac{P(X = x_i, Y = y_j)}{P_X(x_i)}.$$

Preuve Application de la définition de la probabilité conditionnelle à $B : (Y = y_j)$ sachant $A : (X = x_i)$. \square

Exemple 6.3 Une source radioactive émet X particules, où X est une v.a. de Poisson d'intensité λ : chaque particule émise est détectée avec la probabilité p . Soit Y le nombre total de particules détectées. Déterminons la loi de Y .

Pour tout $i \geq k$, $P(Y = k|X = i) = \binom{i}{k} p^k (1-p)^{i-k}$; la loi conditionnelle de Y sachant X est donc binomiale $\mathcal{B}(i, p)$. D'où par la formule de recombinaison :

$$P(Y = k) = \sum_{i \geq k} \binom{i}{k} p^k (1-p)^{i-k} e^{-\lambda} \frac{\lambda^i}{i!} = \frac{e^{-\lambda p} (p\lambda)^k}{k!}.$$

Conclusion : Y est de loi de Poisson de paramètre $p\lambda$.

Définition 6.3 *La moyenne des valeurs de Y conditionnée par l'événement $X = x_i$ est égale à :*

$$\sum_j y_j P_Y(y_j|X = x_i) =_{(notée)} E(Y|X = x_i),$$

dite *espérance ou moyenne conditionnelle de Y sachant $(\mathbf{X} = \mathbf{x}_i)$* .

Théorème 6.6

$$E(Y) = \sum_{i=1}^n P(X = x_i) E(Y|X = x_i).$$

Preuve

$$\begin{aligned} \sum_{i=1}^n P(X = x_i) E(Y|X = x_i) &= \sum_{i=1}^n P(X = x_i) \sum_j y_j P_Y(y_j|X = x_i) \\ &= \sum_j y_j \sum_{i=1}^n P_{X,Y}(x_i, y_j) = E(Y). \end{aligned}$$

□

Exemple 6.4 Soit le vecteur (X, Y) de distribution définie par la figure 6.1.

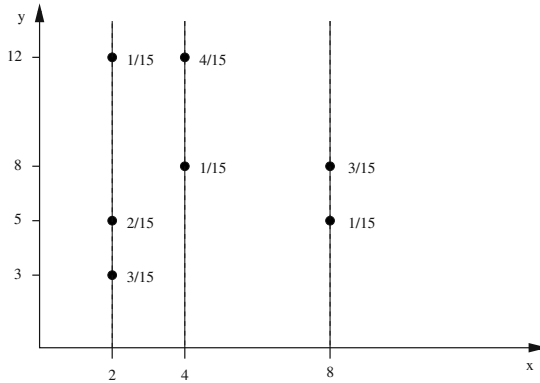


Figure 6.1 – Répartition discrète.

Détermination de quelques probabilités marginales et conditionnelles :

$$P(X = 2) = \frac{1+2+3}{15} = \frac{6}{15} = \frac{2}{5}, P(X = 8) = \frac{3+1}{15} = \frac{4}{15} \text{ ou } P(Y = 5) = \frac{1+2}{15} = \frac{1}{5};$$

$$P(Y=5|X=2) = \frac{1}{3}.$$

$$E(Y|X = 2) = \left(3 \times \frac{3}{15} + 5 \times \frac{2}{15} + 12 \times \frac{1}{15}\right) \times \frac{15}{6} = \frac{31}{6}$$

$$E(Y|X = 4) = \left(8 \times \frac{1}{15} + 12 \times \frac{4}{15}\right) \times \frac{15}{5} = \frac{56}{5}$$

$$E(Y|X = 8) = \left(5 \times \frac{1}{15} + 8 \times \frac{3}{15}\right) \times \frac{15}{4} = \frac{29}{4} \text{ d'où } E(Y) = \frac{116}{15}.$$

Définition 6.4 La *variance de Y conditionnée par $(\mathbf{X} = \mathbf{x}_i)$* est définie par :

$$E((Y - E(Y|X = x_i))^2 | X = x_i) =_{(noté)} V(Y|X = x_i).$$

6.7 ♠♠ Les nombres de véhicules X et Y à destination des villes A et B, passant par un poste de contrôle commun, sont des v.a. indépendantes de Poisson d'intensités respectives a et b . Déterminer la loi conditionnelle de X sachant $(X + Y)$, l'espérance $E(X | X + Y = n)$ et la variance $V(X | X + Y = n)$.

- 6.8 ♠** Soient X, Y des v.a. indépendantes de même loi uniforme sur $\{-1, 0, 1\}$. Calculer $P(X < Y)$ et $E(Y | X < Y)$.

6.2.2 Loi d'une variable absolument continue conditionnée par une variable discrète

On considère des vecteurs aléatoires mixtes dont l'une des composantes Y est discrète tandis que l'autre X est continue.

Définition 6.5 Un vecteur aléatoire (X, Y) défini sur $\mathbb{R} \times \mathbb{D}$, où \mathbb{D} est un espace au plus dénombrable, est mixte si sa loi $P_{(X,Y)}$ est définie par :

$$\forall (B, y_k) \in \beta_{\mathbb{R}} \times \mathbb{D} : P_{(X,Y)}(B, y_k) = \int_B f(x, y_k) dx$$

où f est une fonction mesurable positive qui vérifie $\sum_k \int_{\mathbb{R}} f(x, y_k) dx = 1$.

Les lois marginales de X et Y sont définies par :

$$f_X(x) = \sum_k f(x, y_k);$$

$$P_Y(y_k) = \int_{\mathbb{R}} f(x, y_k) dx.$$

Exemple 6.5 Soit le couple (X, Y) défini sur $\mathbb{R}_+ \times \mathbb{N}$ pour lequel :

$$f(x, k) = e^{-2x} \frac{x^k}{k!}.$$

Il est facile de montrer que : $\int_{\mathbb{R}_+} \sum_k f(x, k) dx = 1$.

$f_X(x) = \sum_k f(x, k) = \sum_k e^{-2x} \frac{x^k}{k!} = e^{-x}$; X est donc une v.a. exponentielle $\mathcal{E}(1)$.

$P_Y(k) = \int_{\mathbb{R}_+} f(x, k) dx = \int_{\mathbb{R}_+} e^{-2x} \frac{x^k}{k!} dx = \left(\frac{1}{2}\right)^{k+1}$; Y est de loi géométrique de paramètre $\frac{1}{2}$. Conclusion : les v.a. X et Y sont dépendantes.

Théorème 6.7 Étant donné le couple mixte (X, Y) , la densité conditionnelle de X sachant $(Y = y_k)$ est définie par :

$$f_{X/(Y=y_k)}(x) = \frac{f(x, y_k)}{\int_{\mathbb{R}} f(x, y_k) dx} (p \cdot p) = \frac{f(x, y_k)}{P_Y(y_k)} \text{ si } P_Y(y_k) \neq 0.$$

Preuve On recherche la loi conditionnelle de X sachant Y sous la forme d'une fonction g vérifiant :

$$\forall k \text{ entier et } B \text{ borélien, } P(X \in B \mid Y = y_k) = \int_B g(x, y_k) dx \quad (*)$$

$$\forall B, P(X \in B \mid Y = y_k) = \frac{P(X \in B, Y = y_k)}{P(Y = y_k)} = \frac{\int_B f(x, y_k) dx}{\int_{\mathbb{R}} f(x, y_k) dx} = \int_B \frac{f(x, y_k) dx}{\int_{\mathbb{R}} f(x, y_k) dx} \quad (**)$$

L'égalité de (*) et (**), pour tout B et k , permet de déduire :

$$g(x, y_k) =_{(p.p.)} \frac{f(x, y_k)}{\int_{\mathbb{R}} f(x, y_k) dx} =_{(notée)} f_{X/(Y=y_k)}(x).$$

□

Dans l'exemple 6.5, $f_{X/Y=k}(x) = 2^{k+1} \cdot \frac{x^k}{k!} \cdot e^{-2x}$; la densité conditionnelle $f_{X/Y=k}$ de X sachant $Y = k$ est donc la densité de la loi gamma de paramètres $(k, 2)$.

6.2.3 Loi d'une variable discrète conditionnée par une variable absolument continue

On se place sous les mêmes hypothèses qu'en 6.2.2.

Théorème 6.8 *La loi de probabilité conditionnelle de la v.a. discrète Y par la v.a. absolument continue X , est discrète et définie pour tout x tel que $f_X(x) \neq 0$ par :*

$$\forall y_k \quad P(Y = y_k \mid X = x) = \frac{f(y_k, x)}{f_X(x)} = \frac{f_{X/Y=y_k}(x) P(Y = y_k)}{f_X(x)}.$$

6.9 ♠ Soit la v.a. Y de loi $\mathcal{B}(n, X)$ où le paramètre X est une v.a. $\mathcal{U}([0, 1])$.

$$P(Y = k \mid X = x) = \binom{n}{k} x^k (1-x)^{n-k}, \forall k \in \{0, 1, 2, \dots, n\}.$$

Déterminer les lois de Y et de X conditionnée par Y . (voir la définition de la fonction *beta* dans le formulaire mathématique.)

Dans l'exemple 6.5, $P(Y = k \mid X = x) = e^{-x} \cdot \frac{x^k}{k!}$; on pourra vérifier que la loi conditionnelle de Y sachant $X = x$ est donc une loi de Poisson de paramètre x .

6.2.4 Loi d'une variable continue conditionnée par une variable continue

Théorème 6.9 *Étant donné le vecteur aléatoire (X, Y) de densité $f_{(X,Y)}$ telle que $\forall x, f_X(x) \neq 0$, la loi conditionnelle de Y sachant $X = x$ est absolument*

continue, de densité :

$$\frac{f_{(X,Y)}(x,y)}{f_X(x)} =_{(noté)} f_{Y/X=x}(y).$$

Preuve Pour tout borélien B appartenant au support de Y et pour tout x tel que $f_X(x) \neq 0$, on désire exprimer $P(Y \in B | x < X \leq x + \Delta x)$ sous la forme $(*) \int_B g_x(y) dy$, où g_x est une densité portée par le support de Y .

$$\begin{aligned} (\forall B) P(Y \in B | x < X \leq x + \Delta x) &= \frac{P((x < X \leq x + \Delta x) \cap (Y \in B))}{P(x < X \leq x + \Delta x)} \\ &= \frac{\int_B f_{(X,Y)}(x,y) \Delta x dy}{f_X(x) \Delta x} = \int_B \frac{f_{(X,Y)}(x,y)}{f_X(x)} dy. \end{aligned}$$

L'identification de cette dernière expression avec $(*)$, pour tout borélien B , entraîne :

$$g_x(y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} (p.p.).$$

□

L'espérance conditionnelle $E(Y | X = x)$ de Y sachant $(X = x)$ s'exprime sous une forme analogue à celle de la définition 6.4 :

$$E(Y | X = x) = \int y f_{Y/X=x}(y) dy.$$

Exemple 6.6 Soit $f_{(X,Y)}(x,y) = \lambda^2 e^{-\lambda y}$ de support $D = \{(x,y) | 0 \leq x \leq y\}$.

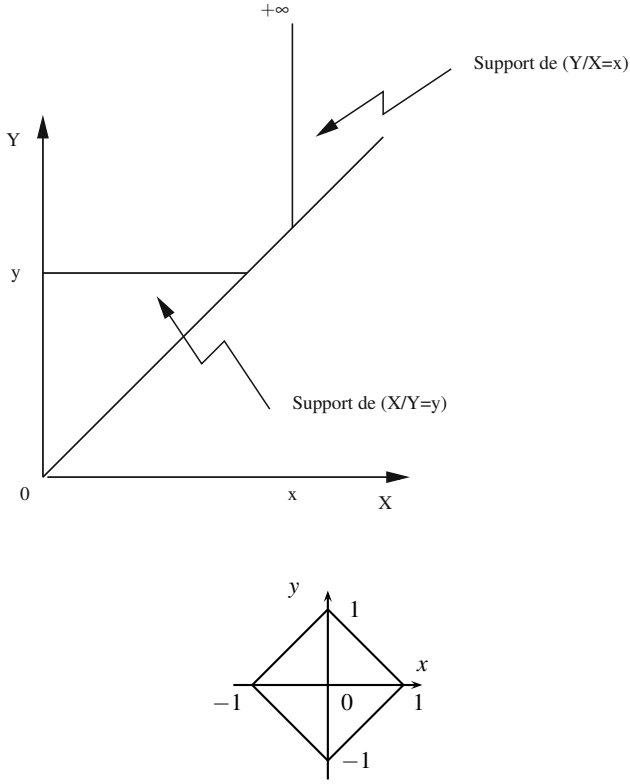
$$f_X(x) = \int_x^{+\infty} \lambda^2 e^{-\lambda y} dy = \lambda e^{-\lambda x} \text{ et } f_Y(y) = \lambda^2 y e^{-\lambda y}.$$

$\forall x \geq 0$, $f_{Y/X=x}(y) = \lambda e^{-\lambda(y-x)} \mathbf{1}_{[x, +\infty[}(y)$ et $\forall y > 0$, $f_{X/Y=y}(x) = \frac{1}{y} \mathbf{1}_{]0,y]}(x)$ qui est la densité de la v.a. uniforme sur $]0,y]$.

Théorème 6.10 (*Caractérisation de l'indépendance de v.a. absolument continues*)

$$X, Y \text{ indépendantes} \iff \begin{cases} \forall y, f_X = f_{X/Y=y} (p.p.), \text{ ou} \\ \forall x, f_Y = f_{Y/X=x} (p.p.), \text{ ou} \\ f_{(X,Y)} = f_X \cdot f_Y (p.p.). \end{cases}$$

6.10 ♠ Soit le vecteur aléatoire (X,Y) uniformément réparti sur le support carré ci-dessous :



(1) Déterminer la loi conjointe, les lois marginales et conditionnelles. (2) Les composantes sont-elles indépendantes ?

6.2.5 Loi d'une variable continue conditionnée par un événement quelconque

Soit $X: (\Omega, \mathcal{T}, P) \rightarrow (\mathbb{R}, \beta_{\mathbb{R}})$ et $E \in \mathcal{T}$ tel que $P(E) \neq 0$.

Définition 6.6 La fonction de répartition conditionnelle de X sachant E est définie par : $\forall x \in \mathbb{R}, P(X \leq x | E) = \frac{P(\{\omega \in \Omega \text{ t.q. } X(\omega) \leq x\} \cap E)}{P(E)} \stackrel{(\text{notée})}{=} F_{X/E}(x)$.

Théorème 6.11

$$f_{X/E}(x) = \frac{P(E | X = x) \cdot f_X(x)}{P(E)}.$$

Preuve

$$P(E | x < X \leq x+h) = \frac{P(x < X \leq x+h | E) \cdot P(E)}{P(x < X \leq x+h)} \simeq \frac{h \cdot f_{X/E}(x) \cdot P(E)}{h \cdot f_X(x)}$$

quand h est proche de 0.

$$\text{D'où : } \lim_{h \downarrow 0} P(E | x < X \leq x+h) = P(E | X = x) = \frac{f_{X/E}(x) \cdot P(E)}{f_X(x)}. \quad \square$$

Le théorème suivant généralise au cas continu la formule de recombinaison ou des probabilités totales que nous rappelons : si $(B_i)_i$ est un système complet d'événements ($\bigsqcup_i B_i = \Omega$), alors pour tout événement A de \mathcal{T} on a $P(A) = \sum_i P(A | B_i) \cdot P(B_i)$.

Théorème 6.12 *Pour toute v.a. X absolument continue définie sur (Ω, \mathcal{T}, P) et tout événement E de \mathcal{T} , on a :*

$$P(E) = \int P(E | X = x) \cdot f_X(x) dx.$$

Preuve Pour tout $h > 0$,

$$P(E) = \sum_{i \in \mathbb{Z}} P(E | X \in](i-1)h, ih]) \cdot P(X \in](i-1)h, ih]) \stackrel{(\text{notée})}{=} P_h(E).$$

Faisons tendre h vers 0 :

$$\begin{aligned} P(X \in](i-1)h, ih]) &\longrightarrow h \cdot f_X(ih), \\ P(E | X \in](i-1)h, ih]) &\longrightarrow P(E | X = ih). \end{aligned}$$

$$P_h(E) \longrightarrow \int_{\mathbb{R}} P(E | X = x) \cdot f_X(x) dx. \quad \square$$

Ce résultat se généralise aux vecteurs aléatoires.

Soit par exemple, (X, Y) de densité $f_{X,Y}$ sur $\mathbb{R}^+ \times \mathbb{R}^+$ et $E = (X/Y \leq 1)$.

$$\begin{aligned} P\left(\frac{X}{Y} \leq 1\right) &= \iint P\left(\frac{X}{Y} \leq 1 | X = x, Y = y\right) \cdot f_{X,Y}(x, y) dx dy \\ &= \int_0^\infty \left(\int_{0 \leq \frac{x}{y} \leq 1} f_{X,Y}(x, y) dy \right) dx \\ &= \int_0^\infty \left(\int_x^\infty f_{X,Y}(x, y) dy \right) dx. \end{aligned}$$

On retrouve évidemment ce résultat grâce à la méthode de calcul de lois (voir chapitre 4) :

$$\begin{aligned} P\left(\frac{X}{Y} \leq 1\right) &= \int_0^1 f_{X/Y}(u) du \\ &= \int_0^1 \left(\int_0^\infty f_{X,Y}\left(x, \frac{x}{u}\right) \cdot \frac{x}{u^2} dx \right) du \\ &= \int_0^\infty \left(\int_0^1 f_{X,Y}\left(x, \frac{x}{u}\right) \cdot \frac{x}{u^2} du \right) dx. \end{aligned}$$

Après avoir posé $y = \frac{x}{u}$, il vient :

$$\begin{aligned} P\left(\frac{X}{Y} \leq 1\right) &= \int_{-\infty}^x \left(\int_0^{\infty} f_{X,Y}(x,y) \cdot \frac{y^2}{x} \cdot \left(-\frac{x}{y^2}\right) dy \right) dx \\ &= \int_0^{\infty} \left(\int_x^{\infty} f_{X,Y}(x,y) dy \right) dx. \end{aligned}$$

6.11 ♠ Résoudre le problème de Buffon (Exercice 3.2) grâce au théorème 6.12.
[Indication : $P(\frac{X}{\cos(\Theta)} < \frac{l}{2}) = \int P(\frac{X}{\cos(\Theta)} < \frac{l}{2} \mid \Theta = \theta) \cdot f_{\Theta}(\theta) d\theta$.]

Notions de fiabilité

La fiabilité d'un système est caractérisée par sa probabilité d'accomplir une tâche donnée dans des conditions déterminées et pendant une période donnée. Dans l'environnement économique et technique actuel où les coûts de défaillance peuvent être très pénalisants, il est nécessaire de développer des études de fiabilité permettant de garantir une maintenance de qualité.

La **durée de vie ou date de première panne** d'un système est décrite par une v.a. T , absolument continue et réelle positive, de fonction de répartition F_T . On supposera que le système une fois réparé retrouve ses qualités initiales.

Définition 6.7 On appelle **fonction de fiabilité ou de survie** d'un système de durée de vie T la fonction qui à tout temps t positif associe la probabilité que le système soit encore en vie au temps t :

$$R_T(t) = P(T > t) = 1 - F_T(t).$$

Définition 6.8 Le **taux de défaillance moyen** $\lambda(t; t+h)$ entre les temps t et $t+h$ est la probabilité que le système tombe en panne entre t et $t+h$ sachant qu'il n'est pas tombé en panne avant t , divisée par h :

$$\begin{aligned} \lambda(t; t+h) &= \frac{P(t < T \leq t+h \mid T > t)}{h} = \frac{P((t < T \leq t+h) \cap (T > t))}{hP(T > t)} \\ &= \frac{P(t < T \leq t+h)}{hP(T > t)} = \frac{F_T(t+h) - F_T(t)}{h(1 - F_T(t))} = \frac{R_T(t) - R_T(t+h)}{hR_T(t)}. \end{aligned}$$

Définition 6.9 Le **taux de défaillance instantané** $\lambda(t)$ au temps t est défini par :

$$\lambda(t) = \lim_{h \rightarrow 0} \lambda(t; t+h) = -\frac{R'_T(t)}{R_T(t)}.$$

Définition 6.10 On désigne par **MTTF**, abréviation de l'anglais mean time to failure, la durée moyenne du bon fonctionnement d'un système et par **MTBF** (mean time between failure) l'intervalle de temps moyen entre deux défaillances consécutives d'un système réparable.

6.12 ♠ ♠ Étude de deux modèles classiques

(1) Si le système n'est pas soumis à un phénomène cumulatif d'usure et que, d'autre part, il n'a pas de défaut de jeunesse, le taux de défaillance peut être considéré constant égal à λ . Établir l'équation différentielle vérifiée par $f_T(t)$ lorsque $\lambda(t) = \lambda$ et la résoudre. Vérifier que le système est sans mémoire et calculer sa MTTF.

(2) Dans le cas, plus réaliste, où le taux de défaillance n'est pas constant, on utilise le modèle de Weibull défini par le taux de défaillance :

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t-\gamma}{\eta} \right)^{\beta-1} \quad \text{si } t > \gamma$$

$$= 0 \quad \text{si } t \leq \gamma \quad (\beta > 0, \eta > 0, \gamma \geq 0).$$

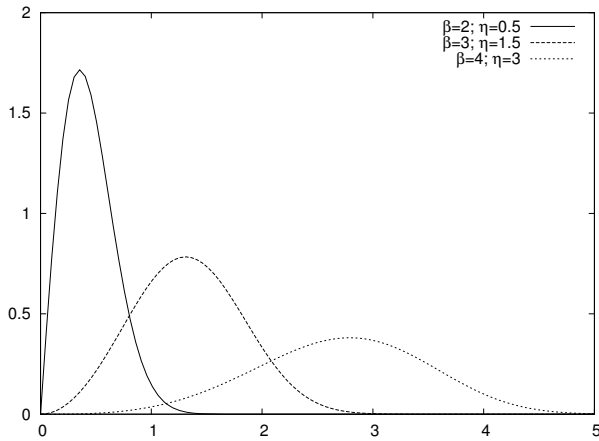


Figure 6.2 – Densités de Weibull.

Caractériser le taux de défaillance selon la valeur de β dans chacun des trois cas : $0 < \beta < 1$, $\beta = 1$, $\beta > 1$. Déterminer la fonction de fiabilité et montrer que $E(T) = \eta\Gamma(1 + \frac{1}{\beta}) + \gamma$.

(3) On suppose que $\gamma = 0$. Démontrer :

$$\beta \in]0, 1[\implies \forall t, s \quad P(T > t+s | T > s) > P(T > t) ;$$

$$\beta > 1 \implies \forall t, s \quad P(T > t+s | T > s) < P(T > t).$$

Interpréter ces résultats.

(4) Soient n systèmes identiques régis par le modèle de Weibull, structurés en série. Calculer la loi de la durée de vie du système en série (T_i étant la durée de vie du système i).

On considère que la durée de vie d'un système est partitionné en trois périodes :

- la période J de rodage ou de jeunesse du système, qui est la plus courte, pendant laquelle il est soumis aux défaillances immédiates dès la mise en marche du système, affectant les composants de moindre qualité ; au cours de cette période $\lambda(t)$ décroît ;

- la vie utile U du système, période pendant laquelle les défaillances sont dues à des causes accidentelles, défauts résiduels correspondante à un taux de défaillance constant qui atteint sa plus petite valeur ;

- la période de vieillesse V du système pendant laquelle il est soumis à une usure croissante due aux effets cumulatifs du vieillissement : $\lambda(t)$ croît.

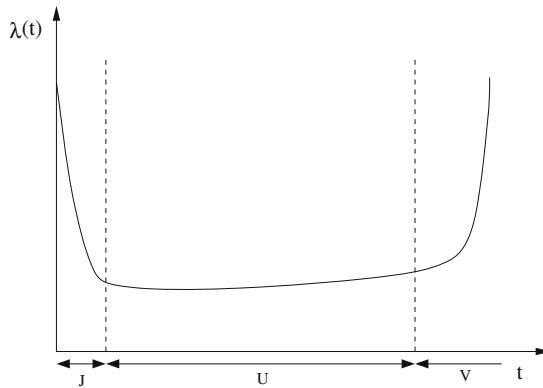


Figure 6.3 – Évolution classique du taux de défaillance selon la courbe dite *en baignoire*.

6.3 Espérance conditionnelle

Étant donné une v.a. Y inobservable, dépendante d'une v.a. X observable, il s'agit de déterminer une nouvelle v.a. fonction de X , notée $\Phi^*(X)$, qui apporte le plus d'information possible sur la v.a. inobservable Y . Dans une première approche, on recherche la fonction Φ^* vérifiant :

$\Phi^*(X)$ minimise $E((Y - \Phi^*(X))^2)$ (les v.a. X et Y sont supposées de moment d'ordre deux fini).

(1) Cas où on recherche $\Phi^*(X)$ sous la forme d'une constante c .

$E((Y - c)^2) = V(Y) + (E(Y) - c)^2$ atteint son minimum pour $c = E(Y)$ qui est donc la meilleure approximation par une constante de la v.a. Y .

(2) Supposons que le vecteur (X, Y) est de densité $f_{X,Y}$:

$$\begin{aligned} E((Y - \Phi(X))^2) &= \int \int (y - \Phi(x))^2 f_{X,Y}(x, y) dx dy \\ &= \int \left[\int (y - \Phi(x))^2 \frac{f_{X,Y}(x, y)}{f_X(x)} dy \right] f_X(x) dx \\ &= \int \left[\int (y - \Phi(x))^2 f_{Y/X=x}(y) dy \right] f_X(x) dx \quad (*). \end{aligned}$$

Cette intégrale positive est minimale si la fonction de $x \int (y - \Phi(x))^2 f_{Y/X=x}(y) dy$ est minimale. Il y a une forte analogie entre la minimisation de l'intégrale (*) et celle de $E((Y - c)^2)$ dans (1), égal à $\int (y - c)^2 f_Y(y) dy$, puisque dans (*), $\Phi(x)$ joue le rôle d'une constante par rapport à la variable d'intégration. On conclut donc que :

$$\Phi^*(x) = \int y f_{Y/X=x}(y) dy = E(Y | X = x).$$

Définition 6.11 La meilleure approximation quadratique d'une v.a. Y par une v.a. X est nommée **espérance conditionnelle de Y sachant X** et notée :

$$E(Y | X).$$

On la détermine à partir de l'expression de $E(Y | X = x)$.

Exemple 6.7

(1) (X, Y) de densité $\lambda^2 e^{-\lambda y}$ et de support $D = \{(x, y) | 0 \leq x \leq y\}$. Étant donné la densité conditionnelle $f_{Y/X=x}(y) = \lambda e^{-\lambda(y-x)} \mathbf{1}_{[x, +\infty[}(y)$, l'espérance de Y conditionnée par $X = x$ est égale à

$$E(Y | X = x) = x + \frac{1}{\lambda},$$

d'où il vient :

$$E(Y | X) = X + \frac{1}{\lambda}.$$

(2) Suite de l'exercice 6.7 :

$$E(X | X + Y = n) = n \frac{a}{a+b} \text{ d'où : } E(X | X + Y) = \frac{a(X+Y)}{a+b}.$$

On vérifie que $E(E(X | X + Y)) = \frac{aE(X+Y)}{a+b} = a$, et de façon analogue que $E(Y | X + Y) = \frac{b(X+Y)}{a+b}$ et $E(E(Y | X + Y)) = b$.

Construction géométrique de l'espérance conditionnelle

Considérons l'espace vectoriel hilbertien $H = L^2(\Omega, \mathcal{T}, P)$ des v.a. réelles possédant un moment d'ordre deux fini, que l'on munit du produit scalaire $(U|V) \equiv E(UV)$ et donc de la norme quadratique $\|U\| = \sqrt{E(U^2)}$.

À toute v.a. $X \in H$, on associe la tribu engendrée \mathcal{T}_X définie par $\{X^{-1}(B) \mid B \in \mathcal{B}_{\mathbb{R}}\}$ et le sous-espace fermé H_X de H défini par $H_X = L^2(\Omega, \mathcal{T}_X, P)$ constitué par les v.a. fonctions de X .

On vient de voir que pour toute v.a. $Y \in H$, la fonction $\Phi(X)$ égale à $E(Y|X)$ minimise $\|Y - \Phi(X)\|$. L'espérance conditionnelle $E(Y|X)$ est déterminée P -presque sûrement, c'est-à-dire à un sous-ensemble négligeable près (voir définition 1.4). Compte tenu des propriétés des espaces hilbertiens H et H_X , on peut appliquer le théorème de la projection unique sur un sous-espace fermé, résultat classique d'analyse fonctionnelle : $\|Y - E(Y|X)\|$ est minimal si et seulement si $E(Y|X)$ est la projection orthogonale unique de Y sur H_X .

L'opérateur espérance $E(\cdot)$ est, dans ce cadre géométrique, l'opérateur qui projette orthogonalement toute v.a. Y sur le sous-espace H_c des v.a. constantes réelles. Dans la figure suivante, H est représenté par l'espace euclidien à trois dimensions, H_X par le plan et H_c par une droite issue de O .

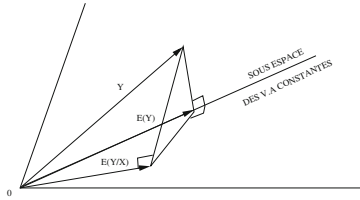


Figure 6.4 – Représentation euclidienne.

Théorème 6.13 (Propriétés de l'espérance conditionnelle)

On suppose que les variables conditionnées ont un moment d'ordre deux fini.

(1) *Linéarité* : $E(aX + bY|Z) = aE(X|Z) + bE(Y|Z)$ où a et b réels.

(2) *Croissance* : si $Y \geq X$ (p.s.), alors pour toute v.a. Z ,

$$E(Y|Z) \geq E(X|Z) \text{ (p.s.)}.$$

(3) Si Y est indépendante de X , alors

$$E(Y|X) = E(Y) \text{ (p.s.)}.$$

(4) Pour toute fonction ϕ bornée :

$$E(\phi(X)Y|X) = \phi(X)E(Y|X).$$

(5) Propriété de l'espérance totale : pour tout vecteur aléatoire (X, Y) ,

$$E(E(Y|X)) = E(Y).$$

(6) Pour toute fonction ϕ bornée, $E(\phi(X)Y) = E(\phi(X)E(Y|X))$.

(7) Majoration de la variance de l'espérance conditionnelle :

$$V(E(Y|X)) \leq V(Y).$$

(Interprétation : si X conditionne Y , toute observation de X apporte une information sur Y et en réduit donc la variance.)

Preuve (1) (2) sont triviales.

(3) La v.a. $E(Y|X)$ est de la forme $\phi(X)$, laquelle est indépendante de Y : elle est nécessairement constante et égale à $E(Y)$.

(4) Pour toute fonction-test ψ , $P_{(X,Y)}$ intégrable,

$$E(\psi(X, Y)|X = x) = E(\psi(x, Y)|X = x) ;$$

prenons $\psi(x, Y) = \phi(x)Y$. Pour tout x ,

$$E(\phi(X)Y|X = x) = E(\phi(x)Y|X = x) = \phi(x)E(Y|X = x),$$

d'où l'égalité attendue.

(5) $E(Y|X)$ est la projection orthogonale de Y sur le sous-espace L_X^2 , que l'on projette ensuite sur le sous-espace des v.a. constantes pour obtenir $E(E(Y|X))$; le théorème des trois perpendiculaires de la géométrie hilbertienne ou euclidienne permet de conclure.

(7) Démonstration qui s'apparente à (5) : il suffit de considérer dans L^2 le triangle rectangle de côtés Y , $E(Y|X)$ et $(Y - E(Y|X))$, et de lui appliquer le théorème de Pythagore. \square

6.13 ♠ Démontrer la propriété (6).

Attention ! L'indépendance de deux v.a. n'implique pas nécessairement l'indépendance de leurs espérances conditionnelles. Montrons cela sur un contre-exemple classique : considérons (X, Y) un vecteur gaussien centré réduit de covariance nulle et les v.a. $S = X + Y$ et $D = X - Y$.

(S, D) est un vecteur gaussien centré, de covariance $E((X + Y)(X - Y))$ nulle, donc de composantes indépendantes et pourtant $E(S|X) = X$ et $E(D|X) = X$, donc S et D ne sont pas conditionnellement indépendantes par rapport à X .

Théorème 6.14 (Régression linéaire de Y par rapport à X)

La meilleure approximation quadratique de Y par une fonction affine de X (supposée de variance non nulle) est la v.a. \hat{Y} définie par :

$$\hat{Y} = \left(\rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \right) X + \left(E(Y) - \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} E(X) \right).$$

La droite de régression de Y par rapport à X est définie par l'équation :

$$y = ax + b, \text{ où } a = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \text{ et } b = E(Y) - aE(X).$$

Preuve Il s'agit de minimiser $\phi(a,b) = E[(Y - (aX + b))^2]_{(a,b) \in \mathbb{R}^2}$.

La fonction $\phi(a,b)$ est convexe; son minimum est donc atteint pour les valeurs de a et b qui annulent ses dérivées partielles :

$$\begin{cases} \frac{\partial \phi}{\partial a} = 0 & \Leftrightarrow E[X(Y - (aX + b))] = 0 \\ \frac{\partial \phi}{\partial b} = 0 & \Leftrightarrow E[Y - (aX + b)] = 0 \end{cases} \implies \begin{cases} E(X^2)a + E(X)b = E(XY) \\ E(X)a + b = E(Y) \end{cases}$$

$$\text{d'où, sachant que } \sigma_X \neq 0 : \begin{cases} a = \frac{E(XY) - E(X)E(Y)}{\sigma_X^2} = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \\ b = E(Y) - aE(X) \end{cases} \quad \square$$

De façon symétrique, la meilleure approximation affine de X par Y est définie par :

$$\hat{X} = \rho_{X,Y} \frac{\sigma_X}{\sigma_Y} Y + E(X) - \rho_{X,Y} \frac{\sigma_X}{\sigma_Y} E(Y).$$

Dans ce cas, l'équation de la régression est définie par $x = \alpha y + \beta$, avec $\alpha = \rho_{X,Y} \frac{\sigma_X}{\sigma_Y}$ et $\beta = E(X) - \alpha E(Y)$.

Dans le cadre de la théorie de la régression linéaire, le couple de v.a. (X, Y) fait l'objet d'observations donnant lieu à un échantillon $(x_i, y_i)_{i=1, \dots, n}$ à partir duquel on estime la pente a et l'ordonnée à l'origine b , en utilisant la méthode des moindres carrés.

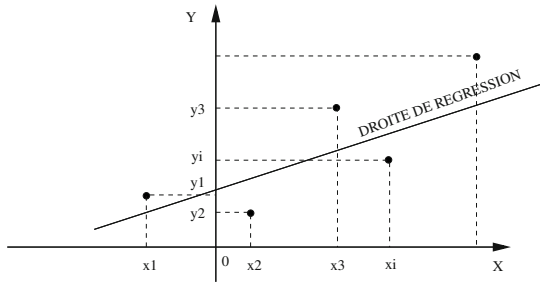


Figure 6.5 – Régression linéaire.

Remarque 6.1 (a) Les droites $y = ax + b$ et $x = \alpha y + \beta$ ne sont généralement pas confondues, sauf dans le cas où $\rho_{X,Y} = \pm 1$.

(b) La relation fonctionnelle entre Y et X peut évidemment relever d'un modèle non affine, par rapport aux paramètres inconnus, de forme analytique quelconque (polynomiale, exponentielle ou autre). Bien que de structure identique, le problème qualifié alors de régression non linéaire, plus délicat à résoudre, nécessite généralement l'utilisation de méthodes numériques.

(c) Le problème de régression linéaire est généralisable au cas d'une v.a. Y conditionnée par un ensemble de n v.a. X_1, X_2, \dots, X_n ; il sera alors question de régression multiple ou multilinéaire.

On rassemble ici les deux types d'approximation quadratique d'une v.a. Y en fonction d'une v.a. X :

Théorème 6.15 *Si (X, Y) est un vecteur aléatoire tel que $E(X^2)$ et $E(Y^2)$ soient finis, les propriétés suivantes sont vérifiées :*

(1) *La meilleure approximation quadratique de Y par une fonction affine dépendant de X est la v.a. \hat{Y} :*

$$\hat{Y} = \left(\rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \right) X + \left(E(Y) - \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} E(X) \right).$$

(2) *L'espérance conditionnelle $E(Y | X)$ est la meilleure approximation quadratique de Y par une fonction quelconque $\varphi(X)$:*

$$\text{pour toute fonction } \varphi : E((Y - E(Y | X))^2) \leq E((Y - \varphi(X))^2).$$

La construction de l'espérance conditionnelle dans le cadre général de la théorie de l'intégration exposée ci-dessous peut ne pas être abordée lors d'une première approche. La structure de cette construction, moins intuitive que la construction hilbertienne, repose sur le théorème suivant :

Théorème 6.16 *Soient Y une v.a. réelle intégrable définie sur l'espace probabilisé (Ω, \mathcal{T}, P) , et \mathcal{T}^* une sous-tribu de \mathcal{T} . Il existe une v.a. Y^* réelle, définie sur Ω vérifiant :*

Y^ est \mathcal{T} -mesurable, P -intégrable ; $\forall E \in \mathcal{T}^*$*

$$\int_E Y(\omega) dP(\omega) = \int_E Y^*(\omega) dP(\omega). \quad (6.1)$$

*Toute v.a. Y^{**} vérifiant (6.1) est P -presque sûrement égale à Y^* ($Y^{**} = Y^*$ sauf éventuellement sur un ensemble de probabilité nulle).*

Définition 6.12 *On appelle espérance conditionnelle de Y sachant \mathcal{T}^* , notée $E(Y | \mathcal{T}^*)$, un représentant de la classe d'équivalence des v.a. Y^* vérifiant (6.1).*

Soit une v.a. réelle intégrable X , distincte de Y , on lui associe la sous-tribu \mathcal{T}_X de \mathcal{T} qu'elle engendre, définie par :

$$\mathcal{T}_X = \{X^{-1}(B) \mid B \in \beta_{\mathbb{R}}\}.$$

D'après le théorème précédent, l'espérance conditionnelle $E(Y \mid \mathcal{T}_X)$ est donc une v.a. Y^* \mathcal{T}_X^* -mesurable, unique P -presque sûrement, vérifiant :

$$\forall E \in \mathcal{T}_X, \quad \int_E Y^*(\omega) dP(\omega) = \int_E Y(\omega) dP(\omega).$$

Un théorème classique de la théorie de la mesure et de l'intégration affirme que Y^* est nécessairement de la forme $Y^* = \varphi^*(X)$ où φ^* est une fonction borélienne. On peut alors énoncer le théorème suivant :

Théorème 6.17 *Pour tout couple de v.a. réelles intégrables (X, Y) , il existe une fonction borélienne φ^* telle que :*

$$\forall E \in \mathcal{T}_X, \quad \int_E Y(\omega) dP(\omega) = \int_E \varphi^*(X)(\omega) dP(\omega). \quad (6.2)$$

$\varphi^*(X)$ est l'espérance conditionnelle construite précédemment.

L'unicité P -presque sûre de Y^* se traduit par l'unicité P_X -presque sûre de la fonction φ^* , qui est un des représentants des fonctions égales à φ^* sauf éventuellement sur un ensemble de probabilité nulle, par rapport à P_X .

Remarque 6.2 Pour $E = \Omega$, l'équation (6.2) s'identifie au théorème de l'espérance totale : $E(Y) = E(E(Y \mid X))$.

En effet :

$$E(Y) = \int_{\Omega} Y(\omega) dP(\omega) = \int_{\mathbb{R}} y dP_Y(y),$$

que l'on calcule autrement :

$$\begin{aligned} \int_{\Omega} E(Y \mid X)(\omega) dP(\omega) &= \int_{X(\Omega)} E(Y \mid X) dP_X = \int_{\mathbb{R}} E(Y \mid X = x) dP_X(x) = E(E(Y \mid X)) \\ &= E(Y). \end{aligned}$$

6.4 Variance conditionnelle

Définition 6.13 *Étant donné deux v.a. X et Y , la variance conditionnelle de Y sachant X est la v.a. définie par :*

$$E((Y - E(Y \mid X))^2 \mid X) =_{(noté)} V(Y \mid X).$$

$V(Y \mid X = x)$ exprime la dispersion quadratique de Y autour de l'espérance conditionnelle $E(Y \mid X = x)$.

Théorème 6.18 (Propriété de la variance totale)

Pour tout couple de v.a. (X, Y) :

$$V(Y) = E(V(Y|X)) + V(E(Y|X)).$$

Comme on l'a déjà mentionné, dans le cadre de la géométrie hilbertienne définie sur $L^2(\Omega, \mathcal{F}, P)$, la propriété de la variance totale équivaut au théorème de Pythagore appliqué au triangle rectangle de côtés Y , $E(Y)$, $E(Y|X)$.

Preuve $V(Y) = E((Y - E(Y|X) + E(Y|X) - E(Y))^2)$
 $= E((Y - E(Y|X))^2) + 2E((Y - E(Y|X))(E(Y|X) - E(Y))) + E((E(Y|X) - E(Y))^2).$

Le dernier terme est égal à $V(E(Y|X))$, le premier terme est égal à $E(V(Y|X))$; quant au terme médian, il est nul, puisque $Y - E(Y|X)$ est orthogonal à H_X , donc à $E(Y|X) - E(Y)$. \square

6.14 ♠ ♠ Espérance et variance de la taille d'une population constituée d'un nombre aléatoire de sous-populations

Soit un nombre aléatoire N de populations de tailles décrites par les v.a. $(X_i)_{i=1, \dots, N}$ indépendantes entre elles, de même loi que X et indépendante de N : posons $S = \sum_{i=1}^N X_i$ étant la taille de la population totale.
 Démontrer : $E(S) = E(N)E(X)$ et $V(S) = E(N)V(X) + V(N)(E(X))^2$.

6.15 ♠ ♠ Soient $(X_n)_n$ une suite de v.a. indépendantes identiquement distribuées, de densité f_X et N une v.a. entière positive, indépendante des X_n . Posons $S_N = \sum_{i=1}^N X_i$.

(1) Déterminer la loi conditionnelle de N sachant $S_N = x$, puis la loi de S_n où n est fixé. Calculer l'espérance de S_N .

(2) Application : $X \sim \mathcal{E}(\lambda)$ et N est la v.a. géométrique $\mathcal{G}(p)$.
 [On rappelle que la somme de n v.a. indépendantes $\mathcal{E}(\lambda)$ suit une loi gamma $\gamma(\lambda, n)$.]

6.16 ♠ ♠ Soient une v.a. exponentielle X de paramètre λ et la v.a. Y dont la densité conditionnelle $f_{Y|X=x}$ par rapport à $X = x$ est égale à $\mathbf{1}_{[x, x+1]}$.

(1) Déterminer la loi de (X, Y) et son support. Calculer $E(Y|X=x)$, $E(Y)$, $E(Y^2|X=x)$, $E(Y^2)$.

(2) Déterminer la densité de Y .

(3) Soit la suite $(T_n)_n$ définie par $T_n = \frac{1}{n} \sum_{i=1}^n (Y_i - \frac{1}{2})$. Déterminer la v.a. vers laquelle la suite $(T_n)_n$ converge en probabilité, les Y_i étant de même loi que Y .

(4) Quelle est la loi limite de la suite définie par $W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \frac{1}{\lambda} - \frac{1}{2})$?

6.5 Loi et espérance conditionnelle des vecteurs gaussiens

On utilise les mêmes notations qu'au chapitre 3.5 auquel il convient de se reporter avant d'aborder ce paragraphe. Le théorème fondamental du conditionnement gaussien nous dit que la meilleure approximation $E(Y|X)$ d'un vecteur gaussien Y , par une fonction du vecteur gaussien conditionnante X , est toujours une fonction affine de la forme $AX + B$.

Théorème 6.19 *Soit le vecteur gaussien de dimension deux (X, Y) , d'espérance mathématique $m = (E(X), E(Y))$ et de coefficient de corrélation $\rho_{(X,Y)}$. L'espérance conditionnelle $E(Y|X)$ s'exprime sous la forme de la fonction affine de X définie par :*

$$E(Y|X) = AX + B \text{ où } A = \frac{\rho_{(X,Y)} \sigma_Y}{\sigma_X} \text{ et } B = E(Y) - AE(X);$$

$$V(Y|X = x) = \sigma_Y^2(1 - \rho_{(X,Y)}^2);$$

$$f_{Y|X=x}(y) = \frac{1}{\sigma_Y \sqrt{(1 - \rho_{(X,Y)}^2)} \sqrt{2\pi}} e^{(-\frac{(y - E(Y|X=x))^2}{2V(Y|X=x)})}.$$

On généralise le théorème précédent au cas de deux vecteurs gaussiens X et Y sous la condition que leur ensemble (X, Y) soit un vecteur gaussien.

Théorème 6.20 (du conditionnement gaussien général) *Soit $(X_1, X_2, \dots, X_{n_1}, Y_1, \dots, Y_{n_2})$ un vecteur gaussien de dimension $n_1 + n_2$ constitué par les vecteurs $X = (X_1, X_2, \dots, X_{n_1})$ et $Y = (Y_1, \dots, Y_{n_2})$.*

Posons : $m_X = E(X)$, $m_Y = E(Y)$, $\Gamma = \begin{pmatrix} \Gamma_X & \Gamma_{X,Y} \\ \Gamma_{Y,X} & \Gamma_Y \end{pmatrix}$ la matrice de covariance de (X, Y) . L'espérance conditionnelle $E(Y|X)$ est le vecteur de dimension n_2 défini par la fonction affine de X

$$E(Y|X) = m_Y + \Gamma_{Y,X} \Gamma_X^{-1} (X - m_X)^t,$$

et de matrice de covariance

$$\Gamma_Y - \Gamma_{Y,X} \Gamma_X^{-1} \Gamma_{X,Y}.$$

Les lois conditionnelles de l'un des vecteurs X ou Y par rapport à l'autre sont gaussiennes. L'espérance de Y sachant $(X = x)$ est définie par :

$$E(Y|X = x) = m_Y + \Gamma_{Y,X} \Gamma_X^{-1} (x - m_X).$$

6.17 ♠♠ Soit le vecteur gaussien centré (X_1, X_2, \dots, X_n) de matrice de covariance Γ , il s'agit de déterminer $E(X_n | X_1, X_2, \dots, X_{n-1})$.

(1) Déterminer $E(X_n X_i)$ pour tout $i = 1, \dots, n-1$; en déduire un système linéaire de $n-1$ équations dont la solution permet d'exprimer

$$E(X_n | X_1, X_2, \dots, X_{n-1}).$$

(2) Appliquer cette méthode au vecteur (X_1, X_2, X_3) pour lequel

$$\Gamma = \begin{pmatrix} 3 & 6 & 2 \\ 6 & 14 & 3 \\ 2 & 3 & 6 \end{pmatrix}.$$

En déduire l'espérance conditionnelle de X_3 sachant (X_1, X_2) .

6.6 Notions d'entropie et d'information

La notion d'entropie fut introduite en 1854 par le thermodynamicien allemand Rudolph Clausius (1822-1888) en thermodynamique, pour quantifier l'état de désordre moléculaire d'un système. La théorie de l'information s'est développée dans les années 1950 autour d'une réflexion critique sur les modèles et les techniques de communication. Le but que s'étaient assigné, en 1948, ses fondateurs Shannon (1916-2001) et Weaver (1894-1978), ingénieurs en génie électrique aux Laboratoires Bell, était de mathématiser les concepts d'entropie et d'information dans le but de résoudre les problèmes posés par le codage et l'optimisation de la transmission des messages. Shannon, s'inspirant de l'expression de l'entropie thermodynamique de Boltzmann, définit d'abord l'entropie d'une loi de probabilité discrète. Il s'agit d'une application H attribuant à toute distribution de probabilité finie (p_1, p_2, \dots, p_n) , son entropie $H(p_1, p_2, \dots, p_n)$, définie par les propriétés suivantes :

(1) continuité par rapport à (p_1, p_2, \dots, p_n) ;

(2) invariance par permutation : $H(p_1, p_2, \dots, p_n) = H(p_{\sigma(1)}, \dots, p_{\sigma(n)})$ pour toute permutation σ de l'ensemble $\{1, 2, \dots, n\}$;

(3) $H(p_1, p_2, \dots, p_n) = H(p_1 + p_2, \dots, p_n) + (p_1 + p_2)H(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2})$.

Shannon construisit la fonction H vérifiant (1), (2), (3) :

$$H(p_1, p_2, \dots, p_n) = - \sum_{i=1}^n p_i \log_2(p_i).$$

Les mathématiciens russes Fadeev et Khinchine démontrèrent en 1957 l'unicité de cette fonction. Le logarithme de base deux (\log_2) permet d'attribuer

une entropie égale à l'unité à la distribution de probabilité $(\frac{1}{2}, \frac{1}{2})$ qui, comme chacun sait, possède l'incertitude maximale : en effet tout autre schéma probabiliste binaire de la forme $(p, 1-p)$ avec $p \neq \frac{1}{2}$, a une issue moins incertaine.

L'entropie $H(p_1, p_2, \dots, p_n)$ est donc une mesure de l'incertitude à laquelle est confronté tout observateur avant la réalisation d'une v.a. discrète de loi (p_1, p_2, \dots, p_n) .

Théorème 6.21 (Propriétés de l'entropie)

(1) $H(p_1, p_2, \dots, p_n)$ est maximale si et seulement si (p_1, p_2, \dots, p_n) est équidistribuée ($\forall i, p_i = \frac{1}{n}$).

(2) $H(1, 0, 0, \dots, 0) = H(0, 1, 0, \dots, 0) = \dots = H(0, 0, \dots, 1) = 0$.

Preuve Pour démontrer (1), démontrons d'abord que pour toute distribution $(p'_1, p'_2, \dots, p'_n)$, $-\sum_{i=1}^n p_i \log_2(p_i) \leq -\sum_{i=1}^n p_i \log_2(p'_i)$ (inégalité de Gibbs).

Sachant que $\forall x$ réel, $\log_2 x \leq x - 1$: $\log_2(\frac{p_i}{p'_i}) \leq \frac{p_i}{p'_i} - 1$. On en déduit :

$\sum_i p_i \log_2(\frac{p_i}{p'_i}) \leq \sum_i (p_i - p'_i) = 0$ et le résultat attendu.

Si $p_i = \frac{1}{n}$, le maximum de H est atteint et vaut $\log_2 n$.

La démonstration et l'interprétation de la propriété (2) sont triviales. \square

6.18 ♠ (1) Démontrer que l'application H vérifie les axiomes (1), (2), (3).

(2) Soit deux urnes contenant des boules : la première contient 4 boules blanches, 3 noires, 2 rouges ; la seconde contient 3 blanches, 5 noires et 1 rouge. Quelle est la distribution de plus grande entropie ? La conclusion confirme-t-elle votre intuition ?

Le concept d'entropie a été étendu de façon naturelle aux distributions de probabilités discrètes infinies ou définies par des densités.

Définition 6.14 Une v.a. X de densité f_X possède l'entropie $H(X)$, dite différentielle :

$$H(X) = \int_{\mathbb{R}} -f_X(x) \ln(f_X(x)) dx,$$

définie si $f_X \ln(f_X) \in L^1(\mathbb{R})$.

(À la différence de l'entropie d'une distribution discrète, l'entropie différentielle peut prendre des valeurs négatives.)

6.19 ♠♠ Caractérisation de certaines lois de probabilités par la maximalité de leurs entropies

(1) Démontrer, grâce à l'équation d'Euler-Lagrange du calcul des variations, que de toutes les v.a. centrées de support \mathbb{R} , de variance σ^2 fixée, la v.a. normale est celle qui maximise l'entropie, laquelle est égale à $\frac{1}{2} \log_2(\sigma^2 2e\pi)$. Il s'agit là d'une remarquable et nouvelle caractérisation de la v.a. normale, parmi de nombreuses autres exposées dans ce cours. Interprété dans le cadre de la théorie des signaux aléatoires, ce résultat nous dit qu'à puissance égale P , le bruit gaussien centré est celui qui a la plus grande entropie. En effet, P est égal à $E(X^2)$: la loi qui maximise l'entropie $\frac{1}{2} \log_2(\sigma^2 2e\pi)$ sous la contrainte $E(X^2) \leq P$ est la loi gaussienne $\mathcal{N}(0, \sigma^2)$.

On démontre de façon analogue que parmi les v.a. définies sur \mathbb{R}_+ de moyenne donnée m , la v.a. exponentielle de paramètre $\lambda = \frac{1}{m}$ est celle qui a l'entropie maximale.

(2) Utiliser la méthode des multiplicateurs de Lagrange, pour démontrer que de toutes les v.a. définies sur \mathbb{N}^* d'espérance m , la v.a. d'entropie maximale est la v.a. géométrique de paramètre $p = \frac{1}{m}$.

Intermède historique : l'entropie en physique statistique

Ludwig Boltzmann (1844-1906), fondateur de la physique statistique, proposa dans le cadre de ses travaux la célèbre formule $S = k \log(W)$ qui met en relation le nombre W d'états possibles d'un grand nombre de particules d'un système donné avec l'entropie physique S du système.

Supposons un ensemble de n particules indiscernables telles que chacune d'elles se trouve dans l'un des états $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_k$: une configuration de l'ensemble des particules est donc décrite par le k -uplet (n_1, \dots, n_k) , où n_i est le nombre de particules occupant l'état \mathcal{E}_i . On suppose n très grand (multiple de 10^{23}). Le nombre W de configurations correspondant à l'état (n_1, \dots, n_k) est égal au coefficient multinomial :

$$W = \frac{n!}{n_1! \dots n_k!}.$$

L'entropie physique moyenne notée H_B est définie par :

$$H_B(n) = \frac{1}{n} \log_2 W.$$

Compte tenu de la grandeur de n , on démontre grâce à la formule de Stirling que :

$$\lim_{n \rightarrow +\infty} H_B(n) = - \sum_{i=1}^k p_i \log_2(p_i), \text{ où } p_i = \frac{n_i}{n}.$$

6.20 ♠♠ Selon le deuxième principe de la thermodynamique, un système isolé en état d'équilibre se trouve toujours dans l'état d'entropie maximale. Supposons n particules et k états possibles d'énergie $\{E_i\}_i$; déterminer la configuration

d'entropie maximale des n particules, sachant que l'énergie moyenne de chaque particule est égale à $E = \sum_{i=1}^k p_i E_i$. [Indication : il s'agit, comme dans l'exercice précédent, d'un problème d'optimisation sous contraintes résoluble par la méthode de Lagrange.] Démontrer que l'on parvient à la distribution de probabilité de Boltzmann :

$$\forall i = 1, \dots, k \quad P(\text{système ait l'énergie } E_i) = C \exp(-\alpha E_i),$$

où C et α sont des constantes positives, que l'on n'essaiera pas d'identifier.

Une approche constructive de l'entropie et de l'information

Si b est un entier et x un réel, tous deux supérieurs à 1, $\log_b(x)$ est une approximation de la longueur du vecteur codant l'écriture de x dans une base de longueur b ; ainsi, pour $b = 2$ et $x = 30$, la partie entière de $\log_2(30)$ est égale à 4, longueur du codage du nombre 30 dans la base $\{0, 1\}$. La propriété de la fonction logarithme $\log_b(x) + \log_b(y) = \log_b(x.y)$ s'interprète en disant que la longueur du vecteur codant le nombre $x.y$ est à peu près égale à la somme des longueurs des codages de x et y .

Ce qui vient d'être dit pour les nombres s'étend aux ensembles : en effet, toute partie d'un ensemble E de n éléments peut être codée par un nombre de bits égal à $b = (\text{Partie entière } \log_2(n)) + 1$ (il suffit de mettre en bijection une partie quelconque de E et le n -uplet (a_1, \dots, a_b) formé de 0 et de 1, qui la décrit).

Considérons un message $M = (m_1, \dots, m_r)$ de longueur r ; il s'écrit avec les lettres d'un alphabet donné $\mathcal{A} = \{a_1, \dots, a_n\}$ et $M \in \mathcal{A}^r$. Soit r_i le nombre d'occurrences de a_i dans le message M , alors $r = \sum_{i=1}^n r_i$, et $\frac{r_i}{r}$ est égal à la fréquence d'occurrences de a_i dans le message M . Ainsi, pour transmettre M , on doit d'abord transmettre r_1, \dots, r_n ce qui nécessite au moins $n \log_2(r)$ bits, puis un nombre au plus égal au nombre $c = \frac{r!}{r_1! \dots r_n!}$ de messages possibles, qui contient $\log_2(c)$ informations. Le nombre total $I(r)$ de bits requis est donc compris entre $\log_2(c)$ et $\log_2(c) + n \log_2(r)$. On démontre, grâce à la formule de Stirling, que la quantité moyenne de bits par symbole transmis est asymptotiquement égale à $-\sum_{i=1}^n p_i \log_2(p_i)$ où $p_i = \lim_{r \rightarrow +\infty} \frac{r_i}{r}$. (voir : B.Bercu et D.Chafaï . *Qu'est-ce que la simulation ?*)

La notion d'information nécessite l'extension de la notion d'entropie à des distributions probabilistes de supports inclus ou égaux à \mathbb{R}^k ; on présente le cas d'un vecteur aléatoire à deux dimensions.

Définition 6.15 *L'entropie du vecteur aléatoire (X, Y) , de loi $(P(X = x_i, Y = y_j) \equiv p_{i,j})_{i \in I, j \in J}$ est définie par :*

$$H(X, Y) = - \sum_{i,j} p_{i,j} \log_2 p_{i,j}.$$

Théorème 6.22

$$(X \text{ et } Y \text{ indépendantes}) \Leftrightarrow (H(X, Y) = H(X) + H(Y)).$$

Preuve $(X \text{ et } Y \text{ indépendantes}) \Rightarrow \forall i, j \quad p_{i,j} = p_{i.} p_{.j}$.

$$\begin{aligned} H(X, Y) &= - \sum_{(i,j)} p_{i,j} \log_2 p_{i,j} = - \sum_{(i,j)} p_{i.} p_{.j} \log_2 (p_{i.} p_{.j}) \\ &= - \sum_{(i,j)} p_{i.} p_{.j} (\log_2(p_{i.}) + \log_2(p_{.j})) \\ &= - \sum_{(i,j)} p_{i.} p_{.j} \log_2(p_{i.}) - \sum_{(i,j)} p_{i.} p_{.j} \log_2(p_{.j}) \\ &= - \sum_i p_{i.} \log_2(p_{i.}) - \sum_j p_{.j} \log_2(p_{.j}) = H(X) + H(Y). \end{aligned}$$

Réciproquement, $-\sum_{(i,j)} p_{i,j} \log_2 p_{i,j} = -\sum_i p_{i.} \log_2(p_{i.}) - \sum_j p_{.j} \log_2(p_{.j})$ entraîne que :

$$- \sum_{(i,j)} p_{i,j} (\log_2 p_{i,j} - \log_2(p_{i.} p_{.j})) = - \sum_{(i,j)} p_{i,j} \cdot (\log_2 \frac{p_{i,j}}{p_{i.} p_{.j}}) = 0.$$

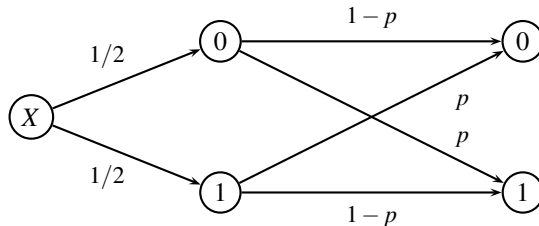
$\frac{p_{i,j}}{p_{i.} p_{.j}} \leq 1$, implique que l'équation précédente n'est vérifiée que si $p_{i,j} = p_{i.} p_{.j}$. \square

Définition 6.16 L'*information mutuelle* entre les événements $(X = x_i)$ et $(Y = y_j)$ est définie par :

$$\log_2 \left(\frac{P(X = x_i | Y = y_j)}{P(X = x_i)} \right) \equiv \text{(notée)} \quad I(x_i, y_j).$$

Exemple 6.8 Le canal binaire symétrique

Il s'agit du canal de transmission mettant en relation un émetteur susceptible d'envoyer un 0 ou un 1 avec la probabilité $\frac{1}{2}$, à un récepteur recevant chacun de ces messages avec la probabilité $(1 - p)$.



On suppose que X décrit le message émis et Y le message reçu.

$$\begin{aligned} P(Y=1, X=1) &= P(Y=1 | X=1)P(X=1) = P(Y=0, X=0) = \frac{1-p}{2}; \\ P(Y=1, X=0) &= P(Y=1 | X=0)P(X=0) = P(Y=0, X=1) = \frac{p}{2}. \end{aligned}$$

Calculons les informations mutuelles :

$$\begin{aligned} I((X=1), (Y=1)) &= I((X=0), (Y=0)) = \log_2(1-p); \\ I((X=1), (Y=0)) &= \log_2 p. \end{aligned}$$

Interprétation :

- si $p = \frac{1}{2}$, toutes les informations mutuelles sont nulles : le message émis et le message reçu sont indépendants ;

- si $p < \frac{1}{2}$, on en déduit que si 1 est reçu, la probabilité pour que 1 ait été émis est supérieure à $\frac{1}{2}$.

Définition 6.17 *L'information moyenne du vecteur aléatoire (X, Y) est définie par :*

$$I(X, Y) = \sum_{(i,j)} p_{i,j} I(x_i, y_j) = \sum_{(i,j)} p_{i,j} \log_2 \frac{p_{i,j}}{p_{\cdot j} p_{i \cdot}}.$$

Définition 6.18 *L'entropie conditionnelle moyenne de X connaissant Y est définie par :*

$$H(X/Y) = H(X, Y) - H(Y) = - \sum_{(i,j)} p_{i,j} \log \frac{p_{i,j}}{p_{\cdot j}}.$$

L'entropie conditionnelle moyenne $H(X|Y)$ mesure l'incertitude résiduelle moyenne de X lorsque l'on connaît Y . L'information $I(X, Y)$ apportée par la réalisation de Y sur X est égale à la diminution d'entropie de X due à la réalisation de Y :

$$I(X, Y) = H(X) - H(X/Y).$$

Théorème 6.23 *(Propriétés de l'entropie et de l'information moyenne d'un vecteur aléatoire)*

(1) *Symétrie : l'information apportée par Y sur X est égale à l'information apportée par X sur Y ; d'où le nom d'information mutuelle.*

$$I(X, Y) = I(Y, X).$$

(2) *Positivité :*

$$I(X, Y) \geq 0.$$

Interprétation : la connaissance d'une occurrence d'une des deux v.a. diminue l'incertitude sur l'occurrence de la deuxième.

(3) *Extensivité : $H(X, Y) \leq H(X) + H(Y)$.*

Interprétation : l'incertitude moyenne des occurrences d'un couple de v.a. est toujours inférieure à la somme des incertitudes de chaque v.a. L'égalité étant vérifiée dans le seul cas où X et Y sont indépendantes.

Illustration : l'entropie d'une langue naturelle

Toute langue naturelle utilise en moyenne chaque lettre de l'alphabet avec une probabilité qui lui est propre et que l'on peut estimer grâce à un échantillon représentatif des textes écrits dans ladite langue. Pour le français, c'est la lettre E qui arrive en tête avec une fréquence égale à 0,175, puis A avec une fréquence égale à 0,084, puis apparaissent successivement S, I et T avec les fréquences respectives 0.079, 0.073 et 0.07 ; l'entropie moyenne d'une lettre vaut 3,98 bits.

Considérant cette fois les couples de lettres consécutives comme occurrences du vecteur aléatoire (X_1, X_2) , on calcule :

$$I(X_2, X_1) = H(X_2) - H(X_2 | X_1) < 3,98.$$

Cette inégalité confirme une loi linguistique : toute occurrence d'une lettre à l'intérieur d'un mot est dépendante des occurrences des lettres voisines précédente et suivante. Par exemple, chacun sait qu'en français la probabilité de trouver deux W ou deux Y consécutifs est nulle. Des linguistes ont poursuivi ce type d'étude pour l'ensemble de n -uplets de lettres : ils ont ainsi pu estimer par exemple des probabilités conditionnelles de la forme $P_{X_3|(X_1, X_2)}$ dans le cas des triplets (X_1, X_2, X_3) parcourant les lettres de l'alphabet. Opérant ensuite par simulation, ils ont généré, grâce à la connaissance des estimations de probabilités conditionnelles, des suites de séquences de lettres d'autant plus proches des mots de la langue considérée que les lettres conditionnantes étaient nombreuses. Dans le cas des triplets d'occurrences de lettres des mots anglais, on obtient des séquences simulées qui ressemblent ou sont même des mots anglais, tels que : *in, no, ist, last, crat, forthe, the, birs*. Intéressant, non ?

6.7 Retour sur la notion de hasard

Une approche contemporaine, fondée sur les suites aléatoires, propose une conception radicalement nouvelle et utile de la notion de hasard. Dans le cadre de ses travaux sur les fondements de la théorie des probabilités, le mathématicien américain von Mises (1883-1953) s'était posé le problème de la caractérisation de toute suite infinie aléatoire équirépartie, composée de 0 et de 1. La question était la suivante : de toutes les suites infinies formées de 0 et de 1, lesquelles peuvent-être considérées comme aléatoirement équiréparties, sachant qu'elles résultent de lancers successifs et indépendants d'une pièce équilibrée où pile vaut 1 et face vaut 0 ?

L'hypothèse la plus naturelle avancée pour caractériser une suite aléatoire est qu'elle ne doit pas résulter de l'application d'une règle de production déterministe, ce qui exclut par exemple les suites $(0,1,0,1,0,1,0,1,...)$

ou $(0,0,1,0,0,0,1,1,0,0,1,0,0,0,1,1,0,\dots)$ de motifs périodiques respectifs égaux à $(0,1)$ et à $(0,0,1,0,0,0,1,1)$. Mais alors, que penser de la suite de Champernowne égale à $(0,1,1,0,1,1,1,0,0,1,0,1,1,1,0,1,\dots)$ d'apparence aléatoire, alors qu'elle est définie par la suite des entiers exprimés en base deux ?

La deuxième hypothèse prise en compte est l'hypothèse de normalité qui postule qu'une suite infinie de 0 et de 1 est dite normale si :

- (a) les fréquences d'apparition de 1 et 0 sont égales à $\frac{1}{2}$;
- (b) les fréquences de chacun des couples $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$ sont égales à $(\frac{1}{2})^2$, et généralement celles des 2^n -uplets sont égales à $(\frac{1}{2})^n$.

On démontre que la condition de normalité est nécessaire, mais non suffisante pour assurer le caractère aléatoire d'une telle suite.

En 1966, le mathématicien Martin-Löf mit un terme (provisoire!) au problème de la caractérisation d'une suite aléatoire. La présentation rigoureuse de sa théorie nécessitant des notions de logique avancée telle que la récursivité ne saurait être abordée ici ; on en résumera toutefois l'idée centrale en disant qu'une suite de 0 et de 1 est aléatoire, si son information est incompressible au sens où l'algorithme de longueur minimale la générant contient plus d'information que la suite qu'il permet de produire. On conçoit en effet qu'il est plus facile de générer la suite $(0,1,0,1,0,1,0,1,0,1,0,\dots)$ qu'une suite produite par un générateur aléatoire.

La mathématisation de la notion de suite aléatoire relève de la théorie de la complexité et fait l'objet de recherches très actives. Pour un exposé accessible et détaillé de ces questions, on pourra consulter par exemple : J.P Delahaye, *Information, complexité et hasard* ou B.Bercu et D.Chafaï, *Qu'est-ce que l'information ?*

Les problèmes 13 et 14 se rapportent à ce chapitre.

6.8 Corrigés des exercices

Exercice 6.1

Soit E l'événement « les deux personnes énoncent la même assertion » et A_i l'événement « le i -ème politicien ne ment pas », alors $E = (A_1 \cap A_2) \cup (\bar{A}_1 \cap \bar{A}_2)$ et la probabilité cherchée est $P((A_1 \cap A_2)/E)$.

Or $P((A_1 \cap A_2) \cap E) = P(A_1 \cap A_2) = P(A_1)P(A_2)$ et $P(E) = P(A_1 \cap A_2) + P(\bar{A}_1 \cap \bar{A}_2) = P(A_1)P(A_2) + P(\bar{A}_1)P(\bar{A}_2)$, car les deux personnes agissent indépendamment. Avec $P(A_1) = \frac{3}{4}$ et $P(A_2) = \frac{4}{5}$, on obtient $P(A_1 \cap A_2) = \frac{12}{20}$ et $P(E) = \frac{12}{20} + \frac{1}{20} = \frac{13}{20}$, d'où :

$$P((A_1 \cap A_2)|E) = \frac{12}{13} \approx 0,92.$$

Exercice 6.2

(1) Propriété d'absence de mémoire.

$$\begin{aligned} P([X > m+n]/[X > n]) &= \frac{P([X > m+n])}{P([X > n])} = \frac{\sum_{k=m+n+1}^{+\infty} (1-p)^{k-1}p}{\sum_{k=n+1}^{+\infty} (1-p)^{k-1}p} \\ &= (1-p)^m = P([X > m]). \end{aligned}$$

(2) La loi exponentielle étant l'analogue continu de la loi géométrique (voir chapitre 2), on pouvait s'attendre à ce résultat.

Exercice 6.3

Taux de réussite $\frac{60 \times 55}{100^2} + \frac{80 \times 45}{100^2} = 69\%$.

Soient A : « être issu de A » et E : « être en échec » :

$$P(A|E) = \frac{P(A \cap E)}{P(E)} = \frac{P(E|A)P(A)}{P(E)} \approx 71\%.$$

Exercice 6.4

Soit M l'événement « être malade du sida » et T_+ l'événement « le test est positif ».

Il s'agit de déterminer $P^{T_+}(\bar{M}) = \frac{P(\bar{M} \cap T_+)}{P(T_+)}$.

On connaît $P(T_+|\bar{M}) = 10^{-3}$, $P(M) = \frac{130}{58\,000} = \frac{13}{5\,800}$, $P(\bar{M}) = 1 - \frac{13}{5\,800} = \frac{5\,787}{5\,800}$ et la fiabilité du test est $P(T_+|M) = 0,998$.

$$P(\bar{M} \cap T_+) = P(T_+|\bar{M})P(\bar{M})$$

$$P(T_+) = P(T_+ \cap M) + P(T_+ \cap \bar{M}) = P(T_+/M)P(M) + P(T_+/\bar{M})P(\bar{M})$$

d'où finalement :

$$\begin{aligned} P(\bar{M}|T_+) &= \frac{P(T_+|\bar{M})P(\bar{M})}{P(T_+/M)P(M) + P(T_+|\bar{M})P(\bar{M})} = \frac{10^{-3} \times \frac{5}{5} \frac{787}{800}}{10^{-3} \times \frac{5787}{800} + 0,998 \times \frac{13}{800}} \\ &= \frac{5,787}{5,787 + 12,974} \approx \frac{5,787}{18,761} \approx 0,31. \end{aligned}$$

Il y a environ 31 % de chances de ne pas avoir le sida alors que le test est positif.

Exercice 6.5

Notations :

rr « la pièce tirée a les deux faces rouges »

rb « la pièce tirée a une face rouge et une face blanche »

bb « la pièce tirée a les deux faces blanches »

R « la face visible de la pièce présentée est rouge »

B « la face visible de la pièce présentée est blanche ».

Il s'agit de calculer $P(rr/R)$.

$$\begin{aligned} P(rr|R) &= \frac{P(rr \cap R)}{P(R)} = \frac{P(R|rr)P(rr)}{P(R|bb)P(bb) + P(R|rb)P(rb) + P(R|rr)P(rr)} \\ &= \frac{\frac{1}{3}}{0 + \frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{2}{3}. \end{aligned}$$

On s'attendait plutôt à une probabilité $1/2$!

Exercice 6.6

• Si le prisonnier ne change pas, pour qu'il gagne, il faut qu'il ait pris la bonne porte dès le début, soit 1 chance sur 3.

• Si le prisonnier change de porte après l'indication, pour gagner, il faut qu'au départ il ait choisi une mauvaise porte : probabilité égale à $2/3$, car alors le directeur est obligé d'ouvrir l'autre mauvaise porte et, en changeant, le prisonnier tombe sur la bonne porte. La meilleure stratégie consiste donc à changer.

Exercice 6.7

$X \sim \mathcal{P}(a)$, $Y \sim \mathcal{P}(b)$ et X et Y sont indépendantes : on sait alors que $X + Y \sim \mathcal{P}(a + b)$, grâce à la stabilité par sommation de v.a. indépendantes poissonniennes.

$$\begin{aligned}
 P^{[X+Y=n]}([X=k]) &= \frac{P([X=k] \cap [X+Y=n])}{P([X+Y=n])} = \frac{P([X=k] \cap [Y=n-k])}{P([X+Y=n])} \\
 &= \frac{P([X=k])P([Y=n-k])}{P([X+Y=n])}
 \end{aligned}$$

par indépendance de X et de Y .

Si $k \notin \llbracket 0, n \rrbracket$, $P^{[X+Y=n]}([X=k]) = 0$, sinon :

$$P^{[X+Y=n]}([X=k]) = \frac{e^{-a} \frac{a^k}{k!} e^{-b} \frac{b^{n-k}}{(n-k)!}}{e^{-(a+b)} \frac{(a+b)^n}{n!}} = C_n^k \frac{a^k b^{n-k}}{(a+b)^n} = C_n^k \left(\frac{a}{a+b} \right)^k \left(\frac{b}{a+b} \right)^{n-k}$$

donc $P_X^{[X+Y=n]}$ est de loi $\mathcal{B}(n, \frac{a}{a+b})$ et $\mathbb{E}^{[X+Y=n]}(X)$ est égal à $n \times \frac{a}{a+b}$. $\frac{a}{a+b}$ est la proportion moyenne de voitures à destination de A ; sachant qu'il y a eu en tout n voitures, le nombre moyen de voitures est $n \times \frac{a}{a+b}$.

Exercice 6.8

$$P(X < Y) = P((X, Y) \in \{(-1, 0), (-1, 1), (0, 1)\}) = 3 \left(\frac{1}{3}\right)^2 = \frac{1}{3}.$$

$$\begin{aligned}
 E(Y|X < Y) &= \sum_y y P(Y=y|X < Y) = \sum_y y \frac{P((Y=y) \cap (Y > X))}{P(X < Y)} \\
 &= -1.P(-1 > X) + 0.P(0 > X) + 1.P(1 > X) = \frac{2}{3}.
 \end{aligned}$$

.

Exercice 6.9

$P([Y=k]) = \binom{n}{k} \int_0^1 x^k (1-x)^{n-k} dx = \frac{n!}{k!(n-k)!} \frac{(n-k)!k!}{(n+1)!} = \frac{1}{n+1}$ donc Y est de loi uniforme sur $\{0, 1, \dots, n\}$.

$$f_{X/Y=k}(x) = \frac{\binom{n}{k} x^k (1-x)^{n-k}}{\left(\frac{1}{n+1}\right)} \mathbf{1}_{[0,1]}(x).$$

Exercice 6.10

(1) $f_{(X,Y)}(x, y) = \frac{1}{2} \mathbf{1}_D(x, y)$ où D est le carré.

$$f_X(x) = \int_{-1}^1 \frac{1}{2} \mathbf{1}_D(x, y) dy = (1+x) \mathbf{1}_{[-1,0]}(x) + (1-x) \mathbf{1}_{[0,1]}(x).$$

Par analogie, $f_Y(y) = (1+y) \mathbf{1}_{[-1,0]}(y) + (1-y) \mathbf{1}_{[0,1]}(y)$.

$f_{X/Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$. Si par exemple, $y \in [0, 1]$,

$$f_{X/Y=y}(x) = \frac{1}{2} \frac{\mathbf{1}_D(x,y)}{(1-y)} = \frac{\mathbf{1}_{[y-1; 1-y]}(x)}{2(1-y)}.$$

(2) Les composantes ne peuvent pas être indépendantes, puisque le support n'est pas égal au produit des supports des composantes.

Exercice 6.12

(1) L'équation $\lambda = -\frac{R'_T(t)}{R_T(t)}$ avec $R_T(0) = 1$ admet la solution $R_T(t) = e^{(-\lambda t)}$: T est, comme on pouvait s'y attendre, de loi exponentielle $\mathcal{E}(\lambda)$.

(2) Pour calculer $\mathbb{E}(T)$, on pose $u = \left(\frac{t-\gamma}{\eta}\right)^\beta$.

$$\mathbb{E}(T) = \int_0^{+\infty} (\eta u^{1/\beta} + \gamma) e^{-u} du = \eta \Gamma\left(1 + \frac{1}{\beta}\right) + \gamma.$$

$$(3) P([T > t+s]/[T > s]) = \frac{R(t+s)}{R(s)} = \exp\left(-\left(\frac{t+\gamma}{\eta}\right)^\beta + \left(\frac{s}{\eta}\right)^\beta\right).$$

Équivalence $P([T > s+t]/[T > s]) > P([T > s])$ si et seulement si $s^\beta - (s+t)^\beta > -t^\beta$.

Posons : $\phi(s) = t^\beta + s^\beta - (s+t)^\beta$.

Premier cas : $0 < \beta < 1$; ϕ croissante et $\lim_{t \rightarrow 0^+} \phi(t) = 0$ donc $\phi(t) > 0$.

Deuxième cas : $\beta > 1$; on aboutit à l'inégalité contraire, par un traitement analogue.

(4) $T = \min_{1 \leq i \leq n} (T_i)$ d'où $R_T(t) = (R(t))^n = \exp\left(-n \left(\frac{t}{\eta}\right)^\beta\right)$: T est donc une v.a. de loi de Weibull $W\left(\beta, \frac{\eta}{n^{1/\beta}}\right)$.

Exercice 6.13

$$\begin{aligned} \mathbb{E}(\phi(X)Y) &= \mathbb{E}(\mathbb{E}(\phi(X)Y | X)) = \int_{\mathbb{R}} \mathbb{E}(\phi(x)Y/[X=x]) f_X(x) dx \\ &= \int_{\mathbb{R}} \phi(x) \mathbb{E}(Y/[X=x]) f_X(x) dx = \mathbb{E}(\phi(X) \mathbb{E}(Y|X)). \end{aligned}$$

Exercice 6.14

$\mathbb{E}(S|N=n) = \mathbb{E}(\sum_{i=1}^N X_i | N=n) = n\mathbb{E}(X)$; d'où $\mathbb{E}(S|N) = N\mathbb{E}(X)$ et $\mathbb{E}(S) = \mathbb{E}(\mathbb{E}(S|N)) = \mathbb{E}(N)\mathbb{E}(X)$.

$$V(S) = \mathbb{E}(S^2) - (\mathbb{E}(N)\mathbb{E}(X))^2 \quad (*)$$

$$\mathbb{E}(S^2) = \sum_n \mathbb{E}(S^2/[N=n]) \cdot P([N=n]); \quad V(S_n) = nV(X) = \mathbb{E}(S_n^2) - (\mathbb{E}(S_n))^2.$$

On en déduit :

$$\mathbb{E}(S^2) = \sum_n (nV(X) + n^2(\mathbb{E}(X))^2) P([N=n]) = V(X)\mathbb{E}(N) + (\mathbb{E}(X))^2\mathbb{E}(N^2),$$

et par (*) : $V(S) = \mathbb{E}(N)V(X) + (\mathbb{E}(X))^2V(N)$.

On aurait pu démontrer ce résultat en utilisant le théorème (6.18) de la variance totale.

Exercice 6.15

(1) D'après le théorème 6.8 :

$$\begin{aligned} P([N=n] | [S_N=x]) &= \frac{f_{S_N/N=n}(x)}{f_{S_N}(x)} P([N=n]) \\ &= \frac{f_{S_N/N=n}(x) P([N=n])}{\sum_k f_{S_N/N=k}(x) P([N=k])}. \end{aligned}$$

(2) $f_{S_N/N=n}$ est la convolution de n densités exponentielles $\mathcal{E}(\lambda)$ et est donc égale à la densité de la loi gamma $\gamma(\lambda, n) : \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)}$. On a donc :

$$\begin{aligned} P([N=n] | [S_N=x]) &= \frac{\frac{\lambda^n x^{n-1} e^{-\lambda x} p(1-p)^{n-1}}{\Gamma(n)}}{\sum_{k=1}^{+\infty} \frac{\lambda^k x^{k-1} e^{-\lambda x} p(1-p)^{k-1}}{\Gamma(k)}} \\ &= \frac{\frac{\lambda^n x^{n-1} (1-p)^{n-1}}{\Gamma(n)}}{\lambda e^{\lambda(1-p)x}} = \frac{(\lambda(1-p)x)^{n-1} e^{-\lambda(1-p)x}}{(n-1)!}. \end{aligned}$$

La v.a. $(N-1)$ conditionnée par $(S_N=x)$ est de loi de Poisson $\mathcal{P}(\lambda(1-p)x)$.

Exercice 6.17

D'après le théorème du conditionnement de vecteurs gaussiens :

$$\mathbb{E}(X_n | X_1, X_2, \dots, X_{n-1}) = \sum_{j=1}^{n-1} \alpha_j X_j. \quad (*)$$

Il s'agit donc de déterminer les $(\alpha_j)_j$. Multiplions les membres de (*) par X_i :

$$\mathbb{E}(X_n | X_1, X_2, \dots, X_{n-1}) \cdot X_i = \sum_{j=1}^{n-1} \alpha_j X_j X_i,$$

d'où $\mathbb{E}(\mathbb{E}(X_n | X_1, X_2, \dots, X_{n-1}) X_i) = \sum_{j=1}^{n-1} \alpha_j \mathbb{E}(X_j X_i) = \mathbb{E}(X_n X_i)$.

Si l'on fait varier i dans $\{1, 2, \dots, n-1\}$, on parvient à un système de $(n-1)$ équations linéaires en les $(\alpha_j)_{1 \leq j \leq n-1}$:

$$\left(\sum_j \alpha_j \mathbb{E}(X_i X_j) = \mathbb{E}(X_n X_i) \right)_{1 \leq i \leq n-1}. \quad (**)$$

Le vecteur étant centré, Γ est la matrice des $(\mathbb{E}(X_i X_j))_{i,j}$.

Application : $(**)$ s'écrit $2 = \alpha_1$, $-1 = 4\alpha_2$, soit $\alpha_2 = -\frac{1}{4}$, d'où $\mathbb{E}(X_3/X_1, X_2) = 2X_1 - \frac{1}{4}X_2$.

Exercice 6.18

(1) Les propriétés 1 et 2 sont triviales ; démontrons la propriété 3, en prenant $n = 3$ pour limiter les calculs.

$$H(p_1 + p_2, p_3) = -(p_1 + p_2) \log_2(p_1 + p_2) - p_3 \log_2(p_3)$$

$$(p_1 + p_2)H\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right) = -p_1 \log_2 p_1 + p_1 \log_2(p_1 + p_2) - p_2 \log_2(p_2) + p_2 \log_2(p_1 + p_2).$$

En sommant membre à membre, on obtient l'axiome 3.

(2) Il est facile de vérifier que la première urne a la plus grande entropie.

Exercice 6.19

(1) Application de la méthode de Lagrange d'optimisation sous contraintes.

(2) relève du calcul des variations sous contraintes auquel nous renvoyons le lecteur. L'équation d'Euler-Lagrange s'écrit :

$$L(f)(x) = f(x) \ln f(x) - \lambda_1 f(x) - \lambda_2 x f(x) - \lambda_3 x^2 f(x)$$

où f est la densité recherchée.

D'où : $f(x) = \exp(\lambda_1 - 1 + \lambda_2 x + \lambda_3 x^2)$.

f symétrique $\implies \lambda_2 = 0$ d'où $f(x)$ de la forme $C_1 \exp(-C_2 x^2)$.

$$\int_{\mathbb{R}} f(x) dx = 1 \implies C_2 = \pi C_1^2.$$

$$V(X) = 1 \implies C_1 = \frac{1}{\sqrt{2\pi}} \text{ d'où } f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Démonstration analogue pour (2) : on parvient à $f(x) = \exp(\lambda_2 x + \lambda_1 - 1)$ puis à $f(x) = \frac{1}{m} \exp\left(-\frac{x}{m}\right)$ ($x \geq 0$).

Chapitre 7

Chaînes de Markov discrètes

Les processus aléatoires ou stochastiques ont été conçus pour modéliser l'évolution temporelle de phénomènes et de systèmes aléatoires. Ils sont décrits par des familles discrètes ou continues de variables ou de vecteurs aléatoires $(X_t)_{t \in \mathbb{T}}$, où \mathbb{T} est l'ensemble des temps d'observation des états du processus. Leur vaste domaine d'application concerne la modélisation et le traitement des signaux et des systèmes aléatoires, des séries de données économiques, des phénomènes météorologiques ou sismiques, etc. Cherchant à généraliser le théorème central limite à des ensembles de v.a. dépendantes, le mathématicien russe Andrei Markov (1856-1922) introduisit la notion de chaînes d'événements et produisit les premiers développements de la théorie qui depuis porte son nom.

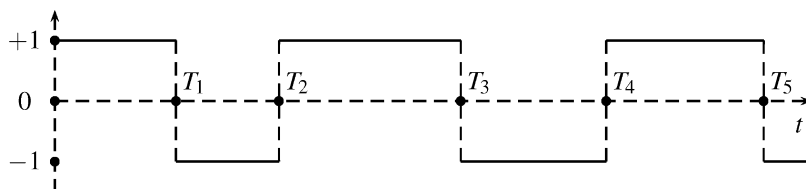
7.1 Introduction aux processus aléatoires

Avant de donner une définition générale des processus aléatoires, on en présente succinctement quelques uns qui illustrent la variété de leurs formes et de leurs champs d'application.

(1) Le signal du téléphone (en anglais : *flip-flop*).

Il s'agit d'un processus $(X_t)_{t \in \mathbb{R}_+}$ à deux états $\{-1, +1\}$ dont l'évolution est régie par les règles suivantes : on suppose qu'au temps $t = 0$, il se trouve dans l'un des deux états, par exemple $(X_0 = 1)$, il y reste une durée aléatoire de loi exponentielle $\mathcal{E}(\lambda)$, et transite brusquement vers l'état -1 dans lequel il séjourne pendant une durée aléatoire indépendante de la précédente et de loi $\mathcal{E}(\mu)$. Ce comportement est itéré indéfiniment ; les instants de transition sont décrits par la suite de v.a. $(T_i)_{i \in \mathbb{N}^*}$.

Graphes d'une réalisation du signal avec $X_0 = +1$:



(2) Le bruit blanc.

Le bruit blanc physique est un phénomène aléatoire généré par des causes externes ou internes aux systèmes qui en perturbent le fonctionnement. Il est présent dans la plupart des systèmes mis en œuvre en traitement du signal, en automatique et plus généralement dans les sciences de l'ingénieur. Le bruit blanc mathématique est le processus aléatoire $(B_t)_{t \in \mathbb{R}_+}$, où chaque v.a. B_t désigne l'amplitude du bruit au temps t . Les variables aléatoires B_t sont centrées et décorrélées entre elles : autrement dit, si proches que soient les temps s et t , $\text{cov}(B_s, B_t) = 0$. Cette hypothèse n'est pas rigoureusement vérifiée par le bruit blanc physique. Le modèle mathématique du bruit blanc est donc une idéalisation du phénomène physique bruit blanc. Il existe plusieurs sortes de bruits blancs, tous caractérisés par une densité spectrale constante ; par exemple, le bruit de grenaille est le modèle d'un courant provoqué par l'arrachement thermique de particules.

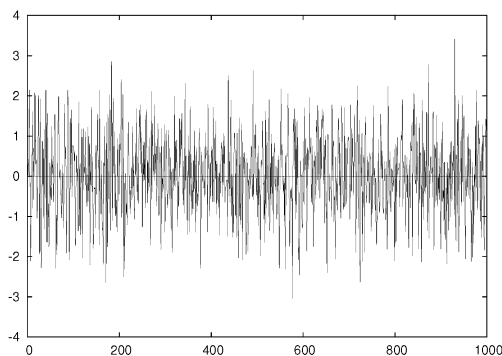


Figure 7.1 – Simulation d'un bruit blanc gaussien.

(3) Processus décrivant le nombre de clients en attente à un guichet.

Ce processus à valeurs dans \mathbb{N} modélise la longueur des files d'attente qui apparaissent à l'entrée des services (guichets d'autoroutes, salles d'attente,...) ou dans les réseaux informatiques ; il fera l'objet d'une étude complète au chapitre 10, consacré aux processus markoviens à temps continu et aux files d'attente.

(4) Marche aléatoire isotrope d'une particule dans un réseau carré du plan.

Il s'agit du modèle décrivant le déplacement d'une particule partant de l'origine d'un repère au temps $t = 0$ et se déplaçant sur un réseau carré infini dont la maille élémentaire est le carré de côté 1 ; à chaque unité de temps elle a la possibilité de faire un saut vers le nord, le sud, l'est ou l'ouest avec la même probabilité $\frac{1}{4}$.

Ci-dessous, le résultat d'une simulation jusqu'au temps $t = 17$, de la marche aléatoire isotrope, définie sur un réseau carré et démarrant à l'origine au temps 0.

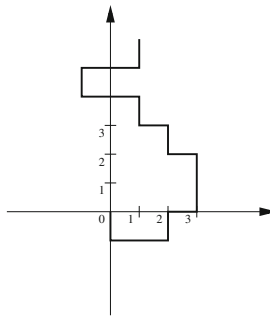


Figure 7.2 – Simulation de la marche aléatoire isotrope dans le plan.

L'étude des processus aléatoires se fait au rythme de questions qui en structurent l'exposé.

(1) Est-il pertinent pour décrire l'évolution de ces processus de vouloir connaître la loi du vecteur $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$, quelle que soit la suite croissante des temps (t_1, \dots, t_n) ? À quelles conditions peut-on se satisfaire de la connaissance des seules probabilités de transition pour décrire l'évolution d'un processus ? Les probabilités de transition sont de la forme :

$$P(X_t = y | X_s = x),$$

où x et y sont des états quelconques, et $s < t$?

(2) Un processus retourne-t-il dans un état qu'il a déjà occupé ? S'il n'y retourne pas sûrement, avec quelle probabilité le fait-il ?

(3) Quelles sont les limites de $E(X_t)$ et de $V(X_t)$ quand t tend vers $+\infty$? Existe-t-il une distribution limite d'occupation de l'ensemble des états lorsque le temps tend vers l'infini ?

(4) À quelle condition la covariance du couple (X_s, X_t) qui nous renseigne sur la dépendance de X_s et X_t ne dépend-elle que de $(t-s)$? Y a-t-il décorrélation des états (X_s, X_t) et donc perte progressive par le système aléatoire de la mémoire de l'état occupé au temps s , lorsque t tend vers $+\infty$?

Définition 7.1 Un **processus aléatoire** est une famille de variables ou de vecteurs aléatoires $(X_t)_{t \in \mathbb{T}}$, indexée par l'ensemble \mathbb{T} des temps dénombrable ou continu, définie sur un espace de probabilité (Ω, \mathcal{F}, P) et à valeurs dans un espace d'états E . La variable ou le vecteur aléatoire X_t décrit l'état du processus au temps t . À toute occurrence ω on fait correspondre la **trajectoire** de la réalisation $(X_t(\omega))_{t \in \mathbb{T}}$ du processus définie par l'application :

$$t \in \mathbb{T} \rightarrow X_t(\omega) \in E.$$

7.2 Définition et caractérisation des chaînes de Markov discrètes

Les chaînes de Markov à temps entier occupent une place centrale dans la théorie des processus aléatoires, tant par le vaste domaine de leurs applications que par le développement de leur théorie.

Définition 7.2 Un processus aléatoire $(X_n)_{n \in \mathbb{N}}$ d'espace d'états $E = \{e_i\}_i$, fini ou dénombrable est une chaîne de Markov homogène discrète si sont vérifiés les deux axiomes suivants :

(1) **Axiome de Markov** : $\forall t_0, t_1, \dots, t_n \in \mathbb{N}$ tels que $t_0 < \dots < t_n$, et pour toute suite $\{e_0, \dots, e_n\}$ finie d'états quelconques de E telle que $P(\bigcap_{i=0}^{n-1} (X_{t_i} = e_i)) \neq 0$, on a :

$$P(X_{t_n} = e_n | \bigcap_{i=0}^{n-1} (X_{t_i} = e_i)) = P(X_{t_n} = e_n | X_{t_{n-1}} = e_{n-1}).$$

Interprétation : la mémoire du processus ne va pas au-delà du dernier état occupé, qui seul conditionne l'état suivant.

(2) **Axiome d'homogénéité** : $\forall s, t \in \mathbb{N}$ tels que $(s < t)$, $\forall u \in \mathbb{N}$ et pour tout couple d'états (e, e') :

$$P(X_{s+u} = e' | X_s = e) = P(X_{t+u} = e' | X_t = e) = P(X_u = e' | X_0 = e).$$

Interprétation : la probabilité de transition $P(X_t = e' | X_s = e)$ d'un état e à un état e' ne dépend que de la durée $(t-s)$ de la transition de e à e' et non du couple (s, t) .

Conséquence de l'axiome d'homogénéité : la **probabilité de transition** $P(X_{n+1} = e_j | X_n = e_i)$ de l'état e_i à l'état e_j en une unité de temps est donc indépendante du temps n , et sera notée p_{ij} .

Définition 7.3 Le *graphe de transition de la chaîne discrète* $(X_n)_n$ est le graphe orienté dont les sommets sont les états $(e_i)_i$ joints deux à deux par l'arc orienté $e_i \rightarrow e_j$, si et seulement si $p_{ij} > 0$. (voir : exemples des pages suivantes.)

Exemple 7.1 L'urne d'Ehrenfest

À la fin du XIX^e, le physicien Ludwig Boltzmann, fondateur de la mécanique statistique, énonça son fameux *théorème H* qui affirme que l'entropie d'un système isolé est une fonction croissante du temps. Or cette proposition qui implique l'irréversibilité de l'évolution du système contredit le principe de réversibilité des équations de la mécanique classique. Les thermodynamiciens critiquèrent durement l'ensemble des travaux de Boltzmann, lui reprochant même de pratiquer des *mathématiques douteuses*. Henri Poincaré porta une dernière estocade en démontrant, dans le cadre de ses travaux sur les systèmes dynamiques, qu'un système repasse toujours, si on attend suffisamment longtemps, dans un voisinage proche de son état initial. Afin de lever ce paradoxe, Paul et Tatiana Ehrenfest proposèrent en 1906 un modèle simplifié de diffusion gazeuse. Il s'agit d'un modèle de diffusion des molécules d'un gaz à travers une paroi poreuse séparant deux compartiments C_1 et C_2 de volumes et de formes identiques, formant un ensemble isolé. Soient la v.a. X_n égale au nombre de molécules contenues dans C_1 au temps n et N le nombre total constant de molécules contenues dans les deux compartiments. On suppose qu'à chaque instant n , une molécule est choisie au hasard parmi les N et change de compartiment.

Cette hypothèse se traduit par :

$$\forall (i, j) \in \{0, 1, 2, \dots, N\}^2 : P(X_{n+1} = j | X_n = i) = \begin{cases} \frac{i}{N} & \text{si } j = i - 1; \\ \frac{N-i}{N} & \text{si } j = i + 1; \\ 0 & \text{sinon.} \end{cases}$$

Le modèle d'urne d'Ehrenfest, qui définit le processus de même nom, vérifie donc les deux axiomes caractéristiques d'une chaîne de Markov :

- (1) L'état du processus au temps $n+1$ ne dépend que de son état au temps n .
- (2) $P(X_{n+1} = j | X_n = i)$ ne dépend pas du temps n .

On démontre que ce processus revisite une infinité de fois son état initial, et qu'il le fait avec des temps de retour d'autant plus grands que l'état initial correspond à un petit nombre de molécules. Si N est un multiple du nombre d'Avogadro égal à $6,02 \cdot 10^{23}$, et que toute transition se fait en une seconde, le temps moyen de retour excéderait largement l'âge de l'univers !

Définition 7.4 La *matrice (ou noyau) de transition* P de la chaîne discrète $(X_n)_n$ est la matrice carrée éventuellement de dimension infinie si E est infini, constituée par les probabilités de transition :

$$P = (p_{ij})_{i,j}.$$

Théorème 7.1 Toute matrice de transition $P = (p_{i,j})_{i,j}$ vérifie les propriétés :

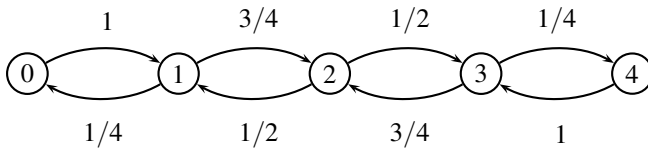
- (1) $\forall i, j \in I, 0 \leq p_{i,j} \leq 1$;
- (2) $\forall i \in I, \sum_{j \in I} p_{i,j} = 1$.

Preuve (2) résulte immédiatement de la formule de recomposition (voir chapitre précédent). □

Exemple 7.2 La matrice de transition de la chaîne d'Ehrenfest définie sur $E = \{0, 1, 2, 3, 4\}$ est égale à :

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Graphes de transition :



Le théorème suivant permet souvent de caractériser ou de simuler une chaîne de Markov (voir : exemple 7.4).

Théorème 7.2 Étant donné une suite $(Y_n)_{n \in \mathbb{N}^*}$ de v.a. indépendantes de même loi à valeurs dans un ensemble E_1 fini ou dénombrable, et la v.a. X_0 indépendante des Y_n à valeurs dans un ensemble E_2 fini ou dénombrable, le processus $(X_n)_{n \in \mathbb{N}}$ défini par :

$$\forall n \in \mathbb{N}, X_{n+1} = f(Y_{n+1}, X_n),$$

est une chaîne de Markov pour toute fonction f de $E_2 \times E_1$ dans E_1 .

Des applications de ce théorème sont données dans l'exercice 7.2 et l'exemple 7.4.

Définition 7.5 La loi initiale $\pi(0)$ d'une chaîne $(X_n)_n$ définie sur l'ensemble des états est la distribution de probabilités d'occupation des états en $t = 0$:

$$\pi(0) = (P(X_0 = e_1), P(X_0 = e_2), \dots, P(X_0 = e_n), \dots).$$

La loi d'occupation des états ou loi de la chaîne au temps k , notée $\pi(k)$, est la distribution définie par :

$$\pi(k) = (P(X_k = e_0), P(X_k = e_1), \dots, P(X_k = e_n), \dots).$$

Définition 7.6 Un état e_i est **absorbant** si $p_{ii} = 1$.

Théorème 7.3 Le temps d'occupation U_i d'un état e_i non absorbant est une v.a. géométrique :

$$\forall k \in \mathbb{N}, P(U_i = k) = p_{ii}^k (1 - p_{ii}).$$

Preuve $(U_i = k)$ décrit l'événement : la chaîne reste en l'état e_i aux temps inférieurs ou égaux à $(k-1)$, et le quitte au temps k . Ces événements de persistance dans l'état e_i étant indépendants entre eux, la probabilité de leur conjonction est égale au produit de leurs probabilités. \square

Théorème 7.4 Soit la chaîne $(X_n)_n$ à valeurs dans l'ensemble E et de matrice de transition P . Pour tout temps $k \in \mathbb{N}^*$ et quels que soient les états x_0, \dots, x_k occupés aux temps $0, 1, \dots, k$:

$$P(X_0 = x_0, X_1 = x_1, \dots, X_k = x_k) = P(X_0 = x_0) p_{x_0, x_1} p_{x_1, x_2} \dots p_{x_{k-1}, x_k}.$$

Preuve Par la formule des conditionnements successifs (Ch 6) on obtient : $P(X_n = x_k, \dots, X_0 = x_0) = P(X_k = x_k | X_{k-1} = x_{k-1}, \dots, X_0 = x_0) \cdot P(X_{k-1} = x_{k-1} | X_{k-2} = x_{k-2}, \dots, X_0 = x_0) \dots P(X_0 = x_0)$. Puis par application de l'axiome de Markov :

$$P(X_k = x_k, \dots, X_0 = x_0) = P(X_k = x_k | X_{k-1} = x_{k-1}) \cdot P(X_{k-1} = x_{k-1} | X_{k-2} = x_{k-2}) \dots P(X_0 = x_0). \quad \square$$

On détermine la probabilité de transition d'un état e_i à un état e_j entre deux instants quelconques m et p grâce à l'équation de Chapman-Kolmogorov.

Théorème 7.5 (Équation de Chapman-Kolmogorov) Soient une chaîne $(X_n)_n$ et les instants m, n , et p vérifiant $m < n < p$.

Pour tout couple d'états e_i, e_j de E , on a :

$$P(X_p = e_j | X_m = e_i) = \sum_{e_k \in E} P(X_p = e_j | X_n = e_k) P(X_n = e_k | X_m = e_i).$$

Preuve On multiplie chacun des deux membres de l'égalité par la probabilité non nulle $P(X_m = e_i)$.

$$\begin{aligned}
 P(X_p = e_j \text{ et } X_m = e_i) &= \sum_{e_k} P(X_p = e_j | X_n = e_k) P(X_n = e_k \text{ et } X_m = e_i) \\
 &= \sum_{e_k} P(X_p = e_j | X_n = e_k, X_m = e_i) \\
 &= \sum_{e_k} P(X_p = e_j \text{ et } X_n = e_k \text{ et } X_m = e_i) \\
 &= P(X_p = e_j \text{ et } X_m = e_i).
 \end{aligned}$$

□

Le théorème fondamental suivant montre que l'évolution aléatoire d'une chaîne est entièrement déterminée par la connaissance de sa loi initiale $\pi(0)$ et de sa matrice de transition P .

Théorème 7.6 (Loi d'évolution d'une chaîne)

La matrice de transition $P^{(k)}$ associée à l'ensemble des probabilités de transition en k unités de temps, constituée des termes $(P(X_k = e_j | X_0 = e_i))_{i,j}$, est définie par :

$$P^{(k)} = \underbrace{P \times P \times \dots \times P}_{k \text{ fois}} = P^k,$$

d'où la loi d'évolution de la chaîne :

$$\forall k, \quad \pi(k) = \pi(0) P^k.$$

Exemple 7.3 Modèle d'Ehrenfest (suite)

Supposons encore que $\pi(0) = (0, 1, 0, 0, 0)$ et calculons la loi $\pi(2)$ de X_2 :

$$\pi(2) = \pi(0).P^2 \text{ avec } P^2 = \begin{pmatrix} \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & \frac{5}{8} & 0 & \frac{3}{8} & 0 \\ \frac{1}{8} & 0 & \frac{3}{4} & 0 & \frac{1}{8} \\ 0 & \frac{3}{8} & 0 & \frac{5}{8} & 0 \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} \end{pmatrix}$$

donc : $\pi(2) = (0, \frac{5}{8}, 0, \frac{3}{8}, 0)$.

7.1 ♠ Étude de la chaîne à deux états

Soit $P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$; supposons que $\pi(0) = (1, 0)$.

Déterminer la loi $\pi(n)$. [Indication : on diagonalisera P en sachant que 1 est toujours une valeur propre d'une matrice de transition.]

[Réponse : $\pi(n) = (\frac{\beta + \alpha(1-\alpha-\beta)^n}{\alpha + \beta}, \frac{\alpha - \alpha(1-\alpha-\beta)^n}{\alpha + \beta}).$]

Interpréter $\pi(n)$ dans le cas où : $\alpha + \beta = 1$, $\alpha + \beta < 1$ et $\alpha = \beta = 1$.

7.2 ♠♠ Soit X_n la v.a. égale à la valeur maximale obtenue après n jets d'un dé.

(1) Utiliser le théorème 7.2 pour démontrer que $(X_n)_n$ est une chaîne de Markov.

(2) Déterminer la matrice de transition.

(3) Démontrer que la limite de $\pi(n)$ quand n tend vers l'infini est égale à $(0,0,0,0,0,1)$.

Exemple 7.4 La marche aléatoire symétrique dans \mathbb{Z} .

On considère la suite $(X_i)_{i \in \mathbb{N}^*}$ de v.a. de Bernoulli, indépendantes, de loi commune : $P(X = 1) = P(X = -1) = \frac{1}{2}$. La marche aléatoire dans l'espace d'états $E = \mathbb{Z}$ est définie par la suite $(S_n)_{n \in \mathbb{N}^*}$ où $S_n = \sum_{i=1}^n X_i$ modélise l'abscisse, au temps n , d'une particule qui était à l'origine au temps $t = 0$, et qui effectue à chaque instant i , un pas de longueur unité dans un sens ou dans l'autre avec la probabilité $\frac{1}{2}$.

Démontrons que $(S_n)_n$ est une chaîne :

$S_{n+1} = S_n + X_{n+1}$, et donc d'après le théorème 7.2 $(S_n)_n$ est une chaîne de Markov définie sur \mathbb{Z} . La matrice de transition de la chaîne est une matrice de dimension infinie, tridiagonale, constituée d'une diagonale formée de 0, bordée par une sous-diagonale et une sur-diagonale constituées de $\frac{1}{2}$.

7.3 Classification des états

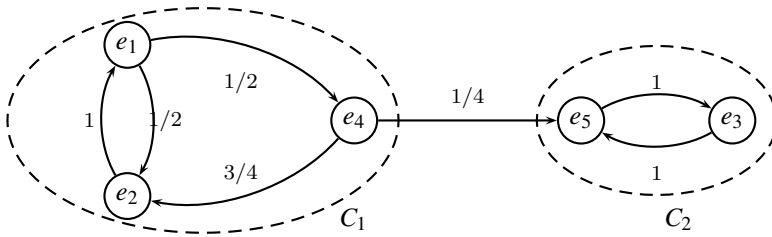
7.3.1 Relation de communication entre états

Définition 7.7 Deux états distincts x et y communiquent s'il existe un instant $n_1 \in \mathbb{N}$ tel que $p_{x,y}^{(n_1)} > 0$ et s'il existe un instant $n_2 \in \mathbb{N}$ tel que $p_{y,x}^{(n_2)} > 0$.

Interprétation : partant de x il est toujours possible d'atteindre y en un temps fini et inversement partant de y on peut atteindre x en un temps fini, non nécessairement par la même suite d'états intermédiaires.

Théorème 7.7 La relation de communication entre les états d'une chaîne est une relation d'équivalence qui induit sur E , une partition en classes de communication à l'intérieur desquelles tous les états communiquent entre eux.

Exemple 7.5 La chaîne de Markov associée au graphe suivant possède deux classes de communication $C_1 = \{e_1, e_2, e_4\}$ et $C_2 = \{e_3, e_5\}$; on remarque que si l'on quitte la première, on n'y retourne plus.



Définition 7.8 Une chaîne est **irréductible** si tous ses états communiquent entre eux; elle est donc constituée d'une seule classe.

7.3.2 États récurrents et transitoires

Définition 7.9 Si au temps 0, la chaîne est en l'état e_i , la variable aléatoire $T_i = \text{Min}\{n \geq 1; X_n = e_i\}$ définit le **temps de premier retour en e_i** .

Définition 7.10 Un état e_i est **récurrent** si partant de e_i on y revient sûrement, ce qui s'exprime par :

$$P((\text{il existe } n \geq 1 \text{ tel que } X_n = e_i) | X_0 = e_i) = 1,$$

soit encore :

$$P(T_i < +\infty | X_0 = e_i) = 1.$$

Un état récurrent est donc visité un nombre infini de fois.

Exemple 7.6 Seuls les états e_3 et e_5 de la chaîne de l'exemple 7.5 sont récurrents.

Exemple 7.7 Considérons la chaîne à 1 000 états, dans laquelle chaque transition s'effectue en une seconde :



La chaîne est irréductible, on peut donc passer d'un état à n'importe quel autre, bien que, par exemple, le passage de l'état e_{1000} à l'état e_1 soit pratiquement improbable et nécessite visiblement un temps moyen fini mais proprement astronomique. La chaîne étant finie et irréductible, on montrera que tous les états sont récurrents et de temps moyens de retour finis.

Définition 7.11 Un état e_i est dit **transitoire** s'il n'est pas récurrent :

$$P(T_i < +\infty | X_0 = e_i) < 1.$$

Un état est transitoire s'il n'est visité qu'un nombre fini de fois.

Par définition, une chaîne irréductible ne contient aucun état transitoire ou absorbant.

Définition 7.12 Une chaîne dont tous les états sont récurrents est dite **récurrente**; une chaîne dont tous les états sont transitoires est dite **transiente**.

Définition 7.13 La probabilité de première transition en n unités de temps, de l'état e_i à l'état e_j , notée $f_{i,j}^{(n)}$, est définie par :

$$f_{i,j}^{(n)} = P(T_j = n \mid X_0 = e_i) = P(X_n = e_j, X_{n-1} \neq e_j, \dots, X_1 \neq e_j \mid X_0 = e_i).$$

$f_{i,i}^{(n)}$ est la probabilité de premier retour à l'état e_i en n unités de temps.

Théorème 7.8 e_i est récurrent $\iff \sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$.

e_i est transitoire $\iff \sum_{n=1}^{\infty} f_{ii}^{(n)} < 1$.

Preuve Évidente par définition de $f_{ii}^{(n)}$. □

Définition 7.14 On appelle **temps moyen de retour en e_j** :

$$\mu_j = E[T_j \mid X_0 = e_j] = \sum_n n \cdot p_{jj}^{(n)}.$$

Il existe deux sortes d'états récurrents : les états récurrents positifs et les états récurrents nuls.

Définition 7.15 Un état e_j est **récurrent positif (ou non nul)** s'il est récurrent et si le temps moyen μ_j de retour est fini. Dans le cas où le temps moyen de retour est infini, l'état est dit **récurrent nul**.

Définition 7.16 La v.a. $N_j = \sum_{n=1}^{\infty} \mathbf{1}_{e_j}(X_n)$ dénombre les passages de la chaîne par l'état e_j ($\mathbf{1}_{e_j}$, fonction indicatrice de l'état e_j , vaut 1 si $X_n = e_j$ et 0 si $X_n \neq e_j$).

Théorème 7.9 Le nombre moyen de retours à l'état e_j , défini par l'espérance conditionnelle $E(N_j \mid X_0 = e_j)$, est égal à :

$$E(N_j \mid X_0 = e_j) = \sum_{n \geq 1} p_{jj}^{(n)}.$$

Preuve $E(\mathbf{1}_{e_j}(X_n) | X_0 = e_j) = P(X_n = e_j | X_0 = e_j) = p_{jj}^{(n)}$. D'où le résultat par sommation. \square

Théorème 7.10 (*Caractérisation des états*)

$$(e_i \text{ transitoire}) \iff \left(\sum_{n \geq 1} p_{ii}^{(n)} < +\infty \right) ;$$

$$(e_i \text{ récurrent nul}) \iff \left(\sum_{n \geq 1} p_{ii}^{(n)} = +\infty \text{ et } \lim_{n \rightarrow +\infty} p_{ii}^{(n)} = 0 \right) ;$$

$$(e_i \text{ récurrent positif}) \iff \left(\sum_{n \geq 1} p_{ii}^{(n)} = +\infty \text{ et } \lim_{n \rightarrow +\infty} p_{ii}^{(n)} > 0 \right) ;$$

Attention ! Les états récurrents nuls ne peuvent exister que dans les chaînes infinies.

Les caractérisations précédentes de la récurrence et de la transience sont surtout utiles à l'étude des chaînes infinies. En effet, dans le cas des chaînes finies, l'interprétation du graphe de transition structuré de façon à faire apparaître distinctement les diverses classes d'équivalence, permet d'identifier facilement les classes récurrentes et transitoires.

Théorème 7.11 *Les propriétés de récurrence et de transience sont des propriétés de classe : une classe de communication est exclusivement récurrente positive, récurrente nulle ou transitoire. Seules les classes récurrentes constituent des sous-chaînes de la chaîne ; cette propriété résulte de la propriété de fermeture d'une classe récurrente, d'où l'on ne peut sortir une fois que la chaîne l'a atteinte.*

Pratiquement, on structurera le graphe d'une chaîne de façon à faire apparaître ses différentes classes de communication, récurrentes ou transitoires : l'interprétation des relations de communication entre les classes fournit des informations pertinentes sur son évolution. On réordonne les états pour obtenir la **matrice canonique de transition** de la forme :

$$\begin{pmatrix} P_T & P_{TR} \\ 0 & P_R \end{pmatrix}$$

où :

P_T est la matrice de transition associée aux classes transientes ;

P_{TR} est la matrice de transition des états transitoires vers les états récurrents ;

P_R est formée des matrices de transition associées aux classes récurrentes.

Exemple 7.8 Soit la chaîne $(X_n)_n$ de matrice de transition P :

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}
e_1		0,7							0,3	
e_2	0,8		0,2							
e_3									1	
e_4				0,4		0,25		0,25	0,05	0,05
e_5						0,2	0,8			
e_6					0,1	0,2	0,7			
e_7					0,7	0,3				
e_8				0,35		0,15	0,1			0,4
e_9	1									
e_{10}				0,5				0,05		0,45

Une fois tracé le graphe de cette chaîne, il est visuellement aisé d'en déterminer les diverses classes :

- une classe transiente $T = \{e_4, e_8, e_{10}\}$;
- une classe récurrente positive 2-périodique : $R_1 = \{e_1, e_2, e_3, e_9\}$;
- une classe récurrente positive aperiodique : $R_2 = \{e_5, e_6, e_7\}$.

En réarrangeant les états selon leur classe d'appartenance, on obtient la matrice canonique qui met en valeur les classes et leurs relations :

	e_4	e_8	e_{10}	e_1	e_2	e_3	e_9	e_5	e_6	e_7
e_4	0,4	0,25	0,05	0	0	0	0,05	0	0,25	0
e_8	0,35	0	0,4	0	0	0	0	0	0,15	0,1
e_{10}	0,5	0,05	0,45	0	0	0	0	0	0	0
e_1	0	0	0	0	0,7	0	0,3	0	0	0
e_2	0	0	0	0,8	0	0,2	0	0	0	0
e_3	0	0	0	0	0	0	1	0	0	0
e_9	0	0	0	1	0	0	0	0	0	0
e_5	0	0	0	0	0	0	0	0	0,2	0,8
e_6	0	0	0	0	0	0	0	0,1	0,2	0,7
e_7	0	0	0	0	0	0	0	0,7	0,3	0

Théorème 7.12 (Propriétés des chaînes finies)

- (1) *Aucun état n'est récurrent nul.*
- (2) *Aucune chaîne finie n'est transiente ; en revanche, certains de leurs états peuvent être transitoires.*
- (3) *Il existe au moins un état récurrent positif.*
- (4) *Si en outre la chaîne est irréductible, elle est récurrente positive.*

Preuve (1) Supposons qu'il existe un état récurrent nul e_j , et C_j sa classe de communication, alors $\forall n \sum_{k \in C_j} p_{j,k}^{(n)} = 1$ (*)

Mais $\lim_{n \rightarrow +\infty} p_{j,k}^{(n)} = 0$, qui est en contradiction avec (*).

(2) Supposons tous les états transitoires, on montre alors que $\forall e_i, e_j$ transitoires, la série $\sum p_{i,j}^{(n)}$ converge, donc $\lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = 0$ (**).

Or $\sum_{j=1}^{Card(E)} p_{i,j}^{(n)} = 1$ ($\forall n \in \mathbb{N}$), donc, par passage à la limite quand $n \rightarrow +\infty$, on arrive à une contradiction de (**).

(3) est conséquence de (1) et (2).

(4) La chaîne ayant au moins un état récurrent d'après (c) et une seule classe est récurrente positive. \square

Exemples de chaînes infinies : les marches aléatoires isotropes

Nous avons défini dans l'exemple 7.4 la marche aléatoire symétrique sur \mathbb{Z} et montré qu'elle était une chaîne de Markov sur l'espace d'états \mathbb{Z} . Dans l'exercice suivant, il s'agit d'en étendre l'étude aux chaînes définies sur les espaces d'états \mathbb{Z}^2 et \mathbb{Z}^3 , et de démontrer les résultats suivants :

- les marches aléatoires isotropes définies sur \mathbb{Z} et \mathbb{Z}^2 sont récurrentes nulles ;
- la marche aléatoire isotrope définie sur \mathbb{Z}^3 est transiente ; donc, partant d'un nœud quelconque du réseau, on n'y retourne qu'un nombre fini de fois.

7.3 ♠♠ Étude des marches aléatoires sur \mathbb{Z} , \mathbb{Z}^2 , \mathbb{Z}^3

(1) Marche aléatoire sur \mathbb{Z} .

On a vu dans l'exemple 7.4 que la marche aléatoire est une chaîne de Markov ; par sa structure même, elle est irréductible. Il suffit donc de caractériser d'après le théorème 7.13 l'un de ses états (par exemple l'origine). On constate d'abord que $\forall m \in \mathbb{N}^*, p_{0,0}^{2m+1} = 0$. On associe à tout chemin de longueur $2m$, partant de l'origine et y revenant, un $2m$ -uplet : démontrer que $p_{0,0}^{(2m)} = \frac{C_{2m}^m}{2^{2m}}$ ainsi que la divergence de la série $\sum_{m \leq 1} p_{0,0}^{(2m)}$ grâce à la formule de Stirling. Conclure.

(2) Marche aléatoire dans \mathbb{Z}^2 .

Une particule part au temps $t = 0$, de l'origine d'un réseau carré ; à chaque unité de temps elle peut faire un saut vers le nord, le sud, l'est ou l'ouest avec une probabilité $\frac{1}{4}$.

Construire la chaîne associée au processus.

[Réponse : $p_{00}^{(2n)} = \sum_{m=0}^n \frac{C_{2n}^m (C_n^m)^2}{4^{2n}}$ asymptotiquement proche de $\frac{1}{\pi n}$ lorsque n est grand.]

Démontrer que la chaîne est récurrente nulle.

(3) Marche aléatoire dans \mathbb{Z}^3 .

On étend le modèle précédent à trois dimensions ; la particule a cette fois six possibilités de déplacements équiprobables.

Démontrer que la chaîne est transiente.

[Réponse : $p_{00}^{(2n)} = \frac{C_{2n}^n}{2^{2n}} \sum_{n_1+n_2+n_3=n} \left(\frac{n!}{n_1!n_2!n_3!3^n} \right)^2$ asymptotiquement proche, quand n tend vers $+\infty$, de $\frac{3\sqrt{3}}{2\pi^{\frac{3}{2}}n^{\frac{3}{2}}}$.]

Le mathématicien et pédagogue hongrois G. Polya (1887-1985) a montré que la probabilité de retour à un état initial quelconque de la marche aléatoire dans \mathbb{Z}^3 était très proche de 0,340. Une marche aléatoire dans \mathbb{Z}^3 ne repasse donc pas par un point déjà visité, approximativement dans deux tiers des cas.

Le paradoxe du jeu de pile ou face

Lorsque deux joueurs A et B s'affrontent au jeu de pile ou face, l'intuition dont on apprendra à se méfier dans ce cours, conduit à penser qu'il y a sur un temps long, alternance de chacun des joueurs dans la situation gagnante ; or il n'en est rien. En effet, on établit qu'étant donné l'intervalle de temps $[0, 2n]$ d'étendue suffisamment grande, l'un des deux joueurs sera en position gagnante avec une probabilité supérieure à $\frac{1}{2}$.

On établit que, sur $2n$ lancers alternatifs, la probabilité p_{2k} pour que l'un des joueurs, par exemple A, mène $2k$ fois est égale à :

$$p_{2k} = \binom{2k}{k} \binom{2(n-k)}{n-k} / 2^{2n}.$$

Si n est très grand, p_{2k} tend vers :

$$\frac{1}{\pi(k(n-k))^{\frac{1}{2}}}.$$

(Utiliser la formule de Stirling pour établir ce résultat).

Soit T_{2n} la v.a. égale au temps qui sépare le temps de dernière visite de l'état 0 du temps $2n$:

$$\begin{aligned} \forall x \in [0, 1] \quad P(T_{2n} \leq 2xn) &\simeq \sum_{k \leq xn} \frac{1}{\pi(k(n-k))^{\frac{1}{2}}} \\ &\simeq \int_0^{xn} \frac{1}{\pi(u(n-u))^{\frac{1}{2}}} du = \frac{2}{\pi} \cdot \text{Arcsin}(x^{\frac{1}{2}}). \end{aligned}$$

On conclut que si n est très grand, la proportion $\frac{T_{2n}}{2n}$ du temps où A gagne dans l'intervalle $[0, 2n]$ est de densité $\frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}} \mathbf{1}_{[0, 1]}(x)$, qui définit la loi arcsinus.

Au cours d'une partie, les joueurs vont-ils alternativement mener au score selon une proportion égale à $\frac{1}{2}$, ou au contraire le hasard va-t-il permettre une domination quasi constante de l'un d'eux ? On peut vérifier par exemple que l'un des joueurs sera gagnant plus des trois quarts du temps avec une probabilité supérieure à $\frac{2}{3}$, ou qu'il a plus de 50 % de chances pour qu'il gagne pendant près de 90 % du temps. L'apparente égalité des chances à laquelle l'intuition commune nous convie dans un jeu équilibré est donc infirmée.

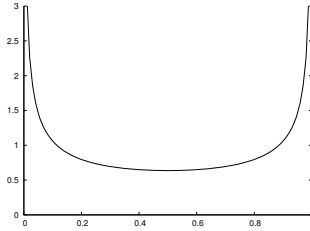


Figure 7.3 – Graphe de la densité arcsinus $\frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}$.

7.4 ♠♠ Soit la chaîne de Markov d'espace d'états \mathbb{N} , de matrice de transition définie par :

$$p_{i,0} = p_i, \quad p_{i,i+1} = 1 - p_i; \\ \forall i \in \mathbb{N} \quad 0 < p_i < 1.$$

- (1) Dessiner le graphe de la chaîne ; en déduire qu'elle est irréductible.
- (2) À quelle condition est-elle récurrente ? [Rappel : si pour tout i , $0 < x_i < 1$, alors le produit infini $\prod_{i=1}^{+\infty} (1 - x_i)$ est nul ssi $\sum_{i=0}^{+\infty} -\ln(1 - x_i)$ diverge.]
- (3) Si $\forall i$, $p_i = p$, montrer que la chaîne est récurrente positive.

7.3.3 États périodiques

Définition 7.17 Un état e_i est **périodique de période d_i** si les temps de retour à e_i sont d_i -périodiques, où d_i est le PGCD de l'ensemble $\{n \in \mathbb{N}^*, \text{ tels que } p_{i,i}^{(n)} > 0\}$. Si $d_i = 1$, l'état e_i est apériodique.

Un état e_i pourvu d'un cycle ($p_{i,i} > 0$) est nécessairement apériodique, puisqu'on peut y retourner en tout temps entier.

Par exemple, la chaîne d'Ehrenfest est périodique de période 2.

Théorème 7.13 La périodicité est une propriété de classe : les états d'une classe sont ou bien tous périodiques de même période ou bien tous apériodiques. Une classe qui possède au moins un cycle est nécessairement apériodique.

Preuve Soit d_i et d_j les périodes de e_i et e_j , et soit $p_{ij}^{(m)} p_{ji}^{(n)} > 0 \Rightarrow p_{ii}^{(n+m)} \geq p_{ij}^{(m)} p_{ji}^{(n)} > 0$ donc d_i divise $(m+n)$. De plus $p_{ii}^{(m+k+n)} \geq p_{ij}^{(m)} p_{jj}^{(k)} p_{ji}^{(n)}$ et donc $p_{jj}^{(k)} > 0 \Rightarrow p_{ii}^{(m+k+n)} > 0$, ce qui implique que si d_j divise k , alors d_j divise $(m+k+n)$; comme d_i divise $(m+n) : (d_j \text{ divise } k) \Rightarrow (d_i \text{ divise } k)$ et par symétrie de i et j , $d_i = d_j$. \square

Exemple 7.9 (1) La chaîne d'Ehrenfest est récurrente positive et 2-périodique.

(2) La chaîne à deux états (voir exemple 7.1) est récurrente positive et apériodique si $\alpha, \beta > 0$.

(3) Dans l'exemple 7.5, la classe C_1 est transiente, apériodique; la classe C_2 est récurrente positive et 2-périodique.

7.4 Distributions stationnaires et distributions limites

Définition 7.18 Une distribution $\pi = (\pi_1, \pi_2, \dots, \pi_i, \dots)$ sur l'ensemble des états est une distribution **stationnaire ou invariante** de la chaîne $(X_n)_n$ de matrice de transition P , si :

$$\pi = \pi.P \quad (*)$$

où $\pi = (\pi_1, \pi_2, \dots, \pi_i, \dots)$ est de somme 1.

$(*)$ est équivalente au système linéaire :

$$\left(\pi_i = \sum_{j \in E} \pi_j p_{j,i} \right)_{i=1,2,\dots}$$

Interprétation : π_i désigne la proportion de temps passé par la chaîne dans l'état e_i lorsque le temps d'évolution de la chaîne devient infiniment long.

Théorème 7.14 Étant donné une chaîne admettant une distribution stationnaire π :

$$\text{si } \pi(0) = \pi \text{ alors } \forall k, \pi(k) = \pi.$$

La loi d'occupation $\pi(k)$ des états au temps k est invariante pour tout k .

Preuve $\pi(k) = \pi(0)P^k$, donc si $\pi(0) = \pi$, $\pi(1) = \pi P = \pi$ et par récurrence, $\pi(k) = \pi$. \square

Remarque 7.1 Une chaîne peut ne pas avoir de distribution stationnaire, en avoir une seule ou plusieurs portées par la ou les classes récurrentes.

Théorème 7.15 *Toute chaîne finie (nécessairement non transiente) admet au moins une distribution stationnaire, qui est unique si la chaîne est irréductible.*

Preuve Rappelons le lemme de Perron-Froebenius : toute matrice stochastique, vérifiant donc $p_{i,j} \geq 0$ et $\sum_j p_{i,j} = 1$ finie, admet un vecteur propre de composantes positives associé à la valeur propre 1. La matrice $P - \mathbf{1}_d$ a des colonnes de sommes nulles, donc 1 est une valeur propre de P ; d'où la conclusion par le lemme. \square

Théorème 7.16 *Si π est une distribution stationnaire, alors pour tout état récurrent e_i :*

$$\pi_i = \frac{1}{\mu_i},$$

où μ_i est le temps moyen de retour à e_i .

Théorème 7.17 *Une chaîne transiente infinie n'admet pas de distribution stationnaire.*

Preuve Supposons qu'il existe une distribution stationnaire π : alors $\pi P = \pi$ et $\pi P^n = \pi$. Pour tout état j et pour tout entier n , $\pi_j = \sum_i \pi_i p_{i,j}^{(n)}$; les états étant transitoires, tous les $p_{i,j}^{(n)}$ tendent vers 0. $\forall j, \pi_j = \lim_{n \rightarrow \infty} \sum_i \pi_i p_{i,j}^{(n)} = \sum_i \lim_{n \rightarrow \infty} \pi_i p_{i,j}^{(n)} = 0$; ce qui est contradictoire avec $\sum_j \pi_j = 1$. \square

Théorème 7.18 (Conditions nécessaires et suffisantes d'existence de distributions stationnaires)

Cas 1 : si la chaîne est irréductible, elle possède une distribution stationnaire unique $\pi = (\frac{1}{\mu_i})_{e_i \in E}$, si et seulement si tous les états sont récurrents positifs. (μ_i est le temps moyen de retour à e_i).

Cas 2 : si la chaîne est quelconque, elle possède une distribution stationnaire π , si et seulement si la chaîne possède au moins une classe récurrente positive. Dans le cas de plusieurs classes récurrentes positives, il existe une distribution stationnaire pour chaque classe.

Exemple 7.10 Marche aléatoire sur \mathbb{N} réfléchi en 0.

La matrice de transition de cette chaîne est définie par :

$$\forall k \in \mathbb{N}^*, p_{k,k+1} = p \text{ et } p_{k,k-1} = 1 - p ; p_{0,1} = 1.$$

La chaîne est infinie, irréductible, de période 2 ; la distribution stationnaire π existe si et seulement si elle est récurrente positive. La résolution de l'équation $\pi P = \pi$ aboutit à $\pi_k = \pi_0 \frac{1}{q} \left(\frac{p}{q}\right)^{k-1}, \forall k \geq 1$; $\sum_k \pi_k$ converge si et seulement si $p < q$,

soit $p < \frac{1}{2}$, condition de récurrence positive. Dans ce cas, $\sum_k \pi_k = \pi_0(1 + \frac{1}{q} \frac{1}{1-p}) = \pi_0 \frac{2q}{q-p}$, d'où : $\pi_0 = \frac{1-2p}{2(1-p)}$ et $\pi_k = \frac{\pi_0}{p} (\frac{p}{1-p})^k$.

Le temps moyen de retour à l'origine est donc égale à : $\mu_0 = \frac{1}{\pi_0} = \frac{2(1-p)}{1-2p}$.

Par exemple, si $p = 0,49$ alors $\mu_0 = 51$; on pourrait démontrer que pour $p = \frac{1}{2}$ la chaîne est récurrente nulle, et que pour $p > \frac{1}{2}$ elle est transiente.

7.5 ♠♠♠ Le nombre de clients qui arrivent à une station-service pendant la durée d'un service de durée égale à l'unité quel que soit le client est une v.a. de Poisson d'intensité λ . Soit X_n le nombre de voitures présentes dans la station à l'instant n , et Y_n le nombre de voitures arrivant entre n et $n+1$.

(1) Démontrer que $(X_n)_n$ est une chaîne de Markov.

(2) Soit $\pi = (\pi_k)_k$ la distribution stationnaire de fonction génératrice G . Écrire les équations vérifiées par les π_k et en déduire G en fonction de la fonction génératrice de la loi de Poisson $\mathcal{P}(\lambda)$. Montrer que $\pi_0 = 1 - \lambda$. Quelle est la condition d'existence de π et le nombre moyen de voitures dans la station en régime stationnaire ?

7.6 ♠♠ Deux machines identiques fonctionnent indépendamment ; chacune pouvant tomber en panne avec la probabilité q . Soit X_n le nombre de machines en panne au début de la n -ième journée.

(1) Si une machine tombe en panne, elle est réparée la nuit suivante et on ne peut réparer qu'une machine par nuit. Décrire la chaîne de Markov $(X_n)_n$. Déterminer la distribution stationnaire.

(2) Même question en supposant qu'une machine en panne n'est réparée que le lendemain et qu'on ne peut réparer qu'une machine dans la journée.

7.7 ♠♠ Chaîne produit

Soient les chaînes de Markov indépendantes $(X_n^{(1)})_n$ et $(X_n^{(2)})_n$ de matrices de transition P_1 et P_2 , d'espaces d'états finis E_1 et E_2 . [$\forall m, n$ distincts, $X_n^{(1)}$ est indépendante de $X_m^{(2)}$]

Démontrer que le processus $X_n = (X_n^{(1)}, X_n^{(2)})$ est une chaîne, dite chaîne produit, et calculer sa matrice de transition. La chaîne produit n'est pas irréductible même si les chaînes $(X_n^{(1)})_n$ et $(X_n^{(2)})_n$ le sont ; par contre on démontre qu'elle est irréductible si et seulement si elle est apériodique.

Définition 7.19 Une chaîne admet une **distribution limite** ou **asymptotique** π^* , si quelle que soit la loi initiale $\pi(0)$:

$$\lim_{n \rightarrow +\infty} \pi(n) = \pi^*.$$

Plus explicitement, pour tout couple d'états (i, j) , $\lim_{n \rightarrow \infty} p_{i,j}^{(n)} = \text{notée } \pi_j^*$ est indépendant de i .

Théorème 7.19 *Toute distribution limite π^* d'une chaîne, si elle existe, est l'unique distribution stationnaire, nécessairement portée par les états récurrents.*

Preuve $\lim_{n \rightarrow +\infty} \pi^{(n)} = \pi^*$ donc $\lim_{n \rightarrow +\infty} \pi^{(n+1)} = \pi^*$; mais $\pi^{(n+1)} = \pi^{(n)}P$, donc par passage à la limite des membres de l'équation, on obtient $\pi^* = \pi^*P$: donc π^* est une distribution stationnaire. Son unicité se prouve par une démonstration classique par l'absurde. \square

Définition 7.20 *Un état récurrent positif et apériodique est dit **ergodique**. Une chaîne irréductible, apériodique et récurrente positive est dite ergodique.*

Théorème 7.20 *Les états d'une classe de communication qui contient au moins un état ergodique sont ergodiques.*

Preuve On sait que les propriétés de récurrence et d'apériodicité sont des propriétés de classe : l'ergodicité l'est donc aussi. Il suffit donc qu'un état de la classe soit ergodique pour que la classe qui le contient le soit. \square

Théorème 7.21 (de Chacon-Ornstein) *Toute chaîne ergodique possède une distribution limite unique $\pi^* = (\pi_i^*)_i$ qui s'identifie à l'unique distribution stationnaire $\pi = (\pi_i)_i$ de la chaîne;*

$$\forall i, j \in E, \quad \lim_{(n \rightarrow +\infty)} p_{j,i}^{(n)} = \pi_i^* = \pi_i = \frac{1}{\mu_i} \text{ où } \pi = \pi P.$$

Exemple 7.11 La chaîne d'Ehrenfest vérifie les hypothèses du théorème, et admet une distribution limite. Considérons le cas étudié dans ce chapitre, où $E = \{0, 1, 2, 3, 4\}$; il est facile de vérifier que la distribution stationnaire limite est $\pi^* = (\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$.

Si la chaîne possède n états, on montre sans difficulté que $\pi^* = (\frac{1}{2^n}, \frac{C_n^1}{2^n}, \frac{C_n^2}{2^n}, \dots, \frac{1}{2^n})$ qui est la distribution binomiale $\mathcal{B}(n; \frac{1}{2})$. Soit X^* la v.a. associée à π^* ; la loi $\mathcal{B}(n; \frac{1}{2})$ est approchée de façon satisfaisante, n étant très grand, par la loi $\mathcal{N}(\frac{n}{2}; \sigma^2 = \frac{n}{4})$. Recherchons la constante c vérifiant :

$$P(\frac{n}{2} - c < X^* < \frac{n}{2} + c) = 99,999 \text{ \%}.$$

$$\text{D'où : } P(-\frac{c}{\sigma} < \frac{X^* - \frac{n}{2}}{\sigma} < \frac{c}{\sigma}) = 99,999 \text{ \%}.$$

La consultation d'une table gaussienne donne $c \simeq 4,8 \cdot \frac{\sqrt{n}}{2}$ et une variation relative égale à $\frac{4,8 \cdot \frac{\sqrt{n}}{2}}{\frac{n}{2}} = \frac{4,8}{\sqrt{n}}$, qui est extrêmement faible lorsque n est grand, ce

qui est le cas dans les situations concrètes où n est un multiple du nombre d'Avogadro $6,02.10^{23}$. Ce résultat corrobore le constat physique d'équilibre des pressions au sein des deux compartiments.

7.8 ♠♠ Modèle markovien de gestion de stocks

Un stock peut contenir au maximum s pièces. La v.a. X_n est le nombre de pièces au début de la n -ième semaine; la v.a. Y_n est le nombre de pièces sorties du stock au cours de la n -ième semaine. On supposera que $X_1 = s$ et que pour tout $j = 0, 1, \dots, i$ $P(Y_n = j | X_n = i) = \frac{1}{1+i}$, quel que soit $n \in \mathbb{N}^*$. Au début de chaque semaine, on complète le stock, mais le délai de livraison est une semaine.

(1) Montrer que $(X_n)_n$ est une chaîne de Markov, et qu'elle admet une distribution stationnaire définie par $\pi_i = \frac{2(i+1)}{(s+1)(s+2)}$.

(2) Quel est le nombre moyen de pièces disponibles ?

Théorème 7.22 *Si $(X_n)_n$ est une chaîne irréductible, apériodique, de distribution stationnaire π , et f une fonction réelle définie sur l'ensemble des états E , alors $E(f(X))$ définie par $\sum_j \pi_j f(e_j)$ est égal à :*

$$E(f(X)) =_{p.s.} \lim_{n \rightarrow +\infty} \frac{1}{n+1} \sum_{i=0}^n f(X_i),$$

où X est la v.a. associée à la distribution stationnaire.

Théorème 7.23 *Soit une chaîne irréductible, récurrente positive, d -périodique et de distribution stationnaire π . Pour tout couple d'états (i, j) :*

$$\begin{aligned} \lim_{n \rightarrow \infty} p_{i,j}^{n,d+k} &= d \cdot \pi_j^* \quad \text{si } k = \min \left\{ q \mid p_{ij}^{(q)} > 0 \right\} ; \\ &= 0 \text{ sinon.} \end{aligned}$$

Résumé des propriétés des chaînes

- (1) Aucune chaîne finie n'est transiente.
- (2) Aucun état d'une chaîne finie n'est récurrent nul.
- (3) Toute chaîne finie possède au moins une classe récurrente positive et donc au moins une distribution stationnaire portée par cette classe.
- (4) Si la distribution limite d'une chaîne quelconque existe, elle est confondue avec son unique distribution stationnaire et est portée par les seuls états récurrents.
- (5) Toute chaîne ergodique, par définition récurrente positive et apériodique, possède une distribution limite égale à la distribution stationnaire.
- (6) Une chaîne infinie transiente ne possède pas de distribution stationnaire.

7.5 Compléments

7.5.1 Estimation de la matrice de transition P et de la loi stationnaire π d'une chaîne finie irréductible

Étant donné une chaîne de Markov $(X_n)_n$ finie irréductible (donc récurrente positive) dont on observe l'évolution, dont on ne connaît que l'ensemble des états, les théorèmes suivants permettent sous certaines conditions d'estimer sa matrice de transition et sa distribution stationnaire, à partir d'une réalisation $(X_i = x_i)_{i=1, \dots, n}$ de la chaîne.

Définition 7.21 Pour tout (i, j) , on définit la v.a. $\widehat{p}_{i,j}^{(n)}$ égale au rapport du nombre de transitions de e_i à e_j pendant la durée $[0, n-1]$, sur le temps de séjour en e_i pendant la même durée :

$$\widehat{p}_{i,j}^{(n)} = \frac{\sum_{t=0}^{n-1} \mathbf{1}_{\{X_t=i, X_{t+1}=j\}}}{\sum_{t=0}^{n-1} \mathbf{1}_{\{X_t=i\}}}.$$

Théorème 7.24 Pour tout (i, j) , $p_{i,j}^{(n)}$ vérifie la propriété :

$$\widehat{p}_{i,j}^{(n)} \xrightarrow{(n \rightarrow \infty)} p_{i,j} \text{ (p.s.)}.$$

Définition 7.22 Pour tout i , on définit la v.a. $\widehat{\pi}_i^{(n)}$ égale à la proportion du temps d'occupation de l'état e_i pendant la durée $[0, n]$: $\widehat{\pi}_i^{(n)} = \frac{1}{n+1} \sum_{t=0}^n \mathbf{1}_{\{X_t=i\}}$.

Théorème 7.25 La v.a. $\widehat{\pi}_i^{(n)}$, vérifie les propriétés :

- (1) $E(\widehat{\pi}_i^{(n)}) \xrightarrow{(n \rightarrow \infty)} \pi_i$;
- (2) $\widehat{\pi}_i^{(n)} \xrightarrow{(n \rightarrow \infty)} \pi_i \text{ (p.s.)}.$

Illustration sur une chaîne à deux états e_1, e_2 : étant donnée la trajectoire observée pendant la durée $[0, 16]$, $(e_1, e_2, e_2, e_1, e_1, e_2, e_1, e_2, e_2, e_1, e_1, e_2, e_2, e_1, e_1)$, il est facile de vérifier que : $\widehat{\pi}_1^{(16)} = \frac{9}{16}$, $\widehat{\pi}_2^{(16)} = \frac{7}{16}$, $\widehat{p}_{1,2}^{(16)} = \frac{4}{8}$, $\widehat{p}_{1,1}^{(16)} = \frac{4}{8}$, $\widehat{p}_{2,1}^{(16)} = \frac{4}{7}$ et $\widehat{p}_{2,2}^{(16)} = \frac{3}{7}$.

7.5.2 Probabilité d'absorption par les classes récurrentes dans le cas d'une chaîne finie

Problème : Étant donné une chaîne finie non irréductible possédant r classes récurrentes R_1, \dots, R_r ; quelle est la probabilité q_j^i , partant d'un état transitoire j , d'atteindre la classe de récurrence R_i , et donc d'y être absorbé ?

Soient :

- $q_j^i = P(\text{d'être absorbé par la classe } R_i, \text{ partant de l'état transitoire } e_j)$;
- $q^{(i)} = {}^t(q_{j_1}^i, \dots, q_{j_t}^i)$ où j_k parcourt l'ensemble E_T des états transitoires;
- P_T , la sous-matrice de transition de l'ensemble T des états transitoires.

Théorème 7.26 (Probabilités d'absorption par les classes récurrentes)

Le vecteur $q^{(i)}$ des probabilités d'absorption par la classe R_i à partir de l'ensemble des états transitoires est solution du système :

$$(I_d - P_T)q^{(i)} = {}^t\left(\sum_{l \in R_i} p_{j_1, l}, \dots, \sum_{l \in R_i} p_{j_t, l}\right).$$

Exemple 7.12 Soit la chaîne, à trois états, de matrice de transition :

$$P = \begin{pmatrix} 1/4 & 5/8 & 1/8 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

La matrice de transition des états transitoires : $P_T = \begin{pmatrix} 1/4 & 5/8 \\ 1/2 & 1/2 \end{pmatrix}$.

On vérifie facilement que le vecteur $q^{(3)} =$ probabilité d'absorption par la seule classe de récurrence réduite ici à l'état e_3 est la solution du système :

$$(I_d - P_T)q^{(3)} = \begin{pmatrix} 1/8 \\ 0 \end{pmatrix}.$$

Il est clair que : $q^{(3)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$: l'état e_3 est donc certainement atteint à partir de e_1 ou e_2 .

7.5.3 Temps moyens de séjour dans les états transitoires

On suppose que la chaîne démarre à partir d'un état transitoire e_i . Quel temps moyen reste-t-elle dans cet état ?

Définition 7.23 Étant donné l'ensemble E_T des états transitoires, soient e_i et e_j deux tels états. Le nombre moyen de visites dans l'état e_j à partir de l'état e_i est égal à $E[N_j | X_0 = e_i]$, où N_j est la v.a. qui compte le nombre de passages de la chaîne par l'état e_j .

Notons : T_S la matrice $(E(N_j|X_0 = e_i))_{i,j}$.

Théorème 7.27

$$T_S = (I_d - P_T)^{-1}.$$

Exemple 7.13 Considérons la chaîne étudiée dans l'exemple (7.9) ; on pourra vérifier que les temps moyens de séjour dans les états transitoires sont égaux à 2,665 dans e_4 , 1,668 dans e_8 et 4,567 dans e_{10} .

Probabilités d'absorption par les états récurrents :

	à partir de	e_4	e_8	e_{10}
par R_1	→	0,152	0,139	0,109
par R_2	→	0,848	0,861	0,891

Temps moyens de retour dans les états de la classe récurrente apériodique $\{e_5, e_6, e_7\}$:

e_5	e_6	e_7
3,06	4,11	2,31

Intermède historique : essor de la théorie des processus aléatoires

Tout au long de ces pages, ont été évoquées les grandes étapes du développement de la théorie des probabilités aboutissant à une première synthèse due à Laplace. Ce n'est qu'au cours de la première moitié du siècle précédent, que le champ d'étude des probabilités s'est profondément transformé. L'établissement de ses bases dans le cadre de la théorie de la mesure et de l'intégration due en grande partie à l'école française dont les protagonistes furent Henri Lebesgue (1872-1941) et Émile Borel (1871-1955), permit au mathématicien soviétique Andrei Kolmogorov (1903-1987) de produire en 1933, dans une célèbre monographie, une formulation axiomatique définitive de la théorie des probabilités.

Chef de file de l'école probabiliste russe, Andrei Markov (1856-1922), qui étudia les processus qui portent son nom, est considéré comme le fondateur de la théorie des processus aléatoires. Débute alors une époque féconde, où se développa la théorisation des processus aléatoires : Lévy (1886-1971) en France, Kolmogorov et Khintchine (1894-1959) en URSS, Doob et Feller (1906-1970) aux États-Unis en furent les créateurs les plus inspirés. Depuis 1950, la théorie des processus stochastiques fait l'objet d'applications puissantes et nombreuses à la théorie du signal, à la physique, à l'économie mathématique et à bien d'autres domaines.

Les problèmes 15 et 16 se rapportent à ce chapitre.

7.6 Corrigés des exercices

Exercice 7.1

Pour déterminer P^n , on diagonalise P .

On sait que $\lambda_1 = 1$ est toujours valeur propre de P , associée au vecteur propre ${}^t(1, \dots, 1)$, car $\sum_{y \in E} p_{x,y} = 1$. Pour trouver l'autre valeur propre, on utilise la propriété $\lambda_1 + \lambda_2 = \text{trace}(P) = 2 - \alpha - \beta$: d'où $\lambda_2 = 1 - \alpha - \beta \neq 1$ si $P \neq I$: P est donc diagonalisable.

Recherche des vecteurs propres : pour $\lambda_2 = 1 - \alpha - \beta$, $(1 - \alpha)x_1 + \alpha x_2 = (1 - \alpha - \beta)x_1$ conduit à $\beta x_1 + \alpha x_2 = 0$.

Soit la matrice de passage Ω égale à $\begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix}$; $\Omega^{-1} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix}$
et

$$\pi(n) = \pi(0)\Omega \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} \Omega^{-1} = \left(\frac{\beta + \alpha(1 - \alpha - \beta)^n}{\alpha + \beta}, \frac{\alpha - \alpha(1 - \alpha - \beta)^n}{\alpha + \beta} \right).$$

Exercice 7.2

Soit Y_i , le résultat du i -ème lancer.

(1) $X_{n+1} = \max(X_n, Y_{n+1})$ est de la forme $\psi(X_n, Y_{n+1})$; d'après le théorème 7.2, $(X_n)_n$ est une chaîne de Markov.

$$(2) E = \{1, 2, 3, 4, 5, 6\} \text{ et } P = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 1/3 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 1/2 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 2/3 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

(3) Soit $\pi(0) = (p_1, p_2, p_3, p_4, p_5, p_6)$.

$$\lim_{n \rightarrow +\infty} P^n = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

d'où : $\lim_{n \rightarrow \infty} \pi(n) = \lim_{n \rightarrow \infty} \pi(0).P^n = (0, 0, 0, 0, 0, 1)$.

Autrement dit : $\lim_{n \rightarrow \infty} X_n = 6$.

Exercice 7.3

(1) $E = \mathbb{Z}$. Pour être de nouveau au point de départ après n étapes, il faut avoir fait autant de pas vers la droite que vers la gauche : ainsi, n doit être pair ($n = 2m$). Il y a autant de trajets possibles que de $2m$ u-plets avec m « d » et m « g » (g et d désignant respectivement un pas à gauche et un pas à droite), soit C_{2m}^m , et ils sont tous de probabilité $(\frac{1}{2})^{2m}$. Ainsi, $p_{x,x}^{(2m)} = C_{2m}^m (\frac{1}{2})^{2m}$ (et $p_{x,x}^{(2m+1)} = 0$).

Pour étudier la nature de la série, on utilise la formule de Stirling quand n tend vers l'infini :

$$p_{xx}^{(2m)} = \frac{(2m)!}{(m!)^2} \left(\frac{1}{2}\right)^{2m} \sim \frac{(2m)^{2m}}{e^{2m}} \sqrt{2\pi \times 2m} \left(\frac{e^m}{m^m \sqrt{2\pi m}}\right)^2 \times \frac{1}{2^{2m}} = \frac{1}{\sqrt{\pi m}}.$$

Ainsi, $p_{xx}^{(2m)} \sim \frac{1}{\sqrt{\pi m}}$ terme général tendant vers 0 d'une série divergente entraîne que la chaîne est récurrente nulle.

(2) $E = \mathbb{Z}^2$. Pour être de nouveau au point de départ après n étapes, il faut avoir fait dans chacune des deux directions autant de pas dans un sens que dans l'autre : ainsi, n doit être pair ($n = 2m$). S'il y a $2k$ pas verticaux (et donc $2m - 2k$ pas horizontaux), alors il doit y avoir k pas vers le haut, k pas vers le bas, $m - k$ pas vers la gauche et $m - k$ pas vers la droite. Ainsi, dans ces cas-là, il y a $C_{2m}^{2k} \times C_{2k}^k \times C_{2m-2k}^{m-k}$ trajets possibles (on a C_{2m}^{2k} façons de choisir les pas verticaux, puis, parmi ceux-ci, on en choisit C_{2k}^k vers le haut et, parmi les $2m - 2k$ horizontaux, on choisit les C_{2m-2k}^{m-k} vers la droite par exemple), et ils sont tous de probabilité $(\frac{1}{4})^{2m}$. Mais, comme k peut prendre toutes les valeurs de 0 à m , il vient :

$$\begin{aligned} p_{xx}^{(2m)} &= \sum_{k=0}^m \frac{(2m)!}{(2k)!(2m-2k)!} \frac{(2k)!}{(k!)^2} \frac{(2m-2k)!}{((m-k)!)^2} \left(\frac{1}{4}\right)^{2m} \\ &= C_{2m}^m \left(\frac{1}{4}\right)^{2m} \left(\sum_{k=0}^m (C_m^k)^2\right) = \left[C_{2m}^m \left(\frac{1}{2}\right)^{2m}\right]^2 \sim \frac{1}{\pi m}, \end{aligned}$$

car $\sum_{k=0}^m (C_m^k)^2 = \sum_{k=0}^m C_m^k C_m^{m-k} = C_{2m}^m$ (on peut par exemple développer $(1+x)^{2m} = (1+x)^m (1+x)^m$ de deux façons différentes et identifier le coefficient de x^m dans chacune des expressions). Même conclusion que précédemment ; la chaîne est récurrente nulle.

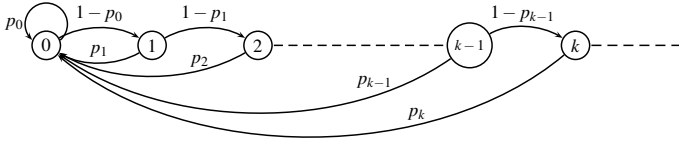
(3) $E = \mathbb{Z}^3$. Même raisonnement qu'en (1) et (2) : après n étapes, il faut avoir fait dans chacune des trois directions autant de pas dans un sens que dans l'autre : ainsi, n doit être pair ($n = 2m$). S'il y a $2k_1$ pas dans la direction 1, $2k_2$ pas dans la direction 2 (et donc $2m - 2k_1 - 2k_2$ pas dans la direction 3),

alors il doit y avoir k_1 pas dans chaque sens de la direction 1, k_2 dans ceux de la direction 2 et $m - k_1 - k_2$ dans ceux de la direction 3. Ainsi, dans ces cas-là, il y a $C_{2m}^{2k_1} \times C_{2m-2k_1}^{2k_2} \times C_{2k_1}^{k_1} \times C_{2k_2}^{k_2} \times C_{2(m-k_1-k_2)}^{m-k_1-k_2}$ trajets possibles et ils sont tous de probabilité $(\frac{1}{6})^{2m}$ (trois directions et deux sens dans chacune, donc six possibilités à chaque pas). On démontre que $p_{xx}^{(2m)}$ est égal à :

$$\begin{aligned} \sum_{k_1, k_2} \frac{(2m)!}{(2k_1)!(2m-2k_1)!} \frac{(2m-2k_1)!}{(2k_2)!(2m-2k_1-2k_2)!} \frac{(2k_1)!}{(k_1!)^2} \frac{(2k_2)!}{(k_2!)^2} \frac{(2m-2k_1-2k_2)!}{((m-k_1-k_2)!)^2} \left(\frac{1}{6}\right)^{2m} \\ = \sum_{k_1, k_2} \frac{(2m)!}{(k_1!)^2 (k_2!)^2 ((m-k_1-k_2)!)^2} \left(\frac{1}{6}\right)^{2m}. \end{aligned}$$

On majore le terme général de cette série par la fraction $\frac{K}{m^{3/2}}$, où K est une constante : on conclut à la convergence de la série et donc à la transience de la marche aléatoire.

Exercice 7.4



(1) La chaîne est irréductible étant donné la structure de son graphe.

(2) Il suffit donc de montrer que l'un des états, par exemple 0, est récurrent pour s'assurer que la chaîne est récurrente.

$P(\text{temps de retour en 0 soit fini}) = 1 - P(\text{temps de retour en 0 soit infini})$

$$= 1 - \prod_{i=0}^{+\infty} (1 - p_i), \text{ donc 0 est récurrent si } \prod_{i=0}^{+\infty} (1 - p_i) = 0.$$

Ce qui équivaut à : $\ln \left(\prod_{i=0}^{+\infty} (1 - p_i) \right) = \sum_{i=0}^{+\infty} \ln(1 - p_i) = -\infty$. Sachant que, si x est

proche de 0, $-\ln(1 - x) \sim x$, la série précédente est de même nature que $-\sum_{i=0}^{+\infty} p_i$.

Conclusion : 0 est récurrent si et seulement si $\sum_{i=0}^{+\infty} p_i = +\infty$.

(3) La chaîne est récurrente positive si l'espérance de T_0 , temps de retour en 0, est finie.

$$P(T_0 = n) = P(T_0 > n - 1) - P(T_0 > n) = p(1 - p)^{n-1},$$

donc T_0 est une v.a. géométrique $\mathcal{G}(p)$ et $E(T_0) = \frac{1}{p}$; la chaîne est donc récurrente positive.

Exercice 7.5

(1) X_n et Y_n sont indépendantes et Y_n suit la loi de Poisson $\mathcal{P}(\lambda)$ ($P([Y_n = j]) = a_j = e^{-\lambda} \frac{\lambda^j}{j!}$). On en déduit :

- si $X_n = 0$, alors $X_{n+1} = Y_n$, donc $P(X_{n+1} = j/X_n = 0) = P(Y_n = j) = a_j$;
- si $X_n \geq 1$, alors $X_{n+1} = X_n + Y_n - 1$ et $P(X_{n+1} = j/X_n = i) = P(Y_n = j - i + 1) = a_{j-i+1}$.

(X_n) est donc une chaîne de Markov de matrice de transition $P = (p_{ij})$ où $p_{0j} = a_j$, $p_{ij} = a_{j-i+1}$ si $1 \leq i$ et $j \geq i - 1$, $p_{ij} = 0$ sinon ;

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & \ddots & \ddots & \end{pmatrix}.$$

(2) $\pi = \pi P$ s'écrit :

$$\begin{cases} \pi_0 = a_0\pi_0 + a_0\pi_1 \\ \pi_1 = a_1\pi_0 + a_1\pi_1 + a_0\pi_2 \\ \pi_2 = a_2\pi_0 + a_2\pi_1 + a_1\pi_2 + a_0\pi_3 \\ \vdots \\ \pi_k = a_k\pi_0 + a_k\pi_1 + \cdots + a_0\pi_{k+1} = a_k\pi_0 + \sum_{i=0}^k a_k\pi_{k-i+1} \\ \vdots \end{cases}$$

Ainsi, en posant $\pi'_i = \pi_{i+1}$, on a $\pi_k = a_k\pi_0 + \sum_{i=0}^k a_k\pi'_{k-i}$ pour tout $k \in \mathbb{N}$ et

$$G(s) = \sum_{k=0}^{+\infty} \pi_k s^k :$$

$$G(s) = \pi_0 A(s) + A(s)B(s),$$

$$\text{où } B(s) = \sum_{i=0}^{+\infty} \pi'_i s^i = \sum_{i=0}^{+\infty} \pi_{i+1} s^i = \frac{1}{s} \sum_{i=0}^{+\infty} \pi_{i+1} s^{i+1} = \frac{1}{s} [G(s) - \pi_0].$$

Donc $sG(s) = \pi_0 sA(s) + A(s)G(s) - \pi_0 A(s)$ et on a bien $G(s) = \frac{\pi_0(s-1)A(s)}{s-A(s)}$.

$G(1+u) = \frac{\pi_0 u e^{\lambda u}}{1+u - e^{\lambda u}} \simeq (\text{Si } u \rightarrow 0) \frac{\pi_0}{1-\lambda} = G(1) = 1$ donc $\pi_0 = 1 - \lambda$, ce qui n'est possible que si $\lambda \in]0, 1[$.

Le nombre moyen de véhicules dans la station en régime stationnaire est alors donné par $G'(1)$, que l'on peut obtenir en faisant un développement limité à l'ordre 1 de $G(1+u)$, où $G(1+u) = (1-\lambda) \frac{ue^{\lambda u}}{1+u - e^{\lambda u}}$.

$$\begin{aligned} 1+u - e^{\lambda u} &= 1+u - \left(1 + \lambda u + \frac{\lambda^2}{2} u^2 + o(u^2)\right) \\ &= (1-\lambda)u \left[1 - \frac{\lambda^2}{2(1-\lambda)} u + o(u)\right], \end{aligned}$$

et $ue^{\lambda u} = u(1 + \lambda u + o(u))$ donc $G(1+u) = \frac{1+\lambda u+o(u)}{1-\frac{\lambda^2}{2(1-\lambda)}u+o(u)} = 1 + \left[\lambda + \frac{\lambda^2}{2(1-\lambda)} \right] u + o(u)$, d'où le nombre moyen de véhicules égal à : $\lambda + \frac{\lambda^2}{2(1-\lambda)}$.

Exercice 7.6

(1) La connaissance de l'état au n -ième jour suffit pour déterminer les probabilités d'occupation des états au $(n+1)$ -ième jour et celles-ci sont indépendantes de n . Ainsi, on a bien une chaîne de Markov telle que :

si $X_n = 0$ alors : $X_{n+1} = 1$ si les deux machines tombent en panne le n -ième jour, et $X_{n+1} = 0$ sinon ;

si $X_n = 1$ alors : $X_{n+1} = 1$ si la machine en service tombe en panne le n -ième jour, et $X_{n+1} = 0$ sinon.

En notant $p_{ij} = P([X_{n+1} = j] / [X_n = i])$, on a $p_{0,1} = q^2$ (les machines fonctionnent indépendamment) ; $p_{0,0} = 1 - q^2$; $p_{1,1} = q$; $p_{1,0} = 1 - q$.

$$P = \begin{pmatrix} 1-q^2 & q^2 \\ 1-q & q \end{pmatrix}.$$

On cherche $\pi = (\pi_0, \pi_1)$ tel que $\pi = \pi P$ et $\pi_0 + \pi_1 = 1$.

Cela donne $\pi_1 = q^2 \pi_0 + q \pi_1$, soit $\pi_1 = \frac{q^2}{1-q} \pi_0$ et $\pi_0 \left(1 + \frac{q^2}{1-q} \right) = 1$. La distribution stationnaire est $\pi = \left(\frac{1-q}{1-q+q^2}, \frac{q^2}{1-q+q^2} \right)$.

(2) Le réparateur ne travaillant plus la nuit, il peut y avoir le matin 0, 1 ou 2 machines en pannes.

si $X_n = 0$, on peut avoir $X_{n+1} = 0$ (0 panne), $X_{n+1} = 1$ (1 panne) ou $X_{n+1} = 2$ (2 pannes) ;

si $X_n = 1$, (une seule machine en service), alors $X_{n+1} = 0$ si elle ne tombe pas en panne le n -ième jour, et $X_{n+1} = 1$ sinon (l'autre remarchera le lendemain) ;

si $X_n = 2$, alors $X_{n+1} = 1$, car la machine réparée remarchera. On a alors

$$P = \begin{pmatrix} (1-q)^2 & 2q(1-q) & q^2 \\ 1-q & q & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

$$\pi = \pi P \text{ et } \pi_0 + \pi_1 + \pi_2 = 1 \text{ donne ici : } \begin{cases} \pi_0 = (1-q)^2 \pi_0 + (1-q) \pi_1 ; \\ \pi_2 = q^2 \pi_0 ; \\ \pi_0 + \pi_1 + \pi_2 = 1. \end{cases} \text{ soit } \pi_1 =$$

$$\frac{2q-q^2}{1-q} \pi_0, \pi_2 = q^2 \pi_0 \text{ et } \pi_0 \left(1 + \frac{2q-q^2}{1-q} + q^2 \right) = 1, \text{ soit } \pi_0 = \frac{1-q}{1+q-q^3}.$$

$$\text{Ainsi, la distribution stationnaire est } \pi = \left(\frac{1-q}{1+q-q^3}, \frac{2q-q^2}{1+q-q^3}, \frac{q^2-q^3}{1+q-q^3} \right).$$

Exercice 7.7

Les axiomes de Markov et de stationnarité sont clairement vérifiés.

Posons pour tout (x_1, y_1) appartenant à l'espace d'états de $(X_n^{(1)})_n$ et pour tout (x_2, y_2) appartenant à l'espace d'états de $(X_n^{(2)})_n$: $A_n = (X_n^{(1)} = x_1)$, $A_{n+1} = (X_{n+1}^{(1)} = y_1)$, $B_n = (X_n^{(2)} = x_2)$ et $B_{n+1} = (X_{n+1}^{(2)} = y_2)$,

$$\begin{aligned} P(A_{n+1} \cap B_{n+1} | A_n \cap B_n) &= \frac{P(A_{n+1} \cap A_n \cap B_{n+1} \cap B_n)}{P(A_n \cap B_n)} \\ &= \frac{P(A_{n+1} \cap A_n)P(B_{n+1} \cap B_n)}{P(A_n)P(B_n)} \\ &= P(A_{n+1} | A_n)P(B_{n+1} | B_n). \end{aligned}$$

$$\begin{aligned} P(X_{n+1}^{(1)} = y_1, X_{n+1}^{(2)} = y_2 | X_n^{(1)} = x_1, X_n^{(2)} = x_2) \\ = P(X_{n+1}^{(1)} = y_1 | X_n^{(1)} = x_1)P(X_{n+1}^{(2)} = y_2 | X_n^{(2)} = x_2). \end{aligned}$$

Exercice 7.8

(1) On a $P([X_{n+1} = k] | [X_n = i]) = P([s - Y_n = k] | [X_n = i]) = P([Y_n = s - k] | [X_n = i]) = \frac{1}{i+1}$ si $0 \leq s - k \leq i$, c'est à dire si $s - i \leq k \leq s$. $P([X_{n+1} = k] | [X_n = i])$ est donc indépendant de n : l'axiome d'homogénéité est bien vérifié. En outre, la connaissance de X_0, \dots, X_{n-1} n'apporterait rien de plus pour déterminer la loi de X_{n+1} , donc l'axiome de Markov est aussi vérifié et $(X_n)_n$ est bien une chaîne de Markov de matrice de transition $P = (p_{ij})$ où $p_{ij} = \frac{1}{i+1}$ si $s - i \leq j \leq s$ et 0 sinon.

Pour $\pi_i = \frac{2(i+1)}{(s+1)(s+2)}$, on a $\sum_{i=0}^s \pi_i = \frac{2}{(s+1)(s+2)} \sum_{i=0}^s (i+1) = \frac{2}{(s+1)(s+2)} \sum_{i=1}^{s+1} i = 1$.

$(\pi P)_j = \sum_i \pi_i p_{ij} = \sum_{i=s-j}^s \frac{2(i+1)}{(s+1)(s+2)} \frac{1}{i+1} = \frac{2}{(s+1)(s+2)} (j+1) = \pi_j$. La chaîne étant irréductible finie, elle admet une unique distribution stationnaire : c'est donc (π_i) où $\pi_i = \frac{2(i+1)}{(s+1)(s+2)}$.

(2) $N = \sum_{i=0}^s i \pi_i = \frac{2}{(s+1)(s+2)} \left[\sum_{i=1}^s i(i+1) \right]$ avec $\sum_{i=1}^s i^2 = \frac{s(s+1)(2s+1)}{6}$ et $\sum_{i=1}^s i = \frac{s(s+1)}{2}$, donc $\sum_{i=1}^s i(i+1) = \frac{1}{6} [s(s+1)(2s+4)] = \frac{s(s+1)(s+2)}{3}$, donc $N = \frac{2s}{3}$ et pour $s = 30$, $N = 20$.

Chapitre 8

Processus de Poisson et de renouvellement

Les processus de Poisson et de renouvellement sont des processus aléatoires de comptage à temps continu qui conviennent à la description des phénomènes dont les occurrences surviennent en des temps successifs aléatoires. Ils sont utiles à la modélisation des files d'attente apparaissant à l'entrée de services et dans les réseaux de communication, et permettent de résoudre des problèmes de maintenance.

8.1 Processus de Poisson homogène

Définition 8.1 *Un **processus de comptage** $(N_t)_{t \geq 0}$ est un processus qui permet de dénombrer les occurrences d'un événement aléatoire donné en fonction du temps : temps d'arrivée des clients à un service, temps d'arrivée de requêtes à un ordinateur, etc.*

N_t est la v.a. égale au nombre d'occurrences appartenant à l'intervalle de temps $[0, t]$; $(N_t - N_s)$ désigne donc le nombre d'occurrences appartenant à l'intervalle de temps $]s, t]$.

Définition 8.2 *Un **processus de Poisson homogène d'intensité λ** est un processus de comptage $(N_t)_{t \geq 0}$ qui vérifie les propriétés suivantes :*

(1) $(N_t)_{t \in \mathbb{R}^+}$ **est à accroissements indépendants** ; les occurrences sur des intervalles de temps disjoints sont donc indépendantes entre elles :

$$\forall (t_i)_{i=1, \dots, n} \text{ vérifiant } 0 < t_1 < t_2 < \dots < t_n,$$

$(N_{t_1} - N_0), (N_{t_2} - N_{t_1}), \dots, (N_{t_n} - N_{t_{n-1}})$ sont des v.a. indépendantes.

(2) $\forall s, t > 0$, la v.a. $(N_{t+s} - N_s)$ est de loi de Poisson de paramètre $\lambda.t$:

$$(\forall k \in \mathbb{N}) \quad P(N_{t+s} - N_s = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Le processus démarrant à l'instant $t = 0$:

$$P(N_t = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

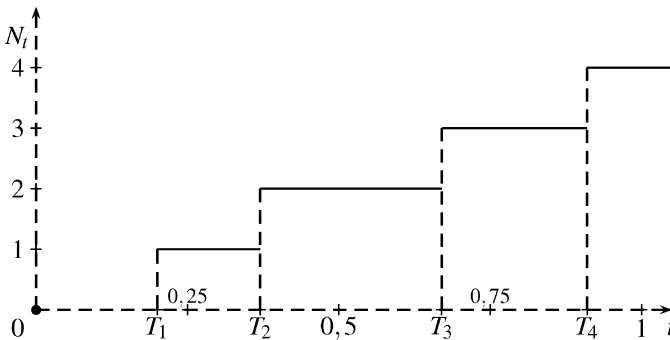
Théorème 8.1 L'intensité d'un processus de Poisson, définie comme étant le nombre moyen d'occurrences par unité de temps, est égale à λ .

Preuve $\forall t \quad \frac{E(N_{t+1} - N_t)}{1} = \frac{E(N_1 - N_0)}{1} = E(N_1) = \lambda$, car $N_1 \sim \mathcal{P}(\lambda)$. □

Exemple 8.1 Le processus de comptage des voitures contrôlées à un guichet d'autoroute, durant un intervalle de temps donné où l'intensité λ est constante, est un processus de Poisson. Supposons que le nombre de voitures franchissant un péage donné est décrit par un processus de Poisson de paramètre $\lambda = 300$ (voitures par heure). Le nombre aléatoire de véhicules franchissant le péage en une minute est décrit par la v.a. de Poisson $N(s + \frac{1}{60}) - N(s)$ de loi :

$$\forall k \quad P(N(s + \frac{1}{60}) - N(s) = k) = e^{-\frac{300}{60}} \frac{(\frac{300}{60})^k}{k!} = \frac{e^{-5} 5^k}{k!}.$$

Graphes d'une réalisation du processus de Poisson d'intensité $\lambda = 5$.



(T_n désigne le temps de réalisation de la n -ième occurrence.)

8.1 ♠♠ Soient $(N_t)_{t \geq 0}$ un processus de Poisson d'intensité λ et deux entiers i, j tels que $i < j$.

(1) Pour tout entier k , déterminer $P(N_u = k \mid (N_s = i) \cap (N_t = j))$ dans chacun des cas suivants : (a) $s \leq t \leq u$; (b) $s \leq u \leq t$; (c) $u \leq s \leq t$.

(2) Le flux des clients arrivant dans un service est décrit par le processus de Poisson d'intensité dix clients/h; déterminer la probabilité pour qu'au bout

d'une demi-heure deux clients soient arrivés, sachant que dans le premier quart d'heure aucun client n'est arrivé et que deux clients sont arrivés durant la première heure.

8.2 ♠♠ Étude du processus de Bernoulli

À une suite de v.a. $(X_n)_n$ indépendantes de même loi de Bernoulli $P(X_n = 1) = p = 1 - P(X_n = 0)$, on associe le processus de Bernoulli $(B_n)_n$, défini par $B_n = \sum_{i=1}^n X_i$.

(1) Démontrer que le processus de Bernoulli est une chaîne de Markov.

(2) Soit T_n le temps d'occurrence du n -ième saut d'amplitude égale à l'unité, effectué par le processus. On pose : $U_n = T_n - T_{n-1}$; déterminer la loi de U_n et démontrer que T_n est de loi de Pascal, définie par : $P(T_n = k) = C_{k-1}^{n-1} p^n (1-p)^{k-n} \forall k \geq n$.

(A) Caractérisation du processus de Poisson homogène fondée sur l'hypothèse des événements rares

Définition 8.3 *Un processus de comptage $(N_t)_{t \geq 0}$ vérifie l'hypothèse des événements rares si pour tout accroissement infinitésimal Δt du temps :*

- (1) $P(N_{\Delta t} = 1) = \lambda \Delta t + o(\Delta t)$;
- (2) $P(N_{\Delta t} \geq 2) = o(\Delta t)$.

On déduit de (1) et (2) : $P(N_{\Delta t} = 0) = 1 - \lambda \Delta t + o(\Delta t)$.

Théorème 8.2 (Caractérisation du processus de Poisson par les hypothèses d'événements rares)

Un processus de comptage à accroissements stationnaires indépendants et vérifiant l'hypothèse des événements rares est un processus de Poisson.

$$\begin{aligned} \text{Preuve } P(N_{t+h} = k) &= P((N_t = k) \cap (N_{t+h} - N_t = 0)) \\ &+ P((N_t = k-1) \cap (N_{t+h} - N_t = 1)) + \sum_{j=0}^{k-2} P((N_t = j) \cap (N_{t+h} - N_t = k-j)). \end{aligned}$$

La propriété caractéristique des processus à accroissements indépendants stationnaires conduit aux égalités :

$$\begin{aligned} P(N_{t+h} = k) &= P(N_t = k)P(N_h = 0) + P(N_t = k-1)P(N_h = 1) + \\ &\sum_{j \leq k-2} P(N_t = j) \underbrace{P(N_h = k-j)}_{o(h)}. \end{aligned}$$

Posons $f_k(t) = P_{N_t}(k)$: $f_k(t+h) = f_k(t)f_0(h) + f_{k-1}(t)f_1(h) + o(h)$ (*), équation de récurrence, résoluble dès que sont connus les deux premiers termes $f_0(h)$ et $f_1(h)$.

$(N_t)_t$ étant un processus stationnaire à accroissements indépendants : $f_0(h) = e^{-\lambda h}$ et $f_1(h) = \lambda h e^{-\lambda h}$. Si h infiniment petit : $f_0(h) = 1 - \lambda h + o(h)$ et $f_1(h) = \lambda h + o(h)$. D'où en portant dans (*) : $f'_k(t) = \lambda(f_{k-1}(t) - f_k(t))$, qui est l'équation de Kolmogorov du processus.

On pourra vérifier que $f_k(t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$ est la solution du système récurrent différentiel. \square

Théorème 8.3 *La v.a. T_n associée au temps de réalisation de la n -ième occurrence vérifie la relation :*

$$\{T_n \leq t\} = \{N_t \geq n\}.$$

8.3 ♠ (1) Déterminer $F_{T_n}(t)$ en fonction de P_{N_t} , puis $f_{T_n}(t)$.

[Solution : loi *gamma* $\gamma(\lambda, n)$.]

(2) Déterminer la covariance et le coefficient de corrélation du couple (N_s, N_t) sachant que $(0 < s < t)$; interpréter ces résultats en fonction des valeurs de s et t . [Utiliser la décomposition : $N_s N_t = N_s(N_t - N_s) + N_s^2$].

Définition 8.4 *Les v.a. $U_i = T_i - T_{i-1}$, $\forall i \geq 2$ et $U_1 = T_1$ sont les intervalles de temps entre deux occurrences consécutives appelés aussi **temps d'interarrivées**.*

(B) Caractérisation du processus de Poisson homogène par l'indépendance des temps d'interarrivées et le caractère exponentiel de leur loi commune.

Elle repose sur le théorème suivant :

Théorème 8.4 (de Dirichlet) *Si (X_1, \dots, X_n) sont des v.a. indépendantes de même densité f_X , alors la statistique d'ordre $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ formée de l'ensemble des X_i rangées par ordre croissant admet la densité :*

$$f_{(X_{(1)}, \dots, X_{(n)})}(x_1, \dots, x_n) = \begin{cases} n! (\prod_{i=1}^n f_X(x_i)) & \text{si } (x_1 < x_2 < \dots < x_n) \\ 0 & \text{sinon.} \end{cases}$$

Preuve Considérons $n = 2$: $X_{(1)} = \inf(X_1, X_2)$ et $X_{(2)} = \sup(X_1, X_2)$. L'événement $(X_{(1)} < x_1, X_{(2)} < x_2)$ est égal à la réunion disjointe des événements

$(X_1 < X_2, X_1 < x_1, X_2 < x_2)$ et $(X_2 < X_1, X_2 < x_1, X_1 < x_2)$ d'où :

$$\begin{aligned} P(X_{(1)} < x_1, X_{(2)} < x_2) &= P(X_1 < X_2, X_1 < x_1, X_2 < x_2) \\ &+ P(X_2 < X_1, X_2 < x_1, X_1 < x_2). \end{aligned}$$

Ce résultat se généralise immédiatement par récurrence au cas de n v.a. L'égalité :

$$P(X_1 < x_1, X_2 < x_2, X_1 < X_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1}(y_1) f_{X_2}(y_2) \mathbf{1}_{(y_1 < y_2)}(y_1, y_2) dy_1 dy_2$$

devient pour n v.a. :

$$P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n, X_1 < X_2 \dots < X_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_{X_1}(y_1) f_{X_2}(y_2) \dots f_{X_n}(y_n) \mathbf{1}_{(y_1 < y_2 \dots < y_n)}(y_1, y_2, \dots, y_n) dy_1 dy_2 \dots dy_n.$$

Donc compte tenu de l'identité des lois des $X_i : F_{(X_{(1)}, \dots, X_{(n)})}(x_1, \dots, x_n) = n! \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_{X_1}(y_1) f_{X_2}(y_2) \dots f_{X_n}(y_n) \mathbf{1}_{(y_1 < y_2 \dots < y_n)}(y_1, y_2, \dots, y_n) dy_1 dy_2 \dots dy_n$ et

$$f_{(X_{(1)}, \dots, X_{(n)})}(x_1, \dots, x_n) = n! f_X(y_1) f_X(y_2) \dots f_X(y_n) \mathbf{1}_{(y_1 < y_2 \dots < y_n)}(y_1, y_2, \dots, y_n).$$

□

Théorème 8.5 *Le processus de comptage $(N_t)_{t \in \mathbb{R}_+}$ est un processus de Poisson d'intensité λ si et seulement si les temps d'inter-arrivées $(U_i)_i$ sont indépendants et de même loi exponentielle $\mathcal{E}(\lambda)$.*

Preuve (1) Condition nécessaire :

Soient n intervalles disjoints $(t_i, t_i + h_i)$ où $t_1 < \dots < t_n$.

$$\begin{aligned} P(T_1 \in (t_1, t_1 + h_1), \dots, T_n \in (t_n, t_n + h_n)) \\ = P(N_{t_1} = 0, N_{t_1+h_1} - N_{t_1} = 1, N_{t_2} - N_{t_1+h_1} = 0, \dots, N_{t_n+h_n} - N_{t_n} = 1) \\ = e^{-\lambda t_1} \times e^{-\lambda h_1} \lambda h_1 \times e^{-\lambda(t_2-t_1-h_1)} \times \dots \times e^{-\lambda(t_n-t_{n-1}-h_{n-1})} \times e^{-\lambda h_n} \\ \simeq e^{-\lambda t_n} \lambda^n h_1 \times \dots \times h_n. \end{aligned}$$

On en déduit la densité du vecteur (T_1, \dots, T_n) aux points (t_1, \dots, t_n) tels que $t_1 < \dots < t_n : f(t_1, \dots, t_n) = \lambda^n e^{-\lambda t_n}$.

Calculons la loi du vecteur (U_1, \dots, U_n) par la méthode du changement de variables (voir chapitre 4) :

$$f_{(U_1, \dots, U_n)}(x_1, \dots, x_n) = \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} \dots \lambda e^{-\lambda x_n} \quad \forall (x_1, \dots, x_n) \in \mathbb{R}_+^n,$$

dont la forme analytique permet de conclure : $f_{(U_1, \dots, U_n)}(u_1, \dots, u_n) = \prod_{i=1}^n f_{U_i}(u_i)$. Les U_i sont donc indépendantes de même densité exponentielle $\lambda e^{-\lambda u}$.

(2) Condition suffisante :

Calcul de la loi de N_t : exprimons d'abord la loi du vecteur (T_1, \dots, T_n) à partir de celle de (U_1, \dots, U_n) . D'après le théorème de Dirchlet :

$$f_{T_1, \dots, T_n}(t_1, \dots, t_n) = \mathbf{1}_{\{0 < t_1 < \dots < t_n\}} \prod_{i=0}^{n-1} \lambda e^{-\lambda(t_{i+1}-t_i)} = \mathbf{1}_{\{0 < t_1 < \dots < t_n\}} \lambda^n e^{-\lambda t_n}.$$

Remarquons que $\forall n : (N_t = n - 1) \equiv (T_i < t \ \forall i \leq n - 1) \text{ et } (T_n > t)$, d'où :

$$\begin{aligned} P(N_t = n - 1) &= \int \mathbf{1}_{\{0 < t_1 < \dots < t_{n-1} < t < t_n\}} \lambda^n e^{-\lambda t_n} dt_1 \dots dt_n \\ &= \int \mathbf{1}_{\{0 < t_1 < \dots < t_{n-1} < t < t_n\}} dt_1 \dots dt_{n-1} \int \lambda^n e^{-\lambda t_n} dt_n \\ &= \frac{1}{(n-1)!} (\lambda t)^{n-1} e^{-\lambda t}. \end{aligned}$$

Montrons que le processus est à accroissements indépendants.

Soient $\Delta_i = [t_i, t_i + dt_i]$, $\Delta = \prod_{i=1}^n \Delta_i$, et $T = (T_1, \dots, T_n)$.

$$P(T \in \Delta | N_t = n) = \frac{P(T \in \Delta, N_t = n)}{P(N_t = n)}.$$

Or $P(T \in \Delta, N_t = n) = \mathbf{1}_{\{0 < t_1 < \dots < t_n\}} \times \lambda^n \times e^{-\lambda t} . dt_1 dt_2 \dots dt_n$, d'où :

$P(T \in \Delta | N_t = n) = \mathbf{1}_{\{0 < t_1 < \dots < t_n\}} \frac{n!}{t^n} dt_1 \dots dt_n$. On utilise la propriété de Dirichlet : donnons-nous une subdivision $(0 = r_0, r_1, \dots, r_k = t)$ de l'intervalle $(0, t)$; la loi des nombres S_i qui tombent dans les intervalles $r_{i+1} - r_i$ est une multinomiale de paramètres $\frac{r_1 - r_0}{t}, \frac{r_2 - r_1}{t}, \dots, \frac{r_k - r_{k-1}}{t}$.

Appliquons ce résultat au nombre des temps T_i tombant dans une telle subdivision $r_1 - r_0, \dots, r_k - r_{k-1}$ et identifions la loi multinomiale de paramètres $\frac{r_1 - r_0}{t}, \frac{r_2 - r_1}{t}, \dots, \frac{r_k - r_{k-1}}{t}$.

$$\begin{aligned} P(N_{r_1} = n_1, N_{r_2} - N_{r_1} = n_2, \dots, N_{r_k} - N_{r_{k-1}} = n_k | N_{r_k} = n) \\ = \frac{n!}{n_1! \dots n_k!} \left(\frac{r_1}{t}\right)^{n_1} \left(\frac{r_2 - r_1}{t}\right)^{n_2} \dots \left(\frac{r_k - r_{k-1}}{t}\right)^{n_k} P(N_{r_1} = n_1, \dots, N_{r_k} - N_{r_{k-1}} = n_k) \\ = \frac{\sum_{i=1}^k n_i}{n_1! \dots n_k!} \frac{r_1^{n_1} (r_2 - r_1)^{n_2} \dots (r_k - r_{k-1})^{n_k}}{\sum_{i=1}^k n_i} \times e^{-\lambda t} \frac{(\lambda t)^{\sum n_i}}{(\sum n_i)!}. \end{aligned}$$

$$\begin{aligned} P(N_{r_1} = n_1, \dots, N_{r_k} - N_{r_{k-1}} = n_k) = \\ e^{-\lambda r_1} \frac{(\lambda r_1)^{n_1}}{n_1!} . e^{-\lambda r_2} \frac{(\lambda (r_2 - r_1))^{n_2}}{n_2!} \dots e^{-\lambda r_k} \frac{(\lambda (r_k - r_{k-1}))^{n_k}}{n_k!}. \end{aligned}$$

$(N_t)_t$ est à accroissements indépendants et chaque accroissement $(N_{r_j} - N_{r_{j-1}})$ est de loi de Poisson de paramètre $\lambda(r_j - r_{j-1})$. Conclusion : $(N_t)_{t \in \mathbb{R}_+}$ est bien un processus de Poisson. \square

8.4 ♠♠♠ Paradoxe de la station de bus

On suppose que les arrivées de bus à une station constituent un processus de Poisson d'intensité λ , ce qui n'est pas réaliste compte tenu de la régulation imposée à un réseau citadin de bus. Un usager arrive à la station au temps θ : soient U le temps qui sépare θ du temps de la prochaine arrivée et V le temps qui sépare θ du temps de la dernière arrivée.

(1) Déterminer la loi de U , puis la loi de V en déterminant $P(V > x)$ dans chacun des deux cas : $0 < x < \theta$ et $x \geq \theta$.

(2) Démontrer l'indépendance des v.a. U et V , en calculant $P((U > y) \cap (V > x))$ dans chacun des deux cas précédents. On constate que $E(U + V)$ est de l'ordre de $2E(T)$, T désignant la durée qui sépare deux arrivées consécutives ; du point de vue d'un usager, l'intervalle moyen séparant deux arrivées de bus est donc deux fois supérieur à $\frac{1}{\lambda}$.

8.5 ♠♠♠ Sur une route à sens unique, l'écoulement des voitures est de loi de Poisson d'intensité $\lambda = \frac{1}{6}$. Un piéton arrive à l'instant t sur le bord de la route.

(1) Probabilité pour qu'il attende, sachant qu'il lui faut 4 secondes pour traverser ?

(2) Quelle est la durée moyenne des intervalles lui permettant de traverser ?

(3) Déterminer la loi du nombre N de voitures qu'un piéton voit passer avant de traverser. En moyenne, combien de temps doit-il attendre avant de traverser ?

Théorème 8.6 (Propriétés des processus de Poisson)

Le processus de Poisson $(N_t)_{t \in \mathbb{R}^+}$ d'intensité λ et de suite $(T_n)_{n \in \mathbb{N}}$ des temps d'occurrences possède les propriétés suivantes.

(1) Absence de mémoire :

$$\forall s, t \quad P(T_n \leq s + t | T_n > s) = P(T_n \leq t).$$

(2) Invariance par translation : la v.a. $T_{N(t)+1} - t$ est de loi exponentielle $\mathcal{E}(\lambda)$ et la suite des temps aléatoires $(T_{N(t)+k} - t)_{k \in \mathbb{N}}$ définit les occurrences d'un processus de Poisson de paramètre λ .

(3) La v.a. $(T_{n+k} - T_n)$ qui sépare k occurrences est de loi gamma $\gamma(\lambda, k)$.

(4) $\forall s, t \ (s < t) \ \forall k \leq n :$

$$P(N_s = k | N_t = n) = C_n^k \left(\frac{s}{t} \right)^k \left(1 - \frac{s}{t} \right)^{n-k}.$$

(5) La loi du vecteur aléatoire (T_1, \dots, T_n) conditionné par $(N_t = n)$ est de densité :

$$f_{(T_1, \dots, T_n)}(t_1, \dots, t_n) = \frac{n!}{t^n} \mathbf{1}_{\{0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq t\}}(t_1, \dots, t_n).$$

(6) Le processus de Poisson est une chaîne de Markov à temps continu, de probabilité de transition $P(N_t = n | N_s = m) = \frac{\lambda^{n-m} (t-s)^{n-m}}{(n-m)!} e^{-\lambda(t-s)}$ où $s < t$, m et n sont des entiers positifs tels que $m \leq n$; on reconnaît la loi de Poisson d'intensité $\lambda(t-s)$.

(7) *Stabilité du caractère poissonnien par sommation finie de processus de Poisson indépendants* : soient n processus de Poisson $(N_t^{(i)})_t$ indépendants d'intensités $(\lambda_i)_i$, alors

$$(\sum_{i=1}^n N_t^{(i)})_t \text{ est un processus de Poisson d'intensité } \sum_{i=1}^n \lambda_i.$$

(8) *Processus de Poisson soumis à effacement* : si chaque occurrence du processus est prise en compte avec une probabilité p , le processus de comptage qui en résulte est un processus de Poisson d'intensité $p\lambda$.

Preuve

$$(1) P(T_n \leq t+s | T_n > s) = \frac{P(s < T_n \leq s+t)}{P(T_n > s)} = \frac{P(T_n \leq s+t) - P(T_n \leq s)}{P(T_n > s)} = 1 - e^{-\lambda t} = P(T_n \leq t).$$

(2) Résulte de la propriété d'absence de mémoire des v.a. exponentielles.

(3) $T_{n+k} - T_n = \sum_{i=n+1}^{n+k} U_i$ somme de k v.a. exponentielles indépendantes de paramètre λ , donc de loi gamma $\gamma(\lambda, k)$, de densité : $f_k(t) = \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t} \mathbf{1}_{\mathbb{R}_+}(t)$.

(4) Se vérifie facilement.

(5) Conséquence directe du théorème 8.4.

(6) La propriété de Markov a été établie dans l'exercice 8.1 (a). \square

8.6 ♠♠ Démontrer les propriétés (7) pour $n = 2$ et (8).

8.7 ♠ Processus de Poisson décalé

Étant donné un processus de Poisson $(N_t)_t$ d'intensité λ , on considère les processus décalés $(N_t^{(1)})_t$ et $(N_t^{(2)})_t$ obtenus respectivement sous chacune des deux hypothèses suivantes.

(1) Chaque temps d'arrivée est décalé par translation d'un même temps constant $\alpha > 0$. $(N_t^{(1)})_t$, le processus obtenu, est-il un processus de Poisson ?

(2) Chaque temps d'arrivée est décalé par translation d'un temps aléatoire décrit par la v.a. X positive, indépendante du processus $(N_t)_t$, de densité f_X ; soit $(N_t^{(2)})_t$ le processus obtenu. Démontrer que $(N_t^{(2)})_t$ n'est pas un processus de Poisson.

8.8 ♠♠ Des bus arrivent à un arrêt selon un processus de Poisson d'intensité λ . De l'arrêt à votre domicile il vous faut t_1 minutes en bus et t_2 minutes à pied. Vous décidez d'attendre le bus s minutes et de partir à pied s'il n'est pas arrivé. Soit T le temps écoulé entre votre arrivée et celle du bus et W le temps mis pour rentrer.

$$(1) \text{ Vérifier que : } W = (T + t_1)\mathbf{1}_{[0,s]}(T) + (s + t_2)\mathbf{1}_{]s,+\infty[}(T).$$

(2) En déduire $E(W)$.

(3) Temps moyen pour rentrer chez vous ? Trouver la valeur de s qui minimise ce temps : considérer trois cas selon la valeur de λ par rapport à $\frac{1}{t_2 - t_1}$.

8.9 ♠♠ Nombre d'occurrences sur un intervalle de temps aléatoire

(1) Calculer la loi du nombre aléatoire X d'occurrences d'un processus de Poisson d'intensité λ , sur une durée U aléatoire de densité f_U . Application au cas où U est de loi exponentielle $\mathcal{E}(\alpha)$.

(2) Déterminer la fonction génératrice $G_X(z)$ de X , et en déduire $E(X)$ et $V(X)$.

Simulation et estimation de l'intensité d'un processus de Poisson

On simule un processus de Poisson d'intensité λ sur un intervalle de temps donné $[0, n]$ de la façon suivante : on répartit selon la loi uniforme sur l'intervalle $[0, n]$, $[\lambda \cdot n]$ (égal au plus grand entier inférieur à $\lambda \cdot n$) points et on note $T_i^{(n)}$ la position du i -ème point dans l'ordre croissant. Une propriété, non énoncée ici, établit que lorsque n tend vers l'infini, la suite $(T_i^{(n)})_i$ converge vers la suite des temps d'occurrences du processus de Poisson d'intensité λ . Supposons maintenant que l'on ait observé n occurrences du processus de Poisson au temps t ; l'estimation de l'intensité λ est donnée par la valeur du rapport $\frac{n}{t}$. On démontre dans le cadre de la théorie de l'estimation que $\frac{n}{t}$, dit *estimateur du maximum de vraisemblance de λ* , est un estimateur de λ sans biais, et de variance minimale. L'estimation de λ est d'autant meilleure que le temps d'observation t est grand. Si on dispose des temps d'occurrences $(T_i = t_i)_{i=1,2,\dots,n}$, il existe un estimateur de $\frac{1}{\lambda}$, égal à $\frac{T_n}{n}$, qui a des qualités équivalentes à l'estimateur de λ précédemment défini.

En pratique, il est souvent plus réaliste de supposer un taux moyen d'arrivée dépendant du temps. C'est par exemple le cas des temps d'arrivée de requêtes dans une unité de traitement informatique et de façon générale des temps d'arrivée dans un service, dont les fréquences sont fonction de la période de temps considérée. D'où la nécessité d'introduire la notion de processus de Poisson non homogène.

Définition 8.5 *Un processus de Poisson non homogène d'intensité $\lambda(t)$ fonction du temps est un processus de comptage non stationnaire et à accroissements indépendants vérifiant les deux axiomes :*

(1) $\forall t \ P(N(t + \Delta t) - N(t) = 1) = \lambda(t) + o(\Delta t)$ où $\lambda(t)$ est une fonction positive ;

(2) $\forall t \ P(N(t + \Delta t) - N(t) \geq 2) = o(\Delta t)$.

Si l'intensité $\lambda(t)$ est strictement positive, l'étude du processus homogène se ramène à l'étude d'un processus de Poisson classique, grâce au théorème suivant :

Théorème 8.7 Si $(N_t)_{t \geq 0}$ est un processus de Poisson non homogène, d'intensité $\lambda(t)$ strictement positive, alors :

$$(1) P(N(s+t) - N(s) = k) = e^{-(\Lambda(s+t) - \Lambda(s))} \frac{(\Lambda(s+t) - \Lambda(s))^k}{k!}.$$

(2) $(N_{\Lambda^{-1}(t)})_t$ est un processus de Poisson homogène d'intensité 1.

Preuve (2) La primitive Λ de λ définie par $\int_0^t \lambda(u) du$ est donc croissante et

$$P(N(\Lambda^{-1}(t+h)) - N(\Lambda^{-1}(t)) = k) = \frac{\Lambda(\Lambda^{-1}(t+h)) - \Lambda(\Lambda^{-1}(t))}{k!} e^{-h} = \frac{h^k}{k!} e^{-h}.$$

□

Définition 8.6 On appelle **processus de Weibull**, le processus de Poisson non homogène de paramètre $\Lambda(t) = (\frac{t}{\alpha})^\beta$ où $\alpha, \beta > 0$.

8.2 Processus de Poisson spatiaux

Les processus de Poisson spatiaux permettent de rendre compte des répartitions aléatoires uniformes de points dans des espaces non bornés de dimension deux ou trois. Ils permettent par exemple de modéliser la répartition des étoiles dans certaines régions de l'espace, de plantes et d'animaux d'une espèce donnée dans une zone donnée. On munit l'espace \mathbb{R}^k , de sa tribu borélienne \mathcal{B}_k et on se donne un ensemble infini de points aléatoirement répartis. À tout borélien B on associe la v.a. $N(B)$ égale au nombre de points contenus dans B .

Définition 8.7 L'ensemble des v.a. $N(B)$ où B est un borélien de \mathbb{R}^k est un **processus de Poisson spatial homogène** d'intensité λ , si sont vérifiés les deux axiomes suivants :

- (1) $\forall B, N(B)$ est de loi de Poisson $\mathcal{P}(\lambda |B|)$ où $|B|$ désigne l'aire ou le volume de B ;
- (2) si les boréliens $(B_i)_i$ sont disjoints, les v.a. $N(B_i)$ sont indépendantes.

On appelle **champ poissonnien** l'espace \mathbb{R}^k muni d'un processus spatial poissonnien.

L'axiome (1) est intuitif : en effet, soit un sous-ensemble borné E_n de \mathbb{R}^k contenant n points ; si on considère que la répartition aléatoire est uniforme, pour tout B appartenant à la tribu \mathcal{B}_{E_n} , $N(B)$ est une v.a. binomiale $\mathcal{B}(n; p)$, où $p = \frac{|B|}{|E_n|}$.

Faisons tendre l'ensemble E_n vers l'espace \mathbb{R}^k , de manière que :

$$\frac{n}{|E_n|} \longrightarrow_{(n \rightarrow +\infty)} \lambda.$$

On en déduit que :

$$p.n = \frac{|B|}{|E_n|} n \longrightarrow_{(n \rightarrow +\infty)} \lambda |B|,$$

qui est la condition de l'approximation de Poisson de la variable binomiale ; $N(B)$ est une v.a. de loi de Poisson $\mathcal{P}(\lambda |B|)$. L'axiome (2) est une généralisation de la propriété des accroissements indépendants.

Théorème 8.8 (*Propriétés des processus de Poisson spatiaux*)

(1) Si B_1 et B_2 sont de volumes égaux, $N(B_1)$ et $N(B_2)$ sont des v.a. de même loi.

(2) Pour tout borélien de volume infinitésimal ΔB :

$$P(N(\Delta B) = 1) \simeq \lambda |\Delta B| ;$$

$$P(N(\Delta B) > 1) \simeq o(|\Delta B|).$$

Preuve La propriété (1) découle du fait que la loi du processus dépend uniquement du volume de B et non de sa forme géométrique.

(2) $P(N(\Delta B) = 1) = e^{-\lambda |\Delta B|} \lambda |\Delta B| \simeq (1 - \lambda |\Delta B|) \lambda |\Delta B| \simeq \lambda |\Delta B|$;

$$P(N(\Delta B) > 1) \simeq P(1 - e^{-\lambda |\Delta B|} = 0) - \lambda |\Delta B| \simeq o(|\Delta B|). \quad \square$$

On remarque l'analogie existant entre la propriété (1) et la propriété de stationnarité des processus de Poisson homogènes ; la propriété (2) est la propriété des événements rares étendue à un espace de dimension quelconque.

8.10 ♠ (1) Démontrer que si $B_1 \subset B_2$: $P(N(B_1) = 1 \mid N(B_2) = 1) = \frac{|B_1|}{|B_2|}$.

(2) Démontrer que la v.a. $(N(B_1) \mid (N(B_2) = n))$ est binomiale $\mathcal{B}(n; \frac{|B_1|}{|B_2|})$.

Théorème 8.9 Étant donné un processus de Poisson d'intensité λ dans le plan et un point quelconque du plan : la distance aléatoire D de ce point au point le plus proche est de densité :

$$f_D(x) = 2\lambda \pi x e^{-\lambda \pi x^2}$$

et d'espérance : $E(D) = \frac{1}{\sqrt{\lambda}}$.

8.11 ♠ (1) Démontrer le théorème. [Indication : déterminer $P(D > x)$.]

(2) Soit D_n la distance au n -ième plus proche voisin ; déterminer f_{D_n} .

8.3 Processus de Poisson homogènes composés

Ces processus sont à la base de l'étude des événements de Poisson qui se produisent de façon groupée comme les arrivées simultanées de voyageurs dans un aéroport ou une gare, ou encore l'arrivée d'informations groupées arrivant dans une unité centrale informatique.

Définition 8.8 *Étant donné $(N_t)_t$ un processus de Poisson homogène d'intensité λ , et $(Y_i)_{i \in \mathbb{N}^*}$ une suite de v.a. discrètes indépendantes de même loi P_Y , on définit le **processus de Poisson composé** $(W_t)_{t \in \mathbb{R}_+}$:*

$$\forall t \geq 0, W_t = \sum_{i=1}^{N_t} Y_i \text{ et } W_t = 0 \text{ si } N_t = 0.$$

Supposons que des groupes de requêtes arrivent dans une unité de traitement informatique selon un processus de Poisson d'intensité λ ; le n -ième groupe contient Y_n requêtes. Si N_t désigne le nombre de groupes arrivés entre le temps 0 et le temps t , le processus $(W_t = \sum_{i=1}^{N_t} Y_i)_t$ est un processus de Poisson composé. L'étude de ce processus nécessite la connaissance de la loi de $W = \sum_{i=1}^N Y_i$ où N est une v.a. de Poisson de paramètre λ . Les théorèmes ci-dessous permettent de calculer les moments d'ordre quelconque de W , ainsi que la loi de W lorsque les Y_i sont des variables entières positives.

Théorème 8.10 *Étant donné un processus de Poisson composé $(W_t)_{t \in \mathbb{R}_+}$, pour tout temps t positif :*

$$\begin{aligned} E(W_t) &= \lambda t E(Y_1) ; \\ V(W_t) &= \lambda t E(Y_1^2). \end{aligned}$$

Preuve $E(W_t) = E(N_t)E(Y_1) = \lambda t E(Y_1)$

$$V(W_t) = (\text{Théo de la variance totale}) E(N_t) V(Y_1) + V(N_t) E(Y_1^2) = \lambda t E(Y_1^2). \quad \square$$

Théorème 8.11 *La fonction génératrice G_{W_t} du processus de Poisson composé $(W_t)_{t \geq 0}$ est égale à :*

$$G_{W_t}(z) = \exp(\lambda t (G_Y(z) - 1)).$$

Preuve $G_{W_t}(z) = \sum_n P(N_t = n) \cdot E(z^{Y_1} \dots z^{Y_n}) = e^{-\lambda t} \sum_{n=0}^{+\infty} (G_Y(z) \lambda t)^n / n! = \exp(\lambda t (G_Y(z) - 1)).$ \square

On rappelle que la connaissance de G_{W_t} permet de déterminer la loi W_t en tout t :

$$\forall k \in \mathbb{N}, P_{W_t}(k) = \frac{1}{k!} \frac{d^k \exp(\lambda t (G_Y(z) - 1))}{dz^k} \Big|_{z=1}.$$

8.4 Processus de renouvellement

Les processus de renouvellement constituent une extension des processus de Poisson au cas où la loi des durées $U_n = T_n - T_{n-1}$ entre deux occurrences consécutives n'est pas nécessairement exponentielle. Ils permettent de modéliser et de résoudre les problèmes de renouvellement des systèmes, dans le cadre d'une politique de maintenance.

8.4.1 Description et propriétés

Définition 8.9 *Étant donné $(T_n)_{n \in \mathbb{N}^*}$ la suite des temps d'occurrence d'un événement donné, où pour tout n , $T_n = \sum_{i=1}^n U_i$; les v.a. U_i sont positives et indépendantes, de même loi définie par leur fonction de répartition commune F . On définit le **processus de renouvellement** $(N_t)_{t \in \mathbb{R}^+}$:*

$$\forall t \geq 0, N_t = \max \{n | T_n \leq t\} = \sum_{n \geq 0} \mathbf{1}_{[0, t]}(T_n).$$

N_t est donc le nombre d'occurrences de l'événement dans l'intervalle $[0, t]$. Si les U_i sont de même loi exponentielle $\mathcal{E}(\lambda)$, le processus de renouvellement est alors un processus de Poisson d'intensité λ .

Définition 8.10 *La **fonction de renouvellement**, qui permet de suivre l'évolution temporelle de la moyenne du processus, est définie par :*

$$t \in \mathbb{R}_+ \longrightarrow E(N(t)) =_{(noté)} m(t).$$

Théorème 8.12 (Équation de renouvellement)

(1) *Les fonctions de répartition F_n des v.a. T_n vérifient l'équation :*

$$F_{n+1}(x) = \int_0^x F_n(x-y) dF_{U_{n+1}}(y).$$

(2) *La fonction de renouvellement m vérifie l'équation de renouvellement :*

$$\forall t > 0, m(t) = F(t) + \int_0^t m(t-x) dF(x),$$

égale à $\sum_{n=1}^{+\infty} F_n(t)$.

Preuve (1) $T_{n+1} = T_n + U_{n+1}$, donc :

$$F_{n+1}(x) = \int_0^x P(T_n + U_{n+1} \leq x | U_{n+1} = y) dF_{U_{n+1}}(y).$$

(2) Par définition $N_t = \sum_{n=1}^{+\infty} \mathbf{1}_{(T_n \leq t)}$ d'où : $E(N_t) = \sum_{n=1}^{+\infty} P(T_n \leq t) = \sum_{n=1}^{+\infty} F_n(t)$.

D'autre part : $E(N_t | T_1 = x) = 0 \times \mathbf{1}_{\{t < x\}} + (1 + E(N_{t-x})) \mathbf{1}_{\{t \geq x\}}$;

$$\begin{aligned} m(t) &= E(N_t) = E(E(N_t | T_1)) \\ &= \int_0^t (1 + m(t-x)) dF(x) = F(t) + \int_0^t m(t-x) dF(x). \end{aligned}$$

□

Théorème 8.13 *La transformée de Laplace de la fonction de renouvellement m est définie par :*

$$\forall s > 0, \quad L(m)(s) = \int_0^{+\infty} e^{-st} \cdot E(N_t) dt = \frac{L(F)(s)}{1 - L(f)(s)} = \frac{L(f)(s)}{s(1 - L(f)(s))}.$$

Preuve $\forall s > 0, L(m)(s) = \sum_{n=1}^{+\infty} L(F_n)(s) = \sum_{n=1}^{+\infty} L(F)(s)(L(f)(s))^{n-1} = \frac{L(F)(s)}{1 - L(f)(s)}$,
puisque $L(F_n)(s) = L(F_{n-1})(s)L(f)(s)$. □

8.12 ♠♠ Étude d'un processus de défaillance

Un système à réponse instantanée est sollicité à des instants aléatoires $(T_n)_n$ constituant un processus de renouvellement de fonction de répartition F . À chaque sollicitation il a la probabilité p d'être défaillant, les défaillances étant indépendantes des sollicitations.

(1) La suite des instants de défaillance est-elle un processus de renouvellement ?

(2) Utiliser la transformation de Laplace pour déterminer la fonction de renouvellement.

(3) Étudier le cas où f est la densité exponentielle $\mathcal{E}(\lambda)$.

8.4.2 Comportement asymptotique d'un processus de renouvellement

Théorème 8.14 (*Propriétés asymptotiques des processus de renouvellement*)

Si $(N_t)_t$ est un processus de renouvellement dont les temps d'interarrivées sont de moyenne $\frac{1}{\mu}$ et de variance σ^2 , alors :

(1) la suite $(\frac{N_t}{t})_{t>0}$ converge p.s. vers $\frac{1}{\mu}$, quand t tend vers $+\infty$;

(2) la suite $(N_t)_{t>0}$ converge en loi vers la v.a. $\mathcal{N}(\frac{t}{\mu}; \frac{\sigma^2 t}{\mu^3})$.

Preuve (1) Soit $T_{N_t} = \sum_{i=1}^{N_t} U_i$; par définition $T_{N_t} < t < T_{N_t+1}$, et donc $\frac{T_{N_t}}{N_t} < \frac{t}{N_t} < \frac{T_{N_t+1}}{N_t}$. Or $\frac{T_{N_t}}{N_t}$ est la moyenne des durées des N_t premiers intervalles entre arrivées consécutives, qui converge p.s. vers $\frac{1}{\mu}$ quand t croît vers l'infini.

(2) Posons $a_t = \frac{t}{\mu} + x\sigma\sqrt{\frac{t}{\mu^3}}$; il vient

$$\begin{aligned} P\left(\frac{N_t - \frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}} < x\right) &= P(N_t < a_t) = P(T_{a_t} > t) \\ &= P\left(\frac{T_{a_t} - a_t\mu}{\sigma\sqrt{a_t}} > -x\left(1 + \frac{x\sigma}{\sqrt{t\mu}}\right)^{-\frac{1}{2}}\right). \end{aligned}$$

La v.a. $\frac{T_{a_t} - a_t\mu}{\sigma\sqrt{a_t}}$ converge vers $\mathcal{N}(0; 1)$ par l'application du théorème de la limite centrale et $-x(1 + \frac{x\sigma}{\sqrt{t\mu}})^{-\frac{1}{2}}$ converge vers $-x$ quand $t \rightarrow +\infty$, d'où la proposition. \square

Si $(N_t)_t$ est un processus de Poisson d'intensité λ , alors $\mu = \sigma^2 = \frac{1}{\lambda}$ et $(\frac{N_t}{t})_t$ converge en loi quand n tend vers $+\infty$, vers la v.a. normale de moyenne et de variance égales à λ . Ce résultat permet de donner une estimation par intervalle de confiance, de la valeur de N_t lorsque t est grand : par exemple pour $t = 100$, la v.a. N_{100} est approchée par la v.a. $\mathcal{N}(m = 100, \sigma^2 = 100)$.

Le graphe de la réalisation d'un processus de Poisson quelconque d'intensité λ , s'enlace de façon aléatoire autour de la droite d'équation $y = \lambda t$: l'écart quadratique moyen avec la droite $y = \lambda t$ est mesuré par la variance égale à $\lambda^2 t$.

Le problème 17 se rapporte à ce chapitre.

8.5 Corrigés des exercices

Exercice 8.1

$$(1) P(N_u = k \mid N_s = i, N_t = j) = \frac{P(N_u = k, N_s = i, N_t = j)}{P(N_s = i, N_t = j)} \text{ où :}$$

$$\begin{aligned} P(N_s = i, N_t = j) &= P(N_s = i, N_t - N_s = j - i) \\ &= P(N_s = i)P(N_t - N_s = j - i) \\ &= P(N_s = i)P(N_{t-s} = j - i) \\ &= e^{-\lambda s} \frac{(\lambda s)^i}{i!} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^{j-i}}{(j-i)!} \\ &= e^{-\lambda t} \frac{\lambda^j s^i (t-s)^{j-i}}{i!(j-i)!}. \end{aligned}$$

a) Cas $s \leq t \leq u$. Pour $i \leq j \leq k$:

$$\begin{aligned} P(N_u = k, N_s = i, N_t = j) &= P(N_s = i, N_t - N_s = j - i, N_u - N_t = k - j) \\ &= P(N_s = i)P(N_t - N_s = j - i)P(N_u - N_t = k - j) \\ &= P(N_s = i)P(N_{t-s} = j - i)P(N_{u-t} = k - j). \end{aligned}$$

$P(N_u = k \mid N_s = i, N_t = j) = P(N_{u-t} = k - j) = e^{-\lambda(u-t)} \frac{(\lambda(u-t))^{k-j}}{(k-j)!}$, qui est donc indépendant de l'événement $(N_s = i)$ antérieur aux événements $(N_t = j)$ et $(N_u = k)$.

b) Cas $s \leq u \leq t$. Pour $i \leq k \leq j$:

$$\begin{aligned} P(N_u = k, N_s = i, N_t = j) &= P(N_s = i, N_u - N_s = k - i, N_t - N_u = j - k) \\ &= P(N_s = i)P(N_u - N_s = k - i)P(N_t - N_u = j - k) \\ &= P(N_s = i)P(N_{u-s} = k - i)P(N_{t-u} = j - k). \end{aligned}$$

On a donc

$$\begin{aligned} P(N_u = k \mid N_s = i, N_t = j) &= \frac{P(N_{u-s} = k - i)P(N_{t-u} = j - k)}{P(N_{t-s} = j - i)} \\ &= \frac{e^{-\lambda(u-s)} \frac{(\lambda(u-s))^{k-i}}{(k-i)!} e^{-\lambda(t-u)} \frac{(\lambda(t-u))^{j-k}}{(j-k)!}}{e^{-\lambda(t-s)} \frac{(\lambda(t-s))^{j-i}}{(j-i)!}}. \end{aligned}$$

$$P([N_u = k \mid N_s = i, N_t = j]) = C_{j-i}^{k-i} \frac{(u-s)^{k-i} (t-u)^{j-k}}{(t-s)^{j-i}}.$$

c) Cas $u \leq s \leq t$. Pour $k \leq i \leq j$:

$$\begin{aligned}
P(N_u = k, N_s = i, N_t = j) &= P(N_u = k, N_s - N_u = i - k, N_t - N_s = j - i) \\
&= P(N_u = k)P(N_s - N_u = i - k)P(N_t - N_s = j - i) \\
&= P(N_u = k)P(N_{s-u} = i - k)P(N_{t-s} = j - i).
\end{aligned}$$

On a donc

$$\begin{aligned}
P(N_u = k, N_s = i, N_t = j) &= \frac{P(N_u = k)P(N_{s-u} = i - k)}{P(N_s = i)} \\
&= \frac{e^{-\lambda u} \frac{(\lambda u)^k}{k!} e^{-\lambda(s-u)} \frac{(\lambda(s-u))^{i-k}}{(i-k)!}}{e^{-\lambda s} \frac{(\lambda s)^i}{i!}}.
\end{aligned}$$

$$P(N_u = k, N_s = i, N_t = j) = C_i^k \frac{u^k (s-u)^{i-k}}{s^i},$$

$$\text{soit } P(N_u = k \mid N_s = i, N_t = j) = C_i^k \frac{u^k (s-u)^{i-k}}{s^i}.$$

(2) *Application* : $u = 1/2$, $k = 1$ et $s = 1/4$, $t = 1$, $i = 0$ et $j = 2$.

On est dans le cas b) ; $u - s = \frac{1}{4}$, $t - u = \frac{1}{2}$, $t - s = \frac{3}{4}$, $k - i = 1$, $j - k = 1$, $j - i = 2$ et la valeur de λ n'a pas d'importance. $p = 2 \times \frac{1}{4} \frac{1}{2} = \frac{1}{4}$, soit $p_1 = \frac{4}{9} \approx 0,44$.

Exercice 8.2

(1) Pour tout n : $B_n = \phi(B_{n-1}, X_n) = B_{n-1} + X_n$ où X_n indépendante des $(B_i)_{i \leq n-1}$; cette propriété caractérise une chaîne de Markov (voir chapitre 7.2).

(2) Pour tout n , $P([U_n = k]) = (1-p)^{k-1}p$: U_n est de loi géométrique.

Démonstration par récurrence de la loi de T_n :

$$\begin{aligned}
P(T_2 = i) &= \sum_{k=1}^{i-1} P(U_1 = k)P(U_2 = i - k) \\
&= \sum_{k=1}^{i-1} (1-p)^{k-1}p(1-p)^{i-k-1}p = p^2 \sum_{k=1}^{i-1} (1-p)^{i-2} \\
&= (i-2)p^2(1-p)^{i-2} = \binom{i-1}{1} p^2 (1-p)^{i-2}.
\end{aligned}$$

On poursuit par une démonstration par récurrence fondée sur l'égalité :

$$P(T_n = k) = \sum_{i=n}^{k-1} P(T_n = k)P(T_{n-1} = i).$$

Exercice 8.3

(1) $F_{T_n}(t) = P([T_n \leq t]) = P([N_t \geq n]) = 1 - P([N_t \leq n-1])$ pour $n \geq 1$ et

$T_0 = 0 : F_{T_n}(t) = 1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!}$ pour $n \geq 1$ et $t > 0$. On a alors :

$$\begin{aligned} f_{T_n}(t) = F'_{T_n}(t) &= \lambda e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} - e^{-\lambda t} \sum_{k=1}^{n-1} \frac{\lambda^k k t^{k-1}}{k!} \\ &= e^{-\lambda t} \left[\sum_{k=0}^{n-1} \frac{\lambda^{k+1} t^k}{k!} - \sum_{k=1}^{n-1} \frac{\lambda^k t^{k-1}}{(k-1)!} \right] \\ &= e^{-\lambda t} \left[\sum_{k=0}^{n-1} \frac{\lambda^{k+1} t^k}{k!} - \sum_{k=0}^{n-2} \frac{\lambda^{k+1} t^k}{k!} \right]. \end{aligned}$$

et finalement $f_{T_n}(t) = \frac{\lambda^n}{(n-1)!} e^{-\lambda t} t^{n-1} \mathbf{1}_{[0, +\infty[}(t)$ qui est la densité de la loi gamma $\gamma(\lambda, n)$.

(2) $\mathbb{E}(N_s N_t) = \mathbb{E}(N_s(N_t - N_s) + N_s^2) = \mathbb{E}(N_s) \mathbb{E}(N_t - N_s) + \mathbb{E}(N_s^2)$, car N_s et $N_t - N_s$ sont indépendantes. De plus, $N_t - N_s$ a même loi que N_{t-s} et si X suit la loi de Poisson $\mathcal{P}(\lambda)$, alors $\mathbb{E}(X) = \text{var}(X) = \lambda$. Comme ici N_t est de loi $\mathcal{P}(\lambda t)$: N_s et $N_t - N_s$ étant indépendantes donc décorrélées :

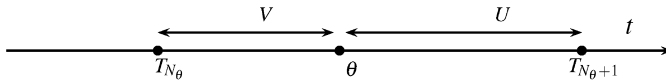
$$\text{cov}(N_s, N_t) = \text{cov}(N_s, N_t - N_s + N_s) = \text{cov}(N_s, N_t - N_s) + \text{var}(N_s) = \text{var}(N_s) = \lambda s.$$

$$\rho(N_s, N_t) = \frac{\text{cov}(N_s, N_t)}{\sqrt{\text{var} N_s} \sqrt{\text{var} N_t}} = \frac{\lambda s}{\lambda \sqrt{st}} = \sqrt{\frac{s}{t}}.$$

En conséquence, plus s est proche de t , plus N_s et N_t sont fortement corrélées, ce qui semble évident.

Exercice 8.4

(1) Loi de $V : \forall x > \theta, P(V > x) = 0$.



$\forall x \in [0, \theta[$, l'événement $[V > x]$ signifie que pendant la durée x précédant θ , il n'y a aucune arrivée :

$$P(V > x) = P(N_\theta - N_{\theta-x} = 0) = P(N_x = 0) = e^{-\lambda x}.$$

Si $x > 0$, $P(V > x) = 0$: d'où la loi de V définie par sa fonction de répartition :

$$F_V(x) = (1 - e^{-\lambda x}) \mathbf{1}_{[0, \theta[}(x) + \mathbf{1}_{[\theta, +\infty[}(x).$$

Loi de U : $P([U > x]) = P([N_{\theta+x} - N_{\theta} = 0]) = P([N_x = 0]) = e^{-\lambda x}$, car $[U > x]$. Pendant la durée x qui suit θ , il n'y a aucune arrivée.

(2) Indépendance de U et V .

Si $0 < x < \theta$ et $y > 0$: $[V > x] \cap [U > y]$ signifie qu'il n'y a aucune arrivée entre $\theta - x$ et $\theta + y$, donc :

$$P([V > x] \cap [U > y]) = e^{-\lambda(x+y)} = e^{-\lambda x} e^{-\lambda y} = P([V > x])P([U > y]).$$

$$\text{Si } x \geq \theta \text{ et } y > 0 : P([V > x] \cap [U > y]) = 0 = P([V > x])P([U > y]).$$

$$\mathbb{E}(U) = \frac{1}{\lambda}; \mathbb{E}(V) = \int_0^{\theta} P([V > s]) ds = \frac{1}{\lambda}(1 - e^{-\lambda\theta}) \text{ donc :}$$

$$\mathbb{E}(U+V) = \frac{2}{\lambda} - \frac{1}{\lambda} e^{-\lambda\theta} > \frac{1}{\lambda} = \mathbb{E}(T).$$

Le paradoxe réside dans le constat que l'écart moyen entre les deux passages consécutifs de bus, l'un avant θ et l'autre après θ , est supérieur à la durée moyenne séparant deux passages consécutifs quelconques.

Exercice 8.5

Le piéton arrive au bord de la route à un instant θ . On note U la v.a.r. qui sépare θ de la prochaine arrivée de voiture.

(1) Le piéton ayant besoin d'au moins 4 secondes pour traverser, il devra attendre si $U \leq 4$.

Or $P([U \leq 4]) = \int_0^4 \lambda e^{-\lambda u} du = [-e^{-\lambda u}]_0^4 = 1 - e^{-4\lambda}$, car on a vu à l'exercice 8.4 que U suit la loi exponentielle $\mathcal{E}(\lambda)$. Comme $\lambda = 1/6$, $P([U \leq 4]) = 1 - e^{-4/6}$ et la probabilité pour qu'il doive attendre est égale à $1 - e^{-2/3} \approx 0,487$.

(2) On cherche $\mathbb{E}^{[U>4]}(U)$, c'est-à-dire $\int u f_U^{[U>4]}(u) du$. On a

$$\begin{aligned} F_U^{[U>4]}(u) &= P^{[U>4]}([U \leq u]) = \frac{P([U \leq u] \cap [U > 4])}{P([U > 4])} \\ &= \begin{cases} 0 & \text{si } u \leq 4 \\ \frac{P([4 < U \leq u])}{P([U > 4])} = \frac{F_U(u) - F_U(4)}{P([U > 4])} & \text{si } u > 4 \end{cases} \end{aligned}$$

$$\text{et donc } f_U^{[U>4]}(u) = \frac{f_U(u)}{P([U > 4])} \mathbf{1}_{]4, +\infty[}(u) = \lambda e^{-\lambda(u-4)} \mathbf{1}_{]4, +\infty[}(u).$$

$$\begin{aligned} \mathbb{E}^{[U>4]}(U) &= \int_4^{+\infty} \lambda u e^{-\lambda(u-4)} du = \int_{v=u-4}^{+\infty} \lambda(v+4) e^{-\lambda v} dv \\ &= \int_0^{+\infty} v \lambda e^{-\lambda v} dv + 4 \int_0^{+\infty} \lambda e^{-\lambda v} dv = \frac{1}{\lambda} + 4, \end{aligned}$$

car $\int_0^{+\infty} v\lambda e^{-\lambda v} dv$ est l'espérance d'une v.a. de loi exponentielle $\mathcal{E}(\lambda)$.

Ainsi, il faut en moyenne 10 secondes pour que le piéton puisse traverser.

(3) Soit N le nombre de voitures que voit passer le piéton avant de pouvoir traverser la route. On a :

- $[N = 0] = [U > 4]$, où U est le temps entre l'arrivée du piéton et la prochaine voiture ;

- $[N = 1] = [U \leq 4] \cap [U_1 > 4]$ où U_1 est le temps entre la première et la deuxième voiture qui se présentent après l'arrivée du piéton ;

- plus généralement, $[N = k] = [U \leq 4] \cap [U_1 \leq 4] \cdots \cap [U_{k-1} \leq 4] \cap [U_k > 4]$ où U_k est le temps entre la k -ième et la $(k+1)$ -ième voiture qui se présentent après l'arrivée du piéton.

Les v.a.r. U, U_1, \dots, U_k étant indépendantes et de même loi exponentielle $\mathcal{E}(\lambda)$,

on a $P([N = 0]) = P([U > 4]) = e^{-4\lambda}$ et

$$\begin{aligned} P([N = k]) &= P([U \leq 4])P([U_1 \leq 4]) \cdots P([U_{k-1} \leq 4])P([U_k > 4]) \\ &= e^{-4\lambda}(1 - e^{-4\lambda})^k = e^{-2/3}(1 - e^{-2/3})^k. \end{aligned}$$

On reconnaît la loi géométrique $\mathcal{G}_0(e^{-2/3})$ définie sur \mathbb{N} définie par : $P(X = k) = p(1 - p)^k$.

$$E(N) = e^{2/3} - 1 \approx 0,948.$$

Le piéton voit donc passer en moyenne une voiture avant de traverser.

Exercice 8.6

(1) Montrons que $(N_t)_t = (N_t^{(1)} + N_t^{(2)})_t$ est un processus de Poisson.

$N_0 = N_0^{(1)} + N_0^{(2)} = 0$; $N_{t_k} - N_{t_{k-1}} = (N_{t_k}^{(1)} - N_{t_{k-1}}^{(1)}) + (N_{t_k}^{(2)} - N_{t_{k-1}}^{(2)})$: les accroissements sont donc indépendants.

Soit $T_1^{(1)}$ et $T_1^{(2)}$ les temps de réalisation de l'événement pour chacun des processus : le temps de réalisation pour le processus somme est égal à $T_1 = \min(T_1^{(1)}, T_1^{(2)})$. Or les v.a. $T_1^{(1)}$ et $T_1^{(2)}$ sont indépendantes de lois $\mathcal{E}(\lambda_1)$ et $\mathcal{E}(\lambda_2)$, donc T_1 est de loi $\mathcal{E}(\lambda_1 + \lambda_2)$. On procèdera de façon analogue pour les temps d'interarrivées quelconques, le théorème de caractérisation 8.5 permettant de conclure.

(2) Soit N'_t le résultat du comptage prenant en compte l'effacement ;

$$P(N'_t = k) = \sum_{i \geq k} P(N'_t = k | N_t = i) P(N_t = i) = \sum_{i \geq k} C_i^k p^k (1-p)^{i-k} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda p} \frac{(\lambda p)^k}{k!}.$$

Montrons que les accroissements sont indépendants. Soit $u < s < t$:

$P(N'_t - N'_s = i \text{ et } N'_s - N'_u = j)$ est égal à : $P((\bigcup_{k \geq i} (k \text{ occurrences et } (k-i) \text{ effacées entre s et t})) \cap (\bigcup_{l \geq j} (l \text{ occurrences et } (l-j) \text{ effacées entre u et s})))$
 $= P(N'_t - N'_s = i)P(N'_s - N'_u = j)$ par indépendance des émissions et des effacements sur $[s, t]$ et $[u, s]$. Le processus vérifie bien les axiomes de la définition 8.2.

Exercice 8.7

(1) $(\forall t), (t < \alpha), P(N_t^{(1)} = 0) = 1$ et $P(N_t^{(1)} = k) = 0, \forall k \geq 1$.

$(\forall t), (t \geq \alpha), P(N_t^{(1)} = k) = P(N_{t-\alpha} = k) = e^{-\lambda(t-\alpha)} \frac{(\lambda(t-\alpha))^k}{k!}$.

L'indépendance des accroissements est évidente : le processus est un processus de Poisson.

(2) Soient $(T_i)_i$ les temps d'arrivée du processus $(N_t)_t$ et $T'_i = T_i + X_i$ où X_i est la v.a. aléatoire associée à la i -ème translation de temps. Il est clair que

$$T'_{i+1} - T'_i = (T_{i+1} - T_i) + (X_{i+1} - X_i),$$

avec $T_{i+1} - T_i$ de loi exponentielle $\mathcal{E}(\lambda)$ et $X_{i+1} - X_i$ de loi non exponentielle : $T'_{i+1} - T'_i$ est de loi non exponentielle et $(N_t^{(2)})_t$ n'est pas un processus de Poisson.

Exercice 8.8 (1) Soit T le temps (aléatoire) écoulé entre notre arrivée et celle du prochain bus. On sait que T suit la loi exponentielle $\mathcal{E}(\lambda)$. Le temps W mis pour rentrer est :

- $T + t_1$ si le bus arrive avant s , c'est-à-dire si $T \leq s$;
- $s + t_2$ si on rentre à pied, c'est-à-dire si $T > s$.

On a donc $W = (T + t_1)\mathbf{1}_{[0, s]}(T) + (s + t_2)\mathbf{1}_{]s, +\infty[}(T) = h(T)$.

$$\begin{aligned} (2) \quad \mathbb{E}(W) &= \int h(t) f_T(t) dt = \int_0^s (t + t_1) \lambda e^{-\lambda t} dt + \int_s^{+\infty} (s + t_2) \lambda e^{-\lambda t} dt \\ &= \left[-te^{-\lambda t} \right]_0^s + \int_0^s e^{-\lambda t} dt + t_1 \left[-e^{-\lambda t} \right]_0^s + (s + t_2) \left[-e^{-\lambda t} \right]_s^{+\infty} \\ &= -se^{-\lambda s} + \frac{1}{\lambda} - \frac{1}{\lambda} e^{-\lambda s} + t_1(1 - e^{-\lambda s}) + se^{-\lambda s} + t_2 e^{-\lambda s} \end{aligned}$$

$$\mathbb{E}(W) = \left(\frac{1}{\lambda} + t_1 \right) + e^{-\lambda s} \left(t_2 - \left(\frac{1}{\lambda} + t_1 \right) \right).$$

(3) $s \mapsto e^{-\lambda s}$ décroît de 1 (pour $s = 0$) à 0 (pour $s \rightarrow +\infty$) donc :

- si $\lambda > \frac{1}{t_2 - t_1}$, c'est-à-dire si $\frac{1}{\lambda} + t_1 < t_2$, le minimum est obtenu lorsque $s \rightarrow +\infty$, et vaut $\frac{1}{\lambda} + t_1$: on a donc intérêt à attendre le bus jusqu'à ce qu'il arrive ;

- si $\lambda < \frac{1}{t_2 - t_1}$, c'est-à-dire si $\frac{1}{\lambda} + t_1 > t_2$, le minimum est obtenu lorsque $s = 0$, et vaut t_2 : on n'a donc pas intérêt à prendre le bus (il n'y en a pas assez souvent) ;

- si $\lambda = \frac{1}{t_2 - t_1}$, c'est-à-dire si $\frac{1}{\lambda} + t_1 = t_2$, $\mathbb{E}(W)$ est indépendant de s et quoi qu'on décide, le temps moyen est égal à t_2 .

Exercice 8.9

$$(1) P(X = k) = \int_0^{+\infty} P(X = k \mid U = u) f_U(u) du = \int_0^{+\infty} \frac{e^{-\lambda u} (\lambda u)^k}{k!} f_U(u) du.$$

Si U est de loi exponentielle de paramètre α ,

$$\begin{aligned} P([X = k]) &= \int_0^{+\infty} \frac{e^{-\lambda u} (\lambda u)^k}{k!} \alpha e^{-\alpha u} du = \frac{\lambda^k}{k!} \alpha \int_0^{+\infty} e^{-(\alpha + \lambda)u} u^k du \\ &= \frac{\lambda^k}{k!} \alpha \frac{1}{(\alpha + \lambda)^{k+1}} \Gamma(k+1) = \left(\frac{\lambda}{\alpha + \lambda} \right)^k \left(\frac{\alpha}{\alpha + \lambda} \right). \end{aligned}$$

Conclusion : X est une v.a. de loi géométrique $\mathcal{G}_0\left(\frac{\alpha}{\alpha + \lambda}\right)$.

$$(2) G_X(s) = \sum_{k=0}^{+\infty} s^k P([X = k]) = \sum_{k=0}^{+\infty} \frac{e^{-\lambda u} (\lambda u s)^k}{k!} f_U(u) du = \int_0^{+\infty} e^{\lambda u(s-1)} f_U(u) du.$$

Ainsi, $G'_X(s) = \int_0^{+\infty} \lambda u e^{\lambda u(s-1)} f_U(u) du$ et $G''_X(s) = \int_0^{+\infty} \lambda^2 u^2 e^{\lambda u(s-1)} f_U(u) du$;

pour $s = 1$, on obtient $\mathbb{E}(X) = G'_X(1) = \lambda \mathbb{E}(U)$ et $G''_X(1) = \lambda^2 \mathbb{E}(U^2)$.

On en déduit : $\text{var}(X) = G''_X(1) + g'_X(1) - G'_X(1)^2$, soit $\text{var}(X) = \lambda^2 \text{var}(U) + \lambda \mathbb{E}(U)$.

Exercice 8.10

$$\begin{aligned} (1) P([N(B_1) = 1] \mid [N(B_2) = 1]) &= \frac{P([N(B_1) = 1] \cap [N(B_2 \setminus B_1) = 0])}{P([N(B_2) = 1])} \\ &= \frac{e^{-\lambda|B_1|} \lambda |B_1| e^{-\lambda|B_2 \setminus B_1|}}{e^{-\lambda|B_2|} \lambda |B_2|} = \frac{e^{-\lambda|B_2|} \lambda |B_1|}{e^{-\lambda|B_2|} \lambda |B_2|} = \frac{|B_1|}{|B_2|}. \end{aligned}$$

(2) Conséquence triviale de (1).

Exercice 8.11

(1) La probabilité qu'aucun point du champ poissonnien n'appartienne au disque de rayon x est :

$$P(D > x) = \frac{(\lambda \pi x^2)^0}{0!} e^{-\lambda \pi x^2} = e^{-\lambda \pi x^2}$$

d'où : $F_D(x) = 1 - e^{-\lambda\pi x^2}$ et $f_D(x) = 2\lambda\pi x e^{-\lambda\pi x^2}$,

(2) Soit maintenant : D_n la distance au n -ième plus proche voisin

$$F_{D_n}(x) = P(D_n \leq x) = e^{-\lambda\pi x^2} \sum_{k=n}^{+\infty} \frac{(\lambda\pi x^2)^k}{k!},$$

d'où : $f_{D_n}(x) = 2\lambda\pi x \frac{(\lambda\pi x^2)^{n-1}}{(n-1)!} e^{-\lambda\pi x^2}$.

Exercice 8.12

(1) Soit Y_1 le premier instant de défaillance : $Y_1 = \sum_{i=1}^N X_i$ où N est de loi géométrique $\mathcal{G}(p)$. Par les hypothèses d'indépendance, le processus de défaillance est un processus de renouvellement de loi identique à celle de Y_1 ;

$$L(f_Y)(s) = \sum_{n=1}^{+\infty} P(N=n) (L(f_X)(s))^n = \frac{pL(f_X)(s)}{1 - qL(f_X)(s)}.$$

$L(f_Y)$ est la transformée de Laplace de f_Y .

(2) La fonction de renouvellement M^* du processus de défaillance admet la transformée de Laplace :

$$L(M^*)(s) = \frac{L(f_Y)(s)}{s(1 - L(f_Y)(s))} = \frac{1}{s} \frac{pL(f_X)(s)}{(1 - L(f_X)(s))} = pL(M)(s).$$

(3) Si les $(X_i)_i$ sont de loi $\mathcal{E}(\lambda)$:

$$L(f_Y)(s) = \frac{p\lambda}{p\lambda + s}.$$

Y est donc de loi exponentielle $\mathcal{E}(p\lambda)$.

Chapitre 9

Chaînes de Markov à temps continu et files d'attente

La notion de chaîne de Markov à temps continu, dénommée aussi processus à sauts markoviens si l'espace d'états est discret, prolonge à la fois la notion de chaîne de Markov à temps entier et la notion de processus de Poisson par l'enrichissement de son espace d'états. Ce chapitre est une introduction à l'étude des systèmes d'attente mis en œuvre dans la modélisation des phénomènes d'attente apparaissant à l'entrée des services tels que les centraux téléphoniques, les réseaux de trafic urbain et aérien et, de façon générale, dans les réseaux de communication. D'un point de vue pratique, une file d'attente apparaît chaque fois que des clients se présentent en des temps aléatoires à une station comportant un ou plusieurs serveurs pour y recevoir un service de durée généralement (mais pas nécessairement) aléatoire de loi connue. Les premiers travaux sur les files d'attente sont dûs au mathématicien danois Erlang (1878-1929), alors qu'il s'intéressait à la modélisation du réseau téléphonique de Copenhague, afin d'en réduire les phénomènes d'attente. La théorie mathématique des files d'attente orientée aujourd'hui vers l'étude des réseaux informatiques et de télécommunication s'est développée par la suite grâce notamment aux travaux de Palm, de Kolmogorov et de Khintchine.

9.1 Chaînes de Markov à temps continu

9.1.1 Définition et propriétés

Définition 9.1 *Le processus aléatoire $(X_t)_{t \geq 0}$ d'espace d'états $E = \{e_i\}_{i \in \mathbb{I}}$, fini ou dénombrable, est une chaîne de Markov à temps continu, si sont vérifiées les deux propriétés :*

$$(1) \text{ propriété de Markov : } \forall (e_1, e_2, \dots, e_n, e_{n+1}) \in E^{n+1}, \forall (t_1, t_2, \dots, t_n, t_{n+1})$$

$\in \mathbb{R}_+^{n+1}$ tels que $t_1 < t_2 < \dots < t_n < t_{n+1}$,

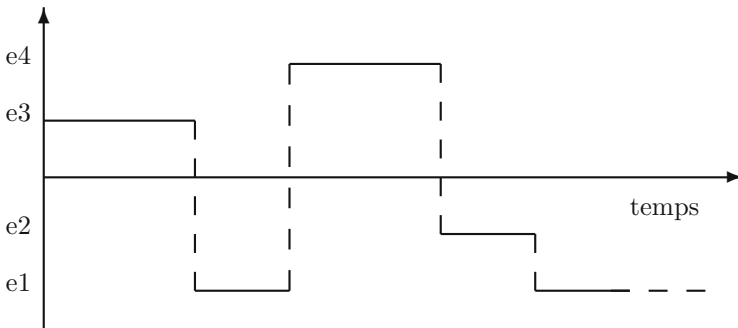
$$P(X_{t_{n+1}} = e_{n+1} \mid X_{t_n} = e_n, \dots, X_{t_1} = e_1) = P(X_{t_{n+1}} = e_{n+1} \mid X_{t_n} = e_n).$$

(2) *homogénéité* : $\forall t_1, t_2, t \in \mathbb{R}_+, \forall e_i, e_j \in E :$

$$P(X_{t_1+t} = e_j \mid X_{t_1} = e_i) = P(X_{t_2+t} = e_j \mid X_{t_2} = e_i) =_{\text{noté}} p_{i,j}(t).$$

Les chaînes de Markov à temps continu étudiées dans ce chapitre sont supposées être **régulières** : leur nombre de transitions en temps fini est presque sûrement fini, ce qui est le cas si la chaîne est récurrente ou finie.

Exemple d'évolution d'une chaîne de Markov à quatre états :



Théorème 9.1 La *matrice de transition* $P(t) = (p_{i,j}(t))_{i,j \in \mathbb{I}}$ vérifie les deux propriétés :

- (1) $\forall t, \forall i, j \in \mathbb{I}, p_{i,j}(t) \geq 0 ;$
- (2) $\forall t, \forall i \in \mathbb{I}, \sum_j p_{i,j}(t) = 1.$

Théorème 9.2 (*Équations de Chapman-Kolmogorov*)

$$\forall s, t \text{ et } \forall i, j \in \mathbb{I}, p_{i,j}(s+t) = \sum_k p_{i,k}(s) p_{k,j}(t),$$

Soit, sous forme matricielle : $P(s+t) = P(s)P(t).$

L'ensemble des matrices de transition $(P(t))_{t \geq 0}$ muni du produit matriciel constitue un semi-groupe, dans lequel l'élément neutre $P(0)$ est la matrice identité. La propriété $P(s+t) = P(s)P(t)$ est analogue à la propriété fondamentale des matrices de transition des chaînes à temps discret :

$$\forall m, n \in \mathbb{N} \quad P(m+n) = P(m)P(n).$$

Preuve

$$\begin{aligned}
 p_{i,j}(s+t) &= P(X_{s+t} = e_j \mid X_0 = e_i) = \sum_k P(X_{s+t} = e_j, X_s = e_k \mid X_0 = e_i) \\
 &= \sum_k P(X_{s+t} = e_j \mid X_s = e_k, X_0 = e_i) P(X_s = e_k \mid X_0 = e_i) \\
 &= \sum_k P(X_{s+t} = e_j \mid X_s = e_k) P(X_s = e_k \mid X_0 = e_i) \\
 &= \sum_k P(X_t = e_j \mid X_0 = e_k) P(X_s = e_k \mid X_0 = e_i) \\
 &= \sum_k p_{i,k}(s) p_{k,j}(t).
 \end{aligned}$$

□

On utilisera désormais l'abréviation **c.m.c.** pour désigner une chaîne de Markov à temps continu.

Hypothèses de continuité et de dérivabilité des $p_{i,j}(t)$

Pour construire les concepts fondamentaux de générateur infinitésimal et d'équations de Kolmogorov associés à toute c.m.c, on supposera vérifiées les hypothèses suivantes de continuité et de dérivabilité des $p_{i,j}(t)$:

$$\mathbf{H}_1 : \forall i \text{ et } j, i \neq j, \lim_{t \downarrow 0^+} p_{i,j}(t) = 0 \text{ et } \forall i, \lim_{t \downarrow 0^+} p_{i,i}(t) = 1.$$

$\mathbf{H}_2 : \forall i \text{ et } j, i \neq j, \lim_{t \downarrow 0^+} \frac{p_{i,j}(t) - p_{i,j}(0)}{t}$ existe et est égale à $p'_{i,j}(0) =_{\text{noté}} a_{i,j}$, dénommé **taux de transition instantané** de e_i à e_j ; $a_{i,j}$ est strictement positif si la transition de e_i à e_j est possible, car la probabilité de transition de e_i à e_j croît en fonction du temps.

$\mathbf{H}_3 : \forall i, \lim_{t \downarrow 0^+} \frac{p_{i,i}(t) - p_{i,i}(0)}{t}$ existe et est égal à $p'_{i,i}(0) =_{\text{noté}} a_{i,i}$ dénommé **taux de permanence** $a_{i,i}$ en l'état e_i est négatif, puisque contrairement à la situation précédente, la probabilité de séjourner dans un état quelconque décroît avec le temps.

Définition 9.2 *Sous les hypothèses précédentes, la matrice $A = (a_{i,j})_{i,j}$ définie par*

$$\lim_{t \downarrow 0} \frac{P(t) - \mathbf{1}_d}{t} = \frac{d}{dt}(P(t))_{(t=0^+)} \text{ existe ;}$$

*elle est nommée **générateur infinitésimal** de la c.m.c. $(X_t)_{t \geq 0}$.*

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdot & \cdot & a_{1,n} \\ a_{2,1} & \cdot & \cdot & \cdot & a_{2,n} \\ \cdot & & & & \\ \cdot & & & & \\ a_{n,1} & a_{n,2} & & & a_{n,n} \end{pmatrix}.$$

Théorème 9.3 *Pour tout état i :*

$$a_{i,i} = - \sum_{j \neq i} a_{i,j}.$$

Preuve $\forall i, \sum_{i=1}^n p_{i,j}(t) = 1$; par dérivation par rapport à t , $\sum_{i=1}^n p'_{i,j}(t) = 0$ et donc $\sum_{i=1}^n p'_{i,j}(0) = \sum_j a_{i,j} = 0$. \square

Théorème 9.4 *Tout processus à accroissements indépendants vérifie la propriété de Markov.*

Preuve Soit la c.m.c. $(X_t)_t$ tel que $X_0 = 0$; désignons X_{t_j} par X_j .

$P(X_n < x_n | X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = P(X_n - X_{n-1} < x_n - x_{n-1} | X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = P(X_n - X_{n-1} < x_n - x_{n-1} | X_1 - X_0 = x_1, \dots, X_{n-1} - X_{n-2} = x_{n-1} - x_{n-2}) = P(X_n - X_{n-1} < x_n - x_{n-1})$, car les $(X_j - X_{j-1})$ sont mutuellement indépendants. $(X_n - X_{n-1})$ étant indépendant du vecteur $(X_1 - X_0, X_2 - X_1, \dots, X_{n-1} - X_{n-2})$, est donc indépendant de la somme de ses composantes. En conclusion, l'expression précédente est égale à : $P(X_n - X_{n-1} < x_n - x_{n-1} | X_{n-1} = x_{n-1}) = P(X_n < x_n | X_{n-1} = x_{n-1})$. \square

Exemple 9.1 Un processus de Poisson $(N_t)_{t \geq 0}$ est une c.m.c., puisqu'il vérifie la propriété de stationnarité et que la propriété des accroissements indépendants implique la propriété de Markov. Déterminons son générateur infinitésimal :

$$\forall t, h > 0, P(N_{t+h} = i+1 | N_t = i) = \lambda h e^{-h\lambda} \text{ et } P(N_{t+h} = i | N_t = i) = 1 - \lambda h e^{-h\lambda},$$

d'où :

$$a_{i,i+1} = \lim_{h \rightarrow 0} \frac{P(N_{t+h}=i+1 | N_t=i) - 0}{h} = \lambda ;$$

$$a_{i,i} = \lim_{h \rightarrow 0} \frac{P(N_{t+h}=i | N_t=i) - 1}{h} = -\lambda.$$

$$A = \begin{pmatrix} -\lambda & \lambda & 0 & \dots & \dots & 0 \\ 0 & -\lambda & \lambda & 0 & & \\ 0 & 0 & -\lambda & \lambda & 0 & \\ \vdots & \ddots & 0 & -\lambda & \lambda & 0 \\ \vdots & & \ddots & 0 & \ddots & \ddots \\ 0 & & & \ddots & \ddots & \ddots \end{pmatrix}.$$

Théorème 9.5 (*Équations de Kolmogorov*)

(1) *Pour tout $t \geq 0$, la matrice de transition $P(t)$ d'une c.m.c. vérifie les équations différentielles ci-dessous :*

(a) équations inverses ou du passé :

$$\forall i, j \quad p'_{i,j}(t) = \sum_{k \in \mathbb{I}} a_{i,k} p_{k,j}(t),$$

$$\text{soit : } P'(t) = A.P(t);$$

(b) équations directes ou du futur :

$$\forall i, j \quad p'_{i,j}(t) = \sum_{k \in \mathbb{I}} p_{i,k}(t) a_{k,j},$$

$$\text{soit : } P'(t) = P(t).A.$$

(2) Lorsque le nombre d'états est fini, $P(t)$ est définie par :

$$P(t) = \sum_{n \geq 0} \frac{(tA)^n}{n!} = e^{tA}.$$

(La forme matricielle ci-dessus a surtout un intérêt théorique, mais sauf cas particuliers, elle n'est pas analytiquement exploitable.)

Preuve (1.a) On dérive l'équation de Chapman-Kolmogorov par rapport à s :

$$\frac{d}{ds} p_{i,j}(s+t) = \sum_k \frac{d}{ds} p_{i,k}(s) p_{k,j}(t); \text{ pour } s \text{ tendant vers } 0 :$$

$$p'_{i,j}(t) = \sum_k p'_{i,k}(0) p_{k,j}(t) = \sum_k a_{i,k} p_{k,j}(t). \text{ On prouve (1.b) de façon analogue.}$$

Schéma formel de démonstration de (2) : $P(t) = (P(\frac{t}{n}))^n$

$$\simeq_{\substack{(n \text{ grand par} \\ \text{rapport } t)}} (P(0) + \frac{t}{n} P'(0))^n = (I_d + \frac{t}{n} A)^n \text{ qui converge vers } e^{At} \text{ quand } n \text{ tend}$$

vers $+\infty$. □

Théorème 9.6 (Loi d'évolution d'une c.m.c.)

(1) Pour tout temps t le vecteur d'évolution $\pi(t)$ vérifie l'équation :

$$\pi(t) = \pi(0)P(t),$$

$$\text{où } \forall i, \pi_i(t) = P(X_t = e_i) \text{ et } \pi(t) = (\pi_1(t), \dots, \pi_n(t)).$$

(2) $\pi(t)$ est solution de l'équation d'évolution :

$$\pi'(t) = \pi(t)A.$$

Preuve (1) Pour tout j , $\pi_j(t) = \sum_i p_{i,j}(t) \pi_i(0)$.

$$(2) \quad \pi'(t) = \pi(0)P'(t) \stackrel{(d'après \text{équ de Kolmogorov})}{=} \pi(0)P(t).A = \pi(t)A. \quad \square$$

Remarque 9.1 Sauf dans le cas d'une c.m.c. ayant un petit nombre d'états ou bien lorsque la matrice A est creuse, ce qui est le cas des processus de naissance et de mort, la résolution analytique de l'équation d'évolution est généralement lourde de calculs. On utilise souvent une méthode analytique ou numérique fondée sur la transformation de Laplace.

Théorème 9.7 (*Caractère exponentiel des temps de séjour*)

Le temps de séjour U_i dans un état quelconque e_i est une v.a. exponentielle de paramètre $-a_{i,i}$.

Preuve Pour les temps t et $h > 0$, h étant petit par rapport à t : $P(U_i > t+h) = P(U_i > t+h \mid U_i > t)P(U_i > t)$ égal à $P(U_i > h)P(U_i > t)$, comme conséquence de l'axiome d'homogénéité. Posons $\Phi(t) = P(U_i > t)$; l'égalité ci-dessus s'écrit $\Phi(t+h) = \Phi(t)\Phi(h)$. Or $\Phi(h) = p_{i,i}(h) + o(h) = 1 + a_{i,i}h + o(h)$, d'où l'on déduit l'équation différentielle $\Phi'(t) = a_{i,i}\Phi(t)$, qui a pour solution $\Phi(t) = e^{a_{i,i}t}$, sachant que $\Phi(0) = 1$.

On conclut que $P(U_i \leq t) = 1 - e^{a_{i,i}t}$ et donc que U_i est de loi exponentielle. \square

On a démontré au chapitre 7 que le temps de séjour dans un état quelconque d'une chaîne de Markov à temps entier est de loi géométrique, laquelle est, comme on le sait (voir chapitre 2.4), l'analogue discret de la loi exponentielle du temps de séjour dans un état d'une c.m.c.

Exemple 9.2 Considérons un système dont la durée de vie T entre deux pannes consécutives est de loi exponentielle $\mathcal{E}(\lambda)$ et dont le temps de réparation R est de loi exponentielle $\mathcal{E}(\mu)$. Le système comporte donc deux états : l'état de marche e_1 et l'état de réparation e_2 . On décrit l'évolution temporelle du système grâce à une c.m.c. à deux états $(X_t)_t$. Déterminons la matrice de transition et le générateur du processus.

Pour t petit, $p_{2,2}(t) = P(X_t = e_2 \mid X_0 = e_2) \approx P(R > t) = e^{-\mu t}$, d'où l'on déduit :

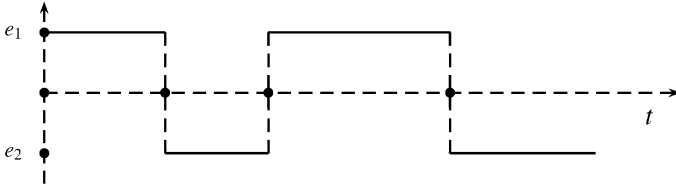
$$a_{2,2} = \left. \frac{dp_{2,2}(t)}{dt} \right|_{t=0} = -\mu ;$$

$$p_{1,1}(t) = P(X_t = e_1 \mid X_0 = e_1) = P(T > t) = e^{-\lambda t}, \text{ d'où : } a_{1,1} = -\lambda.$$

On en déduit le générateur :

$$A = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

Graphe d'évolution de la c.m.c. à deux états :



On calcule facilement $P(t)$ en diagonalisant A .

$$P(t) = e^{At} = \frac{1}{\mu + \lambda} \begin{pmatrix} \mu & \lambda \\ \mu & \lambda \end{pmatrix} + \frac{e^{-(\lambda + \mu)t}}{\mu + \lambda} \begin{pmatrix} \lambda & -\lambda \\ -\mu & \mu \end{pmatrix}, \text{ d'où l'on déduit :}$$

$$\pi(t) = (\pi_1(t), \pi_2(t)) = \pi(0).P(t).$$

Le système démarrant en état de marche : $\pi(0) = (1, 0)$.

$$(\pi_1(t), \pi_2(t)) = \left(\frac{\mu}{\mu + \lambda} + \frac{\lambda}{\mu + \lambda} e^{-(\lambda + \mu)t}, \frac{\lambda}{\mu + \lambda} - \frac{\lambda}{\mu + \lambda} e^{-(\lambda + \mu)t} \right).$$

On aurait pu déterminer $\pi(t)$ en résolvant l'équation d'évolution :

$$\pi'(t) = \pi(t)A.$$

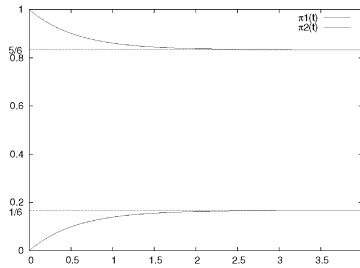


Figure 9.1 – Graphes d'évolution des probabilités $\pi_1(t)$ et $\pi_2(t)$ pour $\lambda = 0,3$ et $\mu = 1,5$.

Lorsque t tend vers $+\infty$, les probabilités $\pi_1(t)$ et $\pi_2(t)$ convergent respectivement vers des probabilités constantes $\frac{\mu}{\mu + \lambda} = \frac{5}{6}$ et $\frac{\lambda}{\mu + \lambda} = \frac{1}{6}$. Pratiquement, sur un temps long, le système est en état de marche pendant 83,3 % du temps.

Théorème 9.8 À toute c.m.c. $(X_t)_t$ on associe une chaîne discrète, dite **chaîne immergée ou incluse** $(Y_n)_n$ définie par :

$$\forall n \in \mathbb{N}^*, Y_n = X_{T_n} ;$$

où T_n est l'instant de la $n^{\text{ième}}$ transition, défini par $\inf\{t > T_{n-1} \mid X_t \neq X_{T_{n-1}}\}$. La matrice de transition $Q = (q_{i,j})_{i,j}$ de la chaîne incluse est définie par :

$$\begin{aligned} \forall i \text{ et } j \text{ distincts} : q_{i,j} &= -\frac{a_{i,j}}{a_{i,i}} \text{ si } j \neq i ; \\ \forall i : q_{i,i} &= 0. \end{aligned}$$

Exemple 9.3 La chaîne incluse associée à la c.m.c. de l'exemple 9.2 est définie par la matrice de transition Q :

$$q_{1,2} = \frac{a_{1,2}}{-a_{1,1}} = \frac{\lambda}{-(-\lambda)} = 1, \quad q_{2,1} = \frac{\mu}{-(-\mu)} = 1.$$

Simulation d'une c.m.c.

Si $\sup_i |a_{i,i}|$ fini, condition de régularité de la chaîne (voir 9.1.1), on simule la chaîne $(X_t)_t$ en appliquant successivement les deux règles ci-dessous :

(1) Lorsque la c.m.c. est dans un état e_i , elle y reste un temps aléatoire de loi exponentielle de paramètre $-a_{i,i}$.

(2) Au terme de ce temps, elle quitte cet état vers l'un des états e_j avec la probabilité $q_{i,j} = \frac{a_{i,j}}{\sum_{k \neq i} a_{i,k}}$.

Les propriétés d'irréductibilité, de récurrence positive ou nulle et de transience d'une c.m.c. sont les mêmes que celles d'une chaîne discrète : le théorème ci-dessous est utile pour les caractériser.

Théorème 9.9 Une c.m.c. $(X_t)_t$ a les mêmes propriétés d'irréductibilité, de récurrence ou de transience que sa chaîne incluse $(Y_n)_n$.

Remarque 9.2 Une c.m.c. ne peut pas être périodique puisque le temps passé dans un état quelconque est décrit par une v.a. continue.

9.1.2 Étude du régime stationnaire

Définition 9.3 Une distribution de probabilités $\pi = (\pi_i)_i$ portée par les états $(e_i)_i$ est **stationnaire** si pour tout j et pour tout $t \geq 0$:

$$\sum_i \pi_i \cdot p_{i,j}(t) = \pi_j,$$

qui s'écrit aussi :

$$\pi \cdot P(t) = \pi.$$

Théorème 9.10 (Équation d'équilibre d'une c.m.c.) Soit une c.m.c. de générateur A , si π est une distribution stationnaire, elle satisfait l'équation d'équilibre :

$$\pi.A = 0,$$

équivalente aux équations

$$\forall j \sum_{i \neq j} \pi_i \cdot a_{i,j} = -\pi_j \cdot a_{j,j},$$

qui expriment l'égalité, en régime stationnaire, du flux moyen entrant dans l'état e_j et du flux moyen sortant vers l'ensemble des autres états. On la détermine en résolvant le système linéaire formé de l'équation d'équilibre et de l'équation de normalisation $\sum_i \pi_i = 1$.

Preuve Par dérivation de l'équation $\pi.P(t) = \pi$ au temps $t = 0$. \square

Théorème 9.11

(1) Une c.m.c. $(X_t)_t$ irréductible et récurrente positive a une unique distribution stationnaire $(\pi_j)_j$ indépendante de la distribution initiale, et égale à $(\pi_j = \lim_{t \rightarrow +\infty} p_{i,j}(t))_j$.

(2) Une c.m.c. transitoire ou récurrente nulle ne possède aucune distribution stationnaire.

9.1 ♠♠ Un système est constitué de deux composants indépendants de durées de vie aléatoires de lois exponentielles $\mathcal{E}(\lambda_i)$ $i = 1, 2$. Lorsque le composant i tombe en panne, il est remplacé par un composant de caractéristiques identiques; la durée du remplacement du composant i défaillant de loi $\mathcal{E}(\mu_i)$ est indépendante de la durée de vie du composant.

Soit $X_t = (X_t^{(1)}, X_t^{(2)})$ le vecteur aléatoire décrivant l'état au temps t du système parmi quatre états possibles : $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$, où 0 désigne l'état de panne et 1 l'état de marche.

(1) Déterminer le générateur du processus de Markov $(X_t)_t$.

[Indication : calculer $\forall t$ et h petit devant t , $P(X_{t+h} = j | X_t = i)$.]

(2) Déterminer sa distribution stationnaire.

9.2 ♠♠ Avions trimoteur et quadrimoteur

Un avion est propulsé par trois réacteurs identiques fonctionnellement indépendants; il continue à voler si deux réacteurs fonctionnent. Soit T_i ($i = 1, 2, 3$) les durées de vie des réacteurs, toutes de loi exponentielle $\mathcal{E}(\mu)$.

(1) Déterminer le taux de panne de l'avion.

(2) Même question dans le cas d'un avion à quatre réacteurs; l'avion continuant à voler si et seulement si trois réacteurs fonctionnent ou bien si deux réacteurs, un sur chaque aile, fonctionnent. Quel est le gain en termes de fiabilité de cette configuration par rapport à la précédente?

9.2 Processus de naissance et de mort

Ce sont des c.m.c. à valeurs dans l'espace d'états \mathbb{N} et dont les seules transitions possibles à partir de l'état n se font vers les états contigus $(n-1)$ et $(n+1)$. Les processus de naissance et de mort sont utilisés pour modéliser les systèmes d'attente et l'évolution de populations. Le processus de Poisson est un processus de naissance et de mort pour lequel la seule transition possible à partir de l'état n se fait vers l'état $(n+1)$.

9.2.1 Définition et propriétés

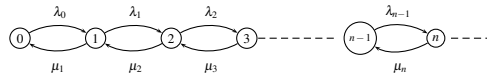
Définition 9.4 *Un processus de naissance et de mort est une c.m.c. à valeurs dans \mathbb{N} , dont le générateur infinitésimal $A = (a_{i,j})_{i,j}$ vérifie :*

- (1) $a_{i,j} = 0$ si $|i-j| \geq 2$;
- (2) $a_{i,i+1} = \text{noté } \lambda_i$ est le **taux de croissance ou de naissance** de l'état e_i ; λ_0 étant égal à $a_{0,1}$;
- (3) $a_{i,i-1} = \text{noté } \mu_i$ est le **taux de décroissance ou de mort** de l'état e_i , pour tout i non nul.

Définition 9.5 *Le générateur A d'un processus de naissance et de mort $(X_t)_t$ est défini par la matrice tridiagonale :*

$$A = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & . & . \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & . \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & . \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 & . \\ 0 & 0 & 0 & \mu_4 & . & . \\ . & . & . & . & . & . \end{pmatrix}.$$

Diagramme de transition :



Le théorème suivant explicite l'équation d'évolution $\pi'(t) = \pi(t)A$ du processus de naissance et de mort.

Théorème 9.12 *Les probabilités $(\pi_n(t))_{n \in \mathbb{N}}$ vérifient le système d'équations différentielles :*

$$\pi'_0(t) = -\lambda_0 \pi_0(t) + \mu_1 \pi_1(t) ;$$

$$\forall n \geq 1, \pi'_n(t) = \lambda_{n-1} \pi_{n-1}(t) - (\lambda_n + \mu_n) \pi_n(t) + \mu_{n+1} \pi_{n+1}(t).$$

Preuve Application directe du théorème 9.5. \square

Théorème 9.13 (Propriétés du temps d'extinction)

Soit T_k le temps qui s'écoule jusqu'à l'extinction, lorsque la taille de la population initiale est k .

(1) La fonction de répartition F_{T_k} de T_k est égale à $P_{k,0}$.

(2) Soit $d_j = \frac{\mu_1 \dots \mu_j}{\lambda_1 \dots \lambda_j}$; la probabilité d'extinction est égale à $\frac{\sum_{j=k}^{+\infty} d_j}{1 + \sum_{j=1}^{+\infty} d_j}$ si $\sum_{j=1}^{+\infty} d_j$ converge et à 1, sinon.

(3) Soit $\rho_j = \frac{1}{d_j \lambda_j}$; le temps moyen de vie jusqu'à l'extinction à partir de l'état k est égal à :

$$E(T_k) = \sum_{j=1}^{+\infty} \rho_j + \sum_{i=1}^{k-1} d_i \sum_{j=i+1}^{+\infty} \rho_j,$$

si $\sum_{j=1}^{+\infty} \rho_j$ converge ; dans le cas contraire, le temps moyen de vie est infini.

Preuve (2) La détermination de la probabilité d'extinction nécessite l'usage de la chaîne induite $(Y_n)_n$ associée au processus de naissance et de mort. On vérifiera que sa matrice de transition est définie par les seuls termes matriciels non nuls : $p_{n,n-1} = \frac{\mu_n}{\mu_n + \lambda_n}$ et $p_{n,n+1} = \frac{\lambda_n}{\mu_n + \lambda_n}$. \square

9.2.2 Étude du régime stationnaire des processus de naissance et de mort

Théorème 9.14 (Équations d'équilibre)

La distribution stationnaire $\pi = (\pi_0, \pi_1, \dots, \pi_n, \dots)$ est solution de l'équation d'équilibre $\pi.A = 0$ équivalente au système linéaire :

$$-\lambda_0 \pi_0 + \mu_1 \pi_1 = 0,$$

$$\forall n \geq 1, \lambda_{n-1} \pi_{n-1} - (\lambda_n + \mu_n) \pi_n + \mu_{n+1} \pi_{n+1} = 0.$$

La distribution stationnaire $(\pi_n)_n$ existe et est unique si et seulement si la série $\sum_{n \geq 1} \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n}$ est convergente. Dans ce cas, elle est définie par :

$$\pi_0 = \left(1 + \sum_{n \geq 1} \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n}\right)^{-1} \text{ et } \forall n \geq 1 : \pi_n = \pi_0 \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n}.$$

Preuve On calcule d'abord π_0 et l'on en déduit par récurrence les valeurs des π_i , $i \geq 1$. \square

9.3 ♠♠ Une entreprise de location de voitures possède N voitures ; chacune d'elles peut tomber en panne avec un taux de panne égal à μ ; la durée de sa réparation est de loi exponentielle $\mathcal{E}(\lambda)$. Soit X_t le nombre de voitures disponibles au temps t .

- (1) Démontrer que $(X_t)_t$ est un processus de naissance et de mort.
- (2) Déterminer la distribution stationnaire π .

9.4 ♠♠

- (1) Étude du processus à croissance pure

Hypothèses : $\lambda_n = n\lambda$ $\mu_n = 0$ $X_0 = 1$. Démontrer par récurrence que pour tout $t \geq 0$ et tout entier positif n , la loi de X_t est géométrique : $p_n(t) = e^{-\lambda t} (1 - e^{-\lambda t})^{n-1}$. (Pour résoudre l'équation différentielle linéaire à laquelle on aboutit, utiliser la méthode de la variation de la constante.)

- (2) Étude du processus à décroissance pure

Hypothèses : $\lambda_n = 0$ et $\mu_n = n\mu$; on suppose que $X_0 = N$. Démontrer que la loi $p_n(t)$ de X_t est la loi binomiale de paramètre $e^{-\mu t}$.

Exemple 9.4 Processus de Ferry-Yule

C'est un processus de naissance et de mort pour lequel les paramètres λ_n sont des fonctions affines de n , de la forme $\lambda_n = n\lambda + \alpha$ et $\mu_n = n\mu$. Supposons que la population initiale comporte N individus ($X_0 = N$), où α est le taux d'immigration du système. Les taux de transition d'un tel processus sont proportionnels à la taille de la population.

(1) Si $\lambda \neq \mu$, on démontre que $p_n(t) = (1 - a(t))(1 - b(t))(b(t))^{n-1}$ où $a(t) = \frac{\mu(e^{(\lambda-\mu)t} - 1)}{(\lambda e^{(\lambda-\mu)t} - \mu)}$ et $b(t) = \frac{\lambda}{\mu} a(t)$.

$$E(X_t) = \begin{cases} \alpha t + N & \text{si } \lambda = \mu \\ N e^{(\lambda-\mu)t} + \frac{\alpha}{\lambda-\mu} (e^{(\lambda-\mu)t} - 1) & \text{si } \lambda \neq \mu \end{cases}$$

(2) Si $\mu = 0$ et $\lambda = 0$, le processus de naissance et de mort s'identifie au processus de Poisson homogène d'intensité α .

9.3 Files d'attente

9.3.1 Définition et exemples

Définition 9.6 Un système de service est constitué d'un espace d'attente où les clients arrivent l'un après l'autre, en formant une file, et d'un espace de service. Il est décrit par les cinq caractéristiques suivantes :

(1) la nature du processus stochastique décrivant le flux des entrées, indépendantes entre elles, déterminé par la loi des intervalles de temps

entre deux arrivées. Si la loi est exponentielle $\mathcal{E}(\lambda)$, le flux est décrit par le processus de Poisson d'intensité λ , qui est le taux moyen d'entrée des clients dans l'unique file d'attente.

(2) **Le nombre s de services mis en parallèle.**

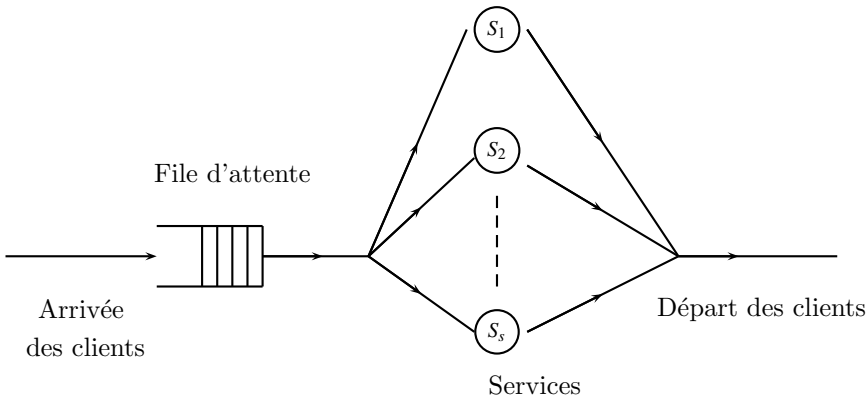
(3) **La loi de probabilité de la durée de service S** , de moyenne $\frac{1}{\mu}$; les durées de service sont toujours des v.a. indépendantes de même loi, qui est souvent mais non nécessairement exponentielle.

(4) **La capacité K du système**, nombre maximal de clients présents dans le système.

(5) **La discipline d'attente ou politique de service.**

La règle la plus fréquemment utilisée s'exprime sous la forme « le premier arrivé est le premier servi », connue sous l'abréviation anglo-saxonne FIFO (first in first out). Cependant, en théorie des réseaux, il existe d'autres politiques, cherchant par exemple à maximiser différentes formes d'unités ou à garantir des propriétés temporelles.

Représentation d'un système d'attente :



La notation dite de Kendall permet de symboliser une file d'attente par la suite de caractères : $A/S/s/K$.

La lettre A caractérise la loi du processus d'entrée et S la loi de service ; par exemple, si on affecte la lettre M en A ou en S , on indique que le flux des entrées est poissonnien ou que les temps de service sont de loi exponentielle. La lettre s dans la notation de Kendall désigne le nombre de services et la lettre K la capacité du système qui s'exprime par un entier strictement positif. (La lettre M a été choisie par Kendall pour exprimer le caractère markovien

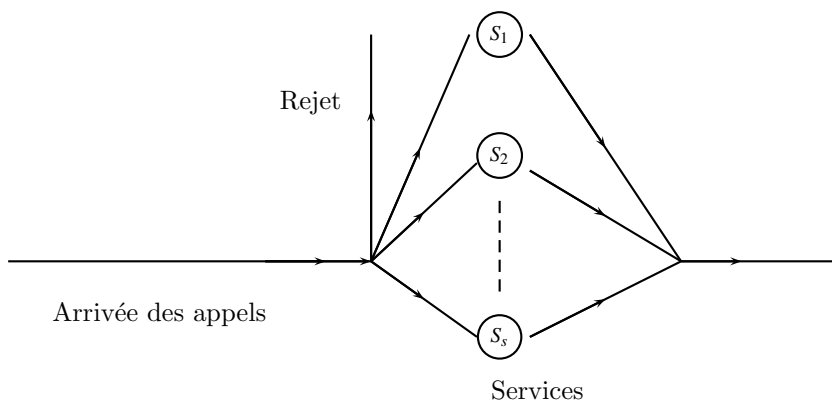
d'un processus dont les temps entre deux arrivées sont indépendants et de loi exponentielle.) Ces quatre symboles peuvent être suivis par un cinquième qui qualifie la politique du service. On peut adjoindre à la nomenclature précédente deux symboles supplémentaires désignant le nombre maximal de clients potentiels et le sigle symbolisant la discipline de service.

Exemples systèmes d'attente classiques

(1) Guichet de poste du type $M/M/s/K_\infty$: un service comporte plusieurs serveurs de mêmes caractéristiques, le client en attente choisit la première station qui se libère, et la file est de capacité bornée.

(2) Central téléphonique (voir figure ci-dessous) ou informatique du type $M/M/s/s$, possédant un nombre s de serveurs ; c'est un système sans attente et à perte, car si un appel n'est pas pris par l'un des serveurs, il est rejeté et doit se représenter.

(3) Station-service, salle d'attente du type $M/M/s/K$ où $K > s$; systèmes à capacité K bornée.



9.3.2 Étude du régime stationnaire des files d'attente

L'étude du régime transitoire des systèmes d'attente généraux est généralement très difficile, voire analytiquement impossible, sauf dans quelques cas comme le système $M/M/1$. Il s'agit d'un moindre mal du point de vue pratique puisque, dans ce domaine comme dans beaucoup d'autres, seuls les régimes stationnaires importent. Soient X_t la v.a. associée au nombre de clients présents à l'instant t dans le système de service et $p_n(t)$ la probabilité $P(X_t = n)$. S'il existe un régime stationnaire, on le décrit par la v.a. X indépendante de t , et on note sa distribution stationnaire par :

$$(P(X = n) = p_n)_{n \in \mathbb{N}}.$$

Définition 9.7 Paramètres descriptifs d'un système d'attente

$\bar{\lambda}$ = taux moyen d'entrée.

$\frac{1}{\mu}$ = temps moyen passé dans une unité de service.

L = nombre moyen $E(X) = \sum_{n \geq 1} n p_n$ de clients dans le système.

L_q = nombre moyen $\sum_{n \geq 1+s} (n-s) p_n$ de clients dans la file.

W = durée moyenne de présence d'un client quelconque dans le système.

W_q = durée d'attente moyenne d'un client quelconque dans la file.

ρ = intensité $\frac{\lambda}{\mu}$ du trafic, dont l'unité est appelée erlang.

u = taux d'utilisation de chaque serveur (proportion du temps pendant lequel il est occupé).

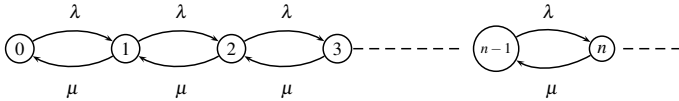
Les paramètres L , L_q , W , W_q et u sont aussi appelés **paramètres de performance du système**.

Théorème 9.15 (Relations entre les paramètres de performance)

$$\begin{aligned}
 (1) \quad L &= L_q + \frac{\bar{\lambda}}{\mu} & (2) \quad L &= \bar{\lambda} \cdot W \text{ (Formule de Little).} \\
 (3) \quad W &= W_q + \frac{1}{\mu} & (4) \quad L_q &= \bar{\lambda} W_q.
 \end{aligned}$$

Étude de la file M/M/1

Le système d'attente le plus simple est constitué d'un serveur et d'une file à capacité illimitée, désigné par $M/M/1$ ou $M/M/1/\infty$. C'est un processus de naissance et de mort de paramètres : $\lambda_n = \lambda$, $\mu_n = \mu$ si $n \neq 0$ et $\mu_0 = 0$. Dans ce cas, le taux moyen d'entrée $\bar{\lambda}$ est égal à λ .


Définition 9.8 (Système de service M/M/1)

(1) Le processus $(T_n)_{n \in \mathbb{N}}$ des temps d'arrivée des clients est supposé poissonnien, d'intensité λ ; les durées $U_n = T_{n+1} - T_n$ entre deux arrivées consécutives, sont donc indépendants et de loi exponentielle $\mathcal{E}(\lambda)$.

(2) Les durées de service $(S_n)_{n \in \mathbb{N}}$ représentées par des v.a. indépendantes sont de même loi exponentielle de paramètre μ définissant le taux de service ; la durée moyenne de service est donc égale à $\frac{1}{\mu}$.

On supposera vérifiée la condition d'existence d'un régime stationnaire : l'intensité ρ du trafic est strictement inférieure à 1.

Théorème 9.16 *La distribution stationnaire $(\pi_j)_j$ est de loi géométrique :*

$$\forall j, \pi_j = (1 - \rho) \rho^j.$$

Preuve Pour tout n , $\lambda_n = \lambda$, μ_n est égal à μ si $n \neq 0$ et à 0 si $n = 0$.

D'autre part : $\pi_n = \pi_0 \rho^n$, d'où : $\pi_n = \rho^n (1 - \rho)$. \square

Théorème 9.17 *La durée d'attente T d'un client dans le système est de loi exponentielle de paramètre $(\mu - \lambda)$.*

Preuve Si, lorsque le client arrive, $(n - 1)$ clients sont dans la file, et un client est en service, son temps d'attente W est égal à la somme de $(n + 1)$ temps d'attente au service, soit $\sum_{i=1}^{n+1} T_i$. En effet, compte tenu de l'absence de mémoire de la loi exponentielle, le temps de service restant au client en train d'être servi vérifie la même loi exponentielle. L'application de la formule de recomposition (Théorème 6.3) permet d'établir : $f_W(t) = \sum_{n=0}^{\infty} \pi_n f_{T_1 + \dots + T_{n+1}}(t)$. On calcule la transformée de Laplace $\mathcal{L}(f_W)(s) = \frac{(1-\rho)\mu}{(1-\rho)\mu+s}$, puis par inversion de ladite transformée, on obtient : $f_W(s) = (\mu - \lambda)e^{-(\mu-\lambda)t}$. \square

Théorème 9.18 *(Caractéristiques du système M/M/1)*

$$\begin{aligned} u &= \rho = \frac{\lambda}{\mu} \\ L &= \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda} = \frac{\rho}{1-\rho} \\ L_q &= \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)} \\ W &= \frac{1}{\mu-\lambda} \\ W_q &= \frac{L}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{\rho}{\mu(1-\rho)}. \end{aligned}$$

Preuve $u = \sum_{n=1}^{+\infty} \pi_n = 1 - \pi_0 = \rho$.

Par définition : $L = \sum_{n \geq 1} n \pi_n = \frac{\rho}{1-\rho}$ et $L_q = \sum_{n \geq 2} (n-1) \pi_n = L - \rho = \frac{\rho^2}{1-\rho}$. Soient T_q le temps d'attente d'un individu dans la file, et E_n l'événement « n clients sont dans le système quand arrive l'individu ».

$$\begin{aligned} P(E_n) &= (1 - \frac{\lambda}{\mu})(\frac{\lambda}{\mu})^n = (1 - \rho) \rho^n \text{ et } E(T_q | E_n) = \frac{n}{\mu}. \\ W_q &= E(T_q) = \sum_{n \geq 0} E(T_q | E_n) P(E_n) = \sum_{n \geq 0} \frac{n}{\mu} (1 - \rho) \rho^n = \frac{1}{\mu} \sum_{n \geq 0} n \rho^n = \frac{L}{\mu}, \\ \text{d'où : } W &= W_q + \frac{1}{\mu} = \frac{L+1}{\mu}. \end{aligned} \quad \square$$

Exemple 9.5 Soit un système d'attente M/M/1 où un client arrive en moyenne tous les quarts d'heure, et dont la durée moyenne de service est 10 minutes.

(1) Calculons les probabilités pour que deux clients, trois clients, k clients soient en attente dans la file, ainsi que la longueur moyenne de la file et le temps moyen d'attente pour un client. (On prend l'heure pour unité.)

$$\pi_2 = \left(\frac{2}{3}\right)^2 \cdot \frac{1}{3} = 0,15 ; \pi_3 = \left(\frac{2}{3}\right)^3 \cdot \frac{1}{3} \simeq 0,1 ; \pi_k = \left(\frac{2}{3}\right)^k \cdot \frac{1}{3}$$

$L_q = \frac{\rho^2}{1-\rho} = \frac{4}{3}$ et $W_q = \frac{\rho}{\mu(1-\rho)} = \frac{1}{3}$ soit 20 minutes d'attente moyenne par client.

(2) Si le taux d'arrivée augmente de 25 %, voyons comment évoluent la longueur moyenne de la file et le temps moyen d'attente pour un client.

Le taux d'arrivée initial est de 4 par heure, et après augmentation il est de 5 par heure ; d'où $\rho = \frac{5}{6}$.

$L_q = \frac{25}{6}$ soit une file moyenne trois fois plus longue que précédemment.

$W_q = \frac{5}{6}$, soit deux fois et demi plus d'attente moyenne.

On constate un phénomène analogue lorsque la durée moyenne de service μ augmente : choisir par exemple $\mu = 4,5$ et $\lambda = 4$.

Conclusion : de façon générale, une augmentation petite de l'intensité ρ au voisinage de 1 se traduit par une augmentation significative de la longueur moyenne de la file et du temps moyen d'attente.

D'un point de vue pratique, il y a donc conflit d'intérêts entre les clients qui souhaitent passer le moins de temps possible en attente et le gestionnaire qui souhaite minimiser le nombre de services. Si le gestionnaire du service souhaite diminuer le risque de perte des clients, il faudra qu'il choisisse un nombre minimal de services de façon à ce que l'intensité du trafic ρ ne soit pas trop proche de 1.

Étude des systèmes d'attente M/M/s/K

(1) Cas où la capacité K est non bornée

On rappelle que la capacité est le nombre maximal de clients en attente ou en service. Ce modèle convient aux situations où l'espace d'attente est de très grande taille, donc d'une capacité beaucoup plus grande que le nombre de serveurs s .

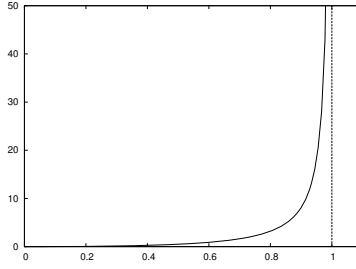


Figure 9.2 – Graphe d'évolution de L_q en fonction de ρ montrant la forte croissance de L_q au voisinage de $\rho = 1$.

Théorème 9.19 (*Caractéristiques du système M/M/s/∞*)

Condition d'existence d'un régime stationnaire : $\rho = \frac{\lambda}{\mu} < s$.

$$(1) \pi_0 = \left(\sum_{k=0}^{s-1} \frac{\rho^k}{k!} + \frac{\rho^s}{(s-1)!} \frac{1}{s-\rho} \right)^{-1}, \quad \pi_k = \begin{cases} \frac{\rho^k}{k!} \pi_0 & \text{si } k \leq s \\ \frac{\rho^k}{s! s^{k-s}} \pi_0 & \text{si } k \geq s. \end{cases}$$

$$(2) L_q = \frac{\pi_0 \rho^{s+1}}{(s-1)!(s-\rho)^2}; \quad W_q = \pi_0 \frac{\rho^s}{\mu (s-1)!(s-\rho)^2}.$$

(3) La probabilité pour que tous les services soient occupés est égale à :

$$\pi_0 \frac{s \rho^s}{s!(s-\rho)}.$$

9.5 ♠♠ Démonstration du théorème

- (1) Tracer le diagramme de transition et écrire les équations bilan.
- (2) Les résoudre dans chacun des cas : $0 \leq k \leq s$ et $k > s$.
- (3) Déterminer L_q .

Exemple 9.6 Optimisation de la production d'un atelier

Un atelier emploie deux ouvriers qui produisent chacun quatre pièces par semaine dans le cadre d'une sous-traitance. La demande étant passée à dix pièces, le chef d'atelier envisage d'embaucher un ouvrier supplémentaire, tandis que les deux ouvriers proposent de produire chacun six pièces si leur salaire est augmenté de la moitié du traitement prévu pour le nouvel ouvrier. Pour qui cette dernière solution est-elle avantageuse ? Déterminons les paramètres de performance dans chacun des cas.

- (1) Embauche d'un ouvrier :
 $s = 3$; soit $\rho = \frac{\lambda}{\mu} = \frac{5}{2}$.

$$\pi_0 = (1 + \rho + \frac{\rho^2}{2} + \frac{\rho^3}{2(3-u)})^{-1} = (1 + \frac{5}{2} + \frac{25}{8} + \frac{125}{8})^{-1} = \frac{4}{89}$$

$$\pi_3 = p_0 \frac{\rho^3}{3!} \cong 0,117; L_q \cong 3,51; L \cong 6,01; W_q \cong 0,35(\text{semaine}); W \cong 0,6.$$

(2) Pas d'embauche :

$$s = 2; \text{ posons } u = \frac{5}{3}. \quad \pi_0 = (1 + u + \frac{u^2}{2-u})^{-1} \cong 0,126; L_q \cong 3,78; L \cong 5,45; W_q \cong 0,38(\text{semaine}); W \cong 0,54.$$

9.6 ♠ Calculer et comparer les compétences des deux systèmes d'attente : $M/M/1$ de paramètres $(\lambda, 2\mu)$ et $M/M/2$ de paramètres (λ, μ) . Montrer qu'il est préférable de disposer d'un service ayant un serveur deux fois plus rapide que d'un service ayant deux serveurs de mêmes caractéristiques.

(2) Cas d'une capacité K finie

Les systèmes d'attente à capacité limitée sont dits à *perte* : lorsqu'un client trouve le ou les services occupés, il est refusé et repart ; c'est typiquement le cas d'un central téléphonique. Dans la notation de Kendall, on les désigne par : $M/M/s/K$, où K est la capacité maximale du système. Le problème 19 en propose l'étude.

Système d'attente $M/G/1$

Ce système d'attente est caractérisé par une durée de service de loi quelconque symbolisée par G en notation de Kendall. On considère donc que le processus d'arrivée est toujours poissonnien d'intensité λ et que le temps de service T_S de densité f , d'espérance $\frac{1}{\mu}$ et de variance σ^2 sont indépendants entre eux. Il s'agit de déterminer la distribution stationnaire du processus $(X_t)_t$.

Soit $(X_{T_i})_i$ la chaîne incluse représentant le nombre de clients dans le système aux instants aléatoires respectifs T_1, T_2, \dots , où le premier client, respectivement le second client, quittent le service.

Soit N_k le nombre de clients entrant dans le système pendant la durée de service du k -ième client ; les N_k sont indépendantes entre elles, de même loi :

$$P(N_k = n) = \int P(N_k = n \mid T_S = t) f(t) dt = \int e^{-\lambda t} \frac{(\lambda t)^n}{n!} f(t) dt \equiv a_n.$$

Les durées de service n'étant plus de loi exponentielle, le processus $(X_t)_t$ n'est plus markovien, alors que la chaîne $(X_{T_i})_i$ l'est, de matrice de transition $P = (p_{i,j})_{i,j}$ définie par :

$$\begin{aligned} p_{0,j} &= a_j \quad \text{si } j \geq 0, \\ p_{i,j} &= a_{j-i+1} \quad \text{si } 1 \leq i \leq j+1, \\ p_{i,j} &= 0 \quad \text{sinon.} \end{aligned}$$

Ainsi :

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & . & . \\ a_0 & a_1 & a_2 & . & . \\ 0 & a_0 & a_1 & & \\ 0 & 0 & a_0 & a_1 & \\ . & . & . & a_0 & a_1 \end{pmatrix}.$$

On peut démontrer que la chaîne $(X_{T_i})_i$ est irréductible, apériodique et récurrente positive, et possède donc d'après le théorème de Chacon-Ornstein une unique distribution stationnaire $(\pi_j)_j$. Grâce à l'utilisation des fonctions génératrices des distributions $(a_i)_i$ et $(\pi_j)_j$, on établit que :

$$L = \left(\frac{1 + C_S^2}{2} \right) \frac{\rho^2}{1 - \rho} + \rho \text{ où } C_S^2 = \mu^2 \sigma^2.$$

L'étude du système d'attente $G/G/1$ où le processus des arrivées n'est pas poissonnien et la durée de service n'est pas de loi exponentielle, se traite grâce à la transformation de Laplace.

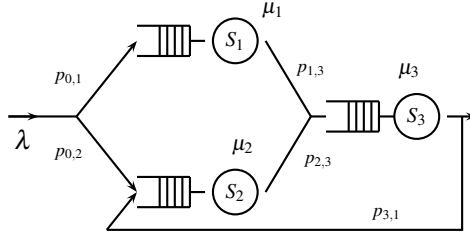
9.3.3 Réseaux de Jackson

De nombreux réseaux de communication ou de services sont constitués de systèmes d'attente communiquant entre eux. En général, il n'y a pas de solution analytique aux problèmes posés par l'évolution temporelle de ces réseaux ; seules des méthodes de simulation permettent d'en faire une approche numérique. Toutefois les réseaux dits de Jackson (du nom de leur créateur), dans lesquels ne circule qu'un type de clients, comme c'est souvent le cas pour les réseaux informatiques, sont modélisables dans le cadre théorique des files d'attente.

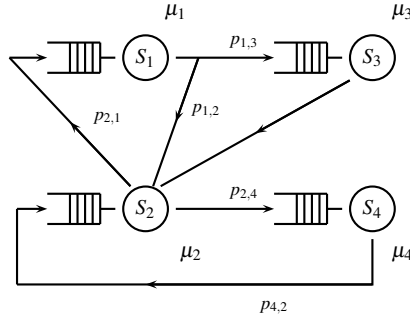
Définition 9.9 *Un réseau de Jackson est un ensemble de K systèmes d'attente $M/M/1$ interconnectés ayant chacun une capacité infinie, de durées moyennes de service $(\mu_i)_{i=1,\dots,K}$, dans lequel ne circule qu'une classe de clients. Au terme du service procuré par la station i , le client (ou l'unité) se transporte à l'une des $K - 1$ stations, par exemple j , avec une probabilité $p_{i,j}$, ou bien sort du réseau avec la probabilité β_i .*

Notons que tout client a la possibilité de revenir à la station i qu'il vient de quitter sans toutefois sortir du système ; dans ce cas $p_{i,i} > 0$. D'autre part, des clients peuvent venir de l'extérieur du système dans la station i , suivant le processus de Poisson d'intensité α_i .

Définition 9.10 *Un réseau de Jackson est **fermé** si $\alpha_i = \beta_i = 0$ pour tout i ; il est **ouvert** sinon. Dans un réseau **ouvert** les clients peuvent entrer et sortir du système ; dans un réseau **fermé**, leur nombre est constant, et ils y évoluent indéfiniment.*



Réseau ouvert



Réseau fermé

Étude des réseaux ouverts à un serveur par station

On appelle λ_j le taux d'arrivée dans la file d'attente j , égale à la somme du taux d'arrivée $p_{0,j}$ des clients venant de l'extérieur et des taux d'arrivée $p_{i,j}$ venant de l'ensemble des files d'attente.

Théorème 9.20 Pour toute station j , les probabilités de transition $(p_{i,j})_{i,j}$ vérifient l'équation de flux ou de trafic du réseau :

$$\alpha_j + \sum_{i=1}^K \lambda_i p_{i,j} = \lambda_j. \quad (*)$$

$\left(\alpha_j + \sum_{i=1}^K \lambda_i p_{i,j} \right)$ est le taux de trafic de la station i , égal au taux d'arrivée λ_j .

Si on désigne l'intensité du flux entrant total par λ et que l'on pose $\lambda_i = \lambda e_i$ où e_i est le taux de visite du service i et $p_{0,i} = \frac{\alpha_i}{\lambda}$, les équations de flux s'écrivent :

$$p_{0,i} + \sum_{j=1}^K e_j p_{j,i} = e_i \quad i = 1, 2, \dots, K. \quad (**)$$

La connaissance des $p_{j,i}$ permet donc de déterminer les taux de visite aux stations e_i , en résolvant le système d'équations linéaires précédent.

Preuve Sachant que $p_{i,j}$ est la probabilité qu'un client aille du service i au service j , $\sum_{j=1}^K \lambda_j p_{i,j}$ est le flux moyen parvenant de l'ensemble du réseau ; d'où le résultat en y ajoutant le flux moyen externe atteignant i . \square

Notons $X_t^{(i)}$ la v.a. associée au nombre de clients dans la station i , au temps t ; $(X_t)_t = (X_t^{(1)}, X_t^{(2)}, \dots, X_t^{(i)}, \dots, X_t^{(K)})_t$ est le processus stochastique vectoriel décrivant l'évolution du nombre de clients dans chacun des services du réseau. Le théorème suivant établit l'existence d'un comportement stationnaire décrit par la distribution stationnaire définie par :

$$P(X^{(1)} = x_1, X^{(2)} = x_2, \dots, X^{(i)} = x_i, \dots, X^{(K)} = x_K).$$

Théorème 9.21 (Régime stationnaire d'un réseau de Jackson ouvert)

(1) Le processus $(X_t)_t$ admet une distribution stationnaire définie par :

$$\forall (x_1, x_2, \dots, x_K) \in \mathbb{N}^K, P(\cap_i (X^{(i)} = x_i)) = \prod_{i=1}^K P(X^{(i)} = x_i) = \prod_{i=1}^K \pi_i$$

où $\pi_i = (1 - \rho_i) \rho_i^{x_i}$.

(2) La file d'attente de la station i est $M/M/1$ de paramètres (λ_i, μ_i) .

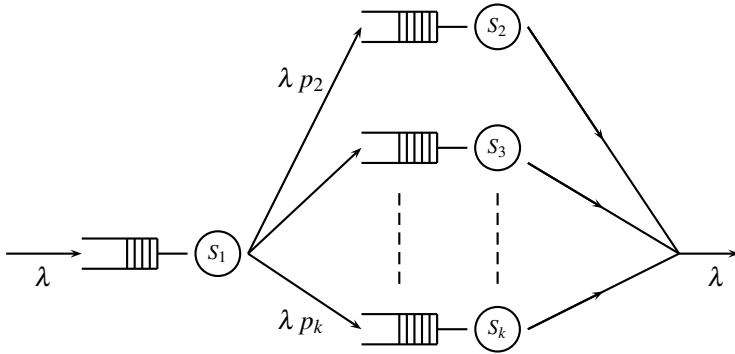
(3) Le réseau est stable si pour tout i : $\lambda_i < \mu_i$.

En régime stationnaire, un réseau de Jackson ouvert de K stations se comporte comme un ensemble indépendant formé des K systèmes d'attente $M/M/1$ qui le constituent.

Les flux à l'entrée des stations ne sont généralement pas poissonniens ; il en est ainsi par exemple lorsque des boucles sont incluses dans le réseau. Toutefois il existe des réseaux où les flux à l'entrée des stations sont poissonniens, comme c'est le cas dans les deux configurations suivantes.

(1) Stations en série (qui ont été étudiés dans l'exercice précédent).

(2) Stations en forme de bifurcation aléatoire : tout client qui sort de la station S_1 a la probabilité p_i d'aller dans la file d'attente S_i où $i = 2, 3, \dots, k$. Cette décomposition probabiliste du flux poissonnien initial d'intensité λ conserve le caractère poissonnien des flux secondaires d'intensités respectives λp_i .



La forme analytique de la distribution stationnaire, $\prod_{i=1}^K P(X_i = x_i)$, est à l'origine de la dénomination **réseau à forme produit** pour désigner un réseau de Jackson stable et ouvert.

Théorème 9.22 (*Caractéristiques d'un réseau de Jackson ouvert*)

(A) Pour chaque système i :

- (1) débit moyen : $\lambda_i = \lambda e_i$;
- (2) nombre moyen de clients : $L_i = \frac{\rho_i}{1-\rho_i}$;
- (3) temps moyen passé dans le système i : $W_i = \frac{L_i}{\lambda_i}$.

(B) Pour l'ensemble du réseau :

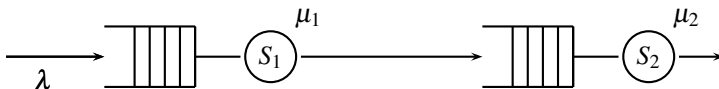
- (1) débit moyen $\lambda = \sum_i \lambda_i p_{i,0}$;
- (2) nombre moyen de clients $L = \sum_i \frac{\rho_i}{1-\rho_i}$;
- (3) temps moyen passé dans le réseau $W = \frac{L}{\lambda}$.

9.7 ♠ Étude des réseaux constitués de deux systèmes d'attente M/M/1 en parallèle et en série.

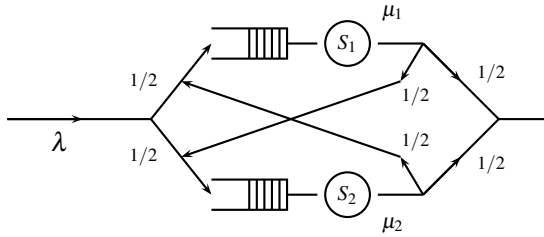
Déterminer la distribution stationnaire et les paramètres de performance.

Remarque 9.3 Des réseaux peuvent avoir des structures différentes et posséder la même distribution stationnaire. C'est par exemple le cas des deux réseaux suivants :

- (1) le réseau formé de deux systèmes de caractéristiques μ_1 et μ_2 , en série.



(2) le réseau ayant le graphe de transition ci-dessous :



Traitement pratique d'un réseau ouvert

- (1) Détermination des taux de service e_i , puis des taux d'arrivée λ_i dans la file d'attente ;
- (2) Vérification des conditions de stabilité ($\lambda_i < \mu_i$, $\forall i$), garantissant l'existence d'un régime stationnaire ;
- (3) Calcul des caractéristiques du réseau.

Réseaux de Jackson ouverts avec K stations possédant plusieurs serveurs

Soit s_i le nombre de serveurs de la station i ; les taux e_i et les taux moyens d'arrivée se calculent de la même façon. La condition d'existence d'un régime stationnaire pour le réseau est :

$$\forall i : \lambda_i < s_i \mu_i.$$

En régime stationnaire, ces réseaux sont toujours équivalents à l'ensemble de K files d'attente $M/M/s_i$ qui les constituent ; la distribution stationnaire a la même forme que dans le cas d'un seul serveur, avec $\rho_i = \frac{\lambda_i}{s_i \mu_i}$.

Réseaux de Jackson fermés (dits de Gordon-Newel)

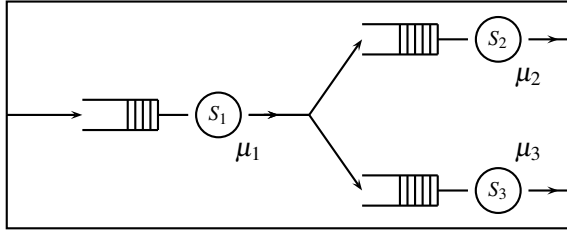
Définition 9.11 *Un réseau de Gordon-Newel est un réseau fermé de K systèmes monoserveurs de taux de service respectifs $(\mu_i)_i$, contenant un nombre constant N de clients, caractérisé par les propriétés :*

- (1) Pour toute station i : $\sum_{j=1}^K p_{i,j} = 1$.
- (2) Pour tout temps t : $\sum_{j=1}^K X_j(t) = N$.
- (3) Pour toute station i , le taux moyen de visite e_i est égal à :

$$e_i = \sum_{j=1}^K e_j p_{j,i}. (*)$$

Le système d'équations (*) est indéterminé, sa résolution nécessite que l'un des taux de visite soit fixé pour l'une des stations dite de référence. Sachant qu'un client passe un nombre illimité de fois par chaque station, le taux moyen de visite e_i de chaque station i est donc égal au nombre moyen de visites de i entre deux passages par la station de référence.

Exemple 9.7 Soit le réseau formé de trois stations interconnectées selon le graphe suivant :



Relations entre les taux moyens de visite : $\frac{1}{2} e_1 = e_2 = e_3$. Donc entre deux passages par S_1 , chaque client passe en moyenne une fois sur deux par S_2 et par S_3 .

Étude du régime stationnaire des réseaux de Gordon-Newel

Théorème 9.23 La distribution stationnaire d'un réseau de Gordon-Newel à K stations est définie par :

$$(*) \quad P(X^{(1)} = n_1, \dots, X^{(K)} = n_K) = \frac{1}{G(K)} \prod_{i=1}^K \left(\frac{e_i}{\mu_i} \right)^{n_i}$$

où $G(K)$ est une constante de normalisation, vérifiant :

$$\sum_{(n_1, \dots, n_K)} P(X_1 = n_1, \dots, X_K = n_K) = G(K),$$

les K -uplets (n_1, \dots, n_K) sont tels que $\sum_{i=1}^K n_i = N$.

D'après le théorème ci-dessus, les v.a. $(X^{(i)})_i$ ne sont pas indépendantes : un réseau de Gordon-Newel n'est pas équivalent à un ensemble de files M/M/1 indépendantes.

La détermination des probabilités stationnaires marginales π_i et des caractéristiques du réseau nécessite souvent, sauf dans quelques cas particuliers, l'utilisation d'algorithmes spécifiques. Il existe deux types d'approche : l'approche fondée sur le calcul de la constante $G(K)$ à l'aide d'algorithmes dits de *convolution*, et l'approche plus récente qui s'est révélée plus efficace, en permettant de déterminer par récurrence les caractéristiques du réseau (voir B.Baynat, *Théorie des files d'attente*).

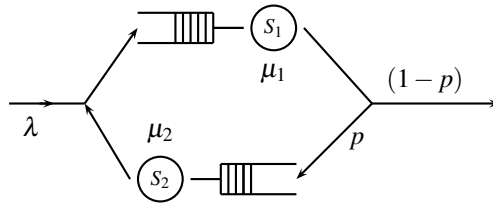
9.8 ♠♠ Un système informatique est constitué d'une unité centrale ($\mu_1 = 2$) et d'une unité entrée/sortie ($\mu_2 = 2$) dans lequel arrivent des requêtes selon un processus de Poisson d'intensité λ , (voir : figure ci-dessous)

(1) Déterminer les taux de visite e_i et les taux d'arrivée λ_i . Pourrait-on prévoir sans calcul que $\lambda_1 > \lambda_2$?

(2) À quelle condition le réseau admet-il une distribution stationnaire ? Déterminer les taux moyens de requêtes dans chacune des stations.

(3) Si $\lambda = 1$ et $p = \frac{1}{3}$, quel est le temps moyen passé par une requête dans la file d'attente de l'unité centrale ?

(4) On suppose qu'il n'y a ni arrivée, ni sortie de requêtes et que N requêtes circulent dans le système. Déterminer la distribution stationnaire du vecteur (X_1, X_2) .



La théorie des réseaux s'étend à des configurations plus générales que les précédentes. C'est le cas des réseaux fermés multiserveurs, des réseaux traitant différentes classes de clients ayant des comportements non standards : clients estimant une file trop longue et refusant d'y entrer, clients arrivant par groupes, constituant un processus de Poisson composé. Leur étude fait l'objet de traités spécialisés.

Les problèmes 8, 19 et 20 se rapportent à ce chapitre.

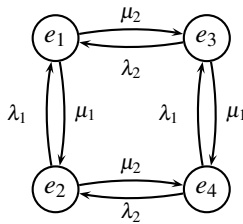
9.4 Corrigés des exercices

Exercice 9.1

(1) Calculons $P(X_{t+h} = e_j \mid X_t = e_i)$ quand h est petit par rapport à t ; on en déduit par dérivation les taux de transition instantanée.

Par exemple, $P(X_{t+h} = (1,0) \mid X_t = (0,0))$ désigne la probabilité que le composant (1) tombé en panne à l'instant t soit en état de marche en $t+h$. Elle est égale à $\int_0^h \mu_1 e^{-\mu_1 u} du \approx \mu_1 h$, d'où un taux de probabilité égal à μ_1 . Le calcul des autres probabilités de transition s'effectue de la même façon. On obtient le générateur :

$$A = \begin{pmatrix} -(\mu_1 + \mu_2) & \mu_1 & \mu_2 & 0 \\ \lambda_1 & -(\lambda_1 + \mu_2) & 0 & \mu_2 \\ \lambda_2 & 0 & -(\lambda_2 + \mu_1) & \mu_1 \\ 0 & \lambda_2 & \lambda_1 & -(\lambda_1 + \lambda_2) \end{pmatrix}.$$

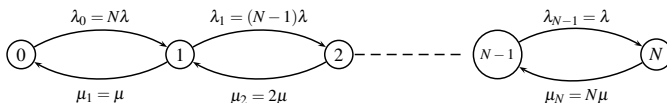


(2) Ce processus de Markov continu est irréductible et récurrent positif : il admet donc une unique distribution stationnaire π solution de l'équation d'équilibre $\pi A = 0$. On obtient $\pi_{0,0} = \frac{\lambda_1 \lambda_2}{D}$; $\pi_{0,1} = \frac{\lambda_1 \mu_2}{D}$; $\pi_{1,0} = \frac{\mu_1 \lambda_2}{D}$; $\pi_{1,1} = \frac{\mu_1 \mu_2}{D}$ où

$$D = (\lambda_1 + \mu_1)(\lambda_2 + \mu_2).$$

Exercice 9.3

(1) $(X_t)_{t \geq 0}$ est un processus de Markov continu dont les seules transitions se font d'un état quelconque vers les états contigus ; c'est donc un processus de naissance et de mort de diagramme de transition :



Taux de transitions : $\lambda_i = (N-i)\lambda$ et $\mu_i = i\mu$ pour $0 \leq i \leq N$. En effet :

« 0 » est l'état « aucune voiture n'est disponible » ; autrement dit, toutes sont en réparation et le taux de sortie d'une voiture des ateliers est donc égal à $N\lambda$. Lorsqu'il y a une seule voiture en marche (état 1), elle peut tomber en panne avec un taux égal à μ .

$$\begin{aligned}(2) \quad \pi_0 &= \left(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \cdots + \frac{\lambda_0\cdots\lambda_{N-1}}{\mu_0\cdots\mu_N}\right) - 1 \\ &= \left(C_N^0 + \frac{\lambda}{\mu}C_N^1 + \cdots + \left(\frac{\lambda}{\mu}\right)^N \frac{N}{1} \frac{N-1}{2} \cdots \frac{1}{N}\right)^{-1} = \left(\frac{\mu}{\mu+\lambda}\right)^N,\end{aligned}$$

d'où par définition des π_j , $\pi_1 = \pi_0 \frac{\lambda_0}{\mu_1} = \left(\frac{\mu}{\mu+\lambda}\right)^N \frac{N\lambda}{\mu}$ et $\pi_j = C_N^j \left(\frac{\mu}{\lambda+\mu}\right)^{N-j} \left(\frac{\lambda}{\lambda+\mu}\right)^j$.

La distribution stationnaire est donc de loi binomiale $\mathcal{B}\left(\frac{\lambda}{\lambda+\mu}; N\right)$.

Exercice 9.4

(1) On peut trouver $p_n(t)$ par récurrence ascendante sur n (la matrice A est triangulaire supérieure).

$p'_0(t) = 0$ donc $p_0(t) = p_0(0) = 0$.

$p'_1(t) = -\lambda p_1(t)$ donc $p_1(t) = Ce^{-\lambda t}$ avec $C = p_1(0) = 1$.

Supposons que $p_{n-1}(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-2}$ pour $n \geq 2$. On a

$$p'_n(t) + n\lambda p_n(t) = (n-1)\lambda e^{-\lambda t}(1 - e^{-\lambda t})^{n-2}.$$

C'est une équation différentielle linéaire du premier ordre avec second membre ; pour la résoudre, on applique la méthode de la variation de la constante. $p_n(t) = C(t)e^{-n\lambda t}$ avec $C'(t)e^{-n\lambda t} = (n-1)\lambda e^{-\lambda t}(1 - e^{-\lambda t})^{n-2}$.

$$\begin{aligned}C'(t) &= (n-1)\lambda e^{(n-1)\lambda t}(1 - e^{-\lambda t})^{n-2} \\ &= (n-1)\lambda e^{\lambda t}(e^{\lambda t} - 1)^{n-2} = \frac{d}{dt} \left((e^{\lambda t} - 1)^{n-1} \right),\end{aligned}$$

donc $C(t) - C(0) = (e^{\lambda t} - 1)^{n-1}$ et $C(0) = p_n(0) = 0$ pour $n \geq 2$. Finalement,

$$p_n(t) = (e^{\lambda t} - 1)^{n-1} e^{-n\lambda t} = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}.$$

X_t suit donc la loi géométrique de paramètre $e^{-\lambda t}$.

(2) $p'_N(t) + N\mu p_N(t) = 0$ donc $p_N(t) = Ce^{-N\mu t}$ avec $C = p_N(0) = 1$.

Supposons maintenant que $p_{n+1}(t) = C_N^{n+1}(e^{-\mu t})^{n+1}(1 - e^{-\mu t})^{N-n-1}$ pour $n \in \{1, \dots, N-1\}$.

$$p'_n(t) + n\mu p_n(t) = (n+1)\mu C_N^{n+1}(e^{-\mu t})^{n+1}(1 - e^{-\mu t})^{N-n-1}.$$

On applique la méthode de la variation de la constante : $p_n(t) = C(t)e^{-n\mu t}$.

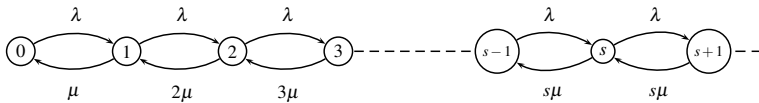
$$\begin{aligned} C'(t) &= e^{n\mu t} \times (n+1)\mu \frac{N!}{(n+1)!(N-n-1)!} e^{-(n+1)\mu t} (1 - e^{-\mu t})^{N-n-1} \\ &= (N-n)\mu C_N^n e^{-\mu t} (1 - e^{-\mu t})^{N-n-1} = \frac{d}{dt} (C_N^n (1 - e^{-\mu t})^{N-n}), \end{aligned}$$

donc $C(t) - C(0) = C_N^n (1 - e^{-\mu t})^{N-n}$ avec $C(0) = p_n(0) = 0$ pour $n < N$ et

$$p_n(t) = C_N^n e^{-n\mu t} (1 - e^{-\mu t})^{N-n}.$$

X_t suit donc la loi binomiale $\mathcal{B}(N, e^{-\mu t})$.

Exercice 9.5



$$(1) \text{ Équations bilan : } \begin{cases} \lambda \pi_0 = \mu \pi_1 ; \\ (\lambda + k\mu) \pi_k = (k+1)\mu \pi_{k+1} + \lambda \pi_{k-1} \text{ pour } 0 < k < s ; \\ (\lambda + s\mu) \pi_k = s\mu \pi_{k+1} + \lambda \pi_{k-1} \text{ pour } k \geq s. \end{cases}$$

$$(2) \begin{cases} \pi_k = \frac{\rho^k}{k!} \pi_0 \text{ pour } k = 0, 1, \dots, s \\ \pi_k = \frac{\rho^k}{s! s^{k-s}} \pi_0 \text{ pour } k \geq s+1 \end{cases} \quad \text{avec } \rho = \frac{\lambda}{\mu} \text{ et}$$

$$\pi_0 = \left(\sum_{k=0}^{s-1} \frac{\rho^k}{k!} + \left(1 - \frac{\rho}{s}\right)^{-1} \frac{\rho^s}{s!} \right)^{-1}.$$

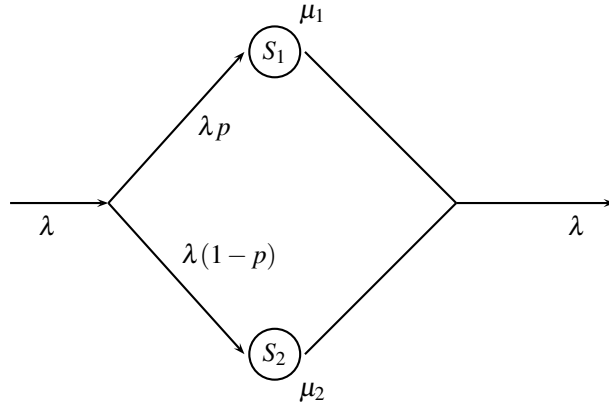
(3)

$$\begin{aligned} L_q &= \sum_{k=s+1}^{+\infty} (k-s) \pi_k = \frac{s^s}{s!} \pi_0 \sum_{k \geq s+1} (k-s) \left(\frac{\rho}{s}\right)^k \\ &= \frac{s^s}{s!} \pi_0 \sum_{m=1}^{+\infty} \left(\frac{\rho}{s}\right)^{m+s} = \left(\frac{\rho}{s}\right)^{s+1} \frac{s^s}{s!} \left(1 - \frac{\rho}{s}\right)^{-2} \\ &= \frac{\rho^{s+1}}{(s-1)!} \frac{\pi_0}{(s-\rho)^2} \end{aligned}$$

et $W_q = \frac{L_q}{\lambda}$. On retrouve les caractéristiques du système d'attente $M/M/1$ en prenant $s = 1$.

Exercice 9.7

(1) Réseau en parallèle :



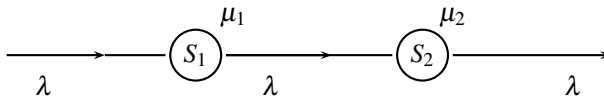
$$\lambda_i = \lambda p_{0,i}.$$

$$e_1 = p_{0,1} \text{ et } e_2 = p_{0,2} = 1 - p; L_1 = \frac{\lambda p}{\mu_1 - \lambda p} \text{ et } L_2 = \frac{\lambda(1-p)}{\mu_2 - \lambda(1-p)}.$$

$$L = L_1 + L_2; W = \frac{L}{\lambda}.$$

Pour μ_1 et μ_2 connus, il est possible de déterminer p , c'est-à-dire d'ajuster le flux vers chacun des services de façon à minimiser $W(p)$.

(2) Réseau en série :



$$L = \frac{\lambda}{\mu_1 - \lambda} + \frac{\lambda}{\mu_2 - \lambda}.$$

Exercice 9.8

$$(1) \begin{cases} e_1 = 1 + e_2 \\ e_2 = pe_1 \end{cases} \text{ donc } e_1 = \frac{1}{1-p} \text{ et } e_2 = \frac{p}{1-p}; \lambda_1 = \lambda e_1 \text{ et } \lambda_2 = \lambda e_2.$$

$$(2) (\lambda_1 < 2 \text{ et } \lambda_2 < 1) \iff \lambda < \min\left(2(1-p); \frac{1-p}{p}\right).$$

$$L_1 = \frac{\lambda_1}{2 - \lambda_1} = \frac{\lambda}{2 - 2p - \lambda}; L_2 = \frac{\lambda p}{1 - p - \lambda p}.$$

$$(3) W_1 = 2.$$

(4) Le réseau étant fermé, la loi stationnaire est définie par :

$$\forall (n_1, n_2) \ (n_1 + n_2 = n) \ P([X_1 = n, X_2 = n_2]) = C. \left(\frac{e_1}{\mu_1}\right)^{n_2} \left(\frac{e_2}{\mu_2}\right)^{n-n_1}.$$

Soit $e_1 = 1$ alors $e_2 = 1$; $P([X_1 = n_1]) = P([X_1 = n, X_2 = n - n_1]) = C \frac{1}{\mu_2^n} \left(\frac{\mu_2}{\mu_1} \right)^{n_1}$,
d'où

$$C = \frac{\mu_1 - \mu_2}{\mu_1^{n+1} - \mu_2^{n+1}} (\mu_1 \mu_2)^n.$$

Chapitre 10

Processus du second ordre

Ce chapitre est une introduction à la théorie des processus et des signaux aléatoires du second ordre complétée par une introduction à la théorie des processus browniens à la notion de bruit blanc, présents dans de nombreux domaines allant de la physique à l'économie.

10.1 Généralités

Définition 10.1 *Processus du second ordre*

Un processus aléatoire à valeurs complexes ou réelles $(X_t)_{t \in \mathbb{T}}$, où \mathbb{T} est un sous-ensemble de temps inclus dans \mathbb{R} , est du **second ordre** si sa puissance instantanée en un temps t quelconque $E(X_t \cdot \overline{X_t}) =_{(noté)} P_t$ est finie pour tout temps $t \in \mathbb{T}$.

Caractéristiques des processus du second ordre :

Espérance m_t au temps t : $E(X_t) = m_t$.

Variance σ_t^2 au temps t : $V(X_t) = \sigma_t^2$.

Puissance P_t instantanée au temps t : $P_t = E(X_t \cdot \overline{X_t}) = \sigma_t^2 + |m_t|^2$.

Autocovariance $R_X(s, t)$ ou $R(s, t)$:

$$\forall s, t \ (s < t) \ R_X(s, t) = \text{cov}(X_s, X_t) = E(X_s \cdot \overline{X_t}) - E(X_s) \cdot E(\overline{X_t}).$$

Autocorrélation :

$$\frac{R_X(s, t)}{\sigma_s \cdot \sigma_t} =_{(noté)} \rho_X(s, t).$$

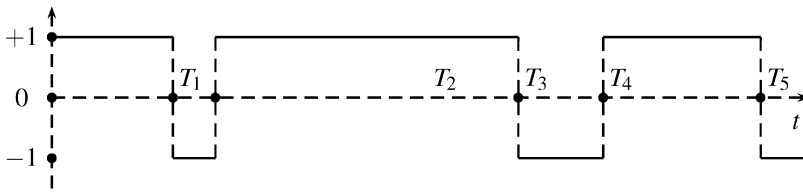
L'autocorrélation $\rho_X(s, t)$ entre les états X_s et X_t du processus nous renseigne sur le degré de dépendance linéaire de X_t par rapport à X_s : si $\rho_X(s, t)$ est proche de $+1$ ou -1 , la corrélation entre les états X_s et X_t est forte, inversement si $\rho_X(s, t)$ est proche de 0 , les états sont très peu corrélés.

Si \mathbb{T} est un intervalle de \mathbb{R} , le processus est dit **analogique** ; si \mathbb{T} est un sous-ensemble au plus dénombrable de \mathbb{R} , par exemple \mathbb{N} , le processus est dit **discret**.

Exemple 10.1 Le processus du télégraphe

Étant donné le processus de Poisson $(N_t)_{t \in \mathbb{R}^+}$ d'intensité λ , on définit le processus du télégraphe $(Z_t)_t$ par $Z_t = (-1)^{N_t}$.

Graphes d'une réalisation démarrant en $Z_0 = 1$



- La moyenne $E(Z_t) = m_t = e^{-2\lambda t}$.

Preuve : $m_t = 1 \times P(Z_t = 1) + (-1) \times P(Z_t = -1)$.

$P(Z_t = 1) = \sum_{k \in \mathbb{N}} P(N_t = 2k) = e^{-\lambda t} \cosh(\lambda t)$;

$P(Z_t = -1) = \sum_{k \in \mathbb{N}} P(N_t = 2k + 1) = e^{-\lambda t} \sinh(\lambda t)$.

Par sommation, on aboutit à : $m_t = e^{-2\lambda t}$.

- La puissance $E(Z_t^2) = 1$ pour tout t .

Preuve : $Z_t^2 = 1$ d'où $E(Z_t^2) = 1$.

- L'autocovariance $R(s, t) = e^{-2\lambda(t-s)}$.

$$\begin{aligned} \text{Preuve : } R(s, t) &= E((-1)^{N_s + N_t}) \\ &= E((-1)^{2N_s + (N_t - N_s)}) = E((-1)^{N_t - N_s}) \\ &= E((-1)^{N_{t-s}}) = m_{t-s} = e^{-2\lambda(t-s)}. \end{aligned}$$

L'autocovariance $R(s, t)$ décroît si $(t - s)$ croît ; pratiquement, on considère que la valeur de Z_t est décorrélée de Z_s lorsque $(t - s)$ a une valeur supérieure à $\frac{3}{\lambda}$.

Définition 10.2 Un processus du second ordre est **centré** si pour tout temps t , sa moyenne m_t est nulle.

Définition 10.3 Une suite de vecteurs aléatoires $(X_n)_{n \in \mathbb{N}}$ converge en moyenne quadratique vers le vecteur aléatoire X si $\lim_{n \rightarrow +\infty} E(\|X_n - X\|^2) = 0$.

On rappelle qu'une suite convergente en moyenne quadratique converge en probabilité et donc en loi.

Définition 10.4 Un processus du second ordre à temps continu $(X_t)_t$ est continu au temps t en moyenne quadratique, si la suite $(X_{t+h_n})_n$ converge en moyenne quadratique vers X_t pour toute suite (h_n) convergente vers 0 :

$$\lim_{n \rightarrow +\infty} E(|X_{t+h_n} - X_t|^2) = 0.$$

Les réalisations d'un processus continu en moyenne quadratique ne sont pas nécessairement continues partout au sens des fonctions. Il en est ainsi du processus de Poisson qui est un processus du second ordre continu en moyenne quadratique, dont les réalisations sont, comme on le sait, des fonctions étagées discontinues aux temps d'occurrence.

10.2 Processus stationnaires du second ordre

De nombreux processus intervenant dans les sciences de l'ingénieur ou en physique sont tels que leur autocovariance $R(s, t)$ ne dépend que de la durée $(t - s)$: si de plus leur espérance est constante, ces processus sont dits faiblement stationnaires conformément à la définition suivante.

Définition 10.5 Un processus du second ordre $(X_t)_{t \in \mathbb{R}}$ est **stationnaire au sens large ou faiblement stationnaire** si :

(1) Pour tout temps t , son espérance $E(X_t)$ est constante,

$$E(X_t) = m ;$$

(2) Pour tout couple de temps (s, t) , l'autocovariance $R(s, t)$ du processus ne dépend que de $(t - s)$; on la notera $R(t - s)$.

Théorème 10.1 La puissance instantanée P_t au temps t d'un processus faiblement stationnaire est constante :

$$\forall t, P_t = R(0) + m^2.$$

Preuve $E(X_t \overline{X_t}) = \text{cov}(X_t, \overline{X_t}) + E(X_t) E(\overline{X_t}) = R(t - t) + m^2 = R(0) + m^2$. \square

Désormais, les processus du second ordre étudiés dans ce chapitre seront supposés être stationnaires au sens large et symbolisés par l'expression SSL.

Théorème 10.2 (Propriétés de l'autocovariance des processus SSL)

- (1) *Parité* : pour tout $u \in \mathbb{T}$, $R(-u) = \overline{R(u)}$.
 (2) *L'autocovariance est majorée par la puissance* : $\forall u, |R(u)| \leq R(0)$.
 (3) *L'autocovariance est une application du type positif* : $\forall n \in \mathbb{N}^*, \forall t_1, \dots, t_n$

tels que $t_1 < t_2 < \dots < t_n$ et $\forall a_1, \dots, a_n \in \mathbb{C}$:

$$\sum_{i=1}^n \sum_{j=1}^n a_i \overline{a_j} R(t_i - t_j) \geq 0.$$

Preuve

- (1) $R(-u) \underset{(\text{déf})}{=} E(X_{t-u} \overline{X_t}) = E(X_t \overline{X_{t+u}}) = \overline{R(u)}$.
 (2) $|E(X_{t+u} \overline{X_t})|^2 \leq_{(\text{Cauchy-Schwarz})} E(|X_{t+u}|^2) E(|\overline{X_t}|^2)$ donc $|R(u)|^2 \leq |R(0)|^2$.
 (3) $\sum_{i=1}^n \sum_{j=1}^n a_i \overline{a_j} R(t_i - t_j) = E(\sum_i a_i X_{t_i} \sum_j \overline{a_j} \overline{X_{t_j}}) = E(|\sum_i a_i X_{t_i}|^2) \geq 0$. \square

Définition 10.6 Un processus du second ordre est dit **asymptotiquement stationnaire au sens large**, si lorsque t tend vers $+\infty$, son espérance m_t converge vers une constante m et si son autocovariance converge lorsque s et t tendent vers $+\infty$, vers une fonction de la forme $R(t-s)$.

Le processus du télégraphe étudié dans l'exemple 10.1 est un processus asymptotiquement SSL.

Il existe un autre type de stationnarité dite stationnarité forte, qui entraîne la stationnarité faible, mais n'a pas d'intérêt pratique dans le cadre de ce cours.

Définition 10.7 Un processus du second ordre $(X_t)_{t \in \mathbb{T}}$ est **fortement stationnaire** si $\forall \tau > 0, \forall m \in \mathbb{N}^*, \forall t_1, \dots, t_m \in \mathbb{T}$:

$$(X_{t_1}, \dots, X_{t_m}) \text{ et } (X_{t_1+\tau}, \dots, X_{t_m+\tau}) \text{ ont même loi.}$$

10.3 Propriétés spectrales des processus SSL

L'étude spectrale des processus SSL s'apparente à l'analyse harmonique des signaux ou des fonctions déterministes et conduit à la notion importante de densité spectrale.

Théorème 10.3 (de Wiener-Khintchine)

La **densité spectrale** S d'un processus SSL $(X_t)_t$ à temps discret (respectivement : à temps continu) de covariance R , est la série (respectivement : la transformée) de Fourier de son autocovariance. On distingue trois cas :

(1) Temps entier :

$$S(\nu) = \sum_{k=-\infty}^{k=+\infty} R(k) e^{-2i\pi\nu k} \text{ et } \forall k \in \mathbb{Z}, \quad R(k) = \int_{-\frac{1}{2}}^{\frac{1}{2}} S(\nu) e^{2i\pi\nu k} d\nu.$$

(2) Temps continu et fonction de covariance R intégrable :

$$S(\nu) = \int_{-\infty}^{+\infty} R(\tau) e^{-2i\pi\nu\tau} d\tau \text{ et } R(\tau) = \int_{-\infty}^{+\infty} S(\nu) e^{2i\pi\nu\tau} d\nu.$$

Dans le cas d'un processus réel :

$$R \text{ est symétrique et } S(\nu) = 2 \int_0^{+\infty} R(\tau) \cos(2\pi\nu\tau) d\tau.$$

(3) Temps continu et fonction de covariance R défini par une distribution tempérée :

$$S = \mathcal{F}(R) \quad (\text{au sens des distributions tempérées}).$$

10.1 ♠♠ Spectre du processus harmonique $(\sum_{k=1}^n A_k e^{(2i\pi\nu_k t)})_t$

Étant donné les v.a. A_k centrées, décorrélées, de variances σ_k^2 et les fréquences ν_k constantes. Démontrer que :

- (1) pour tout t : $E(X_t) = 0$.
- (2) pour tous t_1, t_2 : $E(X_{t_1} X_{t_2}) = E(X_{t_1} \overline{X_{t_2}}) = \sum_{k=1}^n \sigma_k^2 e^{(2i\pi\nu_k(t_1 - t_2))}$ et le processus est donc stationnaire.
- (3) la densité spectrale est définie par la distribution $S(\nu) = \sum_{k=1}^n \sigma_k^2 \delta_{\nu_k}$ qui caractérise le spectre de raies de support fréquentiel $\{\nu_1, \dots, \nu_n\}$; σ_k^2 étant la puissance moyenne du signal $e^{2i\pi\nu_k t}$.

Attention ! Des processus SSL peuvent avoir la même densité spectrale et être différents.

(1) Le signal du téléphone et le signal-réponse obtenu par filtrage RC d'un bruit blanc ont même densité spectrale :

$$f(\nu) = \frac{\lambda}{\lambda^2 + \nu^2 \pi^2}.$$

(2) Le bruit blanc gaussien et le bruit blanc poissonnien constitué les distributions de Dirac dont les supports sont les temps d'occurrence d'un processus poissonnien, ont même le spectre constant.

10.2 ♠♠ Étude spectrale du signal de transmission binaire

Ce signal est décrit par le processus aléatoire $(X_t)_{t \in \mathbb{R}_+}$, qui prend les valeurs $+1$ et -1 sur les intervalles $I_n = [(n-1)a, na]$ avec la probabilité $\frac{1}{2}$ ($a > 0$).

- (1) Tracer une réalisation du processus.
- (2) Le processus est-il SSL ?
- (3) Soit le processus $(Y_t)_{t \in \mathbb{R}_+}$ défini par $Y_t = X_{t-\theta}$ où θ est la v.a. uniforme sur $[0, 1]$; démontrer que $(Y_t)_{t \in \mathbb{R}_+}$ est SSL. Déterminer sa densité spectrale.

10.4 Processus gaussiens stationnaires

Ces processus constituent une extension de la notion de vecteur gaussien à des familles infinies de composantes.

Définition 10.8 *Un processus $(X_t)_{t \in \mathbb{R}_+}$ réel est gaussien si :*

$$\forall n, \forall t_1, \dots, t_n \in \mathbb{R}_+, \text{ le vecteur } (X_{t_1}, \dots, X_{t_n}) \text{ est gaussien.}$$

Par définition un processus gaussien est nécessairement du second ordre, car pour tout i , $E(X_i^2)$ est égal à $V(X_i) + E(X_i)^2$, qui est fini.

Théorème 10.4 (Caractérisation d'un processus gaussien SSL)

Un processus gaussien $(X_t)_{t \in \mathbb{R}_+}$ est SSL si et seulement si :

- (1) $\forall t > 0$, $E(X_t)$ est une constante égale à m ;
- (2) $\forall (t_1, t_2, \dots, t_n)$, la matrice de covariance :

$$\text{Cov}(X_{t_1+h}, \dots, X_{t_n+h}) \text{ est indépendante de } h.$$

Preuve $\forall i, j$ $E(X_{t_i+h}X_{t_j+h}) = \text{cov}(X_{t_i+h}, X_{t_j+h}) + m^2$ est indépendant de h . Inversement, $R(s, t) = \text{cov}(X_s, X_t) = \text{cov}(X_0, X_{t-s})$ fonction de $(t-s)$. \square

Remarque 10.1 Un processus gaussien faiblement stationnaire l'est aussi au sens strict et c'est le seul processus à vérifier cette propriété, analogue à la propriété d'équivalence pour la classe des vecteurs gaussiens, entre l'indépendance des composantes d'un vecteur et leur décorrélation.

10.3 ♠♠ Construction de processus gaussiens classiques

(1) Si X est une v.a. $\mathcal{N}(0, \sigma^2)$ et f une fonction bornée de \mathbb{R}_+ dans \mathbb{R} , démontrer que $(f(t)X)_{t \in \mathbb{R}_+}$ est un processus gaussien.

(2) Démontrer que si $(X_t)_t$ est un processus gaussien, la v.a. Y définie par $\int_a^b \phi(t)X(t)dt$, où ϕ est une fonction continue, est gaussienne. [Indication : commencer par faire une approximation en série de l'intégrale.]

10.5 Processus ergodiques

Il arrive que l'on ne dispose que d'une seule réalisation d'un processus pendant une longue durée. Quelles doivent être les propriétés de ce processus pour que l'on puisse en estimer de façon satisfaisante ses caractéristiques statistiques (moyenne, puissance et autocovariance) à l'aide de cette unique réalisation ?

Définition 10.9 *Caractéristiques temporelles d'un processus*

On suppose que le processus $(X_t)_t$ est indexé par des temps variant dans \mathbb{R} , la transposition des résultats ci-dessous au cas de \mathbb{R}_+ étant immédiate. On suppose de plus que pour toute réalisation $(X_t(\omega))_t$ du processus, la v.a. $\phi_T(\omega) = \frac{1}{2T} \int_{-T}^T X_t(\omega) \cdot dt$ est bien définie (ce qui est le cas si la fonction $t \mapsto X_t(\omega)$ est continue) et converge p.s quand T tend vers $+\infty$, vers une v.a. $m_X(\omega)$, appelée **moyenne temporelle** de la réalisation. On suppose enfin que pour tout ω , $\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T (X_{t+\tau}(\omega) - m_X(\omega))(X_t(\omega) - m_X(\omega))^* dt$ existe, notée $R_\omega(\tau)$ et appelée **autocovariance temporelle** de la réalisation $(X_t(\omega))_t$.

La **moyenne temporelle** $m_X(\omega)$ de la réalisation $(X_t(\omega))_t$ du processus est la v.a. définie par :

- (1) $\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T X_t(\omega) dt$ dans le cas d'un processus à temps continu ;
- (2) $\lim_{N \rightarrow +\infty} \frac{1}{2N+1} \sum_{k=-N}^{k=N} X_k(\omega)$ dans le cas d'un processus à temps entier.

L'**autocovariance temporelle** $R_\omega(\tau)$ de la réalisation $(X_t(\omega))_t$ est définie par :

- (1) $\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T (X_{t+\tau}(\omega) - m_X(\omega))(X_t(\omega) - m_X(\omega))^* dt$ dans le cas d'un processus à temps continu ;
- (2) $\lim_{N \rightarrow +\infty} \frac{1}{2N+1} \sum_{k=-N}^{k=N} (X_{k+\tau}(\omega) - m_X(\omega))(X_k(\omega) - m_X(\omega))^*$ dans le cas d'un processus à temps entier.

Définition 10.10 Un processus SSL $(X_t)_t$ est **faiblement ergodique** par rapport à la moyenne si la moyenne temporelle $m_X(\omega)$ d'une réalisation $(X_t(\omega))_t$ quelconque est égale (p.s.) à sa moyenne m :

$$P(m = m_X(\omega)) = 1.$$

Théorème 10.5 Un processus stationnaire $(X_t)_t$ est **faiblement ergodique** par rapport à la moyenne si :

$$V(\overline{X_\omega}) = \lim_{T \rightarrow +\infty} V\left(\frac{1}{2T} \int_{-T}^T X_t(\omega) dt\right) = 0.$$

Preuve Si $\lim_{T \rightarrow +\infty} V(\frac{1}{2T} \int_{-T}^+ X_t(\omega) dt) = 0$, alors la v.a. $\frac{1}{2T} \int_{-T}^+ X_t(\omega) dt$ converge vers sa moyenne. \square

À l'instar de la propriété de stationnarité, l'ergodicité admet deux versions : l'une au sens faible, qui est la plus pratiquée, l'autre au sens strict (ou fort), impliquant la précédente.

Définition 10.11 Un processus SSL $(X_t)_t$ est **ergodique au sens strict** si, pour tout n , toute suite de temps (t_1, \dots, t_n) et toute fonction ϕ de \mathbb{R}^n dans \mathbb{R} telle que $E(|\phi(X_{t_1}, \dots, X_{t_n})|)$ soit finie, alors :

$$\frac{1}{2N+1} \sum_{i=-N}^N \phi(X_{t_1+i}, \dots, X_{t_n+i}) \text{ converge p.s. vers } E(\phi(X_{t_1}, \dots, X_{t_n})).$$

Le théorème suivant donne deux conditions suffisantes d'ergodicité faible, faciles à vérifier et très utiles.

Théorème 10.6 Un processus SSL $(X_t)_t$ est faiblement ergodique par rapport à la moyenne si l'une des deux conditions est vérifiée :

- (1) $\int_{-\infty}^{+\infty} |R(t)| dt$ est finie ;
- (2) $R(0)$ est fini et $\lim_{|t| \rightarrow +\infty} R(t) = 0$.

Exemple 10.2 (1) Les suites de v.a. $(X_n)_n$ indépendantes de même loi et les processus gaussiens, stationnaires du second ordre, tels que $\lim_{\tau \rightarrow +\infty} R(\tau) = 0$ et $R(0)$ est fini, sont ergodiques.

(2) Le processus harmonique $(\sum_{k=1}^n A_k(\omega) e^{(2i\pi\nu_k t)})_t$ dont l'autocovariance temporelle dépend de l'épreuve ω n'est pas ergodique.

10.4 ♠ Démontrer que le signal sinusoïdal à phase aléatoire, défini par

$X_t = \cos(2\pi\nu t + \Phi)$, où la phase Φ est une v.a. de loi uniforme sur $[0, 2\pi]$, est un SSL faiblement ergodique par rapport à la moyenne.

10.6 Processus de Wiener ou brownien et bruit blanc

Le processus de Wiener porte le nom du mathématicien américain qui en donna une définition axiomatique en 1923. Il est l'aboutissement de la modélisation mathématique du mouvement brownien mis en évidence un siècle auparavant par le botaniste écossais Brown, grâce aux observations qu'il fit du mouvement erratique de particules de pollen s'agitant à la surface d'un liquide

chauffé. C'était le début d'une saga physico-mathématique particulièrement riche dont on donne un aperçu en fin de chapitre. Le processus de Wiener est un processus stationnaire à accroissements indépendants (voir la définition des processus de Poisson au 8.1) caractérisé par la normalité de ses élongations.

Un processus à accroissements indépendants noté PAI, et PAIS s'il est de plus SSL, vérifie la propriété de Markov (voir : théorème 9.4).

Théorème 10.7 *Tout PAIS $(X_t)_{t \in \mathbb{R}}$ centré, tel que $X_0 = 0$, vérifie les propriétés :*

- (1) $\forall t : V(X_t) = K |t| \quad (K > 0) ;$
- (2) $\text{cov}(X_s, X_t) = \begin{cases} K \cdot \min(|s|, |t|) & \text{si } s, t > 0 ; \\ 0 & \text{si } s, t < 0 ; \end{cases}$
- (3) $\forall s, t \text{ tels que } t > s > 0 : \quad \rho_{(s,t)} = \sqrt{\frac{s}{t}}.$

Preuve (1) $\forall t_1, t_2 > 0$,

$$\begin{aligned} V(X_{t_1+t_2}) &= V(X_{t_1+t_2} - X_{t_1} + X_{t_1} - X_0) \\ &= (\text{indépendance des acc}) V(X_{t_1+t_2} - X_{t_1}) + V(X_{t_1} - X_0) \\ &= (\text{stationnarité}) V(X_{t_2}) + V(X_{t_1}) ; \end{aligned}$$

si on pose $V(X_t) = f(t)$, l'équation précédente s'écrit :

(*) $f(t_1 + t_2) = f(t_1) + f(t_2)$, où f est positive donc par (*), croissante. La seule fonction solution de (*) est linéaire : $f(t) = Kt$. Un raisonnement analogue pour des t_i négatifs permet de conclure.

(2) Pour calculer la covariance distinguons trois cas :

$$\begin{aligned} \text{(a) } 0 \leq s \leq t : \text{cov}(X_s, X_t) &= \text{cov}(X_s, (X_t - X_s) + X_s) \\ &= \text{cov}(X_s, (X_t - X_s)) + \text{cov}(X_s, X_s) \\ &= \text{cov}(X_s - X_0, X_t - X_s) + K \cdot s = K \cdot s \end{aligned}$$

car les accroissements $(X_s - X_0, X_t - X_s)$ étant indépendants ;

$$\text{(b) } s \leq t \leq 0 : \text{idem ; on parvient à } \text{cov}(X_s, X_t) = -Kt ;$$

$$\text{(c) } s \leq 0 \leq t : \text{cov}(X_s, X_t) = -\text{cov}(X_0 - X_s, X_t - X_0) = 0.$$

(3) est une conséquence de (1) et (2). □

Définition 10.12 *Le processus de Wiener $(W_t)_{t \geq 0}$ sur \mathbb{R}^+ est défini par les axiomes :*

- (1) $W_0 = 0$;
 (2) $(W_t)_{t \geq 0}$ est un PAIS ;
 (3) $\forall t, W_t$ est de loi $\mathcal{N}(0, \sigma^2 t)$. Si $\sigma = 1$, le processus de Wiener est dit *standard*.

Le théorème ci-dessous donne un ensemble de conditions suffisantes pour qu'un processus gaussien soit un processus de Wiener.

Théorème 10.8 *Un processus gaussien $(W_t)_{t \geq 0}$ est un processus de Wiener si :*

- (1) $\forall t \ E(W_t) = 0, V(W_t) = \sigma^2 t$;
 (2) $\forall s, t \ R(s, t) = \sigma^2 \min(s, t)$.

L'exercice ci-dessous montre qu'un processus de Wiener peut être approché asymptotiquement par une marche aléatoire symétrique.

10.5 ♠♠ Approximation du processus brownien à une dimension par la marche aléatoire symétrique

Soit un processus de Bernoulli $(X_n)_{n \in \mathbb{N}}$ où les X_n sont indépendantes de même loi : $\forall n \in \mathbb{N}, P(X_n = 1) = P(X_n = -1) = 0,5$. Pour tout $t \geq 0$, définissons $Z_t = \Delta x (X_1 + \dots + X_{[\frac{t}{\Delta t}]})$, où Δx et Δt sont des accroissements très petits de x et de t ; $[\frac{t}{\Delta t}]$ représente la partie entière de $\frac{t}{\Delta t}$.

- (1) Démontrer que le processus $(Z_t)_t$ est un PAIS.
 (2) Démontrer que Z_t converge en loi vers la v.a. normale $\mathcal{N}(0, \sigma^2 t)$ et conclure. [Indication : faire tendre Δx et Δt vers 0 de manière que $\frac{\Delta x}{\sqrt{\Delta t}} = \sigma$].
 D'un point de vue physique, $\sigma = \sqrt{2D}$ où D est le coefficient de diffusion de la particule brownienne dans son milieu.

On généralise le processus de Wiener aux espaces \mathbb{R}^2 et \mathbb{R}^3 . Par exemple, lorsque l'espace d'états est \mathbb{R}^2 , $(W_t)_t$ est le modèle d'évolution d'une particule évoluant à la surface d'un liquide, et soumise aux impacts incessants de ses voisines, qui comme on le sait sont fort nombreuses et dissipées. Les composantes sur chacun des axes d'un repère orthonormé d'un processus de Wiener de dimension quelconque dans tout repère sont aussi des processus de Wiener de dimension 1.

Théorème 10.9 (Propriétés du processus de Wiener)

- (1) *Le processus de Wiener est gaussien et ergodique par rapport à la moyenne.*

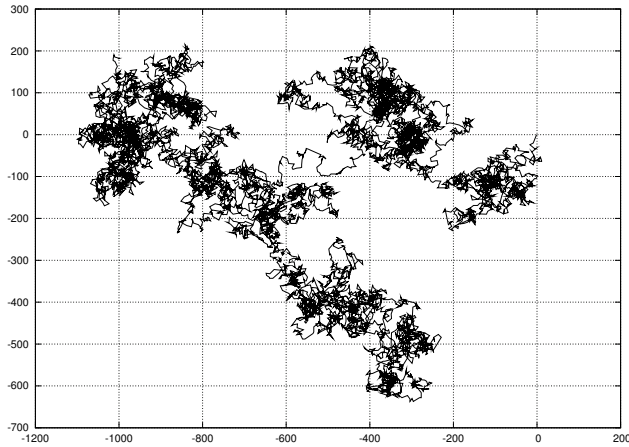


Figure 10.1 – Trajectoire d'un processus de Wiener isotrope d'origine 0 dans un réseau carré.

(2) Étant donné s, t tels que $t > s$, la densité conditionnelle de W_t sachant $W_s = x$, est définie par :

$$f_{W_t/(W_s=x)}(y) = \frac{1}{\sigma\sqrt{2\pi(t-s)}} \exp\left(-\frac{(y-x)^2}{2\sigma^2(t-s)}\right);$$

$$E(W_t|W_s = x) = x; V(W_t|W_s = x) = \sigma^2(t-s).$$

À partir de toute abscisse d'élongation x , atteinte au temps s , l'élongation du processus est de loi gaussienne. L'égalité $E(W_t|W_s = x) = x$ signifie que l'élongation moyenne de W_t , autour d'un point d'abscisse quelconque x atteint en un instant s précédent est nulle; il faut imaginer cela pour toutes les positions x atteintes par le processus. Si l'on pose $Z_t = W_{s+t}$, $(Z_t)_{t \leq 0}$ est le processus de Wiener démarrant au temps 0 en W_s égal à x . L'égalité $V(W_t|W_s = x) = \sigma^2(t-s)$ signifie que la dispersion moyenne quadratique autour de ce même point est fonction linéaire du temps $(t-s)$. Pour tout t , W_t se trouve à l'intérieur ou au voisinage de la parabole d'équation $x^2 = (1,96)^2 \sigma^2 t$, et la probabilité $P(|W_t| \leq 1,96 \sigma \sqrt{t})$ est approximativement égale à 0,95. (Voir la figure suivante)

(3) $(W_t)_t$ est continu (p.s.) et nulle part dérivable (p.s.).

(Les deux propriétés précédentes font de ce processus un candidat rêvé pour la galerie des notions mathématiques longtemps considérées comme « monstrueuses »; c'est ainsi que certains mathématiciens de la fin du XIX^e siècle, et non les moindres, concevaient les fonctions continues et non dérivables, dont des exemples fameux ont été exhibés par Weierstrass et Lebesgue.)

Preuve Montrons que $(W_t)_t$ est continu en probabilité :

$$\forall h > 0, \frac{1}{h} P(|W_{t+h} - W_t| > \varepsilon) < 2 \int_{\varepsilon}^{+\infty} \frac{e^{-\frac{ex}{2h}}}{\sqrt{2\pi h}^{\frac{3}{2}}} dx = 4 \frac{e^{-\frac{\varepsilon^2}{2h}}}{\varepsilon \sqrt{2\pi h}}$$

qui tend vers 0, quand h tend vers 0. Il est plus délicat d'établir qu'il est continu presque sûrement.

Le processus n'est pas dérivable en probabilité ; en effet, soit $Z = \frac{W_{t+h} - W_t}{h}$, Z est une v.a. $\mathcal{N}(0, \frac{\sigma^2}{h})$. $\forall a > 0$, $P(|Z| > a) = P\left(\left|\frac{W_h}{\sigma\sqrt{h}}\right| > \frac{a}{\sigma}\sqrt{h}\right)$, avec $\frac{W_h}{\sigma\sqrt{h}} \sim \mathcal{N}(0, 1)$. Donc, pour tout a $P(|Z| > a)$ converge vers 1 si h tend vers 0 ; d'où la non dérivabilité en probabilité, et par conséquent la non dérivabilité (p.s). Cependant on démontre que les réalisations du processus de Wiener sont dérivables au sens des distributions et que leurs dérivées sont des bruits blancs gaussiens, succinctement présentés à la fin du paragraphe. □

Le processus de Wiener sur l'axe réel est récurrent ; le processus de Wiener bidimensionnel revisite tout voisinage d'un point, aussi petit soit-il, une infinité de fois. Dans le cas tridimensionnel, la propriété précédente n'est pas vérifiée.

Sur la figure 10.2, les graphes dans l'espace de phase, d'une quarantaine de réalisations d'un processus de Wiener de moyenne nulle, qui confirment visuellement l'accroissement moyen de $V(W_t)$ en fonction du temps ainsi que le caractère asymptotiquement parabolique de l'enveloppe des trajectoires.

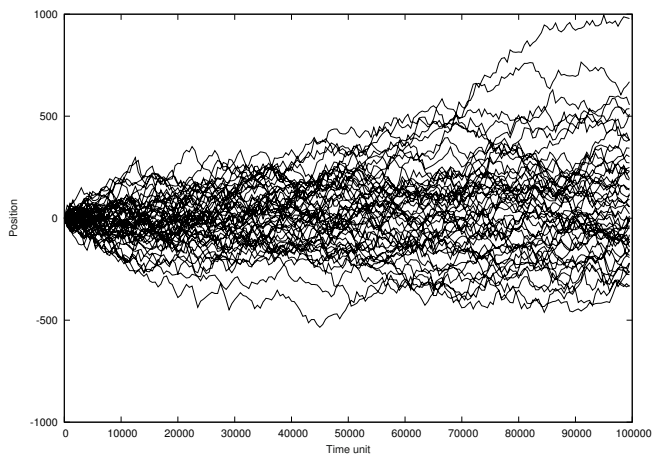


Figure 10.2 – Réalisations d'un processus de Wiener.

10.6 ♠ La densité du processus brownien est solution de l'équation de la chaleur.

Considérons une barre rectiligne de section homogène petite par rapport à la longueur. La chaleur accumulée au temps $t = 0$, dans un petit voisinage d'un point donné distinct des extrémités, résulte de l'agitation thermique d'une multitude de molécules excitées. On sait que la température $T(x, t)$ en un point x au temps t est solution de l'équation de la chaleur : $\frac{\partial T}{\partial t} = \frac{1}{k} \cdot \frac{\partial^2 T}{\partial x^2}$.

Démontrer que la densité conditionnelle $f_{W_t/(W_s=x)}$ du processus brownien est solution de l'équation de la chaleur (la constante k s'identifiant à σ^2). Trouver une contrepartie physique à cette preuve.

Définition 10.13 Le processus $(W_t^D)_{t \geq 0}$ défini par $W_t^D = W_t + \mu t$ est appelé processus de Wiener avec dérive, μ étant le paramètre de dérive.

Théorème 10.10 (Propriétés du processus de Wiener avec dérive)

- (1) Pour tout $t \geq 0$, $W_t^D \sim \mathcal{N}(\mu t, \sigma^2 t)$.
- (2) $(W_t^D)_t$ est un PAIS gaussien.
- (3) Sa fonction d'autocovariance $R(s, t)$ est celle de $(W_t)_{t \geq 0}$.

10.7 ♠♠ Processus brownien intégré

Cet exercice prolonge l'étude du processus gaussien intégré proposée dans l'exercice 10.3. Étant donné le processus brownien standard $(W_t)_t$ de variance $\sigma^2 = 1$, on définit le processus brownien intégré $(Z_t)_t$ de $(W_t)_t$ par :

$$\forall t \geq 0, Z_t = \int_0^t W_s ds.$$

(1) Démontrer que $(Z_t)_t$ est un processus gaussien non stationnaire de covariance $\text{cov}(Z_s, Z_t) = s^2(\frac{t}{2} - \frac{s}{6})$ si $(s < t)$. [Indication : on considérera que $\forall t, Z_t$ est égal à $\lim_{(n \rightarrow +\infty)} \frac{t}{n} \sum_{k=1}^n W(\frac{tk}{n})$, pour démontrer que tout vecteur $(Z_{t_1}, Z_{t_2}, \dots, Z_{t_n})$ est gaussien].

(2) Dédurre de l'expression de la covariance que $\text{cov}(Z_{u+v} - Z_u, Z_u) \neq 0$; en conclure que $(Z_t)_t$ ne vérifie pas la propriété d'indépendance des accroissements. Le processus brownien intégré n'est donc ni stationnaire ni markovien.

Définition 10.14 Un processus de diffusion $(X_t)_{t \in \mathbb{R}_+}$ est un processus à temps continu ayant un espace d'états continu E inclus dans \mathbb{R}^k , markovien et homogène, dont les trajectoires $t \rightarrow X_t$ sont continues presque partout et vérifient les hypothèses suivantes.

- (1) Existence de l'espérance instantanée :

$\forall x \in E, \forall t \in \mathbb{R}_+, \lim_{h \downarrow 0+} (\frac{1}{h} E(X_{t+h} - X_t) \mid X_t = x)$ existe, est notée $\mu(x, t)$ et représente le gradient de l'élongation moyenne au temps t , à partir de l'état x .

(2) *Existence de la variance instantanée :*

$\forall x \in E, \forall t \in \mathbb{R}_+, \lim_{h \downarrow 0^+} \left(\frac{1}{h} E((X_{t+h} - X_t)^2 \mid X_t = x) \right)$ existe, et est notée $\sigma^2(x, t)$.
 $(\mu(x, t)$ et $\sigma^2(x, t)$ sont supposées continues.)

(3) *Propriété de régularité :*

$\forall x \in E, \forall t \in \mathbb{R}_+, \forall \varepsilon > 0, \lim_{h \downarrow 0^+} \left(\frac{1}{h} P(|X_{t+h} - X_t| > \varepsilon \mid X_t = x) \right) = 0$.

Théorème 10.11 *La densité conditionnelle $f_{X_t/X_s=x}$ vérifie l'équation de Chapman-Kolmogorov :*

$\forall (s, u, t)$ réels vérifiant $s < u < t$ et $\forall (x, y) \in \mathbb{R}^k \times \mathbb{R}^k :$

$$f_{X_t/X_s=x}(y) = \int_{\mathbb{R}} f_{X_t/X_u=z}(y) \cdot f_{X_u/X_s=x}(z) dz. \quad (*)$$

Preuve Multiplions chacun des membres par $f_{X_s}(x)$; on obtient pour le premier membre $f_{(X_s, X_t)}(x, y)$ et pour le second membre :

$$\begin{aligned} \int f_{X_t/X_u=z}(y) \cdot f_{(X_s, X_u)}(x, z) \cdot dz &= \int f_{X_t/X_u=z, X_s=x}(y) \cdot f_{(X_s, X_u)}(x, z) \cdot dz \\ &= \int f_{X_s, X_u, X_t}(x, y, z) \cdot dz = f_{(X_s, X_t)}(x, y). \end{aligned}$$

Les deux expressions obtenues étant identiques pour tout x , on conclut à l'égalité (*). \square

Cette équation est analogue à l'équation qui porte le même nom, vérifiée par les chaînes de Markov à temps entier (voir théorème 7.5).

Théorème 10.12 (Équations de diffusion) *Soit $(X_t)_t$ est un processus de diffusion, la densité conditionnelle $f_{X_t/X_s=x}$ vérifie les équations de diffusion :*

$$\text{Équation directe ou de Fokker-Plank : } \frac{\partial f}{\partial t}(y) = -\frac{\partial(\mu(y, t)f)}{\partial y} + \frac{1}{2} \frac{\partial^2(\sigma^2(y, t)f)}{\partial y^2}.$$

$$\text{Équation inverse : } \frac{\partial f}{\partial s}(x) = -\mu(x, s) \frac{\partial f}{\partial x} - \frac{1}{2} \sigma^2(x, s) \frac{\partial^2 f}{\partial x^2}.$$

Exemple 10.3 Deux processus de diffusion classiques

(a) Le processus de Wiener $(W_t)_{t \geq 0}$ est le processus de diffusion pour lequel $\mu(x, t) = 0$ et $\sigma^2(x, s) = \sigma^2$.

(b) Le processus de diffusion d'Ornstein-Uhlenbeck, $(U_t)_{t \geq 0}$, est le processus de diffusion défini par $\mu(x, t) = -\beta x$ et $\sigma^2(x, s) = \sigma^2$. Ce processus a été conçu pour décrire la vitesse d'une particule dans un milieu visqueux, soumise comme dans le cas du mouvement brownien aux chocs incessants des molécules qui lui sont voisines.

10.8 ♠♠ Démontrer que si $(W_t)_t$ est un processus de Wiener, le processus défini par $U(t) = e^{-\beta t} W_{(e^{\beta t}-1)}$ est un processus d'Ornstein-Uhlenbeck.

Le bruit blanc

On sait que la densité spectrale de la lumière blanche est approximativement constante sur tout le spectre visible ; par analogie avec cette propriété, les spécialistes de la théorie du signal ont donné le nom de bruit blanc à tout bruit dont la densité spectrale est constante dans un domaine plus étendu que le domaine de fréquences du système bruité. On définit ci-dessous la notion mathématique de bruit blanc qui est le modèle idéalisé du bruit blanc physique.

Définition 10.15 *Le **bruit blanc** $(B_t)_t$ est un processus centré, stationnaire au sens large, de densité spectrale constante, de covariance $R(\tau) = \delta_0(\tau)$. Donc si $s \neq t$, B_s et B_t sont décorrélées ; d'un point de vue physique, cela signifie que les v.a. B_s et B_t ont un comportement indépendant même lorsque s et t sont très proches. Le bruit blanc gaussien est un bruit blanc tel que pour tout t , B_t est de loi normale $\mathcal{N}(0, \sigma^2)$.*

Le bruit blanc n'est pas un processus du second ordre puisque sa densité spectrale est égale à la transformée de Fourier $\mathcal{F}(\delta_0) = \mathbf{1}_{\mathbb{R}}$ (au sens des distributions tempérées) qui n'est pas un signal de puissance finie. De façon générale, les bruits blancs sont construits à partir de processus à accroissements indépendants et possèdent une densité spectrale constante ; on peut les générer par dérivation temporelle au sens des distributions d'un PAIS du second ordre.

Montrons que le bruit blanc peut être approché par le processus défini comme dérivée du processus de Wiener $(W_t)_{t \geq 0}$. Considérons le processus $(B_t^{(h)})_{t \geq 0}$, pour h fixé strictement positif, défini par :

$$\forall t \geq 0, \quad B_t^h = \frac{W_{t+h} - W_t}{h}.$$

(a) $(B_t^{(h)})_{t \geq 0}$ est centré : $E(B_t^{(h)}) = 0$.

(b) Son autocovariance $R(s, t)$ est égale à : $\frac{\sigma^2}{h^2} |h - |t - s||$ pour tout t tel que $s - h \leq t \leq s + h$.

Démontrons (b) :

$$R(s, t) = \frac{1}{h^2} E((W_{t+h} - W_t)(W_{s+h} - W_s)).$$

Considérons trois configurations possibles pour (s, t) :

• si $t < s - h$ ou $t > s + h$, les intervalles $[t, t + h]$ et $[s, s + h]$ sont disjoints ; la propriété des accroissements finis vérifiée par $(W_t)_t$ implique la nullité de $R(s, t)$;

• si $s - h < t < s$, il est facile d'établir que $R(s, t) = \frac{\sigma^2}{h^2}(h + t - s)$;

• si $s < t < s + h$, $R(s, t) = \frac{\sigma^2}{h^2}(h + s - t)$.

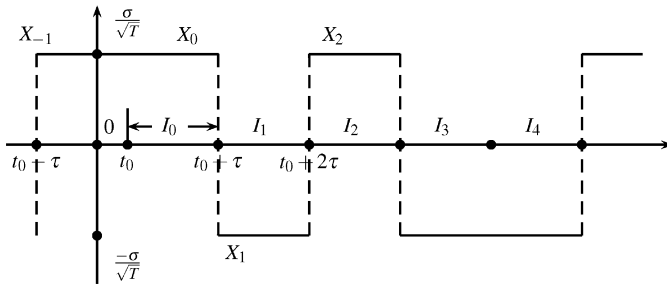
Conclusion : le processus $(B_t^{(h)})_{t \geq 0}$ est stationnaire puisque $R(s, t)$ ne dépend que de $(t - s)$ et, lorsque h devient très petit, $R(t - s)$ converge vers une *fonction* de Dirac. Un traitement rigoureux de cette convergence nécessite l'usage de la théorie des distributions.

On décrit ci-dessous deux bruits blancs qui, bien que de natures différentes, ont une densité spectrale identique.

(1) Soit $(X_t)_t$ un processus qui prend sur chaque intervalle $I_k = [T_0 + k\tau, T_0 + (k+1)\tau[$ où T_0 est une v.a. équirépartie sur $[0, \tau]$, une valeur aléatoire décrite par la v.a. de Bernoulli X_k de loi :

$$P(X_k = \frac{\sigma}{\sqrt{\tau}}) = P(X_k = -\frac{\sigma}{\sqrt{\tau}}) = \frac{1}{2}.$$

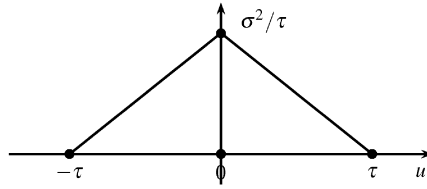
(Les v.a. X_k sont supposées indépendantes.)



Le processus est stationnaire : en effet, grâce au caractère aléatoire de T_0 , la fonction d'autocorrélation $E(X_t X_{t+u})$ ne dépend que de u , notée $R(u)$ de support $[-\tau, +\tau]$.

D'autre part $R(u) = E(X_t X_{t+u} | A) \cdot P(A) + E(X_t X_{t+u} | A^c) \cdot P(A^c)$, où A désigne l'événement « t et $t + u$ sont dans un même intervalle I_k » ; $R(u) = \frac{\sigma^2}{\tau} \times (1 - \frac{|u|}{\tau}) + 0 \times \frac{|u|}{\tau}$. Par parité, on obtient :

$$R(u) = \sigma^2 \left(\frac{\tau - |u|}{\tau^2} \right) \mathbf{1}_{[-\tau, +\tau]}(u)$$



Par application du théorème de la densité spectrale, on obtient :

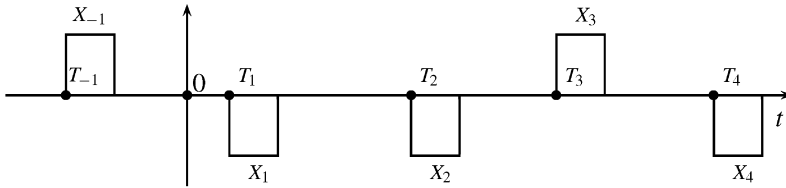
$$S_{\tau}(v) = \sigma^2 \frac{\sin^2(\pi v \tau)}{(\pi v \tau)^2}.$$

La suite des densités spectrales S_{τ} converge vers la fonction constante σ^2 quand τ tend vers 0.

(2) Considérons maintenant la suite des temps d'arrivée $(T_k)_{k \in \mathbb{Z}}$ d'un processus de Poisson défini sur \mathbb{R} , d'intensité λ et tel que $T_0 = 0$; les v.a. Y_k sur les intervalles $I_k = [T_k, T_k + \tau]$ étant des v.a. de Bernoulli indépendantes de loi commune

$$P(Y_k = \frac{\sigma}{\tau\sqrt{\lambda}}) = P(Y_k = \frac{-\sigma}{\tau\sqrt{\lambda}}) = \frac{1}{2}.$$

Le processus $(Y_t)_t$ est décrit par l'une de ses réalisations :



10.9 ♠♠ Étude du processus $(Y_t)_t$

Afin d'éviter le phénomène de recouvrement des intervalles I_k , on suppose que $\tau \ll 1$. Par construction, ce processus est stationnaire d'ordre deux.

Calculer son autocorrélation. Comparer les propriétés des deux processus $(X_t)_t$ et $(Y_t)_t$ quand τ tend vers 0, et conclure.

On remarquera que les réalisations des deux processus lorsque τ tend vers 0 ne sont pas des fonctions mais des distributions au sens de l'analyse fonctionnelle.

10.7 Histoire de la modélisation du mouvement brownien

Robert Brown (1773-1858) ne fut pas le premier à observer cette étrange agitation de particules en suspension dans un liquide; d'autres comme G. Buffon (1707-1788) ou L. Spallanzani (1729-1799) l'avaient fait avant lui. Son mérite est d'en avoir fait une étude systématique, dont il présenta les résultats

en 1831 et dans laquelle il écarta toute tentative d'explication fondée sur l'hypothèse vitaliste alors dominante en biologie. Rappelons que cette conception affirme que les phénomènes vitaux sont irréductibles aux phénomènes physico-chimiques sous-jacents et manifestent l'existence supposée d'une *force vitale*. C'est dans la seconde moitié du XIX^e siècle qu'ont eu lieu de nombreuses expérimentations dont la plupart conduisirent à la conception d'un mouvement dû aux forces attractives ou répulsives.

Les protagonistes de cette saga sont des savants connus comme P. Langevin, A. Einstein ou N. Wiener, et moins connus comme Louis Bachelier (1870-1946) qui soutint une thèse en 1900, sous la direction de Poincaré et dans laquelle il pose les bases de l'analyse financière moderne. En 1888, le physicien français L.G. Gouy, fin observateur du mouvement brownien, en proposa une caractérisation qui fut acceptée par la communauté des physiciens :

- très forte irrégularité et indépendance des trajectoires ;
- mouvement d'autant plus rapide des particules qu'elles sont petites et que le fluide est peu visqueux et de température élevée ;
- mouvement incessant et indépendant de la nature matérielle des particules.

À cette époque, l'hypothèse moléculaire due à Daniel Bernoulli, était opposée à la conception continuiste de la matière défendue par le grand mécanicien allemand E. Mach. En 1905, Einstein et Sutherland, physicien australien dont l'histoire a injustement oublié le nom, proposèrent, indépendamment l'un de l'autre un premier modèle du mouvement brownien fondé sur la théorie cinétique de la chaleur. Dans son article, Einstein voulait prouver l'existence des molécules, en calculer la taille et en déduire le nombre d'Avogadro ; il disposait de deux théories de nature apparemment contradictoire : la théorie de Van't Hoff concernant la pression osmotique, fondée sur l'hypothèse de l'agitation moléculaire et la théorie hydrodynamique de Stokes fondée sur l'hypothèse d'un fluide continu. Il établit l'équation qui exprime le coefficient de diffusion D en fonction de la constante R des gaz parfaits, du nombre d'Avogadro \mathcal{N} , de la température T du fluide, du rayon a de la particule et de sa viscosité η :

$$(1) \quad D = \frac{RT}{6\pi\eta a\mathcal{N}}.$$

Einstein démontra que la densité spatiale de la particule au temps t $f(x, y, z; t)$, définie par :

$P(\text{particule appartienne au borélien } B \text{ au temps } t) = \int \int \int_B f(x, y, z; t). dx. dy. dz,$
était solution de l'équation de diffusion

$$\frac{\partial f(x, y, z; t)}{\partial t} = D\Delta f(x, y, z; t).$$

$f(x, y, z, t)$ est la densité conditionnelle :

$$f_{W_t/(W_0=0)}(x, y, z, t) = \frac{1}{(4\pi Dt)^{\frac{3}{2}}} \exp\left(-\frac{x^2 + y^2 + z^2}{4Dt}\right),$$

où W_t est la position de la particule au temps t .

Einstein calcula ensuite la valeur moyenne quadratique du déplacement selon l'une des coordonnées (par exemple x) de la particule pendant le temps t . En notation moderne, ce résultat s'écrit :

$$(2) E(W_{t,x}^2 | W_0 = 0) = 2Dt,$$

qui met en relation les propriétés de dissipation et de viscosité d'un liquide avec une propriété de fluctuation moyenne, établissant ainsi la linéarité de l'élongation quadratique moyenne en fonction du temps.

Langevin produisit en 1908 une nouvelle et élégante démonstration de cette équation, fondée sur l'équation de l'énergie cinétique moyenne d'une particule dans un fluide :

$$(3) \frac{1}{2}mE(V^2) = \frac{RT}{2\mathcal{N}} \text{ où } V = \text{vitesse aléatoire de la particule.}$$

Démonstration de Langevin : la particule est soumise à une résistance due à la viscosité égale à $-6\pi\eta aV$, ainsi qu'à une force F aléatoire, d'espérance nulle résultant des chocs incessants des molécules sur la particule, supposée avoir une taille bien supérieure à celle des molécules. V est solution de l'équation :

$$(4) m \frac{dV}{dt} = -6\pi\eta aV + F.$$

Multipliée par l'élongation aléatoire X , l'équation (4) se transforme en :

$$\frac{m}{2}(X^2)'' - mV^2 = -\mu \frac{1}{2}(X^2)' + XF,$$

où : $\mu = 6\pi\eta a$.

Langevin calcula l'espérance des deux membres de cette équation et obtint après l'application du théorème de dérivation d'une intégrale :

$$(5) \frac{m}{2}(E(X^2))'' - mE(V^2) = -\mu \frac{1}{2}(E(X^2))' + E(XF).$$

Il postula ensuite que $E(XF) = 0$, ce qui n'est pas acquis, compte tenu de la corrélation possible entre la force F et l'élongation X qu'elle provoque !

Posant $U = \frac{1}{2}(E(X^2))'$, il écrivit (5) sous la forme :

$$(6) \quad \frac{m}{2} \frac{dU}{dt} - \frac{RT}{\mathcal{N}} + \mu U = 0,$$

dont la solution générale est :

$$U(t) = \frac{RT}{\mu \mathcal{N}} + c e^{(-\frac{\mu}{m}t)}.$$

En régime stationnaire, le terme exponentiel tend rapidement vers 0, car $\frac{\mu}{m}$ est très grand, d'où :

$$U(t) = \frac{RT}{\mu \mathcal{N}}.$$

Il en déduit la moyenne du carré de l'élongation sur un intervalle de temps de durée t et retrouva l'équation d'Einstein, en identifiant X^2 à $W_{t,x}^2$:

$$(7) \quad E(X^2) = \frac{2RT}{\mu \mathcal{N}} t = \frac{RT}{\mathcal{N}} \frac{1}{3\pi\eta a} t.$$

On sait actuellement résoudre dans le cadre de la théorie des équations stochastiques l'équation (4) dont la solution est le processus d'Ornstein-Uhlenbeck ; le problème 20 propose une étude de l'équation de Langevin discrète. Pour un approfondissement de la genèse de cette « merveille » qu'est le mouvement brownien, on lira avec profit les articles de J. P. Kahane et de B. Duplantier, dont je me suis inspiré pour ce bref historique. Le physicien polonais Smoluchowski (1872-1917) mit à profit sa conception claire de la cinétique du mouvement brownien pour en donner une conception originale grâce à la notion encore inédite de marche aléatoire isotrope dans le plan. Il démontra que la moyenne du carré de la distance séparant l'origine quittée au temps 0 et le point atteint au temps t par le processus de marche aléatoire est proportionnelle à t , résultat conforme à ceux obtenus par Einstein et Langevin. Puis, par l'étude dynamique du mouvement de la particule, il aboutit à l'équation dite d'Einstein-Sutherland, à un facteur constant multiplicatif près imputable aux approximations faites au cours des calculs. Ce n'est qu'une vingtaine d'années plus tard que Norbert Wiener construisit, dans le cadre de la théorie de la mesure et de l'analyse harmonique, l'objet mathématique (qui porte son nom) décrivant le processus brownien.

10.10 ♠♠ Étude d'un processus SSL à spectre borné

Soit $(X_t)_t$ un processus SSL à temps continu, à spectre borné (dont la densité spectrale est nulle en dehors d'un intervalle $[-v^*, v^*]$).

(1) Démontrer que pour tout τ :

$E((X_{t+\tau} - X_t)^2)$ est égal à $2(R(0) - R(\tau))$ et inférieur à $(2\pi v^* \tau)^2 R(0)$.

(2) Appliquer l'inégalité de Tchebycheff à la v.a. $(X_{t+\tau} - X_\tau)$, pour démontrer que : $\forall t, \tau \geq 0 \quad P(|X_{t+\tau} - X_\tau| > \varepsilon) \leq \frac{(2\pi v^* \tau)^2}{\varepsilon^2} R(0)$. En déduire que la variation d'un tel processus pendant une durée τ est d'autant plus faible que sa puissance est faible, résultat confirmé par l'interprétation physique du phénomène.

Les problèmes 21 et 22 se rapportent à ce chapitre.

10.8 Corrigés des exercices

Exercice 10.1

- (1)
$$\mathbb{E}(X_t) = \sum_k \mathbb{E}(A_k) e^{2i\pi v_k t} = 0.$$
- (2)
$$\begin{aligned} \mathbb{E}(X_{t_1} \overline{X_{t_2}}) &= \sum_{j,k} \mathbb{E}(A_j \overline{A_k}) e^{2i\pi v_j t_1} e^{-2i\pi v_k t_2} \\ &= \sum_{\ell} \mathbb{E}(A_{\ell} \overline{A_{\ell}}) e^{2i\pi v_{\ell}(t_1 - t_2)}, \end{aligned}$$
- car $\mathbb{E}(A_j \overline{A_k}) = 0$ si $j \neq k$, d'où : $R(t_1, t_2) = \sum_{\ell} \sigma_{\ell}^2 \cdot e^{2i\pi v_{\ell}(t_1 - t_2)}.$
- (3) Le spectre associé est défini par la distribution $\sum_k \delta_{v_k} \sigma_k^2.$

Exercice 10.3

(1) Il suffit de montrer que le vecteur $(f(t_1)X, \dots, f(t_n)X)$ est gaussien : ce qui est vrai puisque, pour tout (a_1, a_2, \dots, a_n) réels, $\sum_{i=1}^n a_i f(t_i)X$ est une v.a. gaussienne.

(2) Soit l'équipartition de $[a, b]$ en n intervalles de longueur égale, $(t_1 = a, t_2 \dots, t_n = b)$ vérifiant $t_{i+1} - t_i = \Delta.$

$Y_n = \Delta \sum_{i=1}^n \varphi(t_i) X_{t_i}$; Y_n est une combinaison linéaire finie de v.a. gaussiennes : c'est donc une v.a. gaussienne.

Démontrons que la suite $(Y_n)_n$ converge en moyenne quadratique vers Y :

$$\mathbb{E}((Y - Y_n)^2) = \mathbb{E}(Y^2) - 2\mathbb{E}(Y Y_n) + \mathbb{E}(Y_n^2);$$

On vérifie facilement $\mathbb{E}(Y_n^2)$ et $\mathbb{E}(Y Y_n)$ convergent vers $\mathbb{E}(Y^2)$; d'où la convergence en moyenne quadratique de $(Y_n)_n$ vers Y , qui entraîne la convergence en loi et donc de la normalité de Y .

Exercice 10.4

$$\forall t, \mathbb{E}(X_t) = \frac{1}{2\pi} \int_0^{2\pi} \cos(2\pi v t + \phi) d\phi = 0.$$

$$\mathbb{E}(X_t \cdot X_{t+s}) = \frac{1}{2\pi} \int_0^{2\pi} \cos(2\pi v(t+s) + \phi) \cos(2\pi v t + \phi) d\phi = \frac{1}{2} \cos(2\pi v s). (X_t)_t$$

est donc un SSL centré. Pour tout ϕ fixé, $m_X(\phi) = \frac{1}{2T} \int_{-T}^T \cos(2\pi v t + \phi) dt = 0 = \mathbb{E}(X_t)$: $(X_t)_t$ est donc ergodique par rapport à la moyenne.

Exercice 10.5

(1) Montrons que : $P(Z_s = y, Z_t - Z_s = x - y) = P(Z_s = y)P(Z_t - Z_s = x - y)$ pour x, y entiers, $x \geq y$ et les temps s, t vérifiant $s < t$.

$$\begin{aligned} P(Z_s = y, Z_t - Z_s = x - y) &= P\left(\Delta x \sum_{i=1}^{[s/(\Delta t)]} X_i = y, \Delta x \left(\sum_{j=1}^{[t/(\Delta t)]} X_j - \sum_{i=1}^{[s/(\Delta t)]} X_i\right) = x - y\right) \\ &= P\left(\Delta x \sum_{i=1}^{[s/(\Delta t)]} X_i = y, \Delta x \sum_{j=[s/(\Delta t)]+1}^{[t/(\Delta t)]} X_j = x - y\right) \\ &= P\left(\Delta x \sum_{i=1}^{[s/(\Delta t)]} X_i = y\right) P\left(\Delta x \sum_{j=[s/(\Delta t)]+1}^{[t/(\Delta t)]} X_j = x - y\right). \end{aligned}$$

car les v.a. $\Delta x \sum_{i=1}^{[s/(\Delta t)]} X_i$ et $\Delta x \sum_{j=[s/(\Delta t)]+1}^{[t/(\Delta t)]} X_j$ sont indépendantes.

La stationnarité est triviale à démontrer.

(2) Z_t est une somme finie de v.a. i.i.d. d'espérance nulle et de variance $\frac{(\Delta x)^2}{4}$, donc, par le théorème central limite, Z_t converge en loi vers $\mathcal{N}\left(0; [t/\Delta t] \cdot \frac{(\Delta x)^2}{4}\right)$.

Si $\Delta x \rightarrow 0$ et $\Delta t \rightarrow 0$ de manière que $\frac{\Delta x}{\sqrt{\Delta t}} \rightarrow \sigma$, alors

$$[t/\Delta t] \frac{(\Delta x)^2}{4} = \frac{t}{4} \left(\frac{\Delta x}{\sqrt{\Delta t}} \right)^2 \rightarrow \frac{t\sigma^2}{4}.$$

Donc, le processus $(Z_t)_t$ converge vers le processus de Wiener $(W_t)_t$ où pour tout t , $W_t \sim \mathcal{N}\left(0, \frac{\sigma^2 t}{4}\right) = \mathcal{N}\left(0, \frac{D}{2}t\right)$.

Exercice 10.7

(1) Par approximation : $Z(t) = \lim_{n \rightarrow +\infty} Z_n(t)$ où

$$Z_n(t) = \frac{1}{n} \sum_{k=1}^n W\left(\frac{tk}{n}\right).$$

$Z_n(t)$ est une combinaison linéaire des composantes du vecteur gaussien

$$\left(W\left(\frac{tk}{n}\right)\right)_{k=1, \dots, n};$$

c'est donc une v.a. gaussienne et, par passage à la limite, $Z(t)$ est une v.a. gaussienne d'espérance :

$$\mathbb{E}(Z(t)) = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E}\left(W\left(\frac{tk}{n}\right)\right) = 0.$$

On en déduit grâce à la définition (3.14) que, pour tout r -uplet, $(t_i)_{i=1, \dots, r}$, $(Z_{t_1}, \dots, Z_{t_r})$ est un vecteur gaussien.

Pour $s < t$,

$$\begin{aligned}
 R(s, t) &= \mathbb{E}(Z(s).Z(t)) = \mathbb{E}\left(\int_0^t W(x) dx \int_0^s W(y) dy\right) \\
 &= \int_0^t \int_0^s \mathbb{E}(W(x)W(y)) dx dy = \int_0^s \left(\int_0^y \min(x, y) dx + \int_y^t \min(x, y) dx\right) dy \\
 &= \int_0^s \left(\int_0^y x dx + \int_y^t y dx\right) dy = \int_0^s \left(\frac{y^2}{2} + y(t-y)\right) dy \\
 &= s^2 \left(\frac{t}{2} - \frac{s}{6}\right).
 \end{aligned}$$

Conclusion : $R(s, t)$ n'est pas fonction de $(t-s)$, d'où la non-stationnarité du processus brownien intégré.

$$(2) \operatorname{cov}(Z(t) - Z(s), Z(s)) = \mathbb{E}((Z(t) - Z(s))Z(s)) = s^2 \left(\frac{s}{3} + \frac{t-s}{2}\right) - \frac{s^3}{3} = -\frac{4s^3}{6} + \frac{3s^2 t}{6} \neq 0, \text{ d'où la dépendance des v.a. } Z(t) - Z(s) \text{ et } Z(s).$$

Le processus brownien intégré n'est donc pas un processus de Wiener.

Exercice 10.9

Il existe trois types d'événements :

- A : t et $t + \tau$ appartiennent à un seul I_k .
- B : t et $t + \tau$ sont dans deux I_k distincts.
- C : t ou $t + \tau$ n'est dans aucun I_k .

Les (Y_k) sont indépendantes, d'espérance nulle et de support $[-T, T]$.

Pour tout $\tau \in [0, T]$:

$$R(\tau) = \mathbb{E}(Y(t)Y(t+\tau)/A)P(A) + \mathbb{E}(Y(t)Y(t+\tau)/B)P(B) + \mathbb{E}(Y(t)Y(t+\tau)/C)P(C).$$

Les deux derniers termes sont nuls et,

$$P(A) = P(\ll \text{au moins un } T_k \in [t + \tau - T, t] \gg) = \lambda(T - \tau)(1 + \varepsilon(T)),$$

où $\lim_{T \rightarrow 0} \varepsilon(T) = 0$. Donc, pour T très petit :

$$R(\tau) = \sigma^2 \frac{T - |\tau|}{T^2} \mathbf{1}_{[-T, T]}(\tau)$$

identique à la densité spectrale du processus étudié précédemment.

Les deux processus $(X_t)_t$ et $(Y_t)_t$ ont même autocorrélation. La densité spectrale tend vers σ^2 quand $\tau \rightarrow 0$ et (Y_t) est constitué d'un train d'impulsion de Dirac, aux points $(T_k)_k$ d'amplitude $\frac{\pm \sigma}{\sqrt{\lambda}}$.

Exercice 10.10(1) $(X_t)_t$ étant stationnaire :

$$E((X_{t+\tau} - X_t)^2) = E(((X_{t+\tau} - E(X_{t+\tau})) - (X_t - E(X_t)))^2) = 2(R(0) - R(\tau))$$

$$R(0) - R(\tau) = \int_{-\nu^*}^{\nu^*} S(\nu) d\nu - \int_{-\nu^*}^{\nu^*} S(\nu) e^{2i\pi\nu t} d\nu.$$

Or $R(0) - R(\tau)$ est réel, donc $S(\nu)$ est paire.

$$\begin{aligned} R(0) - R(\tau) &= 2 \int_0^{\nu^*} S(\nu)(1 - \cos(2\pi\nu\tau)) d\nu = 4 \int_0^{\nu^*} S(\nu)(\sin(\pi\nu\tau))^2 d\nu \\ &\leq 4\pi^2 \nu^{*2} \tau^2 \int_0^{\nu^*} S(\nu) d\nu = R(0). \end{aligned}$$

Donc : $2.(R(0) - R(\tau)) \leq 4\pi^2 \nu^{*2} \tau^2 . R(0)$.

$$(2) P(|X_{t+\tau} - X_t| > \varepsilon) \leq \frac{(2\pi\nu^*\tau)^2 R(0)}{\varepsilon^2}.$$

Conclusion : les variations de $(X_t)_t$ sont d'autant plus faibles ($P(|X_{t+\tau} - X_t| > \varepsilon$ est petit) que ν^* , τ et $R(0)$ sont petits.

Chapitre 11

Problèmes

11.1 Énoncés des problèmes

Problème 1 ♠ *Dénombrement de chemins dans un réseau plan régulier carré*

On considère le réseau carré régulier formé par les points de \mathbb{Z}^2 . On appelle chemin une succession de segments de longueur 1, dirigés soit vers le haut, soit vers la droite.

(1) Déterminer le nombre de chemins reliant l'origine au point (m, n) . [Indication : on pourra considérer la bijection entre l'ensemble des chemins et l'ensemble des $(m+n)$ -uplets contenant m lettres « d » et n lettres « h » ; « d » et « h » désignant respectivement les segments unitaires orientés vers la droite et vers le haut.]

(2) Si $m > n$, en déduire que : $C_{m+n}^n = \sum_{k=0}^n C_{n+k}^n \cdot C_{m-k}^{m-n}$.

Problème 2 ♠♠ *Quelques problèmes pour les amateurs de combinatoire*

(1) En combien de parties, un ensemble de n droites dsachant que trois quelconques d'entre elles ne se coupent pas en un même point et que deux quelconques d'entre elles ne sont pas parallèles, partitionnent le plan ? Même question dans le cas de n plans dans l'espace dont trois quelconques d'entre eux ne se coupent jamais selon une même droite ?

(2) Si σ_n^k désigne le nombre de partitions d'un ensemble de n éléments en k classes, démontrer que :

$$\sigma_{n+1}^k = \sigma_n^{k-1} + k \cdot \sigma_n^k.$$

(3) Question ouverte : combien y a-t-il de façons de plier une bande de n timbres ? (Il n'existe pas, semble-t-il, de solution de forme récurrente, à ce jour.)

Problème 3 ♠♠

On effectue une suite de tirages successifs avec remise dans une urne contenant des jetons gagnants (portant le numéro 1), en proportion p ($p \in]0, 1[$), et des jetons perdants (portant le numéro 0), en proportion $q = 1 - p$. On désigne par « série », une suite de numéros identiques.

Soit X la longueur de la première « série » et Y la longueur de la deuxième « série ». Ainsi, par exemple, pour la suite de tirages 111001011..., on a $X = 3$ et $Y = 2$ et pour 01101..., on a $X = 1$ et $Y = 2$. [On rappelle que $\sum_{k=1}^{+\infty} kx^k = \frac{x}{(1-x)^2}$ pour $|x| < 1$.]

- (1) a) Déterminer la loi de la v.a. X ;
 b) Déterminer $E(X)$. Montrer que $E(X)$ est minimale lorsque $p = \frac{1}{2}$.
- (2) a) Montrer que pour tout $(i, j) \in (\mathbb{N}^*)^2$, $P([X = i] \cap [Y = j]) = p^{i+1}q^j + q^{i+1}p^j$;
 b) En déduire la loi de la v.a. Y et montrer que $E(Y) = 2$.
- (3) a) Si $p \neq \frac{1}{2}$, établir que les v.a. X et Y ne sont pas indépendantes [considérer, par exemple, $P([X = 1] \cap [Y = 1])$]. Démontrer que, si $p = \frac{1}{2}$, les v.a. X et Y sont indépendantes ;
 b) Montrer que $\text{cov}(X, Y) = -\frac{(1-2p)^2}{p(1-p)}$.

Problème 4 ♠♠ L'étrange loi de Benford

L'astronome S. Newcomb remarqua en 1881 que l'ensemble des pages des tables de logarithme, qu'on utilisait alors fréquemment, étaient inégalement usées : l'usure allait en décroissant des premières pages comportant les nombres débutant par le chiffre 1, jusqu'aux dernières où les nombres débutent par 9. Il étendit son enquête à d'autres domaines de données scientifiques et socio-économiques et obtint des observations qui confirmèrent ses premiers résultats. Cela signifiait-il que, dans la jungle des nombres brassés par une société, il y aurait plus de nombres commençant par 1, qu'il y en aurait commençant par 2, lesquels seraient plus nombreux que ceux commençant par 3 et ainsi de suite jusqu'à 9 ?

Il proposa la loi de probabilité suivante pour la répartition de la variable X associée au premier chiffre d'un nombre quelconque :

$$\text{Pour tout } i = 1, 2, \dots, 9 \quad P(X = i) = \log_{10}\left(1 + \frac{1}{i}\right).$$

Méconnaissant les travaux de Newcomb, le physicien F. Benford proposa la même loi en 1957, à laquelle la postérité donna (injustement) son nom seul. Il remarqua que cette loi ne convenait pas pour tous les ensembles de nombres tels que les numéros de téléphone d'un pays distribués selon des critères spécifiques, les tailles des êtres humains, les prix d'un type de voitures et généralement les

ensembles de données variant à l'intérieur d'un intervalle de faible étendue. En revanche, des séries numériques telles que $(r^n)_n$, où r est un réel positif tel que $\log_{10}(r)$ ne soit pas un rationnel, ou encore $(n!)_n$ et $(n^n)_n$, ainsi que les échantillons issus de mélanges de distributions numériques aléatoires quelconques suivent la loi de Benford. La loi doit être invariante par rapport aux changements de mesure ou d'unités mis en œuvre dans les diverses régions du monde. Cette idée est à la base de la construction de la loi par R. Pickam en 1961 : la seule probabilité sur les mantisses d'un ensemble de nombres invariante par changement d'échelles est la loi de Benford.

Appelons mantisse d'un réel positif x , le nombre $M(x) \in [1, 9]$, défini par : $x = M(x) \cdot 10^k$ pour un certain entier k . Benford conjectura que la proportion de nombres x tels que $M(x) \in [a, b]$, où $1 \leq a < b \leq 10$, est égale à $(\log(b) - \log(a))$:

$$P(M(x) \in [a, b]) = \log(b) - \log(a).$$

(1) Dédire la loi de Benford de son postulat.

(2) Démonstration de la conjecture, due à Pinkham.

L'invariance par changement d'échelle s'exprime par :

$$(*) \quad P(M(x) \in [a, b]) = P(M(\alpha \cdot x) \in [a, b]).$$

$$\forall x > 0, \alpha > 0 \text{ et les entiers } a \text{ et } b \text{ tel que : } 1 \leq a < b \leq 10.$$

Par définition : $\log_{10} M(x) = \text{Partie fractionnaire } \log_{10}(x)$.

Par exemple : $x = 847,63$, $M(x) = 8,476$, $\log_{10}(x) = 2,928$

$\log_{10}(M(x)) = 0,928 = \text{Partie fractionnaire } (\log_{10}(x)) =_{(notée)} \{\log_{10}(x)\}$.

(2.1) Exprimer les membres de l'équation (*) à l'aide de $\{\log_{10}(x)\}$. Sachant que la partie fractionnaire d'une somme est égale (modulo 1) à la somme des parties fractionnaires, en déduire :

$$P(M(\alpha \cdot x) \in [a, b]) = P(\{\log_{10}(x)\} \in [\log_{10}(a) - \log_{10}(\alpha), \log_{10}(b) - \log_{10}(\alpha)]).$$

(2.2) Un changement de variables transforme l'équation précédente en :

$P(\log_{10}(x) \in [s, t]) = P(\{\log_{10}(x)\} \in [s + v, t + v])$ pour v réel et s, t tels que $0 \leq s < t \leq 1$ et où, $s + v$ et $t + v$ sont définis modulo 1.

En déduire $P(\log_{10}(x) \in [s, t]) = t - s$, puis $P(M(x) \in [a, b]) = \log_{10}(b) - \log_{10}(a)$ et enfin la loi de Benford.

Des tests statistiques fondés sur la loi de Benford ont permis aux experts du fisc américains, depuis les années 1980, de débusquer des fraudes au sein de comptabilités d'entreprises américaines, qui avaient maquillé leurs comptes. Gageons, sans excès de pessimisme de notre part, que depuis ce temps lesdites entreprises et bien d'autres sans doute falsifient leur bilans en tenant compte cette fois de la loi de Benford. Certaines compagnies informatiques ont même

envisagé, dans un souci d'optimisation, de conformer la structure de leurs futurs ordinateurs en fonction des fréquences des entiers 1,2,3,...,9 répartis selon la loi de Benford (voir Janvresse. E., La loi de Benford).

Problème 5 ♠♠ *Étude d'un modèle classique d'évolution : le modèle autorégressif d'ordre un*

Soit une suite finie de v.a. $(X_k)_{k=1,\dots,n}$ vérifiant les propriétés :

- (a) $X_{k+1} = a.X_k + \varepsilon_k$ et $X_0 = 0$;
- (b) pour tout k , ε_k est de loi $\mathcal{N}(0, \sigma^2)$. $(\varepsilon_k)_k$ est le modèle classique d'un bruit discret si les ε_k sont décorréliées : $E(\varepsilon_k.\varepsilon_l) = 0$ si $k \neq l$.

(1) Exprimer X_k en fonction de $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k$ et des puissances de a ; en déduire sa loi, son espérance et sa variance.

(2) Que vaut la corrélation $\rho(X_{k+1}, X_k)$ et sa valeur limite quand n devient très grand, selon les valeurs de a ?

(3) Supposons que a soit inconnu, et que les v.a. X_k soient observées et prennent les valeurs x_k ; comment envisageriez-vous d'estimer a ?

Problème 6 ♠♠♠ *Étude du vecteur gaussien en coordonnées sphériques*

Soit le vecteur normal (X_1, X_2, X_3) de composantes indépendantes, chacune de loi $\mathcal{N}(0, 1)$. Soit T la bijection :

$$(x_1, x_2, x_3) \in \mathbb{R}^3 \longrightarrow T(x_1, x_2, x_3) = (r, \phi, \theta) \in [0, +\infty[\times [0, \pi[\times [-\pi, \pi[,$$

$$\text{avec} \quad \begin{cases} x_1 = r \sin \theta \cos \phi \\ x_2 = r \sin \theta \sin \phi \\ x_3 = r \cos \theta. \end{cases}$$

Soit (R, Φ, Θ) le vecteur aléatoire image de (X_1, X_2, X_3) par T ; on rappelle que $J_{T^{-1}}(r, \phi, \theta) = r^2 \sin \phi$.

(1) Calculer la densité conjointe de (R, Φ, Θ) , ainsi que les densités marginales; en déduire le caractère isotrope du vecteur (X_1, X_2, X_3) . Caractériser la loi de probabilité portée par toute sphère centrée en 0.

(2) Calculer la loi de R^2 .

Problème 7 ♠♠♠ *Étude des vecteurs aléatoires ordonnés dits aussi statistiques d'ordre*

Considérons un vecteur aléatoire de n variables indépendantes X_1, X_2, \dots, X_n de même densité f , auquel on associe le vecteur réordonné $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ défini par :

$$\begin{aligned} X_{(1)} &= \min(X_1, X_2, \dots, X_n); \\ X_{(2)} &= \min((X_1, X_2, \dots, X_n) \setminus X_{(1)}); \\ &\vdots \end{aligned}$$

$$X_{(n)} = \max(X_1, X_2, \dots, X_n).$$

(1) Démontrer que $P(X_{(1)} \leq x_1, X_{(2)} \leq x_2, \dots, X_{(n)} \leq x_n)$ est égal à :

$$\sum_{\sigma \in \text{Perm}\{1, 2, \dots, n\}} P(X_{\sigma(1)} \leq x_1, X_{\sigma(2)} \leq x_2, \dots, X_{\sigma(n)} \leq x_n; X_{\sigma(1)} < X_{\sigma(2)} < \dots < X_{\sigma(n)})$$

où $\text{Perm}(n)$ est l'ensemble des permutations de $\{1, 2, \dots, n\}$. [Envisager d'abord le cas où $n = 2$.]

(2) En déduire :

$$f_{X_{(1)}, X_{(2)}, \dots, X_{(n)}}(x_1, \dots, x_n) = n! \prod_{i=1}^n f(x_i) \cdot \mathbf{1}_{(x_1 < x_2 < \dots < x_n)}(x_1, \dots, x_n).$$

(3) Soit N_x la v.a. égale au nombre des X_i inférieures au réel x . Démontrer que N_x est de loi binomiale. Déterminer la fonction de répartition et la densité de $X_{(k)}$.

Application : les X_i sont de densité uniforme sur $[0, 1]$.

Déterminer la densité de $X_{(k)}$ et son espérance, que l'on calculera grâce à l'égalité : $\int_0^1 x^{p-1} \cdot (1-x)^{q-1} dx = \frac{\Gamma(p) \cdot \Gamma(q)}{\Gamma(p+q)}$, où $p, q > 1$.

(4) Quelle est la limite de la suite $(n \cdot X_{(k)})_n$ pour la convergence en loi, pour k fixé. Trouver intuitivement la loi limite de $X_{(k)}$ lorsque k et n tendent vers l'infini selon le rapport constant $\frac{n}{k} = \alpha$.

Problème 8 ♠♠

On considère trois composants électroniques dont les durées de fonctionnement X_1 , X_2 et X_3 sont supposées indépendantes, de loi exponentielle $\mathcal{E}(\lambda)$. Un appareil utilise les composants 1 et 2 simultanément et lorsque l'un ne fonctionne plus, on le remplace par le troisième.

(1) Déterminer la densité de $(X_1, X_2 - X_1)$. En déduire que :

$$f_{X_2 - X_1}(x) = \frac{\lambda}{2} \left[e^{\lambda x} \mathbf{1}_{]-\infty, 0[}(x) + e^{-\lambda x} \mathbf{1}_{[0, +\infty[}(x) \right].$$

Montrer alors que $Y = |X_2 - X_1|$ a même loi que les X_i .

(2) En utilisant l'indépendance de Y et X_3 , exprimer la densité de $Y - X_3$, puis $F_{Y - X_3}(0)$. [On vérifiera ainsi que le composant 3 a la probabilité $P([X_3 \geq Y]) = \frac{1}{2}$ d'être le dernier à tomber en panne.]

(3) Déterminer la densité de $Z = \min(X_1, X_2)$ puis la densité, l'espérance et la variance de $T = Z + X_3$ [T représente l'instant de panne du composant 3].

Problème 9 ♠♠♠

On considère deux composants électroniques dont les durées de vie X et Y sont supposées indépendantes, de loi exponentielle $\mathcal{E}(\lambda)$.

(1) Un appareil utilise ces composants l'un après l'autre (lorsque l'un ne fonctionne plus, on fait appel à l'autre). Montrer que le couple $(U, V) = (X, \frac{X+Y}{X})$ a pour densité : $f_{U,V}(u, v) = \lambda^2 u e^{-\lambda uv} \mathbf{1}_{]0, +\infty[}(u) \mathbf{1}_{]1, +\infty[}(v)$. En déduire la loi de V , quotient de la durée de vie totale de l'appareil par la durée de vie du premier composant. La variable V admet-elle une espérance ?

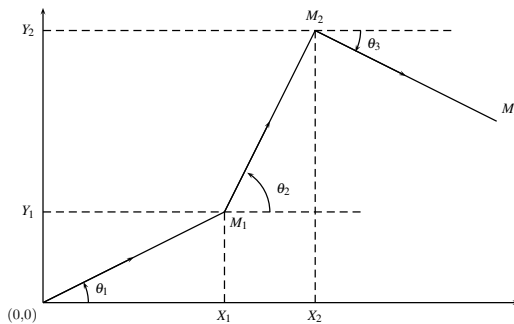
(2) Des composants démarqués présentent un défaut, qui réduit leur durée de vie de moitié, avec une probabilité de 10 %.

On note $Z = TX$, où T est une v.a. indépendante de X , telle que $P([T = 1]) = 0,9$ et $P([T = \frac{1}{2}]) = 0,1$. Que représente Z ? Déterminer $P([Z \leq z])$ et en déduire une densité de la loi de Z . Déterminer la durée de vie moyenne d'un composant démarqué.

Problème 10 ♠♠ *Étude de la marche aléatoire isotrope à pas constants dans le plan*

Le statisticien anglais Karl Pearson introduisit en 1905 la notion de marche aléatoire, afin de rendre compte des migrations de populations d'insectes dans une région donnée. On se propose plus modestement de modéliser le déplacement erratique d'une particule dans le plan, muni d'un repère ortho-normé, effectuant à chaque instant n un pas de longueur constante a dans une direction aléatoire isotropique mesurée par l'angle Θ_n , par rapport à l'axe $x'x$. Les Θ_n sont supposés indépendants entre eux et $(X_0, Y_0) = (0, 0)$.

- (1) Exprimer les coordonnées (X_n, Y_n) au temps n , en fonction des angles Θ_n .
Calculer l'espérance et la variance de X_n et Y_n .
- (2) Sont-elles corrélées ?
- (3) Loi de (X_n, Y_n) quand n est grand ? En déduire la loi de (R_n, T_n) , de T_n et de R_n si $X_n = R_n \cos T_n$ et $Y_n = R_n \sin T_n$. Calculer $E(R_n)$, $E(R_n^2)$ et $V(R_n)$.
Interpréter ces derniers résultats. T_n est l'angle que fait OM_n avec l'axe des abscisses.



Problème 11 ♠♠ *Calcul d'une intégrale simple par la méthode de Monte-Carlo*

Cette méthode d'approximation d'intégrales est une application directe de la loi des grands nombres; elle est très utilisée et d'autant plus efficace par rapport aux méthodes numériques que le nombre k de variables dont dépend la fonction à intégrer est grand et que la fonction est irrégulière. En effet, les méthodes numériques déterministes nécessitent que les fonctions n'aient pas un trop grand nombre de points de discontinuité; d'autre part, le temps de calcul qui leur est nécessaire est égal à une constante élevée à la puissance k , alors qu'il est proportionnel à k pour la méthode de Monte-Carlo. Des méthodes de ce type ont été développées pour résoudre des équations aux dérivées partielles de type transport ou de diffusion comme les équations de Fokker-Planck ou de Kolmogorov. Pour mémoire, on rappelle que c'est Buffon qui en 1777, utilisa sans avoir une idée claire de son assise théorique, la première méthode de Monte-Carlo (voir problème de l'aiguille de Buffon, chapitre 2), afin de donner une estimation convergente de la valeur de π .

Soit une fonction f continue définie sur $[a, b]$, à valeurs dans $[m, M] \subset \mathbb{R}_+$; ($m = \min(f)$, $M = \max(f)$). On note $D = [a, b] \times [m, M]$ et $A = \{(x, y) \in D; f(x) \geq y \geq 0\}$.

L'aire de $A = \int_a^b f(x).dx$ et $p = \frac{\text{aire de } A}{\text{aire de } D}$.

Pour tout $k = 1, \dots, n$, on considère le couple (X_k, Y_k) de v.a. uniformes sur D , et on définit la v.a. Z_k par :

$$Z_k = \begin{cases} 1 & \text{si } Y_k \leq f(X_k) \\ 0 & \text{sinon.} \end{cases}$$

On pose $\overline{Z}_n = \frac{Z_1 + \dots + Z_n}{n}$. (Les couples de v.a. (X_n, Y_n) sont supposés indépendants.)

- (1) Calculer $E(Z_k)$, $\text{Var}(Z_k)$, $E(\overline{Z}_n)$, $\text{Var}(\overline{Z}_n)$.
- (2) En utilisant la loi des grands nombres, démontrer que \overline{Z}_n converge en probabilité vers p .
- (3) En utilisant l'inégalité de Bienaymé-Tchebychev, déterminer le plus petit entier n tel que :

$$P[|\overline{Z}_n - p| > \frac{1}{100}] < \frac{5}{100}.$$

Problème 12 ♠♠♠ *Évolution des suites de records*

Le but de ce problème est de démontrer que les records d'une suite de variables i.i.d. $(X_i)_{i=1,2,\dots,n}$, définis comme étant la suite des maxima relatifs consécutifs de la suite, décrivant par exemple l'évolution d'un paramètre climatique ou d'une performance sportive, devraient évoluer très lentement. On remarquera que compte tenu des modifications profondes subies par le climat

depuis quelques décennies, l'hypothèse d'équidistribution des v.a. $(X_i)_{i=1,2,\dots,n}$ n'est plus recevable de nos jours.

On supposera que les v.a. sont positives de densité f_X et on pose :

$$Y_n = \sup(X_1, X_2, \dots, X_n).$$

(1) Démontrer que pour toute permutation σ des n premiers entiers :

$$P(X_{\sigma(1)} < X_{\sigma(2)} < \dots < X_{\sigma(n)}) = \frac{1}{n!}.$$

(2) Pour tout $n \geq 1$, on définit la v.a. R_n égale à 1 si $Y_n = X_n$ et égale à 0, sinon. Démontrer que $P(R_n = 1) = \frac{1}{n}$.

(3) Soient $S_n = \sum_{i=1}^n R_i$ et $T_n^p = S_{n+p} - S_n$ le nombre de records entre les temps n et $n+p$. Démontrer que $\lim_{n \rightarrow +\infty} E(T_n^p) = 0$. [Indication : on utilisera l'approximation $\sum_{i=1}^n \frac{1}{i} = \ln n + \gamma + o(1)$ où γ est la constante d'Euler.

(4) On admettra le résultat : si $m < n$, $P(R_m = 1, R_n = 1) = \frac{1}{m.n}$. En déduire que $\text{cov}(R_m, R_n) = 0$.

(5) Montrer que $\lim_{n \rightarrow +\infty} \text{Var}(T_n^p) = 0$.

(6) Supposons que $f_X(x) = e^{-x}$. Calculer $P(Y_n < y)$ et en déduire la densité de Y_n ; calculer $E(Y_n)$ lorsque n est grand, en utilisant l'approximation de la question (3). Interpréter ce résultat.

Problème 13 ♠♠

À chaque instant n , il entre Y_n personnes dans une ville donnée et il en part Z_n . Soit X_n le nombre de personnes dans la ville à l'instant n . Chaque personne peut en partir avec la probabilité $p \in]0, 1[$. On suppose que les v.a. Y_n suivent une loi de Poisson $\mathcal{P}(\lambda)$ et sont indépendantes des v.a. X_i et Z_j pour $i \leq n-1$, $j \leq n$. Notons ϕ_n la fonction génératrice de X_n .

(1) Déterminer $\mathbb{E}^{[X_n=i]}(s^{X_n})$; en déduire que $E(s^{X_{n-1}-Z_n} | X_{n-1} = k) = (p + (1-p)s)^k$, puis exprimer ϕ_n en fonction de ϕ_{n-1} .

(2) On pose $K_n(u) = \phi_n(u+1)$. Montrer par récurrence que :

$$K_n(u) = \exp\left(\frac{\lambda u}{p}(1 - (1-p)^n)\right) K_0(u(1-p)^n).$$

En déduire $\lim_{n \rightarrow +\infty} K_n(u)$ puis $\lim_{n \rightarrow +\infty} \phi_n(s)$ et en déduire que X_n converge en loi vers la v.a. de Poisson $\mathcal{P}(\frac{\lambda}{p})$.

Problème 14 ♠♠♠ *Le processus de branchement ou de ramification*

Il s'agit d'un modèle adapté à l'évolution d'une population d'êtres vivants ou de particules (par exemple des neutrons), engendrant à chaque instant n une descendance de taille aléatoire X_n . Historiquement, ce modèle fut proposé par les

mathématiciens anglais Galton et Watson à la fin du XIX^e siècle pour expliquer la disparition avec le temps de certains noms de famille.

On définit : X_n = nombre d'individus à la n -ième génération.

ξ_n^i = nombre de descendants appartenant à la $(n+1)$ -ième génération, de l'individu i appartenant à la n -ième génération.

Les ξ_n^i sont des v.a. entières positives ou nulles, indépendantes et de même loi. On étudiera le processus de ramification, pour lequel la population initiale se réduit à un seul individu ($X_0 = 1$).

(1) Démontrer que : $\forall j, k \in \mathbb{N}, P(X_{n+1} = j | X_n = k) = P(\sum_{l=1}^k \xi_n^l = j)$.

En déduire que $(X_n)_n$ est une chaîne.

(2) Montrer que : $\forall n, g_{n+1}(s) = g_n(g(s))$ où $g(s)$ est la fonction génératrice de ξ et $g_n(s)$ la fonction génératrice de la v.a. X_n .

(3) En déduire les équations de récurrence vérifiées par $E[X_n]$ et $V[X_n]$. On suppose : $X_0 = 1, E[\xi] = m, V[\xi] = \sigma^2$.

$$\text{Montrer que } E[X_n] = m^n \text{ et } V[X_n] = \begin{cases} n \cdot \sigma^2 & \text{si } m = 1 \\ \sigma^2 m^{n-1} \frac{m^n - 1}{m - 1} & \text{si } m \neq 1 \end{cases}$$

Comment évolue la population selon les valeurs de m ?

Problème 15 ♠♠

Un opérateur téléphonique propose à un client deux types d'abonnements mensuels. Soit X_n le type utilisé par le client pour le mois n . On observe que l'abonnement du type 1 est reconduit le mois suivant avec la probabilité $\frac{1}{2}$ et que l'abonnement du type 2 est reconduit avec la probabilité $\frac{1}{4}$.

(1) Déterminer le graphe, la matrice, et, si elle existe, la distribution stationnaire de la chaîne de Markov ainsi obtenue. Quelle est la probabilité d'utiliser l'abonnement 1 au mois de juillet si celui-ci est utilisé au mois d'avril ?

(2) Quel est, en mois, le temps moyen de retour à chaque type d'abonnement ? Retrouver ceux-ci après avoir calculé $P^{(X_0=i)}(T_i = n)$ pour tout $n \in \mathbb{N}$ et pour tout $i \in \{1, 2\}$.

(3) Quel est le coût moyen mensuel de l'abonnement si les tarifs sont de 100 euros pour le type 1 et de 200 euros pour le type 2 ?

(4) Même question, en considérant qu'un changement d'abonnement coûte 50 euros. [On fera intervenir X_n et X_{n-1} avant de passer à la limite.]

Problème 16 ♠♠♠

Lorsqu'une équipe de foot a joué n matches, on note X_n la série de matches en cours sans défaites. Si $X_n = k$, on note p_k la probabilité de ne pas perdre le match suivant. Ainsi, $P([X_{n+1} = k+1] | [X_n = k]) = p_k$ et $P([X_{n+1} = 0] | [X_n = k]) = 1 - p_k$.

(1) Montrer rapidement que (X_n) définit une chaîne de Markov irréductible et faire son graphe. Si T_0 est l'instant de premier retour en 0, exprimer $P([T_0 = n] | [X_0 = 0])$ et donner des conditions pour que la chaîne soit transitoire, récurrente nulle ou récurrente positive (à l'aide des $r_n = p_0 p_1 \cdots p_{n-1}$).

(2) On suppose que $p_k = p$ pour tout k . Déterminer la distribution stationnaire de la chaîne et la longueur moyenne d'une série sans défaite. [*Application numérique* : $p = \frac{1}{2}$.]

Problème 17 ♠

Un radar est placé sur une route où il passe en moyenne cinq véhicules en excès de vitesse par heure. On admet que ces véhicules forment un processus de Poisson (N_t) .

(1) Déterminer la probabilité qu'une voiture ait été prise dans le premier quart d'heure, sachant que deux voitures ont été prises en une heure.

(2) On suppose ici que la durée de fonctionnement du radar T suit une loi exponentielle de moyenne 100 heures. Déterminer la loi du nombre N de véhicules détectés par le radar et le nombre moyen de ces véhicules. Pour cela :

a) expliquer la relation $P^{(T=t)}([N = k]) = P([N_t = k])$;

b) en utilisant $P([N = k]) = \int P^{(T=t)}([N = k]) f_T(t) dt$, établir que N suit une loi géométrique sur \mathbb{N} .

Problème 18 ♠♠♠

Une station assure deux types de service, prioritaire ou non prioritaire. Le taux d'arrivée pour le service prioritaire (respectivement non prioritaire) est égal à λ_1 (respectivement : λ_2). Si un service prioritaire est demandé pendant qu'un service non prioritaire est en cours, la station cesse ce service et s'occupe immédiatement du service prioritaire. Les taux de service prioritaire (respectivement : non prioritaire) sont μ_1 (respectivement : μ_2).

On pose $p_{m,n} = P(m \text{ prioritaires}, n \text{ non prioritaires})$.

(1) Déterminer les équations d'équilibre pour les diverses valeurs de (m, n) .

(2) Déterminer la fonction génératrice $g(s_1, s_2) = \sum_{m \geq 0} \sum_{n \geq 0} s_1^m s_2^n p_{m,n}$, de la distribution de probabilité $(p_{m,n})_{m,n}$.

(3) Faire un développement limité à l'ordre deux de $g(s_1, s_2)$ au voisinage de $(1, 1)$, et en déduire les nombres moyens de clients prioritaires et non prioritaires dans la station.

Problème 19 ♠♠♠ Étude des systèmes d'attente à capacités bornées

(A) Système M/M/1/K

(1) Décrire la file comme une c.m.c. dont on construira le diagramme de transition et définira le générateur. On pose $\rho = \frac{\lambda}{\mu}$.

(2) Démontrer que la distribution stationnaire est définie par $\pi_n = \pi_0 \rho^n$

avec

$$\pi_0 = \left(1 + \frac{(1-\rho^K)\rho}{1-\rho}\right)^{-1} \text{ si } \rho \neq 1 \text{ et } \pi_0 = (1+K)^{-1} \text{ si } \rho = 1.$$

(3) Démontrer que la probabilité pour qu'un client soit rejeté, s'il y en a déjà K dans la file, vaut : $\frac{\rho^K - \rho^{K+1}}{1-\rho^{K+1}}$. En déduire que l'espérance d'attente d'un client est égale à :

$$\frac{L - K\pi_K}{\mu(1 - \pi_K)}.$$

(4) Déterminer la fonction génératrice $g_X(z)$ de la v.a. X de loi $(\pi_n)_n$. En déduire la valeur de $L = E(X)$ qui, on le rappelle, vaut $g'_X(1)$.

[$L = \frac{\rho(1-(K+1)\rho^K + K\rho^{K+1})}{(1-\rho)(1-\rho^{K+1})}$; si K tend vers $+\infty$, on retrouve la valeur de L pour $M/M/1$.]

(B) Système sans attente $M/M/K/K$: dans ce cas, où il n'y a pas de file d'attente, un client qui arrive est immédiatement servi ou bien il est rejeté.

(1) Démontrer que λ_n est égal à λ si $n \leq K$ et à 0 sinon et que μ_n est égal à $n\mu$ si $1 \leq n \leq K$ et à 0, si $n = 0$ ou est strictement supérieur à K . Construire le diagramme de transition et le générateur.

(2) Déterminer la distribution stationnaire et la probabilité de perte.

(3) Déterminer les caractéristiques du système.

(4) Application : soient s ordinateurs qui traitent des requêtes avec une durée moyenne de traitement égale à $\frac{1}{\mu}$; les requêtes arrivent selon un processus de Poisson tel que la durée moyenne séparant deux requêtes est égale à $\frac{1}{\lambda}$. Toute requête qui arrive lorsque les s ordinateurs sont occupés, est refusée. Supposons que $K = 3$, $\lambda = 10/h$ et $\mu = 8$. Déterminer la distribution stationnaire. Quel est le pourcentage de requêtes perdues ?

Problème 20 ♠♠

Un système est constitué de trois composants A , B et C de durées de vie aléatoires indépendantes X_1 , X_2 et X_3 , de lois exponentielles de paramètres λ_1 , λ_2 et λ_3 . On suppose qu'en $t = 0$ les trois composants fonctionnent. La panne du troisième composant ou la panne des deux autres entraînent la panne du système.

(1) Décrire les états du système. Démontrer que le processus $(Y_t)_{t \geq 0}$ associé à l'évolution du système est une c.m.c. ; déterminer son générateur infinitésimal.

(2) Résoudre les équations de Kolmogorov. [Calculer $p_1(t)$, puis, grâce à la transformation de Laplace, $p_2(t)$ et $p_3(t)$.] Si T désigne la durée de vie du système, on utilisera l'égalité : $E(T) = \int_0^{+\infty} P(T > t) dt$.

(3) Calculer la loi et l'espérance de la durée de vie du système.

Problème 21 ♠♠ *Modèle discret de Langevin*

Soit une particule évoluant sur un axe orienté; X_n sa position et V_n sa vitesse à l'instant $n.h$, h étant un pas de temps fixé. En prenant en compte la force de frottement qui s'oppose au déplacement de la particule mais en négligeant la pesanteur, Langevin proposa le modèle dynamique suivant :

$$V_{n+1} - V_n = -\lambda h V_n + Y_{n+1} \lambda \sqrt{2Dh}$$

où λ est le coefficient de frottement, et $(Y_n)_n$ un processus discret gaussien centré de variance σ^2 et de composantes décorrélées. Posons : $a = 1 - h.\lambda$ et $b = \lambda \sqrt{2Dh}$, h étant suffisamment petit pour que $a > 0$.

(1) Étant donné que $X_{n+1} - X_n = h.V_n$, $X_0 = 0$ et $V_0 = v$, déterminer les espérances, les variances et la covariance du vecteur gaussien (X_n, V_n) .

(2) Pour tout t positif, posons $t = nh$; déterminer les caractéristiques de (X_t, V_t) en considérant que n tend vers $+\infty$ et h tend vers 0, et les interpréter quand t devient grand. (voir Malrieu F., Vitesse et position d'une particule.)

Problème 22 ♠♠ *Temps de sortie d'un processus de Wiener unidimensionnel hors d'un intervalle borné*

Soit un processus de Wiener standard $(W_t)_t$, W_t est de loi $\mathcal{N}(0, t^2)$. Étant donné un réel c strictement positif, appelons $T_c = \inf\{t > 0, X_t \geq c\}$ le temps aléatoire de première atteinte de la valeur c .

- (1) Déterminer $P(W_t \geq c)$, puis en déduire F_{T_c} et f_{T_c} .
- (2) Démontrer que $P(T_c < +\infty) = 1$ et $E(T_c) = +\infty$.

11.2 Corrigés des problèmes

Problème 1

(1) Bijection entre l'ensemble des chemins joignant 0 à un point (m, n) et l'ensemble des $(m+n)$ -uplets $(x_1, x_2, \dots, x_{m+n})$ constitués de n éléments « h » et m éléments « d » : il y a donc $\binom{m+n}{m}$ chemins.

(2) Si on suppose que $m > n$, tout chemin allant de 0 à (m, n) se décompose en un chemin de 0 à (n, k) et en un chemin de (n, k) à (m, n) . On en déduit que :

$$\binom{m+n}{n} = \sum_{k=0}^n \binom{n+k}{n} \times \binom{m-n+n-k}{m-n} = \sum_{k=0}^n \binom{n+k}{n} \times \binom{m-k}{m-n}.$$

(On a noté : $\binom{j}{i} = C_j^i$)

Problème 2

(1) Un dessin montre clairement que lorsqu'il y a n droites, la dernière partage en deux parties chacune des n parties obtenues à l'aide de $n-1$ droites. Si P_n est le nombre de parties obtenues pour n droites, on a la relation de récurrence :

$$P_n = 2n + P_{n-1} - n = n + P_{n-1}.$$

Raisonnement analogue pour des plans.

(2) Soit $E_{n-1} = \{1, 2, \dots, n-1\}$: toute partition de l'ensemble E_n admet $\{n\}$ comme singleton ou comme élément d'un sous-ensemble plus grand.

Le nombre de partitions contenant k parties de E_n est donc égal au nombre de partitions contenant $(k-1)$ parties de E_n .

Si une partition de E_n ne contient pas $\{n\}$ comme singleton, on construit une partition contenant $(k-1)$ parties de E_{n-1} à k partitions différentes de E_n en ajoutant $\{n\}$ à chacune des parties de k éléments de cette partition.

Conclusion : $\sigma_{n+1}^k = \sigma_n^{k-1} + k\sigma_n^k$.

Problème 3

On notera A_k l'événement « le k -ième jeton tiré est gagnant ».

(1) a) $X(\Omega) = \mathbb{N}^*$ et $X = k$ signifie que la première série a pour longueur k , c'est-à-dire que les k premiers jetons sont gagnants et le $(k+1)$ -ième perdant ou bien les k premiers jetons sont perdants et le $(k+1)$ -ième gagnant. On a donc :

$[X = k] = (A_1 \cap \dots \cap A_k \cap \bar{A}_{k+1}) \cup (\bar{A}_1 \cap \dots \cap \bar{A}_k \cap A_{k+1})$ et $P([X = k]) = p^k q + q^k p$ si $k \in \mathbb{N}^*$.

$$b) \sum_{k=1}^{+\infty} kP([X=k]) = \sum_{k=1}^{+\infty} kp^k(1-p) + \sum_{k=1}^{+\infty} k(1-p)^k p. \quad g\mathbb{E}(X) = \frac{p}{1-p} + \frac{1-p}{p}.$$

$$\text{Soit } f(x) = \frac{x}{1-x} + \frac{1-x}{x} : f'(x) = \frac{1-x+x}{(1-x)^2} - \frac{1}{x^2} = \frac{x^2-(1-x)^2}{x^2(1-x)^2} = \frac{2x-1}{x^2(1-x)^2}.$$

f' est négative sur $]0, \frac{1}{2}[$, nulle en $\frac{1}{2}$ et positive sur $]\frac{1}{2}, 1[$, et $\mathbb{E}(X) = f(p)$. On a donc :

$\mathbb{E}(X)$ est minimal et vaut 2, lorsque $p = \frac{1}{2}$.

(2) a) $[X=i] \cap [Y=j]$ signifie que la première série est de longueur i et la deuxième de longueur j . On a donc, si $p_{ij} = P([X=i] \cap [Y=j])$:

$$p_{ij} = P[(A_1 \cap \dots \cap A_i \cap \bar{A}_{i+1} \cap \dots \cap \bar{A}_{i+j} \cap A_{i+j+1}) \cup (\bar{A}_1 \cap \dots \cap \bar{A}_i \cap A_{i+1} \cap \dots \cap A_{i+j} \cap \bar{A}_{i+j+1})]$$

et, toujours en utilisant l'indépendance des tirages, on obtient :

$$\text{pour tout } (i, j) \in (\mathbb{N}^*)^2, P([X=i] \cap [Y=j]) = p^{i+1}q^j + q^{i+1}p^j.$$

$$b) P([Y=j]) = \sum_{i=1}^{+\infty} P([X=i] \cap [Y=j]) = \sum_{i=1}^{+\infty} p^{i+1}q^j + \sum_{i=1}^{+\infty} q^{i+1}p^j \text{ avec } \sum_{i=1}^{+\infty} p^{i+1} = \sum_{i=2}^{+\infty} p^i = \frac{p^2}{q} \text{ et } \sum_{i=1}^{+\infty} q^{i+1} = \frac{q^2}{p}. \text{ On en déduit } P([Y=j]) = p^2q^{j-1} + q^2p^{j-1} \text{ pour } j \in \mathbb{N}^*.$$

$$\sum_{j=1}^{+\infty} jP([Y=j]) = \sum_{j=1}^{+\infty} j(p^2q^{j-1} + q^2p^{j-1}) = \frac{p^2}{q} \sum_{j=1}^{+\infty} jq^j + \frac{q^2}{p} \sum_{j=1}^{+\infty} jp^j \text{ soit } \mathbb{E}(Y) = 2.$$

(3) a) $P([X=1] \cap [Y=1]) = p^2q + pq^2 = pq(p+q) = pq$. D'autre part $P([X=1]) = 2pq$ et $P([Y=1]) = p^2 + q^2 = 2p^2 - 2p + 1$. On a donc :

$$P([X=1])P([Y=1]) - P([X=1] \cap [Y=1]) = pq(4p^2 - 4p + 2 - 1) = pq(2p-1)^2.$$

Si $p \neq \frac{1}{2}$, $P([X=1] \cap [Y=1]) \neq P([X=1])P([Y=1])$: X et Y ne sont pas indépendantes.

Si $p = q = \frac{1}{2}$, pour $(i, j) \in (\mathbb{N}^*)^2$, on a $P([X=i] \cap [Y=j]) = \frac{1}{2^{i+j}}$, $P([X=i]) = \frac{1}{2^i}$ et $P([Y=j]) = \frac{1}{2^j}$, donc $P([X=i] \cap [Y=j]) = P([X=i])P([Y=j])$ et les v.a. X et Y sont indépendantes.

$$b) \mathbb{E}(XY) = \sum_{i,j \geq 1} ij p_{ij} = \left(\sum_{i=1}^{+\infty} ip^{i+1} \right) \left(\sum_{j=1}^{+\infty} jq^j \right) + \left(\sum_{i=1}^{+\infty} ip^i \right) \left(\sum_{j=1}^{+\infty} jq^{j+1} \right) = \frac{p^2}{q^2} \frac{q}{p^2} + \frac{p}{q^2} \frac{q^2}{p^2} \text{ donc } \mathbb{E}(XY) = \frac{1}{q} + \frac{1}{p} = \frac{1}{pq} \text{ et } \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{1}{pq} - 2 \frac{p^2+q^2}{pq^2},$$

soit

$$\text{cov}(X, Y) = -\frac{(1-2p)^2}{p(1-p)}.$$

Problème 4

$$(1) P(M(x) \in [i, i+1]) = \log_{10}(i+1) - \log_{10}(i) = \log_{10}\left(1 + \frac{1}{i}\right).$$

$$(2.1) P(M(x) \in [a, b]) = P(\{\log_{10}(x)\} \in [\log_{10}(a), \log_{10}(b)]) ;$$

$$P(M(\alpha.x) \in [a, b]) = P(\{\log_{10}(x) + \log_{10}(\alpha)\} \in [\log_{10}(a), \log_{10}(b)])$$

égal à $P(\{\log_{10}(x)\} \in [\{\log_{10}(a) - \log_{10}(\alpha)\}, \{\log_{10}(b) - \log_{10}(\alpha)\}])$ les bornes de l'intervalle précédent sont à prendre modulo 1.

(2.2) Posons $s = \log_{10} a$, $t = \log_{10} b$, $v = -\log_{10} \alpha$.

L'équation $P(M(x) \in [a, b]) = P(M(\alpha.x) \in [a, b])$ s'écrit :

$$P(\{\log_{10} x\} \in [s, t]) = P([\{\log_{10} x\} \in [s + v, t + v]]),$$

où $s + v$ et $t + v$ sont définis modulo 1.

Ce résultat exprime que la proportion des x tels que $\{\log_{10}(x)\} \in [s, t[$ ne dépend que de l'intervalle $[s, t[$. On montre, en considérant successivement $[s, t[$ de bornes égales à des rationnels, puis des à des réels que :

$$P(\{\log_{10} x\} \in [s, t]) = t - s$$

et en réintroduisant la mantisse $M(x)$:

$$P(M(x) \in [a, b]) = \log_{10} b - \log_{10} a.$$

Problème 5

$$(1) X_k = aX_{k-1} + \varepsilon_k = a(aX_{k-2} + \varepsilon_{k-1}) + \varepsilon_k = a^2X_{k-2} + a\varepsilon_{k-1} + \varepsilon_k.$$

$$\text{Par récurrence : } X_k = \sum_{j=1}^k a^{k-j} \varepsilon_j.$$

Donc X_k est une combinaison de v.a. normales : elle est donc normale, d'espérance $\mathbb{E}(X_k) = \sum_{j=1}^k a^{k-j} \mathbb{E}(\varepsilon_j) = 0$ et de variance $V(X_k) = \sum_{j=1}^k (a^{k-j})^2 V(\varepsilon_j) = \sigma^2 \frac{1-a^{2k}}{1-a^2}$.

(2) $\rho_{X_{k+1}, X_k} = \frac{\text{cov}(X_{k+1}, X_k)}{\sigma_{X_{k+1}} \sigma_{X_k}}$; $\text{cov}(X_{k+1}, X_k) = \mathbb{E}((aX_k + \varepsilon_{k+1})X_k) = a\mathbb{E}(X_k^2) + \mathbb{E}(\varepsilon_{k+1}X_k) = aV(X_k) + \mathbb{E}(\varepsilon_{k+1} \cdot (\varepsilon_k + a\varepsilon_{k-1} + \dots + a^{k-1}\varepsilon_1)) = a\sigma_k^2$, car les $\mathbb{E}(\varepsilon_{k+1}\varepsilon_i)$ pour $i = 1, \dots, k$, sont nuls.

Conclusion :

- si $a = 1$, $\rho_{X_{k+1}, X_k} = \sqrt{\frac{k}{k+1}} \rightarrow 1$ quand $k \rightarrow +\infty$;
- si $a \neq 1$, $\rho_{X_{k+1}, X_k} = a\sqrt{\frac{1-a^{2k}}{1-a^{2k+2}}}$, qui tend vers 1 quand k tend vers $+\infty$, si $a > 1$.

(3) L'estimateur qui convient ici est l'estimateur des moindres carrés : il s'agira donc de minimiser la somme des carrés des erreurs $\sum_{i=1}^k \varepsilon_i^2 = \sum_{i=1}^k (x_i - ax_{i-1})^2$, qui est une fonction convexe de a . Il suffit de rechercher le point stationnaire a^* de la fonction précédente :

$$\left(\frac{\partial}{\partial a} \sum_{i=1}^k (x_i - ax_{i-1})^2 \right)_{a=a^*} = 0 \text{ donne } a^* = \frac{\left(\sum_{i=1}^k x_i x_{i-1} \right)}{\left(\sum_{i=1}^k x_{i-1}^2 \right)}.$$

Problème 6

(1) On va utiliser les coordonnées sphériques

$$\begin{cases} x_1 = r \sin \varphi \cos \theta \\ x_2 = r \sin \varphi \sin \theta \\ x_3 = r \cos \varphi \end{cases}$$

$$h : (x_1, x_2, x_3) \mapsto (r, \varphi, \theta); \quad h^{-1} : (r, \varphi, \theta) \mapsto (x_1, x_2, x_3).$$

$$\begin{aligned} J_{h^{-1}}(r, \varphi, \theta) &= \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \varphi} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \varphi} & \frac{\partial x_2}{\partial \theta} \\ \frac{\partial x_3}{\partial r} & \frac{\partial x_3}{\partial \varphi} & \frac{\partial x_3}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \sin \varphi \cos \theta & r \cos \varphi \cos \theta & -r \sin \varphi \sin \theta \\ \sin \varphi \sin \theta & r \cos \varphi \sin \theta & r \sin \varphi \cos \theta \\ \cos \varphi & -r \sin \varphi & 0 \end{vmatrix} \\ &= r^2 \cos^2 \varphi \cos^2 \theta \sin \varphi + r^2 \sin^3 \varphi \sin^2 \theta + r^2 \cos^2 \varphi \sin^2 \theta \sin \varphi + r^2 \sin^3 \varphi \cos^2 \theta \\ &= r^2 \cos^2 \varphi \sin \varphi + r^2 \sin^3 \varphi = r^2 \sin \varphi. \end{aligned}$$

h est une bijection de $\mathbb{R}^3 \setminus (Oz)$ sur $]0, +\infty[\times]0, \pi[\times]-\pi, \pi[$ sur lequel $J_{h^{-1}} \neq 0$.

$$\begin{aligned} f_{h(X_1, X_2, X_3)}(r, \varphi, \theta) &= f_{X_1, X_2, X_3}(h^{-1}(r, \varphi, \theta)) \times |J_{h^{-1}}(r, \varphi, \theta)| \mathbf{1}_{]0, +\infty[}(r) \mathbf{1}_{]0, \pi[}(\varphi) \mathbf{1}_{]-\pi, \pi[}(\theta) \\ &= f_{X_1, X_2, X_3}(h^{-1}(r, \varphi, \theta)) r^2 \sin \varphi \mathbf{1}_{]0, +\infty[}(r) \mathbf{1}_{]0, \pi[}(\varphi) \mathbf{1}_{]-\pi, \pi[}(\theta) \end{aligned}$$

$$\text{avec } f_{X_1, X_2, X_3}(x_1, x_2, x_3) = f_{X_1}(x_1) f_{X_2}(x_2) f_{X_3}(x_3) = \frac{1}{(\sqrt{2\pi})^3} e^{-\frac{1}{2}(x_1^2 + x_2^2 + x_3^2)}.$$

Or $x_1^2 + x_2^2 + x_3^2 = r^2$, donc :

$$f_{R, \Phi, \Theta}(r, \varphi, \theta) = \frac{1}{(\sqrt{2\pi})^3} e^{-\frac{r^2}{2}} \times r^2 \sin \varphi \mathbf{1}_{]0, +\infty[}(r) \mathbf{1}_{]0, \pi[}(\varphi) \mathbf{1}_{]-\pi, \pi[}(\theta) \quad (a)$$

$$f_R(r) = \iint f_{R, \Phi, \Theta}(r, \varphi, \theta) d\varphi d\theta = \frac{1}{(\sqrt{2\pi})^3} e^{-\frac{r^2}{2}} \times r^2 \mathbf{1}_{]0, +\infty[}(r) \times 2\pi [-\cos \varphi]_0^\pi,$$

soit :

$$f_R(r) = \sqrt{\frac{2}{\pi}} r^2 e^{-\frac{r^2}{2}} \mathbf{1}_{]0, +\infty[}(r), \text{ et par (a) :}$$

(2) $F_{R^2}(u) = P([R^2 \leq u])$. Si $u \leq 0$, $F_{R^2}(u) = 0$. Si $u > 0$:

$$F_{R^2}(u) = P([-\sqrt{u} \leq R \leq \sqrt{u}]) = F_R(\sqrt{u}) - F_R(-\sqrt{u}) = F_R(\sqrt{u}).$$

On en déduit :

$$f_{R^2}(u) = F'_{R^2}(u) = \frac{1}{2\sqrt{u}} f_R(\sqrt{u}) \mathbf{1}_{]0, +\infty[}(u) = \frac{1}{2} \sqrt{\frac{2}{\pi}} \times \frac{1}{\sqrt{u}} \times u e^{-\frac{u}{2}} \mathbf{1}_{]0, +\infty[}(u),$$

soit : $f_{R^2}(u) = \frac{1}{\sqrt{2\pi}} \sqrt{u} e^{-\frac{u}{2}} \mathbf{1}_{]0, +\infty[}(u)$. R^2 est donc de loi gamma $\gamma(\frac{1}{2}, \frac{3}{2})$, d'espérance $E(R^2) = 3$, de variance $V(R^2) = 6$.

$$(\Gamma(\frac{1}{2}) = \sqrt{\pi} \text{ donc } \Gamma(\frac{3}{2}) = \Gamma(\frac{1}{2} + 1) = \frac{1}{2}\Gamma(\frac{1}{2}) = \frac{\sqrt{\pi}}{2}.$$

Problème 7

(1) Envisageons le cas où $n = 2$: $X_{(1)} = \min(X_1, X_2)$ et $X_{(2)} = \max(X_1, X_2)$. L'événement $[X_{(1)} < x_1, X_{(2)} < x_2]$ est égal à l'union disjointe des événements : $[X_1 < X_2, X_1 < x_1, X_2 < x_2]$ et $[X_2 < X_1, X_2 < x_1, X_1 < x_2]$.

$$P([X_{(1)} < x_1, X_{(2)} < x_2]) = P([X_1 < X_2, X_1 < x_1, X_2 < x_2]) + P([X_2 < X_1, X_2 < x_1, X_1 < x_2]).$$

Ce résultat se généralise facilement par récurrence au cas de n v.a., donnant ainsi la formule indiquée.

(2) Dans le cas de deux v.a. :

$$P([X_1 < X_2, X_1 < x_1, X_2 < x_2]) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1}(y_1) f_{X_2}(y_2) \mathbf{1}_{(y_1 < y_2)}(y_1, y_2) dy_1 dy_2.$$

Ce résultat se généralise à n v.a. :

$$P([X_1 < \dots < X_n, X_1 < x_1, \dots, X_n < x_n]) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{X_1}(y_1) \dots f_{X_n}(y_n) \mathbf{1}_{(y_1 < \dots < y_n)}(y_1, \dots, y_n) dy_1 \dots dy_n$$

donc, compte tenu de l'identité des lois des X_i et de la première relation :

$$F_{(X_{(1)}, \dots, X_{(n)})}(x_1, \dots, x_n) = n! \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} \mathbf{1}_{(y_1 < \dots < y_n)}(y_1, \dots, y_n) dy_1 \dots dy_n$$

et

$$f_{(X_{(1)}, \dots, X_{(n)})}(x_1, \dots, x_n) = n! f_{X_1}(x_1) \dots f_{X_n}(x_n) \mathbf{1}_{(x_1 < \dots < x_n)}(x_1, \dots, x_n).$$

(3) L'événement $[X_{(k)} \leq x]$ est équivalent à $[k \text{ au moins des } x_i \text{ sont inférieurs à } x]$, lequel équivaut à $[N_x \geq k] = [N_x = k, N_x = k+1, \dots, N_x = n]$. Or N_x est la v.a. qui compte le nombre d'événements indépendants de la forme $[X_i \leq x]$, tous de même probabilité égale à $F(x)$: en conséquence, N_x est une v.a. binomiale, de paramètre n et $p = F(x)$.

$$\text{Il s'ensuit que } F_{X_{(k)}}(x) = P([N_x \geq k]) = \sum_{i=k}^n \binom{n}{i} (F(x))^i (1 - F(x))^{n-i}.$$

Supposons les X_i de densité uniforme sur $[0, 1]$:

$$\begin{aligned}
 f_{X_{(k)}}(x) &= \sum_{i=k}^n \binom{n}{i} (ix^{i-1}(1-x)^{n-i} - x^i(1-x)^{n-i-1}(n-i)) \\
 &= \sum_{i=k}^n \binom{n}{i} ix^{i-1}(1-x)^{n-i-1} - \sum_{i=k}^n n \binom{n}{i} x^i(1-x)^{n-i-1} \\
 &= n \left(\sum_{i=k-1}^{n-1} \binom{n-1}{i} x^i(1-x)^{n-i-1} - \sum_{i=k}^n \binom{n-1}{i} x^i(1-x)^{n-i-1} \right) \\
 &= n \binom{n-1}{k-1} x^k(1-x)^{n-k}.
 \end{aligned}$$

$E(X_{(k)}) = \frac{k}{n+1}$; pour n suffisamment grand, ce résultat est conforme à l'intuition, car $X_{(k)}$ se situe au voisinage de $\frac{k}{n+1}$.

(4) $f_{nX_{(k)}}(x) = \frac{1}{n} f_{X_{(k)}}\left(\frac{x}{n}\right)$ est approché grâce à la formule de Stirling par :

$$\frac{(n-1)^{n-1} \sqrt{(n-1)2\pi} e^{n-k} x^{k-1} \left(1 - \frac{x}{n}\right)^{n-k}}{n e^{n-1} (n-k)^{n-k} \sqrt{(n-k)2\pi} (k-1)! n^{k-1}},$$

qui converge lorsque n tend vers $+\infty$, vers $\frac{x^{k-1} e^{-x}}{(k-1)!}$, densité d'une loi gamma de paramètres $(1, k)$.

Plus la statistique d'ordre $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ est de grande taille, plus elle a tendance à s'équirépartir sur $[0, 1]$.

Application : $f_{X_{(1)}}(x) = n(1-F(x))^{n-1}f(x) = 10(1-x)^9$, $f_{X_{(2)}}(x) = 90x(1-x)^8$, $f_{X_{(3)}}(x) = 360x^2(1-x)^7, \dots$, $f_{X_{(10)}}(x) = (F(x))^{10} = 10x^9$.

Pour tout $k = 1, 2, \dots, 10$, $f_{X_{(k)}}(x)$ a un unique maximum en $x = \frac{k-1}{9}$, égal à :

$$M_k = \frac{10!}{(k-1)!(10-k)!} \left(\frac{k-1}{9}\right)^{k-1} \left(1 - \frac{k-1}{9}\right)^{10-k}.$$

Remarque : $M_k = M_{11-k}$.

Problème 8

(1) Posons : $U = X_1$, $V = X_2 - X_1$. Soit $h : (x_1, x_2) \mapsto (u, v) = (x_1, x_2 - x_1)$; h est une bijection de \mathbb{R}^2 sur lui-même, avec $h^{-1} : (u, v) \mapsto (x_1, x_2) = (u, u+v)$. $f_{U,V}(u, v) = f_{X_1, X_2}(h^{-1}(u, v)) |J_{h^{-1}}(u, v)| = f_{X_1, X_2}(u, u+v) = f_{X_1}(u) f_{X_2}(u+v)$, car X_1 et X_2 sont indépendantes. On a donc $f_{U,V}(u, v) = \lambda^2 e^{-\lambda(2u+v)} \mathbf{1}_{[0, +\infty[}(u) \mathbf{1}_{[0, +\infty[}(u+v)$ avec $\mathbf{1}_{[0, +\infty[}(u+v) = 1$ si $u+v > 0$, c'est-à-dire si $u > -v$.

On a $f_V(v) = \int f_{U,V}(u, v) du$ avec $f_{U,V}(u, v) = \lambda^2 e^{-2\lambda u} e^{-\lambda v} \mathbf{1}_{[\max(0, -v), +\infty[}(u)$.

$$\begin{aligned} \text{Ainsi, } f_{X_2-X_1}(v) &= \lambda^2 e^{-\lambda v} \left[-\frac{1}{2\lambda} e^{-2\lambda u} \right]_{\max(0, -v)}^{+\infty} \\ &= \frac{\lambda}{2} \left[e^{\lambda v} \mathbf{1}_{]-\infty, 0[}(v) + e^{-\lambda v} \mathbf{1}_{[0, +\infty[}(v) \right]. \end{aligned}$$

$Y = |V|$, donc $F_Y(x) = P([Y \leq x]) = P(|V| \leq x) = P([-x \leq V \leq x]) \mathbf{1}_{[0, +\infty[}(x)$ et $P([-x \leq V \leq x]) = F_V(x) - F_V(-x)$ donc $f_Y(x) = F'_Y(x) = (f_V(x) + f_V(-x)) \mathbf{1}_{[0, +\infty[}(x)$, avec $f_V(x) = \frac{\lambda}{2} e^{-\lambda|x|}$ (fonction paire) et $f_Y(x) = 2f_V(x) \mathbf{1}_{[0, +\infty[}(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0, +\infty[}(x)$.

Y a même loi $\mathcal{E}(\lambda)$ que les X_i .

(2) $P([X_3 \geq Y]) = P([Y - X_3 \leq 0]) = F_{Y-X_3}(0) = F_{X_2-X_1}(0)$, car X_1, X_2, X_3 étant indépendantes, il en est de même de X_3 et $Y = |X_2 - X_1|$ donc $Y - X_3$ a même loi que $X_2 - X_1$ et $F_{Y-X_3}(0) = \int_{-\infty}^0 f_{Y-X_3}(x) dx = \frac{\lambda}{2} \int_{-\infty}^0 e^{\lambda x} dx = \frac{1}{2} [e^{\lambda v}]_{-\infty}^0$, soit $P([X_3 \geq Y]) = \frac{1}{2}$.

(3) $P([Z \leq z]) = 1 - P([Z > z]) = 1 - P([\min(X, Y) > z]) = 1 - P([X > z] \cap [Y > z])$ avec $P([X > z] \cap [Y > z]) = P([X > z])P([Y > z])$, d'où $P([Z \leq z]) = 1 - (1 - F(z))^2$ et $f_Z(z) = 2(1 - F(z))f(z)$, soit $f_Z(z) = 2\lambda e^{-2\lambda z} \mathbf{1}_{[0, +\infty[}(z)$.
 Z et X_3 étant indépendantes, on a :

$$\begin{aligned} f_T(t) &= \int f_Z(z) f_{X_3}(t-z) dz = 2\lambda^2 \int e^{-2\lambda z} e^{-\lambda(t-z)} \mathbf{1}_{[0, +\infty[}(z) \mathbf{1}_{[0, +\infty[}(t-z) dt \\ &= 2\lambda^2 e^{-\lambda t} \int_0^t e^{-\lambda z} dz \mathbf{1}_{[0, +\infty[}(t) = 2\lambda e^{-\lambda t} (1 - e^{-\lambda t}) \mathbf{1}_{[0, +\infty[}(t) \end{aligned}$$

$\mathbb{E}(T) = \mathbb{E}(Z) + \mathbb{E}(X_3) = \frac{1}{2\lambda} + \frac{1}{\lambda}$ et, grâce à l'indépendance $V(T) = V(Z) + V(X_3) = \frac{1}{4\lambda^2} + \frac{1}{\lambda^2}$, soit finalement

$$\mathbb{E}(T) = \frac{3}{2\lambda} \text{ et } V(T) = \frac{5}{4\lambda^2}.$$

Problème 9

(1) Soit $h : (x, y) \mapsto (u, v) = (x, \frac{x+y}{x})$. C'est un C^1 difféomorphisme de $\mathbb{R}^* \times \mathbb{R}$ sur lui-même, avec $h^{-1}(u, v) = (x, y) = (u, uv - u)$. On a :

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \lambda^2 e^{-\lambda(x+y)} \mathbf{1}_{[0, +\infty[}(x) \mathbf{1}_{[0, +\infty[}(y),$$

car X et Y sont indépendantes et $f_{h(X,Y)}(u, v) = f_{X,Y}(h^{-1}(u, v)) |J_{h^{-1}}(u, v)|$ avec $J_{h^{-1}}(u, v) = \begin{vmatrix} 1 & 0 \\ v-1 & u \end{vmatrix} = u$ d'où $f_{U,V}(u, v) = \lambda^2 e^{-\lambda uv} |u| \mathbf{1}_{[0, +\infty[}(u) \mathbf{1}_{[0, +\infty[}(u(v-1))$ avec $u > 0$ et $u(v-1) > 0$ si et seulement si $u > 0$ et $v > 1$. On a donc bien $f_{U,V}(u, v) = \lambda^2 u e^{-\lambda uv} \mathbf{1}_{[0, +\infty[}(u) \mathbf{1}_{[1, +\infty[}(v)$. On en déduit $f_V(v) = \int f_{U,V}(u, v) du = \lambda^2 \int_0^{+\infty} u e^{-\lambda uv} du \mathbf{1}_{[1, +\infty[}(v)$ avec, si $t = \lambda uv$, $\int_0^{+\infty} u e^{-\lambda uv} du = \frac{1}{\lambda^2 v^2} \int_0^{+\infty} t e^{-t} dt = \frac{1}{\lambda^2 v^2}$. D'où $f_V(v) = \frac{1}{v^2} \mathbf{1}_{[1, +\infty[}(v)$ et comme $v f_V(v) \sim \frac{1}{v}$ au voisinage de $+\infty$, V n'admet pas d'espérance.

(2) Z représente la durée de vie d'un composant démarqué (égale à X s'il n'a pas de défaut et à $X/2$ sinon).

$$P([Z \leq z]) = P([TX \leq z]) = P\left(\left([TX \leq z] \cap [T = 1]\right) \cup \left([TX \leq z] \cap [T = \frac{1}{2}]\right)\right).$$

On a alors, en utilisant l'indépendance de X et T , $P([T = 1]) = 0,9$ et $P([T = \frac{1}{2}]) = 0,1$: $P([Z \leq z]) = P([X \leq z] \cap [T = 1]) \cup ([X \leq 2z] \cap [T = \frac{1}{2}]) = 0,9F_X(z) + 0,1F_X(2z)$, soit, avec $F_X(x) = (1 - e^{-\lambda x})\mathbf{1}_{[0, +\infty[}(x)$, d'où : $P([Z \leq z]) = (1 - 0,9e^{-\lambda z} - 0,1e^{-2\lambda z})\mathbf{1}_{[0, +\infty[}(z)$.

En dérivant, on obtient $f_Z(z) = F'_Z(z) = \lambda(0,9e^{-\lambda z} + 0,2e^{-2\lambda z})\mathbf{1}_{[0, +\infty[}(z)$. En remarquant que $P_Z = 0,9\mathcal{E}(\lambda) + 0,1\mathcal{E}(2\lambda)$, on obtient $\mathbb{E}(Z) = 0,9 \times \frac{1}{\lambda} + 0,1 \times \frac{1}{2\lambda}$, soit $\mathbb{E}(Z) = \frac{0,95}{\lambda}$.

Problème 10

$P_{\Theta_k} = \mathcal{U}([0, 2\pi[)$ donc $f_{\Theta_k}(\theta) = \frac{1}{2\pi}\mathbf{1}_{[0, 2\pi[}(\theta)$.

(1) $\mathbb{E}(\cos \Theta_k) = \int \cos \theta f_{\Theta_k}(\theta) d\theta = \frac{1}{2\pi} \int_0^{2\pi} \cos \theta d\theta = 0$. De même,

$$\mathbb{E}(\sin \Theta_k) = \int \sin \theta f_{\Theta_k}(\theta) d\theta = \frac{1}{2\pi} \int_0^{2\pi} \sin \theta d\theta = 0,$$

$$\mathbb{E}(\cos^2 \Theta_k) = \int \cos^2 \theta f_{\Theta_k}(\theta) d\theta = \frac{1}{2\pi} \int_0^{2\pi} \cos^2 \theta d\theta = \frac{1}{2\pi} \int_0^{2\pi} \frac{\cos 2\theta + 1}{2} d\theta = \frac{1}{2},$$

$$\text{var}(\cos \Theta_k) = \mathbb{E}(\cos^2 \Theta_k) - \mathbb{E}(\cos \Theta_k)^2 = \frac{1}{2},$$

$$\mathbb{E}(\sin^2 \Theta_k) = \mathbb{E}(1 - \cos^2 \Theta_k) = \frac{1}{2} \text{ et } \text{var}(\sin \Theta_k) = \frac{1}{2}.$$

$$\mathbb{E}(\cos \Theta_k) = \mathbb{E}(\sin \Theta_k) = 0, \text{ var}(\cos \Theta_k) = \text{var}(\sin \Theta_k) = \frac{1}{2}.$$

$X_0 = 0, X_1 = a \cos \Theta_1, X_2 = a \cos \Theta_1 + a \cos \Theta_2$, soit, par récurrence, $X_n = a \sum_{k=1}^n \cos \Theta_k$

et $Y_n = a \sum_{k=1}^n \sin \Theta_k$.

Par linéarité de l'espérance, on en déduit $\mathbb{E}(X) = \mathbb{E}(Y) = 0$.

$\text{var}(X_n) = a^2 \text{var}\left(\sum_{k=1}^n \cos \Theta_k\right) = a^2 \sum_{k=1}^n \text{var}(\cos \Theta_k) = a^2 \frac{n}{2}$, car les Θ_k sont indépendantes donc les $\cos \Theta_k$ aussi. De même, $\text{var}(Y_n) = a^2 \frac{n}{2}$, d'où $\text{var}(X_n) = \text{var}(Y_n) = \frac{a^2 n}{2}$.

(2) $X_n Y_n = \sum_{1 \leq i, j \leq n} a^2 \cos \Theta_i \sin \Theta_j$, avec :

- si $i \neq j$, $\mathbb{E}(\cos \Theta_i \sin \Theta_j) = \mathbb{E}(\cos \Theta_i) \mathbb{E}(\sin \Theta_j) = 0$;

• $\mathbb{E}(\cos \Theta_i \sin \Theta_i) = \mathbb{E}(\frac{1}{2} \sin 2\Theta_i) = \frac{1}{2\pi} \int_0^{2\pi} \sin \theta \cos \theta d\theta = \frac{1}{2\pi} \left[\frac{\sin^2 \theta}{2} \right]_0^{2\pi} = 0$
donc $\mathbb{E}(X_n Y_n) = 0$ et comme $\mathbb{E}(X_n) = 0$, $\text{cov}(X_n, Y_n) = 0$.

(3) On applique le T.C.L généralisé à la suite des vecteurs aléatoires :

$$(X_n, Y_n) = \left(\sum_{k=1}^n a \cos \Theta_k, \sum_{k=1}^n a \sin \Theta_k \right) :$$

$$P_{(X_n, Y_n)} \approx \mathcal{N} \left((\mathbb{E}(X_n), \mathbb{E}(Y_n)), \Gamma_{X_n, Y_n} = \begin{pmatrix} \text{var}(X_n) & \text{cov}(X_n, Y_n) \\ \text{cov}(X_n, Y_n) & \text{var}(Y_n) \end{pmatrix} \right),$$

donc ici $P_{(X_n, Y_n)} \approx \mathcal{N} \left((0, 0), \begin{pmatrix} \frac{na^2}{2} & 0 \\ 0 & \frac{na^2}{2} \end{pmatrix} \right)$ de densité

$$f_n : (x, y) \mapsto \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\frac{na^2}{2}}} e^{-\frac{x^2}{2\frac{na^2}{2}}} \times \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\frac{na^2}{2}}} e^{-\frac{y^2}{2\frac{na^2}{2}}} = \frac{1}{\pi na^2} e^{-\frac{(x^2+y^2)}{na^2}}.$$

On utilise les coordonnées polaires classiques. Soit $h : (x, y) \mapsto (r, t)$ la bijection de $\mathbb{R}^2 \setminus \{(0, 0)\}$ sur $]0, +\infty[\times]0, 2\pi[$, dont la fonction inverse est $h^{-1} :$

$$(r, t) \mapsto (x, y) = (r \cos t, r \sin t) : \text{d'où le jacobien } J_{h^{-1}}(r, t) = \begin{vmatrix} \cos t & -r \sin t \\ \sin t & r \cos t \end{vmatrix} = r.$$

$$f_{R_n, T_n}(r, t) = f_{X_n, Y_n}(r \cos t, r \sin t) \times r \mathbf{1}_{]0, +\infty[}(r) \mathbf{1}_{]0, 2\pi[}(t).$$

Soit $f_{R_n, T_n}(r, t) = \frac{1}{\pi na^2} e^{-\frac{r^2}{na^2}} r \mathbf{1}_{]0, +\infty[}(r) \mathbf{1}_{]0, 2\pi[}(t)$, et $f_{R_n}(r) = \int f_{R_n, T_n}(r, t) dt$.

$$f_{R_n}(r) = \frac{2}{na^2} r e^{-\frac{r^2}{na^2}} \mathbf{1}_{]0, +\infty[}(r).$$

$$f_{T_n}(t) = \int f_{R_n, T_n}(r, t) dr = \frac{1}{2\pi} \left[-e^{-\frac{r^2}{na^2}} \right]_0^{+\infty} \mathbf{1}_{]0, 2\pi[}(t), \text{ soit } f_{T_n}(t) = \frac{1}{2\pi} \mathbf{1}_{]0, 2\pi[}(t) :$$

$$P_{T_n} = \mathcal{U}(]0, 2\pi[).$$

$$\mathbb{E}(R_n) = \int r f_{R_n}(r) dr = \frac{2}{na^2} \int_0^{+\infty} r^2 e^{-\frac{r^2}{na^2}} dr = \frac{2}{na^2} \times \frac{1}{2} \int_{-\infty}^{+\infty} r^2 e^{-\frac{r^2}{na^2}} dr.$$

Or, si Z est une v.a.r. de loi $\mathcal{N}\left(0, \frac{na^2}{2}\right)$, on a :

$\text{var}(Z) = \frac{1}{\sqrt{2\pi} \sqrt{\frac{na^2}{2}}} \int_{-\infty}^{+\infty} z^2 e^{-\frac{z^2}{na^2}} dz = \frac{na^2}{2}$, d'où $\int_{-\infty}^{+\infty} z^2 e^{-\frac{z^2}{na^2}} dz = \frac{na^2}{2} \times a\sqrt{\pi n}$ et $\mathbb{E}(R_n) = \frac{a}{2} \sqrt{n\pi}$. La distance moyenne de la particule au temps n à l'origine dépend donc linéairement de $\sqrt{2}$.

$$\mathbb{E}(R_n^2) = \int r^2 f_{R_n}(r) dr = \frac{2}{na^2} \int_0^{+\infty} r^3 e^{-\frac{r^2}{na^2}} dr = \int_0^{+\infty} na^2 u e^{-u} du = na^2 \Gamma(2) = na^2$$

et $\mathbb{E}(X_n^2) = \text{var}(X_n) = \frac{na^2}{2}$ (on retrouve ainsi que $\mathbb{E}(R_n^2) = \mathbb{E}(X_n^2) + \mathbb{E}(Y_n^2) = na^2$).
On a alors $\text{var}(R_n) = na^2 - \frac{a^2}{4}n\pi$, soit $\text{var}(R_n) = na^2 \left(1 - \frac{\pi}{4}\right)$.

$$P([R_n > r]) = \int_r^{+\infty} f_{R_n}(u) du = \left[-e^{-\frac{u^2}{na^2}} \right]_r^{+\infty} \text{ d'où } f_{R_n}(r) = 2r.e^{-\frac{r^2}{na^2}}.$$

Problème 11

(1) $Z_k = 1$ si $f(X_k) \geq Y_k$, soit (X_k, Y_k) au-dessous de la courbe (dans A).

$$P([Z_k = 1]) = \frac{\text{aire de } A}{\text{aire de } D} = p,$$

$$\text{aire de } A = \int_a^b (f(x) - m) dx = \int_a^b f(x) dx - m(b-a).$$

$\mathbb{E}(Z_k) = p$ et $\text{var}(Z_k) = p(1-p)$, car $P_{Z_k} = \mathcal{B}(p)$.

$\mathbb{E}(\bar{Z}_n) = \frac{\sum_{i=1}^n \mathbb{E}(Z_i)}{n} = p$ et $\text{var}(\bar{Z}_n) = \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n Z_i\right) = \frac{1}{n} \text{var}(Z_i) = \frac{p(1-p)}{n}$, car les Z_i sont indépendantes. D'où, finalement :

$$\mathbb{E}(\bar{Z}_n) = p \text{ et } \text{var}(\bar{Z}_n) = \frac{1}{n} p(1-p).$$

(2) Par la loi des grands nombres, $\bar{Z}_n \xrightarrow{P} \mathbb{E}(Z_i) = p$.

(3) $P(|\bar{Z}_n - p| \leq 10^{-2}) \geq 0,95$ équivaut à $P(|\bar{Z}_n - p| > 10^{-2}) < 0,05 = 5.10^{-2}$.

$$\text{Or } P(|\bar{Z}_n - p| > 10^{-2}) \leq \frac{1}{10^{-4}} \mathbb{E}((\bar{Z}_n - p)^2) = \frac{\text{var}(\bar{Z}_n)}{10^{-4}} = \frac{p(1-p)}{10^{-4}n}.$$

On cherche n tel que $\frac{p(1-p)}{10^{-4}n} < 5.10^{-2}$, soit $n > \frac{p(1-p)}{5.10^{-6}}$.

Une approximation grossière de p permet de déterminer la taille n de l'échantillon : par exemple si $p \simeq \frac{1}{3}$, $n \geq \frac{1}{3} \frac{2}{3} \frac{1}{5.10^{-6}} \simeq 45.000$.

Si l'on connaît l'intégrale d'une fonction g sur $[a, b]$, g proche de f , alors il suffira d'estimer la surface comprise entre f et g .

Problème 12

(1) Par hypothèse, la densité $f_X(x_1)f_X(x_2)\cdots f_X(x_n)$ du vecteur (X_1, X_2, \dots, X_n) est symétrique : pour toute permutation σ de $\{1, 2, \dots, n\}$, on a :

$$P([X_1 < \cdots < X_n]) = P([X_{\sigma(1)} < \cdots < X_{\sigma(n)}])$$

où les événements $[X_{\sigma(1)} < \cdots < X_{\sigma(n)}]$ sont disjoints. Donc :

$$\sum_{\sigma} P([X_{\sigma(1)} < \cdots < X_{\sigma(n)}]) = n! P([X_1 < \cdots < X_n]) = 1$$

d'où, pour tout σ , $P([X_{\sigma(1)} < \dots < X_{\sigma(n)}]) = \frac{1}{n!}$.

(2) $[R_n = 1] = \bigcup_{\sigma: \sigma(n)=n} [X_{\sigma(1)} < \dots < X_{\sigma(n)}]$ réunion de $(n-1)!$ événements équiprobables.

On conclut que $P([R_n = 1]) = \frac{(n-1)!}{n!} = \frac{1}{n}$.

(3) $\mathbb{E}(S_n) = \sum_{i=1}^n \mathbb{E}(R_i) = \sum_{i=1}^n \frac{1}{i} = \ln n + \gamma + o(1)$ quand n grand, où γ est la constante d'Euler.

$$\mathbb{E}(T_n^p) = \mathbb{E}(S_{n+p}) - \mathbb{E}(S_n)$$

, donc :

$$\mathbb{E}(T_n^p) - (\ln(n+p) - \ln n) = \mathbb{E}(S_{n+p}) - \mathbb{E}(S_n) - \ln\left(1 + \frac{p}{n}\right) \xrightarrow{n \rightarrow +\infty} 0$$

d'où $\mathbb{E}(T_n^p) \xrightarrow{n \rightarrow +\infty} 0$.

(4) $\forall i$, $\mathbb{E}(R_i) = \frac{1}{i}$; $\mathbb{E}(R_m R_n) = P(R_m = 1, R_n = 1) = \frac{1}{m.n}$ d'où : $\text{cov}(R_m, R_n) = 0$.

$$(5) V(T_n^p) = \sum_{i=1}^p V(R_{n+i}) = \sum_{k=1}^p \frac{1}{n+k} - \sum_{k=1}^p \left(\frac{1}{n+k}\right)^2 \xrightarrow{n \rightarrow +\infty} 0.$$

$$(6) P([Y_n \leq y]) = (1 - e^{-y})^n \text{ d'où } f_{Y_n}(y) = n e^{-y} (1 - e^{-y})^{n-1}.$$

$\mathbb{E}(Y_n)$ se calcule en posant $u = (1 - e^{-y})$;

$$\mathbb{E}(Y_n) = \int_0^1 -n \ln(1-u) u^{n-1} du \approx \sum_{i=1}^n \frac{1}{i} \text{ (série harmonique).}$$

Sous l'hypothèse très restrictive d'invariance des lois des v.a. X_k , les records ont une croissance très lente; ce qui n'est évidemment pas le cas, par exemple, des records sportifs ou climatiques.

Problème 13

$$(1) \mathbb{E}^{[X_n=i]}(u^{Y_n}) = (pu + 1 - p)^i \text{ et } \varphi_{Z_n}(s) = e^{\lambda s - \lambda}.$$

La loi conditionnelle de Y_n sachant $X_n = i$ étant la loi binomiale $\mathcal{B}(i, p)$, on a :

$$\begin{aligned} \mathbb{E}^{[X_n=i]}(s^{X_n - Y_n}) &= s^i \mathbb{E}^{[X_n=i]}(s^{-Y_n}) = s^i \mathbb{E}^{[X_n=i]} \left(\left(\frac{1}{s} \right)^{Y_n} \right) \\ &= s^i \left(p \frac{1}{s} + 1 - p \right)^i = (p + (1-p)s)^i. \end{aligned}$$

On a alors : $\varphi_{n+1}(s) = \mathbb{E}(s^{X_{n+1}}) = \mathbb{E}(\mathbb{E}(s^{X_n - Y_n + Z_{n+1}}))$.

$\mathbb{E}^{[X_n=i]}(s^{X_n - Y_n + Z_{n+1}}) = (p + (1-p)s)^i \mathbb{E}(s^{Z_{n+1}})$, car Z_{n+1} est indépendante de X_n et de Y_n . Or Z_{n+1} suit la loi de Poisson $\mathcal{P}(\lambda)$ donc $\mathbb{E}(s^{Z_{n+1}}) = \exp(\lambda(s-1))$ d'où :

$$\mathbb{E}^{X_n}(s^{X_{n+1}}) = \exp(\lambda(s-1))(p + (1-p)s)^{X_n}.$$

Puis, en prenant l'espérance des deux membres :

$$\mathbb{E}(s^{X_{n+1}}) = \exp(\lambda(s-1))\mathbb{E}((p + (1-p)s)^{X_n}),$$

ce qui donne $\varphi_{n+1}(s) = \exp(\lambda(s-1))\varphi_n(p + (1-p)s)$.

(2) On pose $u = s - 1$. On a alors, d'après (1),

$$\varphi_{n+1}(u+1) = \exp(\lambda u)\varphi_n(p + (1-p)(u+1)) = \exp(\lambda u)\varphi_n((1-p)u+1)$$

soit $K_{n+1}(u) = \exp(\lambda u)K_n((1-p)u)$.

- Pour $n = 1$, $K_1(u) = \exp(\lambda u)K_0((1-p)u)$ et la propriété est vérifiée.
- Supposons la propriété vraie au rang n , alors :

$$\begin{aligned} K_{n+1}(u) &= \exp(\lambda u)K_n((1-p)u) \\ &= \exp(\lambda u)\exp\left(\frac{\lambda(1-p)u}{p}(1 - (1-p)^n)\right)K_0(u(1-p)^{n+1}) \\ &= \exp\left(\lambda u\left(1 + \frac{1-p}{p}(1 - (1-p)^n)\right)\right)K_0(u(1-p)^{n+1}) \\ &= \exp\left(\frac{\lambda u}{p}(1 - (1-p)^{n+1})\right)K_0(u(1-p)^{n+1}), \end{aligned}$$

et la propriété est vraie au rang $n+1$. On a donc :

$$K_n(u) = \exp\left(\frac{\lambda u}{p}(1 - (1-p)^n)\right)K_0(u(1-p)^n) \text{ pour tout } n \geq 1.$$

Comme $1-p \in]0, 1[$, $\lim_{n \rightarrow +\infty} (1-p)^n = 0$ et $\lim_{n \rightarrow +\infty} K_n(u) = \exp\left(\frac{\lambda u}{p}\right)K_0(0)$ et $K_0(0) = \varphi_0(1) = 1$: d'où $\lim_{n \rightarrow +\infty} \varphi_n(s) = \exp\left(\frac{\lambda}{p}(s-1)\right)$, car $\varphi_n(s) = K_n(s-1)$. On a donc $\lim_{n \rightarrow +\infty} P_{X_n} = \mathcal{P}\left(\frac{\lambda}{p}\right)$.

Problème 14

(1) X_{n+1} ne dépend que de X_n et de la somme $\sum_{l=1}^k \xi^l$, indépendantes des $(X_i)_{i=1, \dots, n}$: $(X_n)_n$ vérifie donc l'axiome de Markov.

D'autre part $P(X_{n+1} = j | X_n = k) = P(\sum_{l=1}^k \xi^l = j - k)$ est indépendante du temps n : l'axiome d'homogénéité est vérifié.

$$(2) P(X_n = k) = \sum_i P(\sum_{j=1}^i \xi_{n-1}^j = k) \cdot P(X_{n-1} = i)$$

$$\begin{aligned} g_n(s) &= \sum_k \sum_i s^k P(\sum_{j=1}^i \xi_{n-1}^j = k) \cdot P(X_{n-1} = i) \\ &= \sum_i P(X_{n-1} = i) \cdot g(s)^i = g_{n-1}(g(s)). \end{aligned}$$

D'où : $g_n(s) = (g \circ g \circ \dots \circ g)(s)$.

$$(3) E(X_n) = g'_n(1) = g'_{n-1}(g(1)) \cdot g'(1) = m \cdot g'_{n-1}(1).$$

Par récurrence, $E(X_n) = m^m$. On applique la même méthode pour le calcul de la variance. Si $m > 1$, la population croît ; si $m < 1$, elle décroît ; si $m = 1$, sa moyenne est constante et sa variance s'accroît.

Problème 15

(1) La chaîne ainsi obtenue est une chaîne à deux états $E = \{1, 2\}$, de matrice de transition $P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix}$.

C'est une chaîne irréductible finie, donc récurrente positive admettant une unique distribution stationnaire, solution de $\pi = \pi P$ avec $\pi_1 + \pi_2 = 1$. La première équation donne $\pi_1 = \frac{1}{2}\pi_1 + \frac{3}{4}\pi_2$, soit $\pi_1 = \frac{3}{2}\pi_2$. Puis $\frac{3}{2}\pi_2 + \pi_2 = 1 = \frac{5}{2}\pi_2$ donne $\pi_2 = \frac{2}{5}$ et $\pi_1 = \frac{3}{5}$. L'unique distribution stationnaire est donc $\pi = (\frac{3}{5}, \frac{2}{5})$. Pour calculer $p_{1,1}^{(3)}$, on peut faire le calcul direct de P^3 ou bien écrire

$$\begin{aligned} p_{11}^{(3)} &= p_{11}p_{11}p_{11} + p_{11}p_{12}p_{21} + p_{12}p_{21}p_{11} + p_{12}p_{22}p_{21} \\ &= \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^2 \frac{3}{4} + \frac{1}{2} \frac{3}{4} \frac{1}{2} + \frac{1}{2} \frac{1}{4} \frac{3}{4} = \frac{1}{8} + \frac{3}{16} + \frac{3}{16} + \frac{3}{32} = p_{1,1}^{(3)} = \frac{19}{32} = 0,593. \end{aligned}$$

Si l'abonnement d'avril est du type 1, on a 59,3 % des chances de l'avoir aussi en juillet.

(2) On sait que, pour une chaîne irréductible récurrente positive, le temps moyen de retour à l'état i est $\mu_i = \frac{1}{\pi_i}$ où π est l'unique distribution stationnaire. Cela donne ici $\mu_1 = \frac{5}{3}$ et $\mu_2 = \frac{5}{2}$.

$$\begin{aligned} P^{[X_0=1]}([T_1 = 1]) &= p_{1,1} = \frac{1}{2}, \quad P^{[X_0=1]}([T_1 = 2]) = p_{1,2}p_{2,1} = \frac{1}{2} \times \frac{3}{4} \text{ et, pour } n \geq 3, \\ P^{[X_0=1]}([T_1 = n]) &= p_{1,2}(p_{2,2})^{n-2}p_{2,1} = \frac{1}{2} \left(\frac{1}{4}\right)^{n-2} \frac{3}{4} \text{ d'où } G_{T_1}^{[X_0=1]}(s) = \frac{s}{2} + \frac{3}{8} \sum_{n=2}^{+\infty} \frac{s^n}{4^{n-2}} = \\ G_1(s), \text{ soit } G_1(s) &= \frac{s}{2} + \frac{3}{8}s^2 \frac{1}{1-\frac{s}{4}} \text{ et } G'_1(s) = \frac{1}{2} + \frac{3}{4}s \frac{1}{1-\frac{s}{4}} + \frac{3}{8}s^2 \frac{1}{(1-\frac{s}{4})^2} \frac{1}{4} \text{ et } G'_1(1) = \\ &= \frac{1}{2} + 1 + \frac{3}{8} \times \frac{4}{9}. \end{aligned}$$

On retrouve bien $\mu_1 = \mathbb{E}^{[X_0=1]}(T_1) = G'_1(1) = \frac{5}{3}$.

De même, $P^{[X_0=2]}([T_2 = 1]) = p_{2,2} = \frac{1}{4}$, $P^{[X_0=2]}([T_2 = 2]) = p_{2,1}p_{1,2} = \frac{3}{4} \times \frac{1}{2}$ et, pour $n \geq 3$, $P^{[X_0=2]}([T_2 = n]) = p_{2,1}(p_{1,1})^{n-2}p_{1,2} = \frac{3}{4} \left(\frac{1}{2}\right)^{n-2} \frac{1}{2}$ d'où $G_{T_2}^{[X_0=2]}(s) =$

$\frac{s}{4} + \frac{3}{4} \sum_{n=2}^{+\infty} \frac{s^n}{2^{n-1}} = G_2(s)$, soit $G_2(s) = \frac{s}{4} + \frac{3}{8}s^2 \frac{1}{1-\frac{s}{2}}$ et $G'_2(s) = \frac{1}{4} + \frac{3}{4}s \frac{1}{1-\frac{s}{2}} + \frac{3}{8}s^2 \frac{1}{(1-\frac{s}{2})^2} \frac{1}{2}$
 et $G'_2(1) = \frac{1}{4} + \frac{3}{2} + \frac{3}{4} = \frac{5}{2}$.
 On retrouve $\mu_2 = \mathbb{E}^{[X_0=2]}(T_2) = G'_2(1) = \frac{5}{2}$.

(3) On a $\mathbb{E}(C) = 100\pi_1 + 200\pi_2 = 60 + 80$ d'où un coût moyen mensuel de 140 euros.

(4) Si les changements sont payants, on a alors :

$$\begin{aligned} \mathbb{E}(C_n) &= 100 \times P([X_{n-1} = 1] \cap [X_n = 1]) + 150 \times P([X_{n-1} = 2] \cap [X_n = 1]) \\ &+ 200 \times P([X_{n-1} = 2] \cap [X_n = 2]) + 250 \times P([X_{n-1} = 1] \cap [X_n = 2]), \end{aligned}$$

avec $P([X_{n-1} = 1] \cap [X_n = 1]) = \frac{1}{2}\pi_1^{(n-1)}$, $P([X_{n-1} = 2] \cap [X_n = 1]) = \frac{3}{4}\pi_2^{(n-1)}$
 $P([X_{n-1} = 2] \cap [X_n = 2]) = \frac{1}{4}\pi_2^{(n-1)}$, $P([X_{n-1} = 1] \cap [X_n = 2]) = \frac{1}{2}\pi_1^{(n-1)}$
 où $(\pi^{(n-1)})$ désigne la loi de X_{n-1} . Par passage à la limite quand $n \rightarrow +\infty$, il vient :

$$\mathbb{E}(C) = 100 \times \frac{3}{10} + 150 \times \frac{3}{10} + 200 \times \frac{1}{10} + 250 \times \frac{3}{10} = 30 + 45 + 20 + 75 = 170.$$

Si les changements coûtent 50 euros, le coût moyen mensuel est donc de 170 euros.

Problème 16

(1) La valeur de X_n n'est fonction que de X_{n-1} et du résultat du n -ième match (c'est $X_{n-1} + 1$ si le n -ième match n'est pas perdu et c'est 0 si le n -ième match est perdu). Ainsi, l'axiome de Markov est vérifié. En outre, $P(X_n = k+1 \mid X_{n-1} = k) = p_k$ et $P(X_n = 0 \mid X_{n-1} = k) = 1 - p_k$ indépendants de n donc l'axiome d'homogénéité est aussi vérifié. On a donc bien une chaîne de Markov de matrice de transition $(p_{k,j})$:

$$p_{k,j} = \begin{cases} p_k & \text{si } j = k+1 \\ 1 - p_k & \text{si } j = 0 \\ 0 & \text{sinon} \end{cases}$$

$P^{[X_0=0]}([T_0 = 1]) = 1 - p_0$, $P^{[X_0=0]}([T_0 = 2]) = p_0(1 - p_1)$ et, pour $n \geq 3$, il faut aller jusqu'à $n-1$ matches sans défaites, puis perdre, d'où $P^{[X_0=0]}([T_0 = n]) = p_0 p_1 \cdots p_{n-2}(1 - p_{n-1})$.

L'état 0 est récurrent si et seulement si $P^{[X_0=0]}([T_0 < +\infty]) = 1$.

Or $P^{[X_0=0]}([T_0 < +\infty]) = P^{[X_0=0]} \left(\bigcup_{k=1}^{+\infty} [T_0 = k] \right) = \lim_{n \rightarrow +\infty} P^{[X_0=0]} \left(\bigcup_{k=1}^n [T_0 = k] \right)$ avec

$P^{[X_0=0]} \left(\bigcup_{k=1}^n [T_0 = k] \right) = (1 - p_0) + (p_0 - p_0 p_1) + \cdots - p_0 p_1 \cdots p_{n-1} = 1 - r_n$. $r_n = p_0 p_1 \cdots p_{n-1}$ est l'élément d'une suite décroissante de réels positifs, qui converge nécessairement vers une limite $r \geq 0$. L'état 0 est donc récurrent si et seulement

si $r = \lim_n p_0 p_1 \cdots p_{n-1} = 0$.

La chaîne étant irréductible, tous les états sont de même nature. Elle admet une distribution stationnaire si et seulement si elle est récurrente positive.

On cherche donc à résoudre $\pi = \pi P$ avec $P = \begin{pmatrix} 1-p_0 & p_0 & 0 & \cdots & \cdots \\ 1-p_1 & 0 & p_1 & \ddots & \\ 1-p_2 & 0 & 0 & p_2 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}$.

On obtient, à partir de la deuxième équation, $\pi_1 = p_0 \pi_0$, $\pi_2 = p_1 \pi_1, \dots, \pi_n = p_{n-1} \pi_{n-1}, \dots$, puis $\pi_2 = p_0 p_1 \pi_0, \dots, \pi_n = p_0 p_1 \cdots p_{n-1} \pi_0 = r_n \pi_0$ et on doit avoir $\pi_0 \left(1 + \sum_{n=1}^{+\infty} r_n\right) = 1$. On a donc une solution si et seulement si $(\sum r_n)$ converge, et cette solution est alors unique.

La chaîne est donc :

- transitoire si $\lim_n p_0 p_1 \cdots p_{n-1} \neq 0$;
- récurrente nulle si $\lim_n p_0 p_1 \cdots p_{n-1} = 0$ et $(\sum_n p_0 p_1 \cdots p_{n-1})$ diverge;
- récurrente positive si $(\sum_n p_0 p_1 \cdots p_{n-1})$ converge.

(2) Dans le cas où $p_n = p$ pour tout n , on a $r_n = p^n$, avec $p \in]0, 1[$, donc la chaîne est récurrente positive, de distribution stationnaire donnée par $\pi_n = p^n \pi_0$ et donc $\pi_0 \left(1 + \sum_{n=1}^{+\infty} p^n\right) = \frac{\pi_0}{1-p} = 1$, soit $\pi_0 = 1 - p$ et finalement $\pi_n = (1 - p)p^n$ pour tout n .

Le temps moyen de retour à l'état 0 est $\mu_0 = \frac{1}{\pi_0} = 2$ si $p = \frac{1}{2}$; lorsque le temps de retour à 0 est n , la longueur de la série est $n - 1$. Ainsi, si $p = \frac{1}{2}$, il y a en moyenne 1 match sans défaite.

Problème 17

(1) En utilisant successivement les propriétés d'un processus de Poisson (indépendance des accroissements, stationnarité, et N_t de loi $\mathcal{P}(\lambda t)$), il vient :

$$\begin{aligned} P(N_{1/4} = 1 \mid N_1 = 2) &= \frac{P(N_{1/4} = 1, N_1 = 2)}{P(N_1 = 2)} = \frac{P(N_{1/4} = 1, N_1 - N_{1/4} = 1)}{P(N_1 = 2)} \\ &= \frac{P(N_{1/4} = 1)P(N_1 - N_{1/4} = 1)}{P(N_1 = 2)} = \frac{P(N_{1/4} = 1)P(N_{3/4} = 1)}{P(N_1 = 2)} \\ &= \frac{e^{-\frac{\lambda}{4}} \frac{\lambda}{4} \times e^{-\frac{3\lambda}{4}} \frac{3\lambda}{4}}{e^{-\lambda} \frac{\lambda^2}{2!}} = \frac{3 \times 2}{4 \times 4} = \frac{3}{8}. \end{aligned}$$

(2) Pour $T = t$, le radar aura enregistré les véhicules de 0 à t et leur

nombre est exactement N_t . Sachant que $f_T(t) = \frac{1}{100} e^{-\frac{t}{100}} \mathbf{1}_{]0, +\infty[}(t)$, on a alors :

$$\begin{aligned} P([N = n]) &= \int P^{T=t}([N = n]) f_T(t) dt = \int P([N_t = n]) f_T(t) dt \\ &= \int_0^{+\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \frac{1}{100} e^{-\frac{t}{100}} dt = \frac{\lambda^n}{100n!} \int_0^{+\infty} t^n e^{-(\lambda + \frac{1}{100})t} dt \end{aligned}$$

et en posant $u = (\lambda + \frac{1}{100})t$, on a $\int_0^{+\infty} t^n e^{-(\lambda + \frac{1}{100})t} dt = \frac{\Gamma(n+1)}{(\lambda + \frac{1}{100})^{n+1}} = \frac{100^{n+1} n!}{(100\lambda + 1)^{n+1}}$ donc $P([N = n]) = \frac{(100\lambda)^n}{(100\lambda + 1)^{n+1}} = \frac{1}{100\lambda + 1} \left(\frac{100\lambda}{100\lambda + 1} \right)^n$ de la forme pq^n . Ainsi, avec $\lambda = 5$, N suit la loi géométrique $\mathcal{G}_0(\frac{1}{501})$, et $\mathbb{E}(N) = \frac{q}{p} = 500$.

Problème 18

Equations d'équilibre, en régime stationnaire :

- en $(0, 0)$: $(\lambda_1 + \lambda_2)p_0 = \mu_1 p_{10} + \mu_2 p_{01}$,
 - en $(1, 0)$: $(\lambda_1 + \lambda_2 + \mu_1)p_{10} = \mu_1 p_{20} + \lambda_1 p_0$,
 - en $(0, 1)$: $(\lambda_1 + \lambda_2 + \mu_2)p_{01} = \mu_2 p_{02} + \lambda_2 p_0$,
- et, pour $m + n \geq 2$:
- en $(m, 0)$: $(\lambda_1 + \lambda_2 + \mu_1)p_{m0} = \mu_1 p_{m+1,0} + \lambda_1 p_{m-1,0}$,
 - en $(0, n)$: $(\lambda_1 + \lambda_2 + \mu_2)p_{0n} = \mu_1 p_{1n} + \lambda_2 p_{0,n-1} + \mu_2 p_{0,n+1}$,
 - en $(m, n) \in \mathbb{N}^{*2}$: $(\lambda_1 + \lambda_2 + \mu_1)p_{mn} = \mu_1 p_{m+1,n} + \lambda_1 p_{m-1,n} + \lambda_2 p_{m,n-1}$,
 - pour $m \neq 0$: $(\lambda_1 + \lambda_2 + \mu_1)p_{mn} = \mu_1 p_{m+1,n} + \lambda_1 p_{m-1,n} + \lambda_2 p_{m,n-1}$,
 - pour $m = 0$: $(\lambda_1 + \lambda_2 + \mu_2(1 - \delta_{n0}))p_{0n} = \mu_1 p_{1n} + \lambda_2 p_{0,n-1} + \mu_2 p_{0,n+1}$.

En posant $g(s_1, s_2) = \sum_{m=0}^{+\infty} \sum_{n=0}^{+\infty} s_1^m s_2^n p_{mn}$, on a alors :

$$\begin{aligned} (\lambda_1 + \lambda_2 + \mu_1)g(s_1, s_2) - \mu_1 \sum_{n=0}^{+\infty} s_2^n p_{0,n} + \mu_2 \sum_{n=1}^{+\infty} s_2^n p_{0n} &= \frac{\mu_1}{s_1} \left(g(s_1, s_2) - \sum_{n=0}^{+\infty} s_2^n p_{0n} \right) \\ &+ (\lambda_1 s_1 + \lambda_2 s_2)g(s_1, s_2) + \frac{\mu_2}{s_2} \sum_{n=1}^{+\infty} s_2^n p_{0n}. \end{aligned}$$

En regroupant les $g(s_1, s_2)$ d'un côté et les $\sum s_2^n p_{0n}$ de l'autre, on obtient :

$$\begin{aligned} \left(\lambda_1(s_1 - 1) + \lambda_2(s_2 - 1) + \mu_1\left(\frac{1}{s_1} - 1\right) \right) g(s_1, s_2) &= \mu_1\left(\frac{1}{s_1} - 1\right) \left(\sum_{n=0}^{+\infty} p_{0n} s_2^n \right) \\ &- \mu_2\left(\frac{1}{s_2} - 1\right) \left(\sum_{n=1}^{+\infty} p_{0n} s_2^n \right). \end{aligned}$$

Pour faire le développement limité à l'ordre 2 au voisinage de $(1, 1)$, on pose $s_1 = 1 + h$ et $s_2 = 1 + k$, puis $m_1 = \frac{\partial g}{\partial s_1}(1, 1)$ et $m_2 = \frac{\partial g}{\partial s_2}(1, 1)$

$$(\lambda_1 h + \lambda_2 k + \mu_1((1+h)^{-1} - 1))g(1+h, 1+k) = \\ \mu_1((1+h)^{-1} - 1) \left(\sum_{n=0}^{+\infty} p_{0n}(1+k)^n \right) - \mu_2((1+k)^{-1} - 1) \left(\sum_{n=1}^{+\infty} p_{0n}(1+k)^n \right)$$

qui donne :

$$(\lambda_1 h + \lambda_2 k + \mu_1(-h+h^2))(G(1,1) + m_1 h + m_2 k) = \\ \mu_1(-h+h^2) \left(\sum_{n=0}^{+\infty} p_{0n} + k \sum_{n=0}^{+\infty} n p_{0n} \right) \\ - \mu_2(-k+k^2) \left(\sum_{n=1}^{+\infty} p_{0n} + k \sum_{n=1}^{+\infty} n p_{0n} \right) \\ + o(\|(h,k)\|^2).$$

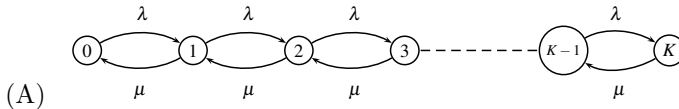
Par unicité du développement limité, on identifie alors les coefficients :

- terme en h : $\lambda_1 - \mu_1 = -\mu_1 \left(\sum_{n=0}^{+\infty} p_{0n} \right)$ avec $\rho_1 = \frac{\lambda_1}{\mu_1}$, on a alors $\sum_{n=0}^{+\infty} p_{0n} = 1 - \rho_1$;
- terme en k : $\lambda_2 = \mu_2 \left(\sum_{n=1}^{+\infty} p_{0n} \right)$ avec $\rho_2 = \frac{\lambda_2}{\mu_2}$, on a alors $\sum_{n=1}^{+\infty} p_{0n} = \rho_2$;
- terme en h^2 : $(\lambda_1 - \mu_1)m_1 + \mu_1 = \mu_1 \left(\sum_{n=0}^{+\infty} p_{0n} \right)$, soit $(\rho_1 - 1)m_1 = -\rho_1$, d'où $m_1 = \frac{\rho_1}{1-\rho_1}$;
- terme en k^2 : $\lambda_2 m_2 = \mu_2 \left(-\sum_{n=1}^{+\infty} p_{0n} + \sum_{n=1}^{+\infty} n p_{0n} \right)$, d'où $\sum_{n=1}^{+\infty} n p_{0n} = \rho_2(m_2 + 1)$;
- terme en hk : $(\lambda_1 - \mu_1)m_2 + \lambda_2 m_1 = -\mu_1 \left(\sum_{n=1}^{+\infty} n p_{0n} \right)$, d'où $\frac{\lambda_2}{\mu_1} m_1 = (1 - \rho_1)m_2 - \rho_2(m_2 + 1)$. On en déduit $(1 - \rho_1 - \rho_2)m_2 = \rho_2 + \rho_2 \frac{\mu_2}{\mu_1} m_1$.

Nombre moyen de clients prioritaires : $m_1 = \frac{\rho_1}{1-\rho_1}$.

Nombre moyen de clients non prioritaires : $m_2 = \frac{\rho_2}{1-\rho_1-\rho_2} \left(1 + \frac{\mu_2}{\mu_1} \frac{\rho_1}{1-\rho_1} \right)$ où l'on a posé $\rho_1 = \frac{\lambda_1}{\mu_1}$ et $\rho_2 = \frac{\lambda_2}{\mu_2}$.

Problème 19



(1) Équations bilan :
$$\begin{cases} \lambda \pi_0 = \mu \pi_1 \\ (\lambda + \mu) \pi_k = \mu \pi_{k+1} + \lambda \pi_{k-1} \text{ pour } 0 < k \leq K-1 \\ \mu \pi_K = \lambda \pi_{K-1} \end{cases}$$

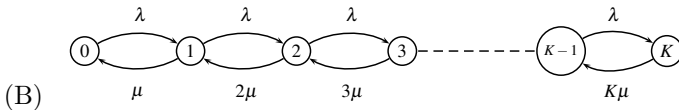
Par récurrence, on obtient $\lambda \pi_k = \mu \pi_{k+1}$ pour $0 \leq k \leq K-1$, soit $\pi_{k+1} = \rho \pi_k$, puis, $\pi_k = \rho^k \pi_0$ pour $0 \leq k \leq K$.

(2) On obtient π_0 en écrivant $\sum_{k=0}^K \pi_k = 1 = \pi_0 \frac{1-\rho^{K+1}}{1-\rho}$ si $\rho \neq 1$, soit $\pi_0 = \frac{1-\rho}{1-\rho^{K+1}}$ si $\rho \neq 1$ (et si $\rho = 1$, $\pi_0 = \frac{1}{K+1}$).

(3) La probabilité cherchée est $\pi_K = \rho^K \pi_0 = \frac{\rho^K - \rho^{K+1}}{1-\rho^{K+1}}$. Le temps moyen d'attente de quelqu'un qui arrive alors qu'il y a k clients est $\frac{k}{\mu}$ et cela se produit avec la probabilité $\frac{\pi_k}{1-\pi_K}$ pour $0 \leq k \leq K-1$. On a donc $W_q = \frac{\sum_{k=0}^{K-1} k}{\mu(1-\pi_K)} = \frac{L-K\pi_K}{\mu(1-\pi_K)}$.

C'est aussi, par la formule de Little, $\frac{L_q}{\lambda(1-\pi_K)}$ avec $L_q = \sum_{k=1}^K (k-1)\pi_k = L - (1-\pi_0)$.

(4) $G_X(z) = \sum_{k=0}^K z^k \pi_k = \pi_0 \frac{1-(\rho z)^{K+1}}{1-\rho z}$. Pour calculer L , on peut soit calculer $G'_X(z)$ pour $z = 1$, soit faire un développement limité à l'ordre 1 de $G_X(z)$ au voisinage de 1, soit utiliser les deux expressions de la question 3. On obtient alors sans problème $L = \frac{\rho(1-(K+1)\rho^K + K\rho^{K+1})}{(1-\rho)(1-\rho^{K+1})}$.



(1) Équations bilan :
$$\begin{cases} \lambda \pi_0 = \mu \pi_1 \\ (\lambda + k\mu) \pi_k = (k+1)\mu \pi_{k+1} + \lambda \pi_{k-1}, 0 < k \leq K-1 \\ K\mu \pi_K = \lambda \pi_{K-1} \end{cases}$$

La distribution stationnaire existe quelles que soient les valeurs de λ et μ . Notons que, si λ est plus grand que μ , alors π_K converge vers 1 : le système a tendance à être saturé.

(2) En posant $\rho = \frac{\lambda}{\mu}$, on a pour $k = 0, 1, \dots, K$, $\pi_k = \frac{\rho^k}{\sum_{i=0}^K \frac{\rho^i}{i!}}$. On peut remarquer que $(\pi_0, \pi_1, \dots, \pi_K)$ est une distribution de Poisson tronquée définie par :

$$\forall k, k \leq K, \pi_k = P(X = k | X \leq K) = \frac{\frac{\rho^k}{k!}}{\sum_{i=0}^K \frac{\rho^i}{i!}}.$$

La probabilité de perte d'un client est égale à la probabilité π_K que tous les serveurs soient occupés :

$$\pi_K = \frac{\frac{\rho^K}{K!}}{\sum_{i=0}^K \frac{\rho^i}{i!}}.$$

(3) Le taux d'entrée $\bar{\lambda}$ est égal à $\lambda(1-\pi_K)$.

$L = \bar{\lambda} W$ (formule de Little) ; $L_q = W_q = 0$.

$$(4) \pi_0 = \left(1 + \rho + \frac{\rho^2}{2} + \frac{\rho^3}{6}\right)^{-1} \text{ avec } \rho = \frac{10}{8} = 1,25 ;$$

$$\pi_0 = (1 + 1,25 + 0,78125 + 0,32552)^{-1}, \text{ soit } \pi_0 \approx 0,298.$$

Problème 20

(1) Le système possède quatre états :

e_1 : « les 3 composants fonctionnent » ; e_2 : « A est en panne, B et C fonctionnent » ; e_3 : « B en panne, A et C fonctionnent » ; e_4 : « le système est en panne ».

$$A = \begin{pmatrix} -(\lambda_1 + \lambda_2 + \lambda_3) & \lambda_1 & \lambda_2 & \lambda_3 \\ 0 & -(\lambda_2 + \lambda_3) & 0 & (\lambda_2 + \lambda_3) \\ 0 & 0 & -(\lambda_1 + \lambda_3) & (\lambda_1 + \lambda_3) \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Équation de Kolmogorov : $p'_1(t) = -(\lambda_1 + \lambda_2 + \lambda_3)p_1(t)$.

D'où $p_1(t) = \exp(-(\lambda_1 + \lambda_2 + \lambda_3)t)$, et $p'_2(t) = \lambda_1 p_1(t) - (\lambda_2 + \lambda_3)p_2(t)$.

Considérons la transformée de Laplace de l'équation différentielle précédente :

$$\mathcal{L}_2(s) = \frac{\lambda_1}{(s + \lambda_2 + \lambda_3)(s + \lambda_1 + \lambda_2 + \lambda_3)} = \frac{1}{s + \lambda_2 + \lambda_3} - \frac{1}{s + \lambda_1 + \lambda_2 + \lambda_3}$$

d'où : $p_2(t) = e^{-(\lambda_2 + \lambda_3)t} - e^{-(\lambda_1 + \lambda_2 + \lambda_3)t}$.

On obtient de la même façon : $p_3(t) = e^{-(\lambda_1 + \lambda_3)t} - e^{-(\lambda_1 + \lambda_2 + \lambda_3)t}$.

D'où, si T est la durée de vie : $P([T > t]) = p_1(t) + p_2(t) + p_3(t) = 1 - p_4(t)$

et

$$\mathbb{E}(T) = \int_0^{+\infty} P([T > t]) dt = \frac{1}{\lambda_1 + \lambda_3} + \frac{1}{\lambda_2 + \lambda_3} - \frac{1}{\lambda_1 + \lambda_2 + \lambda_3}.$$

Problème 21

(1) On suppose que h est tel que $1 > a > 0$.

$$V_{n+1} = aV_n + bY_{n-1}.$$

Entre les instants nh et $(n+1)h$, $X_{n+1} - X_n = hV_n$.

(X_n, V_n) est un vecteur gaussien (stabilité par somme de v.a. gaussiennes).

$$\mathbb{E}(X_n) = v \frac{1 - a^n}{\lambda} ; \quad \mathbb{E}(V_n) = v.a^n$$

$$V(X_n) = \frac{b^2 h^2}{(1 - a)^2} \left(n - 1 - \frac{2 + a}{1 - a^2} + \frac{2a^n}{1 - a} - \frac{a^{2n}}{1 - a^2} \right)$$

$$V(V_n) = \frac{b^2}{1-a^2}(1-a^{2n})$$

$$\text{cov}(X_n, V_n) = \frac{b^2 h}{(1-a)^2} \left(\frac{a}{1+a} - a^n + \frac{a^{2n}}{1+a} \right).$$

(2) Le couple (X_n, V_n) converge en loi vers le vecteur gaussien (X_t, V_t)

$$\mathbb{E}(X_t) = v \frac{1-e^{-\lambda t}}{\lambda}$$

$$\mathbb{E}(V_t) = v.e^{-\lambda t}$$

$$\text{cov}(X_t, V_t) = \begin{pmatrix} 2Dt + \frac{D}{\lambda}(-3 + 4e^{-\lambda t} - e^{-2\lambda t}) & D(1-e^{-\lambda t}) \\ D(1-e^{-\lambda t}) & \lambda D(1-e^{-2\lambda t}) \end{pmatrix}.$$

Quand $t \rightarrow +\infty$, $\mathbb{E}(V_t) \rightarrow 0$, $V(V_t) \rightarrow \lambda D$ et $\lim_{t \rightarrow +\infty} (V_t)$ est de loi $\mathcal{N}(0, \lambda D)$.

Problème 22

(1) $P(W_t \geq c)$ se réduit à $P(W_t \geq c \mid T_c \leq t).P(T_c \leq t) = \frac{1}{2}$

$$P(T_c \leq t) = 2.P(W_t \geq c) = \frac{2}{\sqrt{2.\pi}} \int_{\frac{c}{\sqrt{t}}}^{+\infty} e^{-\frac{x^2}{2}} .dx$$

$$f_{T_c}(t) = \frac{c}{\sqrt{2.\pi}} t^{-\frac{3}{2}} e^{-\frac{c^2}{2t}} (\forall t > 0).$$

(2) $P(T_c < +\infty) = 1$

$$E(T_c) = \int_0^{+\infty} P(T_c > t) dt = \frac{2a^2}{\sqrt{2\pi}} \int_0^{+\infty} \frac{1}{x^2} e^{-\frac{x^2}{2}} .dx$$

$$\text{qui est supérieur à } \frac{2a^2.e^{-\frac{1}{2}}}{\sqrt{2.\pi}} \int_0^{+\infty} \frac{1}{x^2} dx = +\infty.$$

Annexes

Bibliographie

Théorie des probabilités

- BARBE PH., LEDOUX M. (2007) *Probabilité*, Belin
BON P.L., *Fiabilité des systèmes* (1995) Masson
BOUZITAT C., PAGES G., *En passant par hasard : les probabilités de tous les jours* (1999) Vuibert
BOULEAU N., *Probabilités de l'ingénieur* (2002) Hermann
BREIMAN L., *Probability* (1992) Addison-Wesley
COTRELL M. et al, *Exercices de probabilité* (1999) Cassini
COVER T.M., THOMAS J.A., *Elements of information theory* (2006) Wiley
DAUXOIS J., HASSENFORDER C., *Toutes les probabilités et les statistiques* (2004) Ellipses
DELMAS J.P., *Introduction aux probabilités* (1993) Ellipses
FELLER W., *An introduction to probability theory and its application* (1970) Wiley
FOATA D., FUCHS A., *Calcul des probabilités* (2003) Dunod
GIRARDIN V., LIMNIOS N., *Probabilités* (2001) Vuibert
NEVEU J., *Bases mathématiques du calcul de probabilités* (1970) Masson
OUVRARD J.Y., *Probabilités* (1998) Cassini
TOULOUSE P.S., *Thèmes de probabilités et statistiques* (2000) Dunod
TUFFIN B., *La simulation de Monte-Carlo* (2010) Hermès Science-Lavoisier

Processus aléatoires

- BAYNAT B., *Théorie des files d'attente* (2000) Hermes
BERCU B., CHAFAI D., *Modélisation stochastique et simulation* (2007) Dunod
BOULEAU N., *Processus stochastiques et applications* (2000) Hermann
BREMAUD D., *Markov chains* (1998) Springer
CHARBIT M., *Éléments de théorie du signal : aspects aléatoires* (1996) Ellipses
CHUNG K.L., AITSAHLIA F., *Elementary probability theory* (2003) Springer
COCOZZA-THIVENT C., *Processus stochastiques et fiabilité des systèmes* (1997) Springer
DELMAS J.F., JOURDAIN B., *Modèles aléatoires* (2006) Springer
FOATA D., FUCHS A., *Processus stochastiques* (2002) Dunod
GRIMMETT G.G., STIRZAKER D.R., *Probability and random process* (1994) Oxford Science Publication
KLEINROCK L., *Queueing systems*, vol. I et II (1975) Wiley and sons
LEFEBVRE M., *Processus stochastiques appliqués* (2005) Hermann
OKSENDAL B., *Stochastic differential equations* (2003) Springer
PARDOUX E., *Processus de Markov et applications* (2007) Dunod
YCARD B., *Modèles et algorithmes markoviens* (2002) Springer

Analyse fonctionnelle, mesure et intégration

- CAUMEL Y., *Cours d'analyse fonctionnelle et complexe* (2003) Cépaduès
GASQUET C., WITOMSKI P., *Analyse de Fourier et applications* (1996) Dunod
SAMUELIDES M., TOUZILLIER L., *Analyse fonctionnelle* (1989) Cépaduès

Articles

DUPLANTIER B., *Le Mouvement brownien, « divers et ondoyant »* (2005) Séminaire Poincaré I

JANVRESSE E., *La loi de Benford*, Images mathématiques

<http://images.math.cnrs.fr/>

KAHANE J.P., *Le Mouvement brownien*, (1998) Société mathématique de France

MALRIEU F., *Vitesse et position d'une particule* (2009) Préparation à l'agrégation Rennes-I

MANSUY R., *Des dés diaboliques* (2009) Quadrature, n° 74

Quelques résultats particulièrement intéressants

• Répartition des nombres premiers	11
• Problème de Buffon	54
• Construction de la loi de Gauss	38
• Caractérisation géométrique du vecteur gaussien	64
• Le paradoxe du prisonnier	116
• Caractérisation entropique des lois de probabilités	137
• Loi arcsinus ou le paradoxe du jeu de pile ou face	163
• Paradoxe de la station de bus	184
• L'étrange loi de Benford	262
• Marche aléatoire isotropique	266
• Calcul d'intégrale par la méthode de Monte-Carlo	267
• Évolution des suites de records	267

Intermèdes historiques et bibliographiques

• Andreï Kolmogorov	5
• Émile Borel	6
• L'origine de la notion de probabilité	9
• Systèmes générateurs de hasard	11
• Mathématisation du hasard au XVII ^e siècle	28
• Carl Friedrich Gauss	35
• Siméon Denis Poisson	73
• De Jacques Bernoulli à Gauss	94
• Paul Lévy	97
• Les méthodes de Monte-Carlo	100
• Pierre-Simon de Laplace	103
• Le hasard en physique	107
• Essor de la théorie des processus aléatoires	172
• L'entropie en physique statistique	137
• Suites aléatoires	141
• Modélisation du mouvement brownien	251

Théorèmes importants

• Propriétés d'une probabilité	4
• Probabilités totales	5
• Propriétés de l'espérance	24
• Propriétés de la variance	25
• Inégalités classiques	33
• Propriétés de l'espérance et de la covariance d'un vecteur aléatoire	51
• Indépendance des composantes d'un vecteur aléatoire	56
• Propriétés du coefficient de corrélation linéaire	56
• Fonction caractéristique d'une somme de v.a. indépendantes	59
• Probabilité image	73
• Calcul de loi par difféomorphisme	74
• Identification fonctionnelle	75
• Stabilité de la normalité par somme indépendante	76
• Théorème de Lévy	96
• Loi faible des grands nombres	99
• Théorème limite centrale	100
• Formules bayésiennes	115
• Propriétés de l'espérance conditionnelle	128
• Conditionnement gaussien	134
• Chapman-Kolmogorov (chaînes discrètes)	155
• Caractérisation des états récurrents	158
• Chacon-Ornstein	168
• Wiener-Khintchine	238
• Propriétés du processus de Wiener	244
• Équation de Chapman-Kolmogorov (c.m.c.)	204
• Équation d'équilibre (c.m.c.)	211
• Équations de diffusion	248
• Flux dans un réseau ouvert	223

Formulaire mathématique

$$n! \simeq_{(n \text{ grand})} \left(\frac{n}{e}\right)^n \sqrt{2\pi n}.$$

$$\text{Pour tout } p \text{ et } q > 1 \quad \int_0^1 x^{p-1}(1-x)^{q-1} dx = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \equiv_{(\text{noté})} B(p, q).$$

$$\text{Pour tout } p \geq 1 : \int_0^{+\infty} x^{p-1} e^{-x} dx \equiv_{(\text{noté})} \Gamma(p) \text{ égal à } (p-1)! \text{ si } p \text{ entier.}$$

$$\text{Pour tout réel } a \neq 0 : \int_0^{+\infty} e^{-a^2 x^2} dx = \frac{\sqrt{\pi}}{2|a|}$$

$$\sum_{i=1}^n i = \frac{n(n+1)}{2} \equiv S_1$$

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6} \equiv S_2$$

$$\sum_{i=1}^n i^3 = \frac{n^2(n+1)^2}{4} \equiv S_3 = S_1^2.$$

Les symboles \mathcal{F} et \mathcal{L} (ou L) désignent les transformations de Fourier et de Laplace appliquées à des fonctions dont les transformées correspondantes sont supposées exister.

$$\mathcal{F}(f)(v) = \int_{-\infty}^{+\infty} f(t) \exp(-2i\pi vt) dt$$

$$\mathcal{F}\left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right)\right)(v) = \exp(-2\pi^2 v^2 \sigma^2)$$

$$\mathcal{F}\left(\exp\left(-\frac{t}{T}\right) \mathbf{1}_{\mathbb{R}_+}(t)\right)(v) = \frac{T}{1+2i\pi vT}.$$

$$\text{Pour tout } a > 0, \quad \mathcal{F}\left(f\left(\frac{t}{a}\right)\right)(v) = a\mathcal{F}(f)(av)$$

$$\mathcal{F}(f \star g) = \mathcal{F}(f) \cdot \mathcal{F}(g)$$

$$\mathcal{L}(f)(s) = \int_0^{+\infty} f(t) e^{-st} dt.$$

$$\text{Pour tout } t > a > 0, \quad \mathcal{L}(f(t-a))(s) = \exp(-pa) \mathcal{L}(f)(s)$$

$$a > 0 ; \quad \mathcal{L}\left(-\frac{t^2}{2a^2}\right)(s) = \sqrt{\frac{\pi}{2}} a \exp\left(\frac{a^2 s^2}{2}\right) \operatorname{erfc}\left(\frac{as}{\sqrt{2}}\right)$$

$$\mathcal{L}(f \star g) = \mathcal{L}(f) \cdot \mathcal{L}(g).$$

Index

Symbols

Événements, 2

A

Absence de mémoire, 31
Absorbant (état), 155
Apériodique (état), 164
Arrangement, 8
Autocorrélation, 235

B

Benford (loi), 262
Bernoulli, 55
Borel, 1
Box-Muller (méthode de), 82
Brown, 242
Bruit blanc, 150
Buffon (problème de), 54

C

Cardano, 9
Chaîne de Markov, 149
 ergodique, 168
 homogène, 152
 induite (ou incluse), 213
 irréductible, 158
 récurrente, 160
Chacon-Ornstein (théorème), 168
Changement de variables (méthode),
 74
Chapman-Kolmogorov (équation), 155

Codage, 135
Coefficient
 de corrélation, 56
 multinomial, 9
Combinaison, 8
 avec répétition, 8
Combinatoire, 7
Complémentaire, 2
Conditionnements successifs (théorème),
 116
Convergence, 93
 en loi, 93
 en moyenne quadratique, 93
 en probabilité, 93
 presque sûre, 93
Convolution, 54
Covariance, 50
 d'un vecteur aléatoire, 50
 de deux v.a., 56
Cycle, 164

D

Densité, 29
 conditionnelle, 119
 conjointe, 61
 de probabilité, 29
 marginale, 52
 spectrale, 150
diagramme, 21
Difféomorphisme, 74
Doob, 172
Droite de régression, 130

E

Ecart-type, 25
Einstein, 108
Ellipse de dispersion, 63
Entropie, 135
Equation
 d'équilibre, 213
 de flux, 223
Erlang, 203
Espérance conditionnelle, 117
Espace

- fondamental, 2
 - probabilisé, 6
 - produit, 20
- Événements
 - certain, impossibles, 2
 - disjoints, incompatibles, 2
 - indépendants, 6
 - mutuellement indépendants, 7
 - rare (hypothèse des), 181
- F**
- Fiabilité, 78, 124
- Files d'attente, 150
- Fonction
 - caractéristique, 58
 - d'erreur, 37
 - de répartition, 21
 - de renouvellement, 191
 - de survie, 124
 - génératrice, 25, 57
 - mesurable, 52
- Formule
 - de Bayes, 115
 - de Stirling, 9
 - des erreurs relatives, 57
- G**
- Gauss, 35
- Gaussien
 - vecteur, 60
 - vecteur (dégénéré ou non), 62
- Gestion de stocks, 169
- Grandes déviations, 102
- I**
- Identité de Poincaré, 4
- Inégalité
 - de Bienaymé-Tchebycheff, 267
 - de Cauchy-Schwarz, 56
 - de Jensen, 33
 - de Markov, 33
- Indépendance, 6
 - conditionnelle, 115
 - mutuelle, 7
- Indicatrice, 30
- Information, 135
- Intensité (d'une loi de Poisson), 26
- J**
- Jackson (réseaux), 222
- K**
- Kendall (notation), 215
- Kolmogorov, vii
- L**
- Lévy (théorème), 96
- La Vallée-Poussin (théorème), 12
- Langevin, 252
- Laplace, 95
- Limite Centrale (théorème), 34
- Little (formule), 217
- Loi
 - Arcsinus, 297
 - binomiale, 21
 - conjointe, 46
 - de Bernoulli, 58
 - de Cauchy, 32
 - de Fisher-Snedecor, 83
 - de Gauss, 34, 95
 - de Pareto, 37
 - de Poisson, 26
 - de probabilité, 46
 - de Student, 84
 - de Weibull, 125
 - exponentielle, 30
 - faible des grands nombres, 99
 - forte des grands nombres, 104
 - géométrique, 27
 - gamma, 78
 - gaussienne, 94
 - initiale (d'une chaîne), 155
 - limite, 168
 - marginale, 49
 - multinomiale, 184
 - normale, 34

stationnaire, 165
uniforme, 30

M

M/G/1, 221
M/M/1, 217
M/M/s/K, 219
Méthode
 d'identification fonctionnelle, 75
Marche aléatoire, 266
Matrice
 de covariance, 50
 de transition, 154
Mesure absolument continue, 29
Moivre-Laplace (théorème), 97
Moment
 d'ordre n , 59
 d'une v.a., 59
Monte-Carlo, 81
Multisenseur, 48

N

Nombres premiers (répartition), 11

P

Paradoxe des stations de bus, 184
Paradoxe du prisonnier, 116
Paramètres de performance, 217
Passe-dix (jeu du), 10
Presque sûr (événement), 5
Probabilité, 3
 conditionnelle, 113
 de transition, 152
 produit, 20
Probabilités composées, 95
Processus
 aléatoires, 149
 autorégressif, 264
 d'Ornstein-Uhlenbeck, 254
 de branchement, 268
 de comptage, 179
 de croissance, 214
 de décroissance, 214

de Ferry-Yule, 214
de Poisson, 179
de Poisson décalé, 186
de Poisson non homogènes, 187
de Poisson spatiaux, 188
de renouvellement, 191
du second ordre, 93
ergodique, 241
gaussien, 240

Produit de convolution, 76

R

Réalisation (d'un processus), 152
Récurent (état), 158
 nul, 159
 positif, 159
Régression
 linéaire, 130
 non linéaire, 131
Réseau
 de files d'attente, 222
 de Gordon-Newell, 227
 de Jackson, 222
 de Jakson ouvert, 222
Règle des trois sigmas, 33
Radon-Nykodim (théorème), 29

S

Shannon, 135
Signal du téléphone, 149
Simulations de lois, 81
Stabilisé
 pour la somme de v.a., 72
Stationnaire (distribution), 165
Support d'une probabilité, 30
Système
 complet d'événements, 115
 en parallèle, 79
 en série, 79

T

Taux d'utilisation, 217
Taux de défaillance instantané, 124

Taux de défaillance moyen, 124
Taux de transition instantané, 205
Transformation
 de Fourier, 58
Transition
 graphe, 153
 probabilité, 151
Transitoire (état), 158
Tribu, 3
 borélienne, 3
 produit, 20

U

Urne d'Ehrenfest, 153
Usure, 31

V

v.a.
 absolument continue, 29
 discrète, 21
Variables mixtes, 80
Variance, 24
 d'une v.a., 25
 totale (théorème), 133
Vecteur aléatoire, 46

W

Wiener-Khintchine (théorème), 238